An abstract analysis framework for nonconforming approximations of diffusion problems on general meshes

Léo Agélas*[1], Daniele A. Di Pietro†[1], Robert Eymard‡[2], and Roland Masson§[1]

[1]IFP, 1 & 4 av. du Bois-Préau 92852 Rueil-Malmaison Cedex (France)
[2]Université Paris-Est Marne-la-Vallée 5 bd. Descartes F-77454 Champs-sur-Marne Marne-la-Vallée Cedex 2 (France)

February 11, 2010

Abstract

In this work we propose a unified analysis framework encompassing a wide range of nonconforming discretizations of anisotropic heterogeneous diffusion operators on general meshes. The analysis relies on two discrete function analytic tools for piecewise polynomial spaces, namely a discrete Sobolev-Poincaré inequality and a discrete Rellich theorem. The convergence requirements are grouped into seven hypotheses, each of them characterizing one salient ingredient of the analysis. Finite volume schemes as well as the most common discontinuous Galerkin methods are shown to fit in the analysis. A new finite volume cell-centered method is also introduced.

1 Introduction

Several methods have been developed through the years to solve the single phase Darcy equation, often of non-conforming type. A crucial ingredient is a robust discretization of heterogeneous anisotropic diffusion operators. Indeed, strong anisotropy and heterogeneity are usually present in problems of practical interest, thus demanding an approach robust with respect to both. Moreover, even for simple domains, the low regularity of the diffusion coefficient may affect the regularity of the solution itself. It is thus important for a discretization method to ensure convergence to minimal regularity solutions, i.e. solutions belonging to the natural function spaces in which the weak formulation of the

*leo.agelas@ifp.fr
†daniele-antonio.di-pietro@ifp.fr
‡robert.eymard@univ-mlv.fr
§roland.masson@ifp.fr

PDE is set. Furthermore, it is often desirable to handle general nonconforming meshes, both because end-users may have little or no control over the mesh and because local grid refinement could be required.

In this work we propose a unified analysis framework encompassing a wide range of non-conforming methods which respond to the above requirements. In particular, both Finite Volume (FV) and discontinuous Galerkin (dG) methods will be shown to fit in the framework. The analogies between these two families of discretization methods have often been highlighted, and the present analysis aims at providing a consistent framework for both.

Finite Volume methods have been widely employed in industrial applications because of simplicity of implementation, closeness to physical intuition and reduced computational cost. In recent years, these methods have known an impetuous development thanks to both empirical and theoretical works. In particular, the convergence analysis of FV methods has been dealt with by Eymard, Gallouët, Herbin and co-authors (see e.g. [22,24]), who have derived new discrete functional analysis tools allowing to prove the convergence to minimum regularity solutions. The discrete analysis framework above has been used for a variety of FV methods applied to linear or non-linear problems (see e.g. [4,25]). Within the framework of Mimetic Finite Difference approximations, reduced-cost methods on general meshes have also been developed. These methods rely on different discrete analysis tools than the ones used here, and we refer to Brezzi, Lipnikov, Shashkov and Simoncini [10–12] for a unified analysis.

Discontinuous Galerkin methods were introduced over thirty years ago to approximate hyperbolic and elliptic PDEs (see e.g. [6, 20] for a historical perspective), and they have received extensive attention over the last decade. Up to now, convergence analysis has relied on classical Finite Element tools, yielding asymptotical order estimates but requiring regularity assumptions on the exact solution (see, e.g., Arnold, Brezzi, Cockburn and Marini [6]). In recent works, Buffa and Ortner [13] and Di Pietro and Ern [17] have independently extended the discrete analysis tools presented by Eymard, Gallouët and Herbin [24] to piecewise polynomial function spaces on general meshes. By means of such tools, the convergence analysis of dG discretization of both linear and non-linear problems can be performed in the spirit of [24].

