

High-Order Methods for CFD

R. Abgrall¹ and Mario Ricchiuto²

¹*Institute of Mathematics, University of Zurich, Zurich, Switzerland*

²*Team CARDAMOM, INRIA Bordeaux–Sud-Ouest, Talence, France*

1	Introduction	1
2	A Review of Existing Methods	2
3	A Different Setting: Residual Distribution	8
4	Applications	27
5	Conclusion, Open Challenges	47
	Acknowledgments	48
	Notes	49
	References	49

1 INTRODUCTION

In this chapter we discuss, at some length, high-order methods for advection-dominated problems. Typical examples are the advection diffusion equation (with large Peclet number), the Euler equations, the Navier–Stokes equations, the shallow-water equations, and many problems in geophysical flow, to mention just a few. The list is indeed very long.

In these kinds of problems, one has to deal with contrasting constraints. First, the solution must be accurate. Wherever the solution is smooth, the truncation error must scale as h^{r+1} , where h is the typical size of the mesh elements, and r is an integer. However, it is also well known that, in many cases, the solution is not globally smooth and that it may admit very large local gradients. For example, these may be shock waves

(or slip lines) for the Euler equations, and boundary layers for the Navier–Stokes equations at very high Reynolds number. Other effects may come into play, as, for example, dispersion, as in some geophysical and environmental applications (wave propagation, capillary flows).

In this chapter, we address these issues and give several examples of successful modern high-order methods. The literature on this topic has exploded since the mid-1990s, and it is not possible to give an exhaustive coverage of all that has been achieved, so we had to make choices, which, of course, are biased by our own work.

We will start by reviewing, to the extent possible, the two most popular methods today: WENO (weighted essentially non-oscillatory) finite volume schemes, and the discontinuous Galerkin (DG) methods. We will give some details and, in particular, indicate the main principles. However, we will not discuss all the issues related to these methods. In particular, we set aside the choice of the approximation of the viscous terms: this can be easily found in the many papers that have appeared on the topics or in monographs such as Abgrall and Shu (2016) and di Pietro and Ern (2012). Our main focus is on a less successful approach known as the *residual distribution method*. As also discussed in the companion chapter by Deconinck and Ricchiuto, these schemes share a lot with continuous finite element methods, such as the streamline diffusion method, but also embed properties typical of the finite volume method: a lot of emphasis is put on L^∞ stability constraints, allowing the avoidance of spurious oscillations at discontinuities. In this chapter, however, we extend this analysis, showing how the residual-based philosophy underlying these schemes provides a framework that allows us to embed most (or all) other arbitrary order methods, and work with them under a different light, thus providing more insight into

these methods, and perhaps new alternative constructions. Of course, it also provides a setting to construct different arbitrary order schemes, and we will review the main challenges encountered when doing that, as well as some of the solutions proposed so far to overcome these challenges.

This chapter is organized into four sections. The first one gives a review of the WENO and DG methods. Section 2 develops in detail the residual distribution method, from a historical perspective, and its most recent achievements. Section 3 provides several applications, both for compressible flows and aerodynamics, and for some geophysical flows. A conclusion follows. We hope that the bibliography is rich enough to cover and complete all the topics we have mentioned in the text.

2 A REVIEW OF EXISTING METHODS

We start by considering the following problem:

$$\frac{\partial \mathbf{w}}{\partial t} + \operatorname{div} \mathbf{f}(\mathbf{w}) - \operatorname{div} \mathbf{f}_v(\mathbf{w}, \nabla \mathbf{w}) = 0 \quad (1a)$$

defined in $\Omega \subset \mathbb{R}^d$ with $d = 1, 2, 3$, and $\mathbf{w} : \Omega \times \mathbb{R}^+ \rightarrow \mathcal{D} \subset \mathbb{R}^m$. We need to set up an initial condition

$$\mathbf{w}(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \Omega \quad (1b)$$

and boundary conditions on $\partial\Omega$. In (1a), the flux $\mathbf{f} = (\mathbf{f}_1, \dots, \mathbf{f}_d)$ is assumed to be defined on \mathcal{D} and to be smooth enough. The viscous flux \mathbf{f}_v is assumed to satisfy a similar hypothesis (however, see below).

Here, we are mostly interested in fluid mechanics problems where the Navier–Stokes equation is the canonical example. In that case, the state variable \mathbf{w} needs to satisfy additional constraints: the density and the internal energy need to be positive. This is what is meant by saying that $\mathbf{w} \in \mathcal{D}$.

In this section, we review some of the high-order methods that have been developed in recent years. The research activity in this field has been very intense during the last years, so it is impossible to exhaustively review all that has been done. We had to make choices, and these choices are biased.

The second choice we have made, at least for this section, is to deal with a simplified problem. Instead of (1a), we will deal with

$$\frac{\partial \mathbf{w}}{\partial t} + \operatorname{div} \mathbf{f}(\mathbf{w}) = 0 \quad (2)$$

with the initial condition (1b) and relevant boundary conditions. In fluid mechanics, the simplest system of this kind is the Euler system. There, the state variable is

$$\mathbf{w} = (\rho, \rho \mathbf{u}, E)^T$$

where ρ is the density, \mathbf{u} is the fluid velocity, and E is the total energy, that is, the sum of the internal energy e and the kinetic energy. The flux is given by

$$\mathbf{f} = (\rho \mathbf{u}, \rho \mathbf{u} \otimes \mathbf{u} + p \operatorname{Id}_d, (E + p)\mathbf{u})^T$$

To close the system, we need to define the pressure $p = p(\rho, e)$. General assumptions on this function can be found in Godlewski and Raviart (1996); a typical example is that of perfect gas:

$$p = (\gamma - 1)e$$

It is well known that only weak solutions of (2) can be considered because there is no hope to have regular solutions in general. Hence, we need to consider weak solutions, that is, measurable functions in $L^\infty(\Omega \times \mathbb{R}^+)^m \cap L^1(\Omega \times \mathbb{R}^+)^m$ such that, for any compactly supported regular test function $\varphi \in C_0^1(\Omega \times \mathbb{R}^+)^m$, we have

$$\begin{aligned} \int_{\Omega \times \mathbb{R}^+} \frac{\partial \varphi}{\partial t}(\mathbf{x}, t) \cdot \mathbf{w}(\mathbf{x}, t) d\mathbf{x} dt + \int_{\Omega \times \mathbb{R}^+} \nabla \varphi(\mathbf{x}, t) \cdot \mathbf{f}(\mathbf{w}) d\mathbf{x} dt \\ - \int_{\Omega} \varphi(\mathbf{x}, 0) \cdot \mathbf{w}_0(\mathbf{x}) d\mathbf{x} = 0 \end{aligned} \quad (3)$$

Of course, the initial condition needs to be also in $L^\infty(\Omega)^m \cap L^1(\Omega)^m$. Note we have not taken into account the boundary conditions. This is a complex problem (for a rigorous treatment and also from a practical point of view, we refer to Dubois and Le Floch, 1988) for systems.

It is also well known that the definition of a weak solution is not enough. Even in the scalar case, this does not guarantee the uniqueness of the solution and some selection mechanism is needed. For the system case, the solution is much more complex. To this end, one classically considers an entropy, that is, a strictly convex function S defined on \mathcal{D} such that there exists an entropy flux \mathbf{G} such that

$$\nabla_u \mathbf{G} = \nabla_w S \cdot \nabla_w \mathbf{f}$$

An entropy solution should satisfy, in the sense of distribution, the following inequality:

$$\frac{\partial S}{\partial t} + \operatorname{div} \mathbf{G} \leq 0 \quad (4)$$

In the case of Euler equations, the (mathematical) entropy is given by $S = -\rho s$, where s is the (physical) entropy. The entropy flux is $\mathbf{G} = S\mathbf{u}$. The reader may consult (Harten, 1983; Hughes *et al.*, 1986) for further details.

Equation (3) is the origin of all possible forms of numerical approximation of the system (1). The first thing to do is to approximate the domain Ω . For the sake of simplicity, we assume here that Ω is polygonal. Then we discretize Ω

using meshes. In the vocabulary of unstructured meshes, we consider a tessellation \mathcal{T}_h . The domain Ω is

$$\Omega = \cup_{K \in \mathcal{T}_h} K$$

As usual, we assume that the elements K are nonoverlapping. The elements K will be triangles or quadrangles in two dimensions, tetrahedrons, hexahedrons, pyramids, and so on, in three dimensions, or may have more complicated forms. All depend on the choice made for approximating the solution and the choice of test functions φ .

- *Finite volume schemes*: One considers that \mathbf{w} is constant in each cell,

$$\mathbf{w}_K(t) \approx \frac{1}{|K|} \int_K \mathbf{w}(\mathbf{x}, t) d\mathbf{x}$$

and the test functions are also constant. One can nevertheless get high-order accuracy of the averaged value. This is the topic of Section 2.1.1.

- *Continuous finite elements*: Here we assume a globally continuous approximation \mathbf{w}_h of \mathbf{w} . Typically, for any element, $\mathbf{w}_{h|K}$ is a polynomial of degree k . Because of the continuity requirement, this imposes constraints on the mesh: the intersection of two elements is empty or reduced to a (complete) face, or they are identical. The mesh is said to be conformal. The elements need also, in general, be simplices because of the polynomial approximation. These methods are sketched in Section 2.1.3.
- *dG methods*: Here, the continuity requirement is dropped. This allows a lot of freedom: the mesh need not be conformal and the element can be general, so that mesh refinement becomes simple in principle. These methods are sketched in Section 2.1.2.

In the rest of this section, we first consider the spatial approximation (hence using a semidiscrete formulation), and then we discuss briefly the temporal approximation.

2.1 Space discretizations

2.1.1 ENO and WENO

Here, we consider the finite volume formulation of (3). The states are described by $\{\mathbf{w}_K(t)\}_{K \in \mathcal{T}_h}$. Starting from (3), we get

$$\frac{d}{dt} \int_K \mathbf{w}(\mathbf{x}, t) d\mathbf{x} + \int_{\partial K} \mathbf{f}(\mathbf{w}) \cdot \mathbf{n} d\mathbf{x} = 0$$

Here, \mathbf{n} is the outward unit normal to the boundary ∂K of K . This can be obtained from (3) by first regularizing via mollification φ_ε , the characteristic function of K , and taking the limit when $\varepsilon \rightarrow 0$. We see that $\nabla \varphi_\varepsilon \rightarrow \mathbf{n}$.

Since K is polygonal, denoting a generic face/edge of K by e , we see that an approximation of (3) is

$$\frac{d}{dt} \mathbf{w}_K(t) + \frac{1}{|K|} \sum_{e \in \partial K} \int_e \mathbf{f}(\mathbf{w}_h) \cdot \mathbf{n} de = 0 \quad (5)$$

This relation has not yet a meaning since \mathbf{u}_h is discontinuous across edges. In the normal direction to e , we need to solve the following Riemann problem:

$$\frac{\partial \mathbf{w}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{w}) \cdot \mathbf{n}}{\partial \mathbf{n}} = 0$$

with the initial condition

$$\mathbf{w}(\mathbf{x}, 0) = \begin{cases} \mathbf{w}_K, & \text{if } \mathbf{x} \cdot \mathbf{n} < 0 \\ \mathbf{w}_{K^+}, & \text{else} \end{cases}$$

Here, \mathbf{u}_{K^+} is the state on the other side of e . This problem can be solved either exactly or in an approximate way. A meaning of the edge integral is given thanks to the use of numerical flux $\hat{\mathbf{f}}(\mathbf{u}_K, \mathbf{u}_{K^+})$ (see Godlewski and Raviart, 1996; LeVeque, 2002; Toro, 1997) for an extensive discussion about numerical flux and Riemann solvers). Hence, the finite volume method in its simplest form is

$$\frac{d}{dt} \mathbf{w}_K(t) + \frac{1}{|K|} \sum_{e \in \partial K} \int_e \hat{\mathbf{f}}(\mathbf{w}_{h|K}, \mathbf{w}_{h|K^+}, \mathbf{n}) de = 0 \quad (6)$$

Note that the edge integrals are evaluated via a quadrature formula.

With this, only first-order accuracy can be achieved. Formal high-order accuracy can be obtained by using the MUSCL (monotonic upstream-centered scheme for conservation laws) method due to van Leer (1979). The idea is to consider a polynomial reconstruction of degree p , $\mathcal{R}(u_h)$, within each cell K and to replace (6) by

$$\frac{d}{dt} \mathbf{w}_K(t) + \frac{1}{|K|} \sum_{e \in \partial K} \int_e \hat{\mathbf{f}}(\mathcal{R}(\mathbf{w}_h)_K, \mathcal{R}(\mathbf{w}_h)_{K^+}, \mathbf{n}) de = 0 \quad (7)$$

and the quadrature formula need to be of sufficient order, typically exact for polynomials of degree p .

The design of a reconstruction operator is a research field by itself, since one wants to avoid the Gibbs phenomena where the solution becomes steep or discontinuous. After the seminal work of van Leer, a very large literature has been devoted to this problem. A large part of it is about total variation diminishing schemes (see Sweby's paper; Sweby, 1984), but it is rather difficult to reach higher than second-order accuracy (see Chakravarthy and Osher, 1985 for an attempt in one dimension), and it can be shown that a TVD (total variation diminishing) scheme in more than one dimension,

even in the scalar case, is at most first-order accurate, see Goodman and LeVeque (1985). This negative result has motivated researchers to look for criteria that are less strict than the TVD one, and the most successful method is probably the essentially non-oscillatory method, originally due to Harten and Osher (1987), Harten *et al.* (1987) and then refined by Shu and Osher (1988, 1989). Extension to unstructured meshes can be found in Abgrall (1994). Better stability properties are obtained by the so-called WENO technique (see Liu *et al.*, 1994 for the original reference), which was further refined by Shu and coworkers (see Shu, 2009 for a review).

The principle can be explained by assuming a regular mesh in one dimension. Extension to two dimensions and to more general meshes can be found in Friedrich (1998), Hu and Shu (1999), Zhu *et al.* (2008), for example. Taking a mesh $\{x_j\}_{j \in \mathbb{Z}}$, with $x_j = j\Delta x$, we first define four approximations of a smooth function at $x_{i+1/2} = \frac{x_i + x_{i+1}}{2}$:

- Using the stencil $S^{(1)} = \{x_{i-2}, x_{i-1}, x_i\}$, we have $u_{i+1/2}^{(1)} = \frac{3}{8}u_{i-2} - \frac{5}{4}u_{i-1} + \frac{15}{8}u_i = u(x_{i+1/2}) + O(\Delta x^3)$;
- Using the stencil with $S^{(2)} = \{x_{i-1}, x_i, x_{i+1}\}$, we get $u_{i+1/2}^{(2)} = -\frac{1}{8}u_{i-1} + \frac{3}{4}u_i + \frac{3}{8}u_{i+1} = u(x_{i+1/2}) + O(\Delta x^3)$;
- With the stencil, with $S^{(3)} = \{x_i, x_{i+1}, x_{i+2}\}$, we have $u_{i+1/2}^{(3)} = \frac{3}{8}u_i + \frac{3}{4}u_{i+1} - \frac{1}{8}u_{i+2} + O(\Delta x^3)$;
- Lastly, with $S = \{x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}\}$, we obtain

$$u_{i+1/2}^{(4)} = \frac{3}{128}u_{i-2} - \frac{5}{32}u_{i-1} + \frac{45}{64}u_i + \frac{15}{32}u_{i+1} - \frac{5}{128}u_{i+2} = u(x_{i+1/2}) + O(\Delta x^5)$$

Then we notice that $u_{i+1/2}^{(4)} = \gamma_1 u_{i+1/2}^{(1)} + \gamma_2 u_{i+1/2}^{(2)} + \gamma_3 u_{i+1/2}^{(3)}$, with $\gamma_1 = \frac{1}{16}$, $\gamma_2 = \frac{5}{8}$, and $\gamma_3 = \frac{5}{16}$. Note that $\gamma_1 + \gamma_2 + \gamma_3 = 1$. This enables us to approximate $u(x_{i+1/2})$ by $u_{i+1/2} = w_1 u_{i+1/2}^{(1)} + w_2 u_{i+1/2}^{(2)} + w_3 u_{i+1/2}^{(3)}$ with $w_j \approx \gamma_j$ and $w_j \approx 0$ if a discontinuity exists in $S^{(j)}$. In order to achieve this, we define $w_j = \frac{\tilde{w}_j}{\tilde{w}_1 + \tilde{w}_2 + \tilde{w}_3}$ with $\tilde{w}_j = \frac{\gamma_j}{(\epsilon + \beta_j)^2}$, and the smoothening indicator β_j is

$$\beta_j = \sum_{l=1}^2 \Delta x^{2l-1} \int_{x_{i-1/2}}^{x_{i+1/2}} \left(\frac{d^l}{dx^l} p_j(x) \right)^2 dx$$

Compared to the method we are going to discuss now, for a given formal accuracy, they clearly need the lowest possible storage. The price to pay for this is that the computational stencil becomes quite large. In the case of an irregular mesh, many precautions need to be taken in order to effectively reach the formal accuracy. In the case of an unstructured mesh, the extension is possible but very technical.

2.1.2 Discontinuous Galerkin methods

2.1.2.1 Formulation and basic properties

This class of methods was originally designed by Reed and Hill (1973). The first analysis was done by Lesaint and Raviart (1974) and further refined by Johnson and Pitkäranta (1986). The references Chavent and Cockburn (1989) and mostly Cockburn and Shu (1989a, 1991) and their sequel paved the way to the success of DG methods for hyperbolic problems and the Navier–Stokes equations (among many other applications). Reference Cockburn *et al.* (2000) represents the state of the art in the early 2000, whereas (di Pietro and Ern, 2012) is a more mathematical presentation of the theory, which also contains much information on how to approximate parabolic problems in that framework. It is impossible to give a complete survey of this field because the number of papers has grown exponentially. Again, we will sketch the method for purely hyperbolic problems, and refer to the reader to the reference section of di Pietro and Ern (2012) for more information on how to approximate the second-order terms.

Again, we start from (3). We consider a tessellation of the computational domain Ω , like in the finite volume method, but here we look for solutions that are polynomial of degree $r \geq 0$ in each element. More precisely, we want to approximate the solution in V_h defined by

$$V_h = \{\mathbf{w}_h \in (L^\infty(\Omega))^m \cap (L^1(\Omega))^m, \text{ for any element } K, (\mathbf{w}_h)|_K \in (\mathbb{P}^r(K))^m\}$$

As shown on Figure 1, no continuity requirement is needed, and, moreover, the degree r may depend on the element. Then we apply the weak formulation, taking as test function any $\varphi \in V_h$: for any element K

$$\int_K \frac{\partial \varphi}{\partial t} \cdot \mathbf{w}_h(\mathbf{x}, t) d\mathbf{x} - \int_K \nabla \varphi \cdot \mathbf{f}(\mathbf{w}_h(\mathbf{x}, t)) d\mathbf{x} + \int_{\partial K} \varphi \cdot (\mathbf{f}(\mathbf{w}_h(\mathbf{x}, t)) \cdot \mathbf{n}) dK = 0$$

However, as for the finite volume method, this formulation is meaningless because \mathbf{u}_h appearing in the boundary integral is in general multivalued, so the flux term cannot be given a meaning. The idea contained in Cockburn and Shu (1989a, 1991) is, as for the finite volume method, to introduce a numerical flux $\hat{\mathbf{f}}$. The DG (semidiscrete) formulation is, therefore, find $\mathbf{w}_h \in V_h$ such that for any K and any $\varphi_h \in V_h$

$$\int_K \frac{\partial \varphi}{\partial t} \cdot \mathbf{w}_h(\mathbf{x}, t) d\mathbf{x} - \int_K \nabla \varphi \cdot \mathbf{f}(\mathbf{w}_h(\mathbf{x}, t)) d\mathbf{x} + \int_{\partial K} \varphi \cdot \hat{\mathbf{f}}(\mathbf{w}_{h|K}, \mathbf{w}_{h|K^+}, \mathbf{n}) dK = 0 \quad (8)$$

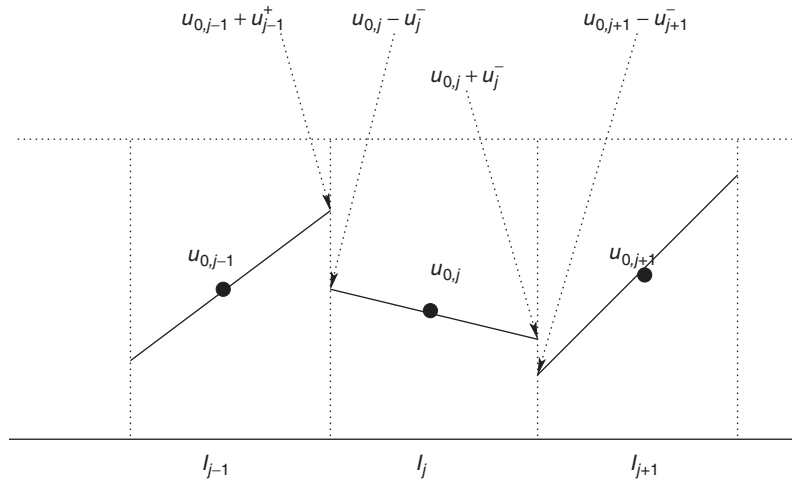


Figure 1. Geometrical representation of a \mathbb{P}^1 approximation. In the element K_j , $u_h = u_{0,j} + \delta u_j(x - x_j)$. x_j is the centroid of K_j , and δ_j is its length. We have set $\delta u_j^\pm = u_h(x_j \pm \frac{\Delta_j}{2}) - u_{0,j}$.

In any element K , $\mathbf{w}_h \in \mathbb{P}^r(K)$. This vector space is spanned by a finite set of polynomial functions:

$$\mathbf{w}_h = \sum_{j=0}^R \mathbf{w}^{(j)} \varphi_j$$

so that we arrive at the following form of the semidiscrete scheme: for any K

$$M_K \frac{d}{dt} \mathbf{W}_K + F(\mathbf{w}_{h|K}) = 0$$

where the mass matrix is

$$(M_K)_{ij} = \int_K \varphi_i \varphi_j \, d\mathbf{x}$$

is clearly invertible. This is a block diagonal matrix, and hence its inversion (needed for time discretization) is rather straightforward. The vector F is defined by its components

$$F_j = - \int_K \nabla \varphi_j \cdot \mathbf{f}(\mathbf{w}_h(\mathbf{x}, t)) d\mathbf{x} + \int_{\partial K} \varphi_j \cdot \hat{\mathbf{f}}(\mathbf{w}_{h|K}, \mathbf{w}_{h|K^+}, \mathbf{n}) d\partial K$$

The choice of the degree of freedom, that is, the choice of the basis function, is an issue by itself. The choices are made depending whether to favor a geometrical interpretation (Lagrange basis), to facilitate the change of polynomial degree within elements (in the case of degree adaptivity), or whether the element shape is completely general (e.g., in the case of Lagrangian hydrodynamics Vilar *et al.*, 2014).

Nonlinear stability

One can show, in the scalar case, that a global entropy inequality can be easily derived, see Jiang and Shu (1994). In the scalar case, a natural entropy is $U(u) = \frac{u^2}{2}$: this is a convex function, and an entropy $g = (g_x, g_y)$ flux satisfies

$$u f'_x = g'_x, \quad u f'_y = g'_y$$

We see that $g_x = u f'_x - \int^u f'_x \, du$, $g_y = u f'_y - \int^u f'_y \, du$. In the following, we set $h_x = \int^u f'_x \, du$ and $h_y = \int^u f'_y \, du$. We wish to establish an inequality of the following type: for any K ,

$$\int_K \frac{\partial U(u_h)}{\partial t} d\mathbf{x} + \int_{\partial K} \hat{\mathbf{g}} \cdot \mathbf{n} \, d\mathbf{x} \leq 0 \quad (9)$$

Here, $\hat{\mathbf{g}} \cdot \mathbf{n}$ is an entropy flux, that is, a numerical flux consistent with g . This inequality simply states that we have a local L^2 energy bound.

For any v_h

$$\int_K \frac{\partial u_h}{\partial t} v_h \, d\mathbf{x} - \int_K f(u_h) \nabla v_h \cdot \mathbf{x} \, d\mathbf{x} + \int_{\partial K} v_h \hat{\mathbf{f}}(u_{h|K}, u_{h|K^-}, \mathbf{n}) d\partial K = 0$$

Then we choose $v_h = u_h$, so that

$$\frac{1}{2} \int_K \frac{\partial (u_h)^2}{\partial t} d\mathbf{x} - \int_K f(u_h) \nabla u_h \, d\mathbf{x} + \int_{\partial K} u_h \hat{\mathbf{f}}(u_{h|K}, u_{h|K^-}, \mathbf{n}) d\partial K = 0 \quad (10)$$

We get

$$\frac{1}{2} \int_K \frac{\partial(u_h)^2}{\partial t} dx + \int_{\partial K} \hat{G}(u_{h|K}, u_{h|K^-}, \mathbf{n}) d\partial K + A_K = 0$$

with

$$\begin{aligned} \hat{G}_{j+1/2} &= \frac{u_{h|K} + u_{h|K^-}}{2} \hat{\mathbf{f}}(u_{h|K}, u_{h|K^-}, \mathbf{n}) \\ &\quad - \frac{1}{2} (g(u_{h|K}) + g(u_{h|K^-})) \cdot \mathbf{n} \end{aligned}$$

This flux is consistent with $g \cdot \mathbf{n}$, and we also have set

$$A_K = \int_{[u_{h|K}, u_{h|K^-}]} (\hat{\mathbf{f}}(u_{h|K}, u_{h|K^-}, \mathbf{n}) - \mathbf{f}(v) \cdot \mathbf{n}) dv$$

Using the mean value theorem, we see that

$$A_K = (u_{h|K} - u_{h|K^-}) (\hat{\mathbf{f}}(u_{h|K}, u_{h|K^-}, \mathbf{n}) - \mathbf{f}(\xi) \cdot \mathbf{n})$$

for a suitable ξ between $u_{h|K}$ and $u_{h|K^-}$. If $\hat{\mathbf{f}}$ is an E-scheme (see Osher, 1984), that is, if for any ξ between u and v , $(\hat{\mathbf{f}}(u, v, \mathbf{n}) - \mathbf{f}(\xi) \cdot \mathbf{n})(u - v) \leq 0$, we see that (9) holds true for any E-scheme. Typical examples are the exact Godunov solver, and the Rusanov scheme

$$\hat{\mathbf{f}}(u, v, \mathbf{n}) = \frac{1}{2} (\mathbf{f}(u) \cdot \mathbf{n} + \mathbf{f}(v) \cdot \mathbf{n}) + \alpha(u - v)$$

for $\alpha \geq \max_{\xi \in [\min(u, v), \max(u, v)]} |\mathbf{f}(\xi) \cdot \mathbf{n}|$

The case of systems is, of course, more complex. One possible solution could be to rewrite the system (1) in term of the entropy variables: $\mathbf{v} = \nabla_{\mathbf{w}} S(\mathbf{w})$, where S is a (mathematical) entropy, in the following form:

$$\mathbf{w} \frac{\partial \mathbf{w}}{\partial t} + \operatorname{div} \mathbf{h}(\mathbf{v}) = 0$$

The change of variables $\mathbf{v} \mapsto \mathbf{w}$ is one to one and does not affect the Rankine–Hugoniot relations. Instead of approximating the state variable \mathbf{w} by piecewise polynomials, we can approximate the entropy with polynomials of degree $r+1$ and define the approximation state as

$$V_h = \{\mathbf{v} \in (L^1(\Omega))^m \cap (L^\infty(\Omega))^m, \mathbf{v} \in \mathbb{P}^{r+1}(K)\}$$

and look for $\mathbf{w}(\mathbf{v})$, with $\mathbf{v} \in V_h$ such that for any $\varphi \in V_h$ and for any K

$$\begin{aligned} \int_K \varphi \frac{\partial \mathbf{w}(\mathbf{v})}{\partial t} - \int_K \nabla \varphi \cdot \mathbf{h}(\mathbf{v}) dx \\ + \int_{\partial K} \varphi \hat{\mathbf{h}}(\mathbf{v}_K, \mathbf{v}_{K^-}, \mathbf{n}) d\partial K = 0 \end{aligned}$$

Clearly, if $\hat{\mathbf{h}}$ is an E-scheme, we have the entropy inequality

$$\int_K \frac{\partial S(\mathbf{v})}{\partial t} + \int_{\partial K} \hat{\mathbf{G}}(\mathbf{v}_K, \mathbf{v}_{K^-}, \mathbf{n}) d\partial K \leq 0$$

for a suitable consistent entropy flux $\hat{\mathbf{G}}$.

Controlling spurious oscillations

Another, and somewhat related issue is to control the Gibbs phenomena: when the numerical solution develops steep gradients, either because the mesh resolution is not sufficient or because discontinuities appear, spurious oscillations will appear. One of the fundamental questions is how to control them as automatically as possible. One of the solutions is to get inspired by what has been done for finite volume schemes, taking into account the negative result of Goodman and LeVeque (1985). In order to describe what has been achieved, let us turn back to the 1D case for the sake of simplicity. In that case, the scheme reduces to

$$\begin{aligned} \int_{K_j} \varphi \frac{du_h}{dt} - \int_{K_j} \varphi' f(u_h) dx + \varphi(x_{j+1/2}) \hat{f}(u_{h,j+1/2}^-, u_{h,j+1/2}^+) \\ - \varphi(x_{j-1/2}) \hat{f}(u_{h,j-1/2}^+, u_{h,j-1/2}^-) = 0 \end{aligned} \quad (11)$$

where $K_j = (x_{j-1/2}, x_{j+1/2})$, $x_j = \frac{x_{j-1/2} + x_{j+1/2}}{2}$, $\Delta_j = x_{j+1/2} - x_{j-1/2}$, and $u_{h,j+1/2}^\pm = u_h(x_j \pm \frac{\delta_j u}{2}) = u_h(x_j) \pm \delta_j^\pm u$.

One of the main remarks that enables us to understand the behavior of methods is what is called Harten's lemma. Instead of considering the semidiscrete case, let us use the fully discretized form; we will come back to this in Section 2.2. Assume we have a sequence $\{u_j^n\}_{j \in \mathbb{Z}, n \in \mathbb{N}}$ that satisfies ($\lambda > 0$)

$$u_j^{n+1} = u_j^n - \lambda \left(C_{j+1/2}(u_{j+1}^n - u_j^n) - D_{j-1/2}(u_j^n - u_{j-1}^n) \right) \quad (12)$$

If for any $j \in \mathbb{Z}$, we have

$$\begin{aligned} C_{j+1/2} \geq 0, \quad D_{j+1/2} \geq 0 \\ \lambda(C_{j+1/2} + D_{j+1/2}) \leq 1 \end{aligned} \quad (13)$$

then the sequence satisfies an L^∞ , an L^1 , and a TVD bound, where the total variation of $u = \{u_j\}_{j \in \mathbb{Z}}$ is

$$\operatorname{TV}(u) = \sum_{j \in \mathbb{Z}} |u_{j+1} - u_j|$$

There is no reason why the arguments u_j^\pm are such that the sequence defined by (12) are such that the conditions (13) are true. To do so, one technique is to introduce a limiter. The simplest is the generalized minmod limiter

$$m(a_1, a_2, \dots, a_m) = \begin{cases} s \min(|a_1|, |a_2|, \dots, |a_m|), \\ \quad \text{if } s = \text{sign}(a_1) = \dots = \text{sign}(a_m) \\ 0, \\ \quad \text{else} \end{cases} \quad (14)$$

The arguments in the flux \hat{f} (11) are modified as follows: We replace δ_j^\pm by

$$(\delta_j^\pm)^{\text{mod}} = m(\delta_j^\pm, \Delta_+ u_{0,j}, \Delta_- u_{0,j})$$

There is a huge literature on this topic. One may quote, among others, (Cockburn and Shu, 1989b; Biswas *et al.*, 1994; Burbeau *et al.*, 2001; Krivodonova *et al.*, 2004; Qiu and Shu, 2005a; Li and Qiu, 2010). Other kinds of polynomial representation can also be used, such as Hermite approximation, see Qiu and Shu 2005b, Balsara *et al.* (2007) and their relation to limiting. Another approach is to combine WENO limiters and the DG method, see Dumbser (2010), Zhu *et al.* (2008).

2.1.3 Stabilized continuous FEM

Another approach for spatial approximation is to consider the following trial space:

$$V_h = \{\mathbf{w}_h \in (L^\infty(\Omega))^m \cap (L^1(\Omega))^m \cap (C^0(\Omega))^m, \\ \text{for any element } K, (\mathbf{w}_h)|_K \in (\mathbb{P}^r(K))^m\}$$

The fundamental difference is that we now require continuity. In the simplest setting one uses elements of V_h as the test function, that is, one looks for a solution \mathbf{w}_h which satisfies for any $\varphi \in V_h$ the weak statement:

$$\int_{\Omega} \varphi \cdot \frac{\partial \mathbf{w}_h}{\partial t} \mathbf{dx} + a(\mathbf{w}_h, \varphi) = 0 \quad (15a)$$

where

$$a(\mathbf{w}, v) = - \int_{\Omega} \nabla v \cdot \mathbf{f}(\mathbf{w}_h) \mathbf{dx} + BC(\mathbf{w}, \varphi) \quad (15b)$$

The problem amounts to solving (with clear notations)

$$M \frac{d\mathbf{w}_h}{dt} + A(\mathbf{w}_h) = 0, \quad M_{ij} = \int_{\Omega} \varphi_i \varphi_j \mathbf{dx}$$

The mass matrix M is also invertible. It is a sparse matrix but is not block diagonal, contrarily to the DG method. In (15b), the operator BC describes the approximation of the weakly enforced boundary conditions. We do not describe it

because it depends on the nature of the boundary conditions: it is problem-dependent.

The method (15) is known to have stability difficulties, so it is better to add to the Galerkin variational form a stabilization operator. There are several forms of this stabilization operator: for example, the stream line operator (Hughes *et al.*, 1986; Johnson and Pitkäranta, 1986), and a jump operator (Burman, 2010). Instead of solving (15), we solve

$$\int_{\Omega} \varphi \cdot \frac{\partial \mathbf{w}_h}{\partial t} + a(\mathbf{w}_h, \varphi) + a_S(\mathbf{w}_h, \varphi) = 0 \quad (16a)$$

where a_S is a stabilization operator.

In the case of the streamline operator, the choice is Hughes *et al.* (1986)

$$a_S(\mathbf{w}_h, \varphi) = \sum_K h_K \int_K (\nabla_{\mathbf{w}} \mathbf{f}(\mathbf{w}_h) \cdot \nabla \varphi) \mathcal{T}_K \\ \times \left(\frac{\partial \mathbf{w}_h}{\partial t} + \nabla_u \mathbf{f}(\mathbf{w}_h) \cdot \nabla \mathbf{w}_h \right) \mathbf{dx} = 0 \quad (16b)$$

where h_K represents the diameter of K , and $\mathcal{T}_K \geq 0$ is a stabilization parameter (or matrix).

In the case of the jump operator, we take (Burman, 2010)

$$a_S(\mathbf{w}_h, \varphi) = \sum_{\text{internal edges}} \gamma_e h_e^2 \int_e \|\nabla_{\mathbf{w}} \mathbf{f}(\mathbf{w}_h)\| [\nabla \mathbf{w}_h] [\nabla \varphi_h] \quad (16c)$$

where $\gamma_e \geq 0$, and h_e is a measure of the edge e . The choice of the stabilization operator is done such that the exact solution also satisfies (16). Note that the structure of the mass matrix is affected in the case of (16b), hence its invertibility is less obvious. In the case of (16c), the mass matrix is not changed but the compactness of the computational stencil is slightly affected. Since these methods share a lot of similarities with the residual distribution methods (indeed, they can be seen as a particular case), we postpone the discussion.

2.2 Temporal discretizations

There are several standard ways of approximating in time, depending on how we look at time with respect to space. Either they are two unrelated parameters, so that one first approximates in space and then in time thanks to the method of lines. Or, one considers the equation

$$\frac{\partial \mathbf{w}}{\partial t} + \text{div } \mathbf{f}(\mathbf{w}) = 0$$

as a space–time divergence applied to the flux $(\mathbf{w}, \mathbf{f}(\mathbf{w}))$. Here, we focus on the explicit method of lines.

A typical example is the well-known method of lines. After having discretized in space, we have to discretize a problem of the form

$$\frac{\partial \mathbf{w}}{\partial t} = L(\mathbf{w}) \quad (17)$$

Depending on the hardness of the problem, or more generally speaking, of the properties we are looking for, one may consider an explicit or implicit scheme. A very popular method is the so-called strong stability preserving (SSP) technique (Gottlieb *et al.*, 2001). If the Euler operator $\mathbf{w} \mapsto \mathbf{v} = \mathbf{w} - \Delta t L(\mathbf{w})$ preserves the L^∞ norm or the L^1 norm or the TVD semi-norm for $\Delta t \leq \Delta t_0$, then the SSP Runge–Kutta method it is built on will also have the same property under the condition of the type $\Delta t \leq C \Delta t_0$. In the cases we are interested in, Δt_0 is defined by mean of a CFL-type condition, and is approximation-dependent.

