The G method for heterogeneous anisotropic diffusion on general meshes

Léo Agélas, Daniele Antonio Di Pietro, Jérôme Droniou

The G method for heterogeneous anisotropic diffusion on general meshes

Léo Agélas* and Daniele A. Di Pietro†

IFP, 1 & 4 av. du Bois-Préau 92852 Rueil-Malmaison Cedex (France)

Jérôme Droniou‡

Université Montpellier 2, Institut de Mathématiques et Modélisation de Montpellier, CC 051, Place Eugène Bataillon 34095 Montpellier Cedex 05 (France)

November 27, 2008

Abstract

In the present work we introduce a new family of cell-centered Finite Volume schemes for anisotropic and heterogeneous diffusion operators inspired by the MPFA L method. A very general framework for the convergence study of finite volume methods is provided and then used to establish the convergence of the new method. Fairly general meshes are covered and a computable coercivity criterion is provided. In order to guarantee consistency in the presence of heterogeneous diffusivity, we introduce a non-standard test space in $H_0^1(\Omega)$ and prove its density. Thorough assessment on a set of anisotropic heterogeneous problems as well as a comparison with classical multi-point Finite Volume methods is provided.

Keywords Finite volume, heterogeneous anisotropic diffusion, MPFA, convergence analysis

1 Introduction

One of the key ingredients for the numerical solution of Darcy equations is the discretization of anisotropic heterogeneous elliptic terms [8]. In the oil industry, the need to improve accuracy in near wellbore regions has prompted the introduction of general unstructured meshes and full permeability tensors. Significant mathematical effort has therefore been devoted to find consistent and robust Finite Volume (FV) discretizations of anisotropic heterogeneous elliptic terms on general meshes. Ideally, a method should (i) be consistent and coercive on general polyhedral meshes as well as robust with respect to the anisotropy and heterogeneity of the permeability tensor; (ii) yield well-conditioned linear systems for which optimal preconditioning strategies can be devised; (iii) have a narrow stencil, both to improve matrix sparsity and to reduce the communication in parallel implementations. The last requirement would speak in favour of cell-centered methods. However, at present time, no unconditionally coercive and consistent compact stencil cell-centered method has been found. Indeed, although several symmetric methods display unconditional coercivity, they either entail severe mesh restrictions, as in [3], or exhibit very large stencils, as in [23, 6].

The so-called Multi Point Flux Approximation (MPFA) methods have been introduced in the middle of the 90s (see, e.g., [2, 21]). The key idea is to obtain consistency on general meshes

*leo.agelas@ifp.fr

†daniele-antonio.di-pietro@ifp.fr

‡droniou@math.univ-montp2.fr

at the expense of a larger stencil while preserving the second order convergence of the classical two-point method. As mentioned, however, coercivity holds only under suitable conditions on both the mesh and the permeability tensor. The compact stencil MPFA L method has been proposed in [5,4] as an improvement of the MPFA O method of [1] both in terms of stencil size and monotonicity properties. The convergence of the MPFA O method has been theoretically proved in [3] on two-dimensional quadrilateral grids and in [10, 9] on general two- and three-dimensional polyhedral meshes. In [26], the equivalence of multi-point methods with the lowest-order Mixed Finite Element method on matching triangular grids has been pointed out, and local coercivity conditions have been proposed. Other relatively inexpensive methods that deserve being mentioned are those developed in the Mimetic Finite Difference framework of [16, 14, 15], as well as the Hybrid Finite Volume scheme of [24] or the Mixed Finite Volume scheme of [19]. The analogies among the three classes of methods have been recently pointed out in [20]. Finally, a unified framework covering both FV and discontinuous Galerkin methods expressed in weak form has recently been introduced in [7] relying on the discrete functional analysis results of [25, 17].

In this work we propose a family of cell-centered schemes generalizing the MPFA L method. The idea is to write the flux through a face as the weighted average of several L-type fluxes corresponding to different stencils. A proper choice of the weights allows to enhance the coercivity of the method, thereby improving robustness with respect to the skewdness of the mesh and to the anisotropy and heterogeneity of the permeability tensor. The provided convergence proof covers more general FV schemes expressed in terms of numerical fluxes and it is inspired by [22, 23]. The relevant requirements are weak flux consistency and coercivity. Convergence is then obtained using the discrete Sobolev embeddings and Rellich theorem proved in [25, §5]. Unlike in [7], where methods in weak formulation are considered, we focus here on the more classical FV flux formulation. The interest of flux formulation is that (i) it provides a natural means to implement new methods in traditional two-point FV codes; (ii) it is more natural for a number of multi-points methods and (iii) it allows to further reduce the set of requirements for convergence (flux continuity, e.g., is not needed).

From a practical viewpoint, the proposed method is a good compromise between accuracy, robustness and computational cost. Indeed, the methods of [16,14,15,24,20] require the introduction of additional face unknowns whose local elimination in terms of cell unknowns is, in general, not possible. While the resulting stability properties are highly appreciable, the increase in computational cost is not affordable in large industrial simulations. Unconditionally coercive cell-centered methods like the ones of [24, §2.2] or of [6] have stencils extending to neighbours of neighbours, which results in denser matrices and stronger memory requirements. Also, in parallel implementations, two layers of ghost cells are needed to ensure communications among subdomains, resulting in heavily penalized scalability (message passing is still considered a bottle-neck when it comes to large industrial cases). More compact methods like the MPFA L method have up to now been based on (sophisticated) euristics rather than on a extensive mathematical analysis. To the best of our knowledge, the present work contains the first convergence proof for the MPFA L method for general meshes and arbitrary heterogeneous anisotropic diffusion tensors. The aim of this paper is also to identify a minimal set of requirements for convergence and investigate the benefits of a deeper mathematical comprehension. As a matter of fact, the resulting MPFA G method outperforms the original version of [5, 4] on a number of representative test cases modeling some of the difficulties encountered in industrial simulations.

In order to avoid artificial regularity assumptions on the permeability tensor in the consistency proof, we have introduced a permeability dependent test space $\mathcal{Q}$ composed of continuous functions with possibly discontinuous gradients but continuous fluxes. This space is proved to be dense in $H_0^1(\Omega)$ following the ideas of [18]. To the best of our knowledge, the idea of selecting a problem-taylored test space as well as the density proof are new.

The work is organized as follows: in §2 we present a general convergence study based on a minimal set of requirements; the analysis applies to fairly general FV methods expressed in terms of numerical fluxes. A forthcoming work will be devoted to showing the extents of such analysis framework. In §3 we present the G method, a generalization of the MPFA L scheme, and show that it fits in the analysis framework of §2; §4 is devoted to numerical tests. The performances of

the proposed method are evaluated against anisotropic and heterogeneous benchmarks on general meshes. A comparison with the MPFA O and L methods as well as with the Success scheme of [24, §2.2] is provided.

2 Abstract framework

2.1 Model problem

Let $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}^*$, be an open bounded connected polygonal domain with boundary $\partial\Omega$ and let $P_\Omega \stackrel{\text{def}}{=} \{\Omega_i\}_{i=1\dots N_\Omega}$ denote a finite partition of Ω into open connected disjoint polygonal subsets. Let Λ be a symmetric tensor-valued function such that (s.t.) (i) $\Lambda_{|\Omega_i} \in [C^2(\overline{\Omega_i})]^{d\times d}$ for all $i = 1\dots N_\Omega$ and (ii) there exists $0 < \alpha_0 < \beta_0$ s.t., for almost every (a.e.) $x \in \overline{\Omega}$, the spectrum of $\Lambda(x)$ is contained in $[\alpha_0, \beta_0]$. Consider the following problem:

$$\begin{cases} \nabla\cdot(-\Lambda\ \nabla\overline{u}) = f & \text{in } \Omega, \\ \overline{u} = 0 & \text{on } \partial\Omega, \end{cases} \tag{1}$$

where $f \in L^r(\Omega)$ with $r > 1$ if $d = 2$ and $r = \frac{2d}{d+2}$ if $d > 2$. The existence and uniqueness of a weak solution $\overline{u} \in H_0^1(\Omega)$ to (1) is a classical result.

Remark 1. Other standard types of boundary conditions can be considered. However, for easiness of presentation, we have preferred to stick to the simpler homogeneous Dirichlet case.

In what follows, we shall provide the definition of a FV discretization of problem (1) as well as an analysis framework covering fairly general (possibly nonconforming) polygonal meshes.

Definition 1 (Admissible family of discretizations). *An admissible family of finite volume discretizations $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ is a triplet $\mathcal{D}_n = (\mathcal{T}_n, \mathcal{E}_n, \mathcal{P}_n)$, where*

(i) $\mathcal{T}_n$ is a finite family of non-empty connected open disjoint subsets of Ω (the cells *or* control volumes*) s.t. $\overline{\Omega} = \cup_{K\in\mathcal{T}_n}\overline{K}$ and $\mathcal{T}_n$ is compatible with P_Ω (each cell is contained in one element of the partition P_Ω). For all $K \in \mathcal{T}_n$, we denote by $\mathrm{m}_K > 0$ its d-dimensional measure (the* volume*) and let $\partial K \stackrel{\text{def}}{=} \overline{K} \setminus K$;*

(ii) $\mathcal{E}_n$ is a finite family of subsets of $\overline{\Omega}$ (the faces*) s.t., for all $\sigma \in \mathcal{E}_n$, σ is a non-empty closed subset of a hyperplane of $\mathbb{R}^d$ with $(d-1)$-dimensional measure $\mathrm{m}_\sigma > 0$ (the* area*), and s.t. the intersection of two different faces has zero $(d-1)$-dimensional measure. We assume that, for all $K \in \mathcal{T}_n$, there exists a subset $\mathcal{E}_K$ of $\mathcal{E}_n$ such that $\partial K = \cup_{\sigma\in\mathcal{E}_K}\sigma$. For all $\sigma \in \mathcal{E}_n$, either $\mathcal{T}_\sigma \stackrel{\text{def}}{=} \{K \in \mathcal{T}_n \,|\, \sigma \in \mathcal{E}_K\}$ has exactly one element and then $\sigma \subset \partial\Omega$ (*boundary face*) or $\mathcal{T}_\sigma$ has exactly two elements (*inner face*); the sets of inner and boundary faces are denoted by $\mathcal{E}_{n,\mathrm{int}}$ and $\mathcal{E}_{n,\mathrm{ext}}$ respectively;*

(iii) $\mathcal{P}_n = (x_K)_{K\in\mathcal{T}_n}$ is a family of points of Ω indexed by $\mathcal{T}_n$ (the cell centers*) s.t. $x_K \in K$ and K is star-shaped with respect to x_K. For all $K \in \mathcal{T}_n$ and for all $\sigma \in \mathcal{E}_K$ we denote by $d_{K,\sigma}$ the Euclidean distance between x_K and the hyperplane supporting σ, and suppose that there exist $0 < \varrho_1 < +\infty$ and $0 < \varrho_2 < +\infty$ independent of n s.t.*

$$\min_{K\in\mathcal{T}_n,\,\sigma\in\mathcal{E}_K} \frac{d_{K,\sigma}}{\mathrm{diam}(K)} \geq \varrho_1, \quad \min_{\sigma\in\mathcal{E}_{n,\mathrm{int}},\,\mathcal{T}_\sigma=\{K,L\}} \frac{\min(d_{K,\sigma}, d_{L,\sigma})}{\max(d_{K,\sigma}, d_{L,\sigma})} \geq \varrho_2. \tag{2}$$

We notice that, by items (ii) and (iii), and since $\frac{\mathrm{m}_\sigma\, d_{K,\sigma}}{d}$ is the measure of the convex hull $\triangle_{K,\sigma}$ of x_K and σ (see Figure 1),

$$\forall K \in \mathcal{T}_n, \quad \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}_\sigma\, d_{K,\sigma} = d\,\mathrm{m}_K\,. \tag{3}$$

The size of the discretization is defined by $h_{\mathcal{D}_n} \stackrel{\text{def}}{=} \sup_{K\in\mathcal{T}_n}\mathrm{diam}(K)$. For all $K \in \mathcal{T}_n$ and $\sigma \in \mathcal{E}_K$, we denote by $\mathbf{n}_{K,\sigma}$ the unit vector normal to σ outward to K. For all $K \in \mathcal{T}_n$, we set

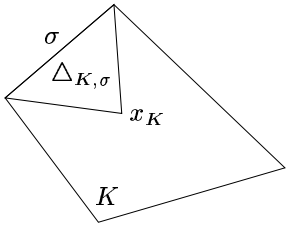


Figure 1: The pyramid convex hull of x_K and σ for $d = 2$.

$\Lambda_K \stackrel{\text{def}}{=} \frac{1}{\mathrm{m}_K} \int_K \Lambda(x)\,\mathrm{d}x$. For all vectors $x \in \mathbb{R}^n$, $n \in \mathbb{N}^*$, the Euclidean norm will be denoted by $|x| \stackrel{\text{def}}{=} \sqrt{x \cdot x}$; for all matrices $A \in \mathbb{R}^n \times \mathbb{R}^n$, $n \in \mathbb{N}^*$, we shall denote by $|A|$ the norm induced by the vector scalar product, i.e., $|A| \stackrel{\text{def}}{=} \sup_{x \in \mathbb{R}^d} \frac{|Ax|}{|x|}$. The vector space of bounded linear operators from E to F will be denoted by $\mathcal{L}(E;F)$.

In what follows, when referring to a generic element $\mathcal{D}_n$ of an admissible family of discretizations $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$, the subscript n will be dropped for easiness of reading if no ambiguity arises. The space of piecewise constant functions on $\mathcal{T}$ is defined as

$$H_{\mathcal{T}}(\Omega) \stackrel{\text{def}}{=} \{v \in L^2(\Omega) \,|\, v_K \stackrel{\text{def}}{=} v_{|K} \in \mathbb{P}^0(K),\ \forall K \in \mathcal{T}\}.$$

For all $\sigma \in \mathcal{E}$, let $I_\sigma \in \mathcal{L}(H_{\mathcal{T}}(\Omega); \mathbb{P}^0(\sigma))$ denote a trace reconstruction operator s.t., for all $v \in H_{\mathcal{T}}(\Omega)$, $I_\sigma v = 0$ if $\sigma \in \mathcal{E}_{\text{ext}}$. The space $H_{\mathcal{T}}(\Omega)$ is endowed with the following norm:

$$||v||_{\mathcal{T},I} \stackrel{\text{def}}{=} \left(\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \frac{\mathrm{m}_\sigma}{d_{K,\sigma}} |I_\sigma v - v_K|^2 \right)^{1/2}.$$

Remark 2. Let $\gamma_\sigma \in \mathcal{L}(H_{\mathcal{T}}(\Omega); \mathbb{P}^0(\sigma))$ be s.t.

$$\forall v \in H_{\mathcal{T}}(\Omega), \quad \begin{cases} \frac{\gamma_\sigma v - v_K}{d_{K,\sigma}} + \frac{\gamma_\sigma v - v_L}{d_{L,\sigma}} = 0 & \text{if } \sigma \in \mathcal{E}_{\text{int}} \text{ with } \mathcal{T}_\sigma = \{K,L\}, \\ \gamma_\sigma v = 0 & \text{if } \sigma \in \mathcal{E}_{\text{ext}}. \end{cases}$$

Then, for all $\widetilde{I}_\sigma \in \mathcal{L}(H_{\mathcal{T}}(\Omega); \mathbb{P}^0(\sigma))$ s.t., for all $v \in H_{\mathcal{T}}(\Omega)$, $\widetilde{I}_\sigma v = 0$ if $\sigma \in \mathcal{E}_{\text{ext}}$,

$$\forall v \in H_{\mathcal{T}}(\Omega), \quad \|v\|_{\mathcal{T},\gamma} \le \|v\|_{\mathcal{T},\widetilde{I}}. \tag{4}$$

Setting, for $\sigma \in \mathcal{E}_{\text{int}}$ with $\mathcal{T}_\sigma = \{K,L\}$, $g_\sigma(y) = \frac{\mathrm{m}_\sigma}{d_{K,\sigma}}|y - v_K|^2 + \frac{\mathrm{m}_\sigma}{d_{L,\sigma}}|y - v_L|^2$, Equation (4) is trivial by noticing that $\gamma_\sigma v$ minimizes g_σ and that $\|v\|^2_{\mathcal{T},\gamma} = \sum_{\sigma\in\mathcal{E}_{\text{int}}} g_\sigma(\gamma_\sigma v) + \sum_{\sigma\in\mathcal{E}_{\text{ext}},\mathcal{T}_\sigma=\{K\}} \frac{\mathrm{m}_\sigma}{d_{K,\sigma}}|v_K|^2$.