In this work we further extend the above results by proposing an abstract set of properties ensuring the convergence of a discretization method to minimal regularity solutions. The analysis framework proposed relies on the discrete functional analysis results of [17,24], where the authors introduce discrete $W^{1,p}$ norms which satisfy discrete Sobolev inequalities and deduce a compactness result for bounded sequences in such norms using the Kolmogorov criterion (see, e.g., [9, Theorem IV.25]). In order to use the compactness results for sequences in piecewise polynomial spaces, we shall assume that, whatever the vector space V_h in which the solution is sought, a reconstruction operator on a suitable piecewise polynomial space is available. The key ideas of the analysis can be summarized as follows:

(i) V_h, is equipped with a norm $\|\cdot\|_{V_h}$ which, for all $v_h \in V_h$, controls the discrete H^1 norm of the piecewise polynomial reconstruction of v_h. As a consequence, bounded sequences in the $\|\cdot\|_{V_h}$ norm yield bounded sequences in the discrete H^1 norm;

(ii) an *a priori* estimate on the discrete solution is derived allowing to infer the strong convergence of a subsequence of (reconstruction of) discrete solutions to a function $u \in L^2(\Omega)$;

(iii) the construction of a discrete gradient weakly converging to ∇u in $[L^2(\Omega)]^d$ allows to prove that the limit u actually belongs to $H_0^1(\Omega)$;

(iv) the convergence of the scheme is finally proved testing against a discrete projection of a smooth function belonging to some convenient dense subspace, say $C_c^\infty(\Omega)$.

Since the exact solution is unique, the convergence of the whole sequence of discrete approximations is deduced. Moreover, stronger convergence results on the discrete gradient can be derived using the dissipative structure of the problem for both symmetric and non-symmetric schemes.

Besides providing a means to analyze existing methods and to develop new ones, the above framework ensures the convergence of arbitrary compositions of compliant methods. This can be particularly useful when one wishes to use a more accurate but expensive methods on a selected region of the domain along with a less accurate but faster method elsewhere.

The paper is organized as follows. §2 introduces the abstract framework, including the assumptions on the mesh family as well as the properties required to prove convergence of a method. The latter are grouped into seven Hypotheses, each of them characterizing one salient ingredient of the analysis. The main result is Theorem 2.2. §3 show some examples of methods which fit in the abstract analysis framework. In particular §3.1 presents a selection of dG methods robust with respect to the heterogeneity and anisotropy of the diffusion tensor; §3.2 deals with a new cell-centered finite volume method; §3.3 investigates a hybrid FV method using both cell- and face-unknowns. For all the methods, a precise definition possibly including further assumptions on the mesh family is followed by the verification of Hypotheses 2.1-2.7.

2 Abstract analysis framework

2.1 Model problem and setting

Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be a bounded polygonal domain and consider the following model problem:

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

where $\nu \in [L^\infty(\Omega)]^{d\times d}$ is s.t. (such that), for a.e. (almost every) $x \in \Omega$, $\nu(x)$ is symmetric and its spectrum $\{\lambda_i(x)\}_{i=1}^d$ is s.t. $0 < \underline{\lambda} \leq \lambda_i(x) \leq \overline{\lambda} < \infty$ and $f \in L^2(\Omega)$. In weak formulation, problem (1) reads: Find $u \in H_0^1(\Omega)$ s.t.

$$a(u, v) = (f, v)_{L^2(\Omega)} \qquad \forall v \in H_0^1(\Omega), \tag{2}$$

where $\mathcal{L}(H_0^1(\Omega) \times H_0^1(\Omega); \mathbb{R}) \ni a(u, v) \overset{\text{def}}{=} (\nu\nabla u, \nabla v)_{[L^2(\Omega)]^d}$. The well-posedness of problem (2) is classical.