The explicit SSP RK schemes are written as the Runge–Kutta schemes

$$\begin{aligned} u^{(0)} &= u^n \\ u^{(i)} &= \sum_{k=0}^{i-1} (\alpha_{i,k} u^{(k)} + \Delta t \beta_{i,k} L(u^{(k)})), \quad i = 1, \dots, m \\ u^{n+1} &= u^{(m)} \end{aligned}$$

where the $\alpha_{i,k}$ and $\beta_{i,k}$ are all *positive*. By consistency, $\sum_{k=0}^{i-1} \alpha_{i,k} = 1$, so that the intermediate stages can be written as a convex combination of the Euler operator. The integer m is the number of stages. Examples of such SSP RK methods are the following:

- Second order in time and $C = 1$ (no degradation of the time step).

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

- Third order in time and $C = 1$.

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

In these examples, the number of stages is equal to the order of the scheme. It can be shown (Gottlieb and Shu, 1998) that there exists no fourth order SSP RK scheme with four stages. Ruuth and Spiteri (2002) developed fourth-order methods with $m = 5, 6, 7$, and 8 stages: for example

$$u^{(1)} = u^n + 0.391752226571890 \Delta t L(u^n)$$

$$\begin{aligned} u^{(2)} &= 0.444370493651235 u^n \\ &\quad + 0.555629506348765 u^{(1)} \\ &\quad + 0.368410593050371 \Delta t L(u^{(1)}) \\ u^{(3)} &= 0.620101851488403 u^n \\ &\quad + 0.379898148511597 u^{(2)} \\ &\quad + 0.251891774271694 \Delta t L(u^{(2)}) \\ u^{(4)} &= 0.178079954393132 u^n \\ &\quad + 0.821920045606868 u^{(3)} \\ &\quad + 0.544974750228521 \Delta t L(u^{(3)}) \\ u^{n+1} &= 0.517231671970585 u^{(2)} \\ &\quad + 0.096059710526147 u^{(3)} \\ &\quad + 0.063692468666290 \Delta t L(u^{(3)}) \\ &\quad + 0.386708617503269 u^{(4)} \\ &\quad + 0.226007483236906 \Delta t L(u^{(4)}) \end{aligned}$$

for which $C = 1.508$.

A rather complete discussion can be found in Gottlieb *et al.* (2001), Gottlieb (2005), Ruuth and Spiteri (2002). Error estimates for explicit Runge–Kutta time stepping can be found in Burman *et al.* (2010), Zhang and Shu (2004, 2010), Meng *et al.* (2015).

3 A DIFFERENT SETTING: RESIDUAL DISTRIBUTION

We now consider the framework known today as *residual distribution*. Its roots can be found in the seminal work of Roe (1981, 1982) on *fluctuation splitting*, and in all the contributions of the 1990s on wave decomposition, hyperbolic elliptic splitting, and multidimensional upwind methods (Roe, 1987, 1990, 1994; Roe and Sidilkover, 1992; Sidilkover and Roe, 1995; Struijs *et al.*, 1991; Deconinck *et al.*, 1993; Nishikawa *et al.*, 2001), and see also Deconinck and Ricchiuto (2007).

We present it here as a general framework to study and unify the “more classical” approaches recalled in Section 2, while giving some additional flexibility to construct new “nonclassical” discretizations.

3.1 Steady hyperbolic problems

Consider the scalar steady-state advection equation

$$\vec{a} \cdot \nabla u = 0 \quad \text{on} \quad \Omega \subset \mathbb{R}^2 \quad (18)$$

where $\nabla \cdot \vec{a} = 0$, and with boundary conditions

$$\int_{\partial\Omega} (\vec{a} \cdot \hat{n})^-(g - u) = \int_{\partial\Omega^-} \vec{a} \cdot \hat{n}(g - u) = 0 \quad (19)$$

To find a numerical approximation of the solution of (18) and (19), on a tessellation of the spatial domain Ω_h , we use a generalization of the fluctuation splitting strategy put forward by Roe. In particular, we start by considering u_h , a continuous nodal finite element approximation of the solution

$$u_h = \sum_{i \in \Omega_h} \varphi_i u_i = \sum_{K \in \Omega_h} u_i \varphi_i \Big|_K \quad (20)$$

For a given degree of freedom i of the continuous collocated finite element expansion, let K_i be the set of elements sharing i as a node, and, similarly, let F_i be the set of mesh faces sharing i . Given an initial guess for the degrees of freedom, we proceed as follows: (cf. Figure 2):

1. For all elements K , compute the *fluctuation/residual*

$$\phi^K = \int_K \vec{a} \cdot \nabla u_h|_K \, d\mathbf{x} \left(\approx - \int_K \partial_t u_h \, d\mathbf{x} \right) \quad (21)$$

2. For all elements K , distribute the fluctuation to the three nodes of K . Let ϕ_j^K denote the *amount* of fluctuation sent to node $j \in K$; then the *conservation/consistency*

requirement is

$$\sum_{j=1}^{j=j_K} \phi_j^K = \phi^K \quad (22)$$

3. For all nodes $i \in \Omega_h$, assemble *signals* from the surrounding elements and evolve toward steady state by some iterative procedure such as, for example

$$u_i^{n+1} = u_i^n - \omega_i \sum_{K \in K_i} \phi_i^K \quad (23)$$

The method described by (21)–(23) aims at providing a solution to the discrete algebraic system

$$\sum_{K \in K_i} \phi_i^K = 0, \quad \forall i \in \Omega_h \quad (24)$$

As formulated, it does not include boundary conditions (BCs). The most general way to introduce them is to consider, for any face $f \in \partial\Omega_h$ the *face fluctuations*:

$$\phi^f = \int_f \vec{a} \cdot \mathbf{n}(g_h^* - u_h) \, df \quad (25)$$

where, to embed the compatibility condition implicit in (19), we have introduced the numerical flux

$$(g^* \vec{a}) \cdot \mathbf{n} = \vec{a} \cdot \mathbf{n} \left[\frac{1 + \text{sign}(\vec{a} \cdot \mathbf{n})}{2} u + \frac{1 - \text{sign}(\vec{a} \cdot \mathbf{n})}{2} g \right]$$

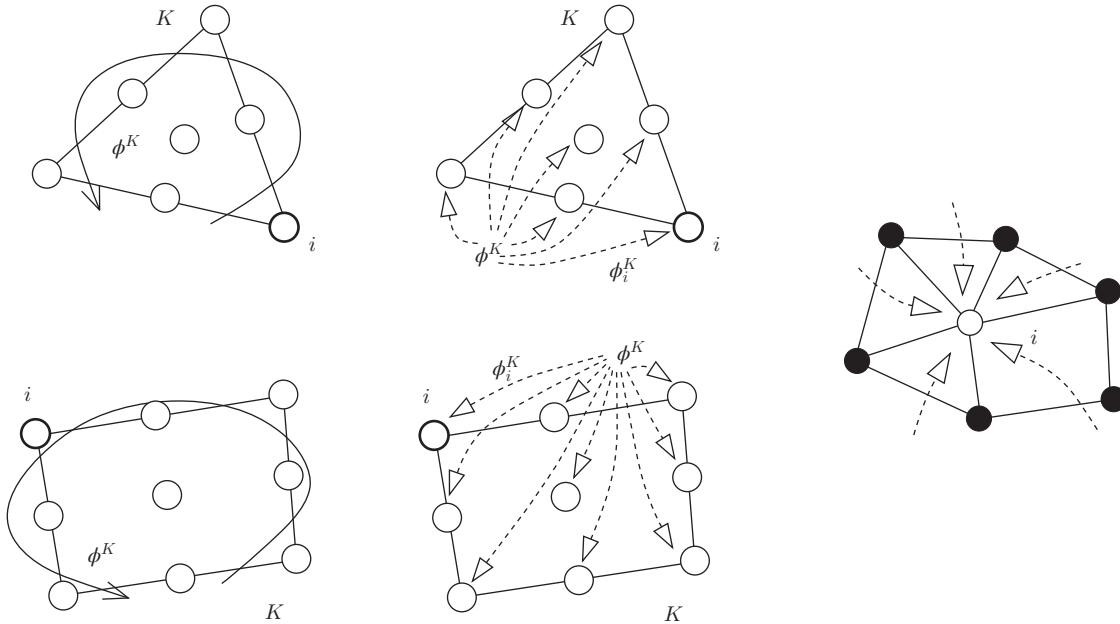


Figure 2. Residual distribution.

Face fluctuations can be split to the degrees of freedom $j \in f$ by means of distributed residuals ϕ_j^f such that

$$\sum_{j=1}^{j_f} \phi_j^f = \phi^f \quad (26)$$

Finally, the complete discrete fluctuation splitting/residual distribution steady equations read

$$\sum_{K \in K_i} \phi_i^K + \sum_{f \in F_i \cap \partial\Omega_h} \phi_i^f = 0 \quad (27)$$

3.1.1 Accuracy conditions

The first formulation of these schemes, on linear triangular elements, relied for the construction of second-order discretizations on the so-called linearity preservation property (Roe, 1987; Paillere and Deconinck, 1997), defined as follows:

Definition 1. (Linearity preservation) Let $\{\beta_j^K\}_{j \in K}$ be a set of distribution coefficients uniformly bounded with respect to h , u_h , ϕ^K , and with respect to the data of the problem (\vec{a} , boundary data, etc.), and verifying the consistency property

$$\sum_{j=1}^{j_K} \beta_j^K = 1 \quad (28)$$

A linearity preserving scheme is one for which

$$\phi_i^K = \beta_i^K \phi^K \quad (29)$$

Proposition 1. Linearity preserving schemes are second-order accurate.

The simple property stated in Definition 1 and Proposition 1 has been known since the late 1980s, but it has taken more than a decade to be formally understood. A more general characterization of the accuracy of these schemes, due to Abgrall (2001) and generalized in Abgrall and Roe (2003), Deconinck and Ricchiuto (2007), Ricchiuto *et al.* (2007), Abgrall and Treflik (2010), Abgrall *et al.* (2014), is the following:

Definition 2. (Truncation error and accuracy) Let ψ be smooth function, $\psi \in C^{r+1}(\Omega)$. Let Ω_h be an unstructured grid composed of nonoverlapping elements. On the generic element $K \in \Omega_h$, consider the r th degree continuous polynomial approximation (20). Let, in particular, $\psi_h = \sum_{j \in K} \psi_j \phi_j$

be the r th degree polynomial approximation of type (20) of ψ , the values ψ_j being obtained by Galerkin projection. Consider now an exact, smooth function $u \in H^{r+1}$ verifying (18) and (19) in a classical sense: $\vec{a} \cdot \nabla u = 0$ in Ω , and $u = g$ on $\partial\Omega^-$. Let u_h be its polynomial approximation of degree r of type (20), obtained by Galerkin projection. Let now $\phi_j^K(u_h)$ and $\phi_j^f(u_h)$ the values of the split residuals obtained when replacing the nodal values of the solution obtained with the scheme by the values u_j of the Galerkin projection of u . We define the integral truncation error $\epsilon(u_h, \psi)$ as

$$\begin{aligned} \epsilon(u_h, \psi) &= \sum_{j \in \Omega_h} \psi_j \left\{ \sum_{K \in K_j} \phi_j^K(u_h) + \sum_{f \in F_j} \phi_j^f(u_h) \right\} \\ &= \sum_{K \in \Omega_h} \sum_{j=1}^{j_K} \psi_j \phi_j^K(u_h) + \sum_{f \in \partial\Omega_h} \sum_{j=1}^{j_f} \psi_j \phi_j^f(u_h) \end{aligned} \quad (30)$$

We say that a scheme is $r+1$ order accurate if it verifies the truncation error estimate

$$|\epsilon(u_h, \psi)| \leq C(\Omega_h) h^{r+1}$$

The following general characterization is possible:

Proposition 2. In d spatial dimensions, a sufficient condition for scheme (27) to be $r+1$ order accurate in the sense of Definition 2 is to simultaneously have

$$\begin{aligned} |\phi_i^K(u_h)| &\leq C_{\Omega_h} h^{r+d} \quad \forall K \in \Omega_h, \quad \forall i \in K \\ |\phi_i^f(u_h)| &\leq C_{\partial\Omega_h} h^{r+d-1} \quad \forall f \in \partial\Omega_h, \quad \forall i \in f \end{aligned} \quad (31)$$

The proof of this property is omitted for brevity. The interested reader can refer to Abgrall and Roe (2003), Deconinck and Ricchiuto (2007), Ricchiuto *et al.* (2007), Abgrall and Treflik (2010), Abgrall *et al.* (2014) for details. The importance of this characterization is that it allows us to provide some design conditions. To see this, first recall that for the solution u of Definition 2, and for its Galerkin projection u_h on the r -degree finite element polynomial space, we can use classical approximation results (Ciarlet and Raviart, 1972; Ern and Guermond, 2004) to show that

$$\|\vec{a} \cdot \nabla u_h - \vec{a} \cdot \nabla u\| \leq C_u h^r \quad \text{in } \Omega_h$$

and, provided that the boundary $\partial\Omega$ is also smooth enough, and provided that $\partial\Omega_h$ is a high-order polynomial rendering

of the exact boundary (Abgrall and Treflik, 2010; Abgrall et al., 2014), we also have¹

$$\|(g^* \vec{a} \cdot \mathbf{n})_h - (g^* \vec{a} \cdot \mathbf{n})\| \leq C_{u,n} h^{r+1}$$

where the norms used are standard L norms, such as the L^2 or the max norm, with no derivatives involved. With the regularity hypotheses made on the mesh, we also have that $|K| = \mathcal{O}(h^d)$ and $|f| = \mathcal{O}(h^{d-1})$, for any K and any f . This, and the fact that $\vec{a} \cdot \nabla u = 0$ and $g^* - u = 0$ for the exact solution, leads to the conclusion that a sufficient condition for a scheme to be $r + 1$ order accurate in the sense of definition 2 is that we can find for any $K \in \Omega_h$ and for any $f \in \partial\Omega_h$ sets of uniformly bounded test functions ω_i^K and ω_i^f , such that

$$\sum_{j=1}^{j_K} \omega_j^K = 1 \quad \text{and} \quad \sum_{j=1}^{j_f} \omega_j^f = 1$$

and that the distribution can be obtained as

$$\begin{aligned} \phi_i^K(u_h) &= \int_K \omega_i^K \vec{a} \cdot \nabla u_h \, d\mathbf{x} \quad \text{and} \\ \phi_i^f(u_h) &= \int_f \omega_i^f \vec{a} \cdot \mathbf{n} (g_h^* - u_h) \, df \end{aligned} \quad (32)$$

Clearly, the linearity preserving schemes of Definition 1 are obtained as the particular case in which the test functions are constant within each element!

As we will see immediately, this consistency analysis applies trivially to classical continuous Galerkin discretizations, as well as to their stabilized counterparts, and to dG methods.

3.1.2 Stability and convergence

The above consistency analysis gives the conditions under which, if convergence with respect to the mesh parameter h is obtained, $r + 1$ convergence rates are expected w.r.t. h for an r th degree polynomial approximation, and in correspondence of sufficiently smooth solutions. The missing piece of information is: how do we make sure that convergence is indeed achieved? A classical finite element convergence analysis would need two main ingredients (Ern and Guermond, 2004): a consistency estimate, which we have provided, and a stability condition, which we have not. If we could provide a stability statement that ensures, for example, that $\forall u_h$ in our approximation space

$$\begin{aligned} &\left| \sum_K \sum_{j=1}^{j_K} u_j \phi_j^K(u_h) + \sum_f \sum_{j=1}^{j_f} u_j \phi_j^f(u_h) \right| \\ &\geq C' \|u_h\|^2, \quad \text{with } 0 < C' < \infty \end{aligned} \quad (33)$$

then using more or less classical arguments (Ern and Guermond, 2004), we could infer the existence of the discrete solution and derive more rigorous estimates for the error associated to this solution.

Unfortunately, *to this day, residual distribution schemes lack a framework for stability analysis*. Some weaker results showing the decay of the solution energy (L^2 norm) during iterations (23) have been shown in several works (Barth, 1996; Abgrall and Barth, 2002; Deconinck and Ricchiuto, 2007). These conditions are, however, not sufficient to say more on the discrete solution.

On the other hand, we are able to rule out some schemes as the following property shows in two space dimensions²:

Proposition 3. (Fall of the $\beta\Phi$ paradigm, 2D advection).
Consider the solution of

$$\vec{a} \cdot \nabla u = 0$$

in two space dimensions, with \vec{a} constant, and with $\partial\Omega$ a collection of straight sides. Any scheme of the form

$$0 = \sum_{K \in K_i} \beta_i^K \phi^K + \sum_{f \in F_i \cap \partial\Omega_h} \beta_i^f \phi^f$$

cannot be free from high-frequency spurious modes whatever the form of β_i^K , if K is a P^k Lagrange triangle with $k > 2$ and if K is a Q^k Lagrange quadrilateral $\forall k \geq 1$.

Proof. For all elements considered, we explicitly show one spurious mode exact solution of the discrete problem with homogeneous boundary conditions. This mode can be added to any grid function without the scheme detecting its presence, thus preserving this unphysical perturbation.

First, recall that for homogeneous boundary conditions and using the hypothesis on $\partial\Omega$

$$\phi^K = \oint_{\partial K} \vec{a} \cdot \mathbf{n} u_h = \sum_{f \in \partial K} \vec{a} \cdot \mathbf{n} \int_f u_h \, df$$

and

$$\begin{aligned} \phi^f &= - \int_f \frac{1 - \text{sign}(\vec{a} \cdot \mathbf{n})}{2} \vec{a} \cdot \mathbf{n} u_h \, df \\ &= - \frac{1 - \text{sign}(\vec{a} \cdot \mathbf{n})}{2} \vec{a} \cdot \mathbf{n} \int_f u_h \, df \end{aligned}$$

so we focus on the approximation of the integrals of u_h over the element faces. Let the number of freedom on each face $f \in \partial K$ be $C_f + 2$. We consider the mode defined by $u_j = 1$

if j is a vertex; otherwise, on each $f \in \partial K$, we set $\forall j \neq v$

$$u_j = -\frac{2}{C_f} \frac{\int_f \varphi_v df}{\int_f \varphi_j df}$$

having denoted with v one of the two vertices forming face f . The mode is compatible with the continuity of the representation and with the adoption of hybrid meshes. For P^k triangles with $k \geq 3$ and Q^k elements with $k \geq 2$, the value of the solution at nodes within the elements remains arbitrary. For this mode, one easily checks that $\phi^K = 0$, $\forall K$, and that $\phi^f = 0$, $\forall f \in \partial\Omega_h$.

The only remaining element is the Q^1 quadrilateral, which is easily checked to suffer from the checkerboard spurious mode in which u oscillates between -1 and 1 on every face. \square

The important consequence of Proposition 3 is that we have to start looking for schemes exploiting the subelemental variation of the discrete solution. A well-known example of such a scheme, perfectly fitting the framework presented, is the Streamline Upwind Petrov Galerkin (SUPG) scheme of Hughes and Brook (1982), Franca *et al.* (1990), Hughes *et al.* (2004) obtained by setting

$$\begin{aligned} \phi_i^K &= \int_K \varphi_i \vec{a} \cdot \nabla u_h \, d\mathbf{x} + \overbrace{\int_K \vec{a} \cdot \nabla \varphi_i \tau_K \vec{a} \cdot \nabla u_h \, d\mathbf{x}}^{\text{Streamline Dissipation}} \\ \text{and } \phi_i^f &= \int_f \varphi_i \vec{a} \cdot \mathbf{n} (g_h^* - u_h) \, df \end{aligned} \quad (34)$$

Stability results for the SUPG scheme can be obtained in the classical sense discussed in the beginning of this section (see, e.g., Szepessy, 1989; Johnson *et al.*, 1990; Johnson and Szepessy, 1990; Bochev *et al.*, 2004; Burman, 2010; Hughes *et al.*, 2004 and references therein), and are based on the positive-semidefinite nature of the bilinear form associated with the streamline dissipation term.

Other examples of schemes that allow overcoming the flaw of Proposition 3 will be given in the following. In general, guidelines to construct such methods can be obtained by considering the convergence of iteration (23). In the simplest setting of scalar advection, if we recast this iteration as the following update for the array of degrees of freedom U

$$U^{n+1} = U^n - \omega(A_h U^n - F)$$

convergence requires that, for some $r < 1$ and for all V

$$\|(\text{Id} - \omega A_h)V\|^2 \leq r\|V\|^2$$

which can be developed into

$$V^t A_h V \geq \frac{1-r}{2\omega} \|V\|^2 + \frac{\omega}{2} \|A_h V\|^2 \geq C_h \|V\|^2 \geq 0$$

leading back to a condition of type (33), and to the necessary (albeit not sufficient) condition

$$V^t A_h V \geq 0 \quad (35)$$

which we will use in the following.

3.1.3 Embedding a discrete maximum principle

Monotonicity of the numerical solution is retained by the so-called local positivity constraints for the distribution. This property is related to positive coefficient theory, which has replaced the TVD theory to construct high-order schemes (Goodman and LeVeque, 1985; Spekreijse, 1987; Barth, 2003).

Definition 3. (Positive scheme) A (locally) positive scheme is one for which

$$\phi_i^K = \sum_{\substack{j \in K \\ j \neq i}} c_{ij}(u_i - u_j), \quad c_{ij} \geq 0 \quad \forall j \in K \quad (36)$$

Positivity is the key to the construction of non-oscillatory schemes (Roe, 1987; Paillere and Deconinck, 1997):

Proposition 4. (Local Positivity and discrete maximum principle). Locally positive schemes, combined with the evolution step (23), verify the discrete maximum principle

$$\min_{j \in K_i} u_j^n \leq u_i^{n+1} \leq \max_{j \in K_i} u_j^n \quad \forall i \in \Omega_h$$

under the following condition:

$$\min_{i \in \Omega_h} \left(\omega_i \sum_{K \in K_i} \sum_{\substack{j \in K \\ j \neq i}} c_{ij} \right) \leq 1$$

Proof. The proof follows from the positivity of the c_{ij} s and time step restriction, and from

$$u_i^{n+1} = \left(1 - \omega_i \sum_{K \in K_i} \sum_{\substack{j \in K \\ j \neq i}} c_{ij} \right) u_i^n + \omega_i \sum_{j \in K_i} \sum_{K \in K_i \cap K_j} c_{ij} u_j^n$$

\square

This characterization can be generalized to fully consistent time-dependent discretizations as we will show later.

3.1.4 A general framework: relation with classical discretization approaches

The formalism introduced in the previous sections for the scalar advection equation can be easily generalized to (systems of) steady nonlinear conservation laws of the form

$$\operatorname{div} \mathbf{f}(\mathbf{w}) = 0 \quad \text{on } \Omega \subset \mathbb{R}^2 \quad (37)$$

with the appropriate boundary conditions on $\partial\Omega$. We now look for a solution satisfying

$$\sum_{K \in \mathcal{K}_i} \phi_i^K + \sum_{f \in F_i \cap \partial\Omega_h} \phi_i^f = 0 \quad (38)$$

where in every element K

$$\sum_{i \in K} \phi_i^K = \oint_{\partial K} \mathbf{f}(\mathbf{w}_h) \cdot \mathbf{n} \, dK = \int_K \operatorname{div} \mathbf{f}(\mathbf{w}_h) \, d\mathbf{x} \quad (39)$$

while on a boundary face f we have

$$\sum_{i \in f} \phi_i^f = \int_f (\hat{\mathbf{f}}(\mathbf{w}_h, \mathbf{g}, \mathbf{n}) - \mathbf{f}(\mathbf{w}_h) \cdot \mathbf{n}) \, df \quad (40)$$

where the numerical flux $\hat{\mathbf{f}}(\mathbf{w}_h, \mathbf{g}, \mathbf{n})$ accounts for the boundary conditions.

This framework defines a sort of *super class* of methods, which allows embedding and has relations with all the discretization approaches introduced in Section 2.

Continuous FEM as residual distribution

The simplest example is perhaps that of the stabilized finite elements discussed in Section 2.1.3. In particular, given a continuous collocated finite element expansion for which we can write $V_h = \operatorname{span}\{\varphi_i\}_{i \in \Omega_h}$, then the (unstabilized) continuous Galerkin method is obtained simply by setting

$$\begin{aligned} \phi_i^K &= \int_K \varphi_i \operatorname{div} \mathbf{f}(\mathbf{w}_h) \, d\mathbf{x}, \\ \phi_i^f &= \int_f \varphi_i (\hat{\mathbf{f}}(\mathbf{w}_h, \mathbf{g}, \mathbf{n}) - \mathbf{f}(\mathbf{w}_h) \cdot \mathbf{n}) \, df \end{aligned}$$

For nodal finite elements, the relation $\sum_{i \in K} \varphi_i = 1$ ensures that consistency is satisfied.

There is, however, a slight catch, which is worth pointing out. The relation between the last definitions and the consistency condition (39) requires that exact integration is performed w.r.t. the assumed polynomial variation of \mathbf{w}_h and the definition of the nonlinear flux \mathbf{f} . This is required to go from the integral of the flux divergence to the boundary integral (39), so that the variational formulation (15b) is recovered. In practice, exact quadrature is never used. The

practical way to handle this issue is to introduce a high-order polynomial representation of the flux \mathbf{f}_h . Based on the accuracy of a given quadrature formula, we can uniquely identify the polynomial degree of such an expansion, built starting from the reconstructed values of \mathbf{w}_h at a sufficient number of flux evaluation points, exactly as was done in the so-called quadrature-free approaches used in DG (Atkins and Shu, 1998; Lockard and Atkins, 1999) and in the most recent flux reconstruction methods (cf. Huynh *et al.*, 2014 and references therein). Based on the exact evaluation of the integrals of this polynomial flux, we can reformulate our consistency conditions as

$$\sum_{i \in K} \phi_i^K = \oint_{\partial K} \mathbf{f}_h(\mathbf{w}_h) \cdot \mathbf{n} \, dK = \int_K \operatorname{div} \mathbf{f}_h(\mathbf{w}_h) \, d\mathbf{x} \quad (41)$$

and

$$\sum_{i \in f} \phi_i^f = \int_f (\hat{\mathbf{f}}^h(\mathbf{w}_h, \mathbf{g}, \mathbf{n}) - \mathbf{f}_h(\mathbf{w}_h) \cdot \mathbf{n}) \, df \quad (42)$$

The choice of the polynomial degree has to respect at least some accuracy constraints, which are easily deduced from the analysis of Section 3.2.1. In particular, for this analysis to apply in the nonlinear case, one must ensure that, for a given smooth flux \mathbf{f} , and for an r th degree finite element approximation

$$\begin{aligned} \|\operatorname{div} \mathbf{f}_h(\mathbf{w}_h) - \operatorname{div} \mathbf{f}\|_K &\leq C_u h^r, \\ \|\mathbf{f}_h \cdot \mathbf{n} - \mathbf{f} \cdot \mathbf{n}\|_f &\leq C_{u,n} h^{r+1} \end{aligned}$$

This requires the flux polynomial to be of degree $r_f \geq r$.

With this modification, the (unstabilized) continuous Galerkin approximation will read

$$\begin{aligned} \phi_i^K &= \int_K \varphi_i \operatorname{div} \mathbf{f}_h(\mathbf{w}_h) \, d\mathbf{x}, \\ \phi_i^f &= \int_f \varphi_i (\hat{\mathbf{f}}^h(\mathbf{w}_h, \mathbf{g}, \mathbf{n}) - \mathbf{f}_h(\mathbf{w}_h) \cdot \mathbf{n}) \, df \end{aligned}$$

while a stabilized variant is readily obtained by including in ϕ_i^K the streamline dissipation term (cf. Section 2.1.3), leading to an SUPG distribution:

$$\begin{aligned} \phi_i^K &= \int_K \varphi_i \operatorname{div} \mathbf{f}_h(\mathbf{w}_h) \, d\mathbf{x} \\ &\quad + \int_K (\nabla_{\mathbf{w}} \mathbf{f}(\mathbf{w}_h) \cdot \nabla \varphi_i) \mathcal{T}_K (\nabla_{\mathbf{w}} \mathbf{f}(\mathbf{w}_h) \cdot \nabla \mathbf{w}_h) \, d\mathbf{x} \end{aligned} \quad (43)$$

where the relation $\sum_{i \in K} \varphi_i = 1$ allows us to show that (42) and (41) are met.

Although different in spirit, the edge-stabilized schemes discussed, for example, in Burman *et al.* (2008, 2010) can also be recast in the formalism above by setting

$$\begin{aligned}\phi_i^K &= \int_K \varphi_i \operatorname{div} \mathbf{f}_h(\mathbf{w}_h) \, d\mathbf{x} \\ &+ \oint_{\partial K} \gamma^{\partial K}(\mathbf{w}_h) [\nabla \mathbf{w}_h] \cdot [\nabla \varphi_i]\end{aligned}\quad (44)$$

where, again, consistency is a consequence of the partition of unity property, while one can easily demonstrate that the accuracy conditions are met provided that $\gamma^{\partial K}(\mathbf{w}_h) = \mathcal{O}(h^2)$, as in (16c).

Finite volume versus residual distribution: local conservation and continuous FEM

We recall here the analogy between residual distribution node and centered finite volume schemes on median dual cells. With the notation of Section 2.1.1, and with reference to Figure 3, we can write the semidiscrete evolution equation for \mathbf{w}_i , the average of \mathbf{w} on cell C_i , as

$$\begin{aligned}|C_i| \frac{d\mathbf{w}_i}{dt} &= - \sum_j \int_{f_{ij}} \hat{f}(\mathcal{R}(\mathbf{w}_h)_i, \mathcal{R}(\mathbf{w}_h)_j, \mathbf{n}_{ij}) \\ &= - \sum_{K \in K_i} \sum_{\substack{j \in K \\ j \neq i}} \int_{f_{ij}^K} \hat{f}(\mathcal{R}(\mathbf{w}_h)_i, \mathcal{R}(\mathbf{w}_h)_j, \mathbf{n}_{ij}) \\ &= - \sum_{K \in K_i} \sum_{\substack{j \in K \\ j \neq i}} \hat{\mathbf{f}}_{ij}^K \cdot \mathbf{n}_{ij}\end{aligned}\quad (45)$$

Local conservation is equivalent now to the condition

$$\hat{\mathbf{f}}_{ij}^K \cdot \mathbf{n}_{ij} + \hat{\mathbf{f}}_{ji}^K \cdot \mathbf{n}_{ji} = 0 \quad (46)$$

Since C_i is a closed polygon, we also have $\sum_{K \in K_i} \sum_{j \in K, j \neq i} \mathbf{n}_{ij} = 0$, which allows us to recast (45) as

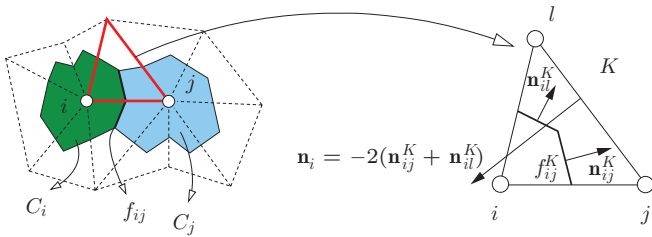


Figure 3. Node centered finite volume.

$$|C_i| \frac{d\mathbf{w}_i}{dt} = - \sum_{K \in K_i} \sum_{\substack{j \in K \\ j \neq i}} (\hat{\mathbf{f}}_{ij}^K - \mathbf{f}_i) \cdot \mathbf{n}_{ij} \quad (47)$$

If we now set

$$\phi_i^K = \sum_{\substack{j \in K \\ j \neq i}} (\hat{\mathbf{f}}_{ij}^K - \mathbf{f}_i) \cdot \mathbf{n}_{ij} \quad (48)$$

we find that the local conservation property (46) implies

$$\sum_{j \in K} \phi_i^K = \frac{1}{2} \sum_{j \in K} \mathbf{f}_j \cdot \mathbf{n}_j = \oint_{\partial K} \mathbf{f}_h \cdot \mathbf{n} \, df = \phi^K$$

with $\mathbf{f}_h = \sum_j \varphi_j \mathbf{f}_j$.

This shows that, for any given higher order finite volume discretization, we may define a residual distribution method consistent with a second-order polynomial flux approximation. While this was known for some time, the reverse is not. In particular, given a definition of the split residuals $\{\phi_j^K\}_{j \in K}$, we may ask if there exists a definition of consistent fluxes expressing local conservation over the median dual cell for the residual distribution method. If we require these fluxes to satisfy (48), then we may write the system

$$\begin{aligned}\hat{\mathbf{f}}_{ij}^K \cdot \mathbf{n}_{ij} + \hat{\mathbf{f}}_{il}^K \cdot \mathbf{n}_{il} &= \phi_i^K - \frac{1}{2} \mathbf{f}_i \cdot \mathbf{n}_i \\ \hat{\mathbf{f}}_{ji}^K \cdot \mathbf{n}_{ji} + \hat{\mathbf{f}}_{jl}^K \cdot \mathbf{n}_{jl} &= \phi_j^K - \frac{1}{2} \mathbf{f}_j \cdot \mathbf{n}_j \\ \hat{\mathbf{f}}_{li}^K \cdot \mathbf{n}_{li} + \hat{\mathbf{f}}_{lj}^K \cdot \mathbf{n}_{lj} &= \phi_l^K - \frac{1}{2} \mathbf{f}_l \cdot \mathbf{n}_l\end{aligned}\quad (49)$$

using local conservation, and setting $\Psi_i^K = \phi_i^K - \mathbf{f}_i \cdot \mathbf{n}_i/2$, we obtain a linear system for $(\hat{\mathbf{f}}_{ij}^K \cdot \mathbf{n}_{ij}, \hat{\mathbf{f}}_{jl}^K \cdot \mathbf{n}_{jl}, \hat{\mathbf{f}}_{li}^K \cdot \mathbf{n}_{li})$:

$$\begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\mathbf{f}}_{ij}^K \cdot \mathbf{n}_{ij} \\ \hat{\mathbf{f}}_{jl}^K \cdot \mathbf{n}_{jl} \\ \hat{\mathbf{f}}_{li}^K \cdot \mathbf{n}_{li} \end{pmatrix} = \begin{pmatrix} \Psi_i^K \\ \Psi_j^K \\ \Psi_l^K \end{pmatrix} \quad (50)$$

The associated matrix has rank 2. We can easily find particular solutions setting to zero one of the unknowns and solving the resulting subsystem. Averaging out the three particular solutions obtained (for symmetry) ends up with the following multidimensional numerical fluxes w.r.t. which the residual distribution scheme is locally conservative on the median dual cell:

$$\begin{aligned}\hat{\mathbf{f}}_{ij}^K \cdot \mathbf{n}_{ij} &= \hat{\mathbf{f}}_{ij}^K \cdot \mathbf{n}_{ij}(\mathbf{w}_i, \mathbf{w}_j, \mathbf{w}_l) = \frac{1}{3} (\Psi_i^K - \Psi_j^K) \\ &= \frac{1}{3} (\phi_i^K - \phi_j^K) - \frac{1}{6} (\mathbf{f}_i \cdot \mathbf{n}_i - \mathbf{f}_j \cdot \mathbf{n}_j)\end{aligned}$$

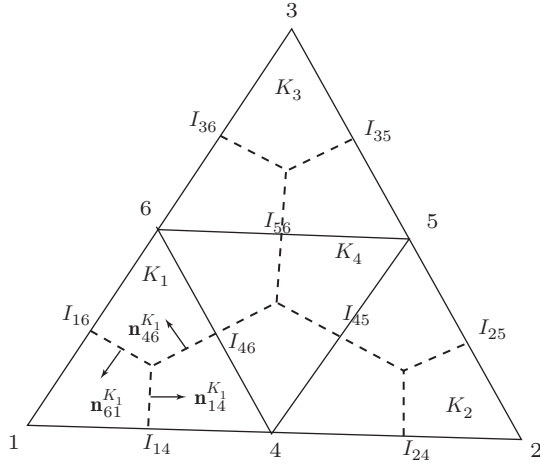


Figure 4. \mathbb{P}^2 residual distribution and finite volumes.

$$\begin{aligned}
 \hat{\mathbf{f}}_{jl}^K \cdot \mathbf{n}_{jl} &= \hat{\mathbf{f}}_{jl}^K \cdot \mathbf{n}_{jl}(\mathbf{w}_i, \mathbf{w}_j, \mathbf{w}_l) = \frac{1}{3}(\Psi_j^K - \Psi_l^K) \\
 &= \frac{1}{3}(\phi_j^K - \phi_l^K) - \frac{1}{6}(\mathbf{f}_j \cdot \mathbf{n}_j - \mathbf{f}_l \cdot \mathbf{n}_l) \\
 \hat{\mathbf{f}}_{li}^K \cdot \mathbf{n}_{li} &= \hat{\mathbf{f}}_{li}^K \cdot \mathbf{n}_{li}(\mathbf{w}_i, \mathbf{w}_j, \mathbf{w}_l) = \frac{1}{3}(\Psi_l^K - \Psi_i^K) \\
 &= \frac{1}{3}(\phi_l^K - \phi_i^K) - \frac{1}{6}(\mathbf{f}_l \cdot \mathbf{n}_l - \mathbf{f}_i \cdot \mathbf{n}_i) \quad (51)
 \end{aligned}$$

These are the *three states' multidimensional numerical fluxes*. Consistency can be formulated as

$$\hat{\mathbf{f}}_{ij}^K \cdot \mathbf{n}_{ij}(\mathbf{w}, \mathbf{w}, \mathbf{w}) = \mathbf{f}(\mathbf{w}) \cdot \frac{\mathbf{n}_j - \mathbf{n}_i}{6}$$

Because for a constant state over the element, we always have $\phi_i^K = 0 \forall i$. Other standard properties of numerical fluxes, for example, Lipschitz continuity, are inherited from the properties of the physical flux \mathbf{f} and of the split residuals ϕ_i^K .

