A mixed finite volume scheme for anisotropic diffusion problems on any grid

Jérôme Droniou, Robert Eymard

A mixed finite volume scheme for anisotropic diffusion problems on any grid

J. Droniou* R. Eymard†

21/04/2005

Abstract

We present a new finite volume scheme for anisotropic heterogeneous diffusion problems on unstructured irregular grids, which simultaneously gives an approximation of the solution and of its gradient. In the case of simplicial meshes, the approximate solution is shown to converge to the continuous ones as the size of the mesh tends to 0, and an error estimate is given. In the general case, we propose a slightly modified scheme for which we again prove the convergence, and give an error estimate. An easy implementation method is then proposed, and the efficiency of the scheme is shown on various types of grids.

Keywords. Finite volume scheme, unstructured grids, irregular grids, anisotropic diffusion problems.

1 Introduction

The computation of an approximate solution for equations involving a second order elliptic operator is needed in so many physical and engineering areas, where the efficiency of some discretization methods, such as finite difference, finite element or finite volume methods, has been proven. The use of finite volume methods is particularly popular in the oil engineering field, since it allows for coupled physical phenomena in the same grids, for which the conservation of various extensive quantities appears to be a main feature. However, it is more challenging to define convergent finite volume schemes for second-order elliptic operators on discretization grids designed for another problem, for which these grids may have been refined or distorted.

For example, in the framework of geological basin simulation, the grids are initially fitted on the geological layers boundaries, which is a first reason for the loss of orthogonality. Then, these grids are modified during the simulation, following the compaction of these layers (see [13]), thus leading to irregular grids, as those proposed by [14]. As a consequence, it is no longer possible to compute the fluxes resulting from a finite volume scheme for a second order operator, by a simple two-point difference across each interface between two neighboring control volumes. Such a two-point scheme is consistent only in the case of an isotropic operator, using a grid such that the lines connecting the centers of the control volumes are orthogonal to the edges of the mesh.

*Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5

†Université de Marne-la-Vallée 5, boulevard Descartes Champs-sur-Marne 77454 Marne-la-Vallée Cedex 2

The problem of finding a consistent expression using only a small number of points, for the finite volume fluxes in the general case of any grid and any anisotropic second order operator, has led to many works (see [1], [2], [3], [13] and references therein; see also [15]). A recent finite volume scheme has been proposed [10, 11], permitting to obtain a convergence property in the case of an anisotropic heterogeneous diffusion problem on unstructured grids, which all the same satisfy the above orthogonality condition. In the case where such an orthogonality condition is not satisfied, a classical method is the mixed finite element method which also gives an approximation of the fluxes and of the gradient of the unknown (see [4], [5], [6], [18] for example, among a very large literature). Unfortunately, the Raviart-Thomas basis is not easily available on control volumes which are not simplices or regular polyhedra (although such a basis can be built on general irregular grids, see [8] and [12], but no easy approximation of these basis functions are known).

We thus propose in this paper an original finite volume method, which can be applicable on any type of grids in any space dimension, with very few restrictions on the shape of the control volumes. The implementation of this scheme is proven to be easy, and no geometric complex shape functions have to be computed. In order to show the mathematical and numerical properties of this scheme, we study here the following problem: find an approximation of $\bar{u}$, weak solution to the following problem:

$$\begin{array}{l} -\mathrm{div}(\Lambda \nabla \bar{u}) = f \text{ in } \Omega, \\ \bar{u} = 0 \text{ on } \partial\Omega, \end{array} \tag{1}$$

under the following assumptions:

$$\Omega \text{ is an open bounded connected polygonal subset of } \mathbb{R}^d,\ d \geq 1, \tag{2}$$

$$\begin{array}{c} \Lambda : \Omega \to \mathcal{M}_d(\mathbb{R}) \text{ is a bounded measurable function such that} \\ \text{there exists } \alpha_0 > 0 \text{ satisfying } \Lambda(x)\xi \cdot \xi \geq \alpha_0 |\xi|^2 \text{ for a.e. } x \in \Omega \text{ and all } \xi \in \mathbb{R}^d, \end{array} \tag{3}$$

and

$$f \in L^2(\Omega). \tag{4}$$

Thanks to Lax-Milgram theorem, there exists a unique weak solution to (1) in the sense that $\bar{u} \in H_0^1(\Omega)$ and the equation is satisfied in the sense of distributions on Ω.

The principle of our scheme, described in Section 2, is the following. We simultaneously look for approximations u_K, $\mathbf{v}_K$ in each control volume K of $\bar{u}$ and $\nabla\bar{u}$, and find an approximation F_σ at each edge σ of the mesh of $\int_\sigma \Lambda(x)\nabla\bar{u}(x) \cdot \mathbf{n}_\sigma \, \mathrm{d}\gamma(x)$, where $\mathbf{n}_\sigma$ is a unit vector normal to σ. The values F_σ must then satisfy the conservation equation in each control volume, and consistency relations are imposed on u_K, $\mathbf{v}_K$ and F_σ. We thus show that these conditions lead to a linear system which, in the general case, has one and only one approximate solution u and $\mathbf{v}$, and, for some particular meshes ("simplicial" meshes), also only one F. We then prove, in this particular case, the convergence of the scheme and an error estimate. We then develop in Section 3 a penalized version of the scheme which can apply on every type of mesh (that was the aim of this work), and which leads to existence and uniqueness properties for $(u, \mathbf{v}, F)$. We provide the mathematical analysis of the convergence of this penalized scheme and give an error estimate. In Section 4, we propose an easy implementation procedure for the penalized scheme, and we use it for the study of some numerical examples. We thus obtain acceptable results on some grids for which it would be complex to use other methods, or to which empirical methods apply but no mathematical results of convergence nor stability have yet been obtained.

2 A first finite volume scheme

2.1 Admissible discretization of Ω

We first present the notion of admissible discretization of the domain Ω, which is necessary to give the expression of the finite volume scheme.

Definition 2.1 [Admissible discretization] *Let Ω be an open bounded polygonal subset of $\mathbb{R}^d$ ($d \geq 1$), and $\partial\Omega = \overline{\Omega} \setminus \Omega$ its boundary. An admissible finite volume discretization of Ω is given by $\mathcal{D} = (\mathcal{M}, \mathcal{E}, \mathcal{P})$, where:*

- *$\mathcal{M}$ is a finite family of non empty open polygonal convex disjoint subsets of Ω (the "control volumes") such that $\overline{\Omega} = \cup_{K\in\mathcal{M}} \overline{K}$.*
- *$\mathcal{E}$ is a finite family of disjoint subsets of $\overline{\Omega}$ (the "edges" of the mesh), such that, for all $\sigma \in \mathcal{E}$, there exists an affine hyperplane E of $\mathbb{R}^d$ and $K \in \mathcal{M}$ with $\sigma \subset \partial K \cap E$ and σ is a non empty open convex subset of E. We assume that, for all $K \in \mathcal{M}$, there exists a subset $\mathcal{E}_K$ of $\mathcal{E}$ such that $\partial K = \cup_{\sigma\in\mathcal{E}_K} \overline{\sigma}$. We also assume that, for all $\sigma \in \mathcal{E}$, either $\sigma \subset \partial\Omega$ or $\overline{\sigma} = \overline{K} \cap \overline{L}$ for some $(K, L) \in \mathcal{M}^2$.*
- *$\mathcal{P}$ is a family of points of Ω indexed by $\mathcal{M}$, denoted by $\mathcal{P} = (\mathbf{x}_K)_{K\in\mathcal{M}}$ and such that, for all $K \in \mathcal{M}$, $\mathbf{x}_K \in K$.*

Some examples of admissible meshes in the sense of the above definition are shown in Figures 1 and 2.

Remark 2.1 *Though the elements of $\mathcal{E}_K$ may not be the real edges of a control volume K (each $\sigma \in \mathcal{E}_K$ may be only a part of a full edge, see figure 2), we will in the following call "edges of K" the elements of $\mathcal{E}_K$.*
Notice that we could also cut each intersection $\overline{K} \cap \overline{L}$ into more than one edge. This would not change our theoretical results but this would lead, for practical implementation, to artificially enlarge the size of the linear systems to solve, which would decrease the efficiency of the scheme.

Remark 2.2 *The whole mathematical study done in this paper applies whatever the choice of the point $\mathbf{x}_K$ in each $K \in \mathcal{M}$. In particular, we do not impose any orthogonality condition connecting the edges and the points $\mathbf{x}_K$. However, the magnitude of the numerical error (and, for some regular or structured types of mesh, its order) does depend on this choice.*
We could also extend our definition to non-planar edges, under some curvature condition. In this case, it remains possible to use the schemes studied in this paper and to prove their convergence.

The following notations are used. The measure of a control volume K is denoted by $\mathrm{m}(K)$; the $(d-1)$-dimensional measure of an edge σ is $\mathrm{m}(\sigma)$. In the case where $\sigma \in \mathcal{E}$ is such that $\overline{\sigma} = \overline{K} \cap \overline{L}$ for $(K, L) \in \mathcal{M}^2$, we denote $\sigma = K|L$. For all $\sigma \in \mathcal{E}$, $\mathbf{x}_\sigma$ is the barycenter of σ.
The set of interior (resp. boundary) edges is denoted by $\mathcal{E}_{\mathrm{int}}$ (resp. $\mathcal{E}_{\mathrm{ext}}$), that is $\mathcal{E}_{\mathrm{int}} = \{\sigma \in \mathcal{E};\ \sigma \not\subset \partial\Omega\}$ (resp. $\mathcal{E}_{\mathrm{ext}} = \{\sigma \in \mathcal{E};\ \sigma \subset \partial\Omega\}$). For all $K \in \mathcal{M}$, we denote by $\mathcal{N}_K$ the subset of $\mathcal{M}$ of the neighboring control volumes (that is, the L such that $\overline{K} \cap \overline{L}$ is an edge of the discretization), and we denote by $\mathcal{E}_{K,\mathrm{ext}} = \mathcal{E}_K \cap \mathcal{E}_{\mathrm{ext}}$.

To study the convergence of the schemes, we need the following two quantities: the size of the discretization

$$\mathrm{size}(\mathcal{D}) = \sup\{\mathrm{diam}(K)\,;\ K \in \mathcal{M}\}$$

and the regularity of the discretization

$$\mathrm{regul}(\mathcal{D}) = \sup\left\{\max\left(\frac{\mathrm{diam}(K)^d}{\rho_K^d}, \mathrm{Card}(\mathcal{E}_K)\right)\ ;\ K \in \mathcal{M}\right\} \tag{5}$$

where, for $K \in \mathcal{M}$, ρ_K is the supremum of the radius of the balls contained in K. Notice that, for all $K \in \mathcal{M}$,

$$\mathrm{diam}(K)^d \leq \mathrm{regul}(\mathcal{D})\rho_K^d \leq \frac{\mathrm{regul}(\mathcal{D})}{\omega_d}\mathrm{m}(K) \tag{6}$$

where ω_d is the volume of the unit ball in $\mathbb{R}^d$. Note also that $\mathrm{regul}(\mathcal{D})$ does not increase in a local refinement procedure.

2.2 The discretization space

Let us assume Assumption (2). Let $\mathcal{D}$ be an admissible discretization in the sense of Definition 2.1. We denote by $H_\mathcal{D}$ the set of functions $\Omega \to \mathbb{R}$ which are piecewise constant on each control volume $K \in \mathcal{M}$. We define the set $L^\nabla_\mathcal{D}$ of all $\mathbf{v} \in H^d_\mathcal{D}$ such that there exists $u \in H_\mathcal{D}$ with

$$\begin{array}{l} \mathbf{v}_K \cdot (\mathbf{x}_\sigma - \mathbf{x}_K) + \mathbf{v}_L \cdot (\mathbf{x}_L - \mathbf{x}_\sigma) = u_L - u_K, \quad \forall K \in \mathcal{M},\ \forall L \in \mathcal{N}_K,\ \text{with } \sigma = K|L, \\ \mathbf{v}_K \cdot (\mathbf{x}_\sigma - \mathbf{x}_K) = -u_K, \quad \forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}_{K,\mathrm{ext}}. \end{array} \tag{7}$$

In the following, we will need some properties on $L^\nabla_\mathcal{D}$. The next one, that we call here "Poincaré's inequality", could also be called an "inf-sup" condition connecting the values of u to that of $\mathbf{v}$, following an analogy of our scheme to a mixed finite element method [16].

Lemma 2.1 [Poincaré's inequality] *Let us assume Assumption (2). Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1, such that* $\mathrm{regul}(\mathcal{D}) \leq \theta$ *for some $\theta > 0$. Then there exists C_1 only depending on d, Ω and θ such that, for all $\mathbf{v} \in L^\nabla_\mathcal{D}$ and $u \in H_\mathcal{D}$ satisfying (7),*

$$\|u\|_{L^2(\Omega)} \leq C_1 \|\mathbf{v}\|_{L^2(\Omega)^d}. \tag{8}$$

As an immediate consequence, for all $\mathbf{v} \in L^\nabla_\mathcal{D}$ there exists one and only one $u \in H_\mathcal{D}$ such that (7) holds; we then define $\psi\ : L^\nabla_\mathcal{D} \to H_\mathcal{D}$ by $\psi(\mathbf{v}) = u$.

Proof.
Let $R > 0$ and $x_0 \in \Omega$ be such that $\Omega \subset B(x_0, R)$ (the open ball of center x_0 and radius R). We extend u by the value 0 in $B(x_0, R) \setminus \Omega$, and we consider $w \in H^1_0(B(x_0, R)) \cap H^2(B(x_0, R))$ such that $-\Delta w(x) = u(x)$, for a.e. $x \in B(x_0, R)$. Denoting $\mathbf{n}_{K,\sigma}$ the unit normal to σ outward to K, we multiply each equation of (7) by $\int_\sigma \nabla w(x) \cdot \mathbf{n}_{K,\sigma}\, \mathrm{d}\gamma(x)$, and we sum on $\sigma \in \mathcal{E}$. Gathering by control volumes, we find

$$\begin{aligned} \sum_{K\in\mathcal{M}} \mathbf{v}_K \cdot \sum_{\sigma\in\mathcal{E}_K} (\mathbf{x}_\sigma - \mathbf{x}_K) \int_\sigma \nabla w(x) \cdot \mathbf{n}_{K,\sigma}\, \mathrm{d}\gamma(x) &= -\sum_{K\in\mathcal{M}} u_K \sum_{\sigma\in\mathcal{E}_K} \int_\sigma \nabla w(x) \cdot \mathbf{n}_{K,\sigma}\, \mathrm{d}\gamma(x) \\ &= -\sum_{K\in\mathcal{M}} u_K \int_K \Delta w(x)\, dx \\ &= \sum_{K\in\mathcal{M}} \mathrm{m}(K) u_K^2 = ||u||^2_{L^2(\Omega)}. \end{aligned}$$

We now compare the left-hand side of this equation, hereafter denoted T_1, with $T_2 = \int_\Omega \mathbf{v}(x) \cdot \nabla w(x)\,\mathrm{d}x$ (note that $|T_2| \le \|\mathbf{v}\|_{L^2(\Omega)^d}\|w\|_{H^1(\Omega)}$). We apply Lemma 5.1 to the vector $\mathbf{G}_K = \frac{1}{\mathrm{m}(K)}\int_K \nabla w(x)\,\mathrm{d}x$ in order to obtain

$$\int_K \nabla w(x)\,dx = \mathrm{m}(K)\mathbf{G}_K = \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}(\sigma)\mathbf{G}_K\cdot\mathbf{n}_{K,\sigma}(\mathbf{x}_\sigma - \mathbf{x}_K),$$

and therefore

$$T_2 = \sum_{K\in\mathcal{M}} \mathbf{v}_K \cdot \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}(\sigma)\mathbf{G}_K\cdot\mathbf{n}_{K,\sigma}(\mathbf{x}_\sigma - \mathbf{x}_K).$$

Hence, setting $\mathbf{G}_{K,\sigma} = \frac{1}{\mathrm{m}(\sigma)}\int_\sigma \nabla w(x)\,\mathrm{d}\gamma(x)$, we get

$$|T_1 - T_2| \le \sum_{K\in\mathcal{M}} |\mathbf{v}_K| \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}(\sigma)\ |\mathbf{G}_K - \mathbf{G}_{K,\sigma}|\ \mathrm{diam}(K).$$

Thanks to the Cauchy-Schwarz inequality, we find

$$(T_1-T_2)^2 \le \left(\sum_{K\in\mathcal{M}} |\mathbf{v}_K|^2 \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}(\sigma)\ \mathrm{diam}(K)\right)\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} \mathrm{m}(\sigma)\ \mathrm{diam}(K)|\mathbf{G}_K - \mathbf{G}_{K,\sigma}|^2\right).$$

We now apply Lemma 5.3, which gives C_2 only depending on d and θ such that

$$(\mathbf{G}_K - \mathbf{G}_{K,\sigma})^2 \le C_2\frac{\mathrm{diam}(K)}{\mathrm{m}(\sigma)}\|w\|^2_{H^2(K)} \tag{9}$$