Remark 2.1. The analysis can be easily extended to $f \in L^r(\Omega)$ with $r \geq \frac{2d}{d+2}$ if $d \geq 3$ and $r > 1$ if $d = 2$; see [17] for the details in the case of dG methods. This requires more general Sobolev inequalities than the one of Hypothesis 2.1, which are proved in [17, 24]. Also, different boundary conditions can be handled with minor modifications, but we have decided to stick to the homogeneous Dirichlet problem for clarity of presentation.

The following definition characterizes an admissible mesh family:

Definition 1 (Admissible mesh family)**.** *Let $\mathcal{H}$ be a countable set. The mesh family $\{\mathcal{T}_h\}_{h\in\mathcal{H}}$, is said to be admissible if the following assumptions are satisfied for all $h \in \mathcal{H}$:*

(i) $\mathcal{T}_h$ is a finite family of non-empty connected (possibly non-convex) open disjoint sets T forming a partition of Ω and whose boundaries are a finite union of parts of hyperplanes. The d-dimensional Lebesgue measure and the diameter of the generic element $T \in \mathcal{T}_h$ will be denoted by $|T|$ and h_T respectively. The representative linear dimension of the discretization will be defined as $h \stackrel{\text{def}}{=} \max_{T\in\mathcal{T}_h} h_T$;

(ii) each $T \in \mathcal{T}_h$ is affine-equivalent to an element of a finite collection of reference elements;

(iii) there is a parameter N_∂ independent of h s.t., for all $h \in \mathcal{H}$, each $T \in \mathcal{T}_h$ has at most N_∂ faces. For all elements $T \in \mathcal{T}_h$, let $\mathcal{F}_h^T$ denote the set of faces of T. A set $F \in \mathcal{F}_h^T$ of non-zero $(d-1)$-dimensional Lebesgue measure $|F|$ is said to be a face of T if F is part of a hyperplane and if either F is located on the boundary of Ω (boundary face) or there is one and only one $T' \in \mathcal{T}_h$ s.t. $F = \mathcal{F}_h^T \cap \mathcal{F}_h^{T'}$ (interface). The diameter of the generic face $F \in \mathcal{F}_h$ will be denoted by h_F;

(iv) there is a parameter ϱ_1 independent of h s.t., for all $T \in \mathcal{T}_h$,

$$\sum_{F\in\mathcal{F}_h^T} h_F|F| \le \varrho_1|T|;$$

The set of boundary faces will be denoted by $\mathcal{F}_h^b$, whereas the interfaces will be collected into the set $\mathcal{F}_h^i$. For every $F = \mathcal{F}_h^{T_1} \cap \mathcal{F}_h^{T_2}$ (the numbering of the elements sharing F is arbitrary but fixed) we let μ_F denote the outward normal to T_1; for all $T \in \mathcal{T}_h$ and for all $F \in \mathcal{F}_h^T$, μ_F^T will denote the outward normal to T. For every $F \in \mathcal{F}_h^T \cap \mathcal{F}_h^b$, both μ_F and μ_F^T will denote the outward normal to Ω. Further assumptions on the mesh family may be required depending on the method considered, and will be specified in the corresponding section.

Remark 2.2. According to Definition 1, (i) the mesh elements are not assumed to be convex, and the mesh may possibly be nonconforming; (ii) in three space dimensions, general hexahedra can be treated by decomposing non-plane faces in a fixed number of plane sub-faces.

Let $\mathcal{T}_h$ denote an element of an admissible mesh family and let $\mathcal{S}_h$ denote a sub-mesh of $\mathcal{T}_h$ (i.e., a mesh obtained by further decomposing the elements of $\mathcal{T}_h$ into polyhedral subelements) depending on the method at hand. We introduce the space of piecewise polynomial functions of total degree less than or equal to $k \ge 0$,

$$P_h^k(\mathcal{X}_h) \stackrel{\text{def}}{=} \{v_h \in L^2(\Omega);\ v_{h|T} \in \mathbb{P}^k(T),\ \forall T \in \mathcal{X}_h\}, \qquad \mathcal{X}_h \in \{\mathcal{T}_h, \mathcal{S}_h\}.$$