A similar construction can be repeated for schemes written on high-order finite elements. To understand how this works, we start from the \mathbb{P}^2 case. We consider the setup shown in Figure 4: the element is split first into four sub-triangles K_1 , K_2 , K_3 , and K_4 . From this subtriangulation, we can construct a dual mesh as in the P^1 case. The dual mesh is a collection of cells C_j whose intersection with an element K defines six subzones, represented by the dashed lines in the figure. The notation used in this case is similar to the one used before: in the sub-triangle K_i , we denote by $\mathbf{n}_{ij}^{K_i}$ the normal to the portion of the face separating the median dual cells C_i and C_j , and by $\mathbf{f}_{ij}^{K_i} \cdot \mathbf{n}_{ij}^{K_i}$ the corresponding local numerical flux.

We can now write down the finite volume equations for each control cell C_j , and then proceed as in the P^1 case to determine the contributions associated with each subelement

K_i . To relate these subelemental residuals to the \mathbb{P} distributed residuals, we sum for each node the contribution from the subelements to which the node belongs. As before, this leads to a system of equations, which reads

$$\begin{aligned}
 &\hat{\mathbf{f}}_{14}^{K_1} \cdot \mathbf{n}_{14}^{K_1} - \hat{\mathbf{f}}_{61}^{K_1} \cdot \mathbf{n}_{61}^{K_1} = \phi_1^K - \mathbf{F}_1^K \\
 &-\hat{\mathbf{f}}_{42}^{K_2} \cdot \mathbf{n}_{42}^{K_2} + \hat{\mathbf{f}}_{25}^{K_2} \cdot \mathbf{n}_{25}^{K_2} = \phi_2^K - \mathbf{F}_2^K \\
 &-\hat{\mathbf{f}}_{53}^{K_3} \cdot \mathbf{n}_{53}^{K_3} + \hat{\mathbf{f}}_{36}^{K_3} \cdot \mathbf{n}_{36}^{K_3} = \phi_3^K - \mathbf{F}_3^K \\
 &-\hat{\mathbf{f}}_{14}^{K_1} \cdot \mathbf{n}_{14}^{K_1} + (\hat{\mathbf{f}}_{41}^{K_1} \cdot \mathbf{n}_{46}^{K_1} - \hat{\mathbf{f}}_{64}^{K_4} \cdot \mathbf{n}_{64}^{K_4}) \\
 &+ (\hat{\mathbf{f}}_{45}^{K_4} \cdot \mathbf{n}_{45}^{K_4} - \hat{\mathbf{f}}_{54}^{K_2} \cdot \mathbf{n}_{54}^{K_2}) + \hat{\mathbf{f}}_{42}^{K_2} \cdot \mathbf{n}_{42}^{K_2} = \phi_4^K - \mathbf{F}_4^K \\
 &-\hat{\mathbf{f}}_{25}^{K_2} \cdot \mathbf{n}_{25}^{K_2} + (\hat{\mathbf{f}}_{54}^{K_2} \cdot \mathbf{n}_{54}^{K_2} - \hat{\mathbf{f}}_{45}^{K_4} \cdot \mathbf{n}_{45}^{K_4}) \\
 &+ (\hat{\mathbf{f}}_{56}^{K_4} \cdot \mathbf{n}_{56}^{K_4} - \hat{\mathbf{f}}_{65}^{K_3} \cdot \mathbf{n}_{65}^{K_3}) + \hat{\mathbf{f}}_{53}^{K_3} \cdot \mathbf{n}_{53}^{K_3} = \phi_5^K - \mathbf{F}_5^K \\
 &-\hat{\mathbf{f}}_{36}^{K_3} \cdot \mathbf{n}_{36}^{K_3} + (\hat{\mathbf{f}}_{65}^{K_3} \cdot \mathbf{n}_{65}^{K_3} - \hat{\mathbf{f}}_{56}^{K_4} \cdot \mathbf{n}_{56}^{K_4}) \\
 &+ (\hat{\mathbf{f}}_{64}^{K_4} \cdot \mathbf{n}_{64}^{K_4} - \hat{\mathbf{f}}_{46}^{K_1} \cdot \mathbf{n}_{46}^{K_1}) + \hat{\mathbf{f}}_{61}^{K_1} \cdot \mathbf{n}_{61}^{K_1} = \phi_6^K - \mathbf{F}_6^K \quad (52)
 \end{aligned}$$

having set

$$\mathbf{F}_i^K = \int_{C_i \cap \partial K} \mathbf{f}(\mathbf{w}_h) \cdot \mathbf{n} \, d\Gamma$$

with \mathbf{w}_h the \mathbb{P} finite element solution, and with the obvious relation

$$\sum_{K \in K_i} \mathbf{F}_i^K = 0$$

due the continuity of the flux. If now we set (to simplify the notation)

$$\begin{aligned}
 \hat{\mathbf{f}}_{14} &:= \hat{\mathbf{f}}_{14}^{K_1} \cdot \mathbf{n}_{14}^{K_1}, & \hat{\mathbf{f}}_{61} &:= \hat{\mathbf{f}}_{61}^{K_1} \cdot \mathbf{n}_{61}^{K_1} \\
 \hat{\mathbf{f}}_{64} &:= \hat{\mathbf{f}}_{64}^{K_4} \cdot \mathbf{n}_{64}^{K_4} - \hat{\mathbf{f}}_{46}^{K_1} \cdot \mathbf{n}_{46}^{K_1}, & \hat{\mathbf{f}}_{42} &:= \hat{\mathbf{f}}_{42}^{K_2} \cdot \mathbf{n}_{42}^{K_2} \\
 \hat{\mathbf{f}}_{25} &:= \hat{\mathbf{f}}_{25}^{K_2} \cdot \mathbf{n}_{25}^{K_2}, & \hat{\mathbf{f}}_{45} &:= \hat{\mathbf{f}}_{45}^{K_4} \cdot \mathbf{n}_{45}^{K_4} - \hat{\mathbf{f}}_{54}^{K_2} \cdot \mathbf{n}_{54}^{K_2} \\
 \hat{\mathbf{f}}_{53} &:= \hat{\mathbf{f}}_{53}^{K_3} \cdot \mathbf{n}_{53}^{K_3}, & \hat{\mathbf{f}}_{36} &:= \hat{\mathbf{f}}_{36}^{K_3} \cdot \mathbf{n}_{36}^{K_3} \\
 \hat{\mathbf{f}}_{56} &:= \hat{\mathbf{f}}_{56}^{K_4} \cdot \mathbf{n}_{56}^{K_4} - \hat{\mathbf{f}}_{65}^{K_3} \cdot \mathbf{n}_{65}^{K_3}
 \end{aligned}$$

and $\Psi_i^K = \phi_i^K - \mathbf{F}_i^K$, we obtain

$$\begin{aligned}
 \hat{\mathbf{f}}_{14} - \hat{\mathbf{f}}_{61} &= \Psi_1^K \\
 -\hat{\mathbf{f}}_{42} + \hat{\mathbf{f}}_{25} &= \Psi_2^K \\
 -\hat{\mathbf{f}}_{53} + \hat{\mathbf{f}}_{36} &= \Psi_3^K \\
 -\hat{\mathbf{f}}_{14} - \hat{\mathbf{f}}_{64} + \hat{\mathbf{f}}_{45} + \hat{\mathbf{f}}_{42} &= \Psi_4^K \\
 -\hat{\mathbf{f}}_{25} - \hat{\mathbf{f}}_{45} + \hat{\mathbf{f}}_{56} + \hat{\mathbf{f}}_{53} &= \Psi_5^K \\
 -\hat{\mathbf{f}}_{36} - \hat{\mathbf{f}}_{56} + \hat{\mathbf{f}}_{64} + \hat{\mathbf{f}}_{61} &= \Psi_6^K \quad (53)
 \end{aligned}$$

System (53) has a very neat interpretation: the sub-triangulation of Figure 4 defines a triangulation of the element K associated with its degrees of freedom. For any edge between two degrees of freedom, say $[i, j]$, we look for fluxes $\hat{\mathbf{f}}_{ij}$ satisfying (53), with the constraint $\hat{\mathbf{f}}_{ij} + \hat{\mathbf{f}}_{ji} = 0$.

In the \mathbb{P} case, one can easily show that the matrix associated with (53) has rank 5, which can be used to obtain a definition of the equivalent finite volume fluxes as was done for linear elements. We only sketch the generalization to P^k elements, which relies on the following main elements:

- Construct a triangulation of K whose vertices are the degrees of freedom of the interpolation.
- Associate with this sub-triangulation a dual tessellation to be used to define local conservation equations:
- set

$$\Psi_i^K = \phi_i^K - \int_{C_i \cap \partial K} \mathbf{f}(\mathbf{w}_h) \cdot \mathbf{n} \, d\Gamma$$

- Write the equations for conservative edge fluxes $\hat{\mathbf{f}}_{ij}$. Assemble a linear system for a subset \mathcal{F} of the ordered couples (i, j) associated with the edges of the sub-triangulation, with \mathcal{F} containing either (i, j) , or (j, i) for any two fixed nodes. We have then
 1. the matrix coefficients of the linear system are

$$\theta_{ij} = \begin{cases} 0, & (i, j) \text{ is not an edge} \\ 1, & (i, j) \text{ is an edge and } (i, j) \in \mathcal{F} \\ -1, & (i, j) \text{ is an edge and } (i, j) \notin \mathcal{F}; \end{cases}$$

2. the i th right-hand side of the system is equal to Ψ_i^K ;
3. the rank of the system matrix is equal to $n_{\text{dof}} - 1$.

Our analysis can be also generalized to three space dimensions, and to other types of elements, since it only relies on the possibility of constructing a set of connected dual cells, which is possible for any mesh. This analysis shows that *whatever the type of element, the approximation of the residual distribution spatial discretization can be reformulated by means of a finite volume approximation defined by a multidimensional flux function of n_{dof} states. Hence, all continuous finite element schemes admitting a residual distribution reformulation are locally conservative.*

3.1.5 WENO-RD and bridge with DG

One can slightly extend the RD formalism. In what was written above, the main assumption is that the approximation is globally continuous. This assumption can be relaxed. Assume that, as for DG, the trial function space is made of functions that are polynomials on each elements, but we relax the continuity assumption. This problem has been

studied in Abgrall and Shu (2009), Abgrall (2010), Hubbard (2008). In each element K , we assume that we have N_K degrees of freedom, say $\{i_j, j = 1, N_K\}$, and assume that we have constructed residuals $\Phi_{i_j}^K(u^h)$. The conservation relation must be relaxed into

$$\sum_{j=1}^{N_K} \Phi_{i_j}^K(\mathbf{w}^h) = \int_{\partial K} \hat{f}(\mathbf{w}^{h,+}, \mathbf{w}^{h,-}, \mathbf{n}) d\partial K \quad (54)$$

where \hat{f} is a consistent numerical flux, and $\mathbf{w}^{h,+}, \mathbf{w}^{h,-}$ are the states on the two sides of the faces that make ∂K . In Abgrall (2010), it is shown how to reformulate a DG method with \mathbb{P}^1 elements in this framework. Though limited to \mathbb{P}^1 element, this approach can be easily extended to higher order of approximation, see also Abgrall (2010) for a more systematic (than Abgrall and Shu, 2009) technique. In the discontinuous case, a simple variant can be found, see Abgrall (2010). This remark makes it possible to use the technique that we describe in Section 3.1.7.

A completely different approach has been pursued in Chou and Shu (2006). Starting from a finite difference-like grid and using WENO reconstruction, these authors have been successful in developing an RD-like approximation for hyperbolic problems.

3.1.6 A general Lax–Wendroff result

One of the key constraints an RD scheme must fulfill is that, for any element or face, the sum of the sub-residual must be equal to the total residuals; these are the conditions (22) and (26).³ These conditions guarantee a Lax–Wendroff like result, see Abgrall and Roe (2003). More precisely, we assume the following:

Assumption 1. The mesh \mathcal{T}_h is conformal and regular. By regular we mean that all elements are roughly of the same size; more precisely, there exist constants C_1 and C_2 such that for any element

$$K, \quad C_1 \leq \sup_{K \in \mathcal{T}_h} \frac{h^2}{|K|} \leq C_2.$$

We introduce the following spaces:

$$V_h^k = \{\mathbf{v}_h \in C^0(\mathbb{R}^d)^p; \mathbf{v}_h|_K \text{ polynomial of degree } k,$$

$$\forall K \in \mathcal{T}_h\}$$

$$X^h = \{\mathbf{v}_h; v_h|_C \text{ constant} \in \mathbb{R}^p, \forall C \in \mathcal{C}_h\}$$

Here, $\mathbf{f}|_K$ denotes the restriction of \mathbf{f} to K . The second assumption is on the residuals.

Assumption 2. Let \mathcal{T}_h be a triangulation satisfying Assumption 1. For any $C \in \mathbb{R}^+$, there exists $C'(C, \mathcal{T}_h) \in \mathbb{R}^+$, which depends only on C and \mathcal{T}_h such that for any $\mathbf{w} \in X^h$, with $\|\mathbf{w}\|_{L^\infty(\mathbb{R}^2)} \leq C$ we have

$$\forall K, \forall i, \|\Phi_i^K\| \leq C'(C, \mathcal{T}_h) h \sum_{j \in K} \|\mathbf{w}(j) - \mathbf{w}(i)\| \quad (55)$$

We assume that the residual $\Phi_i^{K'}$ and the numerical solution satisfy the following conditions:

Assumption 3. There exists an approximation \mathbf{f}^h of the flux \mathbf{f} such that

- (i) $\forall \mathbf{w}^h \in X^h, \Phi^K := \int_K \operatorname{div} \mathbf{f}^h(\mathbf{w}^h) dx = \sum_{i \in K} \Phi_i^K(\mathbf{w}^h),$
- (ii) $\forall \mathbf{w}^h \in X^h, \forall K_1, K_2$ neighbors,

$$\mathbf{f}^h(\mathbf{w}^h)|_{K_1} \cdot \vec{n} = \mathbf{f}^h(\mathbf{w}^h)|_{K_2} \cdot \vec{n} \quad a.e. \text{ on } K_1 \cap K_2$$

where \vec{n} is a normal of $K_1 \cap K_2$.

- (iii) For any $C > 0$, there exists $C'(C)$ such that for any $\mathbf{w}^h \in X^h$ with $\|\mathbf{w}^h\|_{L^\infty(\mathbb{R}^2)} \leq C$, one has for $K \in \mathcal{T}_h$ and $\mathbf{f}_K^h = \mathbf{f}_K^h$, $\|\operatorname{div} \mathbf{f}_K^h(u^h)\| \leq \frac{C'}{h} \sum_{i,j} \|\mathbf{w}_i^h - \mathbf{w}_j^h\|$ a.e. on K .
- (iv) For any sequence $(\mathbf{w}^h)_h$ bounded in $L^\infty(\mathbb{R}^2 \times \mathbb{R}^+)^p$ independent of h and convergent in $L^2_{loc}(\mathbb{R}^2 \times \mathbb{R}^+)^p$ to \mathbf{w} , we have

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

We have the following result:

Theorem 1. Let be $\mathbf{w}_0 \in L^\infty(\mathbb{R}^d)^p$, and \mathbf{w}^h the approximation given by

$$\mathbf{w}_i^{n+1} = \mathbf{w}_i^n - \frac{\Delta}{|C_i|} \left(\sum_{K \ni i} \phi_i^K(\mathbf{w}^{n+1}) + \sum_{F \cap \partial \Omega \ni i} \phi_i^F \right)$$

$$\mathbf{w}_i^0 = \mathbf{w}_0(i)$$

We assume that the scheme satisfies the Assumptions 2 and 3. We also assume there exists a constant C that depends only on C_1 , C_2 , and u_0 and a function $\mathbf{w} \in (L^2(\mathbb{R}^d \times \mathbb{R}^+))^p$ such that

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

$$\lim_h \|\mathbf{w} - \mathbf{w}_h\|_{L^2_{loc}(\mathbb{R}^d \times \mathbb{R}^+)^p} = 0$$

Then, \mathbf{w} is a weak solution of

$$\frac{\partial \mathbf{w}}{\partial t} + \operatorname{div} \mathbf{f}(\mathbf{w}) = 0$$

$$\mathbf{w}(\mathbf{x}, 0) = \mathbf{w}_0(\mathbf{x})$$

The proof is given in Abgrall and Roe (2003).

3.1.7 Construction of nonclassical high-order schemes

Well-posed linear schemes

The simplest method is where the element residuals are split in a symmetric manner, as, for example

$$\phi_i^K = \frac{1}{n_{\text{dof}}} \phi^K$$

This definition verifies trivially all the accuracy criteria and the conditions for the Lax–Wendroff theorem. Nevertheless, it is flawed by the existence of a spurious mode as discussed in Section 3.1.2, which is a clear symptom of a lack of stability. An example of a stabilized method is the Lax–Friedrich’s type distribution

$$\phi_i^{\text{LF}} = \frac{1}{n_{\text{dof}}} \phi^K + \alpha_K(\mathbf{w}_i - \bar{\mathbf{w}}_K) \quad (56)$$

with $\bar{\mathbf{w}}_K$ the arithmetic average of the solution values in K . In the scalar case, we can easily prove the stability of this method in both L^2 and L^∞ norms (cf. Section 3.1.3). This method does not verify the accuracy conditions of Section 3.2.1.

To obtain a stable high-order method, we can follow the ideas of Abgrall *et al.* (2009, 2011), and add to an unstable high-order method a streamline dissipation term

$$\phi_i^K = \beta_i^K \phi^K + \theta_K \int_K (\nabla_{\mathbf{w}} \mathbf{f} \cdot \nabla \varphi_i) \mathcal{T}_K(\nabla_{\mathbf{w}} \mathbf{f} \cdot \nabla \mathbf{w}_h) dx \quad (57)$$

with θ_K a scalar coefficient, and where, for generality, we have replaced $1/n_{\text{dof}}$ by a generic bounded distribution coefficient. Following Abgrall *et al.* (2009, 2011), we seek for rules to define the term $\theta_K \mathcal{T}_K$ for scalar advection. We start by recasting the method obtained with (57) as

$$- \int_{\Omega_h} u_h \vec{a} \cdot \varphi_i dx + \sum_{K \in \mathcal{K}_i} \int_K (\beta_i^K - \varphi_i)(\vec{a} \cdot \nabla u_h) dx$$

$$+ \sum_{K \in \mathcal{K}_i} \theta_K \int_K (\vec{a} \cdot \nabla \varphi_i) \mathcal{T}_K(\vec{a} \cdot \nabla u_h) dx$$

$$= \text{boundary cond.s}$$

We can associate to this method the bilinear form

$$a(v^h, u_h) + \sum_K b_K(v^h, u_h) = l_{bc,s}(v^h)$$

where

$$a(v^h, u_h) = a^{\text{Galerkin}}(v^h, u_h) + \sum_K a_K(v^h, u_h)$$

with

$$a_K(v^h, u_h) = \int_K (v_K^\beta - v^h)(\vec{a} \cdot \nabla u_h) \, d\mathbf{x}$$

and where $b_K(v^h, u_h)$ is the streamline dissipation term

$$b_K(v^h, u_h) = \int_K (\vec{a} \cdot \nabla v^h) \theta_K \mathcal{T}_K(\vec{a} \cdot \nabla u_h) \, d\mathbf{x}$$

For simplicity, and following the ideas of Abgrall *et al.* (2009), we now express the increment $v_K^\beta - v^h$ as a function of ∇v^h and of a (scheme-dependent) element length h_K^β and direction ξ_K^β :

$$\begin{aligned} a_K(v^h, u_h) + b_K(v^h, u_h) &= \int_K (\xi_K^\beta \cdot \nabla v^h) h_K^\beta (\vec{a} \cdot \nabla u_h) \, d\mathbf{x} \\ &+ \int_K (\vec{a} \cdot \nabla v^h) \theta_K \mathcal{T}_K(\vec{a} \cdot \nabla u_h) \, d\mathbf{x} \end{aligned}$$

Finally, we want to define the coefficient $\theta_K \mathcal{T}_K$ such that

$$\begin{aligned} a_K(v^h, u_h) + b_K(u_h, u_h) &= \int_K (\xi_K^\beta \cdot \nabla u_h) h_K^\beta (\vec{a} \cdot \nabla u_h) \, d\mathbf{x} \\ &+ \int_K \theta_K \mathcal{T}_K (\vec{a} \cdot \nabla u_h)^2 \, d\mathbf{x} \geq 0 \end{aligned} \quad (58)$$

In particular, to have (58) satisfied in practice, we consider the fully discrete evaluation of the streamline dissipation term

$$\begin{aligned} &\int_K (\vec{a} \cdot \nabla \varphi_i) \theta_K \mathcal{T}_K (\vec{a} \cdot \nabla u_h) \, d\mathbf{x} \\ &\approx \theta_K \mathcal{T}_K |K| \sum_{x_{\text{quad}}} \omega_{\text{quad}} (\vec{a} \cdot \nabla \varphi_i(x_{\text{quad}})) \\ &\quad \times (\vec{a} \cdot \nabla u_h(x_{\text{quad}})) \end{aligned} \quad (59)$$

where we have assumed for simplicity a constant value of \mathcal{T}_K over each element. We seek now guidelines to choose a quadrature formula.

A necessary condition to have (58) is that the quadratic form

$$q_K(u_h) := |K| \sum_{x_{\text{quad}}} \omega_{\text{quad}} (\vec{a} \cdot \nabla u_h(x_{\text{quad}}))^2$$

must be positive whenever $\vec{a} \cdot \nabla u_h \neq 0$. A sufficient condition for this to be true is that

$$\text{if } \vec{a} \cdot \nabla u_h(x_{\text{quad}}) = 0 \, \forall \, x_{\text{quad}} \text{ then } \vec{a} \cdot \nabla u_h = 0 \quad (60)$$

In this case, we can find positive bounded constants such that

$$C_{1,q} q_K(u_h) \leq \int_K (\vec{a} \cdot \nabla u_h)^2 \, d\mathbf{x} \leq C_{2,q} q_K(u_h)$$

Since $V_h = \text{span}\{\varphi_i\}$ is a finite-dimensional space, the discrete quantity

$$\mathcal{Q}(u_h) = \sum_K q_K(u_h)$$

defines on V_h a norm equivalent to $u_h \mapsto \int_\Omega (\vec{a} \cdot \nabla u_h)^2 \, d\mathbf{x}$. This allows us to prove the following result:

Proposition 5. (Quadrature of the streamline dissipation). *Independent of the values of the weights ω_{quad} , provided that the number of points used to evaluate (59) is large enough to guarantee (60), then we can find $(\theta_K \mathcal{T}_K)_0$ such that the scheme obtained with (57) is well posed whenever $\theta_K \mathcal{T}_K \geq (\theta_K \mathcal{T}_K)_0$.*

Proof. See Abgrall *et al.* (2009). □

In light of this analysis, the set of points used to evaluate (59) need not necessarily be a set of quadrature points, as the relevant condition is not to evaluate the streamline dissipation term exactly but to ensure (60). In this light, the term (59) can be seen as a sort of *filtering term*, allowing the removal of spurious modes and guarantee the well-posedness of the method. In particular, even for constant scalar advection, the number of points sufficient to have (60) is smaller than that of most quadrature/cubature formulas, providing an exact evaluation of the streamline dissipation term, and in any case simpler point values can be used, such as, for example, x_{dof} (cf. Abgrall *et al.*, 2009, 2011).

Another path to avoid the flaw associated with Proposition 3 is to “bring the distribution coefficient β under the integral”. The classical definition associated with SUPG (43) is one example of such a method. However, we can also provide similar generalizations of the so-called multidimensional upwind methods, which have constituted one of the elements of originality of the residual distribution approach (cf. Deconinck and Ricchiuto, 2007 and references therein).

In particular, we consider the method defined in the scalar case by

$$\phi_i^K = \int_K (\vec{a} \cdot \nabla \varphi_i)^+ \gamma_K (\vec{a} \cdot \nabla u_h) \, d\mathbf{x} \quad (61)$$

with $\gamma_K > 0$. If, without loss of generality, we consider the linear advection problem admitting a solution $u > 0$,⁴ we may assume that we seek a discrete solution $u_h \in V_h^+ = \{u_h \in \text{span}\{\varphi_i\} | u_h > 0\}$, and we can in this case write

$$\begin{aligned} \sum_{i \in \Omega_h} u_i \sum_{K \in K_i} \phi_i^K &= \sum_{i \in \Omega_h} \int_K (\vec{a} \cdot \nabla u_h)^+ \gamma_K \vec{a} \cdot \nabla u_h \, d\mathbf{x} \\ &= \sum_{i \in \Omega_h} \int_K (\vec{a} \cdot \nabla u_h)^+ \gamma_K (\vec{a} \cdot \nabla u_h)^+ \, d\mathbf{x} \geq 0 \end{aligned}$$

as $\vec{a} \cdot \nabla u_h = (\vec{a} \cdot \nabla u_h)^+ + (\vec{a} \cdot \nabla u_h)^-$, and $(\vec{a} \cdot \nabla u_h)^- = 0$. This shows that condition (35) is met, giving an indication of the well-posedness of the method, confirmed by all numerical evidence. Note that for the method to satisfy (39), we need to set

$$\gamma_K = \left(\sum_{j \in K} (\vec{a} \cdot \nabla \varphi_j)^+ \right)^{-1}$$

which can be shown to reduce in the P^1 case to

$$\phi_i^K = \beta_i^K \phi^K, \quad \beta_i^K = (\vec{a} \cdot \mathbf{n}_i)^+ / \left(\sum_{j \in K} (\vec{a} \cdot \mathbf{n}_j)^+ \right)$$

which is nothing but the well-known multidimensional upwind LDA scheme introduced by Roe, Deconinck, and coworkers in the 1990s (Deconinck and Ricchiuto, 2007). The generalization (61) is obtained by formally replacing the so-called upwind parameters $k_i = \vec{a} \cdot \mathbf{n}_i/2$ by the i th streamline component of the solution gradient $k_i = \vec{a} \cdot \nabla \varphi_i$ and by performing the distribution of the local residual instead of distributing the integrated cell residual ϕ^K . The extension to nonlinear problems is obtained by replacing in (61) the residual $\vec{a} \cdot \nabla u_h$ with $\nabla \cdot \mathbf{f}_h$, with \mathbf{f}_h a higher order polynomial, of at degree at least $k+1$ (one higher than the solution), built starting from the values of u_h . Refer to D'Angelo *et al.* (2015), Vymazal *et al.* (2015) for more details.

Non-oscillatory methods

The analysis of Section 3.1.3 constitutes the basic artillery used in the past years to construct methods allowing a non-oscillatory approximation of discontinuous solutions. The key element of these constructions is some low (first)-order linear scheme, satisfying (36). A typical example of such a scheme is given by (56). For this scheme,

and in the scalar case, we can readily prove that (36) holds in d space dimensions, as soon as $\alpha_K > h_K^{d-1} \|\nabla_{bbw} \mathbf{f}\|_{L^\infty(K)}$, with h_K a characteristic length scale of the element.

A neat way of producing a formally high-order method starting from (56) is to fabricate uniformly bounded distribution coefficients by applying a nonlinear mapping to the quantities $\phi_i^{\text{LF}}/\phi^K$. An example of such a mapping is the well-known ‘‘PSI limiter’’ (Deconinck and Ricchiuto, 2007)

$$\beta_i^{\text{LLF}} = \frac{(\phi_i^{\text{LF}}/\phi^K)^+}{\sum_{j \in K} (\phi_j^{\text{LF}}/\phi^K)^+} = \frac{(\phi_i^{\text{LF}} \phi^K)^+}{\sum_{j \in K} (\phi_j^{\text{LF}} \phi^K)^+} \quad (62)$$

For the corresponding scheme, one can easily show that

$$\begin{aligned} \phi_i^{\text{LLF}} &= \beta_i^{\text{LLF}} \phi^K = \gamma_i \phi_i^{\text{LF}}, \\ \gamma_i &= \frac{(\phi_i^{\text{LF}}/\phi^K)^+}{\sum_{j \in K} (\phi_j^{\text{LF}}/\phi^K)^+} \phi^K / \phi_i^{\text{LF}} \in [0, 1] \end{aligned} \quad (63)$$

Thus, this limited Lax–Friedrich distribution is by construction stable in the L^∞ norm. Unfortunately, we also know from Proposition 3 that this scheme will not in general converge, and anyway may not converge to the right solution. A cure to this problem has been suggested in Abgrall (2006), Abgrall *et al.* (2009, 2011), and consists in adding the filtering term (59) in smooth regions. The resulting method reads

$$\begin{aligned} \phi_i^{\text{LLFs}} &= \beta_i^{\text{LLF}} \phi^K + \theta(u_h) |K| \\ &\quad \times \sum_{x_{\text{quad}}} \omega_{\text{quad}} (\vec{a} \cdot \nabla \varphi_i(x_{\text{quad}})) \mathcal{T}_K (\vec{a} \cdot \nabla u_h(x_{\text{quad}})) \end{aligned} \quad (64)$$

where $\theta(u_h)$ is defined such that the conditions of Proposition 5 are met in smooth regions while $\theta < \mathcal{O}(h_K)$ in the vicinity of discontinuities. Practical definitions of this term can be found in Abgrall (2006), Abgrall *et al.* (2011), Ricchiuto and Bollermann (2009), Ricchiuto and Abgrall (2010). The extension of this construction to systems is performed by computing the limiter (62) either equation by equation or by a prior projection of the residuals on characteristic directions, and by replacing the advection vector in (64) by the flux Jacobian matrices. A common definition of the scaling matrix \mathcal{T}_K is

$$\mathcal{T}_K = |K| \left(\sum_v (\nabla_{\mathbf{w}} \mathbf{f}(\mathbf{w}_v) \cdot \nabla \varphi_v(\mathbf{x}_v))^+ \right)^{-1}$$

with v the vertices of element K .

An alternative construction consists in adding to a linear high-order and stable scheme a local amount of shock

capturing dissipation. This approach dates back a long way (Hughes and Mallet, 1986). In the framework of residual distribution schemes, it has been reformulated by means of a technique reminiscent of flux limiting in the finite volume context: the nonlinear blending of a linear high-order method with a linear low (first)-order positive coefficient one. For example, blending (56) with a high-order stabilized method would lead to

$$\begin{aligned} \phi_i^K &= \frac{1}{n_{\text{dof}}} \phi^K + \delta(\mathbf{w}_h) \alpha_K(\mathbf{w}_i - \bar{\mathbf{w}}_K) \\ &+ (\text{Id} - \delta(\mathbf{w}_h)) \theta_K |K| \sum_{x_{\text{quad}}} \omega_{\text{quad}} (\vec{a} \cdot \nabla \varphi_i(x_{\text{quad}})) \\ &\times \mathcal{T}_K (\vec{a} \cdot \nabla u_h(x_{\text{quad}})) \end{aligned} \quad (65)$$

where different forms of the stabilization are selected depending on whether \mathbf{w}_h is smooth, in which case $\delta(\mathbf{w}_h) \leq \mathcal{O}(h_K)$, or discontinuous, in which case $\text{Id} - \delta(\mathbf{w}_h) \leq \mathcal{O}(h_K)$. More involved constructions that consider replacing (56) in the blending by (63) have also been proposed, for example, in Ricchiuto (2015).

3.1.8 Handling source terms

The extension of the above framework to the approximation of solutions of

$$\text{div } \mathbf{f}(\mathbf{w}) + \mathbf{s}(\mathbf{w}, \mathbf{x}) = 0 \quad (66)$$

is based on the inclusion of the source in the redefinition of the local element residual, leading to the requirement

$$\sum_{i \in K} \phi_i^K = \oint_{\partial K} \mathbf{f}_h(\mathbf{w}_h) \cdot \mathbf{n} \, d\partial K + \int_K \mathbf{s}_h(\mathbf{w}_h, \mathbf{x}) \, d\mathbf{x} \quad (67)$$

All of the methods described earlier can be extended to this more general setting.

Interesting results can be obtained when \mathbf{s} depends on some given data, say a given field $f(\mathbf{x})$:

$$\mathbf{s}(\mathbf{w}, \mathbf{x}) = \mathbf{s}(\mathbf{w}, f(\mathbf{x}))$$

This is the case in some environmental applications (e.g., shallow-water equations), or when considering the solution of the differential problem on a manifold (see, e.g., Rossmanith *et al.*, 2004 and references therein). Such problems often embed some particular solutions that are characterized by the existence of a set of invariants $\mathbf{v} = \mathbf{v}(\mathbf{w}, f)$ constant throughout the spatial domain. Assuming a sufficient smoothness f of the solution and of the mapping

$(\mathbf{v}, f) \mapsto \mathbf{w}(\mathbf{v}, f)$ in this case, we can write

$$\text{div } \mathbf{f}(\mathbf{w}) = (\nabla_{\mathbf{w}} \mathbf{f} \nabla_{\mathbf{v}} \mathbf{w}) \nabla \mathbf{v} + \underbrace{(\nabla_{\mathbf{w}} \mathbf{f} \nabla_f \mathbf{w}) \nabla f}_{\Lambda(\mathbf{v}, f)}$$

Solutions characterized by the invariance relation $\mathbf{v} = \mathbf{v}_0 = c^t \forall \mathbf{x}$, satisfy (cf. (66))

$$\Lambda(\mathbf{v}_0, f) \nabla f + \mathbf{s}(\mathbf{v}_0, f) = 0 \quad (68)$$

An interesting result concerning this class of solutions for schemes defined by

$$\phi_i^K = \int_K \omega_i^K (\nabla \cdot \mathbf{f}_h(\mathbf{v}_h, f) + \mathbf{s}_h(\mathbf{v}_h, f)) \quad (69)$$

is thus based on a direct approximation of the invariant states, instead of the conserved variables \mathbf{w} . The following is shown in Ricchiuto (2011a,b, 2015):

Proposition 6. (Steady invariants and superconsistency). *Under standard regularity assumptions on the mesh, provided that (69) is true for some test function ω_i^K which is uniformly bounded w.r.t. h , \mathbf{v}_h , element residuals, and w.r.t. to the data of the problem, then for exact integration the scheme defined by (69) preserves exactly the equilibrium (68). For approximate integration, assuming that a flux quadrature exact for approximate polynomial fluxes of degree p_f is used, and a source quadrature exact for approximate polynomial sources of degree p_v , and assuming that $f \in H^{p+1}$ with $\nabla f \in H^p$, and $p > \min(p_f, p_v)$, then the scheme defined by (69) is super-consistent w.r.t. solutions characterized by (68), and, in particular, its consistency is of order $r = \min(p_f + 2, p_v + 3)$.*

This result shows the potential of residual framework considered here in guaranteeing the balance of the flux divergence with complex source terms. Similar and stronger results, especially on simple particular solutions encountered in shallow-water flows, have already been recalled in the previous chapter (Deconinck and Ricchiuto, 2007). Some applications of this property will be shown in the results section.

3.1.9 Handling viscous terms

When considering the advection diffusion equation (with $\nabla \cdot \vec{a} = 0$)

$$\vec{a} \cdot \nabla u - \nabla \cdot (\mathcal{K}(u) \cdot \nabla u) = 0 \quad (70)$$

with $\mathcal{K}(u)$ a positive semidefinite diffusion matrix coefficient, the first idea is to look at it as a standard conservation

relation with an enhanced flux, now

$$\mathbf{f}(u, \nabla u) = \vec{a}u - \mathcal{K}(u) \cdot \nabla u$$

to which, one could apply the same construction as before. However, there is a fundamental difference: if the approximation of the solution is sought to be piecewise polynomial and globally continuous, its gradient will still be piecewise polynomial but will not be globally continuous anymore. One of the fundamental requirements of the previous developments is that the flux on the boundary of the element is single-valued. This can no more be the case here unless something is done.