In view of Remark 2 and of the special nature of γ_σ, the abridged notation $\|\cdot\|_{\mathcal{T}}$ will be used for $\|\cdot\|_{\mathcal{T},\gamma}$ whenever possible. For all $K \in \mathcal{T}$ and for all $\sigma \in \mathcal{E}_K$, let $F_{K,\sigma} \in \mathcal{L}(H_{\mathcal{T}}(\Omega); \mathbb{P}^0(\sigma))$ be a numerical flux function meant to approximate the diffusive flux flowing out K through σ. For all $(u,v) \in [H_{\mathcal{T}}(\Omega)]^2$, define the bilinear form

$$a_{\mathcal{T}}(u,v) \stackrel{\text{def}}{=} \sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(u)(I_\sigma v - v_K).$$

In what follows, we shall consider discretizations of (1) of the form

$$\text{Find } u \in H_{\mathcal{T}}(\Omega) \text{ s.t. } a_{\mathcal{T}}(u,v) = \int_\Omega fv \text{ for all } v \in H_{\mathcal{T}}(\Omega)\ . \tag{5}$$

2.2 Convergence analysis

We introduce the discrete gradient reconstruction $\widetilde{\nabla}_{\mathcal{D}} \in \mathcal{L}(H_{\mathcal{T}}(\Omega);[H_{\mathcal{T}}(\Omega)]^d)$ s.t., for all $K \in \mathcal{T}$ and all $v \in H_{\mathcal{T}}(\Omega)$,

$$(\widetilde{\nabla}_{\mathcal{D}} v)_K \stackrel{\text{def}}{=} \widetilde{\nabla}_{\mathcal{D}} v_{|K} = \frac{1}{\mathrm{m}_K} \sum_{\sigma \in \mathcal{E}_K} \mathrm{m}_\sigma (I_\sigma v - v_K) \mathbf{n}_{K,\sigma}. \tag{6}$$

Equation (3) together with Cauchy-Schwarz inequality yield

$$\|\widetilde{\nabla}_{\mathcal{D}} v\|_{[L^2(\Omega)]^d} \le \sqrt{d} |v|_{\mathcal{T},I} \quad \forall v \in H_{\mathcal{T}}(\Omega). \tag{7}$$

The following result has been proved in [23, §5]:

Lemma 1 (Discrete Sobolev embeddings). *Let $\mathcal{D}$ be an element of a family of discretizations matching Definition 1. Let $q \in [1,+\infty)$ if $d = 2$, and $q \in [1, 2d/(d-2)]$ if $d > 2$. Then, there exists a constant $C_1 > 0$, depending only on Ω, q, ϱ_1 and ϱ_2 s.t.*

$$\|u\|_{L^q(\Omega)} \le C_1 |u|_{\mathcal{T}} \quad \forall u \in H_{\mathcal{T}}(\Omega).$$

Owing to Remark 2, the following theorem can easily be deduced from (7) and the technique of proof of [23, Lemmata 5.6–5.7]:

Lemma 2 (Discrete Rellich theorem). *Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be a sequence of admissible discretizations matching Definition 1 and s.t. $h_{\mathcal{D}_n} \to 0$ as $n \to \infty$, and let $\{v_n\}_{n\in\mathbb{N}}$ be a sequence of $H_{\mathcal{T}_n}(\Omega)$ s.t. there exists $C > 0$ with $|v_n|_{\mathcal{T}_n,I} \le C$ for all $n \in \mathbb{N}$. Then, there exist a subsequence of $\{v_n\}_{n\in\mathbb{N}}$ and a function $\widetilde{v} \in H_0^1(\Omega)$ s.t., as $n \to \infty$, (i) $v_n \to \widetilde{v}$ in $L^q(\Omega)$ for all $q \in [1, 2d/(d-2))$ (and weakly in $L^{2d/(d-2)}(\Omega)$ if $d > 2$); (ii) $\{\widetilde{\nabla}_{\mathcal{D}_n} v_n\}_{n\in\mathbb{N}}$ weakly converges to $\nabla\widetilde{v}$ in $[L^2(\Omega)]^d$.*

The assumptions yielding convergence of the finite volume scheme are gathered in the following

Hypothesis 1. *Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be a family of discretizations matching Definition 1 and s.t. $h_{\mathcal{D}_n} \to 0$ as $n \to \infty$. We suppose that*

(P1) $\mathfrak{D}$ is a dense subspace of $H_0^1(\Omega)$ s.t. $\mathfrak{D} \subset C_0(\overline{\Omega}) \cap C^2(\overline{\Omega_i})$, $i = 1 \dots N_\Omega$, $C_0(\overline{\Omega})$ being the space of continuous functions which vanish on $\partial\Omega$. For all $\varphi \in \mathfrak{D}$, we denote by $\varphi_{\mathcal{T}_n}$ the element of $H_{\mathcal{T}_n}(\Omega)$ defined as follows: For all $K \in \mathcal{T}_n$, $\varphi_{\mathcal{T}_n|K} = \varphi(x_K)$;

(P2) $a_{\mathcal{T}_n}$ is uniformly coercive, i.e., there is $0 < \gamma_1 < +\infty$ independent of n s.t.

$$\forall v \in H_{\mathcal{T}_n}(\Omega), \quad a_{\mathcal{T}_n}(v,v) \ge \gamma_1 |v|^2_{\mathcal{T}_n,I};$$

(P3) the numerical fluxes are weakly consistent on $\mathfrak{D}$, i.e., for all $\varphi \in \mathfrak{D}$,

$$\epsilon_{\mathcal{D}_n}(\varphi) \stackrel{\text{def}}{=} \left(\sum_{K\in\mathcal{T}_n} \sum_{\sigma\in\mathcal{E}_K} \frac{d_{K,\sigma}}{\mathrm{m}_\sigma} \left| F_{K,\sigma}(\varphi_{\mathcal{T}_n}) - \mathrm{m}_\sigma \frac{\int_K \Lambda(x)\nabla\varphi(x)\,\mathrm{d}x}{\mathrm{m}_K} \cdot \mathbf{n}_{K,\sigma} \right|^2 \right)^{\frac{1}{2}} \to 0 \text{ as } n \to \infty.$$

The non standard choice for $\mathfrak{D}$ proposed in §3 allows to weaken the regularity assumptions on Λ with respect to the classical choice $\mathfrak{D} = C_c^\infty(\Omega)$.

Remark 3. Owing to (3), Property (P3) holds for strongly consistent numerical fluxes, i.e. fluxes s.t., for all $\varphi \in \mathfrak{D}$, there is $0 < C_2 < +\infty$ independent of n s.t.

$$\forall K \in \mathcal{T}_n,\ \forall \sigma \in \mathcal{E}_K, \quad \left| F_{K,\sigma}(\varphi_{\mathcal{T}_n}) - \mathrm{m}_\sigma \frac{\int_K \Lambda(x)\nabla\varphi(x)\,\mathrm{d}x}{\mathrm{m}_K} \cdot \mathbf{n}_{K,\sigma} \right| \le C_2\, \mathrm{m}_\sigma\, h_{\mathcal{D}_n}. \tag{8}$$

Remark 4. Finite Volume methods are usually conservative, i.e., for all $v \in H_{\mathcal{T}_n}(\Omega)$ and all $\sigma \in \mathcal{E}_{n,\mathrm{int}}$ with $\mathcal{T}_\sigma = \{K, L\}$, $F_{K,\sigma}(v) + F_{L,\sigma}(v) = 0$. This property is not required to prove Theorem 1 below. However, it is usually needed in the proof of (P2) or even in the definition of

the numerical fluxes. When conservativity holds, problem (5) is equivalent to the (more classical) FV formulation:

$$\text{Find } u \in H_{\mathcal{T}}(\Omega) \text{ s.t. } -\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(u) = \int_K f(x)\,\mathrm{d}x \text{ for all } K \in \mathcal{T},$$

and the bilinear form $a_{\mathcal{T}}$ does not depend on the choice of the trace operators $\{I_\sigma\}_{\sigma\in\mathcal{E}}$.

Proposition 1 (Asymptotic stability of the interpolator). *Under Hypothesis 1, for all* $\varphi \in \mathfrak{D}$,

$$\|\varphi_{\mathcal{T}}\|_{\mathcal{T},I} \le \frac{1}{\gamma_1}\left(\epsilon_{\mathcal{D}}(\varphi) + \beta_0\sqrt{d}|\varphi|_{H^1(\Omega)}\right).$$

Proof. Owing to (6), for all $v \in H_{\mathcal{T}}(\Omega)$, the following integration by parts formula holds:

$$\begin{aligned}
&\sum_{K\in\mathcal{T}_n}\sum_{\sigma\in\mathcal{E}_K} \mathrm{m}_\sigma \frac{\int_K \Lambda(x)\nabla\varphi(x)\,\mathrm{d}x}{\mathrm{m}_K}\cdot\mathbf{n}_{K,\sigma}(I_\sigma v - v_K)\\
&\quad= \sum_{K\in\mathcal{T}_n}\int_K \Lambda(x)\nabla\varphi(x)\,\mathrm{d}x\cdot\left(\frac{1}{\mathrm{m}_K}\sum_{\sigma\in\mathcal{E}_K}\mathrm{m}_\sigma\,\mathbf{n}_{K,\sigma}(I_\sigma v - v_K)\right) \qquad (9)\\
&\quad= \sum_{K\in\mathcal{T}_n}\int_K \Lambda(x)\nabla\varphi(x)\,\mathrm{d}x\cdot(\widetilde{\nabla}_{\mathcal{D}}v)_K = \int_\Omega \Lambda(x)\nabla\varphi(x)\cdot\widetilde{\nabla}_{\mathcal{D}}v(x)\,\mathrm{d}x.
\end{aligned}$$

The above result together with (P2), Cauchy-Schwarz inequality and (7) yield

$$\begin{aligned}
\gamma_1\|\varphi_{\mathcal{T}}\|^2_{\mathcal{T},I} &\le a_{\mathcal{T}}(\varphi_{\mathcal{T}},\varphi_{\mathcal{T}})\\
&= \sum_{K\in\mathcal{T}}\sum_{\sigma\in\mathcal{E}_K}\sqrt{\frac{d_{K,\sigma}}{\mathrm{m}_\sigma}}\left[F_{K,\sigma}(\varphi_{\mathcal{T}}) - \mathrm{m}_\sigma\frac{\int_K\Lambda(x)\nabla\varphi(x)\,\mathrm{d}x}{\mathrm{m}_K}\cdot\mathbf{n}_{K,\sigma}\right]\sqrt{\frac{\mathrm{m}_\sigma}{d_{K,\sigma}}}(I_\sigma\varphi_{\mathcal{T}} - \varphi_K)\\
&\quad+\int_\Omega \Lambda(x)\nabla\varphi(x)\cdot\widetilde{\nabla}_{\mathcal{D}}\varphi_{\mathcal{T}}(x)\,\mathrm{d}x\\
&\le \epsilon_{\mathcal{D}}(\varphi)\|\varphi_{\mathcal{T}}\|_{\mathcal{T},I} + \beta_0|\varphi|_{H^1(\Omega)}\|\widetilde{\nabla}\varphi_{\mathcal{T}}\|_{[L^2(\Omega)]^d} \le \left(\epsilon_{\mathcal{D}}(\varphi) + \beta_0\sqrt{d}|\varphi|_{H^1(\Omega)}\right)\|\varphi_{\mathcal{T}}\|_{\mathcal{T},I}. \quad \square
\end{aligned}$$

Lemma 3 (Uniform *a priori* estimate). *Assume that Hypothesis 1 holds. Then, problem* (5) *is well-posed for each* $n \in \mathbb{N}$, *and the solutions* $u_n \in H_{\mathcal{D}_n}(\Omega)$ *satisfy the following uniform* a priori *estimate:*

$$\|u_n\|_{\mathcal{T}_n,I} \le \frac{C_1}{\gamma_1}\|f\|_{L^r(\Omega)}. \qquad (10)$$

Proof. The well-posedness follows from (P2), which guarantees the inversibility of the linear system corresponding to (5). Using (P2), Hölder's inequality, Lemma 1 and Remark 2, it is inferred that (with $r' \overset{\text{def}}{=} \frac{r}{r-1}$)

$$\gamma_1\|u_n\|^2_{\mathcal{T}_n,I} \le a_{\mathcal{T}_n}(u_n,u_n) = \int_\Omega fu\,\mathrm{d}x \le \|f\|_{L^r(\Omega)}\|u_n\|_{L^{r'}(\Omega)} \le C_1\|f\|_{L^r(\Omega)}\|u_n\|_{\mathcal{T}_n,I}. \quad \square$$

Theorem 1 (Convergence). *Let* $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ *be a family of discretizations satisfying Hypothesis 1 and s.t.* $h_{\mathcal{D}_n} \to 0$ *as* $n \to \infty$. *Then, as* $n \to \infty$, *the sequence of discrete solutions of problem* (5), *say* $\{u_n\}_{n\in\mathbb{N}}$, *converges to the solution* $\overline{u}$ *of* (1) *in* $L^q(\Omega)$ *for all* $q \in [1, 2d/(d-2))$ *(and weakly in* $L^{2d/(d-2)}(\Omega)$ *if* $d > 2$*).*

Proof. Owing to the *a priori* estimate (10) together with Lemma 2, there is $\widetilde{u} \in H^1_0(\Omega)$ s.t., up to a subsequence, (i) $\{u_n\}_{n\in\mathbb{N}}$ converges to $\widetilde{u}$ in $L^q(\Omega)$ for all $q \in [1, 2d/(d-2))$ (and weakly in $L^{2d/(d-2)}(\Omega)$ if $d > 2$) and (ii) $\{\widetilde{\nabla}_{\mathcal{D}_n}u_n\}_{n\in\mathbb{N}}$ weakly converges to $\nabla\widetilde{u}$ in $[L^2(\Omega)]^d$. It only remains to prove that $\widetilde{u} = u$. Let $\varphi \in \mathfrak{D}$. Owing to (7) together with (P2),

$$\|\widetilde{\nabla}_{\mathcal{D}_n}(u_n - \varphi_{\mathcal{T}_n})\|^2_{[L^2(\Omega)]^d} \le d\|u_n - \varphi_{\mathcal{T}_n}\|^2_{\mathcal{T}_n,I} \le \frac{d}{\gamma_1}a_{\mathcal{T}_n}(u_n - \varphi_{\mathcal{T}_n}, u_n - \varphi_{\mathcal{T}_n}) = \frac{d}{\gamma_1}(T_1 + T_2), \quad (11)$$

where $T_1 \stackrel{\rm def}{=} \int_\Omega f(x)(u_n - \varphi_{\mathcal{T}_n})(x)\,\mathrm{d}x$ and $T_2 \stackrel{\rm def}{=} a_{\mathcal{T}_n}(\varphi_{\mathcal{T}_n}, \varphi_{\mathcal{T}_n} - u_n)$. Clearly, by the integrability assumption on f and the weak convergence of $\{u_n\}_{n\in\mathbb{N}}$ to $\widetilde{u}$ in $L^q(\Omega)$ for all $q < +\infty$ if $d = 2$ and for $q = \frac{2d}{d-2}$ if $d > 2$,

$$T_1 \to \int_\Omega f(x)(\widetilde{u} - \varphi)(x)\,\mathrm{d}x \text{ as } n \to \infty. \tag{12}$$

Owing to (9),

$$\begin{aligned} a_{\mathcal{T}_n}(\varphi_{\mathcal{T}_n}, u_n) = &\sum_{K\in\mathcal{T}_n}\sum_{\sigma\in\mathcal{E}_K}\left[F_{K,\sigma}(\varphi_{\mathcal{T}_n}) - \mathrm{m}_\sigma \frac{\int_K \Lambda(x)\nabla\varphi(x)\,\mathrm{d}x}{\mathrm{m}_K}\cdot\mathbf{n}_{K,\sigma}\right](\gamma_\sigma u_n - u_{n,K}) \\ &+ \int_\Omega \Lambda(x)\nabla\varphi(x)\cdot\widetilde{\nabla}_{\mathcal{D}_n} u_n(x)\ dx \stackrel{\rm def}{=} T_{2,1} + T_{2,2}. \end{aligned}$$

Using Cauchy-Schwarz inequality as in the proof of Proposition 1, we have $T_{2,1} \le \epsilon_{\mathcal{D}_n}(\varphi)\|u_n\|_{\mathcal{T}_n,I}$. Thanks to Lemma 3, $\|u_n\|_{\mathcal{T}_n,I}$ is bounded uniformly with respect to n. Thus, by property (P3), $T_{2,1} \to 0$ as $n \to \infty$. Also, using the weak convergence of $\{\widetilde{\nabla}_{\mathcal{D}_n} u_n\}_{n\in\mathbb{N}}$, we conclude that $T_{2,2} \to \int_\Omega \Lambda(x)\nabla\varphi(x)\cdot\nabla\widetilde{u}(x)\,\mathrm{d}x$ as $n \to \infty$. By Proposition 1, $\|\varphi_{\mathcal{T}_n}\|_{\mathcal{T}_n,I}$ is uniformly bounded with respect to n; since $\varphi_{\mathcal{T}_n}$ obviously converges to φ, it is then easy, using Lemma 2, to see that $\widetilde{\nabla}_{\mathcal{D}_n}\varphi_{\mathcal{T}_n}$ weakly converges to $\nabla\varphi$. Proceeding in a similar way as for $a_{\mathcal{T}_n}(\varphi_{\mathcal{T}_n}, u_n)$, we can thus prove that $a_{\mathcal{T}_n}(\varphi_{\mathcal{T}_n}, \varphi_{\mathcal{T}_n}) \to \int_\Omega \Lambda(x)\nabla\varphi(x)\cdot\nabla\varphi(x)\,\mathrm{d}x$ as $n \to \infty$. Therefore,

$$T_2 \to \int_\Omega \Lambda(x)\nabla\varphi(x)\cdot\nabla(\varphi - \widetilde{u})(x)\,\mathrm{d}x \text{ as } n \to \infty. \tag{13}$$

Plugging (12) and (13) into the right hand side of (11) and using the weak convergence of $\widetilde{\nabla}_{\mathcal{D}_n}(u_n - \varphi_{\mathcal{T}_n})$, we conclude that, for all $\varphi \in \mathfrak{D}$,

$$\|\nabla(\widetilde{u} - \varphi)\|^2_{[L^2(\Omega)]^d} \le \frac{d}{\gamma_1}\left(\int_\Omega f(x)(\widetilde{u} - \varphi)(x)\,\mathrm{d}x + \int_\Omega \Lambda(x)\nabla\varphi(x)\cdot\nabla(\varphi - \widetilde{u})(x)\,\mathrm{d}x\right).$$

By virtue of (P1), we can apply this inequality to a sequence $\{\varphi_m\}_{m\in\mathbb{N}} \in \mathfrak{D}$ which tends to $\overline{u}$ in $H^1_0(\Omega)$ and let $m \to \infty$; since $\overline{u}$ solves problem (1), we obtain

$$\|\nabla(\widetilde{u} - \overline{u})\|^2_{[L^2(\Omega)]^d} \le \frac{d}{\gamma_1}\left[\int_\Omega f(x)(\widetilde{u}(x) - \overline{u}(x))\,\mathrm{d}x - \int_\Omega \Lambda(x)\nabla\overline{u}(x)\cdot\nabla(\widetilde{u} - \overline{u})(x)\,\mathrm{d}x\right] = 0,$$

i.e., $\widetilde{u} = \overline{u}$. Due to the uniqueness of the solution of (1), we deduce that the entire sequence $\{u_n\}_{n\in\mathbb{N}}$ converges to $\overline{u}$ in $L^q(\Omega)$ for all $q \in [1, 2d/(d-2))$ (and weakly in $L^{2d/(d-2)}(\Omega)$ if $d > 2$). Observe that the order in which the limits for $n \to \infty$ and $m \to \infty$ are taken cannot be exchanged, the sequence $\{\|(\varphi_m)_{\mathcal{T}_n}\|_{\mathcal{T}_n,I}\}_{m\in\mathbb{N}}$ being possibly unbounded. This concludes the proof. □

2.3 A strongly convergent gradient reconstruction

In practical applications, it is often desirable to dispose of a gradient reconstruction which *strongly* converges to the gradient of the continuous solution (whereas $\widetilde{\nabla}_{\mathcal{D}}$ only provides a weakly convergent approximation of this gradient). Such a reconstruction can be obtained using the same formula as in the Mixed Finite Volume method [19], which is based on the numerical fluxes.

Let $\sigma \in \mathcal{E}$ be a mesh face, and denote by x_σ its barycenter. For all $v \in H_{\mathcal{T}}(\Omega)$, define

$$\overline{\nabla}_{\mathcal{D}} v(x)_{|K} = (\overline{\nabla}_{\mathcal{D}} v)_K \stackrel{\rm def}{=} \frac{1}{\mathrm{m}_K}\Lambda_K^{-1}\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(v)(x_\sigma - x_K) \quad \forall K \in \mathcal{T}. \tag{14}$$

The following geometrical relation holds:

$$\forall \xi \in \mathbb{R}^d\ ,\ \forall K \in \mathcal{T},\quad \sum_{\sigma\in\mathcal{E}_K} \mathrm{m}_\sigma\,\xi\cdot\mathbf{n}_{K,\sigma}(x_\sigma - x_K) = \mathrm{m}_K\,\xi. \tag{15}$$

Equation (15) in fact justifies that $(\overline{\nabla}_{\mathcal{D}} v)_K$ is expected to approximate ∇v on K provided $\{F_{K,\sigma}(v)\}_{\sigma\in\mathcal{E}_K}$ are consistent approximations of the fluxes of $\Lambda\nabla v$.