(notice that $\alpha := \frac{1}{2}\theta^{-1/d} < \mathrm{regul}(\mathcal{D})^{-1/d} \le \rho_K/\mathrm{diam}(K)$ is valid in Lemma 5.3). We also have, for $\sigma\in\mathcal{E}_K$, $\mathrm{m}(\sigma)\le\omega_{d-1}\mathrm{diam}(K)^{d-1}$, where ω_{d-1} is the volume of the unit ball in $\mathbb{R}^{d-1}$. Therefore, according to (6) and since $\mathrm{regul}(\mathcal{D}) \ge \mathrm{card}(\mathcal{E}_K)$ for all $K\in\mathcal{M}$,

$$\begin{aligned}
(T_1-T_2)^2 &\le \left(\sum_{K\in\mathcal{M}} |\mathbf{v}_K|^2 \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}(\sigma)\ \mathrm{diam}(K)\right)\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} C_2\mathrm{diam}(K)^2||w||^2_{H^2(K)}\right)\\
&\le \left(\omega_{d-1}\mathrm{regul}(\mathcal{D})\sum_{K\in\mathcal{M}}|\mathbf{v}_K|^2\mathrm{diam}(K)^d\right)\left(C_2\mathrm{size}(\mathcal{D})^2\mathrm{regul}(\mathcal{D})\|w\|^2_{H^2(\Omega)}\right)\\
&\le \frac{\omega_{d-1}\mathrm{regul}(\mathcal{D})^2}{\omega_d}||\mathbf{v}||^2_{L^2(\Omega)^d}C_2\mathrm{diam}(\Omega)^2\mathrm{regul}(\mathcal{D})\|w\|^2_{H^2(\Omega)}.
\end{aligned}$$

We can now conclude, writing

$$\begin{aligned}
||u||^2_{L^2(\Omega)} = T_1 &\le |T_1 - T_2| + |T_2|\\
&\le \sqrt{\frac{\omega_{d-1}C_2\theta^3}{\omega_d}}\mathrm{diam}(\Omega)\|\mathbf{v}\|_{L^2(\Omega)}||w||_{H^2(\Omega)} + ||\mathbf{v}||_{L^2(\Omega)^d}||w||_{H^1(\Omega)}.
\end{aligned}$$

Since there exists C_3 only depending on d and $B(x_0,R)$ (the ball chosen at the beginning of the proof) such that $||w||_{H^2(\Omega)} \le C_3||u||_{L^2(\Omega)}$, this concludes the proof. □

Lemma 2.2 [Equicontinuity of the translations] *Let us assume Assumption (2). Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1, such that* $\mathrm{regul}(\mathcal{D}) \le \theta$ *for some $\theta > 0$. Then there exists C_4 only depending on d, Ω and θ such that, for all $\mathbf{v} \in L^\nabla_\mathcal{D}$ and $u \in H_\mathcal{D}$ satisfying (7) (which means that $u = \psi(\mathbf{v})$, see Lemma 2.1), for all $\xi \in \mathbb{R}^d$,*

$$\|u(\cdot + \xi) - u\|_{L^1(\mathbb{R}^d)} \le C_4 \|\mathbf{v}\|_{L^1(\Omega)^d} |\xi| \tag{10}$$

(u has been extended by 0 outside Ω).

Proof.
Let $\mathbf{v} \in L^\nabla_\mathcal{D}$ and $u \in H_\mathcal{D}$ (extended by 0 outside Ω) such that (7) holds. For all $\sigma \in \mathcal{E}$, let us define $D_\sigma u = |u_L - u_K|$ if $\sigma = K|L$ and $D_\sigma u = |u_K|$ if $\sigma \in \mathcal{E}_{K,\mathrm{ext}}$. For $(x, \xi) \in \mathbb{R}^d \times \mathbb{R}^d$ and $\sigma \in \mathcal{E}$, we define $\chi(x, \xi, \sigma)$ by 1 if $\sigma \cap [x, x + \xi] \neq \emptyset$ and by 0 otherwise. We then have, for all $\xi \in \mathbb{R}^d$ and a.e. $x \in \mathbb{R}^d$ (the x's such that x and $x + \xi$ do not belong to $\cup_{K\in\mathcal{M}} \partial K$, and $[x, x + \xi]$ does not intersect the relative boundary of any edge),

$$|u(x + \xi) - u(x)| \le \sum_{\sigma\in\mathcal{E}} \chi(x, \xi, \sigma) D_\sigma u.$$

Applying (7), we get

$$|u(x + \xi) - u(x)| \le \left(\begin{array}{l} \displaystyle\sum_{\sigma\in\mathcal{E}_{\mathrm{int}}, \sigma=K|L} \chi(x, \xi, \sigma)(|\mathbf{v}_K||\mathbf{x}_\sigma - \mathbf{x}_K| + |\mathbf{v}_L||\mathbf{x}_L - \mathbf{x}_\sigma|) \\ + \displaystyle\sum_{\sigma\in\mathcal{E}_{\mathrm{ext}}, \sigma\in\mathcal{E}_K} \chi(x, \xi, \sigma)|\mathbf{v}_K||\mathbf{x}_\sigma - \mathbf{x}_K| \end{array} \right)$$

and, gathering by control volumes,

$$|u(x + \xi) - u(x)| \le \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \chi(x, \xi, \sigma) \mathrm{diam}(K) |\mathbf{v}_K|. \tag{11}$$

In order that $\chi(x, \xi, \sigma) \neq 0$, x must lie in the set $\sigma - [0, 1]\xi$ which has measure $\mathrm{m}(\sigma)|\mathbf{n}_\sigma \cdot \xi|$ (where $\mathbf{n}_\sigma$ is a unit normal to σ). Hence,

$$\sum_{\sigma\in\mathcal{E}_K} \int_{\mathbb{R}^d} \chi(x, \xi, \sigma)\, dx \le \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}(\sigma)|\mathbf{n}_\sigma \cdot \xi| \le \mathrm{Card}(\mathcal{E}_K)\omega_{d-1}\mathrm{diam}(K)^{d-1}|\xi|.$$

Thus, integrating (11) on $\mathbb{R}^d$, we find

$$\|u(\cdot + \xi) - u\|_{L^1(\Omega)} \le \omega_{d-1}\mathrm{regul}(\mathcal{D})|\xi| \sum_{K\in\mathcal{M}} \mathrm{diam}(K)^d |\mathbf{v}_K|$$

and we conclude by (6). □

Remark 2.3 *We could prove the property $\|u(\cdot+\xi)-u\|^2_{L^2(\Omega)} \le C\|\mathbf{v}\|^2_{L^2(\Omega)^d}\ |\xi|\ (|\xi|+\mathrm{size}(\mathcal{D}))$, in a similar way as in [9], by introducing the maximum value of* $\mathrm{diam}(K)/\rho_L$, *for all $(K, L) \in \mathcal{M}^2$, in the definition of* $\mathrm{regul}(\mathcal{D})$. *This would allow, in Theorem 2.1, to prove the strong convergence of u_m in $L^2(\Omega)$. Nevertheless, we chose not to do so since the quantity* $\mathrm{diam}(K)/\rho_L$ *cannot remain bounded in a local mesh refinement procedure.*
Note that we shall all the same prove, in a particular case, a strong convergence property in $L^2(\Omega)$ for u, as a consequence of the error estimate.

Lemma 2.3 [Compactness property] *Let us assume Assumption (2). Let $(\mathcal{D}_m)_{m\geq 1}$ be admissible discretizations of Ω in the sense of Definition 2.1, such that* $\mathrm{size}(\mathcal{D}_m) \to 0$ *as* $m \to \infty$ *and* $(\mathrm{regul}(\mathcal{D}_m))_{m\geq 1}$ *is bounded. Let $(\mathbf{v}_m)_{m\geq 1}$ such that, for all $m \geq 1$, $\mathbf{v}_m \in L^{\nabla}_{\mathcal{D}_m}$ and such that $(\mathbf{v}_m)_{m\geq 1}$ is bounded in $L^2(\Omega)^d$.*
Then there exists a subsequence of $(\mathcal{D}_m)_{m\geq 1}$ (still denoted by $(\mathcal{D}_m)_{m\geq 1}$) and $\bar{u} \in H^1_0(\Omega)$ such that the corresponding sequence $(\psi(\mathbf{v}_m))_{m\geq 1}$ converges to $\bar{u}$ weakly in $L^2(\Omega)$ and strongly in $L^q(\Omega)$ for all $q < 2$, and such that $(\mathbf{v}_m)_{m\geq 1}$ converges to $\nabla\bar{u}$ weakly in $L^2(\Omega)^d$.

Proof.
Since $(\mathbf{v}_m)_{m\geq 1}$ is bounded in $L^2(\Omega)^d$, we get the existence of a subsequence of $(\mathcal{D}_m)_{m\geq 1}$, again denoted by $(\mathcal{D}_m)_{m\geq 1}$, and $\bar{\mathbf{v}} \in L^2(\Omega)^d$ such that $(\mathbf{v}_m)_{m\geq 1}$ weakly converges to $\bar{\mathbf{v}}$ in $L^2(\Omega)^d$. Thanks to Lemmas 2.1 and 2.2, we can apply Kolmogorov's theorem on the family $(\psi(\mathbf{v}_m))_{m\geq 1}$: there exists $\bar{u} \in L^2(\Omega)$ and a subsequence of $(\mathcal{D}_m)_{m\geq 1}$, again denoted by $(\mathcal{D}_m)_{m\geq 1}$, such that $(\psi(\mathbf{v}_m))_{m\geq 1}$ converges to $\bar{u}$ weakly in $L^2(\Omega)$ and strongly in $L^1(\Omega)$ (this implies in particular the strong convergence in $L^q(\Omega)$ for all $q < 2$).
We extend $\psi(\mathbf{v}_m)$, $\bar{u}$, $\mathbf{v}_m$ and $\bar{\mathbf{v}}$ by 0 outside Ω and we now prove that $\bar{\mathbf{v}} = \nabla\bar{u}$ in the distributional sense on $\mathbb{R}^d$. This will conclude that $\bar{u} \in H^1(\mathbb{R}^d)$ and, since $\bar{u} = 0$ outside Ω, that $\bar{u} \in H^1_0(\Omega)$.
Let $\mathbf{e} \in \mathbb{R}^d$ and $\varphi \in C^\infty_c(\mathbb{R}^d)$. For simplicity, we drop the index m for $\mathcal{D}_m$, $\mathbf{v}_m$ and $u = \psi(\mathbf{v}_m)$. We multiply each equation of (7) by $\int_\sigma \varphi(x)\,\mathrm{d}\gamma(x)\mathbf{e}\cdot\mathbf{n}_{K,\sigma}$. We sum all these equations and we gather by control volumes, getting $T_3 = T_4$ with

$$T_3 = \sum_{K\in\mathcal{M}} \mathbf{v}_K \cdot \sum_{\sigma\in\mathcal{E}_K} \int_\sigma \varphi(x)\,\mathrm{d}\gamma(x)\ \mathbf{e}\cdot\mathbf{n}_{K,\sigma}(\mathbf{x}_\sigma - \mathbf{x}_K)$$

and

$$T_4 = -\sum_{K\in\mathcal{M}} u_K \sum_{\sigma\in\mathcal{E}_K} \int_\sigma \varphi(x)\,\mathrm{d}\gamma(x)\ \mathbf{e}\cdot\mathbf{n}_{K,\sigma} = -\int_\Omega u(x)\mathrm{div}(\varphi(x)\mathbf{e})\,\mathrm{d}x.$$

We have

$$\lim_{\mathrm{size}(\mathcal{D})\to 0} T_4 = -\int_\Omega \bar{u}(x)\mathrm{div}(\varphi(x)\mathbf{e})\,\mathrm{d}x = -\int_{\mathbb{R}^d} \bar{u}(x)\mathrm{div}(\varphi(x)\mathbf{e})\,\mathrm{d}x$$

(recall that $\bar{u}$ has been extended by 0 outside Ω). We now want to compare T_3 with T_5 defined by

$$T_5 = \sum_{K\in\mathcal{M}} \mathbf{v}_K \cdot \sum_{\sigma\in\mathcal{E}_K} \frac{1}{\mathrm{m}(K)} \int_K \varphi(x)\,\mathrm{d}x\ \mathrm{m}(\sigma)\ \mathbf{e}\cdot\mathbf{n}_{K,\sigma}(\mathbf{x}_\sigma - \mathbf{x}_K).$$

Since there exists C_5 only depending on φ such that

$$\left|\frac{1}{\mathrm{m}(\sigma)}\int_\sigma \varphi(x)\,\mathrm{d}\gamma(x) - \frac{1}{\mathrm{m}(K)}\int_K \varphi(x)\,\mathrm{d}x\right| \leq C_5\mathrm{size}(\mathcal{D}),$$

we get that

$$|T_3 - T_5| \leq C_5\mathrm{size}(\mathcal{D}) \sum_{K\in\mathcal{M}} |\mathbf{v}_K| \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}(\sigma)|\mathbf{x}_\sigma - \mathbf{x}_K|.$$

But $\mathrm{m}(\sigma)|\mathbf{x}_\sigma - \mathbf{x}_K| \leq \omega_{d-1}\mathrm{diam}(K)^d \leq \frac{\omega_{d-1}\mathrm{regul}(\mathcal{D})}{\omega_d}\mathrm{m}(K)$ and, since $\mathrm{card}(\mathcal{E}_K) \leq \mathrm{regul}(\mathcal{D})$, we obtain

$$|T_3 - T_5| \leq C_5\mathrm{size}(\mathcal{D})\frac{\omega_{d-1}\mathrm{regul}(\mathcal{D})^2}{\omega_d}\|\mathbf{v}\|_{L^1(\Omega)}$$

and thus

$$\lim_{\text{size}(\mathcal{D})\to 0} (T_3 - T_5) = 0.$$

Moreover, thanks to Lemma 5.1, we get $T_5 = \int_\Omega \varphi(x)\mathbf{v}(x)\cdot\mathbf{e}\,\mathrm{d}x$ and so $\lim_{\text{size}(\mathcal{D})\to 0} T_5 = \int_\Omega \varphi(x)\bar{\mathbf{v}}(x)\cdot\mathbf{e}\,\mathrm{d}x = \int_{\mathbb{R}^d} \varphi(x)\bar{\mathbf{v}}(x)\cdot\mathbf{e}\,\mathrm{d}x$ ($\bar{\mathbf{v}}$ has been extended by 0 outside Ω). This proves that

$$\int_{\mathbb{R}^d} \varphi(x)\bar{\mathbf{v}}(x)\cdot\mathbf{e}\,\mathrm{d}x = -\int_{\mathbb{R}^d} \bar{u}(x)\mathrm{div}(\varphi(x)\mathbf{e})\,\mathrm{d}x,$$

which completes the proof of the lemma. □

2.3 The scheme

Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1. We consider here the following scheme: finding $(\mathbf{v}, u) \in H_\mathcal{D}^d \times H_\mathcal{D}$ satisfying (7) and a family of real numbers $F = (F_{K,\sigma})_{K\in\mathcal{M},\,\sigma\in\mathcal{E}_K}$ such that

$$F_{K,\sigma} + F_{L,\sigma} = 0, \quad \forall \sigma = K|L \in \mathcal{E}_{\text{int}}, \tag{12}$$

$$m_K \Lambda_K \mathbf{v}_K = \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(\mathbf{x}_\sigma - \mathbf{x}_K), \quad \forall K \in \mathcal{M} \tag{13}$$

(where $\Lambda_K = \frac{1}{m_K}\int_K \Lambda(x)\,\mathrm{d}x$) and

$$-\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} = \int_K f(x)\,\mathrm{d}x, \quad \forall K \in \mathcal{M}. \tag{14}$$

This scheme can be understood the following way: thanks to Lemma 5.1, we expect from (13) that $F_{K,\sigma} \approx \mathrm{m}(\sigma)(\Lambda_K \mathbf{v}_K)\cdot\mathbf{n}_{K,\sigma}$ (i.e. that $F_{K,\sigma}$ is the flux of $\Lambda\mathbf{v}$ through σ). This explains the conservativity imposed in (12). In this setting, (14) is simply the integration on the control volumes of $-\mathrm{div}(\Lambda\mathbf{v}) = f$ (an equation which is to be expected if we think of $\mathbf{v}$ as some approximation of the gradient of the solution to (1)). The equation (7) gives then a discrete u whose discrete gradient is $\mathbf{v}$, and thus u itself stands for an approximation of the solution to (1).

We present in the sequel three types of mathematical analysis. The first one is devoted to the proof of the convergence of this scheme on particular meshes, and to an error estimate. The second one is the proof that, for general meshes, one can ensure the existence and the uniqueness of $u, \mathbf{v}$ and the existence but not the uniqueness of F. For this reason, we develop in a next part the study of a slightly modified version of the scheme, which applies on general meshes and whose additional advantage is to allow an easy implementation.

2.4 Simplicial meshes: convergence of the scheme

A control volume K is a simplex (or is simplicial) if it is the interior of the convex hull of $d+1$ points of $\mathbb{R}^d$ such that no affine hyperplane of $\mathbb{R}^d$ contains all of them, and if the condition $\mathrm{Card}(\mathcal{E}_K) = d+1$ holds. We then call "simplicial meshes" the meshes whose all control volumes are simplicial. We prove here that, for simplicial meshes, there is a unique solution to ((7),(12),(13),(14)) and that this solution converges, as $\text{size}(\mathcal{D}) \to 0$ and $\text{regul}(\mathcal{D})$ stays bounded, to the weak solution to (1).

Let us first state an estimate on the fluxes, which holds under conditions (13)-(14).