There are two ways of solving this issue. Both are similar to what is done in LDG methods. The first step is in each case to rewrite the partial differential equation into a, possibly hyperbolic, first-order system of partial differential equations (PDEs). For the two-dimensional advection–diffusion equation, setting $\mathcal{K} = \nu \text{Id}$, we consider the hyperbolic first-order system

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} + \vec{a} \cdot \nabla u &= \nu(\partial_x p + \partial_y q) \\ \frac{\partial p}{\partial t} &= \frac{1}{T_r}(\partial_x u - p) \\ \frac{\partial q}{\partial t} &= \frac{1}{T_r}(\partial_y u - q) \end{aligned} \quad (71)$$

where p and q are the gradient variables, and T_r is a relaxation time. At the steady state, the system (71) is equivalent to the original equation (70), independent of the parameter T_r , and p, q become equivalent to the derivatives of the unknown. The idea of reformulating a parabolic problem with second-order derivatives as a hyperbolic system such as, for example, (71), is not new, as it dates back to the work of Vernotte (1958), Cattaneo (1958) to study the heat equation. This idea has been efficiently exploited by Nishikawa and Mazaheri to construct schemes for the steady and time-dependent diffusion, advection–diffusion, and Navier–Stokes equations (see, e.g., Nishikawa, 2007, 2010a,b, 2011 and Mazaheri and Nishikawa, 2015, 2016 for recent formulations of residual distribution and DG based on this approach).

There are two ways to approach system (71). The first is to make use of its hyperbolicity, and reuse all the artillery already available. In this case, the overhead of having to introduce the gradient variables can be compensated by a careful design of the scheme that may guarantee the same accuracy for both the solution *and* its derivatives. This may have an impact on the computation of, for example, forces and heat fluxes and allow the use of coarser meshes to provide accurate values of these quantities. This is the path followed in Mazaheri and Nishikawa (2015, 2016), but the

demonstration of its feasibility for practical applications is still in progress.

Another way to exploit the system (71) has been suggested in Nishikawa (2010b, 2011), and developed from scalar advection diffusion up to laminar Navier–Stokes and Reynolds-averaged Navier–Stokes (RANS) equations in Abgrall *et al.* (2014), Abgrall and De Santis (2015), De Santis (2015). In this alternative approach, only the first discrete equation or u_h is kept. This is, of course, a discretization of (70), which, however, depends on values of p and q . These values are now replaced by an appropriate high-order *reconstruction* of the solution derivatives starting from u_h . Simple solutions are possible, as, for example, the use of simple arithmetic averages for the viscous fluxes on element boundaries (see, e.g., Abgrall *et al.*, 2013). However, these simple choices lead to suboptimal accuracy, mostly because one order of accuracy is lost in the evaluation of the gradient. In Abgrall *et al.* (2014), a systematic study of possible recovery methods (arithmetic average, least square, etc.) has been conducted, and the best solution is to take advantage of a local least-squares minimization algorithm and of the existence of super convergence points in the element. At these points, as put forward by Zienkiewicz and Zhu (1987), the gradient is approximated at full order. In the following, this reconstruction will be referred to as SPR-ZZ.

Example of a nonclassical scheme

We now show how to use these ideas to generalize the schemes of Section 3.1.8 for the solution of (70). The schemes obtained are those used in the numerical results we will discuss later.

So we start from scheme (57), assuming for simplicity $\beta_i^K = 1/n_{\text{dof}}$. If we assume to be in the purely diffusive case, we apply this scheme to system (71) and only look at the locally distributed residuals for the first equation we have:

$$\phi_i^{\text{fos}} = \frac{1}{n_{\text{dof}}} \phi^{K,\nu} + \int_K \tau_\nu \nabla \phi_i \cdot (\nabla u_h - (p_h, q_h))$$

where we have assumed the stabilization matrix $\mathcal{T}_K = \delta_\nu \text{Id}$, and where

$$\phi^{K,\nu} = - \oint_{\partial K} \nu(p_h, q_h) \cdot \mathbf{n}$$

Note also that the effect of the relaxation time T_r has been embedded in the δ_ν coefficient.

The trick is now to replace the nodal values of the gradients p and q by accurately reconstructed ones, which we obtain with the SPR-ZZ procedure recalled above. The important part is the definition of the total residual. From the Lax–Wendroff theorem in Section 3.1.7, however, we know that the numerical approximations of both these fluxes must be edge-continuous. The simplest way to achieve that is to

use for the viscous flux the finite element approximation based on the reconstructed nodal gradients. We denote this quantity by $\widetilde{\nabla u_h}$. So, for pure diffusion, the scheme is finally defined by

$$\phi_i^{K,v} = \frac{1}{n_{\text{dof}}} \phi^{K,v} + \int_K \delta_v \nabla \varphi_i \cdot (\nabla u_h - \widetilde{\nabla u_h})$$

where

$$\phi^{K,v} = - \oint_{\partial K} \nu \widetilde{\nabla u_h} \cdot \mathbf{n}$$

and with δ_v having the dimensions of a diffusion coefficient.

For advection diffusion, we can apply the same procedure. Starting with the total residual

$$\phi^K = \oint_{\partial K} (\vec{a} u_h - \nu \widetilde{\nabla u_h}) \cdot \mathbf{n} \, d\Omega$$

we can deduce from the first-order system formulation two types of regularization terms leading to local nodal residual⁵

$$\begin{aligned} \phi_i^K &= \frac{1}{n_{\text{dof}}} \phi^K + \int_K \tau_a \vec{a} \cdot \nabla \varphi_i (\vec{a} \cdot \nabla u_h - \nu \nabla \cdot \nabla u_h) \\ &\quad + \int_K \delta_v \nabla \varphi_i \cdot (\nabla u_h - \widetilde{\nabla u_h}) \end{aligned}$$

The optimal choice of the scaling parameters τ_a and δ_v has been shown to require some dependence on the elemental Re number $Re = \frac{\|\vec{a}\|_h}{\nu}$ (see, e.g., Abgrall *et al.*, 2014; Nishikawa, 2010a; Ricchiuto *et al.*, 2008 and references therein). This is taken into account by setting

$$\begin{aligned} \phi_i^K &= \frac{1}{n_{\text{dof}}} \phi^K \\ &\quad + \xi(Re) \int_K (\vec{a} \cdot \nabla \varphi_i) \tau (\vec{a} \cdot \nabla u_h - \nabla \cdot (\nu \nabla u_h)) \, d\Omega \\ &\quad + (1 - \xi(Re)) \int_K \frac{\nu \delta}{2} (\nabla u_h - \widetilde{\nabla u_h}) \cdot \nabla \varphi_i \, d\Omega \end{aligned} \quad (72)$$

where the function $\xi(Re)$ is such that $\xi(Re) \rightarrow 0$ in the diffusion limit ($Re \rightarrow 0$) and $\xi(Re) \rightarrow 1$ in the advection limit ($Re \rightarrow \infty$).

To account for nonsmooth solutions, one can use the same technique discussed in Section 3.1.8: replace the centered contribution by a nonlinear limited residual, and pre-multiply the stabilization terms by some smoothness sensor, so that the scheme can be generally written in the final general form

$$\begin{aligned} \phi_i^K &= \beta_i^K \phi^K \\ &\quad + \theta(u_h) \xi(Re) \int_K (\vec{a} \cdot \nabla \varphi_i) \tau (\vec{a} \cdot \nabla u_h - \nabla \cdot (\nu \nabla u_h)) \, d\Omega \\ &\quad + \theta(u_h) (1 - \xi(Re)) \int_K \frac{\nu \delta}{2} (\nabla u_h - \widetilde{\nabla u_h}) \cdot \nabla \varphi_i \, d\Omega \end{aligned} \quad (73)$$

where β_i^K is computed following the limiting procedure discussed in Section 3.1.8 in the nonsmooth case.

The numerical scheme obtained for the advection–diffusion scalar equation is then extended to the compressible Navier–Stokes equations. The governing equations read

$$\frac{\partial \mathbf{w}}{\partial t} + \nabla \cdot \mathbf{f}^a(\mathbf{w}) - \nabla \cdot \mathbf{f}^v(\mathbf{w}, \nabla \mathbf{w}) = 0$$

where \mathbf{w} and $\mathbf{f}^a(\mathbf{w})$ are the vector of the conservative variables and the advective flux function, respectively, as defined for the Euler equations, while $\mathbf{f}^v(\mathbf{w}, \nabla \mathbf{w}) = (\mathbf{f}_x^v, \mathbf{f}_y^v)^T$ is the viscous flux function

$$\begin{aligned} \mathbf{f}_x^v(\mathbf{w}, \nabla \mathbf{w}) &= \begin{pmatrix} 0 \\ \tau_{xx} \\ \tau_{xy} \\ u\tau_{xx} + v\tau_{xy} - q_x \end{pmatrix}, \\ \mathbf{f}_y^v(\mathbf{w}, \nabla \mathbf{w}) &= \begin{pmatrix} 0 \\ \tau_{xy} \\ \tau_{yy} \\ u\tau_{xy} + v\tau_{yy} - q_y \end{pmatrix} \end{aligned}$$

where

$$\begin{aligned} \tau_{xx} &= \mu \left(\frac{4}{3} \frac{\partial v_x}{\partial x} - \frac{2}{3} \frac{\partial v_y}{\partial y} \right), \\ \tau_{yy} &= \mu \left(\frac{4}{3} \frac{\partial v_y}{\partial y} - \frac{2}{3} \frac{\partial v_x}{\partial x} \right), \\ \tau_{xy} &= \tau_{yx} = \mu \left(\frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right) \end{aligned}$$

are the components of the stress tensor, with μ being the dynamic viscosity of the fluid, and q_x, q_y the components of the heat flux \mathbf{q} , which is defined as

$$\mathbf{q} = k \nabla T$$

where T is the temperature, and k is the thermal conductivity coefficient. It is well-known that the viscous flux function \mathbf{f}^v is homogeneous with respect to the gradient of the conservative variable $\nabla \mathbf{w}$

$$\mathbf{f}^v(\mathbf{w}, \nabla \mathbf{w}) = \mathbb{K}(\mathbf{w}) \nabla \mathbf{w}$$

with the homogeneity tensor $\mathbb{K}(\mathbf{w}) = \frac{\partial \mathbf{f}^v}{\partial \mathbf{w}}$.

Discretization of the Navier–Stokes equations is straightforward. The total residual on a generic element K is given

by

$$\Phi^K = \oint_{\partial K} (\mathbf{f}^a(\mathbf{w}) - \mathbb{K}(\mathbf{w}) \widetilde{\nabla \mathbf{w}}) \cdot \mathbf{n}$$

with $\widetilde{\nabla \mathbf{w}}$ being the reconstructed gradient of the conservative variables and the boundary integral is computed by the means of a quadrature rule. The total residual is first distributed to all the degrees of freedom (DOFs) of the element using the low-order Rusanov scheme, and subsequently the limitation procedure is applied to obtain a high-order residual, as described in Section 4.2.1. In the last step, the filtering term is added together with the dumping term acting for the viscous part. The complete scheme reads

$$\begin{aligned} \Phi_i^K &= \tilde{\Phi}_i^K + \xi(Re) \int_K (\mathbf{A} \cdot \nabla \varphi_i) \\ &\quad \times \boldsymbol{\tau} (\mathbf{A} \cdot \nabla \mathbf{w}_h - \nabla \cdot (\mathbb{K} \nabla \mathbf{w}_h)) \, d\Omega \\ &\quad + (1 - \xi(Re)) \int_E \frac{1}{2} \mathbb{K}(\nabla \mathbf{w}_h - \widetilde{\nabla \mathbf{w}_h}) \cdot \nabla \varphi_i \, d\Omega \end{aligned} \quad (74)$$

with $\tilde{\Phi}_i^K$ denoting the (unfiltered) centered or nonlinear distribution.

3.2 Time-dependent problems

In this section, we consider the approximation of time-dependent solutions to a system of conservation laws, reading

$$\partial_t \mathbf{w} + \nabla \cdot \mathbf{f}(\mathbf{w}) = 0 \quad \text{on} \quad \Omega \times [0, T_{\text{fin}}] \subset \mathbb{R}^d \times \mathbb{R}^+ \quad (75)$$

As shown in Struijs (1994), and then in Caraeni (2000), Caraeni and Fuchs (2002), and Maerz and Degrez (1996), Ferrante and Deconinck (1997), Abgrall and Mezine (2003), Ricchiuto *et al.* (2005) (cf. also the chapter Deconinck and Ricchiuto, 2007), to obtain high-order schemes for this case, one must carefully design a coupling between the stencil used to approximate the integral of the time derivative and the flux divergence. Some approaches to obtain this coupling are recalled, and a more general prototype is analyzed. The links with other methods are briefly recalled. The first part of the section is devoted to fully implicit methods. We then discuss a path allowing the construction of explicit approaches which do not require the inversion of a mass matrix, or for which this matrix reduces to the symmetric positive-definite Galerkin one.

Note that, compared to the classical stabilized finite element schemes (SUPG, GLS, etc.), here the status of residual distribution type methods is less advanced. Here we discuss some of the most interesting ideas toward generalizing the methods presented for the steady state.

Some research directions to push the limits of the existing constructions will be discussed later.

3.2.1 Implicit prototype for time-dependent solutions

We introduce the time-discretized version of (75) by means of an $(r + 1)$ th order time integration scheme

$$\Gamma^{n+1}(\mathbf{w}) = \sum_{i=0}^p \alpha_i \frac{\delta \mathbf{w}^{n+1-i}}{\Delta t} + \sum_{j=0}^q \theta_j \nabla \cdot \mathbf{f}^{n+1-j} \quad (76)$$

where $\Delta t = \min_n (t^{n+1} - t^n)$, with $\Delta t^{n+1} = t^{n+1} - t^n$, $\delta \mathbf{w}^{n+1} = \mathbf{w}^{n+1} - \mathbf{w}^n$, and $\mathbf{f}^{n+1-j} = \mathbf{f}^{n+1-j}(\mathbf{w}^{n+1-j})$, and the α_i and θ_j coefficients are given by a time integration scheme of choice. This may be a generic stage of a multistage method, or a multistep scheme. Space-time schemes can be embedded in the analysis that follows by appropriate definitions of the α_i s, and of the $\delta \mathbf{w}^{n+1-i}$ to embed eventually jumps in the time direction when using discontinuous in time space-time elements. An important assumption is that the time stepping verifies the conservation identity

$$\sum_{n=0}^N \sum_{i=0}^p \alpha_i \delta \mathbf{w}^{n+1-i} = \mathbf{w}^N - \mathbf{w}^0 = \mathbf{w}(T_{\text{fin}}) - \mathbf{w}_0 \quad (77)$$

We set on every $K \in \Omega_h$

$$\begin{aligned} \Phi^K &= \int_K \Gamma^{n+1}(\mathbf{w}_h) \\ &= \int_K \left(\sum_{i=0}^p \alpha_i \frac{\delta \mathbf{w}_h^{n+1-i}}{\Delta t} + \sum_{j=0}^q \theta_j \nabla \cdot \mathbf{f}_h^{n+1-j} \right) \end{aligned} \quad (78)$$

with \mathbf{w}_h and \mathbf{f}_h continuous finite element polynomial approximations of degree k and (at least) k , respectively. Similarly, on each boundary face f we set

$$\phi^f = \int_f \sum_{j=0}^q \theta_j (\hat{\mathbf{g}} - \mathbf{f}_h)^{n+1-j} \cdot \vec{n} \quad (79)$$

with $\hat{\mathbf{g}}$ being a numerical flux consistent with the BCs.

Similar to the previous sections, we consider the scheme that computes \mathbf{w}_h as the solution of

$$\sum_{K \in K_i} \Phi_i^K + \sum_{f \in F_i} \phi_i^f = 0 \quad (80)$$

where $\forall K$ and $\forall f$

$$\sum_{j \in K} \Phi_j^K = \Phi^K \quad \text{and} \quad \sum_{j \in f} \phi_j^f = \phi^f \quad (81)$$

Consistency analysis

To begin with, we generalize the consistency conditions. To simplify the notation, we consider the scalar case and neglect the boundary conditions, which can be easily embedded in the spatial operator as shown. We will assume some classical regularity properties for the mesh and the time-stepping strategy, namely

$$C_0 \leq \sup_{K \in \Omega_h} \frac{h^2}{|K|} \leq C_1, \quad C'_0 \leq \frac{\Delta t}{h} \leq C'_1 \quad (82)$$

Now, let $\mathbf{w} \in C^{l+1}$ be an exact classical solution of (75), with $l \geq \max(r, k)$, such that

$$\begin{aligned} & \sum_{i=0}^p \alpha_i \frac{\delta \mathbf{w}^{n+1-i}}{\Delta t} + \sum_{j=0}^q \theta_j \nabla \cdot \mathbf{f}^{n+1-j} \\ &= \partial_t \mathbf{w} + \nabla \cdot \mathbf{f} + \mathcal{O}(\Delta t^{r+1}) \end{aligned} \quad (83)$$

We denote by \mathbf{w}_h^m the k th degree continuous finite element projection/interpolation of \mathbf{w}^m .

Consider now $\psi \in C_0^1(\Omega \times [0, T_{\text{fin}}])$, a smooth test function with $\psi|_{\partial\Omega} = 0$. Let ψ_h be its k th degree polynomial finite element projection/interpolation, with ψ_i^n the corresponding values at the chosen degrees of freedom. It is also assumed that (Ciarlet and Raviart, 1972; Ern and Guermond, 2004) there exist constants C''_0, C''_1, C_2 such that

$$\begin{aligned} \|\partial_t \psi_h\|_{L^\infty(\Omega_h)} &\leq C''_0 \|\psi_h(t + \Delta t) - \psi_h(t)\|_{L^\infty(\Omega_h)} \leq C''_0 \Delta t \\ \|\psi_h\|_{L^\infty(\Omega_h)} &\leq C''_1 |\psi_i - \psi_j| \leq \|\nabla \psi_h\|_{L^\infty(\Omega_h)} h \leq C_2 h \end{aligned} \quad (84)$$

We define the following truncation error for scheme (80):

$$\begin{aligned} \epsilon(\mathbf{w}_h, \psi) &:= \sum_{n=0}^N \sum_{i \in \Omega_h} \Delta t^{n+1} \psi_i^{n+1} \sum_{K \in K_i} \Phi_i^K(\mathbf{w}_h) \\ &= \sum_{n=0}^N \sum_{K \in \Omega_h} \sum_{i \in K} \int_{t^n}^{t^{n+1}} \psi_i^{n+1} \Phi_i^K(\mathbf{w}_h) \end{aligned} \quad (85)$$

We introduce the Galerkin splitting in space

$$\Phi_i^G = \int_K \varphi_i \Gamma^{n+1}$$

and note that

$$\sum_{j \in K} (\Phi_j^K - \Phi_j^G) = 0$$

This allows us to recast the error as

$$\epsilon(\mathbf{w}_h, \psi) = \sum_{n=0}^N \int_{t^n}^{t^{n+1}} \left\{ \int_{\Omega_h} \psi_h^{n+1} \Gamma^{n+1}(\mathbf{w}_h) \right.$$

$$\left. + \frac{1}{C_K} \sum_{K \in \Omega_h} \sum_{i,j \in K} (\psi_i - \psi_j)(\Phi_i^K - \Phi_i^G) \right\} \quad (86)$$

Multiplying (83) by ψ_h and integrating over space and time, we can get

$$\begin{aligned} & \sum_{n=0}^N \int_{t^n}^{t^{n+1}} \int_{\Omega_h} \psi_h^{n+1} \Gamma^{n+1}(\mathbf{w}) \\ &= \sum_{n=0}^N \Delta t \mathcal{O}(\Delta t^{r+1}) = \mathcal{O}(\Delta t^{r+1}) \end{aligned}$$

So the error can be estimated as

$$\epsilon(\mathbf{w}_h, \psi) = \text{I} + \text{II} + \text{III} + \mathcal{O}(\Delta t^{r+1})$$

$$\begin{aligned} \text{I} &= \sum_{n=0}^N \int_{t^n}^{t^{n+1}} \int_{\Omega_h} \psi_h^{n+1} \sum_{i=0}^p \alpha_i \frac{\delta(\mathbf{w}_h - \mathbf{w})^{n+1-i}}{\Delta t} \\ \text{II} &= \sum_{n=0}^N \int_{t^n}^{t^{n+1}} \sum_{j=0}^q \int_{\Omega_h} \psi_h^{n+1} \nabla \cdot (\mathbf{f}_h - \mathbf{f})^{n+1-j} \\ \text{III} &= \sum_{n=0}^N \int_{t^n}^{t^{n+1}} \frac{1}{C_K} \sum_{K \in \Omega_h} \sum_{i,j \in K} (\psi_i - \psi_j)(\Phi_i^K - \Phi_i^G) \end{aligned}$$

By estimating each of the terms, we obtain the conditions of the cell and boundary splittings, allowing us to preserve the $\mathcal{O}(\Delta t^{r+1})$ appearing on the right-hand side. This is readily done by using the hypotheses on the regularity of u and standard interpolation results (Ciarlet and Raviart, 1972; Ern and Guermond, 2004). In particular, for term I, we can use hypothesis (77) to write

$$\begin{aligned} \text{I} &= \sum_{n=0}^N \int_{t^n}^{t^{n+1}} \int_{\Omega_h} \sum_{i=0}^p \alpha_i \frac{\delta(\psi_h \mathbf{w}_h - \psi_h \mathbf{w})^{n+1-i}}{\Delta t} \\ &+ \sum_{n=0}^N \int_{t^n}^{t^{n+1}} \int_{\Omega_h} \sum_{i=0}^p \alpha_i (\psi_h^{n+1} - \psi_h^{n-i+1/2}) \\ &\times \frac{\delta(\mathbf{w}_h - u)^{n+1-i}}{\Delta t} \\ &- \sum_{n=0}^N \int_{t^n}^{t^{n+1}} \int_{\Omega_h} (\mathbf{w}_h - u)^{n-i+1/2} \sum_{i=0}^p \alpha_i \frac{\delta(\psi_h)^{n+1-i}}{\Delta t} \\ &= \int_{\Omega_h} (\psi_h(\mathbf{w}_h - u))(T_{\text{fin}}) - \int_{\Omega_h} (\psi_h(\mathbf{w}_h - u))_0 \\ &+ \sum_{n=0}^N \int_{t^n}^{t^{n+1}} \int_{\Omega_h} \sum_{i=0}^p \alpha_i (\psi_h^{n+1} - \psi_h^{n-i+1/2}) \\ &\times \frac{\delta(\mathbf{w}_h - u)^{n+1-i}}{\Delta t} \end{aligned}$$

$$- \sum_{n=0}^N \int_{t^n}^{t^{n+1}} \int_{\Omega_h} (\mathbf{w}_h - u)^{n-i+1/2} \sum_{i=0}^p \alpha_i \frac{\delta(\psi_h)^{n+1-i}}{\Delta t}$$

Using (84), and the regularity of u , we can now bound this term as

$$\begin{aligned} |I| &= \mathcal{O}(h^{k+1}) + C \frac{T_{\text{fin}}}{\Delta t} \Delta t \mathcal{O}(h^{k+1}) C_0'' \sup_{i=1,p} |\alpha_i| \\ &= \mathcal{O}(h^{k+1}) \end{aligned}$$

The analysis of the remaining terms is practically identical to the one of Section 3.2.1, and hence omitted for brevity (the interested reader can refer to Ricchiuto (2011a) for details). The final result is the following:

Proposition 7. (Accuracy of RD, unsteady case). *Under Assumption (82) on time stepping, given a $(k+1)$ th order continuous polynomial approximation of the unknown and of the fluxes, and an $(r+1)$ th order accurate time integration scheme, scheme (80) verifies the truncation error estimate*

$$|\epsilon(\mathbf{w}_h, \psi)| \leq \mathcal{O}(h^{p+1}), \quad p = \min(k, r)$$

provided that

$$\sup_{K \in \Omega_h} \sup_{i \in K} |\Phi_i^K(\mathbf{w}_h)| = \mathcal{O}(h^{p+d}) \quad (87)$$

whenever \mathbf{w}_h is the finite element projection/interpolation of a smooth exact solution. In this case we say that the scheme is $(p+1)$ th order accurate.

Moreover, we have the following estimate:

Lemma 1. (Consistency estimate, time-dependent case) *Under the hypotheses of Proposition 7, the following consistency estimates hold:*

$$\Gamma^{n+1}(\mathbf{w}_h) = \mathcal{O}(h^k) + \mathcal{O}(\Delta t^{r+1}), \quad \Phi^K(\mathbf{w}_h) = \mathcal{O}(h^{p+d}) \quad (88)$$

Proof. The proof is easily obtained by considering that, due to (83),

$$\Gamma^{n+1}(\mathbf{w}_h) = \mathcal{O}(\Delta t^{r+1}) + \Gamma^{n+1}(\mathbf{w}_h) - \Gamma^{n+1}(\mathbf{w})$$

By its definition, and under the hypotheses made, one can easily check that $\Gamma^{n+1}(\mathbf{w}_h) - \Gamma^{n+1}(\mathbf{w}) = \mathcal{O}(h^k)$. The estimate on $\Phi^K(\mathbf{w}_h)$ is trivially obtained upon integration of Γ^{n+1} . \square

As a consequence, we have the following corollary:

Corollary 1. (High-order residual schemes) *Under the hypotheses of Proposition 7, a sufficient condition for a*

scheme of the form (80) to be $(p+1)$ th order accurate is that there exists a test function ω_i uniformly bounded w.r.t. h , \mathbf{w}_h , $\Gamma^{n+1}(\mathbf{w}_h)$, and w.r.t. the data of the problem, such that

$$\Phi_i^K(\mathbf{w}_h) = \int_K \omega_i \Gamma^{n+1}(\mathbf{w}_h) \quad (89)$$

Examples of implicit high-order schemes

Typical examples of high-order methods are obtained with the natural extension to the time-dependent case of SUPG-type methods (see e.g., Hughes and Tezduyar, 1984; Shakib and Hughes, 1991; Hughes *et al.*, 2004; Chalot and Normand, 2010 and references therein). Some notable examples of less classical high-order schemes exploit (89) with $\omega_i = \beta_i^K$ constant per element. The first of such *accuracy-preserving* schemes can be found in the work of Caraeni (2000), Caraeni and Fuchs (2002), up to third order of accuracy for the Navier–Stokes equations, and more recently in Rossiello *et al.* (2007) where the third-order scheme of Caraeni was blended with a monotone one via a flux-corrected transport (FCT) procedure to provide oscillation-free high-order solutions of the compressible Euler equations.

Other nonclassical constructions have tried to exploit the similarities between stabilized finite elements and RD methods with constant distribution coefficients. The objective of these works is to find clever definitions of mass matrices/test functions guaranteeing the satisfaction of (89). This was the initial idea behind the work of Maerz and Ferrante at the von Karman Institute Maerz and Degrez (1996), Ferrante and Deconinck (1997), which was later pursued first in Abgrall and Mezine (2003), Mezine (2002), and then in Mezine *et al.* (2003), Ricchiuto *et al.* (2004, 2005), Ricchiuto and Abgrall (2006), Ricchiuto and Bollermann (2009) (see also Deconinck and Ricchiuto, 2007). This has provided interesting results, but so far only for second-order methods.

Finally, examples of space-time RD schemes up to third order are discussed in Ricchiuto *et al.* (2003), Koloszár *et al.* (2011) with monotonicity-preserving extensions discussed in Abgrall and Mezine (2003), Abgrall *et al.* (2005), Ricchiuto *et al.* (2005), and Hubbard and Ricchiuto (2011).

All these works use almost exactly the same techniques developed for steady problems, treating the time derivative either as a source term or as an additional space direction. The potential of these methods is that they may allow preservation of monotonicity unconditionally w.r.t. the time step size, which is very interesting when considering local mesh refinement (see, e.g., Hubbard and Ricchiuto, 2011; Sarmany *et al.*, 2013), or stiff problems

(viscous terms, chemical reactions, etc.). The drawback of this formulation is that the nonlinear stabilization involved depends on the unknown solution at the new time level, thus ruling out *a priori* simpler, genuinely explicit time-marching methods, often preferred in the hyperbolic case. Some exceptions to this rule exist, such as, for example, Taylor–Galerkin, and Lax–Wendroff type methods, which can also be recast in a residual distribution formalism (see, e.g., Hubbard and Roe, 2000; Ricchiuto and Deconinck, 1999; Rossiello *et al.*, 2009, and Deconinck and Ricchiuto, 2007).

A technique to side step this issue and construct some nonclassical genuinely explicit monotone and high-order residual methods is discussed in the next section.

3.2.2 *Genuinely explicit time advancement for residual methods*

The main idea here is to start from a prototype high-order scheme, which we will write in general as (boundary conditions are neglected for simplicity)

$$\begin{aligned} & \int_{\Omega_h} \omega_i(\mathbf{w}_h) (\partial_t \mathbf{w}_h + \nabla \cdot \mathbf{f}_h(\mathbf{w}_h)) \\ & + \sum_{K \in \Omega_h} \oint_{\partial K} \gamma^{\partial K}(\mathbf{w}_h) [\nabla \mathbf{w}_h] \cdot [\nabla \varphi_i] = 0 \end{aligned}$$

with $[\cdot]$ a jump of a quantity, as in (16c). The weight ω_i in the first term is better expressed as a composition of local restrictions $\omega_i = \sum_K \omega_i^K$, and depends on the specific method chosen. For SUPG-type schemes, we can write $\omega_i^K = \varphi_i^K + \gamma_i^K(\mathbf{w}_h)$, with the first term only depending on the mesh. For other methods such as RD schemes, similar decompositions may be invoked; however, these are not unique (Ricchiuto and Abgrall, 2010). Other definitions can be obtained by considering variational multiscale stabilization techniques or bubble functions (see Hughes *et al.*, 2004 for a review). The last term in the method is one of the possible forms of edge stabilization (Burman *et al.*, 2008, 2010). Because of the presence of the $\partial_t \mathbf{w}_h$ term in the residual $r(\mathbf{w}_h)$ and of the continuity of the approximation, the first term will lead to a global mass matrix in the resulting system of ODEs. This matrix, in general, depends on the discrete solution \mathbf{w}_h , and, in the case of RD schemes, is neither uniquely defined nor guaranteed to be invertible (Ricchiuto and Abgrall, 2010).

The first idea to simplify things came originally from Ricchiuto and Abgrall (2010), and requires the introduction of some discrete approximation of the ODE system. As was done before, we consider a semidiscretization in time, and

the semidiscrete residual we write here as

$$\begin{aligned} \Gamma^{n+1} &= \Gamma^{n+1}(\mathbf{w}_h^{n+1}; \{\mathbf{w}_h^{(s)}\}) \\ &= \alpha_{-1} \mathbf{w}_h^{n+1} + \sum_{s=0}^S \alpha_s \mathbf{w}_h^{(s)} + \Delta t \sum_{s=0}^S \theta_s \nabla \cdot \mathbf{f}_h(\mathbf{w}_h^{(s)}) \end{aligned} \quad (90)$$

with the $\mathbf{w}_h^{(s)}$ values being either those computed from previous time steps (multistep scheme) or from some previous predictor stages (multistage). Note that the two summations on the right-hand side are independent of the unknown \mathbf{w}_h^{n+1} . As before, for an r th-order accurate method in time, the local truncation error relation will be of the type $\Gamma^{n+1} = \mathcal{O}(\Delta t^{r+1}) = \mathcal{O}(h^{r+1})$, if (82) hold, as is always the case for explicit time schemes. If we proceeded as in the last section, we would plug Γ^{n+1} in the spatial discretization, and the term $\alpha_{-1} \mathbf{w}_h^{n+1}$ would lead to the inversion of a (nonlinear) mass matrix. However, in Ricchiuto and Abgrall (2010) it was proved that *given a k th-order accurate approximation in space, and an r th-order accurate approximation in time, provided that the ratio $\Delta t/h$ is uniformly bounded, the space-time discretization*

$$\begin{aligned} & \int_{\Omega_h} \varphi_i(\Gamma^{n+1}(\mathbf{w}_h^{n+1}; \{\mathbf{w}_h^{(s)}\}) - \tilde{\Gamma}^{n+1}(\{\mathbf{w}_h^{(s)}\})) \\ &= - \int_{\Omega_h} \omega_i(\mathbf{w}_h) \tilde{\Gamma}^{n+1}(\{\mathbf{w}_h^{(s)}\}) \\ & \quad - \Delta t \sum_{s=0}^S \beta_s \sum_{K \in \Omega_h} \oint_{\partial K} \gamma^{\partial K}(\mathbf{w}_h^{(s)}) [\nabla \mathbf{w}_h] \cdot [\nabla \varphi_i] \end{aligned}$$

verifies a truncation error/consistency estimate of the type $\epsilon = \mathcal{O}(h^p)$, with $p = \min(k+1, r+1)$, provided that for a smooth exact solution, the modified semidiscrete residual $\tilde{\Gamma}^{n+1}$ verifies the consistency estimate

$$\tilde{\Gamma}^{n+1} = \mathcal{O}(h^{p-1})$$

The first practical use of this reduced consistency requirement for $\tilde{\Gamma}^{n+1}$ was to modify a given time discretization to obtain residual expressions one order lower. For example, for the classical third-order RK3 method, one has (Ricchiuto and Abgrall, 2010)

$$\text{first step} \quad \begin{cases} \Gamma_{\text{RK3}}^{(1)} = \mathbf{w}^{(1)} - \mathbf{w}^n + \nabla \cdot \mathbf{f}(\mathbf{w}^n) \\ \tilde{\Gamma}_{\text{RK3}}^{(1)} = \nabla \cdot \mathbf{f}(\mathbf{w}^n) \end{cases}$$

second step

$$\begin{cases} \Gamma_{\text{RK3}}^{(2)} = \mathbf{w}^{(2)} - \mathbf{w}^n + \frac{\Delta t}{4} \times (\nabla \cdot \mathbf{f}(\mathbf{w}^n) + \nabla \cdot \mathbf{f}(\mathbf{w}^{(1)})) \\ \tilde{\Gamma}_{\text{RK3}}^{(2)} = \frac{\mathbf{w}^{(1)} - \mathbf{w}^n}{2} + \frac{\Delta t}{4} \times (\nabla \cdot \mathbf{f}(\mathbf{w}^n) + \nabla \cdot \mathbf{f}(\mathbf{w}^{(1)})) \end{cases}$$

$$\text{final step} \begin{cases} \Gamma_{\text{RK3}}^{n+1} = \mathbf{w}^{n+1} - \mathbf{w}^n \\ \quad + \frac{\Delta t}{6} (\nabla \cdot \mathbf{f}(\mathbf{w}^n) + 4\nabla \cdot \mathbf{f}(\mathbf{w}^{(2)}) + \nabla \cdot \mathbf{f}(\mathbf{w}^{(1)})) \\ \tilde{\Gamma}_{\text{RK3}}^{n+1} = 2(\mathbf{w}^{(2)} - \mathbf{w}^n) \\ \quad + \frac{\Delta t}{6} (\nabla \cdot \mathbf{f}(\mathbf{w}^n) + 4\nabla \cdot \mathbf{f}(\mathbf{w}^{(2)}) + \nabla \cdot \mathbf{f}(\mathbf{w}^{(1)})) \end{cases}$$

For the extrapolated backward differencing method (eBDF3), one finds (Klosa, 2012; Klosa *et al.*, 2016)

$$\begin{aligned} \Gamma_{\text{eBDF3}}^{n+1} &= \frac{11}{6} \mathbf{w}^{n+1} - 3\mathbf{w}^n + \frac{3}{2} \mathbf{w}^{n-1} - \frac{1}{3} \mathbf{w}^{n-2} \\ &\quad + \Delta t (3\nabla \cdot \mathbf{f}(\mathbf{w}^n) - 3\nabla \cdot \mathbf{f}(\mathbf{w}^{n-1})) + \nabla \cdot \mathbf{f}(\mathbf{w}^{n-2}) \\ \tilde{\Gamma}_{\text{eBDF3}}^{n+1} &= \frac{5}{2} \mathbf{w}^n - 4\mathbf{w}^{n-1} + \frac{3}{2} \mathbf{w}^{n-2} \\ &\quad + \Delta t (3\nabla \cdot \mathbf{f}(\mathbf{w}^n) - 3\nabla \cdot \mathbf{f}(\mathbf{w}^{n-1})) + \nabla \cdot \mathbf{f}(\mathbf{w}^{n-2}) \end{aligned}$$

Equation (90) can be also seen as a defect correction method in which a lower order residual is used as a means of approximating solutions of a high-order one.