Lemma 4 (Consistency). *Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ denote a family of discretizations satisfying Hypothesis 1 and s.t. $h_{\mathcal{D}_n}\to 0$ as $n\to\infty$. Then, for all $\varphi\in\mathfrak{D}$,*

$$\lim_{n\to\infty}\|\overline{\nabla}_{\mathcal{D}_n}\varphi_{\mathcal{T}_n}-\nabla\varphi\|^2_{[L^2(\Omega)]^d}=0.$$

Proof. For all $n\in\mathbb{N}$, for all $K\in\mathcal{T}_n$ and for all $y\in K$, formula (15) applied to $\xi=\frac{1}{\mathrm{m}_K}\int_K\Lambda(x)\nabla\varphi(x)\,\mathrm{d}x$ yields

$$\frac{1}{\mathrm{m}_K}\Lambda_K^{-1}\int_K\Lambda(x)\nabla\varphi(x)\,\mathrm{d}x=\frac{1}{\mathrm{m}_K}\Lambda_K^{-1}\sum_{\sigma\in\mathcal{E}_K}\left(\frac{\mathrm{m}_\sigma}{\mathrm{m}_K}\int_K\Lambda(x)\nabla\varphi(x)\,\mathrm{d}x\right)\cdot\mathbf{n}_{K,\sigma}(x_\sigma-x_K).$$

Let $T_{K,\sigma}(\varphi_{\mathcal{T}})\stackrel{\text{def}}{=}F_{K,\sigma}(\varphi_{\mathcal{T}})-\left(\frac{\mathrm{m}_\sigma}{\mathrm{m}_K}\int_K\Lambda(x)\nabla\varphi(x)\,\mathrm{d}x\right)\cdot\mathbf{n}_{K,\sigma}$. Using Cauchy-Schwarz inequality we get

$$\begin{aligned}
&|(\overline{\nabla}_{\mathcal{D}_n}\varphi_{\mathcal{T}_n})_K-\nabla\varphi(y)|\\
&\quad=\left|\frac{1}{\mathrm{m}_K}\Lambda_K^{-1}\sum_{\sigma\in\mathcal{E}_K}T_{K,\sigma}(\varphi_{\mathcal{T}})(x_\sigma-x_K)+\frac{1}{\mathrm{m}_K}\Lambda_K^{-1}\left(\int_K\Lambda(x)(\nabla\varphi(x)-\nabla\varphi(y))\,\mathrm{d}x\right)\right|\\
&\quad\le\frac{1}{\alpha_0\,\mathrm{m}_K}\left[\sum_{\sigma\in\mathcal{E}_K}\sqrt{\frac{d_{K,\sigma}}{\mathrm{m}_\sigma}}|T_{K,\sigma}(\varphi_{\mathcal{T}_n})|\frac{|x_\sigma-x_K|}{\sqrt{d_{K,\sigma}}}\sqrt{\mathrm{m}_\sigma}+C_3\beta_0\,\mathrm{m}_K\,\mathrm{diam}(K)\right]\\
&\quad\le\frac{1}{\alpha_0\,\mathrm{m}_K}\left(\sum_{\sigma\in\mathcal{E}_K}\frac{d_{K,\sigma}}{\mathrm{m}_\sigma}|T_{K,\sigma}(\varphi_{\mathcal{T}_n})|^2\right)^{\frac12}\left(\sum_{\sigma\in\mathcal{E}_K}\frac{|x_\sigma-x_K|^2}{d_{K,\sigma}}\,\mathrm{m}_\sigma\right)^{\frac12}+\frac{C_3\beta_0}{\alpha_0}\,\mathrm{diam}(K),
\end{aligned}$$

where $C_3\stackrel{\text{def}}{=}\sup_{i=1,\dots,N_\Omega,x\in\Omega_i}|\varphi''(x)|$. As a consequence, using (3),

$$\begin{aligned}
\int_K\left|(\overline{\nabla}_{\mathcal{D}_n}\varphi_{\mathcal{T}_n})_K-\nabla\varphi(y)\right|^2\,\mathrm{d}y&\le\frac{2}{(\alpha_0\varrho_1)^2}\sum_{\sigma\in\mathcal{E}_K}\frac{d_{K,\sigma}}{\mathrm{m}_\sigma}|T_{K,\sigma}(\varphi_{\mathcal{T}_n})|^2\times\frac{1}{\mathrm{m}_K}\sum_{\sigma\in\mathcal{E}_K}\mathrm{m}_\sigma\,d_{K,\sigma}\\
&\quad+2\left(\frac{C_3\beta_0}{\alpha_0}\right)^2\mathrm{m}_K\,\mathrm{diam}(K)^2\\
&\le\frac{2d}{(\alpha_0\varrho_1)^2}\sum_{\sigma\in\mathcal{E}_K}\frac{d_{K,\sigma}}{\mathrm{m}_\sigma}|T_{K,\sigma}(\varphi_{\mathcal{T}_n})|^2+2\left(\frac{C_3\beta_0}{\alpha_0}\right)^2\mathrm{m}_K\,h_{\mathcal{D}_n}^2.
\end{aligned}$$

Finally, summing over $K\in\mathcal{T}_n$,

$$\|\overline{\nabla}_{\mathcal{D}_n}\varphi_{\mathcal{T}_n}-\nabla\varphi\|^2_{[L^2(\Omega)]^d}\le\frac{2d}{(\alpha_0\varrho_1)^2}\epsilon^2_{\mathcal{D}_n}(\varphi)+2\left(\frac{C_3\beta_0}{\alpha_0}\right)^2\mathrm{m}_\Omega\,h^2_{\mathcal{D}_n}.$$

Use (P3) to conclude. □

The convergence of the gradient reconstruction (14) requires the following

Hypothesis 2. *Assume that there is C_4 independent of n s.t.*

$$\forall n\in\mathbb{N},\ \forall v\in H_{\mathcal{T}_n}(\Omega),\quad\sum_{K\in\mathcal{T}_n}\sum_{\sigma\in\mathcal{E}_K}\frac{d_{K,\sigma}}{\mathrm{m}_\sigma}|F_{K,\sigma}(v)|^2\le C_4\|v\|^2_{\mathcal{T}_n,I}.\tag{16}$$

Remark 5. Hypothesis 2 is readily verified on Λ-orthogonal meshes and two point fluxes. A mesh is said to be Λ-orthogonal if, for all $\sigma\in\mathcal{E}$, there exists $x_\sigma\in\sigma$ s.t., for all $K\in\mathcal{T}_\sigma$, $\Lambda_K^{-1}(x_\sigma-x_K)$ is orthogonal to σ. For Λ-orthogonal meshes, we can define two-points consistent fluxes in the sense of (8) as $F_{K,\sigma}(v)=\mathrm{m}_\sigma\,t_\sigma\frac{(\gamma_\sigma v-v_K)}{d_{K,\sigma}}$, where the reals $\{t_\sigma\}_{\sigma\in\mathcal{E}}$ are given by

$$\begin{cases}\frac{d_{K,\sigma}}{\Lambda_K\mathbf{n}_{K,\sigma}\cdot\mathbf{n}_{K,\sigma}}+\frac{d_{L,\sigma}}{\Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}=\frac{d_{K,\sigma}+d_{L,\sigma}}{t_\sigma} & \text{if } \sigma\in\mathcal{E}_{\text{int}} \text{ with } \mathcal{T}_\sigma=\{K,L\}\\ t_\sigma=\Lambda_K\mathbf{n}_{K,\sigma}\cdot\mathbf{n}_{K,\sigma} & \text{if } \sigma\in\mathcal{E}_{\text{ext}} \text{ with } \mathcal{T}_\sigma=\{K\}.\end{cases}$$

Since for all $\sigma\in\mathcal{E}$, $t_\sigma\le\beta_0$, (16) holds with $C_4=\beta_0^2$.

Proposition 2 (Stability of the gradient reconstruction). *Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ denote a family of discretizations matching Definition 1. Let Hypothesis 2 hold. Then,*

$$\forall v \in H_{\mathcal{T}}(\Omega), \quad \|\overline{\nabla}_{\mathcal{D}} v\|_{[L^2(\Omega)]^d} \leq \frac{\sqrt{dC_4}}{\alpha_0 \varrho_1} \|v\|_{\mathcal{T},I}.$$

Proof. Let $v \in H_{\mathcal{T}}(\Omega)$. Cauchy-Schwarz inequality yields

$$\begin{aligned}\|\overline{\nabla}_{\mathcal{D}} v\|^2_{[L^2(\Omega)]^d} &= \sum_{K\in\mathcal{T}} \mathrm{m}_K \left| \frac{1}{\mathrm{m}_K} \Lambda_K^{-1} \sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(v)(x_\sigma - x_K) \right|^2 \\ &\leq \frac{1}{\alpha_0^2} \sum_{K\in\mathcal{T}} \left(\sum_{\sigma\in\mathcal{E}_K} \frac{d_{K,\sigma}}{\mathrm{m}_\sigma} |F_{K,\sigma}(v)|^2 \times \sum_{\sigma\in\mathcal{E}_K} \frac{\mathrm{m}_\sigma \operatorname{diam}(K)^2}{d_{K,\sigma}\, \mathrm{m}_K} \right).\end{aligned}$$

Owing to (3) together with (2), $\sum_{\sigma\in\mathcal{E}_K} \frac{\mathrm{m}_\sigma \operatorname{diam}(K)^2}{d_{K,\sigma}\, \mathrm{m}_K} \leq \frac{d}{\varrho_1^2}$. Conclude using Hypothesis 2. □

Theorem 2 (Strong convergence of the discrete gradient reconstruction). *Let $\overline{u}$ be the solution to (1). Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be a family of meshes matching Definition 1 and s.t. $h_{\mathcal{D}_n} \to 0$ as $n \to \infty$, and denote by u_n the solution of problem (5) on $\mathcal{D}_n$. Finally, let Hypotheses 1 and 2 hold. Then, the sequence $\{\overline{\nabla}_{\mathcal{D}_n} u_n\}_{n\in\mathbb{N}}$ converges to $\nabla\overline{u}$ in $[L^2(\Omega)]^d$.*

Proof. Thanks to Theorem 1 together with Lemma 2, (i) the sequence of weak solutions $\{u_n\}_{n\in\mathbb{N}}$ converges to $\overline{u}$ in $L^q(\Omega)$, for all $q \in [1, 2d/(d-2))$ (and weakly in $L^{2d/(d-2)}(\Omega)$ if $d > 2$) and (ii) the sequence $\{\widetilde{\nabla}_{\mathcal{D}_n} u_n\}_{n\in\mathbb{N}}$ weakly converges to $\nabla\overline{u}$ in $[L^2(\Omega)]^d$. Let $\varphi \in \mathfrak{D}$. For all $n \in \mathbb{N}$,

$$\int_\Omega |\overline{\nabla}_{\mathcal{D}_n} u_n - \nabla\overline{u}|^2 \leq 3 \left(\int_\Omega |\overline{\nabla}_{\mathcal{D}_n} u_n - \overline{\nabla}_{\mathcal{D}_n} \varphi_{\mathcal{T}_n}|^2 + \int_\Omega |\overline{\nabla}_{\mathcal{D}_n} \varphi_{\mathcal{T}_n} - \nabla\varphi|^2 + \int_\Omega |\nabla\varphi(x) - \nabla\overline{u}|^2 \right).$$

Let T_i, $i = 1, \ldots, 3$ denote the addends in the right hand side. Owing to Proposition 2 together with (P2) we have

$$T_1 \leq \frac{dC_4}{(\alpha_0\varrho_1)^2\gamma_1}\, a_{\mathcal{T}_n}(u_n - \varphi_{\mathcal{T}_n}, u_n - \varphi_{\mathcal{T}_n}).$$

Using the same arguments as in the proof of Theorem 1, we infer that

$$\limsup_{n\to\infty} T_1 \leq \frac{dC_4}{(\alpha_0\varrho_1)^2\gamma_1} \left(\int_\Omega f(\overline{u} - \varphi) + \int_\Omega \Lambda\nabla\varphi\cdot(\nabla\varphi - \nabla\overline{u}) \right).$$

Moreover, owing to Lemma 4, $\lim_{n\to\infty} T_2 = 0$ and thus

$$\begin{aligned}\limsup_{n\to\infty} \int_\Omega |\overline{\nabla}_{\mathcal{D}_n} u - \nabla\overline{u}|^2 \leq \\ 3\frac{dC_4}{(\alpha_0\varrho_1)^2\gamma_1} \left(\int_\Omega f(\overline{u} - \varphi) + \int_\Omega \Lambda\nabla\varphi\cdot(\nabla\varphi - \nabla\overline{u}) \right) + 3 \left(\int_\Omega |\nabla\varphi - \nabla\overline{u}|^2 \right).\end{aligned}$$

Finally, since $\mathfrak{D}$ is dense in $H_0^1(\Omega)$, we conclude by letting φ tend to $\overline{u}$ in $H_0^1(\Omega)$. □

3 The G method

In the present section we introduce a family of FV methods generalizing the MPFA L method of [5, 4] and show that it fits in the abstract analysis framework of §2.

3.1 Construction of group gradients

We need the following additional

Definition 2. *Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be a family of meshes matching Definition 1. We further suppose that*

(i) $\mathcal{V}_n$ is a family of points (the vertices *or* nodes*), s.t., for all $K \in \mathcal{T}_n$, for all $U_K \subseteq \mathcal{E}_K$ satisfying* $\mathrm{card}(U_K) \geq d$*, we have $\bigcap_{\sigma\in U_K} \sigma = \emptyset$ or $\bigcap_{\sigma\in U_K} \sigma = s \in \mathcal{V}_n$. For all $s \in \mathcal{V}_n$, we let $\mathcal{E}_s \stackrel{\text{def}}{=} \{\sigma \in \mathcal{E} \mid s \in \sigma\}$ and $\mathcal{T}_s \stackrel{\text{def}}{=} \{K \in \mathcal{T} \mid s \in \overline{K}\}$, and we assume that each σ contains at least one vertex (this could be false in dimension $d = 3$ if, for example, σ is only a piece of a "true face" of a cell);*

(ii) the number of faces sharing one node remains bounded as the mesh is refined, i.e., there exists $\varrho_3 \in \mathbb{N}^*$ s.t.

$$\sup_{n\in\mathbb{N}} \max_{s\in\mathcal{V}_n} \mathrm{card}(\mathcal{E}_s) \leq \varrho_3;$$

(iii) $\widetilde{\mathcal{G}}$ is the finite family of face groups *defined as follows:*

$$\widetilde{\mathcal{G}} \stackrel{\text{def}}{=} \{G \subset \mathcal{E}_K \cap \mathcal{E}_s,\ K \in \mathcal{T}_s,\ s \in \mathcal{V}_n,\ \mathrm{card}(G) = d\}.$$

For each $G \in \widetilde{\mathcal{G}}$, we let $\mathcal{T}_G = \{K \in \mathcal{T},\ G \cap \mathcal{E}_K \neq \emptyset\}$. We also arbitrarily select a cell, which we denote by K_G, s.t. $G \subset \mathcal{E}_{K_G}$.

Remark 6. In many cases, a given group G is contained in a unique $\mathcal{E}_K$ but, in some cases (especially if the discretization has non-convex cells), there can be multiple possible choices for K_G.

Our idea is, for all $v \in H_{\mathcal{T}}(\Omega)$, all group $G \in \widetilde{\mathcal{G}}$, all $\sigma \in G$ and all $K \in \mathcal{T}_\sigma$, to build a "group gradient" $(\nabla_{\mathcal{D}} v)_K^{G,\sigma} \in \mathbb{R}^d$ and use it to define the flux $F_{K,\sigma}(u)$; this gradient could be understood as a *piece* of a full gradient of u on the pyramid $\triangle_{K,\sigma}$, the full gradient (and resulting flux) being obtained as a convex combination of these group gradients corresponding to all the groups G containing σ (see (22)).