Lemma 2.4 *Let us assume Assumptions (2)-(4). Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of definition 2.1, such that* $\mathrm{regul}(\mathcal{D}) \leq \theta$ *for some $\theta > 0$ and $\mathcal{M}$ is a simplicial mesh. Let $\mathbf{v} \in H_{\mathcal{D}}^d$ and a family of real numbers $F = (F_{K,\sigma})_{K\in\mathcal{M},\,\sigma\in\mathcal{E}_K}$ be given such that (13) and (14) hold. Then there exists C_6 only depending on d, Ω, Λ and θ such that*

$$\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} \mathrm{diam}(K)^{2-d} F_{K,\sigma}^2 \leq C_6(||f||_{L^2(\Omega)}^2 + ||\mathbf{v}||_{L^2(\Omega)^d}^2). \tag{15}$$

Proof. For $K \in \mathcal{M}$, let A_K be the $(d+1)\times(d+1)$ matrix whose columns are $(1, \mathbf{x}_\sigma - \mathbf{x}_K)_{\sigma\in\mathcal{E}_K}^T$ (since K is simplicial, it has $d+1$ edges and A_K is indeed a square matrix). The equations (13)-(14) can be written $A_K F_K = E_K$, where $F_K = (F_{K,\sigma})_{\sigma\in\mathcal{E}_K}$ and

$$E_K = \begin{pmatrix} -\int_K f(x)\,\mathrm{d}x \\ \mathrm{m}(K)\Lambda_K \mathbf{v}_K \end{pmatrix}.$$

We now want to estimate $||A_K^{-1}||$ and, in order to achieve this, we divide the rest of the proof in several steps.

Step 1: this step is devoted to allow the assumption $\mathrm{diam}(K) = 1$ in Steps 2 and 3.
Let $K_0 = \mathrm{diam}(K)^{-1}K$. Then $\mathbf{x}_{K,0} = \mathrm{diam}(K)^{-1}\mathbf{x}_K \in K_0$ and the barycenters of the edges of K_0 are $\mathbf{x}_{\sigma,0} = \mathrm{diam}(K)^{-1}\mathbf{x}_\sigma$. Notice also that, if $\rho_{K,0}$ is the supremum of the radius of the balls included in K_0, then

$$\frac{1}{\rho_{K,0}} = \frac{\mathrm{diam}(K_0)}{\rho_{K,0}} = \frac{\mathrm{diam}(K)}{\rho_K} \leq \mathrm{regul}(\mathcal{D})^{1/d} \leq \theta^{1/d}. \tag{16}$$

Let $A_{K,0}$ be the $(d+1)\times(d+1)$ matrix corresponding to K_0, that is to say whose columns are $(1, \mathbf{x}_{\sigma,0} - \mathbf{x}_{K,0})_{\sigma\in\mathcal{E}_K}^T = (1, \mathrm{diam}(K)^{-1}(\mathbf{x}_\sigma - \mathbf{x}_K))_{\sigma\in\mathcal{E}_K}^T$. Since

$$A_K = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \mathrm{diam}(K) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathrm{diam}(K) \end{pmatrix} A_{K,0},$$

we have $||A_K^{-1}|| \leq \sup(1, \mathrm{diam}(K)^{-1})||A_{K,0}^{-1}||$. Hence, an estimate on $||A_{K,0}^{-1}||$ gives an estimate on $||A_K^{-1}||$.

Step 2: estimate on $A_{K,0}$.
By (16), K_0 contains a closed ball of radius $\frac{1}{2}\theta^{-1/d}$. Up to a translation (which does not change the vectors $\mathbf{x}_{\sigma,0} - \mathbf{x}_{K,0}$, and hence does not change $A_{K,0}$), we can assume that this ball is centered at 0. Since $\mathrm{diam}(K_0) = 1$, we have then $\overline{B}(0, \frac{1}{2}\theta^{-1/d}) \subset K_0 \subset \overline{B}(0,1)$.
Let Z_θ be the set of couples $(L, \mathbf{x}_L)$, where L is a simplex such that $\overline{B}(0, \frac{1}{2}\theta^{-1/d}) \subset \overline{L} \subset \overline{B}(0,1)$ and $x_L \in \overline{L}$. Each simplex is defined by $d+1$ vertices in $\mathbb{R}^d$ so Z_θ can be considered as a subset of $P = (\mathbb{R}^d)^{d+1}/S_{d+1} \times \mathbb{R}^d$, where S_{d+1} is the symmetric group acting on $(\mathbb{R}^d)^{d+1}$ by permuting the vectors. As such, Z_θ is compact in P: it is straightforward if we express the condition "the adherence of a simplex contains $\overline{B}(0, \frac{1}{2}\theta^{-1/d})$" as "any point of $\overline{B}(0, \frac{1}{2}\theta^{-1/d})$ is a convex combination of the vertices of the simplex", which is a closed condition with respect to the vertices of the simplex.

For $(L, \mathbf{x}_L) \in Z_\theta$, let $M(L, \mathbf{x}_L)$ be the set of $(d+1) \times (d+1)$ matrices whose columns are, up to permutations, $(1, \mathbf{x}_\sigma - \mathbf{x}_L)^T_{\sigma \in \mathcal{E}_L}$ ($\mathcal{E}_L$ being the set of edges of L and $\mathbf{x}_\sigma$ being the barycenter of σ). $M(L, \mathbf{x}_L)$ can be considered as an element of $\mathcal{M}_{d+1}(\mathbb{R})/S_{d+1}$ (S_{d+1} acting by permuting the columns) and the application $(L, \mathbf{x}_L) \in Z_\theta \to M(L, \mathbf{x}_L) \in \mathcal{M}_{d+1}(\mathbb{R})/S_{d+1}$ is continuous: to see this, just recall that the barycenter of an edge $\sigma \in \mathcal{E}_L$ is $\mathbf{x}_\sigma = \frac{1}{d}\sum_{i=1}^d \mathbf{x}_i$, where $\mathbf{x}_i$ are the vertices of σ (i.e. all vertices but one of L).
If $(L, \mathbf{x}_L) \in Z_\theta$, all the matrices of $M(L, \mathbf{x}_L)$ are invertible. Indeed, assume that such a matrix has a non-trivial element $(\lambda_1, \dots, \lambda_{d+1})$ in its kernel; this leads (denoting $(\sigma_1, \dots, \sigma_{d+1})$ the edges of L) to $\sum_{i=1}^{d+1} \lambda_i = 0$ and $\sum_{i=1}^{d+1} \lambda_i(\mathbf{x}_{\sigma_i} - \mathbf{x}_L) = \sum_{i=1}^{d+1} \lambda_i \mathbf{x}_{\sigma_i} = 0$. Assuming $\lambda_{d+1} \neq 0$, we then can write $\mathbf{x}_{\sigma_{d+1}} = \sum_{i=1}^d \mu_i \mathbf{x}_{\sigma_i}$ with $\sum_{i=1}^d \mu_i = 1$ (since $\mu_i = -\lambda_i/\lambda_{d+1}$). This means that $\mathbf{x}_{\sigma_{d+1}}$ is in the affine hyperplane $\mathcal{H}$ generated by the other barycenters of edges. Note that $\mathcal{H}$ is parallel to σ_{d+1} (this is a straightforward consequence of Thales' theorem at the vertex which does not belong to σ_{d+1}, and of the fact that the barycenters $(\mathbf{x}_{\sigma_1}, \dots, \mathbf{x}_{\sigma_d})$ of the edges are in fact the barycenters of the vertices of the corresponding edge). Therefore $\mathcal{H}$ contains the whole edge σ_{d+1}, because it contains $\mathbf{x}_{\sigma_{d+1}} \in \sigma_{d+1}$. Let $\mathbf{a}$ be the vertex of L which does not belong to σ_{d+1}; $\mathbf{a}$ belongs to σ_1 and we denote $(\mathbf{b}_1, \dots, \mathbf{b}_{d-1})$ the other vertices of σ_1 (which also belong to σ_{d+1}). We have $\mathbf{x}_{\sigma_1} = \frac{1}{d}(\mathbf{a} + \sum_{i=1}^{d-1} \mathbf{b}_i)$, and therefore $\mathbf{a} = d\mathbf{x}_{\sigma_1} - \sum_{i=1}^{d-1} \mathbf{b}_i$; but $d - \sum_{i=1}^{d-1} 1 = 1$ and thus $\mathbf{a}$ belongs to the affine hyperplane generated by $(\mathbf{x}_{\sigma_1}, \mathbf{b}_1, \dots, \mathbf{b}_{d-1})$. Since all these points belong to $\mathcal{H}$, we have $\mathbf{a} \in \mathcal{H}$ and, since $\sigma_{d+1} \subset \mathcal{H}$, all the vertices of L in fact belong to $\mathcal{H}$; L is thus contained in an hyperplane, which is a contradiction with the fact that it contains a non-trivial ball. Thus, for $(L, \mathbf{x}_L) \in Z_\theta$, $M(L, \mathbf{x}_L)$ is in fact an element of $Gl_{d+1}(\mathbb{R})/S_{d+1}$.
The inversion $\mathrm{inv} : Gl_{d+1}(\mathbb{R}) \to Gl_{d+1}(\mathbb{R})$ is continuous; hence, $||\mathrm{inv}(\cdot)|| : Gl_{d+1}(\mathbb{R}) \to \mathbb{R}$ is also continuous. Permuting the columns of a matrix comes down to permuting the lines of its inverse, which does not change the norm; therefore $||\mathrm{inv}(\cdot)|| : Gl_{d+1}(\mathbb{R})/S_{d+1} \to \mathbb{R}$ is well defined and also continuous.
We can now conclude this step. The application $Z_\theta \to Gl_{d+1}(\mathbb{R})/S_{d+1} \to \mathbb{R}$ defined by $(L, \mathbf{x}_L) \to M(L, \mathbf{x}_L) \to ||\mathrm{inv}(M(L, \mathbf{x}_L))||$ is continuous. Since Z_θ is compact, this application is bounded by some C_7 only depending on d and θ. As $(K_0, \mathbf{x}_{K,0}) \in Z_\theta$, this shows that $||A_{K,0}^{-1}|| \leq C_7$.

Step 3: conclusion.
Using the preceding steps, we find $||F_K|| \leq ||A_K^{-1}||\,||E_K|| \leq C_7 \sup(1, \mathrm{diam}(K)^{-1})||E_K||$. Hence,

$$\sum_{K \in \mathcal{M}} \mathrm{diam}(K)^{2-d}||F_K||^2 \leq C_7^2 \sup(\mathrm{diam}(\Omega)^2, 1) \sum_{K \in \mathcal{M}} \mathrm{diam}(K)^{-d}||E_K||^2.$$

But $||E_K||^2 \leq \mathrm{m}(K) \int_K |f(x)|^2 \,\mathrm{d}x + C_8 \mathrm{m}(K)^2 \mathbf{v}_K^2$ with C_8 only depending on Λ. Since $\mathrm{m}(K) \leq \omega_d \mathrm{diam}(K)^d$, this concludes the proof of (15). □

We can now state the existence and estimate on the solution of the scheme.

Lemma 2.5 *Let us assume Assumptions (2)-(4). Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of definition 2.1, such that $\mathcal{M}$ is a simplicial mesh. Then there exists a unique $(\mathbf{v}, u, F)$ such that ((7),(12),(13),(14)) hold. Moreover, for all $\theta \geq \mathrm{regul}(\mathcal{D})$, there exists C_9 only depending on d, Ω, Λ and θ such that*

$$||\mathbf{v}||^2_{L^2(\Omega)^d} \leq C_9 ||f||^2_{L^2(\Omega)} \tag{17}$$

and

$$\sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} \mathrm{diam}(K)^{2-d} F_{K,\sigma}^2 \leq C_9 ||f||^2_{L^2(\Omega)}. \tag{18}$$

PROOF.
We first notice that, since ((7),(12),(13),(14)) is square linear in $(\mathbf{v}, u, F)$, it suffices to prove that any solution satisfies the estimate in order to obtain the existence and uniqueness of the solution (because $f = 0$ then implies $F = 0$ and $\mathbf{v} = 0$ which, in turn, gives $u = 0$ thanks to (8)).

Multiplying (14) by u_K, summing on the control volumes and gathering by edges, we have, thanks to (12),

$$\sum_{\sigma\in\mathcal{E}_{\rm int},\,\sigma=K|L} F_{K,\sigma}(u_L - u_K) + \sum_{\sigma\in\mathcal{E}_{\rm ext},\,\sigma\in\mathcal{E}_K} -F_{K,\sigma}u_K = \int_\Omega u(x)f(x)\,\mathrm{d}x.$$

Using (7) and gathering by control volumes, this gives

$$\begin{aligned}\int_\Omega u(x)f(x)\,\mathrm{d}x &= \sum_{\sigma\in\mathcal{E}_{\rm int},\,\sigma=K|L} F_{K,\sigma}\mathbf{v}_K\cdot(\mathbf{x}_\sigma-\mathbf{x}_K) + F_{L,\sigma}\mathbf{v}_L\cdot(\mathbf{x}_\sigma-\mathbf{x}_L)\\ &\quad+\sum_{\sigma\in\mathcal{E}_{\rm ext},\,\sigma\in\mathcal{E}_K} F_{K,\sigma}\mathbf{v}_K\cdot(\mathbf{x}_\sigma-\mathbf{x}_K)\\ &= \sum_{K\in\mathcal{M}}\mathbf{v}_K\cdot\sum_{\sigma\in\mathcal{E}_K}F_{K,\sigma}(\mathbf{x}_\sigma-\mathbf{x}_K).\end{aligned}$$

Thanks to (13), we then deduce

$$\int_\Omega f(x)u(x)\,\mathrm{d}x = \sum_{K\in\mathcal{M}}\mathrm{m}(K)\Lambda_K\mathbf{v}_K\cdot\mathbf{v}_K = \int_\Omega \Lambda(x)\mathbf{v}(x)\cdot\mathbf{v}(x)\,dx. \tag{19}$$

In particular, by property of Λ, $\alpha_0||\mathbf{v}||^2_{L^2(\Omega)^d} \le ||f||_{L^2(\Omega)}||u||_{L^2(\Omega)}$ and the estimate (17) follows from (8). We then deduce (18), applying Lemma 2.4. □

We can now prove the convergence result.

Theorem 2.1 *Let us assume Assumptions (2)-(4). Let $(\mathcal{D}_m)_{m\ge1}$ be admissible discretization of Ω in the sense of Definition 2.1, such that* $\mathrm{size}(\mathcal{D}_m)\to 0$ *as $m\to\infty$,* $(\mathrm{regul}(\mathcal{D}_m))_{m\ge1}$ *is bounded and $\mathcal{M}_m$ is a simplicial mesh. Let $(\mathbf{v}_m, u_m, F_m)$ be the solution to ((7),(12),(13),(14)) for the discretization $\mathcal{D}_m$. Let $\bar u$ be the weak solution to (1).*
Then, as $m\to\infty$, $\mathbf{v}_m\to\nabla\bar u$ strongly in $L^2(\Omega)^d$ and $u_m\to\bar u$ weakly in $L^2(\Omega)$ and strongly in $L^q(\Omega)$ for all $q<2$.

PROOF.
The assumptions of Lemma 2.3 are satisfied and there exists thus $\bar u\in H^1_0(\Omega)$ such that, up to a subsequence, $\mathbf{v}_m\to\nabla\bar u$ weakly in $L^2(\Omega)^d$ and $u_m\to\bar u$ weakly in $L^2(\Omega)$ and strongly in $L^q(\Omega)$ for all $q<2$.
We now prove that the limit function $\bar u$ is the weak solution to (1). Since any subsequence of $(\mathbf{v}_m, u_m)$ has a subsequence which converges as above, and since the reasoning we are going to make proves that any such limit of a subsequence is the (unique) weak solution to (1), this will

conclude the proof, except for the strong convergence of $\mathbf{v}_m$. In order to simplify the notations, we drop the index m as in the proof of Lemma 2.3.

Let $\varphi \in C_c^\infty(\Omega)$. We multiply (14) by $\varphi(\mathbf{x}_K)$ and we sum on K. Gathering by edges thanks to (12), we get

$$\sum_{\sigma\in\mathcal{E}_{\rm int},\sigma=K|L} F_{K,\sigma}(\varphi(\mathbf{x}_L)-\varphi(\mathbf{x}_K)) = \sum_{K\in\mathcal{M}}\int_K \varphi(\mathbf{x}_K)f(x)\,\mathrm{d}x$$

as long as size($\mathcal{D}$) is small enough (so that $\varphi = 0$ on the control volumes K such that $\partial K \cap \partial\Omega \neq \emptyset$). We set, for $\sigma = K|L$,

$$\varphi(\mathbf{x}_L)-\varphi(\mathbf{x}_K) = \frac{1}{\mathrm{m}(K)}\int_K \nabla\varphi(x)\,\mathrm{d}x\cdot(\mathbf{x}_\sigma-\mathbf{x}_K) + \frac{1}{\mathrm{m}(L)}\int_L \nabla\varphi(x)\,\mathrm{d}x\cdot(\mathbf{x}_L-\mathbf{x}_\sigma) + R_{KL}$$

and we have $|R_{KL}| \le C_\varphi(\mathrm{diam}(K)^2+\mathrm{diam}(L)^2)$. We thus obtain, gathering by control volumes and using (13) (and the fact that $\varphi = 0$ on the control volumes on the boundary of Ω),

$$\int_\Omega \Lambda_\mathcal{D}\mathbf{v}(x)\cdot\nabla\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi_\mathcal{D}(x)\,\mathrm{d}x + T_6, \tag{20}$$

where $\Lambda_\mathcal{D}$ and $\varphi_\mathcal{D}$ are constant respectively equal to Λ_K and $\varphi(\mathbf{x}_K)$ on each mesh K, and

$$|T_6| \le C_\varphi \sum_{\sigma\in\mathcal{E}_{\rm int},\sigma=K|L} |F_{K,\sigma}|(\mathrm{diam}(K)^2+\mathrm{diam}(L)^2) = C_\varphi \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^2|F_{K,\sigma}|. \tag{21}$$

We can write, using Cauchy-Schwarz inequality,

$$\begin{aligned}
|T_6|^2 &\le C_\varphi^2\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{2-d}F_{K,\sigma}^2\right)\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^4\mathrm{diam}(K)^{d-2}\right)\\
&\le C_{10}\mathrm{size}(\mathcal{D})^2\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^d\\
&\le C_{10}\mathrm{size}(\mathcal{D})^2\mathrm{regul}(\mathcal{D})\frac{\mathrm{regul}(\mathcal{D})}{\omega_d}\mathrm{m}(\Omega)
\end{aligned}$$

where, according to Lemma 2.5, C_{10} does not depend on the mesh since regul($\mathcal{D}$) stays bounded (we have also used (6)). Hence, $T_6 \to 0$ as size($\mathcal{D}$) $\to 0$ and we can pass to the limit in (20) to conclude that $\bar{u}$ is a weak solution to (1).

