Mathematical Models and Methods in Applied Sciences
Vol. 14, No. 8 (2004) 1235–1260

World Scientific
www.worldscientific.com

A CELL-CENTERED SECOND-ORDER ACCURATE FINITE VOLUME METHOD FOR CONVECTION–DIFFUSION PROBLEMS ON UNSTRUCTURED MESHES

ENRICO BERTOLAZZI

Dipartimento di Ingegneria Meccanica e Strutturale,
Università di Trento, via Mesiano 77, 38100 Trento, Italy
enrico.bertolazzi@ing.unitn.it

GIANMARCO MANZINI

Istituto di Matematica Applicata e Tecnologie Informatiche,
Consiglio Nazionale delle Ricerche, via Ferrata 1, 27100 Pavia, Italy

Received 28 February 2003
Revised 5 April 2004
Communicated by N. Bellomo

A MUSCL-like cell-centered finite volume method is proposed to approximate the solution of multi-dimensional steady advection–diffusion equations. The second-order accuracy is provided by an appropriate definition of the diffusive and advective numerical fluxes. The method is based on a least squares reconstruction of the vertex values from cell averages. The slope limiter, which is required to prevent the formation and growth of spurious numerical oscillations, is designed to guarantee that the discrete solution of the nonlinear scheme exists. Several theoretical issues regarding the solvability of the resulting discrete problems are thoroughly discussed. Finally, numerical experiments that validate the effectiveness of the approach are presented.

Keywords: Finite volumes; convection diffusion; convergence rate; unstructured meshes; PDE; numerical solvers.

AMS Subject Classification: 65C20, 65N12, 65N22, 76R50, 45L05

1. Introduction

The major difficulty with the numerical approximation of the second-order steady advection–diffusion equation in divergence form is essentially related to the presence of the convection term. It is well known that *central* finite differences lead to schemes that formally have the second order of accuracy but are only stable if the *mesh size* h is rather small (a precise definition of this parameter is given in the following section). First-order accurate upwinding makes it possible to design schemes that are unconditionally stable, i.e. stable not only when h is small, but can also be used in both convection-dominated and diffusive limits. However, this technique introduces a large and often unacceptable numerical diffusion. A special treatment

of the convective term must thus be devised to achieve the goal of constructing approximations that have at least the second-order accuracy and satisfy some form of numerical stability to control spurious oscillations. A number of techniques to do this have been proposed in the literature. A detailed review of the numerical approximations of convection–diffusion models is found in Morton.[19]

Researchers have devoted particular attention to the use of the unstructured meshes of symplectic control volumes (triangles in 2D and tetrahedra in 3D) because they deal well with complex geometries. A systematic treatment of finite difference schemes on triangular meshes is discussed by Heinrich.[15] Cell-centered finite difference approximations were developed for both triangular and quadrilateral meshes, and local refinements by Vassilievski, Petrova, and Lazarov[24] and Lazarov, Makarov and Weinelt.[17] All of the second-order accurate upwind discretizations investigated in Ref. 24 show an unconditional numerical stability that is achieved by satisfying a discrete maximum principle. The error estimates were derived in a mesh-dependent H^1-like norm, and included $\mathcal{O}(h^2)$ super-convergence rates for uniform triangulations. Cell-centered discretizations on tensor-product non-uniform meshes are presented and theoretically analyzed by Weiser and Wheeler,[25] who also derived super-convergence error estimates in a similar norm. Two different approaches, which result in approximations closely related to finite elements, are the *box method* (see Bank and Rose[3] and Hackbusch[12]), and the *finite volume element method* (see Baliga and Patankar,[2] Cai, Mandel and McCormick,[7] Cai,[6] McCormick[18]). These methods were analyzed in the finite element framework, and $\mathcal{O}(h^2)$ super-convergence estimates on uniform meshes were also demonstrated.

Finally, the finite volume scheme analyzed by Coudière, Vila, and Villedieu,[9] and by Coudière and Villedieu[10] is surely worth mentioning. These references are the most pertinent ones to the present work. In the framework of the unstructured finite volume methods, these authors combine a first-order upwind method for the advection term with a central difference discretization for the diffusion term, i.e. the *diamond scheme*. The *a priori* error estimate for the approximation provided by the method is optimal in a discrete H^1-like norm and an $\mathcal{O}(h)$ convergence rate was theoretically demonstrated in Ref. 9. However, the upwinding of the advective flux that was introduced to enforce the unconditional stability of the method adds substantial numerical diffusion to the physical problem. This may cause a significant degradation in the quality of the approximation of cell averages. For this reason, the convergence rate is expected to be first-order.

In this work, we investigate the possibility of increasing the first-order rate up to the second-order rate by carefully designing a non-oscillatory piecewise linear interpolation from approximate cell averages. This kind of approach has been extensively investigated in the case of shock-capturing finite difference schemes on Cartesian meshes for conservation laws and hyperbolic systems. We mention, for example, the MUSCL scheme (see Harten[13]), the ENO scheme (see Harten *et al.*,[14] Shu and Osher[23]), and the WENO scheme (see Shu[22]). As we pointed out in Ref. 5, the cell-wise polynomial reconstruction introduces a numerical anti-diffusion that

balances the first-order upwind diffusion. If the numerical anti-diffusion overcomes the first-order upwind diffusion, a nonlinear instability increases and spurious numerical effects, such as overshoots and undershoots in strong gradient solution regions, appear. The standard approach to dealing with this phenomenon consists of introducing nonlinear feedback using a special constraint, i.e. the slope limiter. In fact, the introduction of a suitable limiter in the scheme design makes it possible to locally monotonize the reconstructed solution and avoid the formation of spurious numerical oscillations when approximating strong gradient solutions.

First, we will outline the key steps in the formulation of this finite volume method. An approximation of the vertex values is defined using the approximate cell averages and the linear least squares algorithm. The approximate vertex values are then combined with the cell averages to formulate cell-centered and face-centered gradients. The cell-centered gradients, which are used in the formulation of the numerical upstream flux of the advective term, are limited following the procedures proposed by Wierse[26] and Hubbard.[16] The reconstruction step and introduction of solution feedback using a multi-dimensional slope limiter imply that this scheme is inherently nonlinear, even if the continuous problem is linear. This nonlinearity is an essential feature of MUSCL-like schemes, and the resulting nonlinear algebraic problem requires an iterative solution procedure, i.e. a fixed-point scheme. On the other hand, the face-centered gradients are used to discretize the numerical diffusive flux using the diamond scheme.[9]

The finite volume scheme created by coupling the diamond scheme and nonlinear reconstruction is conservative. In addition, the discrete problem that results from the application of this scheme to the steady advection–diffusion equation has important theoretical properties. A thorough inspection of the algebraic structure of the discrete operator and the careful design of the limiting strategy allow us to prove the existence of at least one fixed-point solution. We can also prove that the distance between two possibly distinct fixed points must be at most of order $\mathcal{O}(h)$ when it is measured using the mesh-dependent H^1-norm introduced in Ref. 9.

The second-order convergence rate in the approximation of the cell averages is verified in a suitably defined mesh-dependent L^2-norm by numerical experiments. The first-order convergence rate in the solution gradient approximation is measured using the H^1-norm.

The method and theoretical results regarding solvability are formulated and proved in $\mathbb{R}^d$ for $d \geq 2$. Though this issue may seem rather unusual in fluid dynamics applications where model problems are normally defined in $\mathbb{R}^d$ for either $d = 2$ or $d = 3$, we must emphasize its importance. Higher dimensional model problems can be found in different application domains, for example the computational finance one, where d-dimensional Black–Scholes models (which are still based on convection–diffusion type equations) are considered for very large d. For an introduction to this topic see, for example, the book by Wilmott, Dewynne and Howison.[27] Finally, it is worth pointing out that the extension to more general unstructured meshes and multi-dimensional structured ones is straightforward.

The outline of the paper is as follows. The mathematical problem is introduced in Sec. 2. The general framework of this finite volume method is set up in Sec. 3. The least squares-based reconstruction for the approximate vertex values and the formulation of the discrete cell-centered and face-centered gradients are discussed in Sec. 4. In Sec. 5 the finite volume scheme is reformulated as an operator equation that is more suitable for theoretical analysis. The solvability of the discrete problem that results from applying this discretization method to the steady convection–diffusion problem is investigated in Sec. 6. The numerical capabilities of this method are illustrated in Sec. 7. The final conclusions are in Sec. 8.

2. The Advection–Diffusion Equation

Let us consider the convection–diffusion boundary value problem:

find a function $\mathrm{u}(\mathbf{x})$ *satisfying*

$$\begin{aligned} \operatorname{div}(\mathbf{v}\mathrm{u} - \nu\nabla\mathrm{u}) &= \mathrm{s}\,, \quad \text{on } \Omega\,, \\ \mathrm{u} &= \mathrm{g}\,, \quad \text{on } \Gamma\,, \end{aligned} \tag{2.1}$$

where $\Omega \subset \mathbb{R}^d$, $d \geq 2$, is a closed and bounded polytope domain, $\Gamma = \partial\Omega$ its boundary, and $\mathbf{n}$ the outward-oriented vector orthogonal to $\partial\Omega$ a.e. This model equation describes the advective transport of the scalar quantity $\mathrm{u}(\mathbf{x})$ by the velocity field $\mathbf{v}(\mathbf{x})$ and the diffusion process driven by the scalar viscosity field $\nu(\mathbf{x})$. A forcing source term $\mathrm{s}(\mathbf{x})$ may also be present.

Let the field $\mathbf{v}(\mathbf{x})$ and functions $\nu(\mathbf{x})$, $\mathrm{s}(\mathbf{x})$ and $\mathrm{g}(\mathbf{x})$ satisfy the following conditions:

$$\begin{aligned} &\text{(i)} && \nu \geq \beta_0 > 0\,, && \nu \in C^1(\bar{\Omega})\,; \\ &\text{(ii)} && \operatorname{div}\mathbf{v} \geq \beta_1 \geq 0\,, && \mathbf{v} \in C^1(\bar{\Omega})^d\,; \\ &\text{(iii)} && \mathrm{s} \in L^2(\Omega)\,; \\ &\text{(iv)} && \mathrm{g} \in H^{1/2}(\Gamma) \cap C(\Gamma)\,; \end{aligned} \tag{2.2}$$

for suitable real constants β_0 and β_1. In the following we also use the constant

$$\beta_2 = \|\mathbf{v}\|_{L^\infty(\Omega)}\,. \tag{2.3}$$

Under conditions (2.2) a weak reformulation of problem (2.1) is possible in terms of an H^1-coercive bilinear form. In view of the Lax–Milgram lemma argument the weak problem has a unique solution in $H^1(\Omega)$ (see Agmon[1]).

3. The Finite Volume Formulation

In this section, we introduce the finite volume approximation framework and explain the formalism and notations adopted in this paper.

Let us suppose that Ω be covered by a collection of non-empty and non-overlapping multi-dimensional simplices T_i, namely *the mesh*, denoted by $\mathcal{T}_h$. The

mesh $\mathcal{T}_h$ is such that $\bar{\Omega} = \bigcup_{\mathrm{T}_i \in \mathcal{T}_h} \mathrm{T}_i$. These simplices are also called *control volumes* or *cells*, and related quantities are labeled by Roman letters:

— $\mathrm{T}_i \in \mathcal{T}_h$ is the ith simplex of the mesh;
— $|\mathrm{T}_i|$ is the d-dimensional measure of T_i;
— $\mathbf{x}_i$ is the centroid (barycenter) of T_i.