The symbols V_h and Σ_h denote two vector spaces associated with $\mathcal{T}_h$ and $\mathcal{S}_h$ respectively. We assume that $\Sigma_h = [P_h^{k_\Sigma}(\mathcal{S}_h)]^d$ for a fixed $k_\Sigma \ge 0$ depending on the method considered. Also, in what follows, $r_h^V : V_h \to P_h^{k_V}(\mathcal{T}_h)$ will denote a reconstruction operator onto the piecewise polynomial space of degree k_V depending on the method at hand (see Hypothesis 2.1). In particular, for FV methods, $k_V = k_\Sigma = 0$ whereas $k_V \ge k_\Sigma \ge 0$, $k_V \ge 1$ for dG methods.

The symbols $\lesssim$ and $\gtrsim$ will be used in the present section for inequalities that hold up to a positive parameter independent of the mesh size h but possibly depending on the regularity parameters of the mesh family, on ν, k_V and k_Σ. More detailed expressions for these multiplicative constant will be given for each method in §3.

Hypothesis 2.1 (Piecewise polynomial reconstruction r_h^V). *For a fixed $k_V \geq 0$ depending on the actual discretization method, there is a reconstruction operator $r_h^V : V_h \to P_h^{k_V}(\mathcal{T}_h)$ which maps every element $v_h \in V_h$ onto a piecewise polynomial function $r_h^V v_h \in P_h^{k_V}(\mathcal{T}_h)$.*

We define the following bilinear form

$$\mathcal{L}(V_h \times V_h; \mathbb{R}) \ni a_h(u_h, v_h) \stackrel{\text{def}}{=} (\nu G(u_h), \widetilde{G}(v_h))_{[L^2(\Omega)]^d} + j_h(u_h, v_h), \qquad (3)$$

where $G \in \mathcal{L}(V_h; \Sigma_h)$ and $\widetilde{G} \in \mathcal{L}(V_h; \Sigma_h)$ are linear gradient reconstructions whose properties will be detailed in Hypotheses 2.3, 2.4 and 2.7, We particularly stress the importance of Hypotheses 2.4 and 2.7 below, which provide the basic design criteria for the discrete gradients. whereas $j_h \in \mathcal{L}(V_h \times V_h; \mathbb{R})$ is a stabilizing bilinear form meant to ensure the coercivity of a_h. We focus on the following family of approximations for problem (2): Find $u_h \in V_h$ s.t.

$$a_h(u_h, v_h) = (f, r_h^V v_h)_{L^2(\Omega)} \qquad \forall v_h \in V_h. \qquad (4)$$

2.2 Discrete Rellich theorem

The piecewise polynomial space $P_h^{k_V}(\mathcal{T}_h)$, $k_V \geq 0$, must be equipped with a discrete H^1 norm $\|\cdot\|_{1,2,h}$ s.t. the following hypothesis is satisfied:

Hypothesis 2.2 (Compactness). *Let $\{p_h\}_{h\in\mathcal{H}}$ be a sequence in $P_h^{k_V}(\mathcal{T}_h)$, $k_V \geq 0$, bounded in the corresponding $\|\cdot\|_{1,2,h}$ norm. Then, the family $\{p_h\}_{h\in\mathcal{H}}$ is relatively compact in $L^2(\Omega)$ (and also in $L^2(\mathbb{R}^d)$ taking $p_h = 0$ outside Ω).*

Norms satisfying Hypothesis 2.2 will be defined in eqs. (20) and (27) below.