First, for all $\sigma \in \mathcal{E}$, $\mathcal{T}_\sigma = \{K, L\}$, we require that, if $\mathcal{T}_\sigma = \{K, L\}$, the values v_K, v_L and the gradient reconstruction $(\nabla_{\mathcal{D}} v)_K^{G,\sigma}$, $(\nabla_{\mathcal{D}} v)_L^{G,\sigma}$ yield the same value of v on σ, that is to say

$$v_K + (\nabla_{\mathcal{D}} v)_K^{G,\sigma} \cdot (x - x_K) = v_L + (\nabla_{\mathcal{D}} v)_L^{G,\sigma} \cdot (x - x_L) \quad \forall x \in \sigma.$$

For boundary faces, we ask that the value obtained at $x = x_\sigma$, barycenter of σ be zero. Second, we would like the resulting fluxes to be conservative, i.e.,

$$\Lambda_K (\nabla_{\mathcal{D}} v)_K^{G,\sigma} \cdot \mathbf{n}_{K,\sigma} + \Lambda_L (\nabla_{\mathcal{D}} v)_L^{G,\sigma} \cdot \mathbf{n}_{L,\sigma} = 0.$$

These two sets of equations are not sufficient to define uniquely the group gradients (and thus to estimate them, which is fundamental in the study of the numerical method). We therefore add another constraint, giving a particular role to the cell K_G selected for the group G: we ask that $(\nabla_{\mathcal{D}} v)_{K_G}^{G,\sigma}$ does not depend on $\sigma \in G$, and we denote by $(\nabla_{\mathcal{D}} v)_{K_G}^{G}$ the common value of this group gradient for all $\sigma \in G$. The discrete gradients are thus defined, as in [5, 4], by: For all $G \in \widetilde{\mathcal{G}}$ and all $\sigma \in G \cap \mathcal{E}_{\mathrm{int}}$, with $\mathcal{T}_\sigma = \{K_G, L\}$,

$$\begin{cases} v_{K_G} + (\nabla_{\mathcal{D}} v)_{K_G}^{G} \cdot (x - x_{K_G}) = v_L + (\nabla_{\mathcal{D}} v)_L^{G,\sigma} \cdot (x - x_L) \quad \forall x \in \sigma, \\ \Lambda_{K_G} (\nabla_{\mathcal{D}} v)_{K_G}^{G} \cdot \mathbf{n}_{K_G,\sigma} + \Lambda_L (\nabla_{\mathcal{D}} v)_L^{G,\sigma} \cdot \mathbf{n}_{L,\sigma} = 0, \end{cases} \tag{17}$$

and for all $\sigma \in G \cap \mathcal{E}_{\mathrm{ext}}$,

$$v_{K_G} + (\nabla_{\mathcal{D}} v)_{K_G}^{G} \cdot (x_\sigma - x_{K_G}) = 0. \tag{18}$$

Lemma 5. *The gradient reconstruction $(\nabla_{\mathcal{D}} v)_{K_G}^{G}$ defined by* (17) *and* (18) *can be obtained solving a linear system of the form*

$$\mathcal{A}_G X_G = \mathcal{B}_G(v), \tag{19}$$

where the rows of $\mathcal{A}_G \in \mathbb{R}^{d\times d}$ *are built from the following family of vectors of* $\mathbb{R}^d$

$$\left(\left\{\frac{\Lambda_L \mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(x_L - x_{K_G}) + \Lambda_{K_G}\mathbf{n}_{K_G,\sigma} + \Lambda_L \mathbf{n}_{L,\sigma}\right\}_{\sigma\in G\cap\mathcal{E}_{\rm int}}, \left\{\frac{\Lambda_{K_G}\mathbf{n}_{K_G,\sigma}\cdot\mathbf{n}_{K_G,\sigma}}{d_{K_G,\sigma}}(x_\sigma - x_{K_G})\right\}_{\sigma\in G\cap\mathcal{E}_{\rm ext}}\right)$$

and $\mathcal{B}_G(v) \in \mathbb{R}^d$ *is obtained from the family of vectors*

$$\left(\left\{\frac{\Lambda_L \mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(v_L - v_{K_G})\right\}_{\sigma\in G\cap\mathcal{E}_{\rm int}}, \left\{\frac{\Lambda_{K_G}\mathbf{n}_{K_G,\sigma}\cdot\mathbf{n}_{K_G,\sigma}}{d_{K_G,\sigma}}(-v_{K_G})\right\}_{\sigma\in G\cap\mathcal{E}_{\rm ext}}\right).$$

Proof. Let $v \in H_{\mathcal{T}}(\Omega), G \in \mathcal{G}$, $\sigma \in G\cap\mathcal{E}_{\rm int}$ with $\mathcal{T}_\sigma = \{K_G, L\}$. Observe that, if $\mathbf{v} \stackrel{\rm def}{=} (\nabla_{\mathcal{D}} v)^G_{K_G} - (\nabla_{\mathcal{D}} v)^{G,\sigma}_L \neq 0$, the first equation of (17) is the equation of an hyperplane of $\mathbb{R}^d$ orthogonal to $\mathbf{v}$; satisfying this equation for all $x \in \sigma$ is equivalent to imposing that σ is contained in this hyperplane, and thus that $\mathbf{v}$ and $\mathbf{n}_{K_G,\sigma}$ are colinear (this is of course also true if $\mathbf{v} = 0$). As a consequence, taking $y_\sigma \in \sigma$, the first equation in (17) is equivalent to the following linear system (in which $\lambda^G_\sigma \in \mathbb{R}$ is an additional unknown):

$$\begin{cases} (\nabla_{\mathcal{D}} v)^G_{K_G} - (\nabla_{\mathcal{D}} v)^{G,\sigma}_L = \lambda^G_\sigma \mathbf{n}_{K_G,\sigma}, \\ v_{K_G} - v_L + (\nabla_{\mathcal{D}} v)^{G,\sigma}_L \cdot x_L - (\nabla_{\mathcal{D}} v)^G_{K_G}\cdot x_{K_G} = -\lambda^G_\sigma \mathbf{n}_{K_G,\sigma}\cdot y_\sigma. \end{cases}$$

Since $(y_\sigma - x_L)\cdot\mathbf{n}_{K_G,\sigma} = -d_{L,\sigma}$, solving for λ^G_σ we obtain

$$\begin{cases} \lambda^G_\sigma = -\frac{R_{L,\sigma}(v)}{d_{L,\sigma}}, \\ (\nabla_{\mathcal{D}} v)^{G,\sigma}_L = (\nabla_{\mathcal{D}} v)^G_{K_G} - \frac{R_{L,\sigma}(v)}{d_{L,\sigma}}\mathbf{n}_{L,\sigma}, \end{cases}$$

with $R_{L,\sigma}(v) \stackrel{\rm def}{=} v_L - v_{K_G} - (\nabla_{\mathcal{D}} v)^G_{K_G}\cdot(x_L - x_{K_G})$. Using these expressions, the second equation of (17) can be rewritten as

$$\left[\Lambda_{K_G}\mathbf{n}_{K_G,\sigma} + \Lambda_L \mathbf{n}_{L,\sigma} + \frac{\Lambda_L \mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(x_L - x_{K_G})\right]\cdot(\nabla_{\mathcal{D}} v)^G_{K_G} = \frac{\Lambda_L \mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(v_L - v_{K_G}).$$

Finally, the linear system (17)–(18) is equivalent to:

$$\begin{cases} (\nabla_{\mathcal{D}} v)^{G,\sigma}_L = (\nabla_{\mathcal{D}} v)^G_{K_G} - \dfrac{R_{L,\sigma}(v)}{d_{L,\sigma}}\mathbf{n}_{L,\sigma}\,, \quad \forall \sigma \in G\cap\mathcal{E}_{\rm int}\,,\ \mathcal{T}_\sigma = \{K_G, L\}, \\ \left[\dfrac{\Lambda_L \mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(x_L - x_{K_G}) + \Lambda_{K_G}\mathbf{n}_{K_G,\sigma} + \Lambda_L \mathbf{n}_{L,\sigma}\right]\cdot(\nabla_{\mathcal{D}} v)^G_{K_G} \\ \qquad = \dfrac{\Lambda_L \mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(v_L - v_{K_G})\,, \quad \forall \sigma \in G\cap\mathcal{E}_{\rm int}\,,\ \mathcal{T}_\sigma = \{K_G, L\}, \\ \dfrac{\Lambda_{K_G}\mathbf{n}_{K_G,\sigma}\cdot\mathbf{n}_{K_G,\sigma}}{d_{K_G,\sigma}}(\nabla_{\mathcal{D}} v)^G_{K_G}\cdot(x_\sigma - x_{K_G}) = \dfrac{\Lambda_{K_G}\mathbf{n}_{K_G,\sigma}\cdot\mathbf{n}_{K_G,\sigma}}{d_{K_G,\sigma}}(-v_{K_G}), \quad \forall \sigma \in G\cap\mathcal{E}_{\rm ext}. \end{cases} \tag{20}$$

The assert follows. □

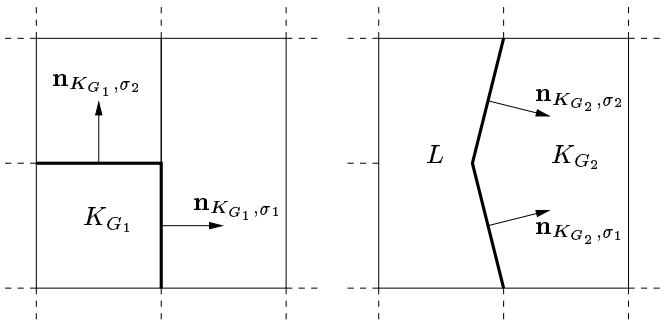
In order to construct the gradient, we therefore need to consider the set of groups s.t. the matrix $\mathcal{A}_G$ is invertible, that is to say

$$\mathcal{G} \stackrel{\rm def}{=} \{G \in \widetilde{\mathcal{G}} \,|\, \mathcal{A}_G \text{ is invertible}\}.$$

We shall also need a symbol for the family of groups containing a give face $\sigma \in \mathcal{E}$. We thus let

$$\forall \sigma \in \mathcal{E}, \quad \mathcal{G}_\sigma \stackrel{\rm def}{=} \{G \in \mathcal{G} \,|\, \sigma \in G\},$$

and we assume throughout the rest of the present section that all the $\mathcal{G}_\sigma$ are non-empty.



(a) $G_1 \in \mathcal{G}_\sigma$ (b) $G_2 \notin \mathcal{G}_\sigma$

Figure 2: Two examples of face groups of $\widetilde{\mathcal{G}}$ respectively belonging and not belonging to $\mathcal{G}_\sigma$.

Remark 7. Figure 2 shows two examples of groups respectively belonging and not belonging to $\mathcal{G}$ (in the case where Λ is constant). Indeed, since $\Lambda_{K_{G_2}} = \Lambda_L$, the terms $\Lambda_{K_{G_2}} \mathbf{n}_{K_{G_2},\sigma_i}$ and $\Lambda_L \mathbf{n}_{L,\sigma_i}$ cancel out each other in each line of $\mathcal{A}_{G_2}$ and, since the cell L on the other side of σ_1 and σ_2 is the same, both lines of $\mathcal{A}_{G_2}$ are colinear to $x_L - x_{K_{G_2}}$ (this matrix is therefore singular). The non-convexity of cells can be a cause to the singularity of some $\mathcal{A}_G$ (but this does not block the use of the G method since, even in this case, the non-emptiness of all $\mathcal{G}_\sigma$ often holds).

Finally, we define in the following two lemmata the space playing the role of $\mathfrak{D}$ in Hypothesis 1 and state its density, and we establish the consistency on this space of the group gradients (this will give (P3)).

Lemma 6 (Density of a space of test-functions). *Let $\mathcal{Q}$ be the space of functions $\varphi : \overline{\Omega} \to \mathbb{R}$ s.t.*

(i) (φ is continuous and piecewise regular*) $\varphi \in C_0(\overline{\Omega})$ and, for all $i = 1, \ldots, N_\Omega$, $\varphi \in C^2(\overline{\Omega_i})$,*

*(ii) (*the tangential derivatives of φ are continuous through the interfaces of P_Ω*) for all $i, j = 1, \ldots, N_\Omega$, for all vector $\mathbf{t}$ parallel to $\partial\Omega_i \cap \partial\Omega_j$, $(\nabla\varphi)_{|\overline{\Omega_i}} \cdot \mathbf{t} = (\nabla\varphi)_{|\overline{\Omega_j}} \cdot \mathbf{t}$, where $(\nabla\varphi)_{|\overline{\Omega_i}}$ refers to the value of $\nabla\varphi$ on $\partial\Omega_i$ computed from the values on $\overline{\Omega_i}$,*

*(iii) (*the flux of $\nabla\varphi$ directed by $\Lambda\mathbf{n}$ is continuous through the interfaces of P_Ω*) for all $i, j = 1, \ldots, N_\Omega$ s.t. $\partial\Omega_i \cap \partial\Omega_j$ has dimension $d-1$, $(\Lambda\nabla\varphi)_{|\overline{\Omega_i}} \cdot \mathbf{n}_i + (\Lambda\nabla\varphi)_{|\overline{\Omega_j}} \cdot \mathbf{n}_j = 0$ on $\partial\Omega_i \cap \partial\Omega_j$, where $\mathbf{n}_i$ is the outer normal to Ω_i.*

Then, $\mathcal{Q}$ is dense in $H_0^1(\Omega)$.

Proof. See Appendix A. □

Lemma 7 (Consistency of the group gradients). *Let $\mathcal{D}$ be an element of a family of discretizations satisfying Hypothesis 1. For all $\varphi \in \mathcal{Q}$, there exists a real $C_5 > 0$ which only depends on ϱ_1, ϱ_2, Λ and φ s.t., for all $G \in \mathcal{G}$, all $\sigma \in G$ and all $K \in \mathcal{T}_\sigma$,*

$$|(\nabla_{\mathcal{D}}\varphi_{\mathcal{T}})_K^{G,\sigma} - \nabla\varphi(x_K)| \leq C_5 \ (1 + |\mathcal{A}_G^{-1}|) \max_{K \in \mathcal{T}_G} \mathrm{diam}(K).$$

Proof. See Appendix B. □

3.2 Numerical fluxes

We choose $\{\theta_\sigma^G\}_{\sigma \in \mathcal{E}, G \in \mathcal{G}_\sigma}$ a set of weights s.t.

$$\text{For all } \sigma \in \mathcal{E}, \text{ for all } G \in \mathcal{G}_\sigma,\ 0 \leq \theta_\sigma^G \leq 1 \text{ and, for all } \sigma \in \mathcal{E},\ \sum_{G \in \mathcal{G}_\sigma} \theta_\sigma^G = 1. \tag{21}$$

The numerical fluxes are then defined as follows: For all $K \in \mathcal{T}$, for all $\sigma \in \mathcal{E}_K$,

$$F_{K,\sigma}(u) \stackrel{\text{def}}{=} \sum_{G \in \mathcal{G}_\sigma} \theta_\sigma^G F_{K,\sigma}^G(u), \quad F_{K,\sigma}^G(u) \stackrel{\text{def}}{=} \mathrm{m}_\sigma \Lambda_K (\nabla_{\mathcal{D}} u)_K^{G,\sigma} \cdot \mathbf{n}_{K,\sigma}. \tag{22}$$

Remark 8. Notice that the subfluxes $F_{K,\sigma}^G$ are conservative (second equation in (17)), and thus the whole fluxes $F_{K,\sigma}$ themselves are also conservative.

Specific methods are obtained from (22) by defining a suitable criterion to compute the family of weights $\{\theta_\sigma^G\}_{\sigma \in \mathcal{E}, G \in \mathcal{G}_\sigma}$.

Example 1 *(MPFA L method).* The MPFA L method can be obtained as follows: For all $\sigma \in \mathcal{E}$, let $\widetilde{G} \in \mathcal{G}_\sigma$ be the group satisfying the criterion proposed in [5] and set $\theta_\sigma^{\widetilde{G}} = 1/\mathrm{card}(\{\mathrm{s} \in \mathcal{V}, \mathrm{s} \in \sigma\})$ and $\theta_\sigma^G = 0$ for $\widetilde{G} \neq G \in \mathcal{G}_\sigma$.

Example 2. The alternative choice used in the numerical examples of §4 is designed so as to enhance the coercivity of the method. For each group $G \in \mathcal{G}$, define the space $\mathcal{H}_{\mathcal{T}_G} \stackrel{\text{def}}{=} \{u_K \in \mathbb{R},\ K \in \mathcal{T}_G\}$ endowed with the semi-norm

$$\|u\|_{\mathcal{T}_G}^2 \stackrel{\text{def}}{=} \sum_{K \in \mathcal{T}_G} \sum_{\sigma \in \mathcal{E}_K \cap G} \frac{\mathrm{m}_\sigma}{d_{K,\sigma}} (\gamma_\sigma u - u_K)^2.$$

For all $u,\ v \in \mathcal{H}_{\mathcal{T}_G}$ set $a_{\mathcal{T}_G}(u,v) = \sum_{K \in \mathcal{T}_G} \sum_{\sigma \in \mathcal{E}_K \cap G} F_{K,\sigma}^G(u)(\gamma_\sigma v - v_K)$. For each $G \in \mathcal{G}$ define

$$\gamma_2 \stackrel{\text{def}}{=} \inf_{\{u \in \mathcal{H}_{\mathcal{T}_G},\ \|u\|_{\mathcal{T}_G} = 1\}} a_{\mathcal{T}_G}(u,u)$$

The computation of the parameter γ_2 requires to evaluate the eigenvalues of a local matrix of $\mathbb{R}^{d \times d}$ associated with the bilinear form $a_{\mathcal{T}_G}$, and its cost is negligible. Indeed, by conservativity of the subfluxes,

$$\begin{aligned} a_{\mathcal{T}_G}(u,u) &= \sum_{\sigma \in G} \sum_{K \in \mathcal{T}_\sigma} F_{K,\sigma}^G(u)(\gamma_\sigma u - u_K) \\ &= \sum_{\sigma \in G \cap \mathcal{E}_{\mathrm{int}}, \mathcal{T}_\sigma = \{K_G, L\}} F_{K_G,\sigma}^G(u)(u_L - u_{K_G}) + \sum_{\sigma \in G \cap \mathcal{E}_{\mathrm{ext}}, \mathcal{T}_\sigma = \{K_G\}} F_{K_G,\sigma}^G(u)(\gamma_\sigma u - u_{K_G}) \end{aligned}$$

Since for all $\sigma \in \mathcal{E}_{\mathrm{int}}$ with $\mathcal{T}_\sigma = \{K_G, L\}$, we have $u_L - u_{K_G} = \frac{d_{K_G,\sigma} + d_{L,\sigma}}{d_{K_G,\sigma}} (\gamma_\sigma u - u_{K_G})$, if we let $d_\sigma = d_{K_G,\sigma} + d_{L,\sigma}$ for all $\sigma \in \mathcal{E}_{\mathrm{int}}$ with $\mathcal{T}_\sigma = \{K_G, L\}$ and $d_\sigma = d_{K_G,\sigma}$ for all $\sigma \in \mathcal{E}_{\mathrm{ext}}$, then we can deduce that

$$a_{\mathcal{T}_G}(u,u) = \sum_{\sigma \in G} \frac{d_\sigma}{d_{K_G,\sigma}} F_{K_G,\sigma}^G(u)(\gamma_\sigma u - u_{K_G}).$$

Now, by (19), $F_{K_G,\sigma}^G(u)$ depends linearly on $\{u_L - u_{K_G}\}_{L \in \mathcal{T}_G \setminus \{K_G\}}$ (and u_{K_G} if $\sigma \in \mathcal{E}_{\mathrm{ext}}$), and it can therefore be written as

$$F_{K_G,\sigma}^G(u) = \sum_{\sigma' \in G} a_{\sigma,\sigma'}^G \frac{d_{\sigma'}}{d_{K_G,\sigma'}} (\gamma_{\sigma'} u - u_{K_G})$$

where $\{a_{\sigma,\sigma'}^G\}_{\sigma,\sigma' \in G \times G}$ is a family of reals. We obtain that

$$a_{\mathcal{T}_G}(u,u) \quad = \quad \sum_{(\sigma,\sigma')\ \in G \times G} \frac{d_\sigma}{d_{K_G,\sigma}} \frac{d_{\sigma'}}{d_{K_G,\sigma'}} a_{\sigma,\sigma'}^G (\gamma_{\sigma'} u - u_{K_G})(\gamma_\sigma u - u_{K_G}).$$

We denote by $X^G(u)$ the vector of size d defined by the family $\left\{ \frac{\sqrt{d_\sigma\, \mathrm{m}_\sigma}}{d_{K_G,\sigma}} (\gamma_\sigma u - u_{K_G}) \right\}_{\sigma \in G}$ and by A^G the matrix of size d defined by the family of reals $\left\{ \sqrt{\frac{d_\sigma\, d_{\sigma'}}{\mathrm{m}_\sigma\, \mathrm{m}_{\sigma'}}} a_{\sigma,\sigma'}^G \right\}_{\sigma,\sigma' \in G \times G}$. Then, we can write $a_{\mathcal{T}_G}(u,u)$ under the form

$$a_{\mathcal{T}_G}(u,u) = (A^G X^G(u)) \cdot X^G(u)$$

or again,

$$a_{\mathcal{T}_G}(u,u) = \left(\frac{A^G + (A^G)^t}{2} X^G(u)\right) \cdot X^G(u) \tag{23}$$

where $(A^G)^t$ is the transpose matrix of A^G. From (23), we deduce that γ_2 is the smallest eigenvalue of the matrix $\frac{A^G+(A^G)^t}{2}$ because the Euclidean norm of the vector $X^G(u)$ is exactly equal to $\|u\|_{\mathcal{T}_G}$. For a given $\epsilon > 0$, let

$$\begin{cases} g_\epsilon(x) = \frac{\epsilon^2}{\epsilon - x} & \text{if } x < 0, \\ g_\epsilon(x) = x + \epsilon & \text{otherwise,} \end{cases}$$

and, for all $G \in \mathcal{G}$, define $\beta^G = g_\epsilon(\gamma_2)$. The weights are defined as

$$\theta_\sigma^G = \frac{\beta^G}{\sum_{G' \in \mathcal{G}_\sigma} \beta^{G'}} \quad \forall G \in \mathcal{G},\ \forall \sigma \in G.$$

Therefore, for a given $G \in \mathcal{G}$, the larger γ_2, the more the subfluxes $\{F_{K,\sigma}^G\}_{K \in \mathcal{T}_G,\ \sigma \in \mathcal{E}_K \cap G}$ will contribute to the global fluxes $\{F_{K,\sigma}\}_{K \in \mathcal{T},\ \sigma \in \mathcal{E}_K \cap G}$. In the numerical tests of §4 we have taken $\epsilon = 0.1$.