The mesh vertex set is denoted by $\mathcal{V}_h$ and is the union of $\mathcal{V}_h^{\mathrm{int}}$ and $\mathcal{V}_h^{\mathrm{bnd}}$, the sets of internal and boundary vertices. Vertex-related quantities are labeled by Greek letters:

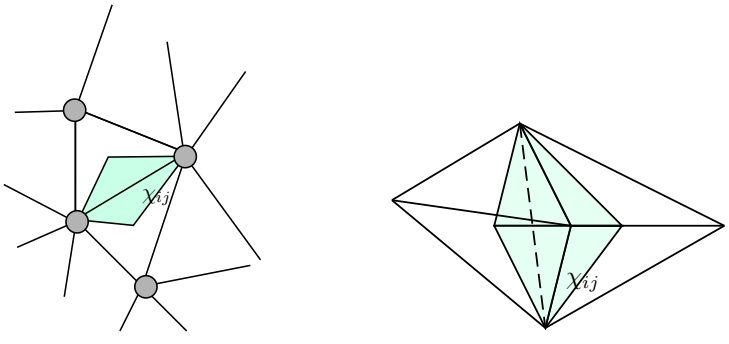
— $\mathrm{v}_\alpha \in \mathcal{V}_h$ is the αth vertex of the mesh;
— $\mathbf{x}_\alpha$ is the position vector of v_α.

The mesh face set is denoted by $\mathcal{F}_h$ and is the union of $\mathcal{F}_h^{\mathrm{int}}$ and $\mathcal{F}_h^{\mathrm{bnd}}$, the sets of internal and boundary faces. Face related quantities are labeled by a couple of Roman indices:

— $\mathrm{f}_{ij} \in \mathcal{F}_h^{\mathrm{int}}$ is the internal face common to the simplices T_i and T_j;
— $\mathrm{f}_{ij} \in \mathcal{F}_h^{\mathrm{bnd}}$ is a boundary face of the simplex T_i, i.e. $\mathrm{f}_{ij} \subset \partial\mathrm{T}_i \cap \partial\Omega$; the second index refers to a suitable numbering system (a sort of fictitious *ghost cell*);
— $|\mathrm{f}_{ij}|$ is the $(d-1)$-dimensional measure of the face f_{ij};
— $\mathbf{x}_{ij} \in \mathrm{f}_{ij}$ is the centroid (barycenter) of f_{ij};
— $\mathbf{n}_{ij}$ is the unit normal to the face f_{ij} pointing out of the cell T_i;
— $\tilde{\mathbf{x}}_{ij} \in \mathrm{f}_{ij}$ is the orthogonal projection of $\mathbf{x}_i$ onto the hyperplane containing the face f_{ij};
— χ_{ij} is the *diamond cell* defined by

$$\chi_{ij} = \begin{cases} \mathrm{hull}\,\{\mathbf{x}_i, \mathbf{x}_j, \mathrm{f}_{ij}\}\,, & \text{for } \mathrm{f}_{ij} \in \mathcal{F}_h^{\mathrm{int}}\,, \\ \mathrm{hull}\,\{\mathbf{x}_i, \mathrm{f}_{ij}\}\,, & \text{for } \mathrm{f}_{ij} \in \mathcal{F}_h^{\mathrm{bnd}}\,. \end{cases}$$

Figure 1 depicts χ_{ij} for an internal face in the 2D and 3D cases.



(a) (b)

Fig. 1. The diamond cell χ_{ij}: (a) 2D case, (b) 3D case.

The following geometric quantities are also used in the paper:

— $h_{ij} = \mathbf{n}_{ij} \cdot (\tilde{\mathbf{x}}_{ij} - \mathbf{x}_i)$ is the distance between $\tilde{\mathbf{x}}_{ij}$ and $\mathbf{x}_i$;
— $H_{ij} = \mathbf{n}_{ij} \cdot (\mathbf{x}_j - \mathbf{x}_i) = h_{ij} + h_{ji}$ is the diamond-cell "characteristic height" for the internal face $\mathrm{f}_{ij} \in \mathcal{F}_h^{\mathrm{int}}$.

If we extend the definition of H_{ij} to the case of the boundary face $\mathrm{f}_{ij} \in \mathcal{F}_h^{\mathrm{bnd}}$ by setting $H_{ij} = h_{ij}$, the d-dimensional measure of the diamond cell can be written as $|\chi_{ij}| = H_{ij}|\mathrm{f}_{ij}|/d$.

The stencils of the control volumes and mesh vertices that summations are usually taken over are represented by sets of cell and vertex indices, respectively denoted by the suitably indexed Greek letters σ and ν:

— σ_i is the index set of the first neighbor cells to T_i;
— σ_α is the index set of the first neighbor cells to v_α;
— ν_i is the index set of the vertices forming T_i;
— ν_α is the index set of the vertices connected to v_α;
— ν_{ij} is the index set of the vertices forming the face f_{ij}; coherently with the definition of internal and boundary vertices, we also introduce the sets

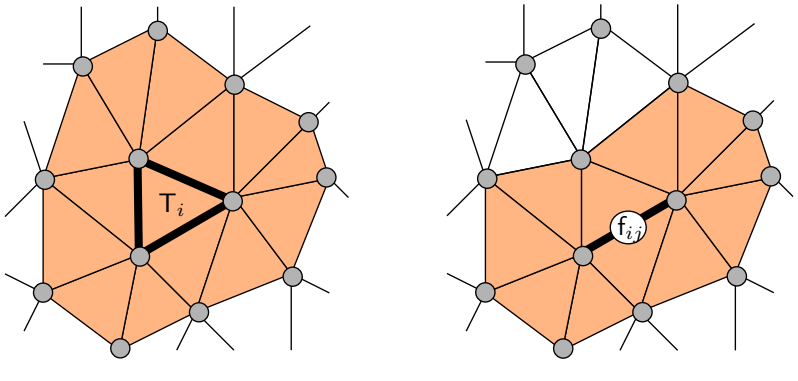
$$\nu_{ij}^{\mathrm{int}} = \{\alpha \in \nu_{ij} | \mathrm{v}_\alpha \in \mathcal{V}_h^{\mathrm{int}}\} \quad \text{and} \quad \nu_{ij}^{\mathrm{bnd}} = \{\alpha \in \nu_{ij} | \mathrm{v}_\alpha \in \mathcal{V}_h^{\mathrm{bnd}}\}\,.$$

The following shortcuts

$$\sigma\nu_i = \bigcup_{\alpha \in \nu_i} \sigma_\alpha\,, \quad \sigma\nu_{ij} = \bigcup_{\alpha \in \nu_{ij}} \sigma_\alpha\,,$$

are also used in the paper and are the stencils depicted in Fig. 2.

In the rest of the paper, summations will also be taken over σ_i', which is the set of indices of fictitious cells used for labeling the faces of T_i on the boundary Γ. The union $\sigma_i \cup \sigma_i'$ represents the set of indices labeling all the faces of the cell boundary



(a) (b)

Fig. 2. The complex stencil in the 2D case: (a) the stencil $\sigma\nu_i$, (b) the stencil $\sigma\nu_{ij}$.

$\partial \mathrm{T}_i$. We denote the absolute value of the scalar x by $|x|$ and the Euclidean norm of the vector $\mathbf{x}$ by $|\mathbf{x}|$.

Assumption 3.1. (Mesh regularity) We take into account families of triangulations $\{\mathcal{T}_h\}$ that are regular in the following sense. Let ρ be the minimum radius of the spheres inscribed in $\mathrm{T}_i \in \mathcal{T}_h$, and h the supremum of the mesh control volume diameters. We assume that

$$\begin{aligned} &\text{(a)} \quad (h/\rho)^d \leq C_{\mathrm{reg}}\,, \quad \text{and } C_{\mathrm{reg}} \text{ is independent of } h\,; \\ &\text{(b)} \quad \tilde{\mathbf{x}}_{ij} \in \mathrm{f}_{ij}\,, \qquad\qquad \text{for } \mathrm{f}_{ij} \in \mathcal{F}_h\,. \end{aligned}$$

Assumption 3.1(a) is rather standard in finite element analysis (see Ciarlet[8] and Raviart and Thomas[20]). Assumption 3.1(b) is always satisfied by the weakly acute triangulations[8] or by more general triangulations under suitable angle conditions. In view of Assumption 3.1(b), $\tilde{\mathbf{x}}_{ij} \in \mathcal{F}_h$ can be expressed as a convex combination of the face vertices v_α, $\alpha \in \nu_{ij}$. From Assumption 3.1, the following inequalities are immediately obtained:

$$\begin{aligned} \operatorname{card}\{\sigma_\alpha\} &\leq C_{\mathrm{reg}}\,, \\ |\mathrm{T}_i|\,|\mathrm{T}_j|^{-1} &\leq C_{\mathrm{reg}}\,, \\ |h_{ij}|\,|h_{ji}|^{-1} &\leq C_{\mathrm{reg}}\,, \\ h h_{ij}^{-1} &\leq C\,, \end{aligned} \tag{3.1}$$

where the constant C is independent of h and depends on d.

The design of our finite volume approximation starts as usual by reformulating the governing equations (2.1) on the generic cell T_i in integral form. By applying the Gauss–Green divergence theorem, we have the set of balance equations

$$\int_{\partial \mathrm{T}_i} (\mathbf{v}\mathrm{u} - \nu \nabla \mathrm{u}) \cdot \mathbf{n}\, dS = \int_{\mathrm{T}_i} \mathrm{s}\, dV\,, \quad \text{for } \mathrm{T}_i \in \mathcal{T}_h\,. \tag{3.2}$$

Let $\mathbf{u}$ be the vector of (unknown) cell averages whose ith component, namely u_i, is the approximation of the cell averaged solution within the cell T_i.

The finite volume discretization mimics equations (3.2) by introducing a set of *discrete* balance equations that correlate each u_i to the balance of the numerical fluxes across its control volume boundary $\partial \mathrm{T}_i$. Thus, the finite volume scheme reads as

$$\frac{1}{|\mathrm{T}_i|} \sum_{j \in \sigma_i} |\mathrm{f}_{ij}|\, (\mathcal{F}_{ij}(\mathbf{u}) + \mathcal{G}_{ij}(\mathbf{u})) + \frac{1}{|\mathrm{T}_i|} \sum_{j \in \sigma_i'} |\mathrm{f}_{ij}|\, \mathcal{B}_{ij}(\mathbf{u}) = \mathrm{s}_i\,, \quad \text{for } \mathrm{T}_i \in \mathcal{T}_h\,. \tag{3.3}$$

The first summation term in Eq. (3.3) takes into account the numerical advective and diffusive flux integrals, respectively denoted by $\mathcal{F}_{ij}(\mathbf{u})$ and $\mathcal{G}_{ij}(\mathbf{u})$, on the control volume interfaces f_{ij}. The second summation term takes into account the contribution to the numerical flux balance from the boundary conditions. Finally, s_i is the cell-average of the source term in (2.1).

The functional dependence on $\mathbf{u}$ of $\mathcal{F}_{ij}(\mathbf{u})$, $\mathcal{G}_{ij}(\mathbf{u})$ and $\mathcal{B}_{ij}(\mathbf{u})$ will be carefully developed in the next sections to define properly the finite volume scheme.

4. The Least Squares-Based Reconstruction

A second-order finite volume discretization requires an $\mathcal{O}(h^2)$ estimate of the solution for the numerical advective flux and an $\mathcal{O}(h)$ estimate of the solution gradient for the numerical diffusive flux at any control volume interface. These estimates are based on a linear least squares reconstruction of the vertex-centered values from the cell-averaged ones. In fact, by using the approximations of cell averages and the reconstructed values at the control volume vertices, we can define the cell-centered *limited* gradient $\boldsymbol{\mathcal{G}}_i(\mathbf{u})$ and the $d+1$ face-centered gradients $\boldsymbol{\mathcal{G}}_{ij}(\mathbf{u})$, for $j \in \sigma_i \cup \sigma_i'$. Then, the limited gradient $\boldsymbol{\mathcal{G}}_i(\mathbf{u})$ allows us to define the linearly reconstructed approximation

$$u_r(\mathbf{x}) = u_i + \boldsymbol{\mathcal{G}}_i(\mathbf{u}) \cdot (\mathbf{x} - \mathbf{x}_i)\,, \quad \mathbf{x} \in \mathrm{T}_i\,, \tag{4.1}$$

and this latter one is used to define the numerical advective flux operator. On the other hand, the $d+1$ face-centered gradients $\boldsymbol{\mathcal{G}}_{ij}(\mathbf{u})$ are used to define the numerical diffusive flux operator.