It remains to prove that the convergence of $\mathbf{v}$ is strong. We use (19): this equality implies, since $u \to \bar{u}$ weakly in $L^2(\Omega)$ and $\bar{u}$ is a weak solution to (1), that $\int_\Omega \Lambda(x)\mathbf{v}(x)\cdot\mathbf{v}(x)\,dx \to \int_\Omega f(x)\bar{u}(x)\,dx = \int_\Omega \Lambda(x)\nabla\bar{u}(x)\cdot\nabla\bar{u}(x)\,\mathrm{d}x$ as size($\mathcal{D}$) $\to 0$. But $N(\mathbf{w})^2 = \int_\Omega \Lambda(x)\mathbf{w}(x)\cdot\mathbf{w}(x)\,\mathrm{d}x$ is a norm on $L^2(\Omega)^d$, coming from a scalar product (defined by $(\Lambda+\Lambda^T)/2$) and equivalent to the usual norm. Since $\mathbf{v} \to \nabla\bar{u}$ weakly in $L^2(\Omega)^d$ and $N(\mathbf{v}) \to N(\nabla\bar{u})$ as size($\mathcal{D}$) $\to 0$, this proves that, in fact, $\mathbf{v} \to \nabla\bar{u}$ strongly in $L^2(\Omega)^d$. □

In the case where the solution to (1) is regular, we can also derive an error estimate.

Theorem 2.2 *Let us assume Assumptions (2)-(4). Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1, such that* regul($\mathcal{D}$) $\le \theta$ *for some $\theta > 0$ and $\mathcal{M}$ is a simplicial*

mesh. Let $(\mathbf{v}, u, F)$ *be the solution to ((7),(12),(13),(14)) for the discretization* $\mathcal{D}$ *and* $\bar{u}$ *be the weak solution to (1). We assume that* $\Lambda \in C^1(\overline{\Omega})^{d\times d}$ *and* $\bar{u} \in C^2(\overline{\Omega})$.
Then there exists C_{11} *only depending on* d, Ω, $\bar{u}$, Λ *and* θ *such that*

$$\|\mathbf{v} - \nabla\bar{u}\|_{L^2(\Omega)^d} \leq C_{11}\mathrm{size}(\mathcal{D}) \tag{22}$$

and

$$\|u - \bar{u}\|_{L^2(\Omega)} \leq C_{11}\mathrm{size}(\mathcal{D}). \tag{23}$$

Proof. In this proof, we denote by C_i (for all integer i) various real numbers which can depend on d, Ω, $\bar{u}$, Λ and θ, but not on size($\mathcal{D}$). We denote, for all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$, $\bar{u}_K = \bar{u}(\mathbf{x}_K)$, $\bar{u}_\sigma = \bar{u}(\mathbf{x}_\sigma)$,

$$\bar{F}_{K,\sigma} = \int_\sigma \Lambda(x)\nabla\bar{u}(x)\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma(x),$$

$$\bar{\mathbf{v}}_K = \frac{1}{\mathrm{m}(K)}\Lambda_K^{-1}\sum_{\sigma\in\mathcal{E}_K}\bar{F}_{K,\sigma}(\mathbf{x}_\sigma - \mathbf{x}_K)$$

(notice that Λ_K is indeed invertible since, from (3), $\Lambda_K \geq \alpha_0$). Note that, thanks to Lemma 5.1, we have

$$|\bar{\mathbf{v}}_K - \nabla\bar{u}(x)| \leq C_{12}\mathrm{diam}(K), \quad \forall x \in K\,,\ \forall K \in \mathcal{M}. \tag{24}$$

We thus get

$$\bar{\mathbf{v}}_K\cdot(\mathbf{x}_\sigma - \mathbf{x}_K) = \bar{u}_\sigma - \bar{u}_K + R_{K,\sigma}, \quad \forall K \in \mathcal{M},\ \forall\sigma \in \mathcal{E}_K,$$

with $|R_{K,\sigma}| \leq C_{13}\mathrm{diam}(K)^2$ for all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}_K$. Since $\bar{u}$ is a classical solution to (1), we have

$$-\sum_{\sigma\in\mathcal{E}_K}\bar{F}_{K,\sigma} = \int_K f(x)\,\mathrm{d}x, \quad \forall K \in \mathcal{M}.$$

We denote, for all $K \in \mathcal{M}$ and all $\sigma \in \mathcal{E}_K$, $\widehat{u}_K = u_K - \bar{u}_K$, $\widehat{\mathbf{v}}_K = \mathbf{v}_K - \bar{\mathbf{v}}_K$ and $\widehat{F}_{K,\sigma} = F_{K,\sigma} - \bar{F}_{K,\sigma}$ and we get

$$-\sum_{\sigma\in\mathcal{E}_K}\widehat{F}_{K,\sigma} = 0, \quad \forall K \in \mathcal{M}, \tag{25}$$

$$m_K\Lambda_K\widehat{\mathbf{v}}_K = \sum_{\sigma\in\mathcal{E}_K}\widehat{F}_{K,\sigma}(\mathbf{x}_\sigma - \mathbf{x}_K), \quad \forall K \in \mathcal{M}, \tag{26}$$

$$\begin{aligned}&\widehat{\mathbf{v}}_K\cdot(\mathbf{x}_\sigma - \mathbf{x}_K) + \widehat{\mathbf{v}}_L\cdot(\mathbf{x}_L - \mathbf{x}_\sigma) + R_{K,\sigma} - R_{L,\sigma} = \widehat{u}_L - \widehat{u}_K,\\ &\qquad\qquad \forall K \in \mathcal{M},\ \forall L \in \mathcal{N}_K,\ \text{with } \sigma = K|L,\\ &\widehat{\mathbf{v}}_K\cdot(\mathbf{x}_\sigma - \mathbf{x}_K) + R_{K,\sigma} = -\widehat{u}_K, \quad \forall K \in \mathcal{M},\ \forall\sigma \in \mathcal{E}_{K,\mathrm{ext}}.\end{aligned} \tag{27}$$

We then multiply (25) by $\widehat{u}_K$ and (27) by $\widehat{F}_{K,\sigma}$. Using the conservativity of the fluxes $\widehat{F}_{K,\sigma}$ and (26), we get

$$\sum_{K\in\mathcal{M}} m_K\Lambda_K\widehat{\mathbf{v}}_K\cdot\widehat{\mathbf{v}}_K = -\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} R_{K,\sigma}\widehat{F}_{K,\sigma}.$$

We therefore get, thanks to the Cauchy-Schwarz inequality,

$$\begin{aligned}\alpha_0||\widehat{\mathbf{v}}||^2_{L^2(\Omega)^d} &\leq \sum_{K\in\mathcal{M}} m_K\Lambda_K\widehat{\mathbf{v}}_K\cdot\widehat{\mathbf{v}}_K \\ &\leq \left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{2-d}\widehat{F}^2_{K,\sigma}\right)^{1/2}\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{d-2}R^2_{K,\sigma}\right)^{1/2}.\end{aligned}$$

We can then apply Lemma 2.4, which holds for $\widehat{F}$ and $\widehat{\mathbf{v}}$ (setting $f=0$) and, recalling that $|R_{K,\sigma}|\leq C_{13}\mathrm{diam}(K)^2$, we get

$$||\mathbf{v}-\bar{\mathbf{v}}||_{L^2(\Omega)^d} = ||\widehat{\mathbf{v}}||_{L^2(\Omega)^d} \leq C_{14}\mathrm{size}(\mathcal{D}). \tag{28}$$

The estimate (22) then follows from (24), which implies $||\bar{\mathbf{v}}-\nabla\bar{u}||_{L^\infty(\Omega)^d}\leq C_{15}\mathrm{size}(\mathcal{D})$.
We now set $\nu_K=\frac{1}{\mathrm{m}(K)}$ for all $K\in\mathcal{M}$. With this definition of ν, (27) implies that $(\widehat{\mathbf{v}},R,\widehat{u})\in L_\nu(\mathcal{D})$, as defined in Section 3. In order to apply Lemma 3.1, we need to compute

$$N_2(\mathcal{D},\nu,R)^2=\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{2d-2}\nu_K^2R^2_{K,\sigma}\mathrm{m}(K).$$

We have

$$N_2(\mathcal{D},\nu,R)^2\leq\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{2d-2}\mathrm{m}(K)\frac{C_{13}^2\mathrm{diam}(K)^4}{\mathrm{m}(K)^2}\leq C_{16}\mathrm{size}(\mathcal{D})^2. \tag{29}$$

Since, by Lemma 3.1,

$$\|\widehat{u}\|_{L^2(\Omega)}\leq C_{17}\left(\|\widehat{\mathbf{v}}\|_{L^2(\Omega)^d}+N_2(\mathcal{D},\nu,R)\right),$$

the estimate (23) follows from (28), (29) and an easy comparison between $\bar{u}_K$ and the values of $\bar{u}$ on K. Note that this also proves in this case the convergence of u to $\bar{u}$ in $L^2(\Omega)$ and not only in $L^q(\Omega)$ for all $q\in[1,2)$. □

Remark 2.4 *We could derive as well, in the case $d\leq 3$, an error estimate in the case where $\bar{u}\in H^2(\Omega)$, following some ideas developed in [9] for example. This remark is also valid for Lemma 3.2.*

2.5 General meshes: existence of a discrete solution

In the case of general meshes, we do not know how to prove the convergence of the discrete solution to the weak solution of (1) (but see the penalized scheme in Section 3). However, in the case where Λ is symmetric, we can prove that there exists a solution to ((7),(12),(13),(14)), and we can give some properties on this solution.

Definition 2.2 [Problem 1] *Let us assume Assumption (2)-(4) and that $\Lambda(x)$ is symmetric for a.e. $x\in\Omega$. Let $\mathcal{D}$ be a discretization of Ω in the sense of Definition 2.1. We say that $\mathbf{v}$ is the solution of Problem 1 if $\mathbf{v}\in L^\nabla_\mathcal{D}$ and*

$$J(\mathbf{v})=\inf_{\mathbf{w}\in L^\nabla_\mathcal{D}}J_\mathcal{D}(\mathbf{w}),$$

where $J_{\mathcal{D}} : L^{\nabla}_{\mathcal{D}} \to \mathbb{R}$ *is defined by*

$$J_{\mathcal{D}}(\mathbf{w}) = \frac{1}{2}\int_\Omega \Lambda(x)\mathbf{w}(x)\cdot\mathbf{w}(x)\,\mathrm{d}x - \int_\Omega f(x)\psi(\mathbf{w})(x)\,\mathrm{d}x.$$

If $\mathbf{v} \in L^{\nabla}_{\mathcal{D}}$ *is such a solution, then it satisfies:*

$$\forall \mathbf{w} \in L^{\nabla}_{\mathcal{D}}, \quad \int_\Omega \Lambda(x)\mathbf{v}(x)\cdot\mathbf{w}(x)\,\mathrm{d}x = \int_\Omega f(x)\psi(\mathbf{w})(x)\,\mathrm{d}x.$$

Notice that, since $J_{\mathcal{D}}$ is strictly convex and coercitive (ψ is linear) on a finite dimensional vector space ($L^{\nabla}_{\mathcal{D}}$), the existence and uniqueness of a minimizer is obvious.

Lemma 2.6 *Let us assume Assumptions (2)-(4) and that* $\Lambda(x)$ *is symmetric for a.e.* $x \in \Omega$. *For all discretization* $\mathcal{D}$, *let* $\mathbf{v} \in L^{\nabla}_{\mathcal{D}}$ *be the solution of Problem 1. Then there exists at least one family of real numbers* $(F_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$ *such that (12),(13),(14) hold.*

Proof. The solution of Problem 1 is a pair $(\mathbf{v}, u) \in H^d_{\mathcal{D}} \times H_{\mathcal{D}}$ which satisfies the minimum value of $\frac{1}{2}\int_\Omega \mathbf{v}(x)\cdot\Lambda(x)\mathbf{v}(x)\,\mathrm{d}x - \int_\Omega f(x)u(x)\,\mathrm{d}x$ under the constraints

$$\begin{aligned}
&\mathbf{v}_K\cdot(\mathbf{x}_\sigma - \mathbf{x}_K) + \mathbf{v}_L\cdot(\mathbf{x}_L - \mathbf{x}_\sigma) = u_L - u_K, \quad \forall K \in \mathcal{M},\ \forall L \in \mathcal{N}_K,\\
&\mathbf{v}_K\cdot(\mathbf{x}_\sigma - \mathbf{x}_K) = -u_K, \quad \forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}_{K,\mathrm{ext}}.
\end{aligned}$$

Let us choose, for all $\sigma \in \mathcal{E}_{\mathrm{int}}$, one of the two pairs $(K, L) \in \mathcal{M}^2$ such that $\sigma = K|L$, denoted by $(K(\sigma), L(\sigma))$, and for all $\sigma \in \mathcal{E}_{\mathrm{ext}}$, let us denote by $K(\sigma)$ the element $K \in \mathcal{M}$ such that $\sigma \in \mathcal{E}_{K,\mathrm{ext}}$. Let us introduce, for all $(\mathbf{v}, u, (F_\sigma)_{\sigma\in\mathcal{E}})$ the Lagrangian

$$\begin{aligned}
\mathcal{L}(\mathbf{v}, u, (F_\sigma)_{\sigma\in\mathcal{E}}) = \ & \frac{1}{2}\int_\Omega \Lambda(x)\mathbf{v}(x)\cdot\mathbf{v}(x)\,\mathrm{d}x - \int_\Omega f(x)u(x)\,\mathrm{d}x\\
&- \sum_{\sigma\in\mathcal{E}_{\mathrm{int}}} F_\sigma(u_{K(\sigma)} - u_{L(\sigma)} + \mathbf{v}_{K(\sigma)}\cdot(\mathbf{x}_\sigma - \mathbf{x}_{K(\sigma)}) + \mathbf{v}_{L(\sigma)}\cdot(\mathbf{x}_{L(\sigma)} - \mathbf{x}_\sigma))\\
&- \sum_{\sigma\in\mathcal{E}_{\mathrm{ext}}} F_\sigma(u_{K(\sigma)} + \mathbf{v}_{K(\sigma)}\cdot(\mathbf{x}_\sigma - \mathbf{x}_{K(\sigma)})).
\end{aligned}$$

We now define $F_{K,\sigma}$, for all $K \in \mathcal{M}$, and all $\sigma \in \mathcal{E}_K$. If $\sigma \in \mathcal{E}_{\mathrm{int}}$, we set $F_{K,\sigma} = F_\sigma$ if $K = K(\sigma)$, else $F_{K,\sigma} = -F_\sigma$. If $\sigma \in \mathcal{E}_{\mathrm{ext}}$, we set $F_{K,\sigma} = F_\sigma$. For all $K \in \mathcal{M}$, we then obtain

$$\frac{\partial\mathcal{L}}{\partial u_K}(\mathbf{v}, u, (F_\sigma)_{\sigma\in\mathcal{E}}) = -\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma} - \int_K f(x)\,\mathrm{d}x,$$

and, defining $\mathbf{w}_K \in \mathbb{R}^d$ by $w_K^{(i)} = \frac{\partial\mathcal{L}}{\partial\mathbf{v}_K^{(i)}}(\mathbf{v}, u, (F_\sigma)_{\sigma\in\mathcal{E}})$, for all $i = 1, \ldots, d$,

$$\mathbf{w}_K = m_K\Lambda_K\mathbf{v}_K - \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(\mathbf{x}_\sigma - \mathbf{x}_K).$$

We now remark that there exists at least one family of Lagrange multipliers $(F_\sigma)_{\sigma\in\mathcal{E}}$ such that these partial derivatives vanish (it suffices to consider an extremal family of independent constraints, and to complete the multiplier by 0 on the remaining ones). Then $(F_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$ is such that (12),(13),(14) hold, which concludes the proof of the Lemma. □

Remark 2.5 (Non-uniqueness of $(F_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$) *In the general case, there is no uniqueness property of $(F_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{E}_K}$. Indeed consider a simple cartesian mesh in 2D; on one of the control volumes, put the fluxes $+1$ on two parallel sides and -1 on the two other sides; extend then these fluxes to the other squares by conservativity: you get a family of fluxes in the kernel of (12),(13),(14) (in fact, this gives a full description of this kernel, which is of dimension 1 in this particular case).*
Nevertheless, in the general case, u and $\mathbf{v}$ remain unique and satisfy the same $L^2(\Omega)$ estimate and Poincaré's inequality. However, the numerical computation of these values is quite complex, because it demands to solve a linear system which is not invertible (though we can ensure that it has at least one solution, by Lemma 2.6). Moreover, one of the main interests of Finite Volume schemes is to provide meaningful discrete fluxes; hence, schemes for which the fluxes are not unique (such as ((7),(12),(13),(14)) on some particular meshes, as we have seen above) are to be avoided. This is why, on general meshes, the penalized version of the scheme given in the next section is preferred.