3.3 Convergence

Property (P2) is only conditionally verified by non symmetric methods. We propose a computable criterion issued from the stronger assumption that coercivity holds locally around each vertex $s \in \mathcal{V}$. For a given discretization and diffusion tensor, this assumption can be checked numerically by computing the eigenvalues of a small linear system of size $\mathrm{card}(\mathcal{E}_s) \leq \varrho_3$ for each vertex $s \in \mathcal{V}$.

Let $s \in \mathcal{V}$, and set $\mathcal{H}_{\mathcal{T}_s} \stackrel{\text{def}}{=} \{u_K \in \mathbb{R},\ K \in \mathcal{T}_s\}$. The space $\mathcal{H}_{\mathcal{T}_s}$ is endowed with the semi-norm

$$\|u\|_{\mathcal{T}_s}^2 \stackrel{\text{def}}{=} \sum_{K \in \mathcal{T}_s} \sum_{\sigma \in \mathcal{E}_s \cap \mathcal{E}_K} \frac{\mathrm{m}_\sigma}{d_{K,\sigma}} (\gamma_\sigma u - u_K)^2.$$

Denote by $a_{\mathcal{T}_s}$ the bilinear form defined as follows: For all $u,\ v \in \mathcal{H}_{\mathcal{T}_s}$,

$$a_{\mathcal{T}_s}(u,v) \stackrel{\text{def}}{=} \sum_{G \in \mathcal{G}, G \subset \mathcal{E}_s} \sum_{K \in \mathcal{T}_G} \sum_{\sigma \in \mathcal{E}_K \cap G} \theta_\sigma^G F_{K,\sigma}^G(u)(\gamma_\sigma v - v_K).$$

Lemma 8. *Let there be a positive constant γ_3 s.t.*

$$\min_{s \in \mathcal{V}} \inf_{\{v \in \mathcal{H}_{\mathcal{T}_s} \,|\, \|v\|_{\mathcal{T}_s} = 1\}} a_{\mathcal{T}_s}(v,v) \geq \gamma_3. \tag{24}$$

Then, for all $u \in \mathcal{H}_{\mathcal{T}}$, $a_{\mathcal{T}}(u,u) \geq \gamma_3 \|u\|_{\mathcal{T}}^2$.

Proof. For all $u \in \mathcal{H}_{\mathcal{T}}$ and $s \in \mathcal{V}$, let $u_s \stackrel{\text{def}}{=} (u_K)_{K \in \mathcal{T}_s} \in \mathcal{H}_{\mathcal{T}_s}$. Since any given group G only belongs to one particular $\mathcal{E}_s$, it is easy to see that $a_{\mathcal{T}}(u,u) = \sum_{s \in \mathcal{V}} a_{\mathcal{T}_s}(u_s, u_s)$, and thus that $a_{\mathcal{T}}(u,u) \geq \gamma_3 \sum_{s \in \mathcal{V}} ||u_s||_{\mathcal{T}_s}^2$. The assert then follows from

$$\sum_{s \in \mathcal{V}} \|u_s\|_{\mathcal{T}_s}^2 \geq \|u\|_{\mathcal{T}}^2,$$

which is straightforward since, for all $K \in \mathcal{T}$ and for all $\sigma \in \mathcal{E}_K$, $\mathrm{card}(\{s \in \mathcal{V} \,|\, \sigma \in \mathcal{E}_s\}) \geq 1$. □

Theorem 3 (Convergence). *Let $\{\mathcal{D}_n\}_{n \in \mathbb{N}}$ be a family of meshes matching Definitions 1 and 2 and s.t. $h_{\mathcal{D}_n} \to 0$ as $n \to \infty$. Suppose, furthermore, that (24) holds with γ_3 not depending on $n \in \mathbb{N}$ and that there exists $\gamma_4 < +\infty$ s.t.*

$$\forall n \in \mathbb{N},\ \forall \sigma \in \mathcal{E}_n,\ \sum_{G \in \mathcal{G}_\sigma} \theta_\sigma^G |\mathcal{A}_G^{-1}| \leq \gamma_4. \tag{25}$$

Then, as $n \to \infty$, the sequence $\{u_n\}_{n \in \mathbb{N}}$ of discrete solutions of problem (5) with numerical fluxes defined by (22) converges to the solution $\overline{u}$ of (1) in $L^q(\Omega)$ for all $q \in [1, 2d/(d-2))$ (and weakly in $L^{2d/(d-2)}(\Omega)$ if $d > 2$).

Proof. It suffices to verify the requirements listed in Hypothesis 1. According to Lemma 6, the choice $\mathfrak{D} = \mathcal{Q}$ meets (P1). The property (P2) holds (with $I_\sigma = \gamma_\sigma$) under (24), by Lemma 8. Finally, the consistency of the fluxes (P3) can be obtained by proving the strong consistency (see Remark 3). Indeed, for all $n \in \mathbb{N}$, $K \in \mathcal{T}_n$ and $\sigma \in \mathcal{E}_K$, (22), (21), Lemma 7 and (25) yield

$$\begin{aligned}\left|F_{K,\sigma}(\varphi_{\mathcal{T}_n}) - \frac{1}{\mathrm{m}_K}\int_K \Lambda(x)\nabla\varphi(x)\cdot \mathrm{m}_\sigma\,\mathbf{n}_{K,\sigma}\right| &\le |F_{K,\sigma}(\varphi_{\mathcal{T}_n}) - \Lambda_K\nabla\varphi(x_K)\,\mathrm{m}_\sigma\,\mathbf{n}_{K,\sigma}| \\ &\quad + \left|\frac{1}{\mathrm{m}_K}\int_K \Lambda(x)(\nabla\varphi(x) - \nabla\varphi(x_K))\cdot \mathrm{m}_\sigma\,\mathbf{n}_{K,\sigma}\right| \\ &\le \mathrm{m}_\sigma\beta_0 \sum_{G\in\mathcal{G}_\sigma}\theta_\sigma^G\left|(\nabla_{\mathcal{D}_n}\varphi_{\mathcal{T}_n})_K^{G,\sigma} - \nabla\varphi(x_K)\right| \\ &\quad + \mathrm{m}_\sigma\beta_0 \sup_{x\in K}|\nabla\varphi(x) - \nabla\varphi(x_K)| \\ &\le (C_5\gamma_4 + C_3)\mathrm{m}_\sigma\beta_0 h_{\mathcal{D}_n},\end{aligned}$$

where $C_3 = \sup_{x\in P_\Omega}|\varphi''(x)|$. □

3.4 Convergence of the gradient reconstruction

In order to prove the convergence of the gradient reconstruction (14), Hypothesis 2 must be verified; this can be achieved by adding a rather benign assumption on the mesh families.

Hypothesis 3. *$\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ is a mesh family matching Definition 1 and there exists a non-negative constant ϱ_4 independent of n s.t.*

$$\max_{K\in\mathcal{T}_n}\max_{\sigma\in\mathcal{E}_K}\frac{\mathrm{diam}(K)^{d-1}}{\mathrm{m}_\sigma} \le \varrho_4.$$

This assumption and (2) allow to uniformly bound the cardinals of $\mathcal{G}_\sigma$. Indeed, since each cell K is star-shaped with respect to x_K we have, for all $\sigma \in \mathcal{E}_K$ and all $x \in \sigma$, $(x - x_K)\cdot\mathbf{n}_{K,\sigma} = d_{K,\sigma} \ge \varrho_1\mathrm{diam}(K)$ and thus, by Stokes' formula,

$$\begin{aligned}d\,\mathrm{m}_K = \int_K \mathrm{div}(x - x_K)\,\mathrm{d}x &= \sum_{\sigma\in\mathcal{E}_K}\int_\sigma (x - x_K)\cdot\mathbf{n}_{K,\sigma} \\ &\ge \varrho_1\mathrm{diam}(K)\sum_{\sigma\in\mathcal{E}_K}\mathrm{m}_\sigma \ge \frac{\varrho_1}{\varrho_4}\mathrm{diam}(K)^d\mathrm{card}(\mathcal{E}_K).\end{aligned}$$

As $\mathrm{m}_K \le \omega_d\mathrm{diam}(K)^d$, ω_d being the volume of the unit ball in $\mathbb{R}^d$, this shows that $\mathrm{card}(\mathcal{E}_K) \le \frac{d\omega_d\varrho_4}{\varrho_1}$; but, if $\mathcal{T}_\sigma = \{K, L\}$, $\mathcal{G}_\sigma$ is contained in the set of families of d faces chosen in $\mathcal{E}_K$ or $\mathcal{E}_L$, and there thus exists C_6 only depending on ϱ_4, ϱ_1 s.t.

$$\max_{\sigma\in\mathcal{E}_n}\mathrm{card}(\mathcal{G}_\sigma) \le C_6. \tag{26}$$

Lemma 9. *Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be a family of discretizations matching Definition 1 and 2 and assume that Hypothesis 3 holds. We also assume that* (25) *is satisfied. Then,* (16) *holds.*

Proof. For simplicity of notation, the subscript n will be suppressed throughout the proof, which holds for a generic element of the mesh family $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$. Let $v \in H_{\mathcal{T}}(\Omega)$

(i) For $G \in \mathcal{G}$, we estimate $|(\nabla_{\mathcal{D}}v)_{K_G}^G|$. Since $(\nabla_{\mathcal{D}}v)_{K_G}^G$ solves (19), and recalling that $\gamma_\sigma v = 0$ for all $\sigma \in \mathcal{E}_{\mathrm{ext}}$,

$$|(\nabla_{\mathcal{D}}v)_{K_G}^G|^2 \le |\mathcal{A}_G^{-1}|^2\beta_0^2\left[\sum_{\sigma\in G\cap\mathcal{E}_{\mathrm{int}},\mathcal{T}_\sigma=\{K_G,L\}}\frac{(v_L - v_{K_G})^2}{d_{L,\sigma}^2} + \sum_{\sigma\in G\cap\mathcal{E}_{\mathrm{ext}}}\frac{(\gamma_\sigma v - v_{K_G})^2}{d_{K_G,\sigma}^2}\right].$$

Observe that, for all $\sigma \in \mathcal{E}_{\rm int}$, denoting $\mathcal{T}_\sigma = \{K, L\}$ we have $\frac{v_L - v_K}{d_{K,\sigma} + d_{L,\sigma}} = \frac{\gamma_\sigma v - v_K}{d_{K,\sigma}}$. As a consequence, by (2),

$$\frac{(v_L - v_K)^2}{d_{L,\sigma}^2} \leq \frac{(\gamma_\sigma v - v_K)^2}{d_{K,\sigma}^2} \left(1 + \frac{1}{\varrho_2}\right)^2 \tag{27}$$

and

$$\mathrm{m}_{K_G} |(\nabla_{\mathcal{D}} v)_{K_G}^G|^2 \leq |\mathcal{A}_G^{-1}|^2 \beta_0^2 \left(1 + \frac{1}{\varrho_2}\right)^2 \sum_{\sigma \in G} \frac{\mathrm{m}_{K_G}}{d_{K_G,\sigma}^2} (\gamma_\sigma v - v_{K_G})^2. \tag{28}$$

(ii) For $G \in \mathcal{G}$, $\sigma \in G \cap \mathcal{E}_{\rm int}$ with $\mathcal{T}_\sigma = \{K_G, L\}$, we estimate $|(\nabla_{\mathcal{D}} v)_L^{G,\sigma}|$. Owing to (20),

$$|(\nabla_{\mathcal{D}} v)_L^{G,\sigma}|^2 \leq 2 \left[|(\nabla_{\mathcal{D}} v)_{K_G}^G|^2 \left(1 + \frac{|x_L - x_{K_G}|}{d_{L,\sigma}}\right)^2 + \frac{|v_L - v_{K_G}|^2}{d_{L,\sigma}^2} \right]. \tag{29}$$

Using (2) and observing that $|x_L - x_{K_G}| \leq \mathrm{diam}(L) + \mathrm{diam}(K_G)$, we have that

$$\frac{|x_L - x_{K_G}|}{d_{L,\sigma}} \leq \frac{1}{\varrho_1} \left(1 + \frac{1}{\varrho_2}\right). \tag{30}$$

Then, from (29) together with (28), (30) and (27), we deduce that there exists $C_7 > 0$ which solely depends on β_0, ϱ_1 and ϱ_2 s.t.

$$\mathrm{m}_L |(\nabla_{\mathcal{D}} v)_L^{G,\sigma}|^2 \leq C_7 (1 + |\mathcal{A}_G^{-1}|^2) \frac{\mathrm{m}_L}{\mathrm{m}_{K_G}} \sum_{\sigma' \in G} \frac{\mathrm{m}_{K_G}}{d_{K_G,\sigma'}^2} (\gamma_{\sigma'} v - v_{K_G})^2. \tag{31}$$

(iii) Stability. The fact that K is star-shaped with respect to x_K and the definition of $d_{K,\sigma}$ show that K contains the ball centered at x_K and with radius $\inf_{\sigma \in \mathcal{E}_K} d_{K,\sigma}$; we deduce from (2) that $C_8 \mathrm{diam}(K)^d \leq \mathrm{m}_K \leq C_9 \mathrm{diam}(K)^d$ (with C_8 and C_9 only depending on ϱ_1) and thus that, if K and L are two neighborhing grid cells, there exists $C_{10} > 0$ only depending on ϱ_2 and ϱ_1 s.t. $\frac{\mathrm{m}_L}{\mathrm{m}_K} \leq C_{10}$. Hence, thanks to (28) and (31), there exists a real $C_{11} > 0$ solely depending on ϱ_1, ϱ_2 and β_0 s.t.

$$\forall \sigma \in G\,,\ \forall K \in \mathcal{T}_\sigma, \quad \mathrm{m}_K |(\nabla_{\mathcal{D}} v)_K^{G,\sigma}|^2 \leq C_{11} (1 + |\mathcal{A}_G^{-1}|^2) \sum_{\sigma' \in G} \frac{\mathrm{m}_{K_G}}{d_{K_G,\sigma'}^2} (\gamma_{\sigma'} v - v_{K_G})^2. \tag{32}$$

Furthermore, observe that, for all $\sigma' \in G$, K_G belongs to $\mathcal{T}_{\sigma'}$, so that

$$\frac{\mathrm{m}_{K_G}}{d_{K_G,\sigma'}^2} (\gamma_{\sigma'} v - v_{K_G})^2 \leq \sum_{K \in \mathcal{T}_{\sigma'}} \frac{\mathrm{m}_K}{d_{K,\sigma'}^2} (\gamma_{\sigma'} v - v_K)^2. \tag{33}$$

We infer from (3) and the definition (22) of the fluxes that

$$\begin{aligned} T \stackrel{\rm def}{=} \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \frac{d_{K,\sigma}}{\mathrm{m}_\sigma} |F_{K,\sigma}(v)|^2 &= \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \frac{d_{K,\sigma}}{\mathrm{m}_\sigma} \left| \mathrm{m}_\sigma \Lambda_K \sum_{G \in \mathcal{G}_\sigma} \theta_\sigma^G (\nabla_{\mathcal{D}} v)_K^{G,\sigma} \cdot \mathbf{n}_{K,\sigma} \right|^2 \\ &\leq d \beta_0^2 \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \mathrm{m}_K \left| \sum_{G \in \mathcal{G}_\sigma} \theta_\sigma^G (\nabla_{\mathcal{D}} v)_K^{G,\sigma} \cdot \mathbf{n}_{K,\sigma} \right|^2. \end{aligned}$$

Equation (25) gives in particular $\theta_\sigma^G |\mathcal{A}_G^{-1}| \leq \gamma_4$ and thus, by Cauchy-Schwarz inequality, (26) and (32), we deduce

$$\begin{aligned} T &\leq d \beta_0^2 \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \mathrm{m}_K \sum_{G \in \mathcal{G}_\sigma} (\theta_\sigma^G)^2 \left| (\nabla_{\mathcal{D}} v)_K^{G,\sigma} \right|^2 \times \mathrm{card}(\mathcal{G}_\sigma) \\ &\leq d \beta_0^2 C_6 C_{11} (1 + \gamma_4^2) \sum_{K \in \mathcal{T}} \sum_{\sigma \in \mathcal{E}_K} \sum_{G \in \mathcal{G}_\sigma} \sum_{\sigma' \in G} \frac{\mathrm{m}_{K_G}}{d_{K_G,\sigma'}^2} (\gamma_{\sigma'} v - v_{K_G})^2. \end{aligned}$$

We then permute some sums and use the fact that each $\mathcal{T}_\sigma$ has one or two elements to obtain C_{12} only depending on ϱ_1, ϱ_2, β_0 and γ_4 s.t.

$$T \le C_{12} \sum_{\sigma\in\mathcal{E}} \sum_{K\in\mathcal{T}_\sigma} \left(\sum_{G\in\mathcal{G}_\sigma} \sum_{\sigma'\in G} \frac{\mathrm{m}_{K_G}}{d^2_{K_G,\sigma'}} (\gamma_{\sigma'} v - v_{K_G})^2 \right)$$
$$\le 2C_{12} \sum_{\sigma\in\mathcal{E}} \sum_{G\in\mathcal{G}_\sigma} \sum_{\sigma'\in G} \frac{\mathrm{m}_{K_G}}{d^2_{K_G,\sigma'}} (\gamma_{\sigma'} v - v_{K_G})^2 = 2C_{12} \sum_{G\in\mathcal{G}} \sum_{\sigma\in G} \sum_{\sigma'\in G} \frac{\mathrm{m}_{K_G}}{d^2_{K_G,\sigma'}} (\gamma_{\sigma'} v - v_{K_G})^2.$$

But each group G has cardinal d and thus, by (33),

$$T \le 2C_{12}d \sum_{G\in\mathcal{G}} \sum_{\sigma'\in G} \frac{\mathrm{m}_{K_G}}{d^2_{K_G,\sigma'}} (\gamma_{\sigma'} v - v_{K_G})^2 = 2C_{12}d \sum_{\sigma'\in\mathcal{E}} \sum_{G\in\mathcal{G}_{\sigma'}} \frac{\mathrm{m}_{K_G}}{d^2_{K_G,\sigma'}} (\gamma_{\sigma'} v - v_{K_G})^2$$
$$\le 2C_{12}d \sum_{\sigma'\in\mathcal{E}} \left(\sum_{K\in\mathcal{T}_{\sigma'}} \frac{\mathrm{m}_K}{d^2_{K,\sigma'}} (\gamma_{\sigma'} v - v_K)^2 \times \mathrm{card}(\mathcal{G}_{\sigma'}) \right).$$

Hypothesis 3, Equation (2) and $\mathrm{m}_K \le C_9 \mathrm{diam}(K)^d$ imply that $\frac{\mathrm{m}_K}{d^2_{K,\sigma'}} \le \frac{C_9\varrho_4}{\varrho_1} \frac{\mathrm{m}_{\sigma'}}{d_{K,\sigma'}}$ and we thus infer

$$\sum_{K\in\mathcal{T}} \sum_{\sigma\in\mathcal{E}_K} \frac{d_{K,\sigma}}{\mathrm{m}_\sigma} |F_{K,\sigma}(v)|^2 \le 2C_{12}dC_6 \frac{C_9\varrho_4}{\varrho_1} \|v\|^2_{\mathcal{T}},$$

which concludes the proof. □

The following result is a direct consequence of Theorem 2 together with Lemma 9.