The finite volume logic of the reconstruction process is summarized as follows:

$$\begin{aligned}
\text{cell averages} &\Rightarrow \text{vertex point values}\\
\text{vertex point values} &\Rightarrow \text{cell-centered gradient}\\
&\Rightarrow \text{numerical advective flux}\\
\left.\begin{array}{r}\text{cell averages}\\ \text{vertex point values}\end{array}\right\} &\Rightarrow \text{face-centered gradients}\\
&\Rightarrow \text{numerical diffusive flux}\,.
\end{aligned}$$

In the rest of the section we describe each reconstruction case in detail.

4.1. *Reconstruction of vertex values*

The reconstructed value u_α at the internal or boundary vertex v_α is an affine linear combination of the approximate control volume averages of the surrounding cells,

$$u_\alpha = \begin{cases} \displaystyle\sum_{k \in \sigma_\alpha} \mathrm{W}_k^\alpha u_k & \text{for } \mathrm{v}_\alpha \in \mathcal{V}_h^{\mathrm{int}}\,,\\ \mathrm{g}(\mathbf{x}_\alpha) & \text{for } \mathrm{v}_\alpha \in \mathcal{V}_h^{\mathrm{bnd}}\,. \end{cases} \tag{4.2}$$

The weights W_k^α provided by the linear least squares method are such that[9]

$$\sum_{k \in \sigma_\alpha} \mathrm{W}_k^\alpha = 1 \quad \text{and} \quad \sum_{k \in \sigma_\alpha} \mathrm{W}_k^\alpha (\mathbf{x}_k - \mathbf{x}_\alpha) = 0\,.$$

If the mesh is regular, these weights are also uniformly bounded[9]; i.e.

$$|\mathrm{W}_k^\alpha| \le C_{\text{weight}}\,,$$

where C_{weight} is a real positive constant independent of h.

4.2. *Reconstruction of cell-centered gradients*

The cell-centered gradient $\boldsymbol{\mathcal{G}}_i(\mathbf{u})$ is reconstructed by using the Gauss–Green theorem applied to the control volume T_i; $\boldsymbol{\mathcal{G}}_i(\mathbf{u})$ is then limited by multiplying by a limiter factor $l_i(\mathbf{u})$. This factor is estimated by a variant of the central limiter technique with the projection step proposed by Wierse[26] and Hubbard.[16] Further details about this limiting strategy are discussed in Sec. 6.

Let us introduce the symbols $|\mathrm{f}_\alpha|$, which is the measure of the face f_α opposite to the vertex v_α and $\mathbf{n}_\alpha$, which is the outward-oriented unit vector orthogonal to f_α. The multi-dimensional cell-centered gradient formula reads as

$$\boldsymbol{\mathcal{G}}_i(\mathbf{u}) = -\frac{l_i(\mathbf{u})}{d|\mathrm{T}_i|}\sum_{\alpha\in\nu_i} u_\alpha|\mathrm{f}_\alpha|\mathbf{n}_\alpha\,, \tag{4.3}$$

where $l_i(\mathbf{u})$ is the central limiter factor.

4.3. *Reconstruction of face-centered gradients*

In this section, we present an alternative derivation of face gradients that is completely equivalent to the usual one, which directly exploits the Gauss–Green theorem applied to a piecewise constant gradient vector on diamond cells. The motivation is that in the general case for $d > 2$, the diamond-cell gradient formula is readily derived. The two definitions perfectly coincide when $d = 2$; see, for example, the formulation discussed by Coudière, Vila, Villedieu.[9]

Let us first define the normal component of the *one-side* face-centered gradient $\boldsymbol{\mathcal{G}}_{ij}(\mathbf{u})$ associated to the face f_{ij}. This value is reconstructed by using the Gauss–Green theorem applied to the control volume defined by the centroid of T_i and the vertices of the face f_{ij}.

After some algebraic manipulations, the formula needed to reconstruct face gradients becomes

$$\boldsymbol{\mathcal{G}}_{ij}(\mathbf{u}) = \frac{\tilde{u}_{ij} - u_i}{h_{ij}}\mathbf{n}_{ij} + \{\text{tangential terms}\}\,, \tag{4.4}$$

where

$$\tilde{u}_{ij} = \sum_{\alpha\in\nu_{ij}} \tilde{\lambda}_\alpha^{ij} u_\alpha\,, \quad \tilde{\mathbf{x}}_{ij} = \sum_{\alpha\in\nu_{ij}} \tilde{\lambda}_\alpha^{ij}\mathbf{x}_\alpha\,. \tag{4.5}$$

In view of Assumption 3.1(b), which implies that $\tilde{\mathbf{x}}_{ij} \in \mathrm{f}_{ij}$, the barycentric coordinates $\{\tilde{\lambda}_\alpha^{ij}\}$ are non-negative. Consequently, $\tilde{u}_{ij}$ is a *convex* linear interpolation of the vertex values u_α for $\alpha \in \nu_{ij}$.

In order to have a *conservative* definition of the numerical diffusive flux $\mathcal{G}_{ij}(\mathbf{u})$, one single definition of the face gradient is required. To this purpose, when f_{ij} is an internal face, we average the two one-sided numerical gradients $\boldsymbol{\mathcal{G}}_{ij}(\mathbf{u})$ and $\boldsymbol{\mathcal{G}}_{ji}(\mathbf{u})$, respectively defined within the cells T_i and T_j adjacent to the face f_{ij}. In other words,

$$\boldsymbol{\mathcal{G}}^{\diamond}_{ij}(\mathbf{u}) = W_{ij}\boldsymbol{\mathcal{G}}_{ij}(\mathbf{u}) + W_{ji}\boldsymbol{\mathcal{G}}_{ji}(\mathbf{u})\,, \tag{4.6}$$

where the non-negative *diamond scheme* weights are defined by

$$W_{ij} = 1 - W_{ji} = \frac{h_{ij}}{H_{ij}}\,. \tag{4.7}$$

The normal component of the face-centered gradient associated to the face f_{ij}, namely $\boldsymbol{\mathcal{G}}^{\diamond}_{ij}(\mathbf{u})\cdot\mathbf{n}_{ij}$, is given by inserting (4.4) into (4.6) and using the weights (4.7):

$$\boldsymbol{\mathcal{G}}^{\diamond}_{ij}(\mathbf{u})\cdot\mathbf{n}_{ij} = \frac{\tilde{u}_{ij} - \tilde{u}_{ji} + u_j - u_i}{H_{ij}}\,.$$

In view of (4.5) we reformulate the face-centered gradient as

$$\boldsymbol{\mathcal{G}}^{\diamond}_{ij}(\mathbf{u})\cdot\mathbf{n}_{ij} = \frac{1}{H_{ij}}(u_j - u_i) + \frac{1}{H_{ij}}\sum_{\alpha\in\nu_{ij}}(\tilde{\lambda}^{ij}_{\alpha} - \tilde{\lambda}^{ji}_{\alpha})u_{\alpha}\,. \tag{4.8}$$

Substituting the expression of the vertex reconstructed value u_α given by (4.2) in (4.8) yields

$$\boldsymbol{\mathcal{G}}^{\diamond}_{ij}(\mathbf{u})\cdot\mathbf{n}_{ij} = \frac{1}{H_{ij}}\sum_{k\in\sigma\nu_{ij}}\tilde{\omega}^{ij}_{k}u_k + \mathrm{g}^{\diamond}_{ij}\,, \tag{4.9}$$

where, in (4.9), the following summation rule is used first

$$\sum_{\alpha\in\nu^{\mathrm{int}}_{ij}}\sum_{k\in\sigma_\alpha} = \sum_{k\in\sigma\nu_{ij}}\sum_{\alpha\in\nu_k\cap\nu^{\mathrm{int}}_{ij}}\,,$$

and the averaging weights $\tilde{\omega}^{ij}_k$ and displacement term $\mathrm{g}^{\diamond}_{ij}$ are then given by:

$$\mathrm{g}^{\diamond}_{ij} = \frac{1}{H_{ij}}\sum_{\alpha\in\nu^{\mathrm{bnd}}_{ij}}(\tilde{\lambda}^{ij}_{\alpha} - \tilde{\lambda}^{ji}_{\alpha})\;\mathrm{g}(\mathbf{x}_\alpha)\,,$$

$$\tilde{\omega}^{ij}_{k} = \sum_{\alpha\in\nu_k\cap\nu^{\mathrm{int}}_{ij}}\mathrm{W}^{\alpha}_{k}(\tilde{\lambda}^{ij}_{\alpha} - \tilde{\lambda}^{ji}_{\alpha}) + \delta_{kj} - \delta_{ki}\,,$$

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j\,,\\ 0 & \text{otherwise}\,. \end{cases}$$

When f_{ij} is a boundary face, the face-centered gradient is set to the available one-sided face-centered gradient, i.e. $\boldsymbol{\mathcal{G}}^{\diamond}_{ij}(\mathbf{u}) = \boldsymbol{\mathcal{G}}_{ij}(\mathbf{u})$.

5. The Vector Formulation of the Finite Volume Method

In this section, we use the piecewise linear reconstruction and the cell and face gradients that have been introduced in Sec. 4 to derive the advective and diffusive numerical flux. Then, the finite volume method (3.3) is reformulated in a compact vector form.

5.1. *The numerical advective flux*

First, for each face $\mathrm{f}_{ij} \in \mathcal{F}_h$ let us introduce the *upwind velocities*, that are defined by the standard relations $v_{ij}^{\pm} = (v_{ij} \pm |v_{ij}|)/2$, where

$$v_{ij} = \frac{1}{|\mathrm{f}_{ij}|} \int_{\mathrm{f}_{ij}} \mathbf{v} \cdot \mathbf{n}_{ij}\, dS \tag{5.1}$$

and $\mathbf{n}_{ij}$ is the unit orthogonal vector to f_{ij}.