Note, however, that (90) still requires the inversion of the Galerkin mass matrix, which, even though symmetric positive-definite, is not an inverse monotone matrix. This may destroy all the efforts made in the construction of a shock-capturing mechanism in the method. The solution is to constrain the choice of finite element spaces to those allowing the lumping of this matrix. Several choices exist, either based on standard Lagrange elements on cubature grids with strictly positive cubature weights (Cohen *et al.*, 2001; Giraldo and Taylor, 2006; Xu, 2011; Mulder, 2013), or on non-Lagrange elements having a property similar to that the Bezier basis proposed in Abgrall and Treflik (2010) (see also Rogers, 2001, Ch. 5). Whatever the choice, this approach leads to a fully explicit space-time discretization, reading

$$\begin{aligned} |\mathcal{V}_i| &\left(\alpha_{-1} \mathbf{w}_i^{n+1} + \sum_{s=0}^S \tilde{\alpha}_s \mathbf{w}_i^{(s)} \right) \\ &= - \int_{\Omega_h} \omega_i(\mathbf{w}_h) \tilde{\Gamma}^{n+1}(\{\mathbf{w}_h^{(s)}\}) \\ &\quad - \Delta t \sum_{s=0}^S \theta_s \sum_{K \in \Omega_h} \oint_{\partial K} \gamma^{\partial K}(\mathbf{w}_h^{(s)}) [\nabla \mathbf{w}_h] \cdot [\nabla \varphi_i] \end{aligned}$$

with $|\mathcal{V}_i|$ being a nodal volume depending on the areas of the surrounding elements and on the quadrature weights induced

by the finite element basis, and with the $\tilde{\alpha}_s$ obtained from the “defect-correction” in time $\Gamma - \tilde{\Gamma}$.

This construction provides genuinely explicit variants of all well-known stabilized continuous finite elements (SUPG, GLS, VMS, etc.), as well as of nonlinear residual distribution schemes discussed in this chapter. Thorough numerical validations have been reported in Ricchiuto and Abgrall (2010), Ricchiuto (2015), Klosa (2012), Klosa *et al.* (2016). Some examples will be provided in the following sections.

4 APPLICATIONS

4.1 Scalar examples

We start with a few scalar convergence tests to check some of the theoretical aspects discussed in this chapter. Consider the approximation of solutions of the steady scalar advection equation (18) on the domain $\Omega = [0, 1]^2$, with $\vec{a} = (0, 1)$, and with inlet condition $u(x, 0) = \sin^2(\kappa \pi x)$.

We start with a result taken from Abgrall *et al.* (2009). The test aims at verifying the analysis of Section 3.1.7. The grid convergence has been run for $\kappa = 1$ with the nonlinear LLFs scheme (64), and with different evaluation strategies for the streamline dissipation or *filtering* term. In particular, the discrete term in (59) is taken as the arithmetic average of its value in a certain set of points. Note that, with the exception of linear polynomials, this evaluation does not give in general any k_{exact} quadrature formula. Table 1 shows the impact of under-evaluating this term. For a \mathbb{P}^2 finite element approximation, first-order accuracy is obtained unless a three-point stencil is used. Similarly, for the \mathbb{P}^3 finite element approximation, a stencil of at least six points is required. Provided that the number of points is large enough, we see that, indeed, we recover the expected second-, third-, and fourth-order rates, even though the expressions used to evaluate the streamline dissipation are not obtained from a high-order quadrature formula.

The next example, taken from Vymazal (2016) (see also Vymazal *et al.*, 2015; D’Angelo *et al.*, 2015), aims at verifying the convergence rates obtained with the “variable- β ” LDA (61). Polynomial approximations up to degree $k = 7$ are tested using meshes with roughly the same number of DOFs in all cases (from ≈ 2000 for the coarsest mesh to $\approx 32\,000$ for the finest). The simulations are run with $\kappa = 5$. The results, summarized in Table 2, show that indeed the method converges with a rate between $k + 1/2$ and $k + 1$. For $k = 6$ and $k = 7$, converging results have been obtained only by using the optimized collocation of the DOFs based on the warp-and-blend procedure discussed in Warburton (2006). Computations on standard Lagrange elements with equally spaced DOFs did not converge for $k > 5$.

Table 1. Scalar advection: grid convergence for the LLFs scheme (64).

	$k = 1$ Filter: \mathbb{P}^0 dof	$k = 2$ Filter: \mathbb{P}^0 dof	$k = 2$ Filter: \mathbb{P}^1 dof	$k = 3$ Filter: \mathbb{P}^1 dof	$k = 3$ Filter: \mathbb{P}^2 dof
h	L^2	L^2	L^2	L^2	L^2
1/25	0.50493E-02	0.25122E-01	0.32612E-04	2.17274E-02	0.12071E-05
1/50	0.14684E-02	0.12935E-01	0.48741E-05	1.13486E-02	0.90642E-07
1/100	0.41019E-03	0.83978E-02	0.66019E-06	5.83347E-03	0.53860E-08
Average rate	1.790	0.7904	2.812	0.9292	3.914

Verification of the analysis of Section 3.1.7: impact of the number of evaluation points for the “filtering term”.

Source: Reproduced with permission from Abgrall *et al.* (2009). © Elsevier, 2009.

Table 2. Scalar advection: convergence for the variable β LDA (61). See also Vymazal *et al.* (2015), D’Angelo *et al.* (2015).

k	N_{dof}	h	ϵ_{L^2}	Rate
1	2 094	0.02185	3.49E-02	—
	8 124	0.01109	7.44E-03	2.24
	32 546	0.00554	1.36E-03	2.46
2	2 189	0.02137	1.37E-02	—
	8 217	0.01103	2.19E-03	2.65
	32 181	0.00557	3.04E-04	2.88
(equi-spaced)	3	2 113	0.02175	—
		8 347	0.01095	4.16
(equi-spaced)		33 520	0.00546	3.89
4	2 017	0.02227	2.57E-03	—
	8 593	0.01079	9.94E-05	4.71
	(equi-spaced)	32 553	0.00554	4.55
5	2 381	0.02049	1.15E-03	—
	8 611	0.01078	2.83E-05	5.36
	(equi-spaced)	33 546	0.00546	5.75
6	2 317	0.02077	6.68E-04	—
	8 293	0.01098	7.30E-06	7.01
	(warp-blend)	33 073	0.00550	6.67
7	2 633	0.01949	3.82E-04	—
	9 430	0.01030	2.44E-06	7.92
	(warp-blend)	34 427	0.00539	8.51

Source: Reproduced with permission from Vymazal (2016). © M. Vymazal, 2016.

4.2 External aerodynamics

In this section, we report a couple of results from Abgrall *et al.* (2011) for compressible fluids without viscous effect (Euler equations) and from Abgrall *et al.* (2014), Abgrall and de Santis (2015) for the Navier–Stokes case. The interested reader may consult (De Santis, 2015) for information and results for the turbulent case (Spalart and Allmaras model).

4.2.1 Euler equations

Method: from scalar to systems

So far, we have only dealt with scalar problems: the computation of the residual distribution parameters is done via

arithmetic and logical operations on a scalar. This cannot be as simple for systems because dividing vectors has no meaning.

The method that is followed was introduced in Abgrall (2006). The idea is as follows: given an element K , we first consider an average state $\bar{\mathbf{w}}$. The choice of this average state does not seem to be essential, and we take the arithmetic mean. From this, one can evaluate the Jacobians of the flux at this state, say $A(\bar{\mathbf{w}})$. The next step is to choose a direction \mathbf{d} . Again, the choice does not seem to be essential, and for fluid dynamic problems we consider the normalized velocity except when the velocity vanishes. In that case, we take an arbitrary direction. Once this is done, we compute the eigenvectors $\{r_j\}_{j=1,\dots,m}$ of the matrix $A(\bar{\mathbf{w}}) \cdot \mathbf{d}$. Any vector X can be decomposed on this basis as

$$X = \sum_{i=1}^m \ell_i(X) r_i$$

The eigenvectors r_j are often called the right eigenvectors, while the linear forms ℓ_j are often called the left eigenvectors of $A(\bar{\mathbf{w}}) \cdot \mathbf{d}$.

We start from the LLF residual, $\{\Phi_j\}_{j=1,K}$, where K is the number of DOFs in K . For any eigenvector r_i , we consider the quantities

$$\{\ell_i(\Phi_j)\}_{j=1,\dots,K}$$

that clearly satisfy

$$\sum_{j=1}^N \ell_i(\Phi_j) = \ell_i(\Phi)$$

Because of this, we interpret these quantities as residual, and we can apply the technique of Section 3.1.7 to evaluate, for any $j = 1, \dots, K$,

$$(\ell_i(\Phi_j))^* = \beta_j^i \ell_i(\Phi),$$

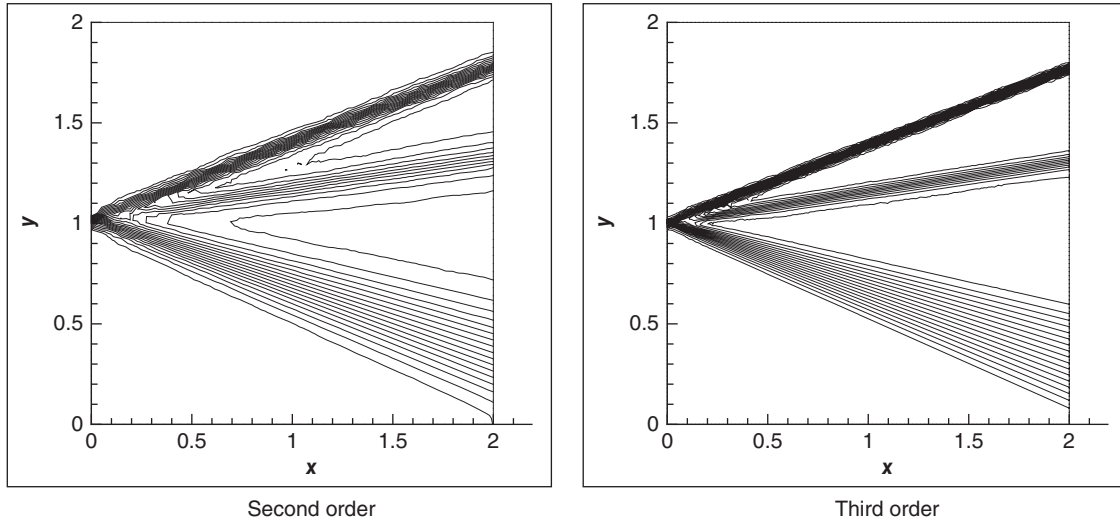


Figure 5. Jet problem: isolines of the density, second- and third-order LLxFf scheme. All the degrees of freedom are plotted, and the same isolines are also plotted. (Reproduced with permission from Abgrall *et al.*, 2011. © Elsevier, 2011.)

where, for example, β_j^i is evaluated via the PSI “limiter” (62). Once this is done, we define

$$\Phi_j^* = \sum_{i=1}^m (\mathcal{L}_i(\Phi_j))^*$$

which satisfies the accuracy requirements. If needed (and this is generally the case), one can add a least-squares filtering term

$$\begin{aligned} \Phi_j^{**} &= \Phi_j^* + \theta(\mathbf{w}_h) |K| \sum_{x_{\text{quad}}} \omega_{\text{quad}} (A(\bar{\mathbf{w}}) \cdot \nabla \varphi_i(x_{\text{quad}})) \\ &\quad \times \mathcal{T}_K (A(\bar{\mathbf{w}}) \cdot \nabla \mathbf{w}_h(x_{\text{quad}})) \end{aligned}$$

where

$$\mathcal{T}_K^{-1} = \sum_{j=1}^N |A(\mathbf{w}) \cdot \nabla \varphi_j(x_{\text{quad}})|$$

In Abgrall (2001), the matrix $\sum_{j=1}^N |A(\bar{\mathbf{w}}) \cdot \nabla \varphi_j(x_{\text{quad}})|$ is always invertible except when the velocity defined by $\bar{\mathbf{w}}$ is zero. However, in that case the matrices

$$A(\bar{\mathbf{w}} \cdot \nabla \varphi_i(x_{\text{quad}})) \mathcal{T}_K$$

can always be defined, see Abgrall (2001) for details.

Applications

These results are taken from Abgrall *et al.* (2011). The meshes use triangles only unless specified.

In our first example, the domain is a square $\Omega = [0, 1]^2$. The boundary conditions are as follows:

- If $y > 0.5$ and $x = 0$, the Mach number is set to $M_\infty = 4$, the density $\rho_\infty = 0.5$, and the velocity is $(u_\infty = M_\infty c_\infty, 0)$ with $c_\infty = \sqrt{\gamma p_\infty / \rho_\infty}$.
- If $y \leq 0.5$ and $x = 0$, the Mach number is set to 2.4, the velocity is $(u_\infty, 0)$, and the density set to 1.
- The other boundaries are assumed to be supersonic.

In such a configuration, the flow is steady and supersonic. We have a shock wave at the bottom, followed by a slip line and then a fan, see Figure 5. Since the flow is supersonic, the x -coordinate plays the role of time: if one makes a cross section $x = \text{const}$, we have a self-similar solution of the same type as what one gets for a one-dimensional shock tube. It is clear that there is no oscillation at all on the density. The same conclusion holds for the other variables (not displayed).

The next example is the classical flow at $M_\infty = 0.35$ over a sphere. In that case, the flow is symmetric with respect to the x -axis of the domain, but also with respect to the y -axis. We have run this case with a second-order scheme, a third-order scheme, and again the second-order scheme on the mesh that has the same degrees of freedom as those of the \mathbb{P}^2 scheme. In other words, we subdivided each triangle into four smaller triangles whose vertices are those of the large triangle and the mid-edge points. The initial mesh has 2719 nodes, 5308 elements, and 100 nodes on a cylinder. It is displayed in Figure 6.

We see in Figure 7, which displays the pressure coefficient isolines, the improvement of the solution quality when the scheme is upgraded from second order to third order. More important, the same figure indicates clearly that the second-order scheme on the refined mesh gives less accurate

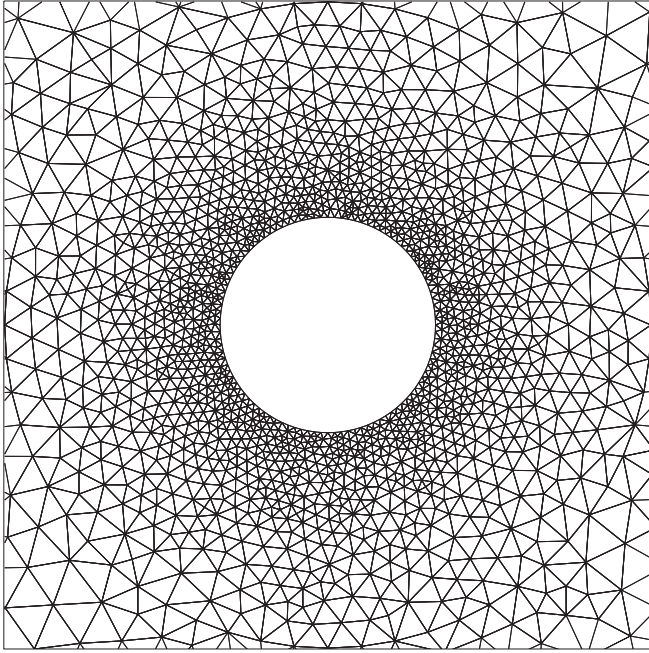


Figure 6. Subsonic sphere problem: zoom of the mesh for the sphere problem. The mesh has no symmetry. (Reproduced with permission from Abgrall *et al.*, 2011. © Elsevier, 2011.)

results than the third-order one. Note that we have the same DOFs in both cases.

We have rerun this test case on a hybrid mesh using the second-order and third-order schemes. In both cases, the same DOFs are used (i.e., we use the DOFs of the sub-triangulation for the second-order scheme). The results are shown in Figure 8. The mesh uses 81 points on the sphere. We get the same conclusions as before.

Our next example is a flow over an NACA012 airfoil. It is transonic, and has the following conditions at infinity: $M = 0.8$, angle of attack of 1.25° . The mesh has 10 959 points and 21 591, corresponding to 43 509 degrees of freedom.

In Figure 9, we have displayed the Mach number, the pressure coefficients, and relative entropy deviation for the third-order version of the scheme. The solutions are fine. Note, however, a non-physical overshoot in the entropy across the upper shock.

We have run many other tests (results not shown). If we compare the second-order solution run with a mesh constructed from the mesh we have used where the element is sub-triangulated so that we have the same number of degrees of freedom, we can see an excellent agreement between the solutions, but with a main difference. In both cases the shock is with one element, but one element for the third-order solution is roughly twice as large as an element for the second-order one. Hence, the shock looks more diffused

in the third-order case. However, the entropy levels are much lower, as we have already seen in the two sphere subsonic case.

Another case is the Ringleb flow. It was devised by Ringleb (1940) in 1940, see von Mises (1958) for derivation of more general solutions. This is an isentropic, irrotational, two-dimensional flow. It is defined from the streamline function (θ is the velocity angle with respect to a given direction, and v is the norm of the velocity) $\psi = \frac{\sin \theta}{v}$. From this, it is possible to get the explicit form of the streamlines

$$x = \frac{1}{2} \frac{1}{\rho} \left(\frac{1}{v^2} - \frac{2}{k^2} \right) + \frac{J}{2}$$

$$y = \pm \frac{1}{k\rho v} \sqrt{1 - \left(\frac{q}{k} \right)^2}$$

with

$$k = \frac{1}{\phi} \text{ a constant on any stream line,}$$

$$J = \frac{1}{c} + \frac{1}{3c^2} + \frac{1}{5c^2} - \frac{1}{2} \log \left(\frac{1+c}{1-c} \right)$$

$$c = \sqrt{1 - \frac{\gamma-1}{2} q^2}, \quad \rho = c^{2/(\gamma-1)}$$

The pressure is determined by the equal-entropy assumption. We see that the isotach lines are the circles

$$\left(x - \frac{J}{2} \right)^2 + y^2 = \frac{1}{4\rho^2 q^4}$$

From this, it is possible to determine the exact solution: given a point (x, y) , we determine the speed of sound c such that (x, y) belongs to the circle of center $(J(c)/2, 0)$ and radius $1/(2(\rho q^2))$. Once this is done, we can get all the other values.

We have run this case in the (symmetric) domain defined by

- the circle $q = 0.3$ on the top and the bottom,
- the extreme stream lines $k = 0.4$ and $k = 0.8$.

The simulation was conducted with two series of meshes. The first one is made of quads cut into two triangles, always in the same direction. The mesh is then made symmetric. In the second one, we only consider the quads. In both cases, we have $2 \times P$ points on the streamlines $k = 0.3$ and 0.8 and P points on the circles $q = 0.3$. Here we have taken $P = 15, 30, 60$, and 100 . The error in the L^2 norm for the density is shown in Figure 10. We see a slope of -3 for the third-order scheme and -1.5 for the second-order scheme. We also note that, though the formal accuracy in both cases is as expected, the effective accuracy on the quad meshes is much superior to that obtained for triangular meshes.

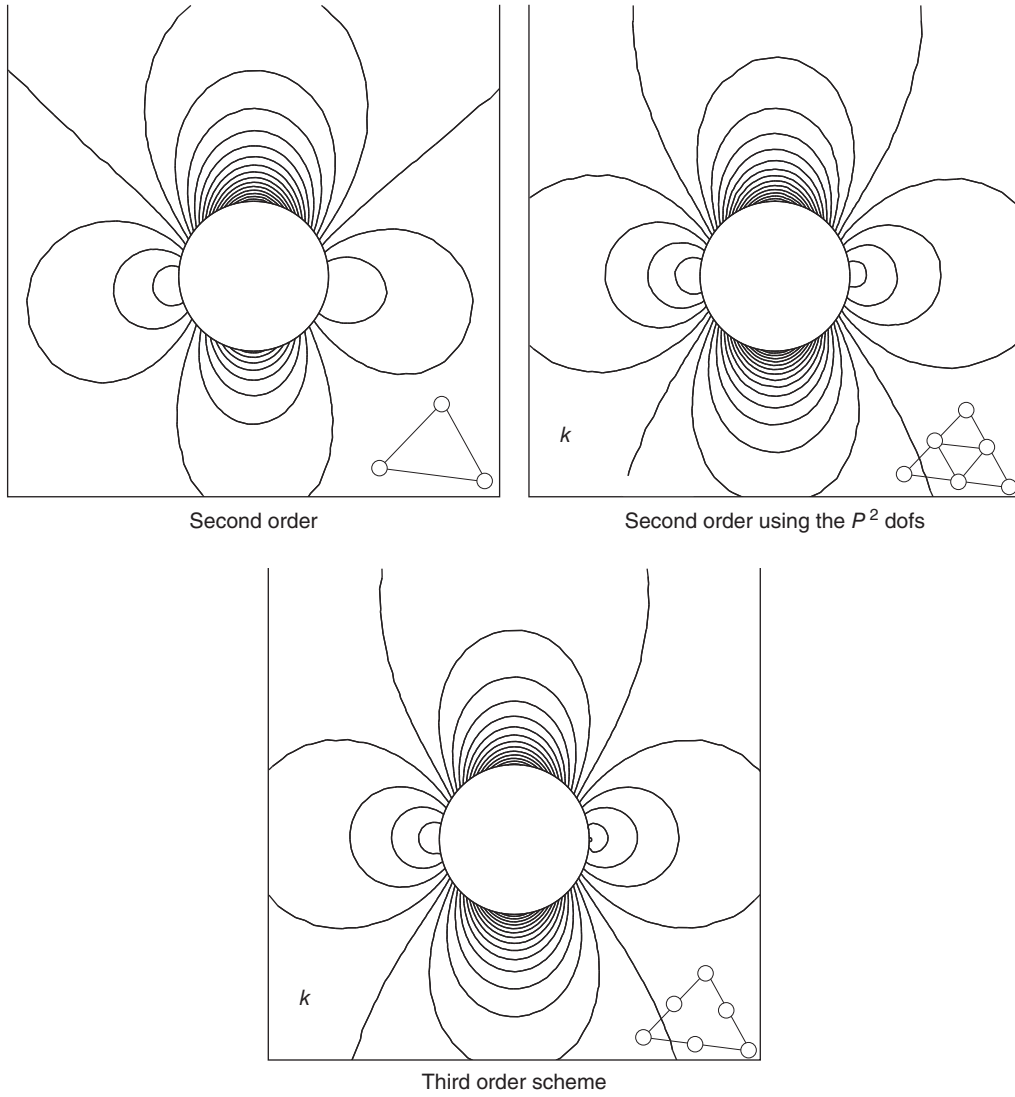


Figure 7. Subsonic sphere problem: isolines of the pressure coefficient. We have the same isolines on each figure. (Reproduced with permission from Abgrall *et al.*, 2011. © Elsevier, 2011.)

We have run the same scheme on a scramjet-like configuration using a hybrid mesh, as shown in Figure 11.

This example has already been run in Abgrall (2006). The inflow Mach number is set to 3.5. The geometry is such that many waves coexist and interact in very complex flow patterns. This situation is particularly clear on the upper part of the internal body, where shocks, fans, and their reflection interact because of the wall. Again, in both cases, the same number of DOFs was used. Once again, the scheme was run starting from a uniform flow configuration. Figure 12 shows the Mach number isolines. As expected, there is no real difference between the solutions since the flow is basically made of shock, fans, slip lines, and constant states: this is

not an accuracy case, but a case that shows that, despite the flow complexity, the third-order scheme is robust.

However, one can see a small difference between the solutions: the slip line created by the interaction of two shocks after the blade is slightly more twisted for the third-order scheme than the second-order one. We also see that the resolution of the discontinuities is in both case approximately one cell width.

4.2.2 Navier–Stokes equations

We report here again the results taken from Abgrall and de Santis (2015). The scheme and problems have already

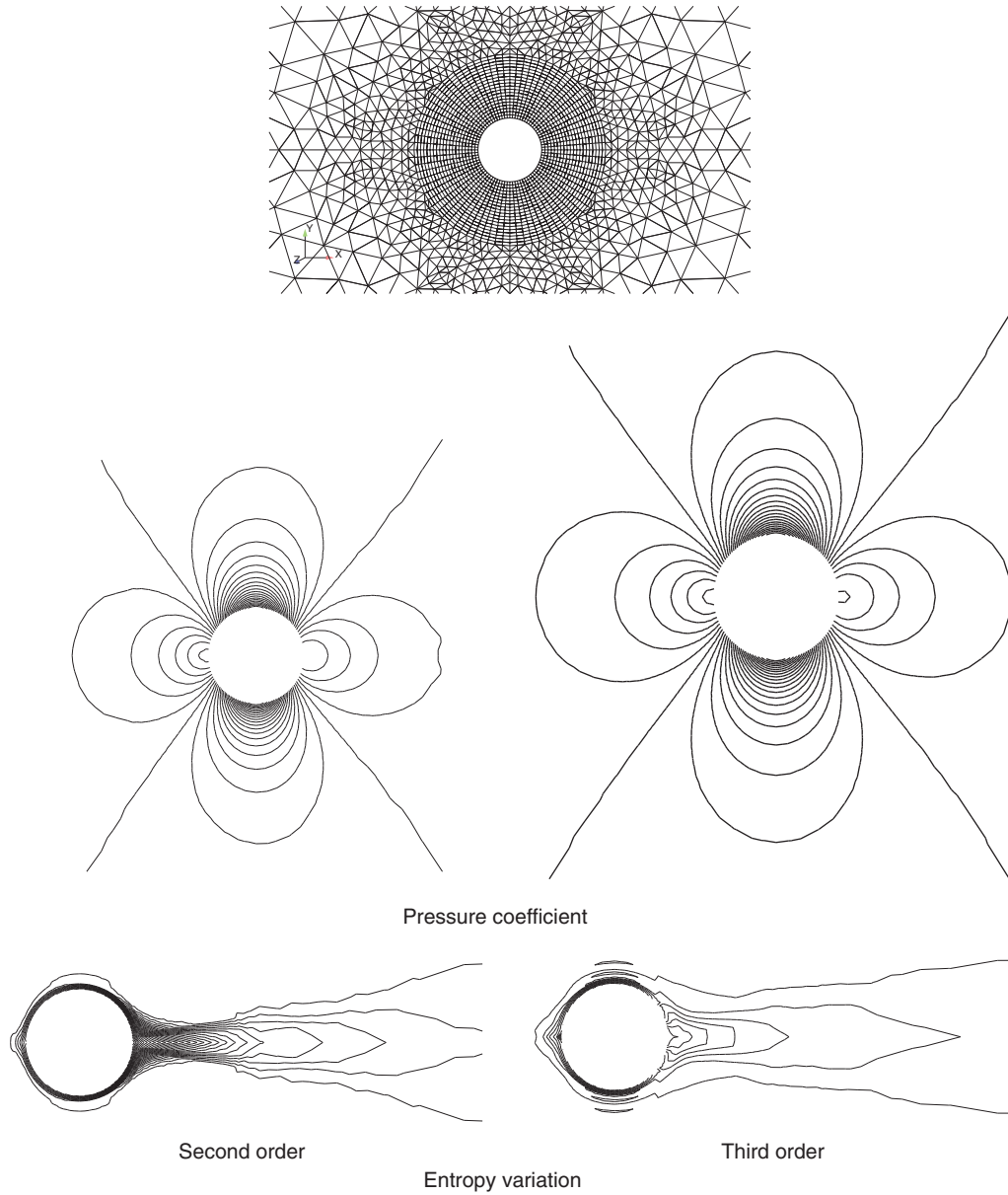


Figure 8. Subsonic sphere problem, hybrid mesh: pressure coefficient and entropy variation on a hybrid mesh, $M_\infty = 0.35$. (Reproduced with permission from Abgrall *et al.*, 2011. © Elsevier, 2011.)

been discussed. For more details, the reader may also consult (Abgrall *et al.*, 2014). The filtering term has to be more elaborate in order to take into account the viscous terms.

The first example is the classical test case consisting of a subsonic viscous flow over a NACA-0012 airfoil at zero angle of attack. The free stream Mach number is 0.5, and the Reynolds number is 5000. This is a widely used test case for two-dimensional laminar flows; a distinctive feature of this test case is a steady separation bubble near the trailing edge of the airfoil. An example of computational grid is displayed

in Figure 13. The grid extends about 50 chords away from the airfoil. The airfoil boundary is considered adiabatic and without slip, and is represented by piecewise quadratic elements; the far-field boundary condition is applied on the outer boundary of the domain (see Abgrall and de Santis, 2015 for a precise description of the boundary conditions approximation as well as details on the steady-state solver). The steady state is considered to be reached when the L^2 norm of the density residual drops by 10 orders of magnitude compared to the initial value.

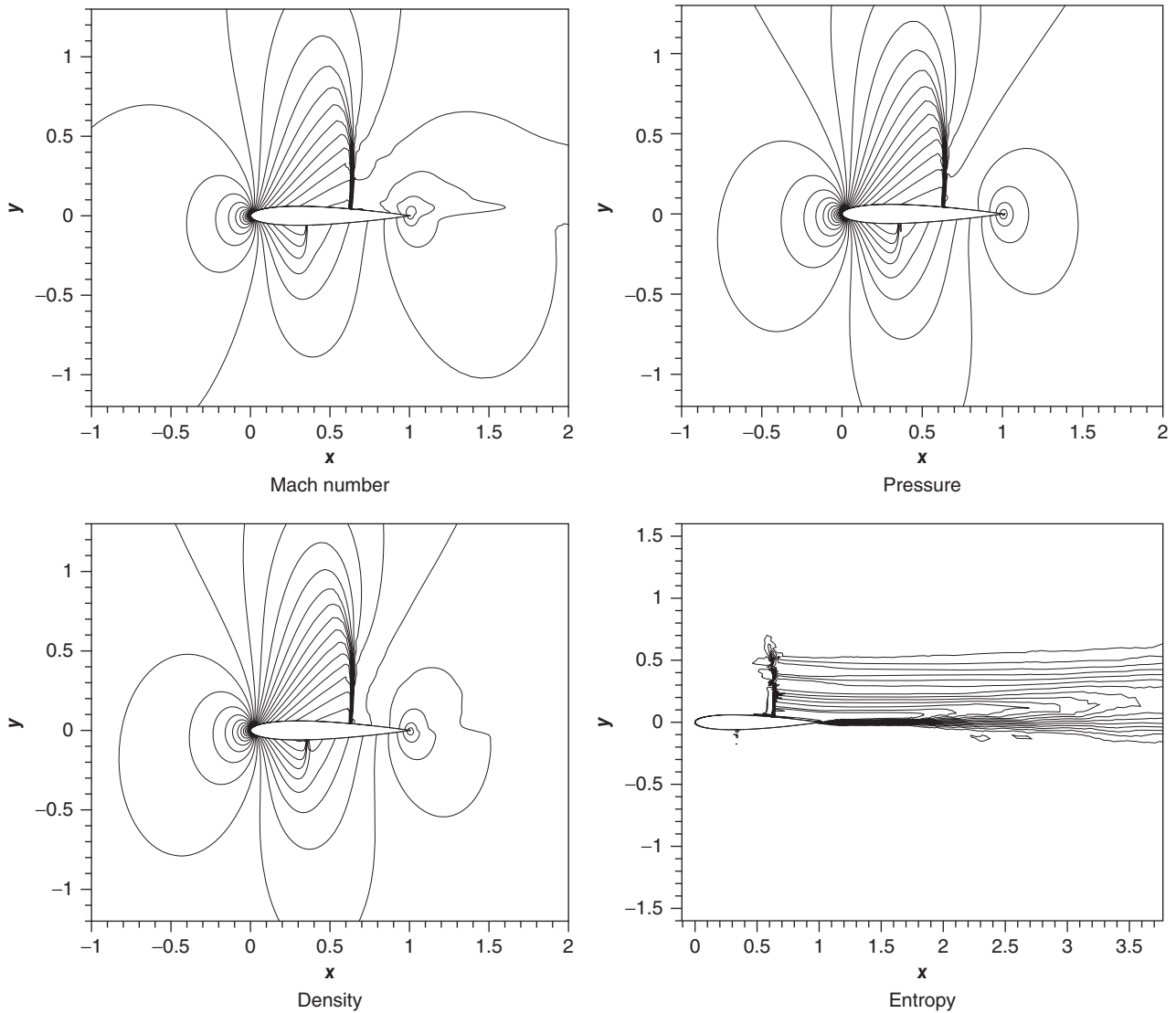


Figure 9. Transonic NACA012 problem. Isolines of the Mach number, pressure, density, and entropy for the NACA012 case. (Reproduced with permission from Abgrall *et al.*, 2011. © Elsevier, 2011.)

Figure 14 shows the solutions computed with the linear scheme and the SPR-ZZ gradient reconstruction, for \mathbb{P}^1 and \mathbb{P}^2 elements. The solution with the Φ^1 elements was computed on a grid obtained with \mathbb{P}^2 elements (4216 elements) and splitting each \mathbb{P}^2 triangle with four \mathbb{P}^1 triangles in such a way the number of DOFs for the second- and third-order simulation was exactly the same. Note that in Figure 14 although there is not much difference in the Mach number contours between the second- and third-order simulations, the streamlines near the trailing edge are very different, and only the third-order scheme is able to reproduce the symmetric recirculation bubble. For the same simulations, Figures 15 and 16 show the pressure and skin friction coefficient profiles, respectively. Note the

better regularity of the solution of the third-order simulation compared to the second-order one for the same number of DOFs.

The second example is a steady laminar flow at high angle of attack, around a delta wing with sharp edges. As the flow passes the leading edge, it rolls up and creates a big vortex structure, which is convected far behind the wing; at the same time, near the leading edge a smaller secondary vortex appears. A free stream Mach number $M = 0.5$ is considered; the Reynolds number, based on the root chord of the wing is $Re = 4000$; and the angle of attack is $\alpha = 12.5^\circ$.

The geometry of the delta wing is shown in Figure 17, together with an example of a coarse grid used for the simulations. The grid consists of tetrahedra; finer levels of

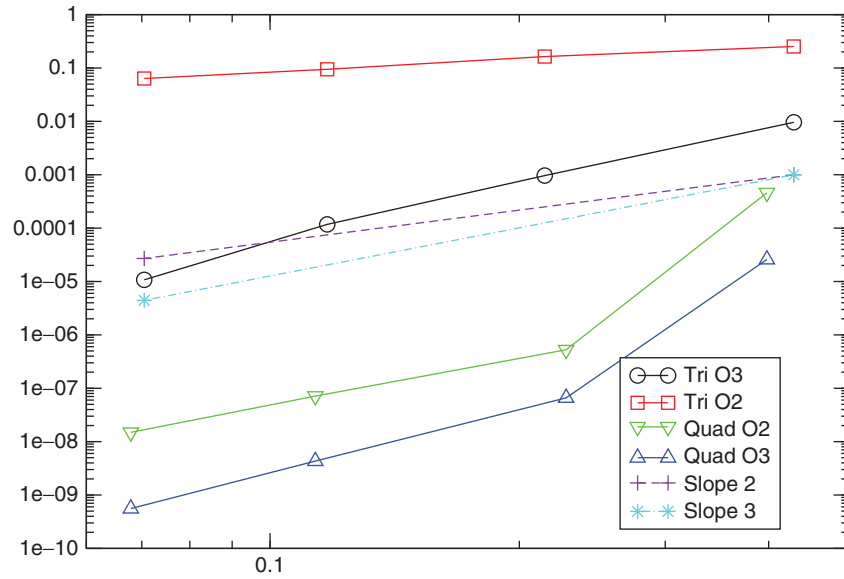


Figure 10. Ringleb flow problem. L^2 error on the density for the Ringleb flow. Tri stands for triangle, Quad for quadrangle. O2 stands for second order, and O3 for third order. (Reproduced with permission from Abgrall *et al.*, 2011. © Elsevier, 2011.)

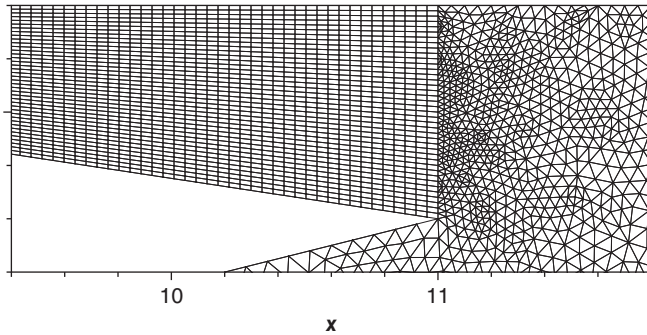


Figure 11. Zoom of the mesh for the scramjet problem. (Reproduced with permission from Abgrall *et al.*, 2011. © Elsevier, 2011.)

grids are obtained by uniformly splitting each tetrahedron of the coarser level with eight tetrahedra. Note the presence of very stretched elements on the wing. The wing surface is treated as a no-slip adiabatic wall, and the vertical plane intersecting the root of wing is treated as a symmetry plane, while far-field boundary conditions are applied on the outer boundary of the domain.