Theorem 4. *Let $\overline{u}$ be the solution to (1). Let $\{\mathcal{D}_n\}_{n\in\mathbb{N}}$ be a family of meshes matching Definitions 1 and 2 and s.t. $h_{\mathcal{D}_n} \to 0$ as $n \to \infty$ and denote by u_n the solution of (5) with numerical fluxes defined by (22) on $\mathcal{D}_n$. Then, if (25) and Hypothesis 3 hold, the sequence $\{\overline{\nabla}_{\mathcal{D}_n} u_n\}_{n\in\mathbb{N}}$ converges to $\nabla\overline{u}$ in $[L^2(\Omega)]^d$.*

4 Numerical tests

The objective of this section is to assess the performance of the method described in Example 2 on challenging diffusion problems combining mild or strong anisotropy, heterogeneity and distorted or skewed meshes. For the sake of completeness, a comparison is provided against (i) the method of [24, §2.2] refered to as Success; (ii) the MPFA O method of [1] and (iii) the MPFA L method of [5, 4], also described in Example 1. In the first test case, we consider the Dirichlet problem associated with the following exact solution featuring anisotropic permeability:

$$\bar{u} = \sin(\pi x)\sin(\pi y), \quad \Lambda = \mathrm{diag}(0.1, 1).$$

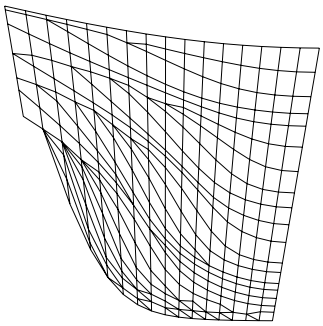
In the second test case, we consider the Dirichlet problem associated with the following exact solution featuring heterogeneous anisotropic permeability:

$$\bar{u} = \begin{cases} \sin(b\pi x)\sin(c\pi y) & \text{if } x \le \delta, \\ \sin(b\pi\delta)\sin(c\pi y) + \pi b \dfrac{a_1}{a_2} \cos(b\pi\delta)\sin(c\pi y)(x-\delta) & \text{otherwise} \end{cases}$$

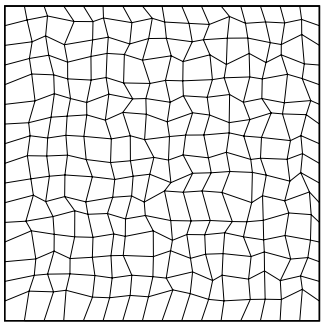
and

$$\Lambda = \begin{cases} \mathrm{diag}(a_1, b_1) & \text{if } x \le \delta, \\ \mathrm{diag}(a_2, b_2) & \text{otherwise,} \end{cases}$$

where $b = \frac{1}{1.7}$, $c = 1.9$, $a_1 = 1$, $b_1 = 10$, $a_2 = 5, b_2 = 1, \delta = 0.5$. Both tests have been run on (i) a family of Corner Point Geometry basin meshes with erosion (see Figure 3(a)) and (ii) a family of randomly distorted quadrangular meshes of $(0, 10) \times (0, 1)$ (see Figure 3(b)).



(a) Basin mesh. The actual aspect ratio is 10:1 (x:y)



(b) Randomly perturbed quadrangular mesh

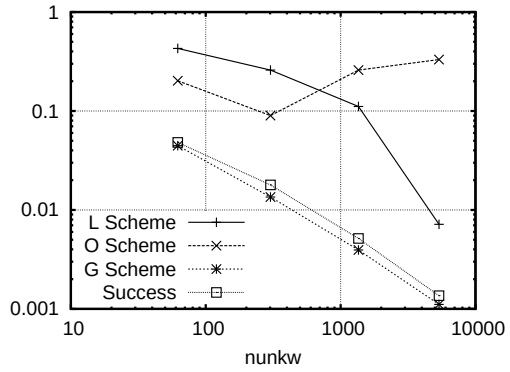
Figure 3: Mesh families.

The following indicators have been considered: `l2err`, the L^2 error; `ergrad`, the L^2 error on the gradient; `nit`, the number of preconditioned GMRes iterations; `nzmat`, the number of nonzero matrix entries; `umin`, the mininum of the discrete solution; `umax`, the maximum of the discrete solution. The number of degrees of freedom is denoted by `nunkw`. Blown up methods with respect to one indicator are not plotted to keep the scale readable. The linear systems have been solved with a direct solver for the indicators `l2err`, `umin`, `umax` and `ergrad`, whereas the GMRes algorithm from PETSc [13,12,11] with Hypre BoomerAMG preconditioner (see www.llnl.gov/CASC/hypre) has been used for `nit`. The stopping criterion required the preconditioned residual norm to be smaller than 10^{-7}. As expected, while sometimes displaying better accuracy, the Success scheme of [24, §2.2] has much denser matrices.

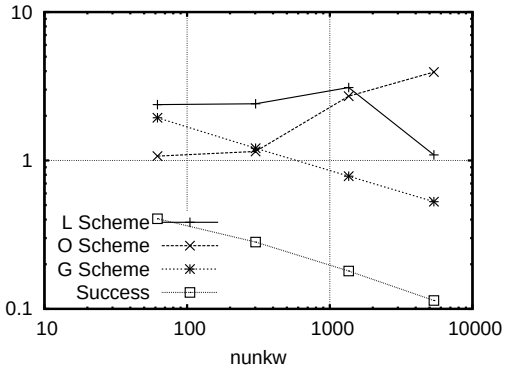
A Proof of Lemma 6

The proof is trivial if $\Lambda \in C(\overline{\Omega})$ since, in this case, $C_c^\infty(\Omega)$ is contained in $\mathcal{Q}$. The difficulty comes from the possible discontinuities of Λ through the interfaces of P_Ω, in which case item (iii) of Lemma 6 is not easy to obtain and might impose discontinuity of $\nabla\varphi$ through these interfaces. The proof is made in several steps, following the idea of [18]: we first eliminate the singularities (vertices if $d = 2$, vertices and edges if $d = 3$, etc.) of the boundaries of the open sets $\{\Omega_i\}_{1\le i\le N_\Omega}$ by showing that we only need approximate functions which vanish around these singularities; then we reason on each $\overline{\Omega_i}$, approximating a given function by functions having the same value on the boundary and vanishing derivatives in the direction $\Lambda\mathbf{n}$; gluing these approximations together, we obtain a function in $\mathcal{Q}$ which is close to the initial given function.

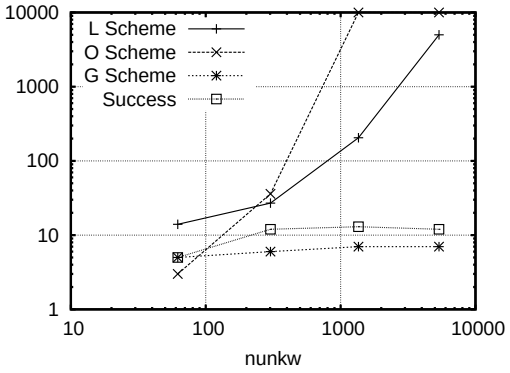
(i) Elimination of the singularities of $\{\Omega_i\}_{1\le i\le N_\Omega}$. First of all we notice that, since $C_c^\infty(\Omega)$ is dense in $H_0^1(\Omega)$, the result of the lemma follows if we prove that functions in $\mathcal{Q}$ approximate, in $H_0^1(\Omega)$, any $\psi \in C_c^\infty(\Omega)$.
Let S be the set of singularities of $\cup_{i=1}^{N_\Omega}\partial\Omega_i$ (i.e. affine parts of dimension $d-2$ or less: the vertices in dimension $d = 2$, the vertices and edges if $d = 3$, etc.); it is known that S has a 2-capacity equal to 0 and we can therefore find a sequence of functions $\gamma_n \in C_c^\infty(\mathbb{R}^d;[0,1])$ s.t. $\gamma_n \to 0$ in $H^1(\mathbb{R}^n)$ as $n \to \infty$ and, for all $n \in \mathbb{N}$, $\gamma_n \equiv 1$ on a neighborhood of S. If $\psi \in C_c^\infty(\Omega)$ and $\psi_n = (1-\gamma_n)\psi \in C_c^\infty(\Omega)$, then $\psi_n \to \psi$ in $H_0^1(\Omega)$ and, for all n, $\psi_n \equiv 0$ on a neighborhood of S. Hence, denoting by $C_{c,S}^\infty(\Omega)$ the set of functions in $C_c^\infty(\Omega)$ which vanish on neighborhoods of S, the proof of the lemma is complete if we can approximate, in $H_0^1(\Omega)$, elements of $C_{c,S}^\infty(\Omega)$ by elements of $\mathcal{Q}$.



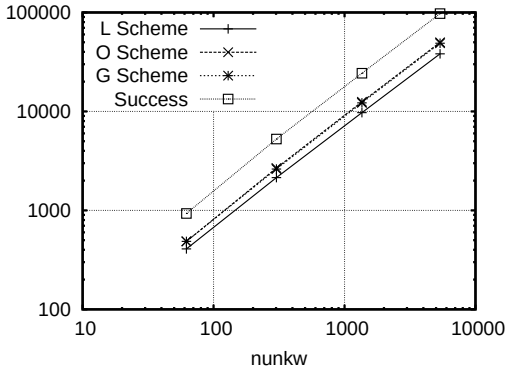
(a) `l2err`



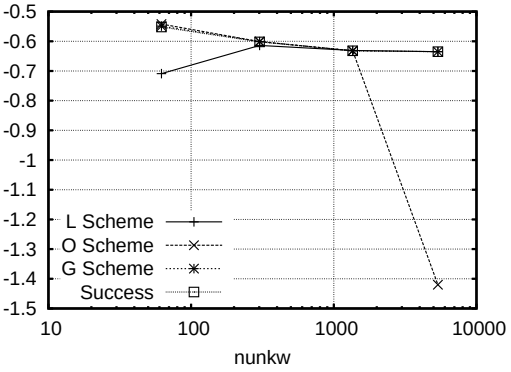
(b) `ergrad`



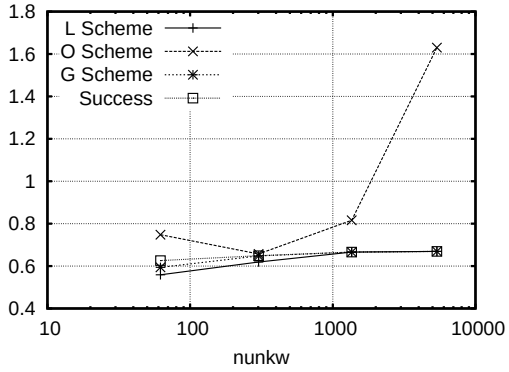
(c) `nit`



(d) `nzmat`

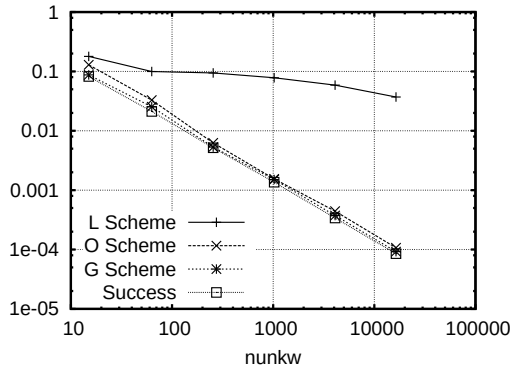


(e) `umin`=-0.6349341

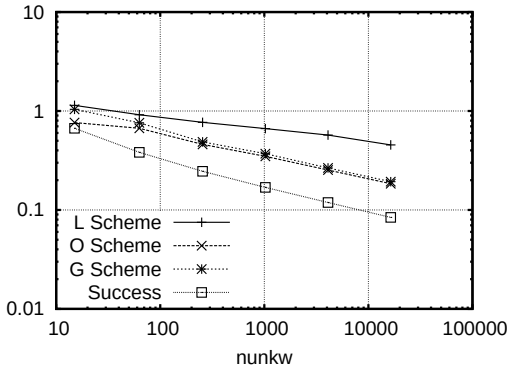


(f) `umax`=0.6687703

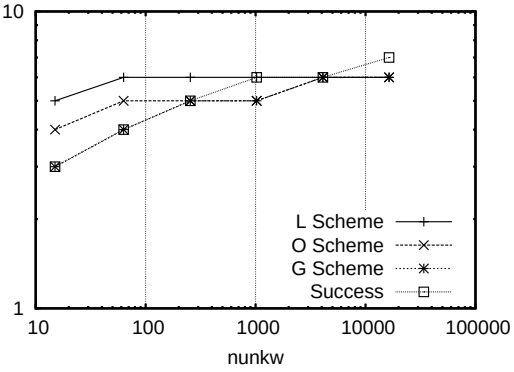
Figure 4: Numerical results for test case 1 on the basin mesh family.



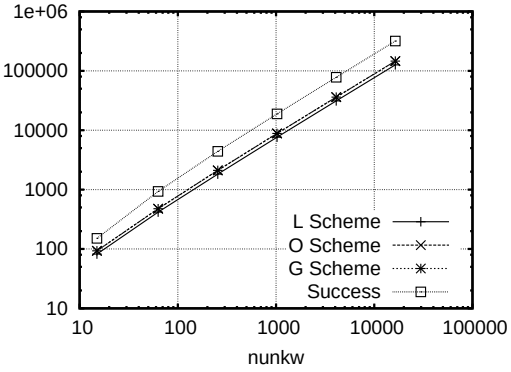
(a) `l2err`



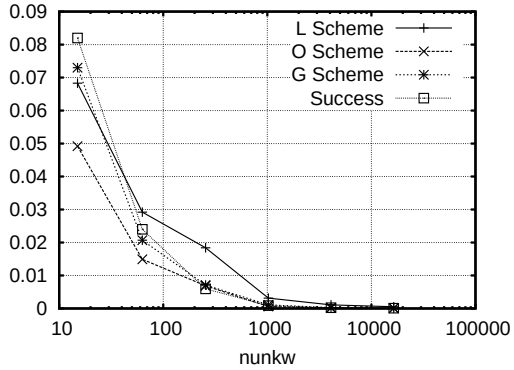
(b) `ergrad`=1



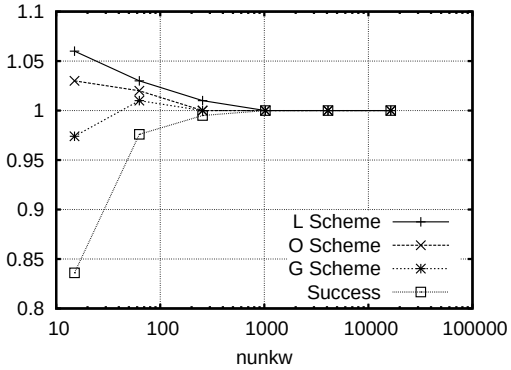
(c) `nit`



(d) `nzmat`

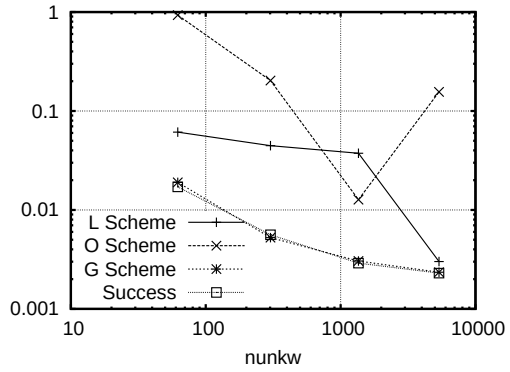


(e) `umin`=0

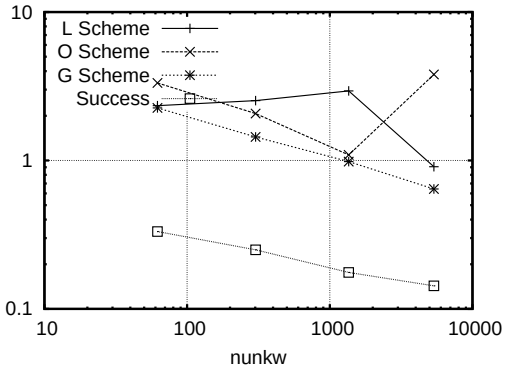


(f) `umax`=1

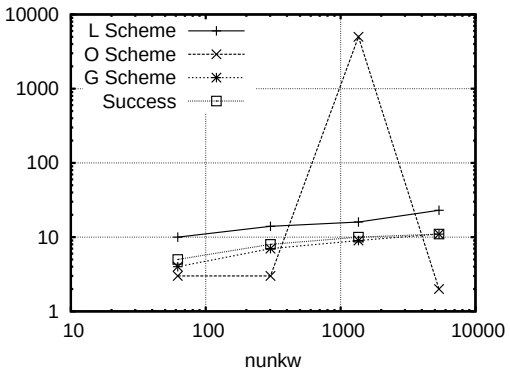
Figure 5: Numerical results for test case 1 on the randomly perturbed mesh family.