We use u_{ij} and u_{ji} to denote the approximate solution at the face center $\mathbf{x}_{ij} = \mathbf{x}_{ji}$ calculated using the linear reconstruction (4.1) within the adjacent cells T_i and T_j, respectively. Therefore,

$$u_{ij} = u_i + (\mathbf{x}_{ij} - \mathbf{x}_i) \cdot \boldsymbol{\mathcal{G}}_i(\mathbf{u}) \quad \text{and} \quad u_{ji} = u_j + (\mathbf{x}_{ji} - \mathbf{x}_j) \cdot \boldsymbol{\mathcal{G}}_j(\mathbf{u})\,. \tag{5.2}$$

By using (5.1) and (5.2), the average flux of the advective term on the *internal* face $\mathrm{f}_{ij} \in \mathcal{F}_h^{\mathrm{int}}$ is

$$\begin{aligned} \mathcal{F}_{ij}(\mathbf{u}) &= v_{ij}^{+} u_{ij} + v_{ij}^{-} u_{ji}\,, \\ &= \underbrace{v_{ij}^{+} u_i + v_{ij}^{-} u_j}_{\text{1st-order term}} + \underbrace{v_{ij}^{+} (\mathbf{x}_{ij} - \mathbf{x}_i) \cdot \boldsymbol{\mathcal{G}}_i(\mathbf{u}) + v_{ij}^{-} (\mathbf{x}_{ji} - \mathbf{x}_j) \cdot \boldsymbol{\mathcal{G}}_j(\mathbf{u})}_{\text{2nd-order correction}}\,, \end{aligned} \tag{5.3}$$

where we have underlined the first-order upwind term and the nonlinear second-order downwind correction determined by the reconstruction. Using (5.3), the integral balance for the upwind flux can be reformulated in the more compact matrix/vector form

$$\frac{1}{|\mathrm{T}_i|} \sum_{j \in \sigma_i} |\mathrm{f}_{ij}| \mathcal{F}_{ij}(\mathbf{u}) = (\mathbf{F}\mathbf{u} - \mathbf{r}^{\mathrm{int}}(\mathbf{u}))\Big|_i\,, \quad \text{for } \mathrm{T}_i \in \mathcal{T}_h\,,$$

by introducing the matrix $\mathbf{F} = \{\mathrm{F}_{ij}\}$, which takes into account the first-order upwind terms, and the nonlinear vector $\mathbf{r}^{\mathrm{int}}(\mathbf{u}) = \{\mathrm{r}_i^{\mathrm{int}}(\mathbf{u})\}$ which puts the second-order downwind corrections together. Since $v_{ij}^{-} = -v_{ji}^{+}$, the matrix $\mathbf{F}$ and the vector $\mathbf{r}^{\mathrm{int}}(\mathbf{u})$ are

$$\mathrm{F}_{ij} = \frac{1}{|\mathrm{T}_i|} \begin{cases} \displaystyle\sum_{k \in \sigma_i} |\mathrm{f}_{ik}|\, v_{ik}^{+}\,, & \text{if } i = j\,, \\ -|\mathrm{f}_{ij}|\, v_{ji}^{+}\,, & \text{for } j \in \sigma_i\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{5.4}$$

$$\mathrm{r}_i^{\mathrm{int}}(\mathbf{u}) = \frac{1}{|\mathrm{T}_i|} \sum_{j \in \sigma_i} |\mathrm{f}_{ij}|\, (v_{ij}^{+} (\mathbf{x}_i - \mathbf{x}_{ij}) \cdot \boldsymbol{\mathcal{G}}_i(\mathbf{u}) - v_{ji}^{+} (\mathbf{x}_j - \mathbf{x}_{ij}) \cdot \boldsymbol{\mathcal{G}}_j(\mathbf{u}))\,.$$

5.2. *The numerical diffusive flux*

The numerical *diffusive* flux is defined by reformulating the viscous normal derivative $\nu\,\mathbf{n}\cdot\nabla \mathrm{u}$ in (3.2) as $\nu_{ij}\boldsymbol{\mathcal{G}}^{\diamond}_{ij}(\mathbf{u})\cdot\mathbf{n}_{ij}$, where $\boldsymbol{\mathcal{G}}^{\diamond}_{ij}(\mathbf{u})\cdot\mathbf{n}_{ij}$ is the normal component of the reconstructed gradient on the face f_{ij} and

$$\nu_{ij} = \frac{1}{|\mathrm{f}_{ij}|}\int_{\mathrm{f}_{ij}} \nu(\mathbf{x})\,dS\,.$$

Its contribution to the flux balance of the cell T_i, namely the term $\mathcal{G}_{ij}(\mathbf{u})$ in (3.3), is evaluated by applying the second-order mid-point quadrature rule to the face integral of the numerical *diffusive* flux. This yields $\mathcal{G}_{ij}(\mathbf{u}) = -\nu_{ij}\boldsymbol{\mathcal{G}}^{\diamond}_{ij}(\mathbf{u})\cdot\mathbf{n}_{ij}$, and by applying the summation rule

$$\sum_{j\in\sigma_i}\sum_{k\in\sigma\nu_{ij}} = \sum_{k\in\sigma\nu_i}\sum_{j\in\sigma_i\cap\sigma\nu_k}\,,$$

in (4.9) we obtain

$$\sum_{j\in\sigma_i}|\mathrm{f}_{ij}|\mathcal{G}_{ij}(\mathbf{u}) = -\sum_{k\in\sigma\nu_i}u_k\sum_{j\in\sigma_i\cap\sigma\nu_k}\frac{\nu_{ij}|\mathrm{f}_{ij}|}{H_{ij}}\tilde{\omega}^{ij}_k - \sum_{j\in\sigma_i}\nu_{ij}|\mathrm{f}_{ij}|\mathrm{g}^{\diamond}_{ij}\,. \tag{5.5}$$

The discrete flux balance (5.5) takes the vector form

$$\frac{1}{|\mathrm{T}_i|}\sum_{j\in\sigma_i}|\mathrm{f}_{ij}|\mathcal{G}_{ij}(\mathbf{u}) = (\mathbf{G}\mathbf{u}-\mathbf{g})\Big|_i\,,\quad \text{for } \mathrm{T}_i\in\mathcal{T}_h\,,$$

where the *diffusion matrix* $\mathbf{G}=\{\mathrm{G}_{ij}\}$ and the *diffusion vector* $\mathbf{g}=\{\mathrm{g}_i\}$ are

$$\begin{aligned}
\mathrm{G}_{ij} &= \frac{1}{|\mathrm{T}_i|}\begin{cases} -\displaystyle\sum_{l\in\sigma_i\cap\sigma\nu_j}\nu_{il}|\mathrm{f}_{il}|\tilde{\omega}^{il}_j/H_{il}\,, & \text{for } j\in\sigma\nu_i\,,\\ 0 & \text{otherwise}\,,\end{cases}\\
\mathrm{g}_i &= \frac{1}{|\mathrm{T}_i|}\sum_{j\in\sigma_i}\nu_{ij}|\mathrm{f}_{ij}|\mathrm{g}^{\diamond}_{ij}\,.
\end{aligned} \tag{5.6}$$

5.3. *The boundary data and source term contribution*

The boundary flux $\mathcal{B}_{ij}(\mathbf{u})$ which appears in the second summation term of (3.3) is

$$\mathcal{B}_{ij}(\mathbf{u}) = \begin{cases} v^{+}_{ij}u_{ij} + v^{-}_{ij}\mathrm{g}(\mathbf{x}_{ij}) - \nu_{ij}(\mathrm{g}(\tilde{\mathbf{x}}_{ij}) - u_i)/h_{ij}\,, & \text{for } j\in\sigma'_i\,,\\ 0\,, & \text{otherwise}\,.\end{cases} \tag{5.7}$$

By taking into account (5.2) and (5.7), the boundary face contribution to the discrete flux balance of T_i is

$$\frac{1}{|\mathrm{T}_i|}\sum_{j\in\sigma'_i}|\mathrm{f}_{ij}|\mathcal{B}_{ij}(\mathbf{u}) = (\mathbf{B}\mathbf{u}-\mathbf{b}-\mathbf{r}^{\mathrm{bnd}}(\mathbf{u}))\Big|_i\,,\quad \text{for } \mathrm{T}_i\in\mathcal{T}_h\,,$$

where

$$\mathrm{B}_{ij} = \frac{1}{|\mathrm{T}_i|} \begin{cases} \displaystyle\sum_{l \in \sigma_i'} |\mathrm{f}_{il}|(v_{il}^+ + \nu_{il}/h_{il})\,, & \text{for } j = i\,, \\ 0\,, & \text{otherwise}\,, \end{cases}$$

$$\mathrm{b}_i = \frac{1}{|\mathrm{T}_i|} \sum_{j \in \sigma_i'} |\mathrm{f}_{ij}| \left[\frac{\nu_{ij}}{h_{ij}} \mathrm{g}(\tilde{\mathbf{x}}_{ij}) - v_{ij}^- \mathrm{g}(\mathbf{x}_{ij}) \right], \tag{5.8}$$

$$\mathbf{r}_i^{\mathrm{bnd}}(\mathbf{u}) = \frac{1}{|\mathrm{T}_i|} \sum_{j \in \sigma_i'} |\mathrm{f}_{ij}| v_{ij}^+ (\mathbf{x}_i - \mathbf{x}_{ij}) \cdot \boldsymbol{\mathcal{G}}_i(\mathbf{u})\,.$$

The discretization of the source term s in (2.1) is taken into account by $\mathbf{s} = \{\mathrm{s}_i\}$, where

$$\mathrm{s}_i = \frac{1}{|\mathrm{T}_i|} \int_{\mathrm{T}_i} \mathrm{s}(\mathbf{x})\, dV\,. \tag{5.9}$$

5.4. *The vector formulation*

By using the definitions introduced so far in Eqs. (5.4), (5.6), (5.8), (5.9), and setting $\mathbf{r}(\mathbf{w}) = \mathbf{r}^{\mathrm{int}}(\mathbf{w}) + \mathbf{r}^{\mathrm{bnd}}(\mathbf{w})$, the finite volume method, which has been introduced as a flux balance in Eq. (3.3), takes the vector form:

find $\mathbf{w} \in \mathbb{R}^{\mathrm{card}\{\mathcal{T}_h\}}$ such that

$$(\mathbf{F} + \mathbf{G} + \mathbf{B})\mathbf{w} = \mathbf{b} + \mathbf{g} + \mathbf{s} + \mathbf{r}(\mathbf{w})\,. \tag{5.10}$$

6. Solvability of the Finite Volume Approximation

In this section, we show that the matrix $\mathbf{F} + \mathbf{G} + \mathbf{B}$ in (5.10) is non-singular by construction. However, as the discrete problem is nonlinear due to the dependence of $\mathbf{r}(\mathbf{w})$ on $\mathbf{w}$, its solvability has to be carefully investigated. The main results are stated in Theorems 6.7 and 6.8 and claim that a solution to (5.10) always exists and is asymptotically unique in the approximation process for $h \to 0$.

Let us introduce the set of piecewise constant functions T_h defined on a given triangulation $\mathcal{T}_h$; i.e.

$$\mathrm{T}_h = \{\mathrm{w}_h : \Omega \to \mathbb{R},\ \text{such that } \mathrm{w}_h|_{\mathrm{T}_i} = w_i,\ \text{for } \mathrm{T}_i \in \mathcal{T}_h\}\,.$$

The set T_h is isomorphic to $\mathbb{R}^{\mathrm{card}\{\mathcal{T}_h\}}$ and an identification can be easily established between every function $\mathrm{w}_h \in \mathrm{T}_h$ and the $\mathrm{card}\{\mathcal{T}_h\}$-size vector $\mathbf{w}$, indicated by the corresponding bold letter, such that $w_i = \mathrm{w}_h|_{\mathrm{T}_i}$ for $\mathrm{T}_i \in \mathcal{T}_h$. In the rest of the paper, we use $\mathbf{w}$ or w_h to denote the same vector depending on the context it belongs to.

Let us introduce the two mappings from T_h to itself defined by

$$\mathcal{L}_h \mathrm{w}_h \stackrel{\mathrm{def}}{\equiv} (\mathbf{F} + \mathbf{G} + \mathbf{B})\mathbf{w}\,,$$

$$f_h(\mathrm{w}_h) \stackrel{\mathrm{def}}{\equiv} \mathbf{b} + \mathbf{g} + \mathbf{s} + \mathbf{r}(\mathbf{w})\,,$$

using Eqs. (5.4), (5.6), (5.8), (5.9) and setting $\mathbf{r}(\mathbf{w}) = \mathbf{r}^{\rm int}(\mathbf{w}) + \mathbf{r}^{\rm bnd}(\mathbf{w})$. The discrete problem (3.3) can be formulated in the operator form:

$$\text{find} \quad \mathrm{u}_h \in \mathrm{T}_h \quad \text{such that } \mathcal{L}_h \mathrm{u}_h = f_h(\mathrm{u}_h)\,. \tag{6.1}$$

In the following sections, we discuss the limiting strategy and its influence on the existence and uniqueness of the discrete solution of the nonlinear problem that results from the proposed finite volume discretization.