The solution is initialized with a uniform flow. The lower order solution is used as the initial solution for the third-order computation. For this test case, the linear scheme is used with the SPR-ZZ gradient reconstruction method. Figure 18 shows the streamlines and Mach number contours, at different stations, of the third-order solution on the finest grid.

Figure 19 shows the drag and lift coefficients computed with linear and quadratic elements on three uniformly refined grids. For comparison, the reference values computed in Leicht and Hartmann (2010) by extrapolating the results obtained with a higher order DG method are also shown. Observing the convergence of the drag coefficient in term of DOFs, it can be noted that there is no significant gain in using a higher order approximation with respect to the second-order one. This behavior can be due to the singularity at the leading edge of the wing, which might mask the benefits of a higher order approximation with a uniform mesh refinement. Regarding the convergence of the lift coefficient, a clear benefit of using a higher order approximation can be seen, because the big vortex structure over the wing is better captured with higher order elements.

As the last test example, the interaction of an oblique shock wave with a laminar boundary layer is considered. The aim of this test is to show the non-oscillatory properties of the nonlinear scheme in presence of discontinuities of the solution and, at the same time, the capability to maintain the accuracy required for the discretization of the boundary layer.

The test consists in a laminar boundary layer developing over a flat plate and an incident shock impinging the boundary layer. Since the flow is supersonic, a shock appears at the leading edge of the flat plate, which interacts with the oblique shock. Furthermore, at the impinging point, the incident shock produces a separation of the boundary layer, the shock is then reflected, and an expansion fan appears,

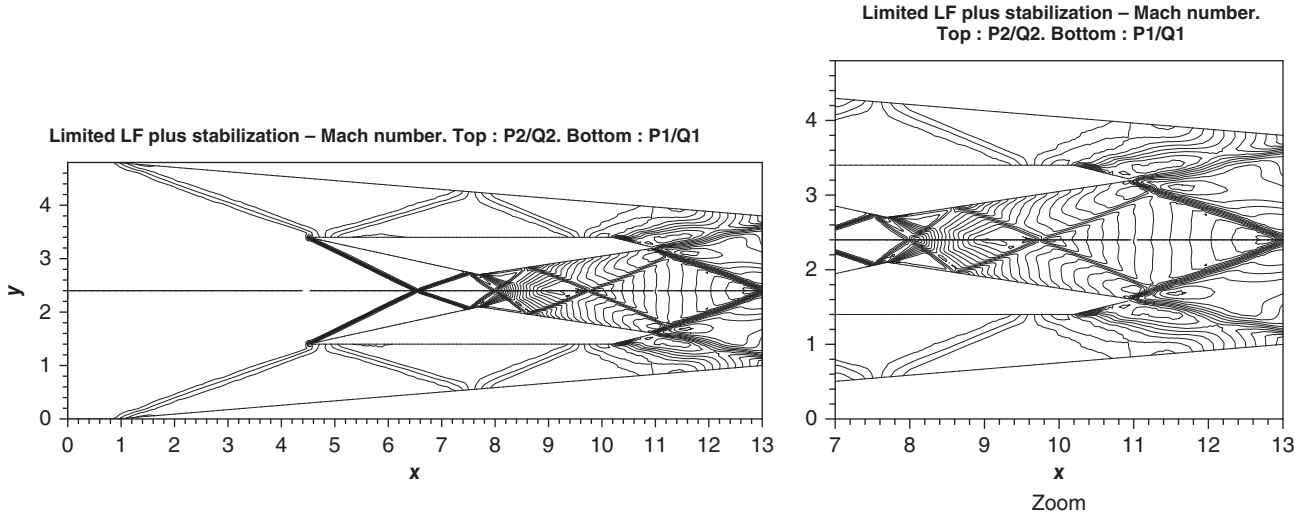


Figure 12. Scramjet problem. Mach number distribution. (Top) third-order solution. (Bottom) Second-order solution. The same isolines are plotted. Reproduced with permission from Abgrall *et al.*, 2011. © Elsevier, 2011.)

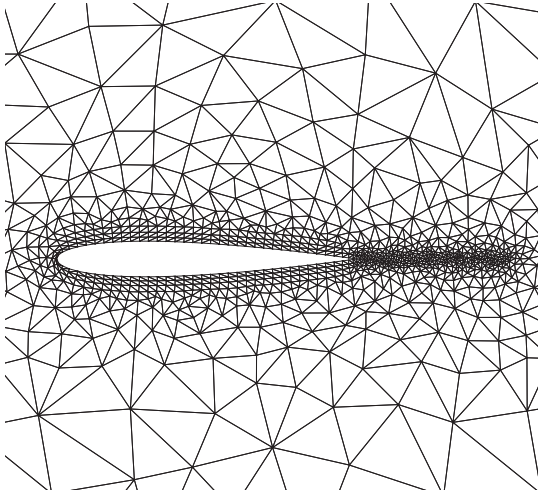


Figure 13. Example of the computational grid used for the NACA-0012 test case. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

turning the flow toward the wall and causing a reattachment of the boundary layer, as shown in Figure 20.

In the numerical simulations, the oblique shock is generated by imposing the incoming supersonic flow state on the lower part of left boundary, while another supersonic state is imposed on the upper part of the left boundary and on the top boundary; this state is computed using the relations of the oblique shocks, such that the incident shock has a certain angle of incidence θ_s . The height of the computational domain is 0.94, while the range of the domain in the x -direction is $[-0.2, 2]$. The flat plate has length $L = 2$, with

the leading edge of flat plate at $x = 0$. Along the plate, the no-slip adiabatic wall boundary condition is applied, while on the remaining part of the bottom boundary the symmetry boundary condition is applied. On the right boundary, the outflow boundary condition is applied, see Figure 20. The inflow states are chosen such that the free-stream Mach number is $M = 2.15$ and the angle of the incident shock is $\theta_s = 30.8^\circ$; in this case, the impingement point would be at center of the plate for an inviscid fluid. The Reynolds number based on the free-stream values and the distance between the plate leading edge and the inviscid shock impingement point is 1×10^5 .

The nonlinear scheme with the SPR-ZZ gradient recovery strategy is used to perform the numerical simulations at second and third order of accuracy. The computational domain is generated from the triangulation of a 90×85 structured grid; the first number refers to the number of elements on the horizontal boundaries, with 80 elements along the plates; the second number refers to the number of elements on the vertical boundaries. The element distribution is uniform on the x -direction, while along the y -direction a nonuniform distribution of the elements is used, with a mesh spacing $\Delta y = 0.5 \times 10^{-3}$ near the bottom boundary. For comparison, a second-order simulation is also performed on a finer grid with the same number of DOFs of the third-order simulation on the coarse grid. The simulation is initialized with a uniform solution, and the second-order solution is used as the initial solution for the third-order approximation. Except in the case of the second-order simulation on the coarse grid, for which the initial residual is reduced by 10 orders of magnitude,

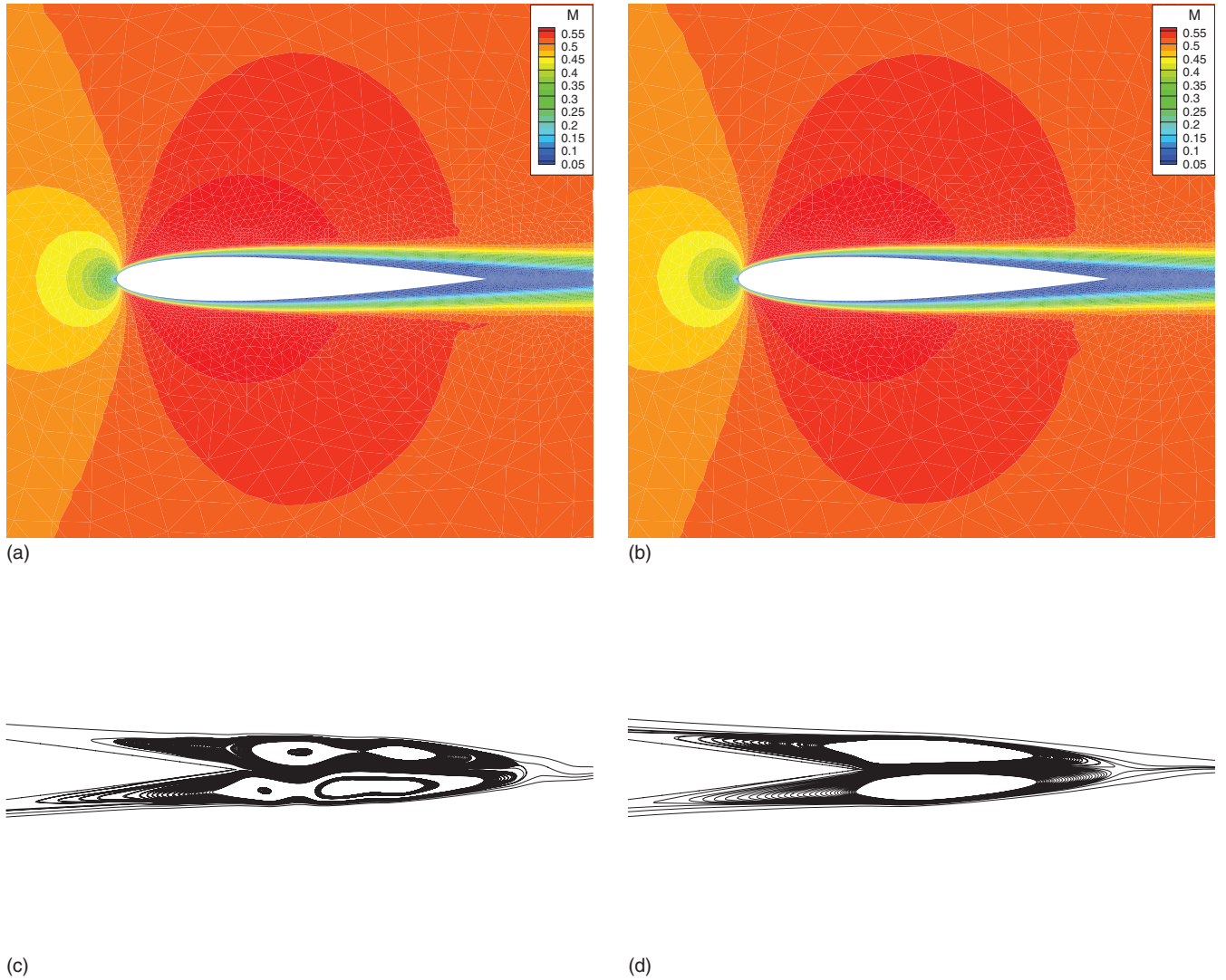


Figure 14. Mach number contours (a and b) and streamlines near the trailing edge (c and d) for the second-order (a and c) and third-order (b and d) linear scheme. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

the residual for the third-order and second-order simulations on the finer grid could not be reduced by more than 8 orders of magnitude.

Figure 21(a) shows the contours of the pressure for the third-order simulation; all the features of this problem are well represented. Figure 21(b) shows a zoom of the solution where the incident shock impinges the boundary layer. Two features are evident: the reflection of the incident shock and the recirculation bubble as a consequence of the separation, and the subsequent reattachment of the boundary layer produced by the incident shock and the expansion fan.

The profiles of density, pressure, and Mach number along the lines at $y = 0.29$ and $y = 0.15$ are reported in Figure 22.

Note that the third-order scheme gives a very sharp and monotone representation of the discontinuities, and also smooth portions of the solution are better represented compared to the second-order solution. It is important to remember that smooth and discontinuous solutions are treated within the same nonlinear scheme without any special treatment or tuning parameter. For a fair comparison, the solution obtained with the second-order scheme on a finer mesh is also reported. It is worth noticing that, although mesh refinement gives an improvement of the numerical solution, the level of accuracy obtained with the second-order scheme is still lower than that obtained with the third-order scheme for the same number of DOFs.

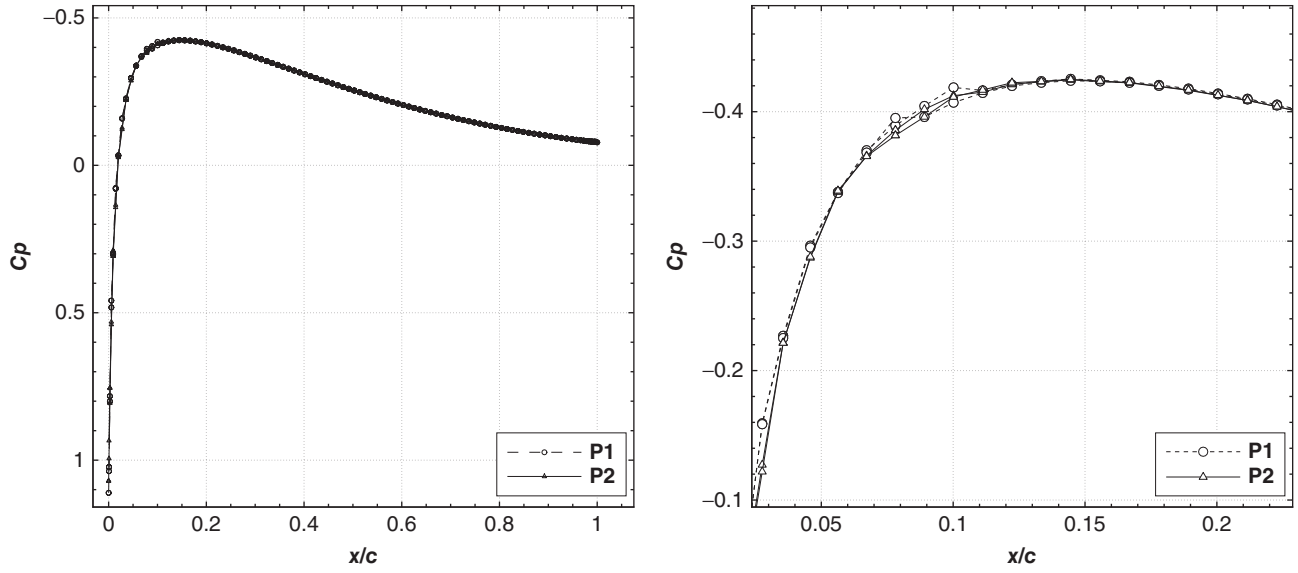


Figure 15. Pressure coefficient along the whole NACA-0012 airfoil for the second- and third-order simulations with the same number of DOFs. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

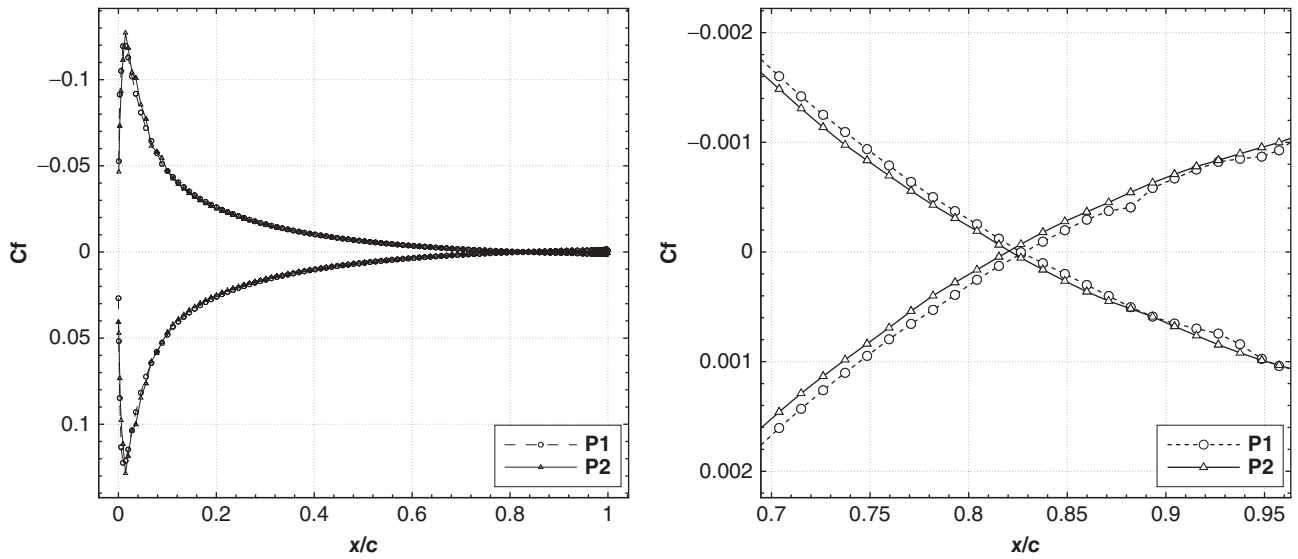


Figure 16. Skin friction coefficient along the whole NACA-0012 airfoil for the second- and third-order simulations with the same number of DOFs. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

Finally, Figure 23 shows the values of the pressure and the friction coefficient along the plate. The oscillations near the point $x = 0$ are due to the singularity of the solution at the leading edge of the flat plate, but they are limited to only a small region around the leading edge. The third-order scheme seems less sensitive to this singularity compared to the second-order simulations. The separation bubble can easily be detected by the negative values of friction coefficients. Note also the pressure plateau in the detached zone.

4.3 Free surface flows

The framework presented in this chapter has proved quite interesting to construct discrete approximations of systems of PDEs modeling free surface flows, namely the shallow-water equations and dispersive enhancements (Boussinesq and/or Green–Naghdi equations). Early work on steady hydrostatic flows had been reported in the Ph.D. thesis of Hubbard (see, e.g., Garcia-Navarro *et al.*, 1995; Hubbard and Baines, 1997,

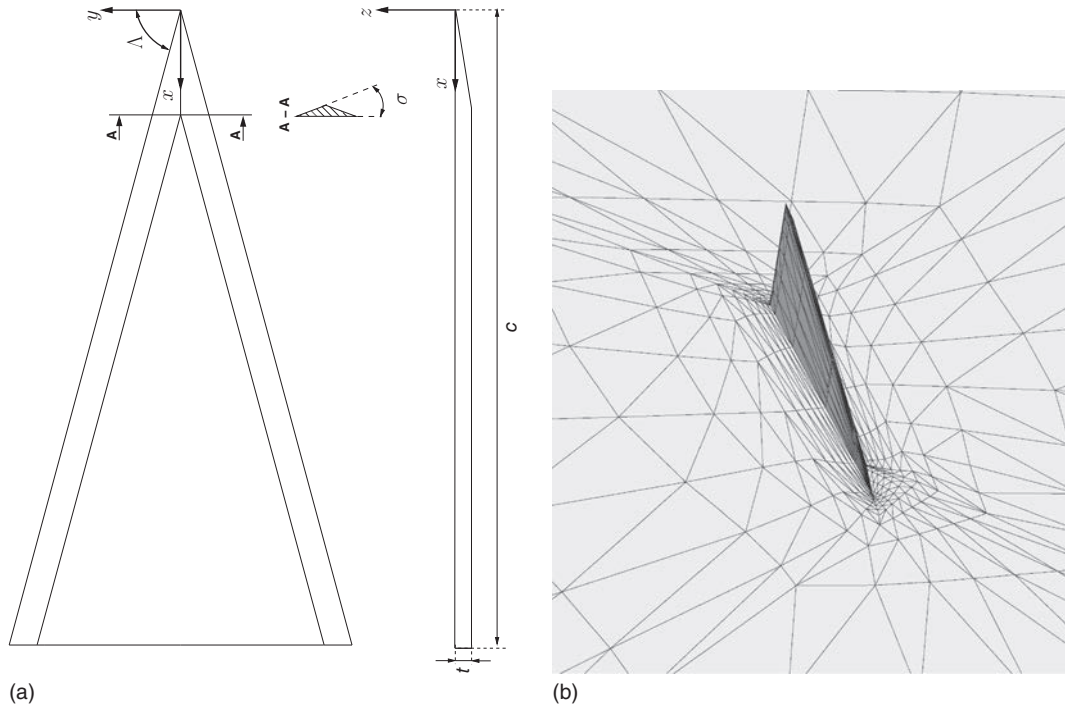


Figure 17. (a) Bottom and side views of the model of the delta wing: $\Lambda = 75^\circ$, $\sigma = 60^\circ$, and $t/c = 0.024$. (b) Coarse mesh of tetrahedra used for the simulations. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

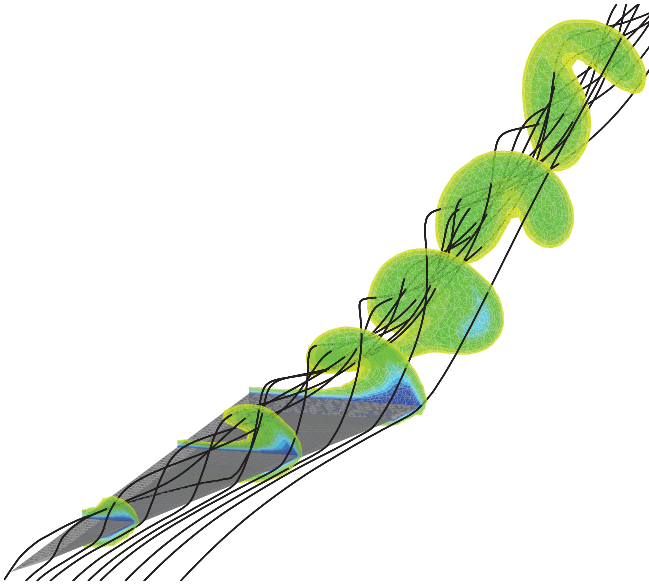


Figure 18. Streamlines and slices of Mach number contours along and behind the delta wing for a third-order simulation on a fine grid. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

and also in the paper by Brufau and Garcia-Navarro, 2003). More recent works, combining high order of accuracy in space and time, the preservation of moving steady states,

robust handling of dry areas, and dispersive extensions, can be found in Ricchiuto *et al.* (2007), Ricchiuto and Bollermann (2009), Ricchiuto (2011b, 2015), Sarmany *et al.* (2013), Ricchiuto and Filippini (2014), and Filippini *et al.* (2016).

4.3.1 *Inundation of a complex three-dimensional beach*

The first example we consider involves the solution of the shallow-water equations, which is taken from Ricchiuto (2015). In this paper, the nonlinear stabilized Lax–Friedrich’s method was combined with the fully explicit time-marching strategy discussed in Section 3.2, and modified to allow a (provable) preservation of the nonnegativity of the water depth. To illustrate the capabilities of the method obtained, we consider a standard benchmark in the oceanography community involving the tsunami runup onto a complex three-dimensional beach. The so-called Monai valley benchmark aims at simulating a scaled down laboratory experiment reproducing the impact of the tsunami wave that hit the Okushiri island in Japan in 1993. The bathymetry and inlet data are available on the web page of the Third International Workshop on Long Wave Runup Models ISEC (see also NOAA Center for Tsunami Research; Liu *et al.*, 2008), with the data relative to the time series of the water level in three gauges close to the

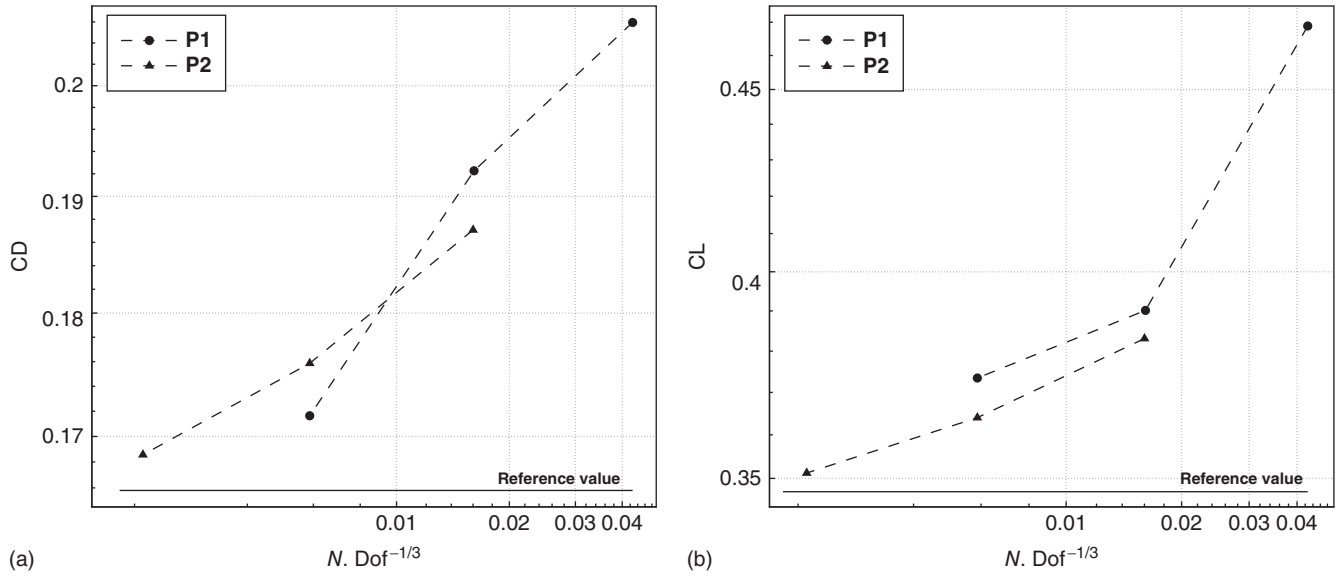


Figure 19. Drag (a) and lift (b) coefficients as function of DOFs for the delta wing simulation with linear and quadratic elements. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

shore. The shape of the bathymetry and of the inlet wave, as well as the position of the three wave gauges, are shown in the left, middle, and right pictures in Figure 24. In the observations (ISEC; NOAA Center for Tsunami Research; Liu *et al.*, 2008), the highest runup is of 32m, and it occurs in the region of the Monai valley where the bathymetry is steepest. For clarity, this region is encircled in the results presented in the following.

The results obtained are summarized in Figures 25 and 26. The top row in the first figure shows the initial withdrawal of the water, followed by the arrival of the main wave. The bottom row shows how, after hitting the beach, the wave gets reflected, and a large wave travels toward the right to hit the steepest slopes in the region of the Monai village. As already mentioned, the highest runup observed is about 32m and it has been observed in the region of the Monai valley, highlighted by a yellow circle in the figure. This is well reproduced by the simulations.

Lastly, we report in Figure 26 the time history of the water level in gauge 5, comparing simulated and measured values, and the runup plot, showing clearly that the deepest inundation point is the region of the Monai village.

4.3.2 Approximation of moving steady states

The super consistency property discussed in Section 3.1.8 also has applications in shallow-water flows. In this case, the state vector \mathbf{w} is defined by the quantities H , the water depth, and \bar{q} , the volume flux $\bar{q} = H\bar{u}$, with \bar{u} the depth-averaged flow velocity. A known steady state involving moving water

is the pseudo-one-dimensional flow characterized by $\bar{q} = \bar{q}_0 = c^t$, and $\mathcal{E} = \mathcal{E}_0$, with \mathcal{E} the total energy $g(H + b) + \bar{u} \cdot \bar{u}/2$, and $b = b(x, y)$ the bathymetry. This solution allows us to check numerically Proposition 6. To do this, we consider the tests discussed in Ricchiuto (2011b, 2015).

The first involves a small perturbation of the initial steady state over a bathymetry with C^1 regularity obtained as a series of ribs defined by truncated \sin^2 functions. The evolution of the perturbation on an irregular triangulation is studied. The typical result is shown in Figure 27, showing a 3D view of the free surface level. The left picture is obtained with a standard scheme based on a \mathbb{P}^1 approximation of the state vector \mathbf{w} and of the bathymetry. The right result is obtained with the scheme based on a direct approximation of the total energy \mathcal{E} and of the flux \bar{q} , and with a higher order approximation and quadrature of the bathymetric gradient. The improvement is quite remarkable.

The second test consists in verifying the property of Proposition 6 by computing, on irregular triangulations, the solution error at a finite time when starting from the exact nodal steady state. This is done with bathymetries of increasing smoothness and with volume and edge quadrature strategies of increasing accuracy. The results are summarized in Figure 28 in which (a), (b), (c), and (d) show the grid convergence obtained on bathymetries with different regularity when using quadrature strategies with errors of order h^2 , h^4 , h^6 , and h^8 , respectively. The last column shows the error convergence on a fixed mesh when increasing the accuracy of the quadrature. In particular, picture (e) is obtained on the coarsest mesh used in the convergence study, while

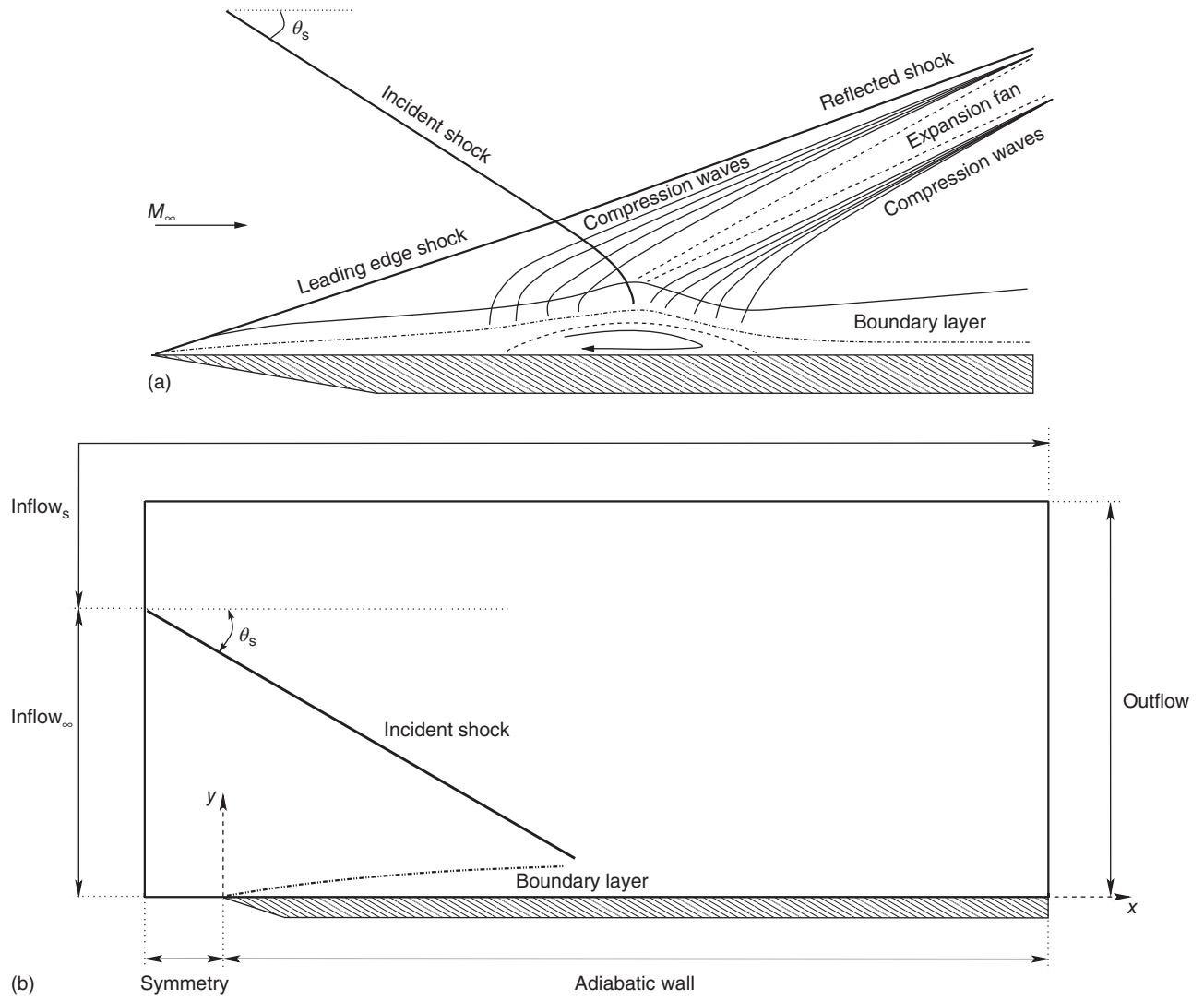


Figure 20. Schematic representation of the wave pattern (a) and computational domain with boundary conditions (b) for the shock wave/boundary layer interaction problem. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

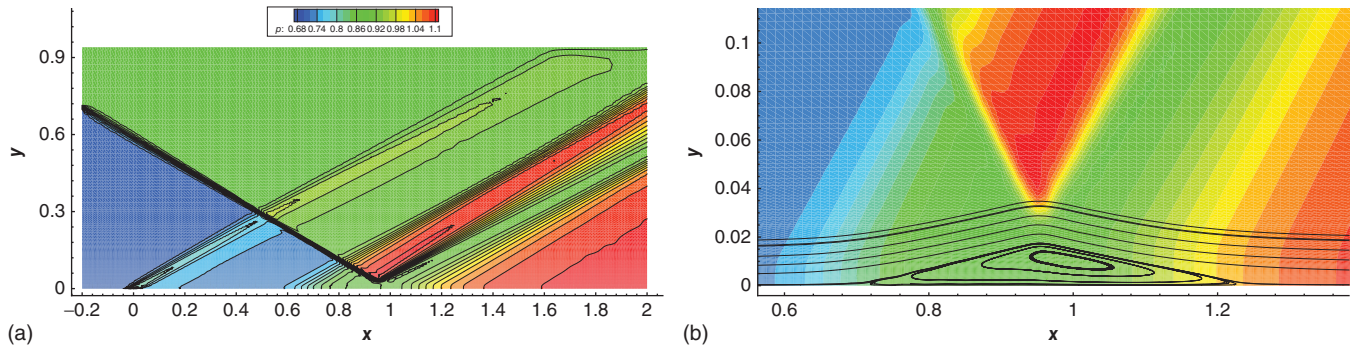


Figure 21. (a) Contours of the pressure obtained with the third-order scheme for the shock/boundary layer interaction. (b) Zoom of the solution near the impinging point of the shock with the boundary layer. Streamlines are also reported to show the separation bubble. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

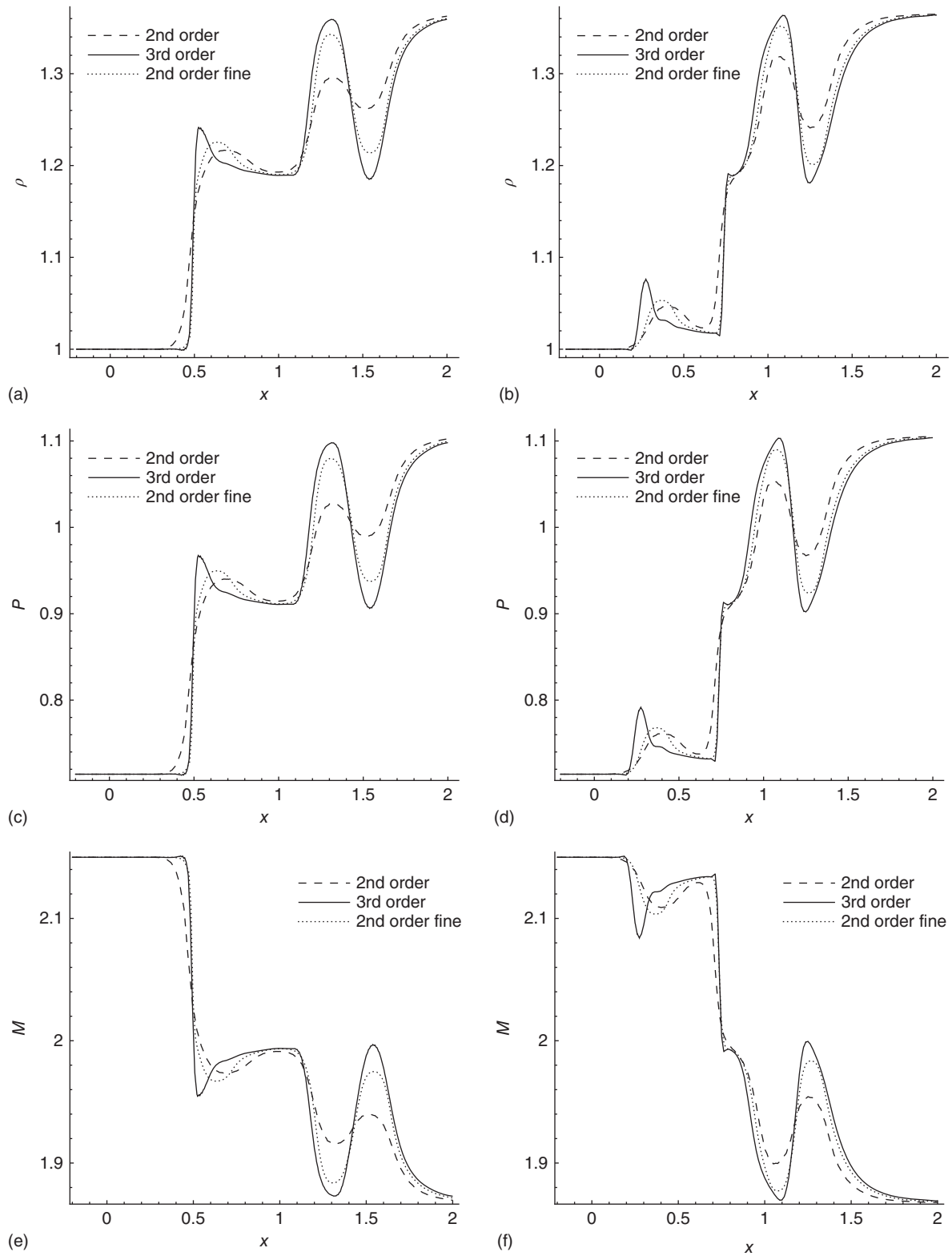


Figure 22. Density, pressure, and Mach number profiles along the line $y = 0.29$ (a, c, e) and the line $y = 0.15$ (b, d, f) for the shock/boundary layer interaction problem. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

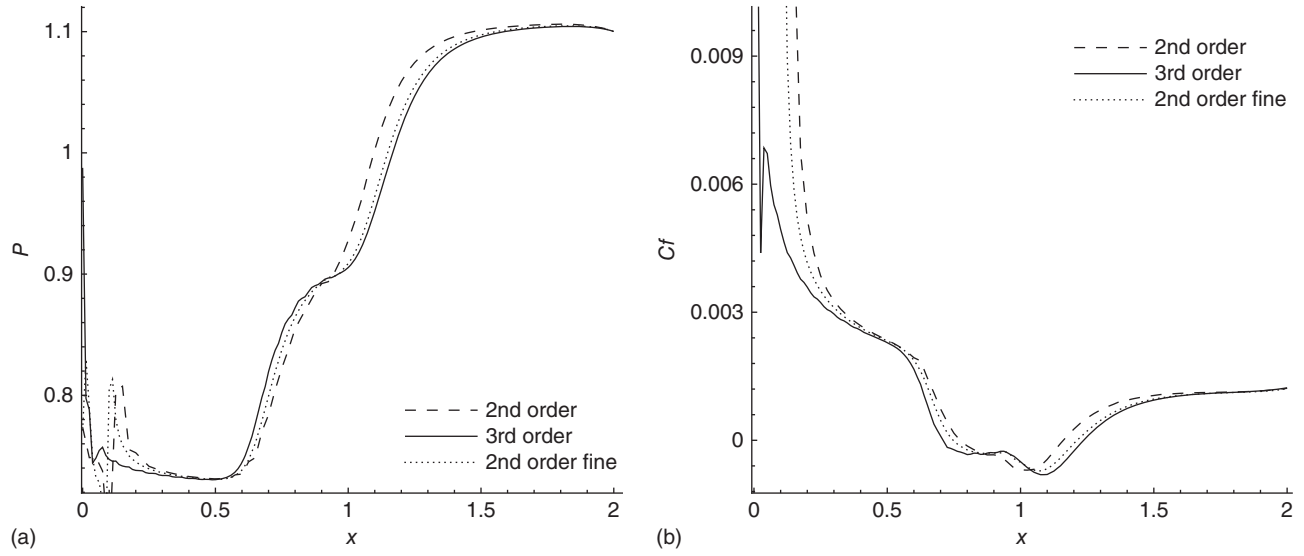


Figure 23. Pressure (a) and skin friction (b) profiles along the flat plate for the shock/boundary layer interaction problem. (Reproduced with permission from Abgrall and de Santis (2015). © Elsevier, 2015.)