Lemma 2.1 (Discrete Sobolev-Poincaré inequality). *Let $\{\mathcal{T}_h\}_{h\in\mathcal{H}}$ be a mesh family compliant with Definition 1 and let us suppose that Hypothesis 2.2 holds. Then, for all $p_h \in P_h^{k_V}$, $k_V \geq 0$,*

$$\|p_h\|_{L^2(\Omega)} \lesssim \|p_h\|_{1,2,h}. \qquad (5)$$

Proof. For the sake of simplicity, let $\mathcal{H} = \mathbb{N}$ and $h_n \to 0$ as $n \to \infty$. We proceed by contradiction. Let us admit that, for all $C > 0$, there is $n \in \mathbb{N}$ and $p_{h_n} \in \mathcal{T}_{h_n}$ s.t. $\|p_{h_n}\|_{L^2(\Omega)} > C\|p_{h_n}\|_{1,2,h_n}$. In particular, we can take $C = n$ and set $\widetilde{p}_{h_n} \stackrel{\text{def}}{=} p_{h_n}/\|p_{h_n}\|_{1,2,h}$, so that

$$\|\widetilde{p}_{h_n}\|_{L^2(\Omega)} > n, \qquad \|\widetilde{p}_{h_n}\|_{1,2,h_n} = 1. \qquad (6)$$

As n increases, the L^2 norm of $\widetilde{p}_{h_n}$ increases, whereas its $\|\cdot\|_{1,2,h}$ norm remains bounded. According to Hypothesis 2.2, $\{\widetilde{p}_{h_n}\}_{n\in\mathbb{N}}$ is thus relatively compact in $L^2(\Omega)$, and we can extract a subsequence $\{\widetilde{p}_{h_{\varphi(n)}}\}_{n\in\mathbb{N}}$ which converges to some $\overline{p}$ in $L^2(\Omega)$. As a consequence, $\|\widetilde{p}_{\varphi(n)}\|_{L^2(\Omega)} \to \|\overline{p}\|_{L^2(\Omega)}$ as $n \to \infty$, which is in contradiction with (6). ☐

A direct proof of the Poincaré-Friedrichs inequality on broken Sobolev spaces has been given in [5, 8, 24]; broken Sobolev embeddings have been derived by Lasis and Süli [26, 27] in the Hilbertian case; broken Sobolev embeddings in the non-Hilbertian case have been recently presented in [17]

Hypothesis 2.3 ($\|\cdot\|_{V_h}$ norm)**.** *The vector space V_h is equipped with an inner product norm $\|\cdot\|_{V_h}$ s.t., for all $v_h \in V_h$,*

$$\|r_h^V v_h\|_{1,2,h} \lesssim \|v_h\|_{V_h}, \tag{7}$$

$$\|G(v_h)\|_{[L^2(\Omega)]^d} + \|\widetilde{G}(v_h)\|_{[L^2(\Omega)]^d} \lesssim \|v_h\|_{V_h}. \tag{8}$$

Inequality (7) will be used to derive an estimate for the piecewise polynomial reconstruction of the solution in terms of the discrete H^1 norm $\|\cdot\|_{1,2,h}$. This will, in turn, ensure the boundedness of the sequence of the reconstructed discrete solutions of (4) on the mesh family $\{\mathcal{T}_h\}_{h\in\mathcal{H}}$, a key ingredient to infer a compactness result. Inequality (8) states that bounded sequences in the $\|\cdot\|_{V_h}$ norm yield bounded sequences of gradient approximations in the L^2 norm.

The following assumption contains the major requirement for the gradient $\widetilde{G}$, i.e., weak consistency for suitable sequences in V_h:

Hypothesis 2.4 (Weak convergence of $\widetilde{G}$)**.** *Let $\{v_h\}_{h\in\mathcal{H}}$, be a sequence in V_h s.t. $\{r_h^V v_h\}_{h\in\mathcal{H}}$ converges to $v \in L^2(\Omega)$ in $L^2(\mathbb{R}^d)$ (prolonging $r_h^V v_h$ to zero outside Ω) and $\{\widetilde{G}(v_h)\}_{h\in\mathcal{H}}$ is bounded in the $[L^2(\mathbb{R}^d)]^d$ norm. Then, for all $\Phi \in [C_c^\infty(\mathbb{R}^d)]^d$,*

$$\lim_{h\to 0} \int_{\mathbb{R}^d} \widetilde{G}(v_h)\cdot\Phi = -\int_{\mathbb{R}^d} v\nabla\cdot\Phi.$$