In particular, we show that $\mathcal{L}_h$ is a coercive operator under suitable assumptions, and that at least one discrete solution exists and this solution is unique in the approximation limit for $h \to 0$.

To analyze these issues theoretically, we use the L^2 mesh-dependent scalar product, its derived norm, and the H^1-norm defined by

$$\begin{aligned}
(\mathrm{u}_h, \mathrm{w}_h)_{\mathcal{T}_h} &= \sum_{\mathrm{T}_i \in \mathcal{T}_h} |\mathrm{T}_i| u_i w_i\,, \\
\|\mathrm{w}_h\|_{\mathcal{T}_h} &= \sqrt{(\mathrm{w}_h, \mathrm{w}_h)_{\mathcal{T}_h}}\,, \\
\|\mathrm{w}_h\|_{\mathcal{T}_h,1} &= \left(\sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\rm int}} \frac{|\mathrm{f}_{ij}|}{H_{ij}} (w_j - w_i)^2 + \sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\rm bnd}} \frac{|\mathrm{f}_{ij}|}{h_{ij}} w_i^2 \right)^{1/2}.
\end{aligned}$$

Let us introduce the set of piecewise constant functions F_h defined for a given triangulation $\mathcal{T}_h$; i.e.

$$\mathrm{F}_h = \{\mathrm{z}_h : \Omega \to \mathbb{R}, \text{ such that } \mathrm{z}_h|_{\chi_{ij}} = \mathrm{z}_{ij}, \text{ for } \mathrm{f}_{ij} \in \mathcal{F}_h\}\,.$$

The set F_h is isomorphic to $\mathbb{R}^{\mathrm{card}\{\mathcal{F}_h\}}$ and an identification can be easily established between every function $\mathrm{z}_h \in \mathrm{F}_h$ and the $\mathrm{card}\{\mathcal{F}_h\}$-size vector $\mathbf{z}$, indicated by the corresponding bold letter, such that $z_{ij} = \mathrm{z}_h|_{\chi_{ij}}$ for $\mathrm{f}_{ij} \in \mathcal{F}_h$. For F_h we also define the L^2 mesh-dependent norm:

$$\|\mathrm{z}_h\|_{\mathcal{F}_h} = \left(\sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\rm int}} |\mathrm{f}_{ij}| H_{ij} z_{ij}^2 + \sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\rm bnd}} |\mathrm{f}_{ij}| h_{ij} z_{ij}^2 \right)^{1/2}.$$

We will also use the inequality:

$$\begin{aligned}
&\sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\rm int}} |\mathrm{f}_{ij}| H_{ij} (|a_{ij} \pm a_{ji}|)^2 + \sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\rm bnd}} |\mathrm{f}_{ij}| h_{ij} |a_{ij}|^2 \\
&\qquad \le 2d(1 + C_{\rm reg}) \sum_{\mathrm{T}_i \in \mathcal{T}_h} |\mathrm{T}_i| \max_{j \in \sigma_i \cup \sigma_i'} |a_{ij}|^2\,,
\end{aligned} \tag{6.2}$$

that holds for every $\mathrm{a}_h \in \mathrm{F}_h$.

6.1. *The limiting strategy*

The reconstructed gradient $\boldsymbol{\mathcal{G}}_i(\mathrm{u}_h)$ in (4.1) includes a limiting factor $l_i(\mathbf{u})$, which is the maximum value in [0, 1] such that the reconstructed solution u_r defined in (4.1) satisfies:

$$\min_{\alpha\in\nu_i} u_\alpha \le u_r(\mathbf{x}_{ij}) \le \max_{\alpha\in\nu_i} u_\alpha\,, \quad \text{for } j \in \sigma_i \cup \sigma_i' \tag{6.3}$$

and the reconstructed gradient is bounded as follows:

$$\|\boldsymbol{\mathcal{G}}(\mathrm{u}_h)\|_{\mathcal{T}_h} \le C_{\mathrm{lim}}\,, \tag{6.4}$$

where $\boldsymbol{\mathcal{G}}(\mathrm{u}_h) = \{\boldsymbol{\mathcal{G}}_i(\mathrm{u}_h) | \mathrm{T}_i \in \mathcal{T}_h\}$ and C_{lim} is a positive real constant independent of h. The procedure for evaluating l_i is:

— set $\widetilde{l_i}$ for $\mathrm{T}_i \in \mathcal{T}_h$ to the maximum value in [0, 1] such that (6.3) is satisfied;
— set $l_i = c\widetilde{l_i}$ where c is the maximum value in [0, 1] such that (6.4) is satisfied.

It should be pointed out that the left-hand side of (6.4) is the T_h-norm of the finite volume approximation of the solution gradient $\nabla\mathrm{u}$ allowing (6.4) to control a discrete definition of the total variation of u_h.

Furthermore, it is worth pointing out that both conditions (6.3), (6.4) are satisfied by taking $l_i = 1$ when a linear solution is reconstructed. This fact occurs because the least squares method discussed in Sec. 4 provides the exact value of the solution at the mesh vertices [(Eq. (4.1)] and of the gradient solution within the mesh cells [Eq. (4.3)]. For this reason, the finite volume scheme proposed in this paper is formally exact for linear solutions even if the reconstruction process is limited.

The term $\mathbf{r}(\mathrm{u}_h)$ defined in Eqs. (5.4) and (5.8) shows the properties described in the next two lemmata.

Lemma 6.1. $\mathbf{r}(\mathrm{u}_h)$ *is a continuous mapping of* u_h.

Proof. This statement derives from the fact that $\mathbf{r}(\mathrm{u}_h)$ is a composition of three continuous operations: the least-squares reconstruction of the vertex values, the reconstructions of the cell-centered gradients, and the application of the limiter. □

Lemma 6.2. *A real constant* $M > 0$ *independent of the mesh size such that* $\|\mathbf{r}(\mathrm{u}_h)\|_{\mathcal{T}_h} \le M$ *exists.*

Proof. First, let us introduce the notations

$$\Delta u_{ij} = (\mathbf{x}_{ij} - \mathbf{x}_i) \cdot \boldsymbol{\mathcal{G}}_i(\mathrm{u}_h)\,, \quad \Delta u_{ji} = (\mathbf{x}_{ji} - \mathbf{x}_j) \cdot \boldsymbol{\mathcal{G}}_j(\mathrm{u}_h)\,.$$

In view of (5.4) and (2.3), applying the Cauchy–Schwartz inequality, and rewriting the summations over the set of edges of the mesh, it follows that

$$\|\mathbf{r}(\mathrm{u}_h)\|^2_{\mathcal{T}_h} = \sum_{\mathrm{T}_i\in\mathcal{T}_h} |\mathrm{T}_i| \left[\frac{1}{|\mathrm{T}_i|}\left(\sum_{j\in\sigma_i} |\mathrm{f}_{ij}|[v^+_{ij}\Delta u_{ij} + v^-_{ij}\Delta u_{ji}] + \sum_{j\in\sigma'_i} |\mathrm{f}_{ij}| v^+_{ij}\Delta u_{ij}\right)\right]^2,$$
$$\leq 2(d+1)\beta_2{}^2 \times (\star)\,,$$

where

$$(\star) = \sum_{\mathrm{f}_{ij}\in\mathcal{F}_h^{\mathrm{int}}} \left(\frac{|\mathrm{f}_{ij}|^2}{|\mathrm{T}_i|} + \frac{|\mathrm{f}_{ij}|^2}{|\mathrm{T}_j|}\right)(|\Delta u_{ij}|^2 + |\Delta u_{ji}|^2) + \sum_{\mathrm{f}_{ij}\in\mathcal{F}_h^{\mathrm{bnd}}} \frac{|\mathrm{f}_{ij}|^2}{|\mathrm{T}_i|}|\Delta u_{ij}|^2\,.$$

Since $|\mathrm{T}_i| = (d+1)h_{ij}|\mathrm{f}_{ij}|/d$, by multiplying and dividing by H_{ij} and exploiting the fact that from (3.1) it follows that $H_{ij}/h_{ij} \leq 1 + C_{\mathrm{reg}}$, we have

$$(\star) \leq \frac{d(1+C_{\mathrm{reg}})}{d+1}\left[\sum_{\mathrm{f}_{ij}\in\mathcal{F}_h^{\mathrm{int}}} |\mathrm{f}_{ij}| H_{ij} \frac{|\Delta u_{ij}|^2 + |\Delta u_{ji}|^2}{H_{ij}^2} + \sum_{\mathrm{f}_{ij}\in\mathcal{F}_h^{\mathrm{bnd}}} |\mathrm{f}_{ij}| h_{ij} \frac{|\Delta u_{ij}|^2}{h_{ij}^2}\right].$$

Using (6.2) with $a_{ij} = \Delta u_{ij}/H_{ij}$ and $|\Delta u_{ij}| \leq h|\boldsymbol{\mathcal{G}}_i(\mathrm{u}_h)|$ yields

$$(\star) \leq \frac{2d^2(1+C_{\mathrm{reg}})^2 C^2}{d+1} \sum_{\mathrm{T}_i\in\mathcal{T}_h} |\mathrm{T}_i|\,|\boldsymbol{\mathcal{G}}_i(\mathrm{u}_h)|^2\,,$$

where $h/H_{ij} \leq C$ is used from (3.1). In view of (6.4), the statement of the lemma is true for $M = 2d\beta_2 C_{\mathrm{lim}} C(1 + C_{\mathrm{reg}})$. □

6.2. *Coercivity and Poincaré's inequality*

The coercivity of the finite volume operator $\mathcal{L}_h$ can be stated under suitable mesh assumptions when $d = 2$.[9,11] The basic result of this subsection is stated in Proposition 6.5 and generalizes the similar result of formula 25, Sec. 4.1 of Refs. 9, 11 to the case $d > 2$. Essentially, a sufficient condition is formulated for the reconstructed gradient at the internal faces to ensure the coercivity of the discrete operator $\mathcal{L}_h$. The satisfaction of this condition for the approximate finite volume solution is clearly related to the deformation of the mesh as it depends on the fact that the orthogonal projections of the cell centers $\mathbf{x}_i$ and $\mathbf{x}_j$ on $\mathrm{f}_{ij} \in \mathcal{F}_h^{\mathrm{int}}$ may not coincide on the triangulations allowed by Assumption 3.1. It should be noted that the requirement stated by Proposition 6.5 is trivially met on several kind of meshes, such as, for example, the Voronoi ones. Finally, it is important to say that the numerical investigations that have been extensively performed on unstructured meshes suggest that this condition might be unnecessary. This is also perfectly in agreement with what Coudière, Vila, and Villedieu[9] have already observed.

Definition 6.3. The discrete operator $\mathcal{L}_h$ is said to be *coercive* if there exist a real positive constant C_{coerc}, which is independent of h, and a real positive constant h_0 such that the inequality

$$C_{\text{coerc}}\|\mathrm{w}_h\|^2_{\mathcal{T}_h,1} \le (\mathrm{w}_h, \mathcal{L}_h \mathrm{w}_h)_{\mathcal{T}_h}\,, \quad \text{for any } \mathrm{w}_h \in \mathrm{T}_h$$

is true for every $h \le h_0$.