3 A penalized scheme

3.1 The discretization space

Let us assume Assumption (2). Let $\mathcal{D}$ be an admissible discretization and $\nu = (\nu_K)_{K\in\mathcal{M}}$ be a family of positive numbers. We denote by $\mathcal{F}$ the set of real numbers $(F_{K,\sigma})_{K\in\mathcal{M},\,\sigma\in\mathcal{E}_K}$. We define the set $L_\nu(\mathcal{D})$ of all $(\mathbf{v}, F, u) \in H_{\mathcal{D}}^d \times \mathcal{F} \times H_{\mathcal{D}}$ such that

$$\begin{aligned} &\mathbf{v}_K\cdot(\mathbf{x}_\sigma-\mathbf{x}_K)+\mathbf{v}_L\cdot(\mathbf{x}_L-\mathbf{x}_\sigma)+\nu_K\mathrm{m}(K)F_{K,\sigma}-\nu_L\mathrm{m}(L)F_{L,\sigma}=u_L-u_K,\\ &\qquad\qquad\forall K\in\mathcal{M},\ \forall L\in\mathcal{N}_K,\ \text{with } \sigma=K|L,\\ &\mathbf{v}_K\cdot(\mathbf{x}_\sigma-\mathbf{x}_K)+\nu_K F_{K,\sigma}\mathrm{m}(K)=-u_K,\quad \forall K\in\mathcal{M},\ \forall\sigma\in\mathcal{E}_{K,\mathrm{ext}}. \end{aligned} \tag{30}$$

As before, we will need the following properties on this discretization space.

Lemma 3.1 [Poincaré's Inequality] *Let us assume Assumption (2). Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1, such that* $\mathrm{regul}(\mathcal{D})\le\theta$ *for some $\theta>0$. Let $(\nu_K)_{K\in\mathcal{M}}$ be a family of positive real numbers. Then there exists C_{17} only depending on d, Ω and θ such that, for all $(\mathbf{v}, F, u) \in L_\nu(\mathcal{D})$,*

$$\|u\|_{L^2(\Omega)}\le C_{17}\left(\|\mathbf{v}\|_{L^2(\Omega)^d}+N_2(\mathcal{D},\nu,F)\right), \tag{31}$$

where we have noted $N_2(\mathcal{D},\nu,F)=\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{2d-2}\nu_K^2F_{K,\sigma}^2\mathrm{m}(K)\right)^{1/2}$.

Proof.
We use the same reasoning and notations as in the proof of Lemma 2.1. After multiplying (30) by $\int_\sigma\nabla w(x)\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma(x)$, summing on the edges and gathering by control volumes, we find $T_1+T_7=\|u\|^2_{L^2(\Omega)}$, where T_1 is the same as in the proof of Lemma 2.1 and

$$\begin{aligned} T_7 &= \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\nu_K F_{K,\sigma}\mathrm{m}(K)\int_\sigma\nabla w(x)\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma(x)\\ &= \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\nu_K F_{K,\sigma}\mathrm{m}(K)\mathrm{m}(\sigma)\mathbf{G}_{K,\sigma}\cdot\mathbf{n}_{K,\sigma}. \end{aligned}$$

From the proof of Lemma 2.1, we now how to bound T_1. We thus have to study T_7, comparing it with

$$T_8 = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \nu_K F_{K,\sigma} \mathrm{m}(K)\mathrm{m}(\sigma)\mathbf{G}_K\cdot\mathbf{n}_{K,\sigma}.$$

We get, using (9),

$$\begin{aligned}
&(T_7-T_8)^2\\
&\quad\le \left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} \mathrm{diam}(K)\mathrm{m}(\sigma)\nu_K^2F_{K,\sigma}^2\mathrm{m}(K)^2\right)\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\frac{\mathrm{m}(\sigma)}{\mathrm{diam}(K)}|\mathbf{G}_K-\mathbf{G}_{K,\sigma}|^2\right)\\
&\quad\le \left(\omega_{d-1}\omega_d\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{2d}\nu_K^2F_{K,\sigma}^2\mathrm{m}(K)\right)\mathrm{regul}(\mathcal{D})C_2||w||^2_{H^2(\Omega)}\\
&\quad\le \omega_{d-1}\omega_d\mathrm{diam}(\Omega)^2N_2(\mathcal{D},\nu,F)^2\mathrm{regul}(\mathcal{D})C_2||w||^2_{H^2(\Omega)}.
\end{aligned}$$

On the other hand, we have

$$\begin{aligned}
T_8^2 &\le \left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{m}(\sigma)^2\nu_K^2F_{K,\sigma}^2\mathrm{m}(K)\right)\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{m}(K)|\mathbf{G}_K|^2\right)\\
&\le \omega_{d-1}^2N_2(\mathcal{D},\nu,F)^2\left(\mathrm{regul}(\mathcal{D})\sum_{K\in\mathcal{M}}\mathrm{m}(K)|\mathbf{G}_K|^2\right)\\
&\le \omega_{d-1}^2N_2(\mathcal{D},\nu,F)^2\mathrm{regul}(\mathcal{D})\|w\|^2_{H^1(\Omega)}.
\end{aligned}$$

We then write, using the bound on T_1 obtained at the end of the proof of Lemma 2.1,

$$\begin{aligned}
||u||^2_{L^2(\Omega)} &= T_1+T_7\\
&\le T_1+|T_7-T_8|+|T_8|\\
&\le \sqrt{\frac{\omega_{d-1}C_2\theta^3}{\omega_d}}\mathrm{diam}(\Omega)\|\mathbf{v}\|_{L^2(\Omega)}||w||_{H^2(\Omega)}+||\mathbf{v}||_{L^2(\Omega)^d}||w||_{H^1(\Omega)}\\
&\quad+\sqrt{\omega_{d-1}\omega_dC_2\theta}\,\mathrm{diam}(\Omega)N_2(\mathcal{D},\nu,F)\|w\|_{H^2(\Omega)}+\omega_{d-1}\sqrt{\theta}\,N_2(\mathcal{D},\nu,F)\|w\|_{H^1(\Omega)}
\end{aligned}$$

and we conclude as in the proof of Lemma 2.1, using the fact that $||w||_{H^2(\Omega)}\le C_3||u||_{L^2(\Omega)}$ (with C_3 only depending on d and Ω). □

Lemma 3.2 [Equicontinuity of the translations] *Let us assume Assumption (2). Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1, such that* $\mathrm{regul}(\mathcal{D})\le\theta$ *for some $\theta>0$. Let $(\nu_K)_{K\in\mathcal{M}}$ be a family of positive real numbers. Then there exists C_{18} only depending on d, Ω and θ such that, for all $(\mathbf{v},F,u)\in L_\nu(\mathcal{D})$ and all $\xi\in\mathbb{R}^d$,*

$$\|u(\cdot+\xi)-u\|_{L^1(\mathbb{R}^d)}\le C_{18}\left(\|\mathbf{v}\|_{L^1(\Omega)^d}+N_1(\mathcal{D},\nu,F)\right)|\xi|, \tag{32}$$

where $N_1(\mathcal{D},\nu,F)=\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{d-1}\nu_K|F_{K,\sigma}|\mathrm{m}(K)$ (and u has been extended by 0 outside Ω).

PROOF. The proof is similar to that of Lemma 2.2. We introduce the same notation $\chi(x,\xi,\sigma)$. Applying (30), we get, for a.e. $x \in \mathbb{R}^d$, $|u(x+\xi)-u(x)| \le T_9(x)+T_{10}(x)$ with

$$T_9(x) = \left(\begin{array}{l} \displaystyle\sum_{\sigma\in\mathcal{E}_{\rm int},\sigma=K|L} \chi(x,\xi,\sigma)(|\mathbf{v}_K||\mathbf{x}_\sigma-\mathbf{x}_K| + |\mathbf{v}_L||\mathbf{x}_L-\mathbf{x}_\sigma|) \\ \displaystyle + \sum_{\sigma\in\mathcal{E}_{\rm ext},\sigma\in\mathcal{E}_K} \chi(x,\xi,\sigma)|\mathbf{v}_K||\mathbf{x}_\sigma-\mathbf{x}_K| \end{array} \right)$$

and

$$T_{10}(x) = \left(\begin{array}{l} \displaystyle\sum_{\sigma\in\mathcal{E}_{\rm int},\sigma=K|L} \chi(x,\xi,\sigma)\left(\nu_K \mathrm{m}(K)|F_{K,\sigma}| + \nu_L \mathrm{m}(L)|F_{L,\sigma}|\right) \\ \displaystyle + \sum_{\sigma\in\mathcal{E}_{\rm ext},\sigma\in\mathcal{E}_K} \chi(x,\xi,\sigma)\nu_K \mathrm{m}(K)|F_{K,\sigma}| \end{array} \right).$$

The handling of $T_9(x)$ is similar to what is done in the proof of Lemma 2.2, and we obtain $\int_{\mathbb{R}^d} T_9(x)\,\mathrm{d}x \le C_4||\mathbf{v}||_{L^1(\Omega)^d}|\xi|$. We have

$$T_{10}(x) = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \chi(x,\xi,\sigma)\nu_K \mathrm{m}(K)|F_{K,\sigma}|$$

and, to bound this expression, we write $\int_{\mathbb{R}^d} \chi(x,\xi,\sigma)\,\mathrm{d}x = \mathrm{m}(\sigma)|\mathbf{n}_\sigma\cdot\xi| \le \omega_{d-1}\mathrm{diam}(K)^{d-1}|\xi|$ (for $\sigma\in\mathcal{E}_K$), which gives

$$\int_{\mathbb{R}^d} T_{10}(x)\,\mathrm{d}x \le \omega_{d-1}|\xi| \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \mathrm{diam}(K)^{d-1}\nu_K|F_{K,\sigma}|\mathrm{m}(K)$$

and concludes the proof. □

Lemma 3.3 [Compactness property] *Let us assume Assumption (2). Let $(\mathcal{D}_m)_{m\ge1}$ be admissible discretizations of Ω in the sense of Definition 2.1, such that* $\mathrm{size}(\mathcal{D}_m) \to 0$ *as $m\to\infty$ and* $(\mathrm{regul}(\mathcal{D}_m))_{m\ge1}$ *is bounded. Let $(\mathbf{v}_m, F_m, u_m, \nu_m)_{m\ge1}$ be such that $(\mathbf{v}_m, F_m, u_m) \in L_{\nu_m}(\mathcal{D}_m)$, $(\mathbf{v}_m)_{m\ge1}$ is bounded in $L^2(\Omega)^d$ and $N_2(\mathcal{D}_m,\nu_m,F_m)\to0$ as $m\to\infty$ (N_2 has been defined in Lemma 3.1).*
Then there exists a subsequence of $(\mathcal{D}_m)_{m\ge1}$ (still denoted by $(\mathcal{D}_m)_{m\ge1}$) and $\bar u \in H^1_0(\Omega)$ such that the corresponding sequence $(u_m)_{m\ge1}$ converges to $\bar u$ weakly in $L^2(\Omega)$ and strongly in $L^q(\Omega)$ for all $q<2$, and such that $(\mathbf{v}_m)_{m\ge1}$ converges to $\nabla\bar u$ weakly in $L^2(\Omega)^d$.

PROOF.
Notice first that, for all discretization $\mathcal{D}$, for all $\nu = (\nu_K)_{K\in\mathcal{M}}$ positive number and for all $F = (F_{K,\sigma})_{K\in\mathcal{M},\,\sigma\in\mathcal{E}_K}$,

$$\begin{aligned} N_1(\mathcal{D},\nu,F) &= \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \mathrm{diam}(K)^{d-1}\nu_K|F_{K,\sigma}|\mathrm{m}(K) \\ &\le \left(\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \mathrm{diam}(K)^{2d-2}\nu_K^2 F_{K,\sigma}^2 \mathrm{m}(K) \right)^{1/2} \left(\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}(K) \right)^{1/2} \\ &\le N_2(\mathcal{D},\nu,F)\mathrm{regul}(\mathcal{D})^{1/2}\mathrm{m}(\Omega)^{1/2}. \end{aligned}$$

Hence, if $N_2(\mathcal{D},\nu,F)$ and $\mathrm{regul}(\mathcal{D})$ are bounded, so is $N_1(\mathcal{D},\nu,F)$.

Owing to this, we can reason as in the proof of Lemma 2.3: the hypotheses and Lemmas 3.1, 3.2 allow to extract a subsequence such that $\mathbf{v_m} \to \bar{\mathbf{v}}$ weakly in $L^2(\Omega)^d$ and $u_m \to \bar{u}$ weakly in $L^2(\Omega)$ and strongly in $L^q(\Omega)$ for all $q<2$. To prove that $\bar{u}\in H^1_0(\Omega)$ and $\nabla\bar{u}=\bar{\mathbf{v}}$, we still follow the proof of Lemma 2.3 (omitting the index m). Since

$$\left|\int_\sigma \varphi(x)\,\mathrm{d}\gamma(x)\mathbf{e}\cdot\mathbf{n}_{K,\sigma}\right| \le C_\varphi \mathrm{m}(\sigma),$$

we only have to prove that $T_{11}=\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{m}(\sigma)\nu_K|F_{K,\sigma}|\mathrm{m}(K)$ tends to 0 (this quantity bounds the additional term, with respect to the proof of Lemma 2.3, which appears when multiplying (30) by $\int_\sigma \varphi(x)\,\mathrm{d}\gamma(x)\,\mathbf{e}\cdot\mathbf{n}_{K,\sigma}$). But, as noticed at the beginning of this proof,

$$T_{11}\le \omega_{d-1}N_1(\mathcal{D},\nu,F)\le \omega_{d-1}\mathrm{m}(\Omega)^{1/2}\mathrm{regul}(\mathcal{D})^{1/2}N_2(\mathcal{D},\nu,F),$$

which completes the proof of the lemma, by assumption on the discretizations. □

3.2 The scheme

Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1 and $\nu=(\nu_K)_{K\in\mathcal{M}}$ be positive numbers. We consider the scheme defined by (30), the conservativity property

$$F_{K,\sigma}+F_{L,\sigma}=0,\quad \forall\sigma=K|L\in\mathcal{E}_{\mathrm{int}}, \tag{33}$$

the condition

$$m_K\Lambda_K\mathbf{v}_K=\sum_{\sigma\in\mathcal{E}_K}F_{K,\sigma}(\mathbf{x}_\sigma-\mathbf{x}_K),\quad \forall K\in\mathcal{M}, \tag{34}$$

and the relation

$$-\sum_{\sigma\in\mathcal{E}_K}F_{K,\sigma}=\int_K f(x)\,\mathrm{d}x,\quad \forall K\in\mathcal{M}. \tag{35}$$

We now prove the existence and uniqueness of a solution to this scheme, and give an estimate on this solution.

Lemma 3.4 *Let us assume Assumptions (2)-(4). Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1. Let $(\nu_K)_{K\in\mathcal{M}}$ be a family of positive real numbers. Then there exists one and only one $(\mathbf{v},F,u)\in L_\nu(\mathcal{D})$ solution of ((30),(33),(34),(35)). Moreover, for all $\nu_0>0$, for all $\beta_0\ge\beta\ge 2-2d$ such that $\nu_K\le\nu_0\mathrm{diam}(K)^\beta$ $(\forall K\in\mathcal{M})$ and for all $\theta\ge\mathrm{regul}(\mathcal{D})$, this solution satisfies*

$$\|\mathbf{v}\|^2_{L^2(\Omega)^d}+\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\nu_K F^2_{K,\sigma}\mathrm{m}(K)\le C_{19}||f||^2_{L^2(\Omega)} \tag{36}$$

where C_{19} only depends on d, Ω, α_0, θ, ν_0 and β_0.

Proof. Notice first that, since ((30),(33),(34),(35)) is square and linear in $(\mathbf{v},F,u)$, it suffices to prove the estimate in order to obtain the existence and uniqueness of the solution (because $f=0$ then implies $F=0$ and $\mathbf{v}=0$, and thus $u=0$ by (31)).