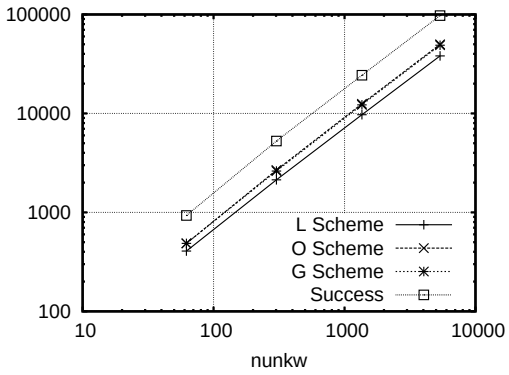
(a) `l2err`



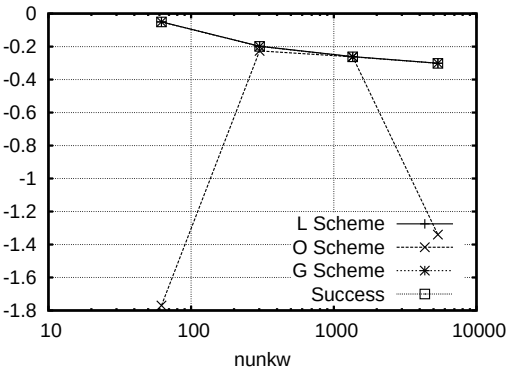
(b) `ergrad`



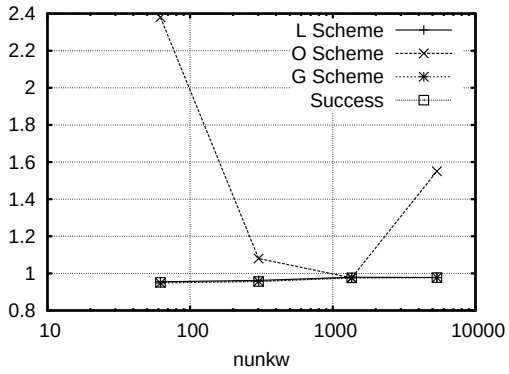
(c) `nit`



(d) `nzmat`



(e) `umin`=-0.3013243



(f) `umax`=1.114539

Figure 6: Numerical results for test case 2 on the basin mesh family.

(ii) Reduction to a Ω_i. Let $\psi \in C^\infty_{c,S}(\Omega)$ and assume that, for all $1 \leq i \leq N_\Omega$, there exists a sequence $\varphi^i_n \in C^2(\overline{\Omega_i})$ which converges to ψ in $H^1(\Omega_i)$ as $n \to \infty$ and s.t., for all $n \in \mathbb{N}$, $\varphi^i_n = \psi$ and $(\Lambda \nabla \varphi^i_n)_{|\overline{\Omega_i}} \cdot \mathbf{n}_i = 0$ on $\partial\Omega_i$. Define then $\varphi_n : \overline{\Omega} \to \mathbb{R}$ as the function equal to φ^i_n on $\overline{\Omega_i}$ for all $i = 1, \ldots, N_\Omega$; since $\varphi^i_n = \varphi^j_n = \psi$ on $\partial\Omega_i \cap \partial\Omega_j$, φ_n is well defined and continuous on $\overline{\Omega}$, it is C^2 on each $\overline{\Omega_i}$, it vanishes on $\partial\Omega$ (on which $\psi = 0$) and the tangential derivatives of φ_n are continuous through the interfaces of P_Ω (for all $\mathbf{t}$ parallel to $\partial\Omega_i \cap \partial\Omega_j$, the values of $(\nabla\varphi_n)_{|\overline{\Omega_i}} \cdot \mathbf{t}$ and $(\nabla\varphi_n)_{|\overline{\Omega_j}} \cdot \mathbf{t}$ on $\partial\Omega_i \cap \partial\Omega_j$ can be computed using only the values of $\varphi^i_n = \varphi^j_n = \psi$ on $\partial\Omega_i \cap \partial\Omega_j$, and are therefore equal). The continuity of φ_n accross the boundary of Ω_i for each i moreover ensures that $\nabla\varphi_n$ has no singularity on these boundaries and it is therefore simply the function equal to $\nabla\varphi^i_n$ on Ω_i for all i; hence, $\varphi_n \to \psi$ in $H^1_0(\Omega)$. Finally, the fluxes $\Lambda\nabla\varphi_n \cdot \mathbf{n}$ are clearly continuous through the interfaces of P_Ω since they vanish on either side of each such interface $\partial\Omega_i \cap \partial\Omega_j$.
To conclude the proof, it remains to find the convenient approximations $\{\varphi^i_n\}_{n\geq 1}$ of $\psi \in C^\infty_{c,S}(\Omega)$ on $\overline{\Omega_i}$.

(iii) Approximation on $\overline{\Omega_i}$. Let $\psi \in C^\infty_{c,S}(\Omega)$ and let $\mathcal{O}$ be an open set containing S s.t. $\psi \equiv 0$ on a neighborhood of $\overline{\mathcal{O}}$. Let $(F_l)_{1\leq l\leq r}$ be the faces of Ω_i (i.e. the affine parts of $\partial\Omega_i$ of dimension $d-1$); for all $1 \leq l \leq r$, we denote by $\mathbf{n}_l$ the unit normal to F_l pointing inside Ω_i and we define the C^2 function $f_l : \mathbb{R} \times F_l \to \mathbb{R}^d$ by

$$\forall t \in \mathbb{R},\ \forall y \in F_l\ ,\ f_l(t,y) = y + t\Lambda(y)\mathbf{n}_l. \tag{34}$$

If $(t,y) \in \mathbb{R} \times F_l$ and $(t',y') \in \mathbb{R} \times F_l$ are s.t. $f_l(t,y) = f_l(t',y')$ then, since $(y-y')\cdot\mathbf{n}_l = 0$, one has $t\Lambda(y)\mathbf{n}_l\cdot\mathbf{n}_l = t'\Lambda(y')\mathbf{n}_l\cdot\mathbf{n}_l$ and thus

$$y - y' = t\Lambda(y)\mathbf{n}_l\cdot\mathbf{n}_l \left(\frac{\Lambda(y')\mathbf{n}_l}{\Lambda(y')\mathbf{n}_l\cdot\mathbf{n}_l} - \frac{\Lambda(y)\mathbf{n}_l}{\Lambda(y)\mathbf{n}_l\cdot\mathbf{n}_l} \right). \tag{35}$$

Letting $\varepsilon > 0$ be smaller than the inverse of the Lipschitz constant of $y \to \beta_0 \frac{\Lambda(y)\mathbf{n}_l}{\Lambda(y)\mathbf{n}_l\cdot\mathbf{n}_l}$ (which is well-defined since $\Lambda(y)\mathbf{n}_l\cdot\mathbf{n}_l > \alpha_0$ for all y), (35) can happen with $y \neq y'$ only if $|t| \geq \varepsilon$. Hence, f_l is one-to-one on $(-\varepsilon,\varepsilon)\times F_l$. We also notice that $\Lambda(y)\mathbf{n}_l$ is uniformly transverse to the hyperplane H_l containing F_l (this is again $\Lambda(y)\mathbf{n}_l \cdot \mathbf{n}_l \geq \alpha_0$) and thus that, upon reducing ε, the Jacobian matrix of f_l at any $(t,y) \in (-\varepsilon,\varepsilon)\times F_l$ is invertible.
Let $\mathcal{V}_l$ be an open neighborhood of $\overline{F_l\backslash\mathcal{O}}$ in F_l s.t. $\mathrm{dist}(\mathcal{V}_l, S) > 0$; the preceding reasoning shows that f_l is a C^2-diffeomorphism from $(-\varepsilon,\varepsilon)\times\mathcal{V}_l$ to $f_l((-\varepsilon,\varepsilon)\times\mathcal{V}_l)$, an open set in $\mathbb{R}^d$ containing in particular $f_l(\{0\}\times\overline{F_l\backslash\mathcal{O}}) = \overline{F_l\backslash\mathcal{O}}$. Since $\Lambda(y)\mathbf{n}_l$ points inside Ω_i (one more time, $\Lambda(y)\mathbf{n}_l\cdot\mathbf{n}_l > 0$) and $\mathrm{dist}(\mathcal{V}_l, S) > 0$, upon reducing again ε if needed, we also see that $\mathcal{U}_l \stackrel{\mathrm{def}}{=} f_l([0,\varepsilon[\times\mathcal{V}_l)$ is contained in $\overline{\Omega_i}$ and is a neighborhood of $\overline{F_l\backslash\mathcal{O}}$ in $\overline{\Omega_i}$ (see Figure 7 for a representation of some sets appearing in this proof). Moreover, for all $x \in \mathcal{U}_l$, if $x = f_l(t,y)$ for $(t,y) \in [0,\varepsilon[\times\mathcal{V}_l$ then $\mathrm{dist}(x,H_l) = (x-y)\cdot\mathbf{n}_l = t\Lambda(y)\mathbf{n}_l\cdot\mathbf{n}_l$ and thus $0 \leq t \leq \frac{1}{\alpha_0}\mathrm{dist}(x,H_l)$. This shows that

$$\forall x \in \mathcal{U}_l\ ,\ \text{ if } (t,y) = (f_{l|[0,\varepsilon[\times\mathcal{V}_l})^{-1}(x) \text{ then } |x-y| \leq \frac{\beta_0}{\alpha_0}\mathrm{dist}(x,H_l). \tag{36}$$

Let us define ψ_l on $\mathcal{U}_l$ s.t.

$$\psi_l(f_l(t,y)) = \psi(y) \quad \text{for all } (t,y) \in [0,\varepsilon[\times\mathcal{V}_l. \tag{37}$$

ψ_l belongs to $C^2(\mathcal{U}_l)$ and $\psi_l = \psi$ on $\mathcal{V}_l$ (because $f_l(0,y) = y$); derivating (37) with respect to t, taking $t = 0$ and using (34) we also have

$$0 = \frac{d}{dt}(\psi_l(f_l(t,y)))_{|t=0} = \nabla\psi_l(y)\cdot\Lambda(y)\mathbf{n}_l = \Lambda(y)\nabla\psi_l(y)\cdot\mathbf{n}_l \quad \text{for all } y \in \mathcal{V}_l. \tag{38}$$

As ψ vanishes on a neighborhood of $\overline{\mathcal{O}}$, there exists a neighborhood $\mathcal{N}_l$ of $\mathcal{V}_l\cap\mathcal{O}$ in $\mathcal{V}_l$ s.t. $\psi = 0$ on $\mathcal{N}_l$; (37) then implies $\psi_l = 0$ on $f_l([0,\varepsilon[\times\mathcal{N}_l)$ which is, f_l being a diffeomorphism, a neighborhood in $\mathcal{U}_l$ of $f_l(\{0\}\times(\mathcal{V}_l\cap\mathcal{O})) = \mathcal{V}_l\cap\mathcal{O}$; to sum up,

$$\psi_l = 0 \text{ on a neighborhood of } \mathcal{V}_l\cap\mathcal{O} \text{ in } \mathcal{U}_l. \tag{39}$$

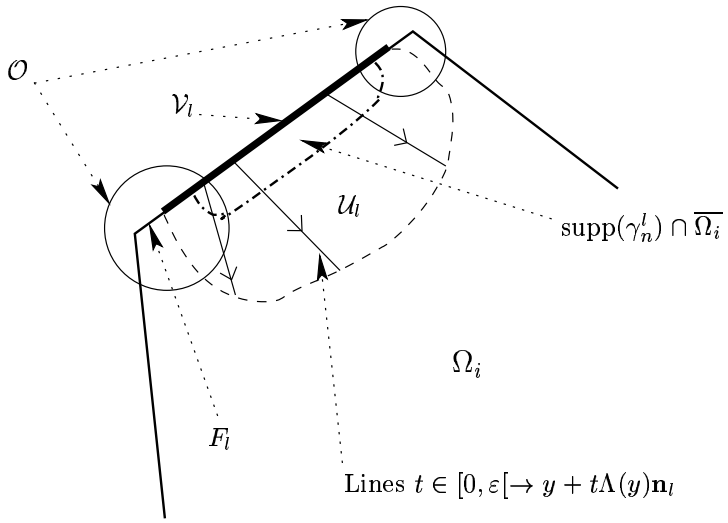


Figure 7: Various sets appearing in the proof of Lemma 6.

For $1 \leq l \leq r$, we take a sequence $\gamma_n^l \in C_c^\infty(\mathbb{R}^d; [0,1])$ s.t., for all $n \in \mathbb{N}$, $\gamma_n^l \equiv 1$ on a neighborhood of $\overline{F_l \backslash \mathcal{O}}$ and

$$\gamma_n^l \equiv 0 \text{ on } \{x \in \mathbb{R}^d \,,\ \mathrm{dist}(x, F_l \backslash \mathcal{O}) \geq 1/n\} \quad \text{and} \quad ||\nabla \gamma_n^l||_{L^\infty(\mathbb{R}^d)} \leq C_{13} n, \tag{40}$$

with C_{13} not depending on n. If n is large, $\mathrm{supp}(\gamma_n^l) \cap \overline{\Omega_i}$ is a compact subset of $\mathcal{U}_l$ and $\gamma_n^l \psi_l$ can therefore be extended to $\overline{\Omega_i}$ by 0 outside $\mathcal{U}_l$ without loosing smoothness; we then define $\varphi_n = \sum_{l=1}^r \gamma_n^l \psi_l + (1 - \sum_{l=1}^r \gamma_n^l)\psi \in C^2(\overline{\Omega_i})$. Since $\psi_l = \psi$ on $\mathcal{V}_l$ and, for n large enough, γ_n^l vanishes on $\partial\Omega_i$ outside $\mathcal{V}_l$, we have $\varphi_n = \psi$ on $\partial\Omega_i$ for such n. Still considering large n, on a neighborhood of $\overline{F_l \backslash \mathcal{O}}$ in $\overline{\Omega_i}$ we have $\gamma_n^l = 1$ and $\gamma_n^k = 0$ if $k \neq l$ and therefore, on such a neighborhood, $\varphi_n = \psi_l$; (38) thus shows that $\Lambda \nabla \varphi_n \cdot \mathbf{n} = 0$ on $\cup_{l=1}^r F_l \backslash \mathcal{O} = \partial\Omega_i \backslash \mathcal{O}$; since all the $\gamma_n^l \psi_l$ and ψ vanish on a neighborhood of $\partial\Omega_i \cap \mathcal{O}$ in $\overline{\Omega_i}$ (see (39)), we obviously also have $\Lambda \nabla \varphi_n \cdot \mathbf{n} = 0$ on $\partial\Omega_i \cap \mathcal{O}$, and thus on the whole boundary of Ω_i. It remains to prove that $\varphi_n \to \psi$ in $H^1(\Omega_i)$ as $n \to \infty$; in order to achieve this, we write $\varphi_n - \psi = \sum_{l=1}^r \gamma_n^l (\psi_l - \psi)$ and use (36), (37) and the smoothness of ψ to see that, if $\mathrm{dist}(x, F_l \backslash \mathcal{O}) \leq 1/n$, then $|\psi_l(x) - \psi(x)| \leq C_{14}/n$ with C_{14} not depending on n or x (because $x = f_l(t,y)$ with $y \in \mathcal{V}_l$ s.t. $|x - y| \leq \beta_0/(\alpha_0 n)$); we infer from (40) that

$$||\gamma_n^l (\psi_l - \psi)||_{L^2(\Omega_i)} \leq \frac{C_{14}}{n} \mathrm{meas}(\Omega_i)^{1/2}$$

and

$$||\nabla(\gamma_n^l (\psi_l - \psi))||_{L^2(\Omega_i)} \leq C_{13} C_{14} \mathrm{meas}(\Omega_i \cap \mathrm{supp}(\gamma_n^l))^{1/2} + ||\nabla(\psi_l - \psi)||_{L^2(\Omega_i \cap \mathrm{supp}(\gamma_n^l))}.$$

Since $\mathrm{meas}(\Omega_i \cap \mathrm{supp}(\gamma_n^l)) \to 0$ as $n \to \infty$, this concludes the proof that $\varphi_n \to \psi$ in $H^1(\Omega_i)$.

Remark 9. The proof shows that Λ need not be C^2 on the whole of each $\overline{\Omega_i}$, only on the affine parts of $\partial\Omega_i$ (and the reader can check that the rest of the paper only requires the C^1 regularity of Λ on each $\overline{\Omega_i}$).

B Proof of Lemma 7

Proposition 3. *Let $\mathcal{D}$ be a generic element of a family of discretizations satisfying Hypothesis 1. Let $\varphi \in \mathcal{Q}$, $\sigma \in \mathcal{E}_{\mathrm{int}}$ with $\mathcal{T}_\sigma = \{K, L\}$ and $y_\sigma \in \sigma$. Then $\nabla\varphi(x_K) - \nabla\varphi(x_L)$ can be decomposed as follows:*

$$\nabla\varphi(x_K) - \nabla\varphi(x_L) = \mu_\sigma \mathbf{n}_{K,\sigma} + \tau_\sigma \mathbf{t}_\sigma, \tag{41}$$

where $|\mathbf{t}_\sigma| = 1$, $\mathbf{t}_\sigma \cdot \mathbf{n}_{K,\sigma} = 0$ *and the reals* μ_σ, τ_σ *verify*

$$|\tau_\sigma| \leq C_{15} \left[\mathrm{diam}(L) + \mathrm{diam}(K)\right], \tag{42}$$

$$\mu_\sigma = -\frac{W_K(x_L)}{d_{L,\sigma}} + \tau_\sigma \frac{\mathbf{t}_\sigma \cdot (y_\sigma - x_L)}{d_{L,\sigma}} + \frac{W_K(y_\sigma) - W_L(y_\sigma)}{d_{L,\sigma}}, \tag{43}$$

with

$$W_K(x) \stackrel{\mathrm{def}}{=} \varphi(x) - \varphi(x_K) - \nabla\varphi(x_K) \cdot (x - x_K) \tag{44}$$

and $C_{15} \stackrel{\mathrm{def}}{=} \max(|\varphi''|_{L^\infty(K)}, |\varphi''|_{L^\infty(L)})$

Proof. The vector $\mathbf{t}_\sigma$ is obviously the normed orthogonal projection of $\nabla\varphi(x_K) - \nabla\varphi(x_L)$ on the hyperplane parallel to σ, and the reals μ_σ, τ_σ are given by the formulæ

$$\mu_\sigma = (\nabla\varphi(x_K) - \nabla\varphi(x_L)) \cdot \mathbf{n}_{K,\sigma}, \quad \tau_\sigma = (\nabla\varphi(x_K) - \nabla\varphi(x_L)) \cdot \mathbf{t}_\sigma.$$

Since

$$\begin{aligned} -W_K(x_L) + W_K(y_\sigma) - W_L(y_\sigma) &= -\varphi(x_L) + \varphi(x_K) + \nabla\varphi(x_K) \cdot (x_L - x_K) \\ &\quad + \varphi(y_\sigma) - \varphi(x_K) - \nabla\varphi(x_K) \cdot (y_\sigma - x_K) \\ &\quad - \varphi(y_\sigma) + \varphi(x_L) + \nabla\varphi(x_L) \cdot (y_\sigma - x_L) \\ &= \nabla\varphi(x_K) \cdot (x_L - y_\sigma) + \nabla\varphi(x_L) \cdot (y_\sigma - x_L) \\ &= (\nabla\varphi(x_K) - \nabla\varphi(x_L)) \cdot (x_L - y_\sigma), \end{aligned}$$

we can use (41) and the fact that $(x_L - y_\sigma) \cdot \mathbf{n}_{K,\sigma} = d_{L,\sigma}$ to re-write μ_σ under the form

$$\mu_\sigma = -\frac{W_K(x_L)}{d_{L,\sigma}} + \tau_\sigma \frac{\mathbf{t}_\sigma \cdot (y_\sigma - x_L)}{d_{L,\sigma}} + \frac{W_K(y_\sigma) - W_L(y_\sigma)}{d_{L,\sigma}}.$$

The face σ is completely contained either in one element of the partition P_Ω or in an interface of this partition; using then either the regularity of φ inside each element of the partition or the continuity of its tangential derivatives through the interfaces of P_Ω, we can re-write τ_σ under the form

$$\tau_\sigma = (\nabla\varphi(x_K) - \nabla\varphi(y_\sigma)) \cdot \mathbf{t}_\sigma + (\nabla\varphi(y_\sigma) - \nabla\varphi(x_L)) \cdot \mathbf{t}_\sigma$$

and the proof is complete since φ is C^2 on $\overline{K}$ and $\overline{L}$. □

Proposition 4 (Flux "quasi-continuity"). *Let* $\mathcal{D}$ *be a generic element of a family of discretizations satisfying Hypothesis 1 and* $\varphi \in \mathcal{Q}$. *For all* $G \in \mathcal{G}$, $\nabla\varphi(x_{K_G})$ *is the solution of a linear system of equations of the form*

$$\mathcal{A}_G Y_G = \mathcal{B}_G(\varphi_\mathcal{T}) + \mathcal{C}_G(\varphi),$$

where $\varphi_\mathcal{T} \in H_\mathcal{T}(\Omega)$ *is defined by the family* $\{\varphi(x_K)\}_{K \in \mathcal{T}}$, $\mathcal{A}_G$ *and* $\mathcal{B}_G(\varphi_\mathcal{T})$ *are the matrices defined in Lemma 5 and the vector* $\mathcal{C}_G(\varphi)$ *verifies*

$$|\mathcal{C}_G(\varphi)| \leq C_1 \max_{K \in \mathcal{T}_G} \mathrm{diam}(K) \tag{45}$$

with $C_1 > 0$ *which only depends on* ϱ_1, ϱ_2, Λ *and* φ.