Lemma 6.3. *For any* $\mathrm{u}_h \in \mathrm{T}_h$ *we have*

$$(\mathrm{u}_h, \mathcal{L}_h \mathrm{u}_h)_{\mathcal{T}_h} \ge \frac{\beta_1}{2}\|\mathrm{u}_h\|^2_{\mathcal{T}_h} + \beta_0 \|\mathrm{u}_h\|^2_{\mathcal{T}_h,1} - |\mathcal{S}_h(\mathrm{u}_h)|\,, \tag{6.5}$$

where

$$\mathcal{S}_h(\mathrm{u}_h) = \sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\text{int}}} \frac{|\mathrm{f}_{ij}|\nu_{ij}}{H_{ij}} (u_i - u_j)(\tilde{u}^0_{ij} - \tilde{u}^0_{ji})\,,$$

is the tangential skewness, and $\tilde{u}^0_{ij} = \sum_{\alpha \in \nu^{\text{int}}_{ij}} \tilde{\lambda}^{ij}_\alpha u_\alpha$.

Proof. Let $\mathbf{T} = \operatorname{diag}\{|\mathrm{T}_i|, \mathrm{T}_i \in \mathcal{T}_h\}$. In view of (5.4), we have

$$\begin{aligned}
2\mathbf{u}^T\mathbf{T}\mathbf{F}\mathbf{u} &= 2\sum_{\mathrm{T}_i \in \mathcal{T}_h} u_i \sum_{j \in \sigma_i} |\mathrm{f}_{ij}|[v^+_{ij}u_i + v^-_{ij}u_j]\,,\\
&= \sum_{\mathrm{T}_i \in \mathcal{T}_h} u_i^2 \sum_{j \in \sigma_i} |\mathrm{f}_{ij}| v_{ij} + \sum_{\mathrm{T}_i \in \mathcal{T}_h} \sum_{j \in \sigma_i} |\mathrm{f}_{ij}|\,|v_{ij}|(u_i^2 - u_i u_j)\,,\\
&= \sum_{\mathrm{T}_i \in \mathcal{T}_h} u_i^2 \int_{\mathrm{T}_i} \operatorname{div} \mathbf{v}\, dV - \sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\text{bnd}}} |\mathrm{f}_{ij}| u_i^2 v_{ij} + \sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\text{int}}} |\mathrm{f}_{ij}|\,|v_{ij}|\,(u_i - u_j)^2\,,\\
&\ge \beta_1 \|\mathrm{u}_h\|^2_{\mathcal{T}_h} - \sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\text{bnd}}} |\mathrm{f}_{ij}| u_i^2 v_{ij}\,.
\end{aligned}$$

In view of (5.6) and recalling (4.8), we have

$$\begin{aligned}
\mathbf{u}^T\mathbf{T}\mathbf{G}\mathbf{u} &= \sum_{\mathrm{T}_i \in \mathcal{T}_h} u_i \sum_{j \in \sigma_i} \frac{|\mathrm{f}_{ij}|\nu_{ij}}{H_{ij}}[(u_i - u_j) - (\tilde{u}^0_{ij} - \tilde{u}^0_{ji})]\,,\\
&= \sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\text{int}}} \frac{|\mathrm{f}_{ij}|\nu_{ij}}{H_{ij}}[(u_i - u_j)^2 - (u_i - u_j)(\tilde{u}^0_{ij} - \tilde{u}^0_{ji})]\,,\\
&\ge \beta_0 \|\mathrm{u}_h\|^2_{\mathcal{T}_h,1} - \mathcal{S}_h(\mathrm{u}_h)\,.
\end{aligned}$$

In view of (5.8), we have

$$\mathbf{u}^T\mathbf{T}\mathbf{B}\mathbf{u} = \sum_{\mathrm{T}_i \in \mathcal{T}_h} u_i \sum_{j \in \sigma'_i} |\mathrm{f}_{ij}| \left[v^+_{ij}u_i + \nu_{ij}\frac{u_i}{h_{ij}}\right] = \sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\text{bnd}}} u_i^2\,|\mathrm{f}_{ij}| \left[v^+_{ij} + \frac{\nu_{ij}}{h_{ij}}\right].$$

By using the previous three inequalities, we have

$$
\begin{aligned}
(\mathrm{u}_h, \mathcal{L}_h \mathrm{u}_h)_{\mathcal{T}_h} &= \mathbf{u}^T \mathbf{T}(\mathbf{F}+\mathbf{G}+\mathbf{B})\mathbf{u}\,, \\
&\geq \frac{\beta_1}{2}\|\mathrm{u}_h\|^2_{\mathcal{T}_h} + \beta_0 \|\mathrm{u}_h\|^2_{\mathcal{T}_h,1} - \mathcal{S}_h(\mathrm{u}_h) \\
&\quad + \sum_{\mathrm{f}_{ij}\in\mathcal{F}_h^{\mathrm{bnd}}} u_i^2 |\mathrm{f}_{ij}| \left[\underbrace{v_{ij}^+ - \frac{1}{2} v_{ij}}_{=|v_{ij}|/2} + \frac{\nu_{ij}}{h_{ij}} \right].
\end{aligned}
$$

The lemma statement follows by noting that the last term is non-negative. □

Proposition 6.5. (Coercivity[9,11]) *Let us assume that there exists a real positive constant* $\gamma < 1$, *which is independent of* h, *and a real positive constant* h_0 *such that the inequality* $|\mathcal{S}_h(\mathrm{u}_h)| \leq \gamma\beta_0 \|\mathrm{u}_h\|^2_{\mathcal{T}_h,1}$ *is true for every* $h \leq h_0$. *Then,* $\mathcal{L}_h$ *is a* coercive *operator and* $(1-\gamma)\beta_0$ *can be taken as coercivity constant.*

Proof. From Eq. (6.5)

$$
\begin{aligned}
(\mathrm{u}_h, \mathcal{L}_h \mathrm{u}_h)_{\mathcal{T}_h} &\geq \frac{\beta_1}{2}\|\mathrm{u}_h\|^2_{\mathcal{T}_h} + \beta_0 \|\mathrm{u}_h\|^2_{\mathcal{T}_h,1} - \gamma\beta_0 \|\mathrm{u}_h\|^2_{\mathcal{T}_h,1} \\
&\geq (1-\gamma)\beta_0 \|\mathrm{u}_h\|^2_{\mathcal{T}_h,1}\,.
\end{aligned}
$$

□

The following proposition is the d-dimensional version of a lemma found in Eymard *et al.* for $d = 2, 3$ (see p. 38 of Ref. 11). The extension to $d > 3$ is straightforward and is omitted.

Proposition 6.6. (Discrete Poincaré's inequality) *The following inequality holds*:

$$
\|\mathrm{u}_h\|_{\mathcal{T}_h} \leq \mathrm{diam}\{\Omega\} \|\mathrm{u}_h\|_{\mathcal{T}_h,1}\,, \quad \textit{for } \mathrm{u}_h \in \mathrm{T}_h\,.
$$

6.3. *Existence and quasi-uniqueness of the finite volume approximation*

Theorem 6.7. *If* $\mathcal{L}_h$ *is a coercive operator, the problem defined by Eq.* (6.1) *has at least one solution.*

Proof. First, it should be noted that $\mathcal{L}_h$ is non-singular because it is a coercive operator.

$$
\Phi(\mathrm{w}_h) = \mathcal{L}_h^{-1} f_h(\mathrm{w}_h)
$$

is continuous because from Lemma 6.1 it follows that $f_h(\mathrm{w}_h)$ depends continuously on w_h. Let us denote $\mathrm{p}_h = \Phi(\mathrm{w}_h)$. Then, p_h is the solution of $\mathcal{L}_h \mathrm{p}_h = f_h(\mathrm{w}_h)$ and we have

$$
(\mathrm{p}_h, \mathcal{L}_h \mathrm{p}_h)_{\mathcal{T}_h} = (\mathrm{p}_h, f_h(\mathrm{w}_h))_{\mathcal{T}_h}\,. \tag{6.6}
$$

In view of the coercivity of $\mathcal{L}_h$ and by using the Poincaré's Proposition 6.6, we have

$$C_{\text{coerc}}\,\text{diam}\{\Omega\}^{-2}\|\mathrm{p}_h\|^2_{\mathcal{T}_h} \le (\mathrm{p}_h, \mathcal{L}_h\mathrm{p}_h)_{\mathcal{T}_h}\,. \tag{6.7}$$

By using the Cauchy–Schwartz inequality it follows that

$$(\mathrm{p}_h, f_h(\mathrm{w}_h))_{\mathcal{T}_h} \le \|\mathrm{p}_h\|_{\mathcal{T}_h}\ \|f_h(\mathrm{w}_h)\|_{\mathcal{T}_h}\,.$$

By combining (6.6) and (6.7) and using the inequality of Lemma 6.2, we finally obtain

$$C_{\text{corec}}\,\text{diam}\{\Omega\}^{-2}\|\mathrm{p}_h\|_{\mathcal{T}_h} \le \|f_h(\mathrm{w}_h)\|_{\mathcal{T}_h} \le \|\mathbf{b}+\mathbf{g}+\mathbf{s}\|_{\mathcal{T}_h} + M\,.$$

Hence, Φ is a continuous mapping from the convex compact set

$$\{\mathrm{w}_h \in \mathrm{T}_h,\ \|\mathrm{w}_h\|_{\mathcal{T}_h} \le (\|\mathbf{b}+\mathbf{g}+\mathbf{s}\|_{\mathcal{T}_h} + M)\ \text{diam}\{\Omega\}^2/C_{\text{coerc}}\}$$

into itself. The result follows from the Brouwer fixed-point theorem (see Zeidler[28]). □

Theorem 6.8. *If* u_h *and* $\tilde{\mathrm{u}}_h$ *are two* distinct *solutions of problem* (6.1), *then the following inequality holds*:

$$\|\mathrm{u}_h - \tilde{\mathrm{u}}_h\|_{\mathcal{T}_h,1} \le C\|\boldsymbol{\mathcal{G}}(\mathrm{u}_h) - \boldsymbol{\mathcal{G}}(\tilde{\mathrm{u}}_h)\|_{\mathcal{T}_h}\,h\,,$$

where $\boldsymbol{\mathcal{G}}(\mathrm{u}_h)$ *and* $\boldsymbol{\mathcal{G}}(\tilde{\mathrm{u}}_h)$ *are the reconstructed cell gradients defined in* (4.3) *and* C *is independent of* h. *In particular, if condition* (6.4) *is satisfied, there holds*

$$\|\mathrm{u}_h - \tilde{\mathrm{u}}_h\|_{\mathcal{T}_h,1} \le 2CC_{\text{lim}}h\,,$$

i.e. all the possible solutions are in a ball of radius $\mathcal{O}(h)$ *in the norm* $\|\cdot\|_{\mathcal{T}_h,1}$.