Multiply (35) by u_K, sum on the control volumes and gather by edges using (33); multiply (30) by $F_{K,\sigma}$, sum on the edges and gather by control volumes still using (33). This gives, by (34),

$$\begin{aligned}\int_\Omega \mathbf{v}(x)\cdot\Lambda(x)\mathbf{v}(x)\,\mathrm{d}x + \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\nu_K F_{K,\sigma}^2\mathrm{m}(K) &= \int_\Omega f(x)u(x)\,\mathrm{d}x \\ &\leq ||f||_{L^2(\Omega)}||u||_{L^2(\Omega)}.\end{aligned} \tag{37}$$

Using Young's inequality and Lemma 3.1, we deduce that, for all $\varepsilon>0$,

$$\begin{aligned}\alpha_0||\mathbf{v}||^2_{L^2(\Omega)^d} + \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\nu_K F_{K,\sigma}^2\mathrm{m}(K) \leq \frac{1}{2\varepsilon}||f||^2_{L^2(\Omega)} + \varepsilon C_{17}^2||\mathbf{v}||^2_{L^2(\Omega)^d} \\ +\varepsilon C_{17}^2\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{2d-2}\nu_K^2F_{K,\sigma}^2\mathrm{m}(K).\end{aligned} \tag{38}$$

Since $\nu_K\leq\nu_0\mathrm{diam}(K)^\beta$, we have $\nu_K\mathrm{diam}(K)^{2d-2}\leq\nu_0\mathrm{diam}(K)^{\beta+2d-2}\leq\nu_0\mathrm{diam}(\Omega)^{\beta+2d-2}\leq\nu_0\sup(1,\mathrm{diam}(\Omega)^{\beta_0+2d-2})$ (recall that $\beta+2d-2\geq 0$). Hence, (38) gives

$$\begin{aligned}\alpha_0||\mathbf{v}||^2_{L^2(\Omega)^d} + \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\nu_K F_{K,\sigma}^2\mathrm{m}(K) \leq \frac{1}{2\varepsilon}||f||^2_{L^2(\Omega)} + \varepsilon C_{17}^2||\mathbf{v}||^2_{L^2(\Omega)^d} \\ +\varepsilon\nu_0\sup(1,\mathrm{diam}(\Omega)^{\beta_0+2d-2})C_{17}^2\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\nu_K F_{K,\sigma}^2\mathrm{m}(K).\end{aligned}$$

Taking $\varepsilon=\min(\frac{\alpha_0}{2C_{17}^2},\frac{1}{2\nu_0\sup(1,\mathrm{diam}(\Omega)^{\beta_0+2d-2})C_{17}^2})$ concludes the proof of the lemma. □

We now prove the convergence, as $\mathrm{size}(\mathcal{D})\to 0$ and with a suitable choice of $(\nu_K)_{K\in\mathcal{M}}$, of the solution to ((30),(33),(34),(35)) to the weak solution of (1).

Theorem 3.1 *Let us assume Assumptions (2)-(4). Let $(\mathcal{D}_m)_{m\geq 1}$ be admissible discretizations of Ω in the sense of Definition 2.1, such that $\mathrm{size}(\mathcal{D}_m)\to 0$ as $m\to\infty$ and $(\mathrm{regul}(\mathcal{D}_m))_{m\geq 1}$ is bounded. Let $\nu_0>0$ and $\beta\in(2-2d,4-2d)$ be fixed. For all $m\geq 1$, let $(\mathbf{v}_m,F_m,u_m)$ be the solution to ((30),(33),(34),(35)) for the discretization $\mathcal{D}_m$, setting $\nu_K=\nu_0\mathrm{diam}(K)^\beta$ for all $K\in\mathcal{M}_m$. Let $\bar{u}$ be the weak solution to (1).*
Then, as $m\to\infty$, $\mathbf{v}_m\to\nabla\bar{u}$ strongly in $L^2(\Omega)^d$ and $u_m\to\bar{u}$ weakly in $L^2(\Omega)$ and strongly in $L^q(\Omega)$ for all $q<2$.

Proof.
For the simplicity of the notations, we omit the index m. First, thanks to Estimate (36) and since $\nu_K=\nu_0\mathrm{diam}(K)^\beta$, we get

$$\begin{aligned}N_2(\mathcal{D},\nu,F)^2 &= \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{2d-2}\nu_K^2F_{K,\sigma}^2\mathrm{m}(K) \\ &= \nu_0\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K}\mathrm{diam}(K)^{\beta+2d-2}\nu_K F_{K,\sigma}^2\mathrm{m}(K) \\ &\leq \nu_0\mathrm{size}(\mathcal{D})^{\beta+2d-2}C_{20}\end{aligned}$$

where C_{20} does not depend on the discretization $\mathcal{D}$ (recall that $\mathrm{regul}(\mathcal{D})$ is bounded). Since $\beta+2d-2>0$, this last quantity tends to 0, and so does $N_2(\mathcal{D},\nu,F)$, as $\mathrm{size}(\mathcal{D})\to 0$. Hence,

still using (36), we see that the assumptions of Lemma 3.3 are satisfied: there exists $\bar{u} \in H_0^1(\Omega)$ such that, up to a subsequence and as $\mathrm{size}(\mathcal{D}) \to 0$, $\mathbf{v} \to \nabla\bar{u}$ weakly in $L^2(\Omega)^d$ and $u \to \bar{u}$ weakly in $L^2(\Omega)$ and strongly in $L^q(\Omega)$ for $q < 2$.
Since (35) is similar to (14), (33) is similar to (12) and (34) is similar to (13), we can reason as in the proof of Theorem 2.1 and we arrive at (20). It remains to prove that $T_6 \to 0$ as $\mathrm{size}(\mathcal{D}) \to 0$ (to see that $\bar{u}$ is a weak solution to (1)), and that $\mathbf{v}$ strongly converges.

We have, from (21),

$$\begin{aligned} |T_6|^2 &\leq \left(C_\varphi \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \mathrm{diam}(K)^2 |F_{K,\sigma}| \right)^2 \\ &\leq C_\varphi^2 \left(\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \nu_K F_{K,\sigma}^2 \mathrm{m}(K) \right) \left(\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \frac{\mathrm{diam}(K)^4}{\nu_K \mathrm{m}(K)} \right) \\ &\leq C_{21} \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \frac{\mathrm{diam}(K)^4}{\nu_K \mathrm{m}(K)^2} \mathrm{m}(K) \end{aligned} \tag{39}$$

where, according to (36), C_{21} does not depend on the mesh since $\mathrm{regul}(\mathcal{D})$ stays bounded. But $\nu_K = \nu_0 \mathrm{diam}(K)^\beta$ and $\mathrm{diam}(K)^d \leq \frac{\mathrm{regul}(\mathcal{D})}{\omega_d} \mathrm{m}(K)$, so that

$$\frac{\mathrm{diam}(K)^4}{\nu_K \mathrm{m}(K)^2} \leq \frac{\mathrm{regul}(\mathcal{D})^2 \mathrm{diam}(K)^{4-\beta}}{\omega_d^2 \nu_0 \mathrm{diam}(K)^{2d}} = \frac{\mathrm{regul}(\mathcal{D})^2}{\omega_d^2 \nu_0} \mathrm{diam}(K)^{4-2d-\beta}.$$

Since $4 - 2d - \beta > 0$, we deduce from (39) that

$$|T_6|^2 \leq C_{21} \frac{\mathrm{regul}(\mathcal{D})^2}{\omega_d^2 \nu_0} \mathrm{size}(\mathcal{D})^{4-2d-\beta} \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}(K) \leq \frac{C_{21} \mathrm{regul}(\mathcal{D})^3 \mathrm{m}(\Omega)}{\omega_d^2 \nu_0} \mathrm{size}(\mathcal{D})^{4-2d-\beta}$$

and this quantity tends to 0 as $\mathrm{size}(\mathcal{D}) \to 0$, which concludes the proof that $\bar{u}$ is a weak solution to (1).

The strong convergence of $\mathbf{v}$ to $\nabla\bar{u}$ is a consequence of (37). From this equation, and defining $N(\mathbf{w})^2 = \int_\Omega \Lambda(x)\mathbf{w}(x)\cdot\mathbf{w}(x)\,\mathrm{d}x$ as in the proof of Theorem 2.1, we have $N(\mathbf{v})^2 \leq \int_\Omega f(x)u(x)\,\mathrm{d}x$ and thus

$$\limsup_{\mathrm{size}(\mathcal{D})\to 0} N(\mathbf{v})^2 \leq \lim_{\mathrm{size}(\mathcal{D})\to 0} \int_\Omega f(x)u(x)\,\mathrm{d}x = \int_\Omega f(x)\bar{u}(x)\,\mathrm{d}x = N(\nabla\bar{u})^2 \tag{40}$$

(we use the fact that $u \to \bar{u}$ weakly in $L^2(\Omega)$ and that $\bar{u}$ is the weak solution to (1)). But N is a norm on $L^2(\Omega)^d$ and $\mathbf{v} \to \nabla\bar{u}$ weakly in $L^2(\Omega)^d$ as $\mathrm{size}(\mathcal{D}) \to 0$, so that $N(\nabla\bar{u}) \leq \liminf_{\mathrm{size}(\mathcal{D})\to 0} N(\mathbf{v})$. We conclude with (40) that $N(\mathbf{v}) \to N(\nabla\bar{u})$ as $\mathrm{size}(\mathcal{D}) \to 0$ and, therefore, the weak convergence of $\mathbf{v}$ to $\nabla\bar{u}$ in $L^2(\Omega)^d$ is in fact strong. □

Remark 3.1 *As a consequence of (37) and the strong convergence of* $\mathbf{v}$ *to* $\nabla\bar{u}$*, we see that* $\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{E}_K} \nu_K F_{K,\sigma}^2 \mathrm{m}(K) \to 0$ *as* $\mathrm{size}(\mathcal{D}) \to 0$*. This strengthens Lemma 3.4 which only states that this quantity is bounded.*

Remark 3.2 *In a similar way as above, we could get the convergence of u_m to $\bar{u}$ in $L^2(\Omega)$ by assuming a uniform regularity property for the mesh. Thanks to the error estimate below, we nevertheless get this strong convergence in some particular cases, with no additional hypothesis on the discretization.*

We now derive an error estimate, which also could be extended to the case $d \leq 3$ and $\bar{u} \in H^2(\Omega)$ following some arguments of [9].

Theorem 3.2 *Let us assume Assumptions (2)-(4). Let $\mathcal{D}$ be an admissible discretization of Ω in the sense of Definition 2.1, such that* $\mathrm{size}(\mathcal{D}) \leq 1$ *and* $\mathrm{regul}(\mathcal{D}) \leq \theta$ *for some $\theta > 0$. We take $\nu_0 > 0$ and $\beta \in (2-2d, 4-2d)$ and, for all $K \in \mathcal{M}$, we let $\nu_K = \nu_0 \mathrm{diam}(K)^\beta$. Let $(\mathbf{v}, F, u)$ be the solution to ((30),(33),(34),(35)). Let $\bar{u}$ be the weak solution to (1). We assume that $\Lambda \in C^1(\overline{\Omega})$ and $\bar{u} \in C^2(\overline{\Omega})$.*
Then there exists C_{22} only depending on d, Ω, $\bar{u}$, Λ, θ and ν_0 such that

$$\|\mathbf{v} - \nabla \bar{u}\|_{L^2(\Omega)^d} \leq C_{22} \mathrm{size}(\mathcal{D})^{\frac{1}{4}\min(\beta+2d-2, 4-2d-\beta)} \tag{41}$$

and

$$\|u - \bar{u}\|_{L^2(\Omega)} \leq C_{22} \mathrm{size}(\mathcal{D})^{\frac{1}{4}\min(\beta+2d-2, 4-2d-\beta)} \tag{42}$$

(note that the maximum value of $\frac{1}{4}\min(\beta + 2d - 2, 4 - 2d - \beta)$ is $\frac{1}{4}$, obtained for $\beta = 3 - 2d$).

Proof.
The proof is similar to that of Theorem 2.2, and we use the same notations. We have the same relations as in the proof of Theorem 2.2, except for (27) which becomes

$$\begin{aligned} &\widehat{\mathbf{v}}_K \cdot (\mathbf{x}_\sigma - \mathbf{x}_K) + \widehat{\mathbf{v}}_L \cdot (\mathbf{x}_L - \mathbf{x}_\sigma) + \nu_K \mathrm{m}(K) F_{K,\sigma} + R_{K,\sigma} \\ &\qquad\qquad -\nu_L \mathrm{m}(L) F_{L,\sigma} - R_{L,\sigma} = \widehat{u}_L - \widehat{u}_K, \\ &\qquad\qquad \forall K \in \mathcal{M},\ \forall L \in \mathcal{N}_K,\ \text{with } \sigma = K|L, \\ &\widehat{\mathbf{v}}_K \cdot (\mathbf{x}_\sigma - \mathbf{x}_K) + \nu_K F_{K,\sigma} \mathrm{m}(K) + R_{K,\sigma} = -\widehat{u}_K, \quad \forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}_{K,\mathrm{ext}}. \end{aligned} \tag{43}$$

We then get, multiplying (25) by $\widehat{u}_K$, (43) by $\widehat{F}_{K,\sigma}$ and using (26),

$$\sum_{K \in \mathcal{M}} m_K \Lambda_K \widehat{\mathbf{v}}_K \cdot \widehat{\mathbf{v}}_K + \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} \nu_K F_{K,\sigma}^2 \mathrm{m}(K) = T_{12} - T_{13} + T_{14}, \tag{44}$$

where

$$T_{12} = \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} \nu_K F_{K,\sigma} \bar{F}_{K,\sigma} \mathrm{m}(K),$$

$$T_{13} = \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} R_{K,\sigma} F_{K,\sigma},$$

$$T_{14} = \sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} R_{K,\sigma} \bar{F}_{K,\sigma}.$$

Since $|\bar{F}_{K,\sigma}| \leq C_{23} \mathrm{m}(\sigma) \leq C_{23} \omega_{d-1} \mathrm{diam}(K)^{d-1}$ and $|R_{K,\sigma}| \leq C_{13} \mathrm{diam}(K)^2$, it is straightforward to see that $|T_{14}| \leq C_{12} \mathrm{size}(\mathcal{D})$. Thanks to the Cauchy-Schwarz inequality, we get

$$T_{12}^2 \leq \left(\sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} \nu_K F_{K,\sigma}^2 \mathrm{m}(K) \right) \left(\sum_{K \in \mathcal{M}} \sum_{\sigma \in \mathcal{E}_K} \nu_K \bar{F}_{K,\sigma}^2 \mathrm{m}(K) \right).$$

Using Lemma 3.4 with $\beta_0 = 4-2d \geq \beta$, we thus obtain $T_{12}^2 \leq C_{19}||f||^2_{L^2(\Omega)} C_{24}\text{size}(\mathcal{D})^{\beta+2d-2}$. We also have

$$\begin{aligned} T_{13}^2 &\leq \left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} \nu_K F_{K,\sigma}^2 \text{m}(K)\right)\left(\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} \text{m}(K)\frac{R_{K,\sigma}^2}{\nu_K \text{m}(K)^2}\right) \\ &\leq C_{19}||f||^2_{L^2(\Omega)} C_{25}\text{size}(\mathcal{D})^{4-2d-\beta}. \end{aligned}$$

Gathering these estimates in (44) leads to

$$||\widehat{\mathbf{v}}||^2_{L^2(\Omega)} \leq C_{26}\left(\text{size}(\mathcal{D}) + \text{size}(\mathcal{D})^{\frac{1}{2}(\beta+2d-2)} + \text{size}(\mathcal{D})^{\frac{1}{2}(4-2d-\beta)}\right) \tag{45}$$

and (41) follows, using the fact that $\text{size}(\mathcal{D}) \leq 1$ and that $||\bar{\mathbf{v}} - \nabla\bar{u}||_{L^\infty(\Omega)^d} \leq C_{15}\text{size}(\mathcal{D})$.
We now set $\widetilde{F}_{K,\sigma} = F_{K,\sigma} + \frac{R_{K,\sigma}}{\nu_K \text{m}(K)}$ for all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}$ and we estimate $N_2(\mathcal{D},\nu,\widetilde{F})$ the following way:

$$\begin{aligned} N_2(\mathcal{D},\nu,\widetilde{F})^2 &= \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} \text{diam}(K)^{2d-2}\nu_K^2 \widetilde{F}_{K,\sigma}^2 \text{m}(K) \\ &\leq 2\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} \text{diam}(K)^{2d-2}\nu_K^2 F_{K,\sigma}^2 \text{m}(K) \\ &\quad +2\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{E}_K} \text{diam}(K)^{2d-2}\nu_K^2 \text{m}(K)\frac{C_{13}^2 \text{diam}(K)^4}{(\nu_K \text{m}(K))^2} \\ &\leq C_{27}(\text{size}(\mathcal{D})^{\beta+2d-2} + \text{size}(\mathcal{D})^2) \end{aligned} \tag{46}$$

(we have used (36)). We can apply Lemma 3.1, since (43) implies that $(\widehat{\mathbf{v}}, \widetilde{F}, \widehat{u}) \in L_\nu(\mathcal{D})$: we obtain

$$\|\widehat{u}\|_{L^2(\Omega)} \leq C_{17}\left(\|\widehat{\mathbf{v}}\|_{L^2(\Omega)^d} + N_2(\mathcal{D},\nu,\widetilde{F})\right)$$

and (42) follows from (45), (46) and an easy estimate between $\bar{u}_K$ and the values of $\bar{u}$ on K. □

Remark 3.3 *This error estimate is not sharp, and the numerical results below show a much better order of convergence.*

4 Implementation

We present the practical implementation in the case where $\Lambda(x)$ is symmetric for a.e. $x \in \Omega$, though it is valid for any Λ.