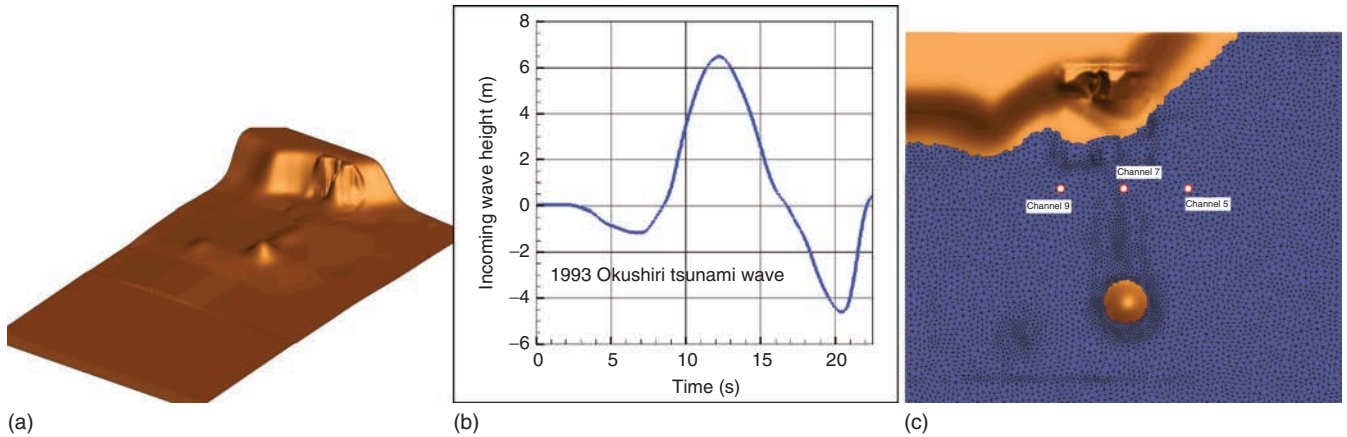


Figure 24. Monai valley benchmark. (a) Bathymetry. (b) Incoming wave. (c) Position of the wave gauges. (Reproduced with permission from Ricchiuto (2015). © Elsevier, 2015.)

picture (f) on the finest. The underlying approximation is P^1 . Not only this result confirms the super consistency analysis but it also shows that, for exact quadrature, the residual approach would yield exact preservation of the steady state.

For additional examples involving other steady state solutions, the interested reader is referred to Ricchiuto (2011a, b, 2015).

4.3.3 Residual-based stabilized methods for dispersive waves

Another challenging application in free surface flows is the inclusion of non-hydrostatic effects in depth-averaged

models. The interested reader may consult the review papers (Kirby, 2003; Brocchini, 2013) and the book (Lannes, 2013) for an overview of the modeling issues. Concerning numerics, the typical form of a depth-averaged Boussinesq-type model is

$$\partial_t K + \nabla \cdot \mathbf{f}(\mathbf{w}) + \mathbf{s}(\mathbf{w}, \mathbf{x}) = 0 \quad (91)$$

where the quantity $K(\mathbf{w})$ is related to the state vector by

$$\mathbf{w} - \mathcal{T}(\mathbf{w}) = K \quad (92)$$

where $\mathcal{T}(\cdot)$ is a nonlinear elliptic operator. Here, physical dispersion is present in the PDE. The challenge is thus

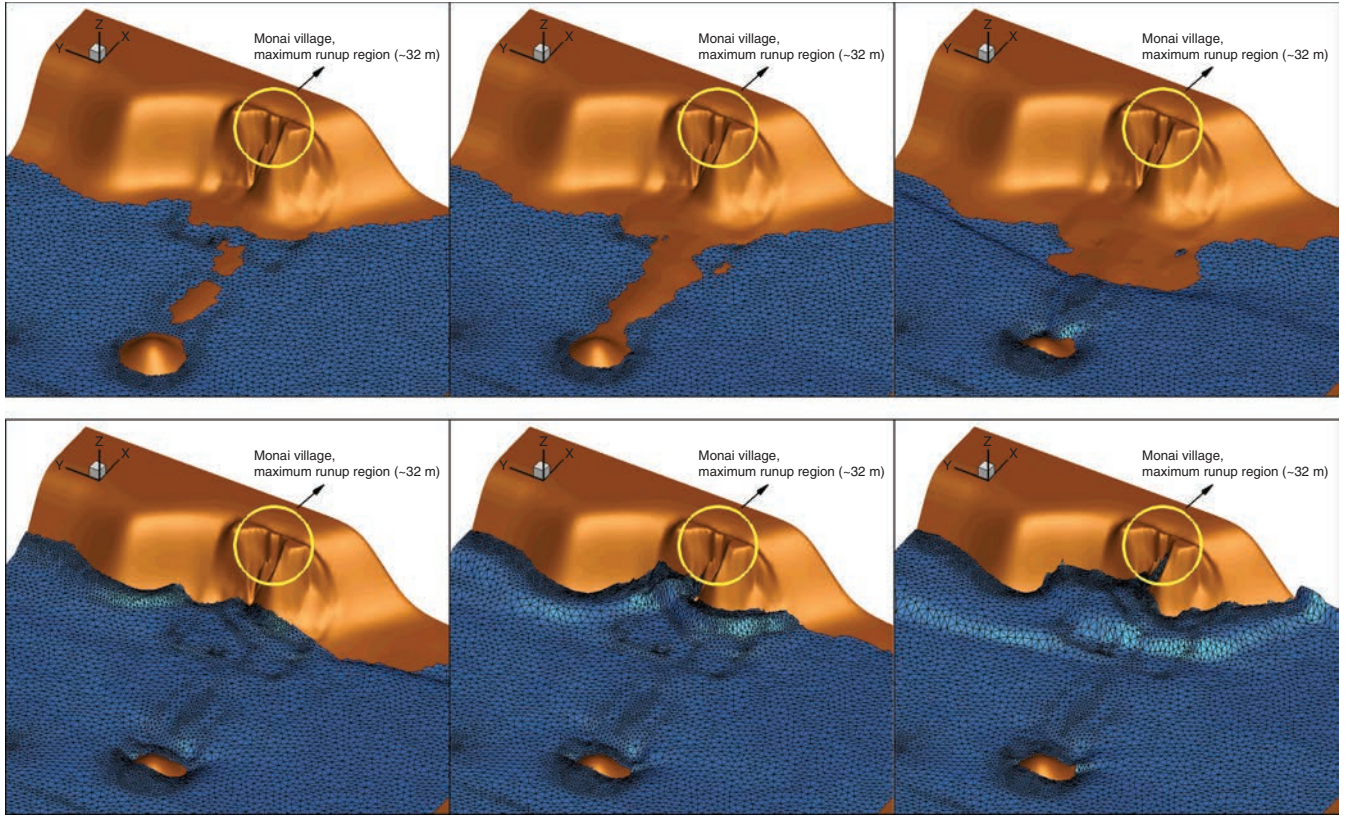


Figure 25. Monai valley benchmark. 3D view of the inundation process. (Reproduced with permission from Ricchiuto (2015). © Elsevier, 2015.)

to design a numerical method with low dissipation and very low dispersion errors to allow long-time integration of propagating waves while, however, guaranteeing a sufficient degree of dissipation to avoid spurious modes. The use of some stability mechanism is also required, as the term $\mathcal{T}(\mathbf{w})$ is often neglected locally to recover the hyperbolic shallow-water equations, and model breaking regions as moving bores (Tonelli and Petti, 2011; Bonneton *et al.*, 2011; Kazolea *et al.*, 2014; Filippini *et al.*, 2016). The requirement is then to have a low dissipation/dispersion method that is capable of handling both the parabolic Boussinesq equations and the hyperbolic shallow-water ones, with eventually capabilities for capturing shocks and dry areas.

This has led to the work presented in Ricchiuto and Filippini (2014), Bacigaluppi *et al.* (2014), Filippini *et al.* (2016), which has tried to extend upwind and multidimensional upwind residual-based stabilization techniques to these systems. The main idea is to decouple the approximation of the two subproblems above. The elliptic step (92) is solved with a standard C^0 Galerkin method, while an upwind scheme is used in the evolution step (91). The work discussed in the references shows evidence that this

approach is a sound one, and provided the hyperbolic step is solved with at least third-order of accuracy, the elliptic phase can be solved with a second-order method without affecting the dispersion accuracy. This generalizes on unstructured grids, and to residual-based stabilized method, an idea proposed in the finite difference context by Wei and Kirby (1995). The schemes obtained all reduce in one dimension to a streamline upwind method stabilizing the Galerkin approximation of the first-order PDE (91) with cell integrals depending on the residual of (91), and on the sign of the shallow-water Jacobians. In two space dimensions, both a standard streamline upwind formulation and a multidimensional upwind variant based on the LDA method (61) have been proposed in Ricchiuto and Filippini (2014).

We present here three results. The first is the characterization of the accuracy of the methods obtained. Figure 29 shows the results relative to a second-order Galerkin approximation of (92) and a third-order streamline upwind (SU) or fourth-order Galerkin (cG) approximation of (91). In particular, the left picture provides a numerical convergence study on a propagating solitary wave. Despite the second-order

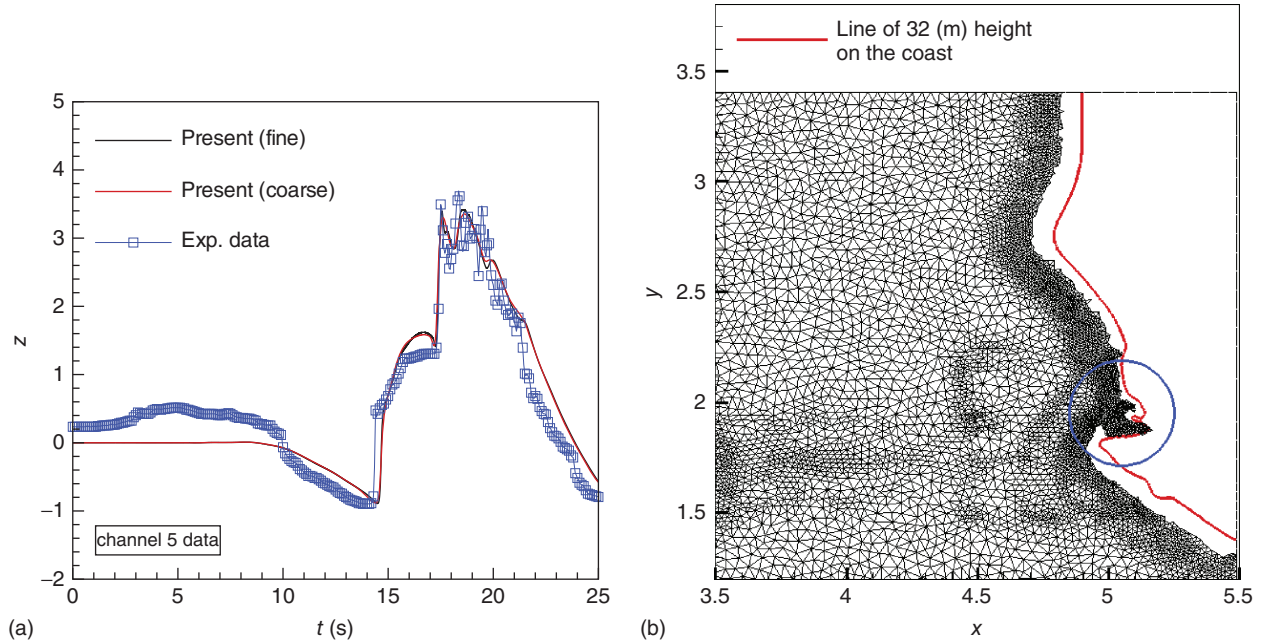


Figure 26. Monai valley benchmark. Time series in gauge 5 (a), and runup plot (b). (Reproduced with permission from Ricchiuto (2015). © Elsevier, 2015.)

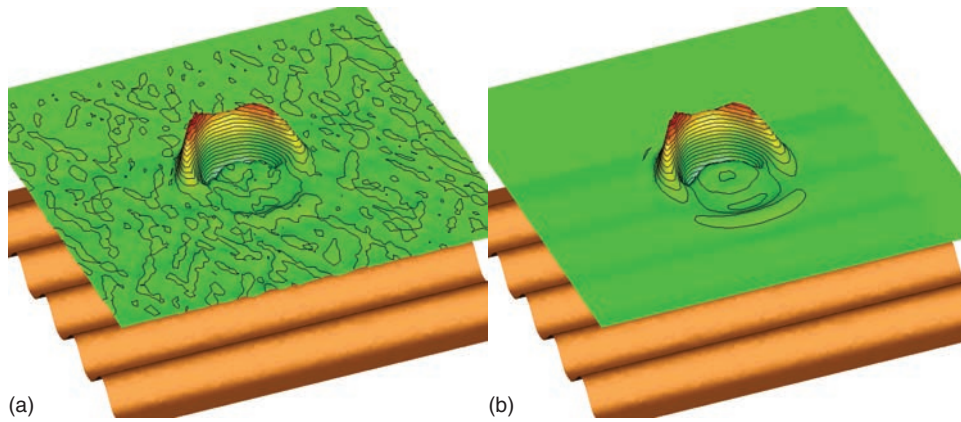


Figure 27. Moving steady states: evolution of a small perturbation in a homo-energetic steady state. 3D plot of the free surface. (a) Approximation in physical variables. (b) Approximation in steady invariants. (Reproduced with permission from Ricchiuto (2015). © Elsevier, 2015.)

treatment of the elliptic term, the overall accuracy measured for a propagating solution is 3. More importantly, the middle and right pictures study the dispersion errors of the schemes and compare them to second- and fourth-order finite differencing. The result shows two important features: both the cG and SU are as good as or better than the fourth-order finite difference method; and for propagating solutions, the upwind SU stabilization actually improves the dispersion properties of the scheme providing lower dispersion errors, especially for shorter waves.

The second result tests the ability of the proposed method to correctly reproduce the energy exchange between different harmonics when monochromatic waves shoal on a 2D circular shelf. This is a standard benchmark for multidimensional Boussinesq-type codes (see Sørensen and Madsen, 1992; Beji and Nadaoka, 1996; Walkley and Berzins, 2002; Sørensen, 2004 *et al.*; Eskilsson *et al.*, 2006; Tonelli and Petti, 2009; Kazolea *et al.*, 2012 and references therein). In Whalin (1971), experiments were conducted in several configurations involving values of the period and amplitude

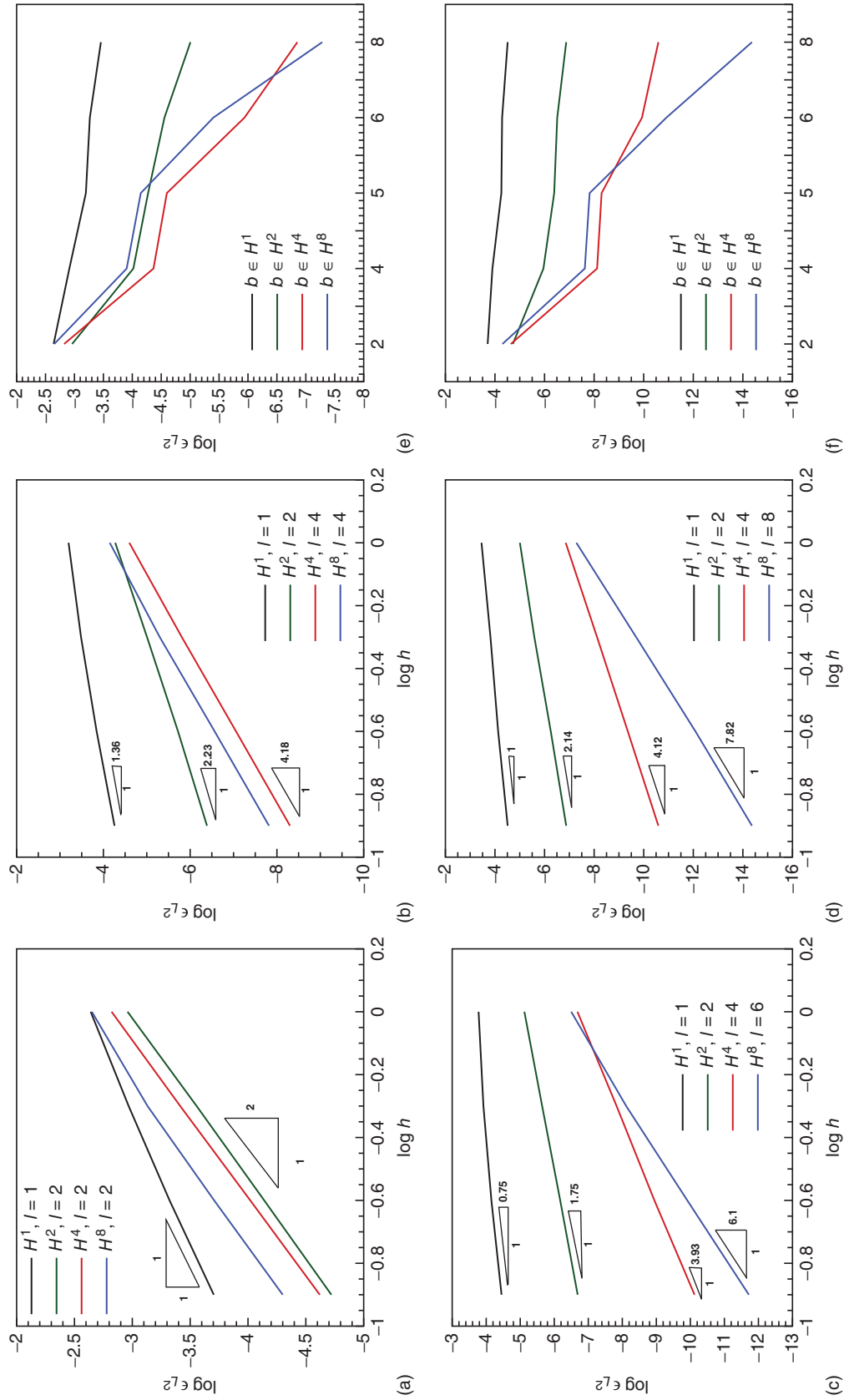


Figure 28. Moving steady states: super consistency of the scheme. (a–d) Grid convergence for different quadrature strategies. (e and f) Quadrature convergence on the coarsest (e) and finest (f) grid. (Reproduced with permission from Ricchiuto (2015). © Elsevier, 2015.)

for the incoming monochromatic wave. Here we discuss the results for case (i) with $T = 2$ s, $A = 0.0075$ m, $h_0/\lambda = 0.117$; case (ii) with $T = 1$ s, $A = 0.0195$ m, $h_0/\lambda = 0.306$. The first case has a relatively weak dispersive character, but presents an important energy transfer to higher harmonics. The second case is quite demanding, as it involves a higher dispersion degree outside the validity of the most simple Boussinesq models. Figure 30 summarizes the results obtained by solving the enhanced Boussinesq equations of Schaffer and Madsen (1995) on unstructured triangulations. Both a “classical” streamline upwind stabilization and a multidimensional one based on the LDA distribution (61) were tested. The pictures clearly show that these multidimensional stabilized methods have a high potential in resolving the energy transfer between harmonics, also in the more demanding cases.

Finally, we show the results obtained on an experiment carried out in Berkhoff *et al.* (1982) and involving the refraction and diffraction of monochromatic waves over a complex bathymetry. A sketch of the experiment is shown in Figure 31(a). The bathymetry involves a shoal presenting a constant angle with the main incoming wave direction, with an elliptic bump, which leads to a complex multidimensional wave pattern that involves dispersive effects both in the main wave direction and along the orthogonal direction. As shown in the sketch in Figure 31, the experiments provide the normalized time average of the water height in eight different sections. Profiting from the general formulation used here, the problem is solved on an unstructured triangulation refined in correspondence to the sampling region, as shown in Figure 31(b).

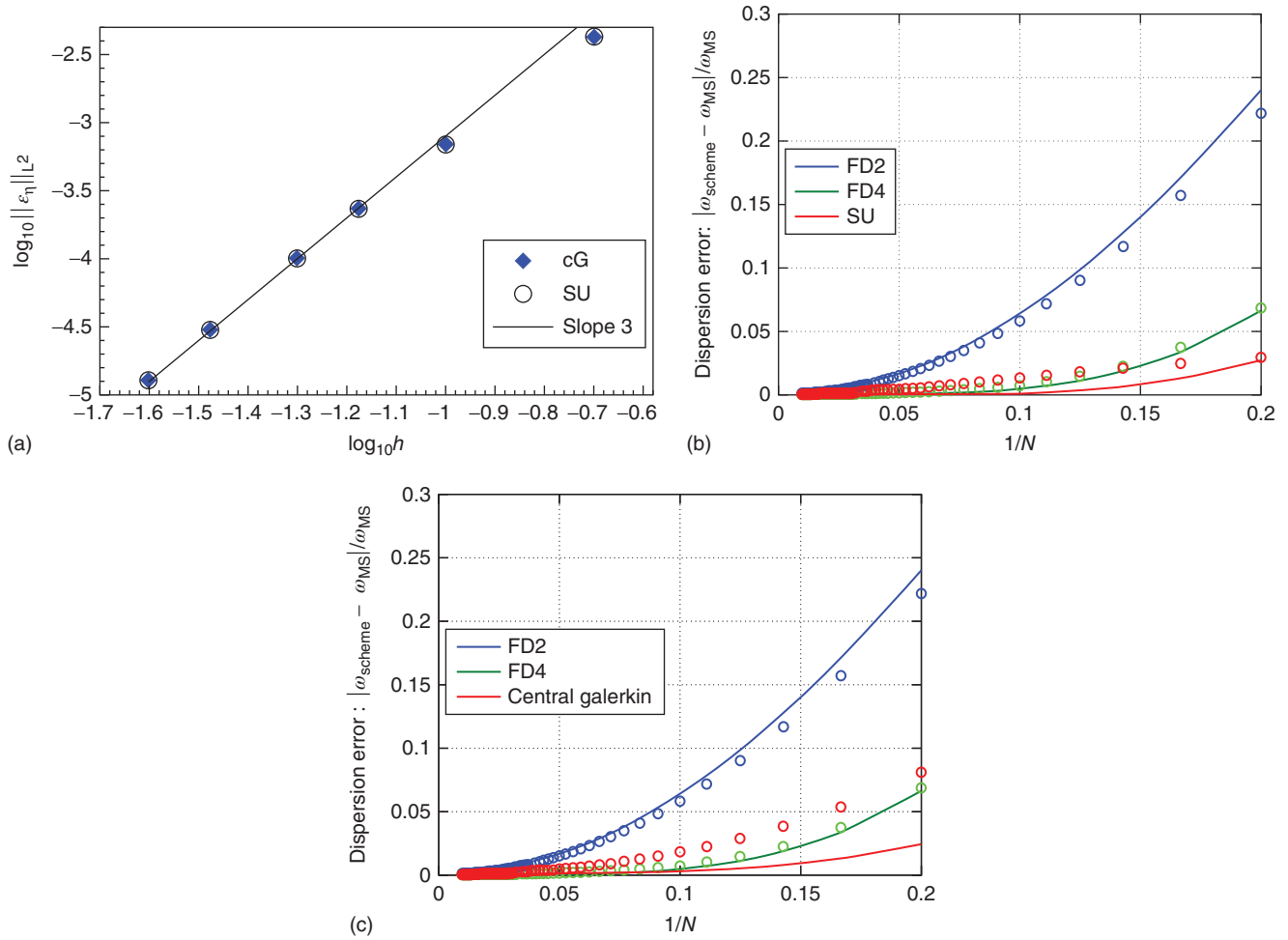


Figure 29. Accuracy of residual approximation of dispersive equations. (a) Grid convergence study on an exact solitary wave solution. (b and c) Dispersion error in function of the number of nodes per wavelength for the upwind stabilized and unstabilized scheme, and comparison with second- and fourth-order finite differencing. Solid line: $kh_0 = 0.5$ (long wave). Circles: $kh_0 = 2.6$ (“short” wave). (Reproduced with permission from Ricchiuto and Filippini (2014). © Elsevier, 2014.)

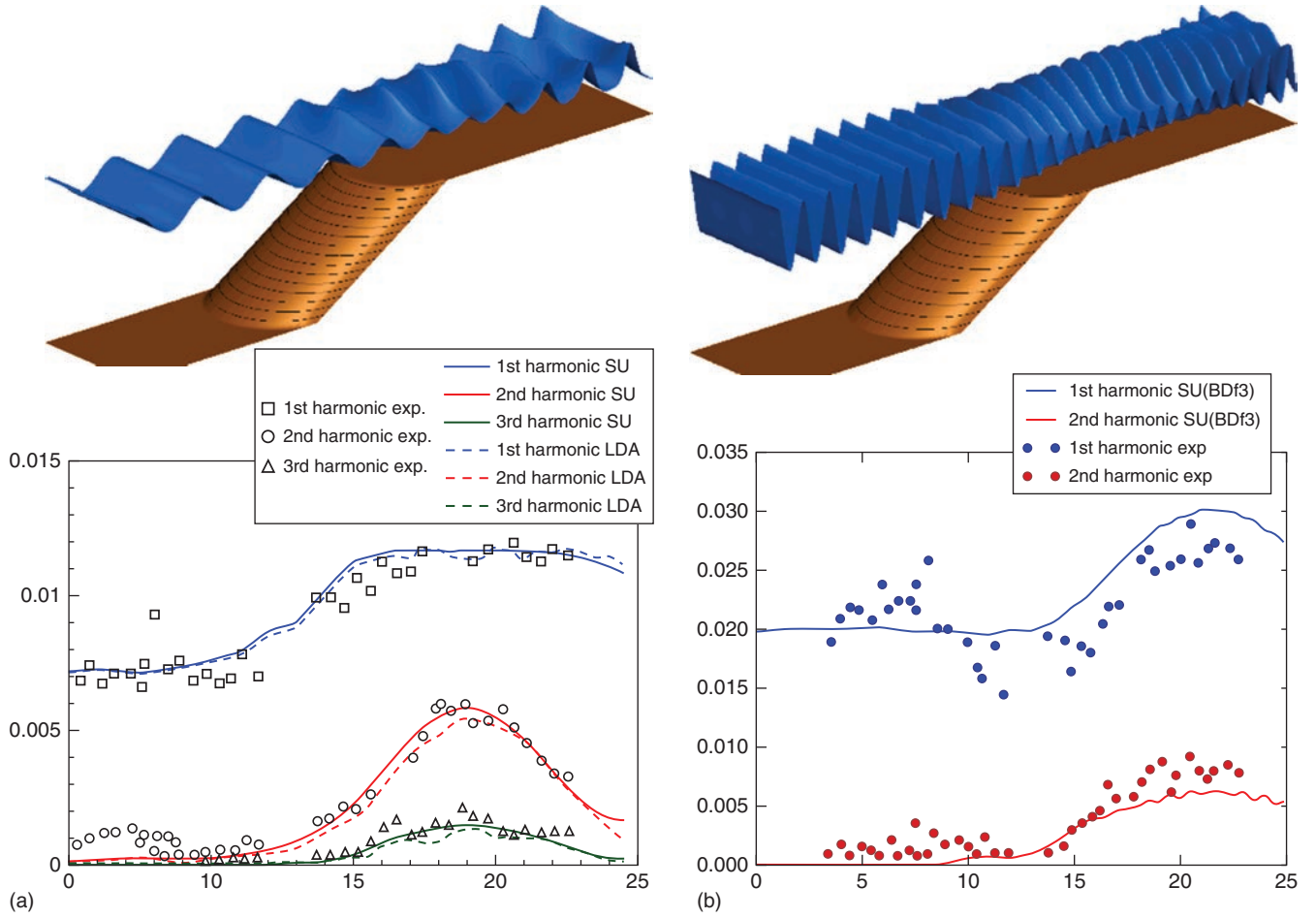


Figure 30. Wave diffraction on a circular shoal, case (i) (a) and case (ii) (b). (Reproduced with permission from Ricchiuto and Filippini (2014). © Elsevier, 2014.)

On the same figure, the typical instantaneous wave pattern obtained is also shown. One can clearly see the effect of the submerged feature in diffracting the incoming waves. To provide a more quantitative appreciation of the result, the comparison with the experiments is shown for three of the eight sections in Figure 32. The results, again obtained with two different upwind (and multidimensional upwind) stabilization approaches, confirm the potential of residual-based methods in capturing complex dispersive wave phenomena.

5 CONCLUSION, OPEN CHALLENGES

Over the years, the residual distribution technique has proven that continuous finite elements allow the same flexibility as discontinuous finite elements. The stencils are comparable, in particular for viscous calculations, and fewer degrees

of freedom are always needed, even though the difference between completely discontinuous approximation and continuous ones tend to become smaller and smaller as the polynomial degree increases. We have also shown that all these methods are *locally conservative*, contrary to common belief. The techniques developed here show that the schemes are very robust. We could not show all possible results, but simulations for hypersonic flows are possible without major difficulties. We have also shown that iterative convergence to machine zero is possible even for turbulent flows, see (De Santis, 2015).

However, all the problems have not been solved so far:

- High-order and unsteady problems. This chapter has presented a couple of solutions for geophysical flows. Other examples, related to compressible flow problems, can be found in Ricchiuto and Abgrall (2010),

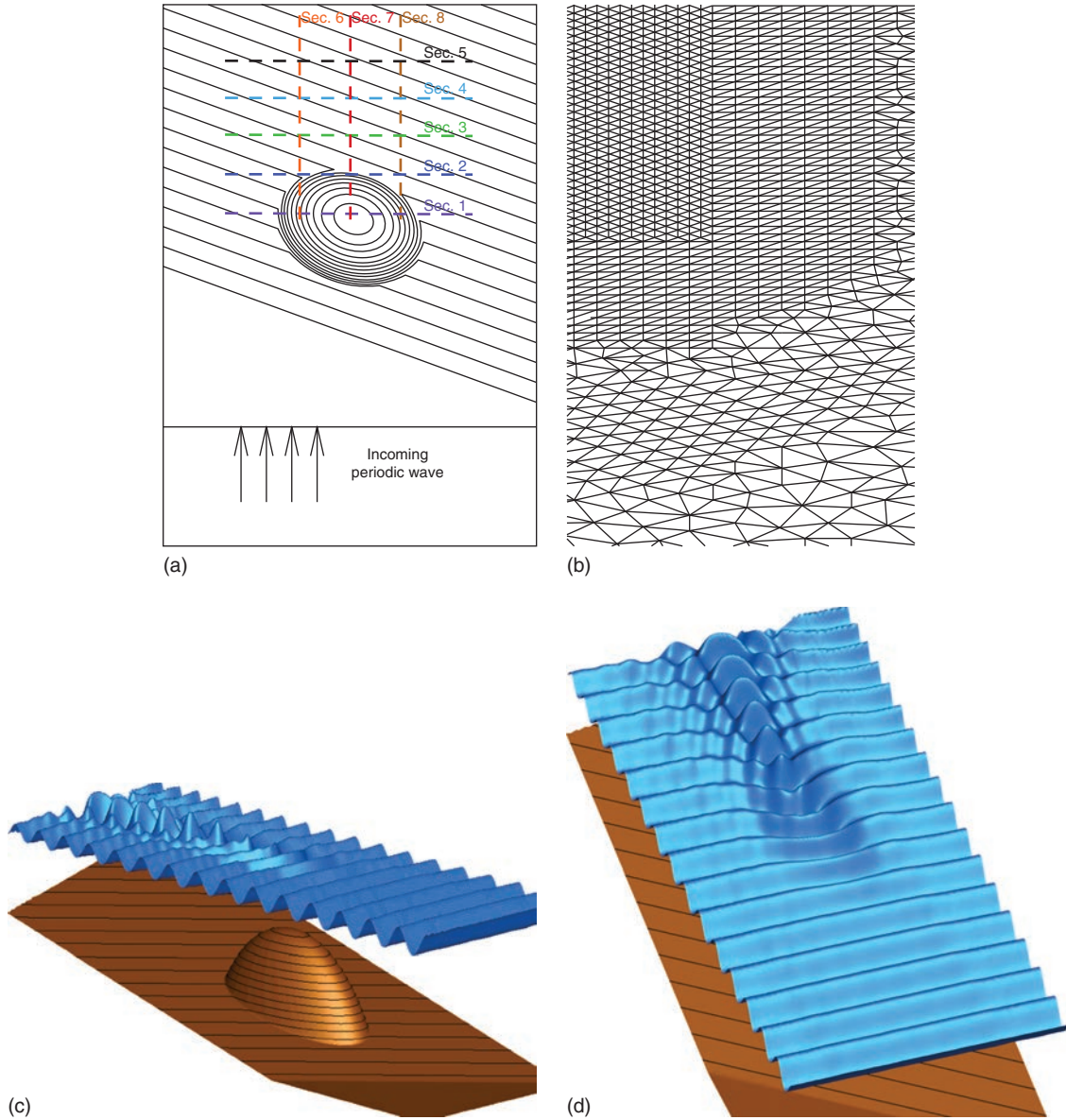


Figure 31. Wave scattering on an elliptic shoal. Problem sketch (a), close up of the grid (b), and instantaneous wave patterns (c and d). (Reproduced with permission from Ricchiuto and Filippini (2014). © Elsevier, 2014.)

where a fully explicit method is described, or (Sarmany *et al.*, 2013; Abgrall *et al.*, 2005) for implicit technique. Considering now higher than second order in time, research is still needed, but see Abgrall *et al.* (2016c) for a fully explicit (i.e., mass matrix free) technique for linear problems. The same technique can be applied for nonlinear problems.