Proof. Let u_h and $\tilde{\mathrm{u}}_h$ be two solutions of problem (6.1). From the coercivity of $\mathcal{L}_h$ stated in Proposition 6.5, there follows

$$\begin{aligned} C_{\text{coerc}}\|\mathrm{u}_h - \tilde{\mathrm{u}}_h\|^2_{\mathcal{T}_h,1} &\le (\mathrm{u}_h - \tilde{\mathrm{u}}_h, \mathcal{L}_h(\mathrm{u}_h - \tilde{\mathrm{u}}_h))_{\mathcal{T}_h}\,, \\ &= (\mathrm{u}_h - \tilde{\mathrm{u}}_h, f_h(\mathrm{u}_h) - f_h(\tilde{\mathrm{u}}_h))_{\mathcal{T}_h}\,. \end{aligned} \tag{6.8}$$

By introducing the notations

$$z_i = u_i - \tilde{u}_i\,, \quad \Delta z_{ij} = (\mathbf{x}_{ij} - \mathbf{x}_i)\cdot(\boldsymbol{\mathcal{G}}_i(\mathrm{u}_h) - \boldsymbol{\mathcal{G}}_i(\tilde{\mathrm{u}}_h))\,,$$

the right-hand side of (6.8) takes the form

$$\begin{aligned} &(\mathrm{u}_h - \tilde{\mathrm{u}}_h, f_h(\mathrm{u}_h) - f_h(\tilde{\mathrm{u}}_h))_{\mathcal{T}_h} \\ &\quad= \sum_{\mathrm{T}_i\in\mathcal{T}_h} z_i \left[\sum_{j\in\sigma_i} |\mathrm{f}_{ij}|(v^+_{ij}\Delta z_{ij} + v^-_{ij}\Delta z_{ji}) + \sum_{j\in\sigma'_i} |\mathrm{f}_{ij}|v^+_{ij}\Delta z_{ij}\right], \\ &\quad= \sum_{\mathrm{f}_{ij}\in\mathcal{F}^{\text{int}}_h} |\mathrm{f}_{ij}|(v^+_{ij}\Delta z_{ij} + v^-_{ij}\Delta z_{ji})(z_i - z_j) + \sum_{\mathrm{f}_{ij}\in\mathcal{F}^{\text{bnd}}_h} |\mathrm{f}_{ij}|v^+_{ij}\Delta z_{ij}z_i\,, \end{aligned}$$

where the final expression is obtained by transforming the summations over the mesh control volumes into summations over the mesh faces. By using the standard Hölder inequality, we have

$$(\mathrm{u}_h - \tilde{\mathrm{u}}_h, f_h(\mathrm{u}_h) - f_h(\tilde{\mathrm{u}}_h))_{\mathcal{T}_h} \leq \|\mathrm{u}_h - \tilde{\mathrm{u}}_h\|_{\mathcal{T}_h,1} \sqrt{(\star)}\,, \tag{6.9}$$

where

$$\begin{aligned}(\star) = &\sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\mathrm{int}}} |\mathrm{f}_{ij}| H_{ij} (v_{ij}^+ \Delta z_{ij} - v_{ji}^+ \Delta z_{ji})^2 \\ &+ \sum_{\mathrm{f}_{ij} \in \mathcal{F}_h^{\mathrm{bnd}}} |\mathrm{f}_{ij}| h_{ij} (v_{ij}^+ \Delta z_{ij})^2\,. \end{aligned} \tag{6.10}$$

Equation (6.10) can be transformed as

$$(\star) \leq 2dh^2 {\beta_2}^2 (1 + C_{\mathrm{reg}}) \sum_{\mathrm{T}_i \in \mathcal{T}_h} |\mathrm{T}_i|\, |\boldsymbol{\mathcal{G}}_i(\mathrm{u}_h) - \boldsymbol{\mathcal{G}}_i(\tilde{\mathrm{u}}_h)|^2 \tag{6.11}$$

by using (6.2) with $a_{ij} = v_{ij}^+ \Delta z_{ij}$, the inequality (2.3), and

$$|v_{ij}^+ \Delta z_{ij}| \leq h \beta_2 |\boldsymbol{\mathcal{G}}_i(\mathrm{u}_h) - \boldsymbol{\mathcal{G}}_i(\tilde{\mathrm{u}}_h)|\,.$$

The theorem statement follows by combining (6.8), (6.9) and (6.11) with $C = \beta_2 (2d(1 + C_{\mathrm{reg}})/C_{\mathrm{coerc}})^{1/2}$. □

7. Numerical Experiments

In this section, we provide the numerical results for the scheme discussed in this paper. The performance of the method, in terms of order of accuracy and non-oscillatory shock-capturing behavior, is studied for two different test problems.

The first problem has a smooth solution by suitably setting the boundary conditions and the source term in the model equation. This case is considered to experimentally assess the convergence rate on a wide range of values of the viscosity coefficient. Let us highlight that the expected convergence rate is 2 for the approximation of solution cell averages and 1 for the approximation of solution cell-averaged gradients.

Instead, the second test case is chosen to demonstrate the shock-capturing capabilities of this finite volume scheme in solving problems whose solutions display internal and boundary layers.

The mesh manager used to construct the finite-volume approximation was implemented by means of P2MESH (see Bertolazzi and Manzini[4]), a free software package conceived for the fast development of finite volume and finite element codes on 2D *unstructured* meshes.

All the computational grids are generated by refining a basic coarse grid based on the recursive strategy described below. The coarsest unstructured grid is generated by the Delaunay mesh generator TRIANGLE developed by Shewchuk,[21] and is made up of 87 vertices, 139 triangles, 225 edges, 33 boundary edges. This mesh is

indicated by the label $M = 1$ as it is the first level mesh. Given a mesh at level M, we partition each triangle into four equally sized subtriangles by connecting its edge midpoints. This nested mesh is the next refinement level mesh, i.e. the mesh at the refinement level $M + 1$. This strategy produces a sequence of nested grids such that the grid at level $M + 1$ has exactly four times the number of triangles of the coarser M-level grid. Furthermore, though the mesh size is exactly halved at each refinement step, the aspect ratio of the triangular mesh cells is preserved.

For this sequence of meshes we measure:

— the error of the cell-averaged solution $\mathcal{E}_h = \|\mathrm{u}_h - \mathcal{A}_{\mathcal{T}_h}(\mathrm{u})\|_{\mathcal{T}_h}$;
— the error of the cell-centered gradient $\mathcal{E}_{\mathrm{G},h} = \|\boldsymbol{\mathcal{G}}(\mathrm{u}_h) - \mathcal{A}_{\mathcal{T}_h}(\nabla \mathrm{u})\|_{\mathcal{T}_h}$,

where $\mathcal{A}_{\mathcal{T}_h}(\cdot)$ denotes the L^2-orthogonal projector on the space of piecewise constant functions defined on the triangulation $\mathcal{T}_h$.

The experimental rate of convergence is numerically estimated by comparing the errors on two successive mesh level calculations. In other words, the convergence rate of the cell-averaged solution approximation is given by applying the standard formula

$$\mathrm{Rate}(h, h/2) = \log_2 \frac{\mathcal{E}_h}{\mathcal{E}_{h/2}}\,.$$

The same formula with $\mathcal{E}_{\mathrm{G},h}$, instead of $\mathcal{E}_h$, makes it possible to estimate the convergence rate of the cell-averaged gradient approximations.

7.1. *Test case* 1

In the first test case we impose a forcing term according to the analytical solution

$$\mathrm{u}(x, y) = 16\, x(1 - x)\, y(1 - y)\,,$$

for the computational domain $\Omega = [0, 1] \times [0, 1]$, where

$$v(x, y) = \begin{bmatrix} y(1 - y) \\ 0 \end{bmatrix}, \quad \nu(x, y) = \kappa(1 + x^2)\,.$$

Tables 1 and 2 show the results of this test case for respectively the first- and second-order versions of the scheme, and the viscosity coefficient κ decreasing from 10^4 to 10^{-4}. In both tables, the first column gives the value of κ and the second one the mesh size level M. In Table 2, the third column reports the number of iterations required to achieve convergence when solving the nonlinear problem of the second-order method. The other columns, whose labels are self-explanatory, report the measured errors and convergence rates.

In the nonlinear version of the scheme, the iterations are stopped when two successive iterates $(\mathrm{u}_h)^k$ and $(\mathrm{u}_h)^{k+1}$ satisfy the condition

$$\|(\mathrm{u}_h)^{k+1} - (\mathrm{u}_h)^k\|_\infty \leq 10^{-12}\,.$$

Table 1. Approximation errors of test case 1, first-order scheme.

		Cell averages		Gradients	
κ	M	$\mathcal{E}_h$	Rate	$\mathcal{E}_{\mathrm{G},h}$	Rate
	1	1.104×10^{-2}	—	4.529×10^{-1}	—
	2	2.635×10^{-3}	2.066	2.222×10^{-1}	1.027
10^4	3	6.308×10^{-4}	2.062	1.104×10^{-1}	1.008
	4	1.542×10^{-4}	2.032	5.513×10^{-2}	1.003
	5	3.802×10^{-5}	2.020	2.753×10^{-2}	1.002
	1	1.103×10^{-2}	—	4.529×10^{-1}	—
	2	2.630×10^{-3}	2.068	2.222×10^{-1}	1.010
10^2	3	6.280×10^{-4}	2.066	1.105×10^{-1}	1.023
	4	1.553×10^{-4}	2.039	5.551×10^{-2}	1.005
	5	3.731×10^{-5}	2.034	2.753×10^{-2}	1.002
	1	1.011×10^{-2}	—	4.534×10^{-1}	—
	2	2.209×10^{-3}	2.195	2.222×10^{-1}	1.028
10^0	3	5.103×10^{-4}	2.114	1.106×10^{-1}	1.008
	4	1.913×10^{-4}	1.415	5.515×10^{-2}	1.003
	5	1.027×10^{-5}	0.897	2.754×10^{-2}	1.002
	1	5.635×10^{-2}	—	5.710×10^{-1}	—
	2	3.436×10^{-2}	0.714	3.101×10^{-1}	0.881
10^{-2}	3	2.007×10^{-2}	0.775	1.658×10^{-1}	0.904
	4	1.107×10^{-2}	0.858	8.638×10^{-2}	0.940
	5	5.855×10^{-3}	0.919	4.419×10^{-2}	0.967
	1	1.033×10^{-1}	—	8.155×10^{-1}	—
	2	5.011×10^{-2}	1.043	6.223×10^{-1}	0.389
10^{-4}	3	2.523×10^{-2}	0.989	4.580×10^{-1}	0.442
	4	1.326×10^{-2}	0.928	3.144×10^{-2}	0.543
	5	7.002×10^{-3}	0.921	2.041×10^{-2}	0.624

This stopping criterion is quite restrictive and not optimized at all; instead, a less conservative and expensive one could be investigated.

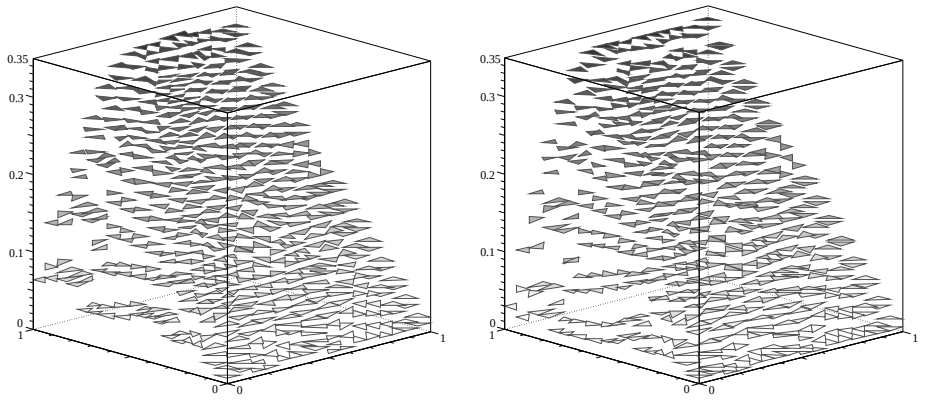
Tables 1 and 2 display an excellent agreement of the experimental convergence rates with the theoretical ones. In particular, we note that the first-order version of the method (see Table 1) shows a second-order of accuracy in the cell average approximation for the greatest values of κ for the coarsest meshes, while the correct first-order rate of convergence is achieved when the mesh is refined. This behavior is reasonable because, in the earliest calculations, the error in the approximation of the diffusive term is likely to be predominant over the error of the convective one. On the other hand, the expected second-order of accuracy is confirmed by the results in Table 2, i.e. when the nonlinear scheme is considered. Finally, it is worth pointing out that in both versions the convergence rate of error of the gradient approximation is the first-order of accuracy as was theoretically expected.

Table 2. Approximation errors of test case 2, second-order scheme.