Disposing of a weakly converging gradient allows to prove the following result concerning the regularity of the limit of a converging sequence in V_h:

Theorem 2.1 (Discrete Rellich theorem)**.** *Let $\{v_h\}_{h\in\mathcal{H}}$ be a sequence in V_h bounded in the $\|\cdot\|_{V_h}$ norm. Then, (i) $\{r_h^V v_h\}_{h\in\mathcal{H}}$ is relatively compact in $L^2(\Omega)$; (ii) if $r_h^V v_h \to v$ in $L^2(\Omega)$ as $h \to 0$, then $v \in H_0^1(\Omega)$.*

Proof. Owing to the assumptions of the theorem together with (7), there is $C \in \mathbb{R}_+$ s.t.

$$\|r_h^V v_h\|_{1,2,h} \lesssim \|v_h\|_{V_h} \le C \qquad \forall h \in \mathcal{H}.$$

As a consequence, the sequence $\{r_h^V v_h\}_{h\in\mathcal{H}}$ is bounded in the $\|\cdot\|_{1,2,h}$ norm. Owing to Hypothesis 2.2, it is possible to extract a subsequence converging to some v in $L^2(\Omega)$ and also in $L^2(\mathbb{R}^d)$ provided we prolong $r_h^V v_h$ by zero outside Ω. Moreover, (8) yields, for all $h \in \mathcal{H}$,

$$\|\widetilde{G}(v_h)\|_{[L^2(\Omega)]^d} \lesssim \|v_h\|_{V_h} \le C.$$

We thus conclude that there exists a $\tau \in [L^2(\Omega)]^d$ to which the sequence $\{\widetilde{G}(v_h)\}_{h\in\mathcal{H}}$ converges (up to a subsequence) in $[L^2(\Omega)]^d$ and also in $[L^2(\mathbb{R}^d)]^d$. On the other hand, the sequence $\{r_h^V v_h\}_{h\in\mathcal{H}}$ satisfies the assumptions of Hypothesis 2.4, so that $\tau = \nabla v$, which concludes the proof. □

2.3 *A priori* estimate on the solution

Let $\pi_h^V : C^0(\overline{\Omega}) \to V_h$ denote an interpolator onto V_h whose properties will be detailed in Hypotheses 2.5 and 2.7. In what follows, π_h^V will be applied to functions of $C_c^\infty(\Omega)$, which is used as a pivot space.

Hypothesis 2.5 (Stabilization j_h)**.** *The bilinear form j_h is symmetric, positive semidefinite and continuous with respect to the $\|\cdot\|_{V_h}$ norm, i.e.,*

$$j_h(u_h, v_h) \lesssim \|u_h\|_{V_h} \|v_h\|_{V_h} \qquad \forall (u_h, v_h) \in [V_h]^2. \tag{9}$$

Furthermore, the following consistency property holds:

$$\lim_{h\to 0} j_h(\pi_h^V \varphi, \pi_h^V \varphi) = 0 \qquad \forall \varphi \in C_c^\infty(\Omega). \tag{10}$$

The following Cauchy-Schwarz type inequality is an immediate consequence of Hypothesis 2.5:

$$|j_h(u_h, v_h)| \lesssim [j_h(u_h, u_h)]^{1/2} [j_h(v_h, v_h)]^{1/2}. \tag{11}$$

Hypothesis 2.6 (Coercivity of a_h)**.** *For all $v_h \in V_h$, $a_h(v_h, v_h) \gtrsim \|v_h\|_{V_h}^2$.*

The coercivity of the bilinear form a_h is an essential ingredient of the analysis, since it allows to obtain an estimate of the solution for use in the discrete Rellich Theorem 2.1.