Proof. For a cell K, let W_K be the function defined by (44). Since φ is C^2 regular on the closure of each element of P_Ω and since each cell is completely contained in one of these elements, there exists $C_{16} > 0$ only depending on φ s.t., for all $K \in \mathcal{T}$,

$$|W_K(x)| \leq C_{16} \mathrm{diam}(K)^2 \quad \text{for all } x \in \overline{K}. \tag{46}$$

For all $\sigma \in \mathcal{E}_{\mathrm{int}}$ with $\mathcal{T}_\sigma = \{K, L\}$ and $y_\sigma \in \sigma$, we apply Proposition 3 to decompose $\nabla\varphi(x_K) - \nabla\varphi(x_L)$ (note that the $W_K(x_L)$ appearing in (43) *is in general not* of order 2 with respect to the

size of the mesh, since $x_L \notin \overline{K}$ and φ is not regular accross the boundary of some cells). Since $\varphi \in \mathcal{Q}$, we can also write

$$\frac{1}{\mathrm{m}_\sigma}\int_\sigma (\Lambda\nabla\varphi)_{|\overline{K}}(x)\cdot\mathbf{n}_{K,\sigma}\,\mathrm{d}x + \frac{1}{\mathrm{m}_\sigma}\int_\sigma (\Lambda\nabla\varphi)_{|\overline{L}}(x)\cdot\mathbf{n}_{L,\sigma}\,\mathrm{d}x = 0 \tag{47}$$

and, $\nabla\varphi$ and Λ being C^1 on the closure of each control volume, we deduce from (47) that the real $\zeta_\sigma(\varphi) = \Lambda_K\nabla\varphi(x_K)\cdot\mathbf{n}_{K,\sigma} + \Lambda_L\nabla\varphi(x_L)\cdot\mathbf{n}_{L,\sigma}$ verifies

$$|\zeta_\sigma(\varphi)| \le C_{17}(\mathrm{diam}(K) + \mathrm{diam}(L)), \tag{48}$$

where $C_{17} > 0$ depends only on φ, Λ.

Let us now consider $G \in \mathcal{G}$, $\sigma \in G \cap \mathcal{E}_{\mathrm{int}}$ and use these preliminary remarks with $K = K_G$ and L s.t. $\mathcal{T}_\sigma = \{K_G, L\}$. By definition of $\zeta_\sigma(\varphi)$ and (41),

$$\begin{aligned}(\Lambda_L\mathbf{n}_{L,\sigma} + \Lambda_{K_G}\mathbf{n}_{K_G,\sigma})\cdot\nabla\varphi(x_{K_G}) &= \Lambda_L\mathbf{n}_{L,\sigma}\cdot\nabla\varphi(x_{K_G}) - \Lambda_L\mathbf{n}_{L,\sigma}\cdot\nabla\varphi(x_L) + \zeta_\sigma(\varphi)\\ &= -\Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}\mu_\sigma + \tau_\sigma\Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{t}_\sigma + \zeta_\sigma(\varphi).\end{aligned}$$

Equation (43) and the definition of $W_{K_G}(x_L)$ then show

$$\begin{aligned}&(\Lambda_L\mathbf{n}_{L,\sigma} + \Lambda_{K_G}\mathbf{n}_{K_G,\sigma})\cdot\nabla\varphi(x_{K_G})\\ &\quad= \frac{\Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(\varphi(x_L) - \varphi(x_{K_G})) - \frac{\Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}\nabla\varphi(x_{K_G})\cdot(x_L - x_{K_G})\\ &\qquad+\zeta_\sigma(\varphi) - \frac{\Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(\tau_\sigma\mathbf{t}_\sigma\cdot(y_\sigma - x_L) + W_{K_G}(y_\sigma) - W_L(y_\sigma)) + \tau_\sigma\Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{t}_\sigma\end{aligned}$$

and therefore

$$\begin{aligned}&\left(\Lambda_L\mathbf{n}_{L,\sigma} + \Lambda_{K_G}\mathbf{n}_{K_G,\sigma} + \frac{\Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(x_L - x_{K_G})\right)\cdot\nabla\varphi(x_{K_G}) =\\ &\qquad\qquad\frac{\Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(\varphi(x_L) - \varphi(x_{K_G})) + c_\sigma(\varphi)\end{aligned}$$

with $c_\sigma(\varphi) = \zeta_\sigma(\varphi) - \frac{\Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{n}_{L,\sigma}}{d_{L,\sigma}}(\tau_\sigma\mathbf{t}_\sigma\cdot(y_\sigma - x_L) + W_{K_G}(y_\sigma) - W_L(y_\sigma)) + \Lambda_L\mathbf{n}_{L,\sigma}\cdot\mathbf{t}_\sigma\tau_\sigma$. If $\sigma \in G \cap \mathcal{E}_{\mathrm{ext}}$, using the definition of $W_{K_G}(x_\sigma)$, we have

$$\frac{\Lambda_{K_G}\mathbf{n}_{K_G,\sigma}\cdot\mathbf{n}_{K_G,\sigma}}{d_{K_G,\sigma}}\nabla\varphi(x_{K_G})\cdot(x_\sigma - x_{K_G}) = \frac{\Lambda_{K_G}\mathbf{n}_{K_G,\sigma}\cdot\mathbf{n}_{K_G,\sigma}}{d_{K_G,\sigma}}(-\varphi(x_{K_G})) + c_\sigma(\varphi)$$

with $c_\sigma(\varphi) = -\frac{\Lambda_{K_G}\mathbf{n}_{K_G,\sigma}\cdot\mathbf{n}_{K_G,\sigma}}{d_{K_G,\sigma}}W_{K_G}(x_\sigma)$.

We deduce that $\nabla\varphi(x_{K_G})$ is the solution of the linear system of equations

$$\mathcal{A}_G Y_G = \mathcal{B}_G(\varphi_\mathcal{T}) + \mathcal{C}_G(\varphi),$$

where $\mathcal{C}_G(\varphi)$ is the vector of $\mathbb{R}^d$ defined by $\{c_\sigma(\varphi)\}_{\sigma\in G}$. Thanks to (48), (42) and (46), there exists $C_{18} > 0$ which only depends on ϱ_1, ϱ_2, Λ and φ s.t., for all $\sigma \in G$ with $\mathcal{T}_\sigma = \{K_G, L\}$, $|c_\sigma(\varphi)| \le C_{18}(\mathrm{diam}(L) + \mathrm{diam}(K_G))$. The proof is complete. □

We are now in a position to prove Lemma 7. Let W_K the function defined by (44) and recall that (46) holds. Since $(\nabla_\mathcal{D}\varphi_\mathcal{T})^G_{K_G}$ is the solution of the linear system (19) with $v = \varphi_\mathcal{T}$, we can deduce from Proposition 4 that $\nabla\varphi(x_{K_G}) - (\nabla_\mathcal{D}\varphi_\mathcal{T})^G_{K_G}$ is the solution of the linear system $\mathcal{A}_G Z_G = \mathcal{C}_G(\varphi)$ where the vector $\mathcal{C}_G(\varphi)$ satisfies (45). We obtain

$$|\nabla\varphi(x_{K_G}) - (\nabla_\mathcal{D}\varphi_\mathcal{T})^G_{K_G}| \le C_1|\mathcal{A}_G^{-1}|\max_{K\in\mathcal{T}_G}\mathrm{diam}(K). \tag{49}$$

For all $\sigma \in G \cap \mathcal{E}_{\rm int}$ with $\mathcal{T}_\sigma = \{K_G, L\}$, thanks to (20) with $v = \varphi_\mathcal{T}$, we have

$$(\nabla_\mathcal{D}\varphi_\mathcal{T})_L^{G,\sigma} = (\nabla_\mathcal{D}\varphi_\mathcal{T})_{K_G}^G - \frac{R_{L,\sigma}(\varphi_\mathcal{T})}{d_{L,\sigma}}\mathbf{n}_{L,\sigma},$$

where $R_{L,\sigma}(\varphi_\mathcal{T}) = \varphi(x_L) - \varphi(x_{K_G}) - (\nabla_\mathcal{D}\varphi_\mathcal{T})_{K_G}^G\cdot(x_L - x_{K_G})$. Thanks to Proposition 3, we can deduce that

$$\begin{aligned}
\nabla\varphi(x_L) - (\nabla_\mathcal{D}\varphi_\mathcal{T})_L^{G,\sigma} &= \nabla\varphi(x_{K_G}) + \mu_\sigma\mathbf{n}_{L,\sigma} - \tau_\sigma\mathbf{t}_\sigma - (\nabla_\mathcal{D}\varphi_\mathcal{T})_{K_G}^G + \frac{R_{L,\sigma}(\varphi_\mathcal{T})}{d_{L,\sigma}}\mathbf{n}_{L,\sigma}\\
&= \nabla\varphi(x_{K_G}) - (\nabla_\mathcal{D}\varphi_\mathcal{T})_{K_G}^G + \frac{R_{L,\sigma}(\varphi_\mathcal{T}) - W_{K_G}(x_L)}{d_{L,\sigma}}\mathbf{n}_{L,\sigma}\\
&\quad + \left(\tau_\sigma\frac{\mathbf{t}_\sigma\cdot(y_\sigma - x_L)}{d_{L,\sigma}} + \frac{W_{K_G}(y_\sigma) - W_L(y_\sigma)}{d_{L,\sigma}}\right)\mathbf{n}_{L,\sigma} - \tau_\sigma\mathbf{t}_\sigma\\
&= \nabla\varphi(x_{K_G}) - (\nabla_\mathcal{D}\varphi_\mathcal{T})_{K_G}^G + (\nabla\varphi(x_{K_G}) - (\nabla_\mathcal{D}\varphi_\mathcal{T})_{K_G}^G)\cdot\frac{(x_L - x_{K_G})}{d_{L,\sigma}}\mathbf{n}_{L,\sigma}\\
&\quad + \left(\tau_\sigma\frac{\mathbf{t}_\sigma\cdot(y_\sigma - x_L)}{d_{L,\sigma}} + \frac{W_{K_G}(y_\sigma) - W_L(y_\sigma)}{d_{L,\sigma}}\right)\mathbf{n}_{L,\sigma} - \tau_\sigma\mathbf{t}_\sigma.
\end{aligned}$$

Using then (49), (46) and (2), we can deduce that there exists a real $C_{19} > 0$ which only depends on ϱ_1, ϱ_2, Λ and φ s.t.

$$|\nabla\varphi(x_L) - (\nabla_\mathcal{D}\varphi_\mathcal{T})_L^{G,\sigma}| \le C_{19}(1 + |\mathcal{A}_G^{-1}|)\max_{K\in\mathcal{T}_G}\operatorname{diam}(K),$$

and the proof is complete.

References

[1] I. Aavatsmark, *An introduction to multipoint flux approximations for quadrilateral grids*, Comput. Geosci. **6** (2002), 405–432.

[2] I. Aavatsmark, T. Barkve, Ø. Bøe, and T. Mannseth, *Discretization on non-orthogonal, curvilinear grids for multi-phase flow*, Prov. of the 4th European Conf. on the Mathematics of Oil Recovery (Røros, Norway), vol. D, 1994.

[3] I. Aavatsmark, G.T. Eigestad, R.A. Klausen, M.F. Wheeler, and I. Yotof, *Convergence of a symmetric MPFA method on quadrilateral grids*, 2007, Submitted.

[4] I. Aavatsmark, G.T. Eigestad, B.T. Mallison, and J.M. Nordbotten, *A compact multipoint flux approximation method with improved robustness*, Numer. Methods Partial Differential Equations **24** (2008), no. 5, 1329–1360.

[5] I. Aavatsmark, G.T. Eigestad, B.T. Mallison, J.M. Nordbotten, and E. Øian, *A new finite volume approach to efficient discretization on challeging grids*, Proc. SPE 106435, Houston, 2005.

[6] L. Agélas and D.A. Di Pietro, *A symmetric finite volume scheme for anisotropic heterogeneous second-order elliptic problems*, Finite Volumes for Complex Applications V (R. Eymard and J.-M. Hérard, eds.), John Wiley & Sons, 2008, pp. 705–716.

[7] L. Agélas, D.A. Di Pietro, R. Eymard, and R. Masson, *An abstract analysis framework for nonconforming approximations of anisotropic heterogeneous diffusion*, Preprint available at http://hal.archives-ouvertes.fr/hal-00318390/fr, September 2008, Submitted.

[8] L. Agélas, D.A. Di Pietro, and R. Masson, *A symmetric and coercive finite volume scheme for multiphase porous media flow with applications in the oil industry*, Finite Volumes for Complex Applications V (R. Eymard and J.-M. Hérard, eds.), John Wiley & Sons, 2008, pp. 35–52.

[9] L. Agelas and R. Masson, *Convergence of finite volume MPFA O type schemes for heterogeneous anisotropic diffusion problems on general meshes*, Preprint available at http://hal.archives-ouvertes.fr/hal-00340159/fr, May 2008, Submitted.

[10] ———, *Convergence of the finite volume MPFA O scheme for heterogeneous anisotropic diffusion problems on general meshes*, C. R. Acad. Sci. Paris, Sér. I (2008), no. 346, 1007–1012.

[11] Satish Balay, Kris Buschelman, Victor Eijkhout, William D. Gropp, Dinesh Kaushik, Matthew G. Knepley, Lois Curfman McInnes, Barry F. Smith, and Hong Zhang, *PETSc users manual*, Tech. Report ANL-95/11 - Revision 2.1.5, Argonne National Laboratory, 2004.

[12] Satish Balay, Kris Buschelman, William D. Gropp, Dinesh Kaushik, Matthew G. Knepley, Lois Curfman McInnes, Barry F. Smith, and Hong Zhang, *PETSc Web page*, 2001, www.mcs.anl.gov/petsc.

[13] Satish Balay, William D. Gropp, Lois Curfman McInnes, and Barry F. Smith, *Efficient management of parallelism in object oriented numerical software libraries*, Modern Software Tools in Scientific Computing (E. Arge, A. M. Bruaset, and H. P. Langtangen, eds.), Birkhäuser Press, 1997, pp. 163–202.

[14] F. Brezzi, K. Lipnikov, and M. Shashkov, *Convergence of mimetic finite difference methods for diffusion problems on polyhedral meshes*, SIAM J. Numer. Anal. **45** (2005), 1872–1896.

[15] ———, *Convergence of mimetic finite difference methods for diffusion problems on polyhedral meshes with curved faces*, Math. Mod. Meths. Appli. Sci. (M3AS) **26** (2006), 275–298.

[16] F. Brezzi, K. Lipnikov, and V. Simoncini, *A family of mimetic finite difference methods on polygonal and polyhedral meshes*, Math. Mod. Meths. Appli. Sci. (M3AS) **15** (2005), 1533–1553.

[17] D.A. Di Pietro and A. Ern, *Discrete functional analysis tools for discontinuous Galerkin methods with application to the incompressible Navier-Stokes equations*, Preprint available at http://hal.archives-ouvertes.fr/hal-00278925/fr/, May 2008, Submitted.

[18] J. Droniou, *A density result in Sobolev spaces*, J. Math. Pures Appl. **81** (2002), no. 7, 697–714.

[19] J. Droniou and R. Eymard, *A mixed finite volume scheme for anisotropic diffusion problems on any grid*, Numer. Math. **105** (2006), no. 1, 35–71.

[20] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin, *A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods*, November 2008, Submitted.

[21] M.G. Edwards and C.F. Rogers, *A flux continuous scheme for the full tensor pressure equation*, Prov. of the 4th European Conf. on the Mathematics of Oil Recovery (Røros, Norway), vol. D, 1994.

[22] R. Eymard, T. Gallouët, and R. Herbin, *The finite volume method*, Ph. Charlet and J.L. Lions eds, North Holland, 2000.

[23] ———, *A new finite volume scheme for anisotropic diffusion problems on general grids: convergence analysis*, C. R. Math. Acad. Sci. **344** (2007), no. 6, 3–10.

[24] ———, *Discretization of heterogeneous and anisotropic diffusion problems on general non-conforming meshes. sushi: a scheme using stabilization and hybrid interfaces*, Preprint available at http://hal.archives-ouvertes.fr/, Janvier 2008, Submitted.

[25] R. Eymard, R. Herbin, and J.C. Latché, *Convergence analysis of a colocated finite volume scheme for the incompressible Navier-Stokes equations on general 2D or 3D meshes*, SIAM J. Numer. Anal. **45** (2007), no. 1, 1–36.

[26] M. Vohralík, *Equivalence between lowest-order mixed finite element and multi-point finite volume methods on simplicial meshes*, M2AN Math. Model. Numer. Anal. **40** (2006), no. 2, 367–391, DOI: 10.1051/m2an:2006013.