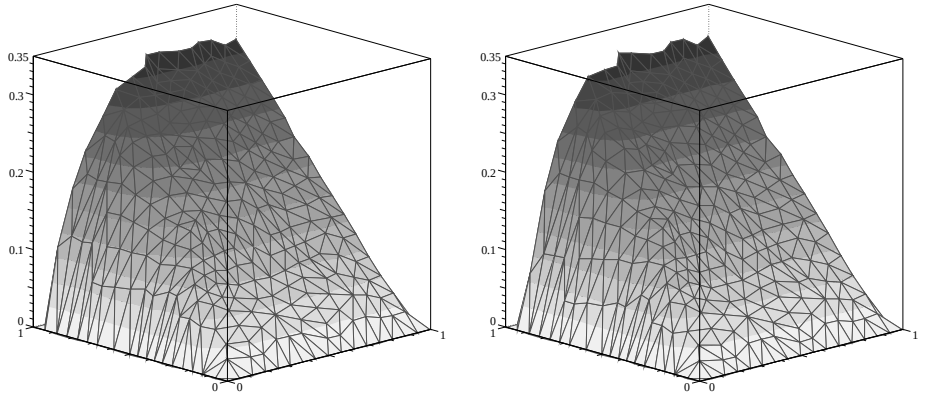
			Cell averages		Gradients	
κ	M	#Iter	$\mathcal{E}_h$	Rate	$\mathcal{E}_{\mathrm{G},h}$	Rate
	1	3	1.104×10^{-2}	—	4.529×10^{-1}	—
	2	3	2.635×10^{-3}	2.066	2.222×10^{-1}	1.027
10^4	3	2	6.332×10^{-4}	2.057	1.104×10^{-1}	1.009
	4	1	1.544×10^{-4}	2.035	5.507×10^{-2}	1.003
	5	2	3.806×10^{-5}	2.020	2.751×10^{-2}	1.001
	1	17	1.103×10^{-2}	—	4.529×10^{-1}	—
	2	16	2.633×10^{-3}	2.066	2.222×10^{-1}	1.027
10^2	3	16	6.329×10^{-4}	2.056	1.104×10^{-1}	1.009
	4	14	1.543×10^{-4}	2.036	5.507×10^{-2}	1.003
	5	10	3.805×10^{-5}	2.019	2.751×10^{-2}	1.001
	1	29	1.109×10^{-2}	—	4.529×10^{-1}	—
	2	26	2.616×10^{-3}	2.083	2.222×10^{-1}	1.027
10^0	3	23	6.302×10^{-4}	2.053	1.104×10^{-1}	1.009
	4	21	1.539×10^{-4}	2.033	5.507×10^{-2}	1.003
	5	18	3.798×10^{-5}	2.018	2.751×10^{-2}	1.001
	1	56	9.465×10^{-3}	—	4.604×10^{-1}	—
	2	42	2.129×10^{-3}	2.156	2.236×10^{-1}	1.041
10^{-2}	3	37	5.378×10^{-4}	1.985	1.106×10^{-1}	1.015
	4	32	1.426×10^{-4}	1.914	5.510×10^{-2}	1.005
	5	27	3.790×10^{-5}	1.912	2.751×10^{-2}	1.002
	1	56	3.571×10^{-2}	—	5.012×10^{-1}	—
	2	42	9.413×10^{-3}	1.923	2.439×10^{-1}	1.039
10^{-4}	3	37	1.830×10^{-3}	2.362	1.181×10^{-1}	1.046
	4	32	4.251×10^{-4}	2.105	5.684×10^{-2}	1.055
	5	27	1.049×10^{-4}	2.019	2.751×10^{-2}	1.046

7.2. *Test case* 2

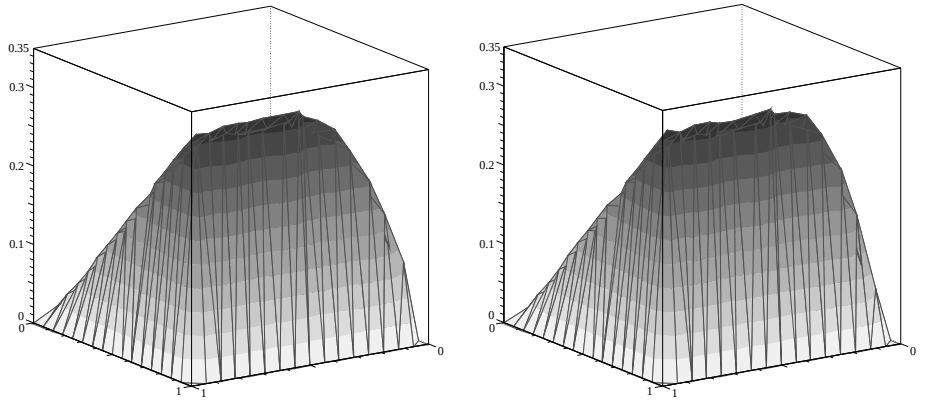
This test case illustrates the non-oscillatory behavior of the method when applied to the resolution of problems with strong gradient solutions. Let us take $\Omega = [0,1] \times [0,1]$, $f = 1$, $\mathbf{v} = (1/3, 1)$, $\mu = 10^{-6}$, and homogeneous Dirichlet conditions, namely $g = 0$ on $\partial\Omega$. Due to this boundary condition, a boundary layer is developed in the analytical solution. The approximate results are displayed in Fig. 3. This figure depicts the piecewise constant approximation of cell averages, and the (continuous) linearly-reconstructed values at the mesh vertices (front and rear views). The pictures on the left show the results provided by the first-order scheme, while the pictures on the right the results provided by the second-order MUSCL method proposed in this paper. The absence of numerical oscillations (overshoots and undershoots) is remarkably evident.



Cell-Average Approximations



Vertex Approximations (rear view)



Vertex Approximations (front view)

Fig. 3. Test case 2: finite volume first-order (left) and second-order (right) approximation of solution cell averages (top row) and linearly-reconstructed vertex values (mid and bottom rows).

8. Conclusions

A MUSCL-like cell-centered finite volume approximation method that applies to the second-order steady advection–diffusion equation in divergence form was formulated. The method was proposed on unstructured d-simplex meshes, where $d \geq 2$ is the spatial dimension. It is conservative and formally second-order accurate in the approximation of the cell-averaged solution values. Vertex-centered solution approximations and cell-centered solution gradient approximations are reconstructed from cell-averaged data by a suitable linear least squares algorithm. Both these approximations are formally second-order accurate. This fact was tested numerically and the method proved to be in excellent agreement with the theoretical predictions.

The reconstruction step and introduction of a suitable limiter to prevent numerical oscillations in the strong gradient regions, which are usual features in shock-capturing MUSCL schemes, imply that the method is nonlinear, even if the model problem is a linear one. This nonlinearity was solved by a simple fixed-point iterative technique. A careful design of the limiting strategy allowed us to prove the existence of at least one fixed point solution with asymptotic uniqueness for $h \to 0$. The asymptotic uniqueness was formulated in the sense that the distance between two possibly distinct fixed points must be at most of order $\mathcal{O}(h)$ when it is measured using a suitable mesh-dependent norm. Finally, the robustness of the method was verified experimentally by approximating of a boundary layer problem for a wide interval of viscosity values ranging from a diffusive situation to a strongly advected-dominated one.

References

1. S. Agmon, *Elliptic Boundary Value Problems*, Mathematical Studies (Van Nostrand, 1965).
2. B. R. Baliga and S. V. Patankar, A new finite-element formulation for convection–diffusion problems, *Numer. Heat Transfer* **3** (1980) 393–409.
3. R. E. Bank and D. J. Rose, Some error estimates for the box method, *SIAM J. Numer. Anal.* **24** (1987) 777–787.
4. E. Bertolazzi and G. Manzini, Algorithm 817 P2MESH: Generic object-oriented interface between 2-D unstructured meshes and FEM/FVM-based PDE solvers, *ACM TOMS* **28** (2002) 101–132.
5. ———, Polynomial reconstructions and limiting strategies in finite volume approximations, in *Finite Volumes for Complex Applications III, Problems and Perspectives* (Hermes Penton Science, 2002).
6. Z. Cai, On the finite volume element method, *Numer. Math.* **58** (1991) 713–735.
7. Z. Cai, J. Mandel and S. McCormick, The finite volume element method for diffusion equations on composite grids, *SIAM J. Numer. Anal.* **28** (1991) 392–402.
8. P. G. Ciarlet, *The Finite Element Method for Elliptic Problems* (North-Holland, 1980).
9. Y. Coudière, J.-P. Vila and P. Villedieu, Convergence rate of a finite volume scheme for a two dimensional convection–diffusion problem, *Math. Model. Numer. Anal.* **33** (1999) 493–516.

10. Y. Coudière and P. Villedieu, Convergence rate of a finite volume scheme for the linear convection–diffusion equation on locally refined meshes, *Math. Model. Numer. Anal.* **34** (1999) 1123–1149.
11. R. Eymard, T. Gallouët and R. Herbin, Finite volume methods, in *Handbook of Numerical Analysis*, Vol. VII (North-Holland, 2000), pp. 713–1020.
12. W. Hackbusch, On first and second-order box scheme, *Computing* **41** (1989) 277–296.
13. A. Harten, High resolution schemes for hyperbolic conservation laws, *J. Comput. Phys.* **49** (1983) 357–393.
14. A. Harten, B. Engquist, S. Osher and S. R. Chakravarthy, Uniformly high-order accurate essentially nonoscillatory schemes. III, *J. Comput. Phys.* **71** (1987) 231–303.
15. B. Heinrich, Finite difference methods on irregular networks (1987).
16. M. E. Hubbard, Multidimensional slope limiters for MUSCL-type finite volume schemes on unstructured grids, *J. Comput. Phys.* **155** (1999) 54–74.
17. R. D. Lazarov, V. L. Makarov and W. Weinelt, On the convergence of difference schemes for the approximation of solutions $u \in W_2^m (m > 0.5)$ of elliptic equations with mixed derivatives, *Numer. Math.* **44** (1984) 223–232.
18. S. McCormick, *Multilevel Adaptive Methods for Partial Differential Equations* (Soc. Indus. Appl. Math., 1989).
19. K. W. Morton, *Numerical Solution of Convection–Diffusion Problems* (Chapman & Hall, 1996).
20. P.-A. Raviart and J.-M. Thomas, *Introduction à l'Analyse Numérique des Équations aux Dérivées Partielles*, Collection Mathématiques Appliquées pour la Maîtrise (Masson, 1983).
21. J. R. Shewchuk, Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator, in *Applied Computational Geometry*: *Towards Geometric Engineering*, *Lecture Notes in Computer Science*, Vol. 1148 (Springer-Verlag, 1996), pp. 203–222. From the First ACM Workshop on Applied Computational Geometry.
22. C.-W. Shu, High order ENO and WENO schemes for computational fluid dynamics, in *High-Order Methods for Computational Physics* (Springer, 1999), pp. 439–582.
23. C.-W. Shu and S. Osher, Efficient implementation of essentially nonoscillatory shock-capturing schemes. II, *J. Comput. Phys.* **83** (1989) 32–78.
24. P. S. Vassilievski, S. I. Petrova and R. D. Lazarov, Finite difference schemes on triangular cell-centered grids with local refinement, *SIAM J. Sci. Statist. Comput.* **13** (1992) 1287–1313.
25. A. Weiser and M. F. Wheeler, On convergence of block-centered finite differences for elliptic problems, *SIAM J. Numer. Anal.* **25** (1988) 351–375.
26. M. Wierse, A new theoretically motivated higher order upwind scheme on unstructured grids of simplices, *Adv. Comput. Math.* **7** (1997) 303–335.
27. P. Wilmott, J. Dewynne and S. Howison, *Option Pricing*: *Mathematical Models and Computation* (Financial Press, 1995).
28. E. Zeidler, *Nonlinear Functional Analysis and its Applications* (Springer-Verlag, 1986).