4.1 Resolution procedure

The size of System ((30),(33),(34),(35)) is equal to $(d+1)\text{Card}(\mathcal{M}) + 2\text{Card}(\mathcal{E}_{\text{int}}) + \text{Card}(\mathcal{E}_{\text{ext}})$. However, it is possible to proceed to an algebraic elimination which leads to a symmetric positive definite sparse linear system with $\text{Card}(\mathcal{E}_{\text{int}})$ unknowns, following the same principles as in the hybrid resolution of a mixed finite element problem (see for example [17]). Indeed, for all $(\mathbf{v}, F, u)$ such that (30) and (34) hold, we define $(u_\sigma)_{\sigma\in\mathcal{E}_K}$ by

$$\mathbf{v}_K \cdot (\mathbf{x}_\sigma - \mathbf{x}_K) + \nu_K F_{K,\sigma}\text{m}(K) = u_\sigma - u_K, \quad \forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}.$$

We thus have that $u_\sigma = 0$ for all $\sigma \in \mathcal{E}_{K,\text{ext}}$. We can then express $(\mathbf{v}, F)$ as a function of $(u_\sigma)_{\sigma\in\mathcal{E}_K}$ and of u, since we have

$$\frac{1}{\mathrm{m}(K)} \sum_{\sigma'\in\mathcal{E}_K} F_{K,\sigma'} \Lambda_K^{-1} (\mathbf{x}_{\sigma'} - \mathbf{x}_K) \cdot (\mathbf{x}_\sigma - \mathbf{x}_K) + \nu_K F_{K,\sigma} \mathrm{m}(K) = u_\sigma - u_K, \\ \forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E},$$

which is, for all $K \in \mathcal{M}$, an invertible linear system with unknown $(F_{K,\sigma})_{\sigma\in\mathcal{E}_K}$, under the form $B_K(F_{K,\sigma})_{\sigma\in\mathcal{E}_K} = (u_\sigma - u_K)_{\sigma\in\mathcal{E}_K}$ where B_K is a symmetric positive definite matrix (thanks to the condition $\nu_K > 0$). We can then write

$$F_{K,\sigma} = \sum_{\sigma'\in\mathcal{E}_K} (B_K^{-1})_{\sigma\sigma'} (u_{\sigma'} - u_K), \quad \forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{E}_K. \tag{47}$$

We then obtain from (35), denoting $b_{K,\sigma'} = \sum_{\sigma\in\mathcal{E}_K} (B_K^{-1})_{\sigma\sigma'}$ and $b_K = \sum_{\sigma'\in\mathcal{E}_K} b_{K,\sigma'}$, that u_K satisfies the relation

$$- \sum_{\sigma'\in\mathcal{E}_K} b_{K,\sigma'} u_{\sigma'} + b_K u_K = \int_K f(x)\,\mathrm{d}x. \tag{48}$$

We have $(b_{K,\sigma'})_{\sigma'\in\mathcal{E}_K} = B_K^{-1}(1)_{\sigma'\in\mathcal{E}_K}$ and therefore we get $b_K = (1)_{\sigma'\in\mathcal{E}_K} \cdot B_K^{-1}(1)_{\sigma'\in\mathcal{E}_K} > 0$ since B_K^{-1} is symmetric positive definite. Reporting the previous linear relations in (33), we find

$$\sum_{\sigma'\in\mathcal{E}_K} \left((B_K^{-1})_{\sigma\sigma'} - \frac{b_{K,\sigma} b_{K,\sigma'}}{b_K} \right) u_{\sigma'} + \sum_{\sigma'\in\mathcal{E}_L} \left((B_L^{-1})_{\sigma\sigma'} - \frac{b_{L,\sigma} b_{L,\sigma'}}{b_L} \right) u_{\sigma'} = \\ \frac{b_{K,\sigma}}{b_K} \int_K f(x)\,\mathrm{d}x + \frac{b_{L,\sigma}}{b_L} \int_L f(x)\,\mathrm{d}x, \quad \forall \sigma = K|L \in \mathcal{E}_{\text{int}}, \tag{49}$$

which is a symmetric linear system, whose unknowns are $(u_\sigma)_{\sigma\in\mathcal{E}_{\text{int}}}$. Let us show that its matrix M is positive. We can write, for all family of real numbers $(u_\sigma)_{\sigma\in\mathcal{E}_{\text{int}}}$,

$$(u_\sigma)_{\sigma\in\mathcal{E}_{\text{int}}} \cdot M\ (u_\sigma)_{\sigma\in\mathcal{E}_{\text{int}}} = \sum_{K\in\mathcal{M}} \left(\sum_{\sigma\in\mathcal{E}_K} \sum_{\sigma'\in\mathcal{E}_K} (B_K^{-1})_{\sigma\sigma'} u_\sigma u_{\sigma'} - \frac{\left(\sum_{\sigma\in\mathcal{E}_K} b_{K,\sigma} u_\sigma\right)^2}{b_K} \right).$$

Thanks to the fact that B_K^{-1} is symmetric positive definite, we get, using the Cauchy-Schwarz inequality,

$$\left((1)_{\sigma\in\mathcal{E}_K} \cdot B_K^{-1}(u_\sigma)_{\sigma\in\mathcal{E}_K}\right)^2 \le \left((1)_{\sigma\in\mathcal{E}_K} \cdot B_K^{-1}(1)_{\sigma\in\mathcal{E}_K}\right) \left((u_\sigma)_{\sigma\in\mathcal{E}_K} \cdot B_K^{-1}(u_\sigma)_{\sigma\in\mathcal{E}_K}\right),$$

which is exactly

$$\left(\sum_{\sigma\in\mathcal{E}_K} b_{K,\sigma} u_\sigma \right)^2 \le b_K \sum_{\sigma\in\mathcal{E}_K} \sum_{\sigma'\in\mathcal{E}_K} (B_K^{-1})_{\sigma\sigma'} u_\sigma u_{\sigma'}.$$

In order to show that M is definite, we simply remark that the preceding reasoning shows that the systems ((30),(33),(34),(35)) and (49) are equivalents. Hence, since ((30),(33),(34),(35)) has a unique solution, so must (49), which means that M is invertible.

We then solve System (49) in the practical implementation of the penalized scheme, using a direct solver. We then compute (u, F) thanks to relations (48) and (47). Moreover, even in the

case of simplicial meshes where the non-penalized scheme could be used, we nevertheless use the penalized scheme in order to obtain the approximate solution using System (49). Note that, in the case of simplicial meshes, letting ν_K tends to 0 leads to a limit of System (49) corresponding to the inversion of each local matrix A_K, which is not the case otherwise since, in the case of a general mesh, A_K can be rectangular. Note that the non-penalized scheme provides a unique solution for $(u, \mathbf{v})$, for which we did not prove the convergence. Nevertheless, we expect that the non-penalized solution is more precise than a significantly penalized one, and therefore we let $\nu_K = 10^{-9}/\mathrm{m}(K)$ for all the following computations (we used a direct method for the inversion of matrices B_K). In all the cases, the points $\mathbf{x}_K$ have been located at the center of gravity of the control volumes.

4.2 Numerical results

All the following numerical results have been obtained for the case $d = 2$, $\Omega = (0,1) \times (0,1)$, $\Lambda = \mathrm{I}_d$ and $\bar{u}(x) = x^{(1)}(1 - x^{(1)})x^{(2)}(1 - x^{(2)})$ for all $x = (x^{(1)}, x^{(2)}) \in \Omega$.

Remark 4.1 *We have also successfully used the scheme for the numerical study of some anisotropic heterogeneous problems. However, we do not present these results here (which are roughly similar to the ones below), preferring for shortness reasons to focus on the application of the scheme to various types of grids.*

We first present in Figure 1 two different simplicial (i.e. triangular) discretizations $\mathcal{D}_{t1}$ and $\mathcal{D}_{t2}$ (in the sense presented above in this paper) used for the computation of an approximate solution for the problem. We also show in Figure 1 the error $e_{\mathcal{D}}$, defined by

$$e_K = \frac{|u_K - \bar{u}(\mathbf{x}_K)|}{\|\bar{u}\|_{L^\infty(\Omega)}}, \quad \forall K \in \mathcal{M},$$

using discretizations $\mathcal{D}_{t1}$ and $\mathcal{D}_{t2}$. Note that these discretizations do not respect the Delaunay condition on a sub-domain of Ω, and that the 4-point finite volume scheme (see [9]) cannot be used on these grids. The grids $\mathcal{D}_{t2}$ and $\mathcal{D}_{t3}$ (which is not represented here) have been obtained from $\mathcal{D}_{t1}$ (containing 400 control volumes) by the respective divisions by 2 and 4 of each edge (there are 1600 control volumes in $\mathcal{D}_{t2}$ and 6400 in $\mathcal{D}_{t3}$). The errors in L^2 norms obtained with these grids are given in the following table.

	$\|u - \bar{u}\|_{L^2(\Omega)}$	$\|\mathbf{v} - \nabla\bar{u}\|_{L^2(\Omega)^d}$
$\mathcal{D}_{t1}$	$5.1\ 10^{-4}$	$1.8\ 10^{-2}$
$\mathcal{D}_{t2}$	$1.9\ 10^{-4}$	$9.0\ 10^{-3}$
$\mathcal{D}_{t3}$	$8.2\ 10^{-5}$	$4.5\ 10^{-3}$
order of convergence	≥ 1	1

We observe that the numerical orders of convergence for $\|u - \bar{u}\|_{L^2(\Omega)}$ and $\|\mathbf{v} - \nabla\bar{u}\|_{L^2(\Omega)^d}$ seem to be equal to 1, and therefore no super-convergence property can reasonably be expected in this case.

We then present in Figure 2 discretizations $\mathcal{D}_{q1}$ and $\mathcal{D}_{q2}$ and error $e_{\mathcal{D}}$ using these grids. Such grids could be obtained using a refinement procedure: for example, in the case of coupled systems, the grid might have been refined in order to improve the convergence on another equation (thanks to some *a posteriori* estimates maybe) and must then be used to solve (1) which is the second

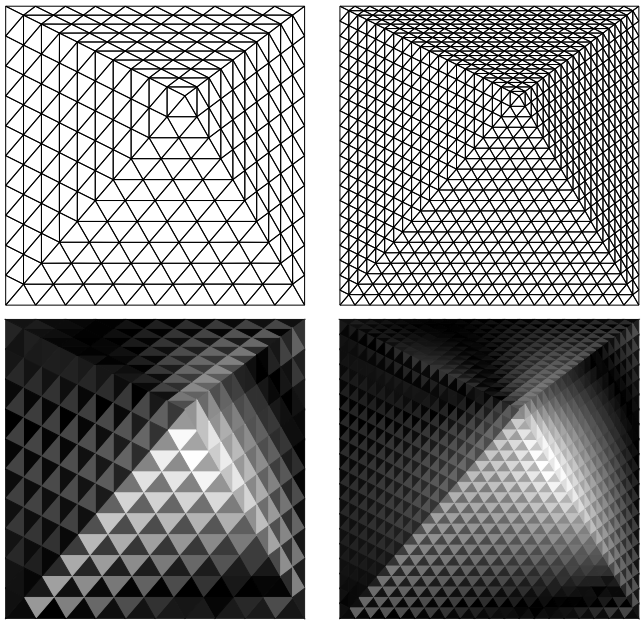

Figure 1: Top: discretizations used for the numerical tests (left: $\mathcal{D}_{t1}$, right: $\mathcal{D}_{t2}$), bottom: error $e_{\mathcal{D}}$ obtained with $\mathcal{D}_{t1}$ (left: black=0, white = 2.2 10^{-2}) and $\mathcal{D}_{t2}$ (right: black=0, white = 8.9 10^{-3}).

part of the system. The grid $\mathcal{D}_{q2}$ has been obtained from $\mathcal{D}_{q1}$ by a uniform division of each edge by 2, and similarly $\mathcal{D}_{q3}$ (not represented here) has been obtained from $\mathcal{D}_{q2}$ in the same way. The respective errors in L^2 norms obtained with these grids are given in the following table.

	$\|u - \bar{u}\|_{L^2(\Omega)}$	$\|\mathbf{v} - \nabla\bar{u}\|_{L^2(\Omega)^d}$
$\mathcal{D}_{q1}$	8.7 10^{-4}	5.8 10^{-3}
$\mathcal{D}_{q2}$	1.7 10^{-4}	1.3 10^{-3}
$\mathcal{D}_{q3}$	3.9 10^{-5}	4.0 10^{-4}
order of convergence	≥ 2	≥ 1

We then observe that the numerical order convergence is better than 2 for $\|u - \bar{u}\|_{L^2(\Omega)}$, which corresponds to a case of a mainly structured grid (there is no significant additional error located at the internal boundaries between the differently gridded subdomains, see Figure 2).
Finally, in Figure 3, we represent grids $\mathcal{D}_\flat$ and $\mathcal{D}_\sharp$ and the error $e_{\mathcal{D}}$ thus obtained. These meshes (which have the same number of control volumes) could correspond to the case of moving meshes (for example, due to a phenomenon of compaction, see [13]). The respective errors in L^2 norms obtained with these grids are given in the following table.

	$\|u - \bar{u}\|_{L^2(\Omega)}$	$\|\mathbf{v} - \nabla\bar{u}\|_{L^2(\Omega)^d}$
$\mathcal{D}_\flat$	2.0 10^{-4}	6.7 10^{-4}
$\mathcal{D}_\sharp$	4.6 10^{-4}	1.8 10^{-3}

We observe that the error is mainly connected to the size of the control volumes, and maybe to some effect of loss of regularity of the mesh.

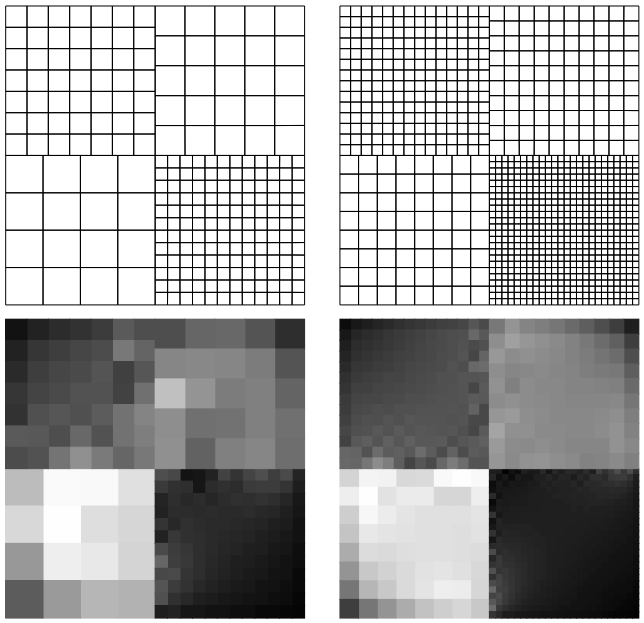

Figure 2: Top: discretizations used for the numerical tests (left: $\mathcal{D}_{q1}$, right: $\mathcal{D}_{q2}$), bottom: error $e_{\mathcal{D}}$ obtained with $\mathcal{D}_{q1}$ (left: black=0, white = 2.7 10^{-2}) and $\mathcal{D}_{q2}$ (right: black=0, white = 5.3 10^{-3}).

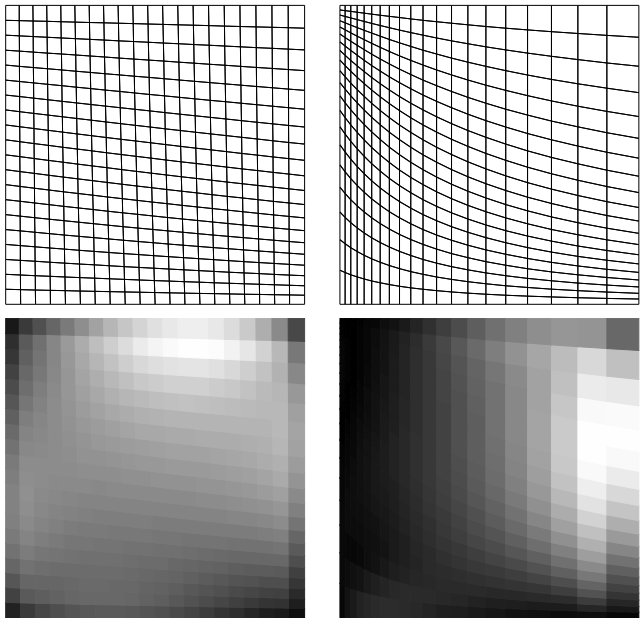

Figure 3: Top: discretizations used for the numerical tests (left: $\mathcal{D}_\flat$, right: $\mathcal{D}_\sharp$), bottom: error $e_{\mathcal{D}}$ obtained with $\mathcal{D}_\flat$ (left: black=0, white = 5.4 10^{-3}) and $\mathcal{D}_\sharp$ (right: black=0, white = 1.5 10^{-2}).

5 Appendix

Lemma 5.1 *Let K be a non empty open convex polygonal set in $\mathbb{R}^d$. For $\sigma \in \mathcal{E}_K$ (the edges of K, in the sense given in Definition 2.1), we let $\mathbf{x}_\sigma$ be the center of gravity of σ; we also denote $\mathbf{n}_{K,\sigma}$ the unit normal to σ outward to K. Then, for all vector $\mathbf{e} \in \mathbb{R}^d$ and for all point $\mathbf{x}_K \in K$, we have*

$$\mathrm{m}(K)\mathbf{e} = \sum_{\sigma \in \mathcal{E}_K} \mathrm{m}(\sigma)\mathbf{e} \cdot \mathbf{n}_{K,\sigma}(\mathbf{x}_\sigma - \mathbf{x}_K).$$

Proof.
We denote by a superscript i the i-th coordinate of vectors and points in $\mathbb{R}^d$. By Stokes formula, we have

$$\mathrm{m}(K)\mathbf{e}^i = \int_K \mathrm{div}((x^i - \mathbf{x}_K^i)\mathbf{e})\,\mathrm{d}x = \sum_{\sigma \in \mathcal{E}_K} \int_\sigma (x^i - \mathbf{x}_K^i)\mathbf{e} \cdot \mathbf{n}_{K,\sigma}\,\mathrm{d}\gamma(x)$$

and the proof is concluded since, by definition of the center of gravity, $\int_\sigma (x^i - \mathbf{x}_K^i)\,\mathrm{d}\gamma(x) = \int_\sigma x^i\,d\gamma(x) - \mathrm{m}(\sigma)\mathbf{x}_K^i = \mathrm{m}(\sigma)\mathbf{x}_\sigma^i - \mathrm{m}(\sigma)\mathbf{x}_K^i$. □

The following lemma is quite similar to Lemma 7.2 in [7], but since we need this result with slightly more general hypotheses than in this reference, we include the full proof for sake of completeness.