Lemma 2.2 (Well-posedness)**.** *Problem* (4) *is well-posed. Furthermore, its solution satisfies the following* a priori *estimates:*

$$\|r_h^V u_h\|_{1,2,h} \lesssim \|u_h\|_{V_h} \lesssim \|f\|_{L^2(\Omega)}. \tag{12}$$

Proof. (i) To prove the well-posedness we use the Lax-Milgram lemma. Using (8) together with (9) we have, for all $(u_h, v_h) \in [V_h]^2$,

$$a_h(u_h, v_h) \lesssim \overline{\lambda} \|u_h\|_{V_h} \|v_h\|_{V_h} + \|u_h\|_{V_h} \|v_h\|_{V_h} \lesssim \|u_h\|_{V_h} \|v_h\|_{V_h},$$

i.e., the bilinear form a_h is continuous in V_h. The Cauchy-Schwarz inequality together with (5) and (7) yield, for all $v_h \in V_h$,

$$(f, r_h^V v_h)_{L^2(\Omega)} \leq \|f\|_{L^2(\Omega)} \|r_h^V v_h\|_{L^2(\Omega)} \lesssim \|f\|_{L^2(\Omega)} \|r_h^V v_h\|_{1,2,h} \lesssim \|f\|_{L^2(\Omega)} \|v_h\|_{V_h}.$$

We conclude using Hypothesis 2.6. (ii) If u_h is the null element of V_h, the estimate is trivially verified. If this is not the case, Hypothesis 2.6 together with the Cauchy-Schwarz inequality, (5) and (7) yield

$$\|r_h^V u_h\|_{1,2,h} \|u_h\|_{V_h} \lesssim \|u_h\|_{V_h}^2 \lesssim a_h(u_h, u_h) \lesssim \|f\|_{L^2(\Omega)} \|u_h\|_{L^2(\Omega)} \lesssim \|f\|_{L^2(\Omega)} \|u_h\|_{V_h},$$

thus concluding the proof. □

2.4 Convergence

The following assumption contains a major requirement for the gradient G, i.e., consistency for smooth functions:

Hypothesis 2.7 (Consistency)**.** *The following results hold:*

$$\|\pi_h^V\varphi\|_{V_h} \lesssim \sigma_\varphi \qquad \forall\varphi\in C_c^\infty(\Omega), \tag{13}$$

$$\lim_{h\to 0}\|(r_h^V\circ\pi_h^V)\varphi-\varphi\|_{L^2(\Omega)}=0 \qquad \forall\varphi\in C_c^\infty(\Omega), \tag{14}$$

$$\lim_{h\to 0}\|\nabla\varphi-G(\pi_h^V\varphi)\|_{[L^2(\Omega)]^d}=0 \qquad \forall\varphi\in C_c^\infty(\Omega), \tag{15}$$

where $\sigma_\varphi>0$ is a parameter depending only on φ and on the mesh regularity parameters.

The above assumptions ensure that we can consistently approximate smooth functions and their gradients on the discrete spaces at hand. The consistency of G stated in (15) allows to prove the following

Lemma 2.3 (Convergence of G)**.** *Let $\{\mathcal{T}_h\}_{h\in\mathcal{H}}$ be a family of admissible meshes. Let u_h denote the unique solution of the discrete problem (4) on $\mathcal{T}_h$. Then, (i) there exists $\widetilde{u}\in H_0^1(\Omega)$ and a subsequence $\{r_h^V u_h\}_{h\in\mathcal{H}}$ converging to $\widetilde{u}$ in $L^2(\Omega)$ as $h\to 0$; (ii) $\{G(u_h)\}_{h\in\mathcal{H}}$ converges to $\nabla\widetilde{u}$ in $[L^2(\Omega)]^d$.*