- Error estimation and adaptation. Some work on adjoint problems in the RD framework has been done by D'Angelo *et al.* (2011, 2015).

- p-Adaptation. Continuous and discontinuous approximation. Some work in that direction has been done in Abgrall *et al.* (2016a, b).

ACKNOWLEDGMENTS

This work would not have been possible without the contributions, suggestions, and help of our collaborators, students, and friends (in alphabetical order: P. Congedo (INRIA), H. Deconinck (VKI), S. D'Angelo (VKI, now Credo

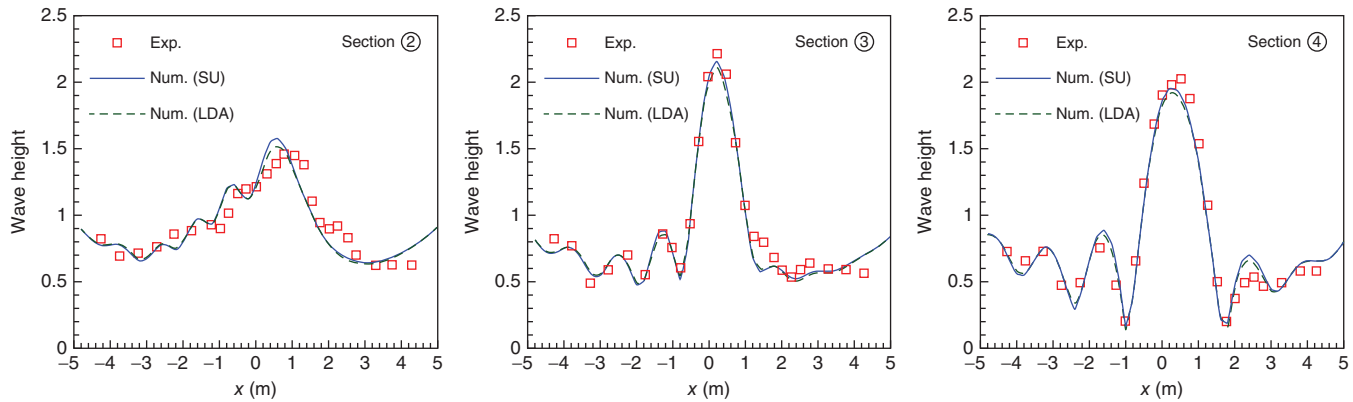


Figure 32. Wave scattering on an elliptic shoal. Time average of the water height: simulations versus experiments. (Reproduced with permission from Ricchiuto and Filippini (2014). © Elsevier, 2014.)

Consulting), D. De Santis (INRIA, now Stanford University), A. Filippini (Inria), M. Hubbard (Nottingham University), A. Larat (INRIA, now Ecole Centrale Paris), M. Vymazal (VKI, now Imperial College), and many others. RA has been conducting this work with the partial support of SNFS grant # 200021_153604. MR has been partly funded by the TANDEM project (reference ANR-11-RSNR-0023-01 French *Programme Investissements d'Avenir*).

NOTES

1. This aspect can be explained in more rigorous terms and made systematic if taken into account from the start.
2. A similar explicit construction can be done in the three space dimensions including high-order tets, prisms, and hexas. Details are left out.
3. The results above can be easily generalized to the conditions (54) provided \hat{f} is Lipschitz continuous.
4. This we may always do in the linear case by rescaling the boundary data g by $M' = |\min_{x \in \Omega} g| + M$, $M > 0$.
5. We have set here $\mathcal{T}_K = \text{diag}(\tau_a, \delta_v, \delta_v)$.

REFERENCES

- Abgrall R. On essentially non-oscillatory schemes on unstructured meshes: analysis and implementation. *J. Comput. Phys.* 1994; **114**(1):45–58.
- Abgrall R. Toward the ultimate conservative scheme: following the quest. *J. Comput. Phys.* 2001; **167**(2):277–315.
- Abgrall R. Essentially non oscillatory residual distribution schemes for hyperbolic problems. *J. Comput. Phys.* 2006; **214**(2):773–808.
- Abgrall R. A residual method using discontinuous elements for the computation of possibly non smooth flows. *Adv. Appl. Math. Mech.* 2010; **2**(1):32–44.
- Abgrall R, Andrianov N and Mezine M. Towards very high-order accurate schemes for unsteady convection problems on unstructured meshes. *Int. J. Numer. Methods Fluids* 2005; **47**(8-9):679–691.
- Abgrall R and Barth TJ. Residual distribution schemes for conservation laws via adaptive quadrature. *SIAM J. Sci. Comput.* 2002; **24**(3):732–769.
- Abgrall R, Beaugendre H, Dobzinsky C and Viville Q. p -adaptation is possible with continuous finite elements: the Euler equations case. *J. Sci. Comput.* 2017, in revision, also https://hal.inria.fr/hal-01225583/file/RR_8808.pdf
- Abgrall R, Pacigaluppi P and Tokareva S. How to avoid mass matrix for linear hyperbolic problems. In *Numerical Mathematics and Advanced Applications- ENUMATH 2015, Lecture notes in Computational Science and Engineering*, Karasözen B, Manguoglu M, Tezer-Sezgin M, Göktepe S and Ugur Ø (eds). Springer, 2016c.
- Abgrall R and de Santis D. Linear and non-linear high order accurate residual distribution schemes for the discretization of the steady compressible Navier-Stokes equations. *J. Comput. Phys.* 2015; **283**:329–359.
- Abgrall R, de Santis D and Ricchiuto M. High-order preserving residual distribution schemes for advection-diffusion scalar problems on arbitrary grids. *SIAM J. Sci. Comput.* 2014a; **36**(3):A955–A983.
- Abgrall R, Krust A, Ricchiuto M and de Santis D. Numerical approximation of parabolic problems by means of residual distribution schemes. *Int. J. Numer. Methods Fluids* 2013; **71**(9):1191–1206.
- Abgrall R, Larat A, Ricchiuto M and Tavé C. A simple construction of very high order non-oscillatory compact schemes on unstructured meshes. *Comput. Fluids* 2009; **38**(7):1314–1323.
- Abgrall R, Larat A and Ricchiuto M. Construction of very high order residual distribution schemes for steady inviscid flow problems on hybrid unstructured meshes. *J. Comput. Phys.* 2011; **230**(11):4103–4136.

- Abgrall R and Mezine M. Construction of second-order accurate monotone and stable residual distribution schemes for unsteady flow problems. *J. Comput. Phys.* 2003; **188**:16–55.
- Abgrall R and Roe PL. High-order fluctuation schemes on triangular meshes. *J. Sci. Comput.* 2003; **19**(3):3–36.
- Abgrall R and Shu C-W. Development of residual distribution schemes for the discontinuous Galerkin method: the scalar case with linear elements. *Commun. Comput. Phys.* 2009; **5**(2-4):376–390.
- Abgrall R and Shu CW (eds). *Chapters for the Handbook of Numerical Methods for Hyperbolic Problems, Handbook of Numerical Analysis*, vol. 17. Elsevier, 2016.
- Abgrall R and Treflik J. An example of high order residual distribution scheme using non-Lagrange elements. *J. Sci. Comput.* 2010; **45**:3–25.
- Atkins HL and Shu C-W. Quadrature-free implementation of discontinuous Galerkin method for hyperbolic equations. *AIAA J.* 1998; **36**:775–782.
- Bacigaluppi P, Ricchiuto M and Bonneton P. A 1D stabilized finite element model for non-hydrostatic wave breaking and run-up. In *Finite Volumes for Complex Applications VII, Springer Proceedings in Mathematics and Statistics*, vol. 77, Fuhrmann J, Ohlberger M and Rohde C (eds). Springer, 2014.
- Balsara D, Altmann C, Munz CD and Dumbser M. A sub-cell based indicator for troubled zones in RKDG schemes and a novel class of hybrid RKDG+HWENO schemes. *J. Comput. Phys.* 2007; **226**(1):586–620.
- Barth TJ. An energy look at the N scheme. Working notes, NASA Ames Research center, CA, USA, 1996.
- Barth TJ. Numerical methods for conservation laws on structured and unstructured meshes. In VKI LS 2003-05, 33rd Computational Fluid dynamics Course. von Karman Institute for Fluid Dynamics, 2003.
- Beji S and Nadaoka K. A formal derivation and numerical modeling of the improved Boussinesq equations for varying depth. *Ocean Eng.* 1996; **23**:691–704.
- Berkhoff JCW, Booy N and Radder AC. Verification of numerical wave propagation models for simple harmonic linear water waves. *Coastal Eng.* 1982; **6**:255–279.
- Biswas R, Devine KD and Flaherty JE. Parallel, adaptive finite element methods for conservation laws. *Appl. Numer. Math.* 1994; **14**:255–283.
- Bochev PB, Gunzburger MD and Shadid JN. Stability of the SUPG finite element method for transient advection-diffusion problems. *Comput. Methods Appl. Mech. Eng.* 2004; **193**(23-26):2301–2323.
- Bonneton P, Chazel F, Lannes D, Marche F and Tissier M. A splitting approach for the fully nonlinear and weakly dispersive Green-Naghdi model. *J. Comput. Phys.* 2011; **230**:1479–1498.
- Brocchini M. A reasoned overview on Boussinesq-type models: the interplay between physics, mathematics and numerics. *Proc. R. Soc. A: Math. Phys. Eng.* 2013; **469**:2160.
- Brufau P and Garcia-Navarro P. Unsteady free surface flow simulation over complex topography with a multidimensional upwind technique. *J. Comput. Phys.* 2003; **186**(2):503–526.
- Burbeau A, Sagaut P and Bruneau Ch-H. A problem-independent limiter for high-order Runge-Kutta discontinuous Galerkin methods. *J. Comput. Phys.* 2001; **169**(1):111–150.
- Burman E. Consistent SUPG-method for transient transport problems: stability and convergence. *Comput. Methods Appl. Mech. Eng.* 2010; **199**(17-20):1114–1123.
- Burman E, Ern A and Fernandez MA. Explicit Runge-Kutta schemes and finite elements with symmetric stabilization for first-order linear PDE systems. *SIAM J. Numer. Anal.* 2010; **48**(6):2019–2042.
- Burman E, Quarteroni A and Stamm B. Interior penalty continuous and discontinuous finite element approximations of hyperbolic equations. *J. Sci. Comput.* 2008; **43**(3):293–312.
- Caraeni DA. *Development of a Multidimensional Upwind Residual Distribution Solver for Large Eddy Simulation of Industrial Turbulent Flows*. PhD thesis, Lund Institute of Technology, 2000.
- Caraeni D and Fuchs L. Compact third-order multidimensional upwind scheme for Navier-Stokes simulations. *Theor. Comput. Fluid Dyn.* 2002; **15**:373–401.
- Cattaneo C. A form of heat-conduction equations which eliminates the paradox of instantaneous propagation. *C.R. Acad. Sci.* 1958; **247**:431–433.
- Chalot F and Normand P-E. Higher-order stabilized finite elements in an industrial Navier-Stokes code. In *ADIGMA - A European Initiative on the Development of Adaptive Higher-Order Variational Methods for Aerospace Applications*, vol. 113, *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, Kroll N, Bieler H, Deconinck H, Couaillier V, van der Ven H and Sorensen K (eds). Springer: Berlin / Heidelberg, 2010; 145–165.
- Chakravarthy SR and Osher SJ. A new class of high accuracy TVD schemes for hyperbolic conservation laws. *AIAA Paper* 85-0363, 1985.
- Chavent G and Cockburn B. The local projection p^0p^1 -discontinuous Galerkin finite element method for scalar conservation laws. *RAIRO Modél. Math. Anal. Numér.* 1989; **23**(4):565–592.
- Chou C-S and Shu C-W. High order residual distribution conservative finite difference WENO schemes for steady state problems on non-smooth meshes. *J. Comput. Phys.* 2006; **214**(3):698–724.
- Ciarlet PG and Raviart PA. General Lagrange and Hermite interpolation in \mathbb{R}^n with applications to finite element methods. *Arch. Ration. Mech. Anal.* 1972; **46**:177–199.
- Cockburn B, Karniadakis GE and Shu CW (eds). *Discontinuous Galerkin Methods- Theory, Computation and Applications, Lecture Notes in Computer Sciences and Engineering*, vol. 11. Springer, 2000.
- Cockburn B and Shu CW. TVB Runge-Kutta local projection discontinuous Galerkin finite element Method for conservation laws. III. *J. Comput. Phys.* 1989a; **84**(1):90–113.
- Cockburn B and Shu CW. TVD Runge-Kutta local projection discontinuous Galerkin Finite element method for conservation laws II: general framework. *Math. Comput.* 1989b; **52**(186):411–435.
- Cockburn B and Shu CW. The local Runge-Kutta local projection discontinuous finite element method for scalar conservation laws. *RAIRO Modél. Math. Anal. Numér.* 1991; **25**(3):337–361.
- Cohen G, Joly P, Roberts JE and Tordjman N. High order triangular finite elements with mass lumping for the wave equation. *SIAM J. Numer. Anal.* 2001; **38**(6):2047–2078.
- D’Angelo S, Ricchiuto M, Abgrall R and Deconinck H. Generalized framework for adjoint error estimation of PG method in linear problems. Research Report RR-7613, INRIA, 2011.

- D'Angelo S, Ricchiuto M and Deconinck H. Adjoint-based error estimation for adaptive Petrov-Galerkin finite element methods. In *38th Lecture Series on Advanced Computational Fluid Dynamics—Adjoint methods and their application in CFD*, von Karman Institute for Fluid Dynamics, 2015.
- Deconinck H and Ricchiuto M. Residual distribution schemes: foundation and analysis. In *Encyclopedia of Computational Mechanics*, Stein E, de Borst R and Hughes TJR (eds). John Wiley & Sons, Ltd., 2007. DOI: 10.1002/0470091355.ecm2054.
- Deconinck H, Roe PL and Struijs R. A multidimensional generalization of Roe's difference splitter for the Euler equations. *Comput. Fluids* 1993; **22**(2/3):215–222.
- De Santis D. High-order linear and non-linear residual distribution schemes for turbulent compressible flows. *Comput. Methods Appl. Mech. Eng.* 2015; **285**:1–31.
- di Pietro DA and Ern A. *Mathematical Aspects of Discontinuous Galerkin Methods, Mathématiques et Applications*, vol. 69. Springer, 2012.
- Dubois F and Le Floch P. Boundary conditions for nonlinear hyperbolic systems of conservation laws. *J. Differ. Equ.* 1988; **71**:93–122.
- Dumbser M. Arbitrary high order PNPM schemes on unstructured meshes for the compressible Navier-Stokes equations. *Comput. Fluids* 2010; **39**(1):60–76.
- Ern A and Guermond J-C. *Theory and Practice of Finite Elements, Applied Mathematical Sciences*, vol. 159. Springer, 2004.
- Eskilsson C, Sherwin SJ and Bergdhal L. An unstructured spectral/hp element model for enhanced Boussinesq-type equations. *Coastal Eng.* 2006; **53**:947–963.
- Ferrante A and Deconinck H. Solution of the unsteady Euler equations using residual distribution and flux corrected transport. Technical Report VKI-PR 97-08, von Karman Institute for Fluid Dynamics, 1997.
- Filippini AG, Kazolea M and Ricchiuto M. A flexible genuinely nonlinear approach for wave propagation, breaking and runup. *J. Comput. Phys.* 2016; **310**:381–417.
- Franca LP, Frey SL and Hughes TJR. Stabilized finite element methods: I. Application to the advective-diffusive model. Technical Report RR-1300, INRIA, 1990.
- Friedrich O. Weighted essentially non-oscillatory schemes for the interpolation of mean values on unstructured grids. *J. Comput. Phys.* 1998; **144**(1):194–212.
- Garcia-Navarro P, Hubbard ME and Priestley A. Genuinely multi-dimensional upwinding for the 2D shallow-water equations. *J. Comput. Phys.* 1995; **121**(1):79–93.
- Giraldo FX and Taylor MA. A diagonal mass-matrix triangular spectral element method based on cubature points. *J. Eng. Math.* 2006; **56**:307–322.
- Godlewski E and Raviart PA. *Numerical Approximation of Hyperbolic Systems of Conservation Laws, Applied Mathematical Sciences*, vol. 118. Springer, 1996.
- Goodman J and LeVeque R. On the accuracy of stable schemes for 2D scalar conservation laws. *Math. Comput.* 1985; **45**(171):15–21.
- Gottlieb S. On high order strong stability preserving Runge–Kutta and multi step time discretizations. *J. Sci. Comput.* 2005; **25**(1):105–128.
- Gottlieb S and Shu C-W. Total variation diminishing Runge–Kutta schemes. *Math. Comput.* 1998; **67**:73–85.
- Gottlieb S, Shu C-W and Tadmor E. Strong stability-preserving high-order time discretization methods. *SIAM Rev.* 2001; **43**(1):89–112.
- Harten A. On the symmetric form of conservation laws with entropy. *J. Comput. Phys.* 1983; **49**:151–164.
- Harten A, Engquist B, Osher S and Chakravarthy SR. Uniformly high order accurate essentially non-oscillatory schemes, III. *J. Comput. Phys.* 1987; **71**(2):231–303.
- Harten A and Osher S. Uniformly high-order accurate nonoscillatory schemes I. *SIAM J. Numer. Anal.* 1987; **24**(2):279–309.
- Hu C and Shu C-W. Weighted essentially non-oscillatory schemes on triangular meshes. *J. Comput. Phys.* 1999; **150**:97–127.
- Hubbard M. Discontinuous fluctuation distribution. *J. Comput. Phys.* 2008; **227**(24):10125–10147.
- Hubbard ME and Baines MJ. Conservative multidimensional upwinding for the steady two-dimensional shallow water equations. *J. Comput. Phys.* 1997; **138**(2):419–448.
- Hubbard M and Ricchiuto M. Discontinuous upwind residual distribution: a route to unconditional positivity and high order accuracy. *Comput. Fluids* 2011; **46**(1):263–269.
- Hubbard M and Roe PL. Compact high resolution algorithms for time dependent advection problems on unstructured grids. *Int. J. Numer. Methods Fluids* 2000; **33**(5):711–736.
- Hughes TJR and Brook A. Streamline upwind Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Comput. Methods Appl. Mech. Eng.* 1982; **32**:199–259.
- Hughes TJR and Mallet M. A new finite element formulation for CFD IV: a discontinuity-capturing operator for multidimensional advective-diffusive systems. *Comput. Methods Appl. Mech. Eng.* 1986; **58**:329–336.
- Hughes ThJR, Franca LP and Mallet M. Finite element formulation for computational fluid dynamics: I symmetric forms of the compressible Euler and Navier Stokes equations and the second law of thermodynamics. *Comput. Methods Appl. Mech. Eng.* 1986; **54**:223–234.
- Hughes TJR, Scovazzi G and Franca LP. *Multiscale and Stabilized Methods*. John Wiley & Sons, Ltd., 2004.
- Hughes TJR and Tezduyar TE. Development of time-accurate finite element techniques for first order hyperbolic systems with emphasis on the compressible Euler equations. *Comput. Methods Appl. Mech. Eng.* 1984; **45**(1-3):217–284.
- Huynh HT, Wang ZJ and Vincent PE. High-order methods for computational fluid dynamics: a brief review of compact differential formulations on unstructured grids. *Comput. Fluids* 2014; **98**:209–220.
- ISEC. Benchmark problem #2, Tsunami runup onto a complex three-dimensional beach. The Third International Workshop on Long-Wave Runup Models, http://isec.nacse.org/workshop/2004_cornell/bmark2.html (accessed 7 December 2016).
- Jiang G and Shu CW. On a cell entropy for Discontinuous Galerkin methods. *Math. Comput.* 1994; **62**(206):531–538.
- Johnson C and Pitkäranta J. An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comput.* 1986; **46**:1–26.

- Johnson C and Szepessy A. On the convergence of a finite element method for a nonlinear hyperbolic conservation law. *Math. Comput.* 1990; **49**:427–444.
- Johnson C, Szepessy A and Hansbo P. On the convergence of shock capturing streamline diffusion finite element methods for hyperbolic conservation laws. *Math. Comput.* 1990; **54**:107–129.
- Kazolea M, Delis AI, Nikolos IK and Synolakis CE. An unstructured finite volume numerical scheme for extended 2d Boussinesq-type equations. *Coastal Eng.* 2012; **69**:42–66.
- Kazolea M, Delis AI and Synolakis CE. Numerical treatment of wave breaking on unstructured finite volume approximations for extended Boussinesq-type equations. *J. Comput. Phys.* 2014; **271**:281–305.
- Kirby JT. Chapter 1 Boussinesq models and applications to nearshore wave propagation, surf zone processes and wave-induced currents. In *Advances in Coastal Modeling, Elsevier Oceanography Series*, vol. 67, Lakhan VC (ed.). Elsevier, 2003; 1–41.
- Klosa J. *Extrapolated BDF Residual Distribution Schemes for the Shallow Water Equations*. Master thesis, 2012.
- Klosa J, Ricchiuto M and Abgrall R. Multi-step and multi-stage explicit time stepping for residual distribution. application to shallow water flows, 2016, in preparation.
- Koloszár L, Villedieu N, Quintino T, Rambaud P, Deconinck H and Anthoine J. Residual distribution method for aeroacoustics. *AIAA J.* 2011; **49**(5):1021–1037.
- Krivodonova L, Xin J, Remacle J-F, Chevaugneand N and Flaherty JE. Shock detection and limiting with discontinuous Galerkin methods for hyperbolic conservation laws. *Appl. Numer. Math.* 2004; **48**:323–338.
- Lannes D. *The Water Waves Problem: Mathematical Analysis and Asymptotics*. American Mathematical Society: Providence, RI, 2013.
- Leicht T and Hartmann R. Error estimation and anisotropic mesh refinement for 3D laminar aerodynamic flow simulations. *J. Comput. Phys.* 2010; **229**(19):7344–7360.
- Lesaint P and Raviart PA. On a finite element method for solving the neutron transport equation. In *Mathematical Aspects of Finite Elements in partial Differential Equations*, Mathematics Resource Center, vol. 33. University of Wisconsin-Madison, Academic Press: New York, 89–123–; 1974.
- LeVeque R (ed.). *Finite Volume Methods for Hyperbolic Problems, Cambridge Texts in Applied Mathematics*. Cambridge University Press, 2002.
- Li G and Qiu J. Hybrid weighted essentially non-oscillatory schemes with different indicators. *J. Comput. Phys.* 2010; **229**:8105–8129.
- Liu X-D, Osher S and Chan T. Weighted essentially non-oscillatory schemes. *J. Comput. Phys.* 1994; **115**(1):200–212.
- Liu PL-F, Yeh H and Synolakis C (eds). *Advanced Numerical Models for Simulating Tsunami Waves and Runup, Advances in Coastal and Ocean Engineering*, vol. 10. World Scientific, 2008.
- Lockard DP and Atkins HL. Efficient implementations of the quadrature free Discontinuous Galerkin method. In *14th AIAA CFD Conference*, Norfolk, VA, 1999.
- Maerz J and Degrez G. Improving time accuracy of residual distribution schemes. Technical Report VKI-PR 96-17, von Karman Institute for Fluid Dynamics, 1996.
- Mazaheri A and Nishikawa H. Improved second-order hyperbolic residual-distribution scheme and its extension to third-order on arbitrary triangular grids. *J. Comput. Phys.* 2015; **300**:455–491.
- Mazaheri A and Nishikawa H. Efficient high-order discontinuous Galerkin schemes with first-order hyperbolic advection-diffusion system approach. *J. Comput. Phys.* 2016. Available online.
- Meng X, Shu CW and Wu B. Optimal error estimates for discontinuous Galerkin methods based on upwind-biased fluxes for linear hyperbolic equations. *Math. Comput.* 2015; **85**(299):1225–1261.
- Mezine M. *Conception de Schémas Distributifs pour l'aérodynamique stationnaire et instationnaire*. PhD thesis, École doctorale de mathématiques et informatique, Université de Bordeaux I, 2002.
- Mezine M, Ricchiuto M, Abgrall R and Deconinck H. Monotone and stable residual distribution schemes on prismatic space-time elements for unsteady conservation laws. In VKI LS 2003-05, 33rd Computational Fluid dynamics Course, von Karman Institute for Fluid Dynamics, 2003.
- Mulder WA. New triangular mass-lumped finite elements of degree six for wave propagation. *Progr. Electromagnet. Res.* 2013; **141**:671–692.
- Nishikawa H. A first-order system approach for diffusion equation. I: Second-order residual-distribution schemes. *J. Comput. Phys.* 2007; **227**(1):315–352.
- Nishikawa H. A first-order system approach for diffusion equation. II: Unification of advection and diffusion. *J. Comput. Phys.* 2010a; **229**(11):3989–4016.
- Nishikawa H. Beyond interface gradient: a general principle for constructing diffusion scheme. In *40th Fluid Dynamics Conference and Exhibit*. AIAA Paper 2010-5093, 2010b.
- Nishikawa H. Robust and accurate viscous discretization via upwind scheme-I: basic principle. *Comput. Fluids* 2011; **49**:62–86.
- Nishikawa H, Rad M and Roe PL. A third-order fluctuation splitting scheme that preserves potential flow. 15th AIAA Computational Fluid Dynamics Conference, Anaheim, CA, USA, June 2001.
- NOAA Center for Tsunami Research. *Tsunami Runup Onto a Complex Three-Dimensional Beach; Monai Valley*. Benchmarks of the NOAA Center for tsunami research, http://nctr.pmel.noaa.gov/benchmark/Laboratory/Laboratory_MonaiValley/.
- Osher S. Riemann solvers, the entropy condition and difference approximations. *SIAM J. Numer. Anal.* 1984; **21**:217–235.
- Paillere H and Deconinck H. Compact cell vertex convection schemes on unstructured meshes. In *Notes on Numerical Fluid Mechanics*, Deconinck H and Koren B (eds). Vieweg-Verlag: Braunschweig, Germany, 1997; 1–50.
- Qiu J and Shu C-W. A comparison of trouble cell indicators for Runge-Kutta discontinuous Galerkin method using WENO limiters. *SIAM J. Sci. Comput.* 2005a; **27**:995–1013.
- Qiu J and Shu C-W. Hermite WENO schemes and their application as limiters for Runge-Kutta discontinuous Galerkin method II: two dimensional case. *Comput. Fluids* 2005b; **34**:642–663.
- Reed WH and Hill TR. Triangular mesh methods for the neutron transport equation. Technical Report LA-UR-73-0479, Los Alamos Scientific Laboratory, 1973, <http://lib-www.lanl.gov/cgi-bin/getfile?00354107.pdf> (accessed 5 December 2016).

- Ricchiuto M. *Contributions to the Development of Residual Discretizations for Hyperbolic Conservation Laws with Application to Shallow Water Flows*. Habilitation à diriger des recherches, Université Sciences et Technologies – Bordeaux I, 2011a.
- Ricchiuto M. On the C-property and generalized C-property of residual distribution for the shallow water equations. *J. Sci. Comput.* 2011b; **48**:304–318.
- Ricchiuto M. An explicit residual based approach for shallow water flows. *J. Comput. Phys.* 2015; **280**:306–344.
- Ricchiuto M and Abgrall R. Stable and convergent residual distribution for time-dependent conservation laws. In *ICCFD4 International Conference on Computational Fluid Dynamics 4*, Ghent, Belgium, July 2006.
- Ricchiuto M and Abgrall R. Explicit Runge-Kutta residual distribution schemes for time dependent problems: second order case. *J. Comput. Phys.* 2010; **229**(16):5653–5691.
- Ricchiuto M, Abgrall R and Deconinck H. Construction of very high order residual distribution schemes for unsteady advection: preliminary results. In VKI LS 2003-05, 33rd Computational Fluid dynamics Course, von Karman Institute for Fluid Dynamics, 2003.
- Ricchiuto M, Abgrall R and Deconinck H. Application of conservative residual distribution schemes to the solution of the shallow water equations on unstructured meshes. *J. Comput. Phys.* 2007; **222**:287–331.
- Ricchiuto M and Bollermann A. Stabilized residual distribution for shallow water simulations. *J. Comput. Phys.* 2009; **228**(4):1071–1115.
- Ricchiuto M, Csík A and Deconinck H. Conservative residual distribution schemes for general unsteady systems of conservation laws. In *ICCFD3 International Conference on Computational Fluid Dynamics 3*, Toronto, Canada, July 2004.
- Ricchiuto M, Csík A and Deconinck H. Residual distribution for general time dependent conservation laws. *J. Comput. Phys.* 2005; **209**(1):249–289.
- Ricchiuto M and Deconinck H. Time accurate solution of hyperbolic partial differential equations using fact and residual distribution. VKI report VKI SR1999-33, 1999.
- Ricchiuto M and Filippini AG. Upwind residual discretization of enhanced Boussinesq equations for wave propagation over complex bathymetries. *J. Comput. Phys.* 2014; **271**:306–341.
- Ricchiuto M, Villedieu N, Abgrall R and Deconinck H. On uniformly high-order accurate residual distribution schemes for advection-diffusion. *J. Comput. Appl. Math.* 2008; **215**(2):547–556.
- Ringleb F. Exakte Lösungen der Differentialgleichungen einer abadiatischen Gassströmung. *Z. Angew. Math. Mech.* 1940; **20**(4):185–198.
- Roe PL. Approximate Riemann solvers, parameter vectors, and difference schemes. *J. Comput. Phys.* 1981; **43**:357–372.
- Roe PL. Fluctuations and signals – a framework for numerical evolution problems. In *Numerical Methods for Fluids Dynamics*, Morton KW and Baines MJ (eds). Academic Press: 1982; 219–257.
- Roe PL. Linear advection schemes on triangular meshes. Technical Report CoA 8720. Cranfield Institute of Technology, 1987.
- Roe PL. The “optimum” upwind advection on a triangular mesh. Technical Report ICASE 90-75. ICASE, 1990.
- Roe PL. Multidimensional upwinding: motivation and concepts. In *VKI Lecture Series: Computational Fluid Dynamics*. VKI-LS 1994-05, 1994.
- Roe PL and Sidilkover D. Optimum positive linear schemes for advection in two and three dimensions. *SIAM J. Numer. Anal.* 1992; **29**(6):1542–1568.
- Rogers DF. *An Introduction to NURBS: with Historical Perspectives*. Morgan Kaufman: San Mateo, CA, 2001.
- Rossiello G, DePalma P, Pascazio G and Napolitano M. Third-order-accurate fluctuation splitting schemes for unsteady hyperbolic problems. *J. Comput. Phys.* 2007; **222**(1):332–352.
- Rossiello G, De Palma P, Pascazio G and Napolitano M. Second-order-accurate explicit fluctuation splitting schemes for unsteady problems. *Comput. Fluids* 2009; **38**(7):1384–1393.
- Rosmanith JA, Bale DS and LeVeque RJ. A wave propagation algorithm for hyperbolic systems on curved manifolds. *J. Comput. Phys.* 2004; **199**:631–662.
- Ruuth SJ and Spiteri RJ. Two barriers on strong-stability-preserving time discretization methods. *J. Sci. Comput.* 2002; **17**(1-4):211–220.
- Sarmany D, Hubbard M and Ricchiuto M. Unconditionally stable space-time discontinuous residual distribution for shallow water flows. *J. Comput. Phys.* 2013; **253**:86–113.
- Schaffer HA and Madsen PA. Further enhancements of Boussinesq-type equations. *Coastal Eng.* 1995; **26**:1–14.
- Shakib F and Hughes TJR. A new finite element formulation for computational fluid dynamics: IX. Fourier analysis of space-time Galerkin/least-squares algorithms. *Comput. Methods Appl. Mech. Eng.* 1991; **87**(1):35–58.
- Shu C-W. High order weighted essentially nonoscillatory schemes for convection dominated problems. *SIAM Rev.* 2009; **51**(1):82–126.
- Shu C-W and Osher S. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *J. Comput. Phys.* 1988; **77**(2):439–471.
- Shu C-W and Osher S. Efficient implementation of essentially non-oscillatory shock-capturing schemes, II. *J. Comput. Phys.* 1989; **83**(1):32–78.
- Sidilkover D and Roe PL. Unification of some advection schemes in two dimensions. Technical Report 95-10. ICASE, 1995.
- Sørensen OR and Madsen PA. A new form of the Boussinesq equations with improved linear dispersion characteristics. Part 2: A slowing varying bathymetry. *Coastal Eng.* 1992; **18**:183–204.
- Sørensen L, Sørensen OR and Schaffer H. Boussinesq-type modelling using an unstructured finite element technique. *Coastal Eng.* 2004; **50**:181–198.
- Spekreijse SP. Multigrid solution of monotone second-order discretizations of hyperbolic conservation laws. *Math. Comput.* 1987; **49**:135–155.
- Struijs R. *A Multi-Dimensional Upwind Discretization Method for the Euler Equations on Unstructured Grids*. PhD thesis, University of Delft, Netherlands, 1994.
- Struijs R, Deconinck H, De Palma P, Roe PL and Powell KG. Progress on multidimensional upwind Euler solvers for unstructured grids. AIAA paper 91x971550, 1991.
- Sweby PK. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM J. Numer. Anal.* 1984; **21**:995–1011.

- Szepessy A. Convergence of a shock-capturing streamline diffusion finite element method for a scalar conservation law in two space dimensions. *Math. Comput.* 1989; **53**:527–545.
- Tonelli M and Petti M. Simulation of wave breaking over complex bathymetries by a Boussinesq model. *J. Hydraul. Res.* 2011; **49**:473–486.
- Tonelli M and Petti M. Hybrid finite volume–finite difference scheme for 2DH improved Boussinesq equations. *Coastal Eng.* 2009; **56**:609–620.
- Toro EF. *Riemann Solvers and Numerical Methods for Fluid Dynamics*. Springer, 1997.
- van Leer B. Toward the ultimate conservative difference scheme V. A second-order sequel to Godunov’s method. *J. Comput. Phys.* 1979; **32**(1):101–136.
- Vernotte P. Les paradoxes de la theorie continue de l’équation de la chaleur. *C.R. Acad. Sci.* 1958; **246**:3154–3155.
- Vilar F, Maire PH and Abgrall R. A discontinuous Galerkin discretization for solving the two-dimensional gas dynamics equations written under total Lagrangian formulation on general unstructured grids. *J. Comput. Phys.* 2014; **276**(1):188–234.
- von Mises R. Unabridged republication of the work first published by Academic Press Inc., Dover 1958.
- Vymazal M. *Very high order residual distribution schemes via variable distribution coefficients*. PhD thesis, Université Libre de Bruxelles and von Karman Institute for Fluid Dynamics, 2016, under review.
- Vymazal M, Koloszar L, D’Angelo S, Villedieu N, Ricchiuto M and Deconinck H. High-order residual distribution and error estimation for steady and unsteady compressible flow. In *IDIHOM: Industrialization of High-Order Methods – A Top-Down Approach*, Notes on Numerical Fluid Mechanics and Multidisciplinary Design, vol. 128, Kroll N, Hirsch C, Bassi F, Johnston C and Hillewaert K (eds). Springer International Publishing, 2015; 381–395.
- Walkley MA and Berzins M. A finite element method for the two-dimensional extended Boussinesq equations. *Int. J. Numer. Methods Fluids* 2002; **39**:865–885.
- Warburton T. An explicit construction of interpolation nodes on the simplex. *J. Eng. Math.* 2006; **56**(3):247–262.
- Wei G and Kirby JT. A time-dependent numerical code for extended Boussinesq equations. *J. Waterw. Port Coastal Ocean Eng.* 1995; **120**:251–261.
- Whalin RW. The limit of applicability of linear wave refraction theory in a convergence zone. Research Report H-71-3. USACE, Waterways Experiment Station: Vicksburg, MS, 1971.
- Xu Y. Gauss-Lobatto integration on the triangle. *SIAM J. Numer. Anal.* 2011; **49**:541–548.
- Zhang Q and Shu CW. Error estimates to smooth solutions of Runge-Kutta discontinuous Galerkin methods for scalar conservation laws. *SIAM J. Numer. Anal.* 2004; **42**(2):641–666.
- Zhang Q and Shu CW. Stability analysis and a priori error estimates for the third order explicit Runge-Kutta discontinuous Galerkin method for scalar conservation laws. *SIAM J. Numer. Anal.* 2010; **48**(3):1038–1063.
- Zhu J, Qiu J, Shu C and Dumbser M. Runge-Kutta discontinuous Galerkin method using WENO limiters II: unstructured meshes. *J. Comput. Phys.* 2008; **227**(9):4330–4435.
- Zienkiewicz OC and Zhu JZ. A simple error estimator and adaptive procedure for practical engineering analysis. *Int. J. Numer. Methods Eng.* 1987; **24**(2):337–357.