Lemma 5.2 *Let K be a non empty open polygonal convex set in $\mathbb{R}^d$. Let E be an affine hyperplane of $\mathbb{R}^d$ and σ be a non empty open subset of E contained in $\partial K \cap E$. We assume that there exists $\alpha > 0$ and $\mathbf{p}_K \in K$ such that $B(\mathbf{p}_K, \alpha\mathrm{diam}(K)) \subset K$. We denote $\triangle_{K,\sigma}$ the convex hull of σ and $\mathbf{p}_K$. Then there exists C_{28} only depending on d and α such that, for all $v \in H^1(K)$,*

$$\left(\frac{1}{\mathrm{m}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}} v(x)\,\mathrm{d}x - \frac{1}{\mathrm{m}(\sigma)}\int_\sigma v(\xi)\,\mathrm{d}\gamma(\xi)\right)^2 \leq \frac{C_{28}\mathrm{dist}(\mathbf{p}_K, E)^2}{\mathrm{m}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}} |\nabla v(x)|^2\,\mathrm{d}x.$$

Proof.
The regular functions being dense in $H^1(K)$ (since K is convex), it is sufficient to prove the lemma for $v \in C^1(\mathbb{R}^N)$. By translation and rotation, we can assume that $E = \{0\} \times \mathbb{R}^{d-1}$, $\sigma = \{0\} \times \widetilde{\sigma}$ with $\widetilde{\sigma} \subset \mathbb{R}^{d-1}$ and that $\mathbf{p}_K = (p_1, 0)$ with $p_1 = \mathrm{dist}(\mathbf{p}_K, E)$.
Notice that, since K is convex and $\partial K \cap E$ contains a non empty open subset of E, K is on one side of E. In particular, $B(\mathbf{p}_K, \alpha\mathrm{diam}(K))$ is also on one side of E (it is contained in K) and

$$p_1 = \mathrm{dist}(\mathbf{p}_K, E) \geq \alpha\mathrm{diam}(K). \tag{50}$$

For $a \in [0, p_1]$, we denote $\widetilde{\sigma}_a = \{z \in \mathbb{R}^{d-1} \mid (a, z) \in \triangle_{K,\sigma}\}$. By definition, $(a, z) \in \triangle_{K,\sigma}$ if and only if there exists $t \in [0, 1]$ and $y \in \widetilde{\sigma}$ such that $t(p_1, 0) + (1-t)(0, y) = (a, z)$; this is equivalent to $t = \frac{a}{p_1}$ and $z = (1-t)y = \left(1 - \frac{a}{p_1}\right) y$. Thus, $\widetilde{\sigma}_a = \left(1 - \frac{a}{p_1}\right)\widetilde{\sigma}$.

For all $y \in \widetilde{\sigma}$ and all $a \in [0, p_1]$, we have

$$v(0, y) - v\left(a, \left(1 - \frac{a}{p_1}\right) y\right) = \int_0^1 \nabla v\left(ta, \left(1 - t\frac{a}{p_1}\right) y\right) \cdot \left(-a, \frac{a}{p_1} y\right)\,\mathrm{d}t.$$

Integrating on $y \in \widetilde{\sigma}$ and using the change of variable $z = \left(1 - \frac{a}{p_1}\right) y$, we find

$$\int_\sigma v(\xi)\,\mathrm{d}\gamma(\xi) - \frac{1}{\left(1-\frac{a}{p_1}\right)^{d-1}} \int_{\widetilde{\sigma}_a} v(a,z)\,\mathrm{d}z = \int_{\widetilde{\sigma}} \int_0^1 \nabla v \left(ta, \left(1 - t\frac{a}{p_1}\right) y\right) \cdot \left(-a, \frac{a}{p_1} y\right) \mathrm{d}t\mathrm{d}y.$$

Multiplying by $\left(1-\frac{a}{p_1}\right)^{d-1}$ and integrating on $a \in [0, p_1]$, we obtain

$$\begin{aligned} &\int_\sigma v(\xi)\,\mathrm{d}\gamma(\xi) \int_0^{p_1} \left(1-\frac{a}{p_1}\right)^{d-1} \mathrm{d}a - \int_0^{p_1} \int_{\widetilde{\sigma}_a} v(a,z)\,\mathrm{d}z\mathrm{d}a \\ &= \int_0^{p_1} \left(1-\frac{a}{p_1}\right)^{d-1} \int_{\widetilde{\sigma}} \int_0^1 \nabla v \left(ta, \left(1 - t\frac{a}{p_1}\right) y\right) \cdot \left(-a, \frac{a}{p_1} y\right) \mathrm{d}t\mathrm{d}y\mathrm{d}a. \end{aligned}$$

But $\int_0^{p_1} \left(1-\frac{a}{p_1}\right)^{d-1} \mathrm{d}a = \frac{p_1}{d}$ and $\mathrm{m}(\triangle_{K,\sigma}) = \frac{\mathrm{m}(\sigma) p_1}{d}$; therefore, dividing by $\mathrm{m}(\triangle_{K,\sigma})$, we find

$$\begin{aligned} &\frac{1}{\mathrm{m}(\sigma)} \int_\sigma v(\xi)\,\mathrm{d}\gamma(\xi) - \frac{1}{\mathrm{m}(\triangle_{K,\sigma})} \int_{\triangle_{K,\sigma}} v(x)\,\mathrm{d}x \\ &= \frac{1}{\mathrm{m}(\triangle_{K,\sigma})} \int_0^{p_1} \left(1-\frac{a}{p_1}\right)^{d-1} \int_{\widetilde{\sigma}} \int_0^1 \nabla v \left(ta, \left(1 - t\frac{a}{p_1}\right) y\right) \cdot \left(-a, \frac{a}{p_1} y\right) \mathrm{d}t\mathrm{d}y\mathrm{d}a. \quad (51) \end{aligned}$$

For all $y \in \widetilde{\sigma}$, we have $|y| = |(0,y)| \le |(0,y) - \mathbf{p}_K| + |\mathbf{p}_K| \le \mathrm{diam}(K) + p_1$ (because $(0,y)$ and $\mathbf{p_K}$ belong to K). By (50), this implies $|y| \le (\frac{1}{\alpha} + 1) p_1$ and thus

$$\begin{aligned} &\left| \int_0^{p_1} \left(1-\frac{a}{p_1}\right)^{d-1} \int_{\widetilde{\sigma}} \int_0^1 \nabla v \left(ta, \left(1 - t\frac{a}{p_1}\right) y\right) \cdot \left(-a, \frac{a}{p_1} y\right) \mathrm{d}t\mathrm{d}y\mathrm{d}a \right| \\ &\le C_{29} \int_0^{p_1} \left(1-\frac{a}{p_1}\right)^{d-1} \int_{\widetilde{\sigma}} \int_0^1 \left| \nabla v \left(ta, \left(1 - t\frac{a}{p_1}\right) y\right) \right| a\,\mathrm{d}t\mathrm{d}y\mathrm{d}a \\ &\le C_{29} \int_0^{p_1} \int_{\widetilde{\sigma}} \int_0^1 \left| \nabla v \left(ta, \left(1 - t\frac{a}{p_1}\right) y\right) \right| a \left(1-\frac{ta}{p_1}\right)^{d-1} \mathrm{d}t\mathrm{d}y\mathrm{d}a \quad (52) \end{aligned}$$

where C_{29} only depends on α (we have used the obvious fact that, for $t \in]0,1[$, $1 - \frac{a}{p_1} \le 1 - \frac{ta}{p_1}$). But, for all $a \in]0, p_1[$, the change of variable

$$\varphi_a : (t,y) \in]0,1[\times \widetilde{\sigma} \to z = \left(ta, \left(1 - t\frac{a}{p_1}\right) y\right) \in \varphi_a(]0,1[\times \widetilde{\sigma})$$

has Jacobian determinant equal to $a\left(1-\frac{ta}{p_1}\right)^{d-1}$ and therefore

$$\int_{\widetilde{\sigma}} \int_0^1 \left| \nabla v \left(ta, \left(1 - t\frac{a}{p_1}\right) y\right) \right| a \left(1-\frac{ta}{p_1}\right)^{d-1} \mathrm{d}t\mathrm{d}y = \int_{\varphi_a(]0,1[\times\widetilde{\sigma})} |\nabla v(z)|\,\mathrm{d}z.$$

Moreover, $(ta, (1-t\frac{a}{p_1})y) = \frac{ta}{p_1}(p_1, 0) + (1 - \frac{ta}{p_1})(0,y)$ with $\frac{ta}{p_1} \in]0,1[$; hence, $\varphi_a(]0,1[\times\widetilde{\sigma}) \subset \triangle_{K,\sigma}$ and we obtain

$$\int_0^{p_1} \int_{\widetilde{\sigma}} \int_0^1 \left| \nabla v \left(ta, \left(1 - t\frac{a}{p_1}\right) y\right) \right| a \left(1-\frac{ta}{p_1}\right)^{d-1} \mathrm{d}t\mathrm{d}y\mathrm{d}a \le p_1 \int_{\triangle_{K,\sigma}} |\nabla v(z)|\,\mathrm{d}z.$$

We introduce this inequality in (52) and use the resulting estimate in (51) to obtain

$$\left|\frac{1}{\mathrm{m}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}} v(x)\,\mathrm{d}x - \frac{1}{\mathrm{m}(\sigma)}\int_\sigma v(\xi)\,\mathrm{d}\gamma(\xi)\right| \le \frac{C_{29}p_1}{\mathrm{m}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}} |\nabla v(x)|\,\mathrm{d}x$$

and the conclusion follows from the Cauchy-Schwarz inequality, recalling that $p_1 = \mathrm{dist}(\mathbf{p}_K, E)$. □

Lemma 5.3 *Let K be a non empty open polygonal convex set in $\mathbb{R}^d$ such that, for some $\alpha > 0$, there exists a ball of radius $\alpha\mathrm{diam}(K)$ contained in K. Let E be an affine hyperplane of $\mathbb{R}^d$ and σ be a non empty open subset of E contained in $\partial K \cap E$. Then there exists C_{30} only depending on d and α such that, for all $v \in H^1(K)$,*

$$\left(\frac{1}{\mathrm{m}(K)}\int_K v(x)\,\mathrm{d}x - \frac{1}{\mathrm{m}(\sigma)}\int_\sigma v(x)\,\mathrm{d}\gamma(x)\right)^2 \le \frac{C_{30}\mathrm{diam}(K)}{\mathrm{m}(\sigma)}\int_K |\nabla v(x)|^2\,\mathrm{d}x.$$

Proof.
Let $B(\mathbf{p}_K, \alpha\mathrm{diam}(K)) \subset K$ and $\triangle_{K,\sigma}$ be the convex hull of $\mathbf{p}_K$ and σ. By Lemma 5.2, we have

$$\left(\frac{1}{\mathrm{m}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}} v(x)\,\mathrm{d}x - \frac{1}{\mathrm{m}(\sigma)}\int_\sigma v(x)\,\mathrm{d}\gamma(x)\right)^2 \le \frac{C_{28}\mathrm{dist}(\mathbf{p}_K, E)^2}{\mathrm{m}(\triangle_{K,\sigma})}\int_K |\nabla v(x)|^2\,\mathrm{d}x.$$

But $\mathrm{m}(\triangle_{K,\sigma}) = \frac{\mathrm{m}(\sigma)\mathrm{dist}(\mathbf{p}_K,E)}{d}$ and $\mathrm{dist}(\mathbf{p}_K, E) \le \mathrm{dist}(\mathbf{p}_K, \sigma) \le \mathrm{diam}(K)$. Therefore,

$$\left(\frac{1}{\mathrm{m}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}} v(x)\,\mathrm{d}x - \frac{1}{\mathrm{m}(\sigma)}\int_\sigma v(x)\,\mathrm{d}\gamma(x)\right)^2 \le \frac{C_{28}d\,\mathrm{diam}(K)}{\mathrm{m}(\sigma)}\int_K |\nabla v(x)|^2\,\mathrm{d}x. \tag{53}$$

Using Lemma 7.1 in [7], we get C_{31} only depending on d such that

$$\left(\frac{1}{\mathrm{m}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}} v(x)\,\mathrm{d}x - \frac{1}{\mathrm{m}(K)}\int_K v(x)\,\mathrm{d}x\right)^2 \le \frac{C_{31}\mathrm{diam}(K)^{d+2}}{\mathrm{m}(\triangle_{K,\sigma})\mathrm{m}(K)}\int_K |\nabla v(x)|^2\,\mathrm{d}x,$$

which implies

$$\left(\frac{1}{\mathrm{m}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}} v(x)\,\mathrm{d}x - \frac{1}{\mathrm{m}(K)}\int_K v(x)\,\mathrm{d}x\right)^2 \le \frac{C_{31}d\,\mathrm{diam}(K)^{d+2}}{\mathrm{m}(\sigma)\mathrm{dist}(\mathbf{p}_K,E)\mathrm{m}(K)}\int_K |\nabla v(x)|^2\,\mathrm{d}x.$$

But, as in the proof of Lemma 5.2, we have $\mathrm{dist}(\mathbf{p}_K, E) \ge \alpha\mathrm{diam}(K)$ (see (50)). Since $\mathrm{m}(K) \ge \omega_d\alpha^d\mathrm{diam}(K)^d$, we deduce that

$$\left(\frac{1}{\mathrm{m}(\triangle_{K,\sigma})}\int_{\triangle_{K,\sigma}} v(x)\,\mathrm{d}x - \frac{1}{\mathrm{m}(K)}\int_K v(x)\,\mathrm{d}x\right)^2 \le \frac{C_{31}d\,\mathrm{diam}(K)}{\omega_d\alpha^{d+1}\mathrm{m}(\sigma)}\int_K |\nabla v(x)|^2\,\mathrm{d}x. \tag{54}$$

The lemma follows from (53) and (54). □

References

[1] I. Aavatsmark, An introduction to multipoint flux approximations for quadrilateral grids. Locally conservative numerical methods for flow in porous media. Comput. Geosci. 6, 405–432 (2002).

[2] I. Aavatsmark, T. Barkve, O. Boe and T. Mannseth, Discretization on unstructured grids for inhomogeneous, anisotropic media. Part I: Derivation of the methods. Journal on Scientific Computing, 19, 1700–1716 (1998).

[3] I. Aavatsmark, T. Barkve, O. Boe and T. Mannseth, Discretization on unstructured grids for inhomogeneous, anisotropic media. Part II: Discussion and numerical results. SIAM Journal on Scientific Computing, 19, 1717–1736 (1998).

[4] T. Arbogast, L.C. Cowsar, M.F. Wheeler and I. Yotov, Mixed finite element methods on nonmatching multiblock grids. SIAM J. Numer. Anal. 37, No.4, 1295–1315 (2000).

[5] T. Arbogast, M.F. Wheeler and I. Yotov, Mixed finite elements for elliptic problems with tensor coefficients as cell-centered finite differences. SIAM J. Numer. Anal. 34, No.2, 828–852 (1997).

[6] G. Chavent, G. Cohen and J. Jaffré, Discontinuous upwinding and mixed finite elements for two-phase flows in reservoir simulation. Comput. Methods Appl. Mech. Eng. 47, 93–118 (1984).

[7] J. Droniou, Error estimates for the convergence of a finite volume discretization of convection-diffusion equations. J. Numer. Math. 11, 1–32 (2003).

[8] J. Droniou, R. Eymard, D. Hilhorst and X. D. Zhou, Convergence of a finite volume - mixed finite element method for a system of a hyperbolic and an elliptic equations. IMA Journal of Numerical Analysis 23, 507–538 (2003).

[9] R. Eymard, T. Gallouët and R. Herbin, Finite Volume Methods. Handbook of Numerical Analysis, Edited by P.G. Ciarlet and J.L. Lions, North Holland 7, 713–1020 (2000).

[10] R. Eymard, T. Gallouët and R. Herbin, A finite volume for anisotropic diffusion problems. Comptes Rendus de l'Académie des Sciences 339, 299–302 (2004).

[11] R. Eymard, T. Gallouët and R. Herbin, A cell-centered finite volume approximation for nondiagonal second order partial derivative operators on 2 or 3D unstructured meshes. submitted, (2005).

[12] R. Eymard, T. Gallouët and R. Herbin, Finite volume approximation of elliptic problems and convergence of an approximate gradient. Appl. Num. Math. 37, 31–53 (2001).

[13] I. Faille, A control volume method to solve an elliptic equation on a two- dimensional irregular mesh. Comput. Methods Appl. Mech. Eng. 100, 275–290 (1992).

[14] D.S. Kershaw, Differencing of the diffusion equation in Lagrangian hydrodynamic codes. J. Comput. Phys. 39, 375–395 (1981).

[15] K. Lipnikov, J. Morel and M. Shashkov, Mimetic finite difference methods for diffusion equations on non-orthogonal non-conformal meshes. (English) J. Comput. Phys. 199, 589–597 (2004).

[16] P.A. Raviart and J.M. Thomas, A mixed finite element method for 2nd order elliptic problems. Math. Aspects Finite Elem. Meth., Proc. Conf. Rome 1975, Lect. Notes Math. 606, 292–315 (1977).

[17] J.E. Roberts and J.M. Thomas,M. Mixed and hybrid methods. Ciarlet, P. G. (ed.) et al. Handbook of numerical analysis. North-Holland 2, 523–639 (1991).

[18] A. Younes, P. Ackerer and G. Chavent, From mixed finite elements to finite volumes for elliptic PDEs in two and three dimensions. Int. J. Numer. Methods Eng. 59, 365-388 (2004).