Proof. (i) Thanks to (12), the sequence $\{r_h^V u_h\}_{h\in\mathcal{H}}$ is bounded in the $\|\cdot\|_{1,2,h}$ norm. According to Theorem 2.1, there is a subsequence of $\{r_h^V u_h\}_{h\in\mathcal{H}}$ (still denoted with the same symbol) and an element $\widetilde{u}\in H_0^1(\Omega)$ s.t. $\{r_h^V u_h\}_{h\in\mathcal{H}}$ converges to $\widetilde{u}$ in $L^2(\Omega)$ as $h\to 0$. (ii) Let $\varphi\in C_c^\infty$ and set $\varphi_h \stackrel{\text{def}}{=} \pi_h^V\varphi$. We have that

$$\|G(u_h)-\nabla\widetilde{u}\|_{[L^2(\Omega)]^d}^2 \leq \\ 3\left[\|G(u_h)-G(\varphi_h)\|_{[L^2(\Omega)]^d}^2+\|G(\varphi_h)-\nabla\varphi\|_{[L^2(\Omega)]^d}^2+\|\nabla\varphi-\nabla\widetilde{u}\|_{[L^2(\Omega)]^d}^2\right].$$

Let S_i, $i\in\{1\ldots 3\}$ denote the terms in the right hand side. Thanks to Hypothesis 2.6 and to the linearity of a_h we have that

$$S_1 \lesssim a_h(u_h,u_h)-a_h(u_h,\varphi_h)-a_h(\varphi_h,u_h)+a_h(\varphi_h,\varphi_h).$$

Owing to (4), $a_h(u_h,u_h)=(f,r_h^V u_h)_{L^2(\Omega)}$ and $a_h(u_h,\varphi_h)=(f,r_h^V\varphi_h)_{L^2(\Omega)}$. As a consequence,

$$\lim_{h\to 0}a_h(u_h,u_h)=(f,\widetilde{u})_{L^2(\Omega)}.$$

Furthermore, using (14), we conclude that

$$0\leq\limsup_{h\to 0}\left|a_h(u_h,\varphi_h)-(f,\varphi)_{L^2(\Omega)}\right|\leq\limsup_{h\to 0}\|f\|_{L^2(\Omega)}\|r_h^V\varphi_h-\varphi\|_{L^2(\Omega)}=0,$$

that is, gathering the above results,

$$\forall\varphi\in C_c^\infty(\Omega),\qquad \lim_{h\to 0}\left[a_h(u_h,u_h)-a_h(u_h,\varphi_h)\right]=(f,\widetilde{u}-\varphi)_{L^2(\Omega)}. \tag{16}$$

To estimate the remaining terms, observe that

$$a_h(\varphi_h,\varphi_h)-a_h(\varphi_h,u_h)=(\nu\nabla\varphi,\widetilde{G}(\varphi_h-u_h))_{[L^2(\Omega)]^d} \\ +(\nu(G(\varphi_h)-\nabla\varphi),\widetilde{G}(\varphi_h-u_h))_{[L^2(\Omega)]^d}+j_h(\varphi_h,\varphi_h-u_h).$$

Owing to Hypothesis 2.4, the term in the first line tends to $(\nu\nabla\varphi, \nabla(\varphi - \widetilde{u}))_{[L^2(\Omega)]^d}$ as $h \to 0$. Let $S \stackrel{\text{def}}{=} (\nu(G(\varphi_h) - \nabla\varphi), \widetilde{G}(\varphi_h - u_h))_{[L^2(\Omega)]^d}$ denote the first term in the second line. The following estimate holds:

$$\begin{aligned}
|S| &\leq \overline{\lambda}\|G(\varphi_h) - \nabla\varphi\|_{[L^2(\Omega)]^d}\|\widetilde{G}(\varphi_h - u_h)\|_{[L^2(\Omega)]^d} \\
&\lesssim \|G(\varphi_h) - \nabla\varphi\|_{[L^2(\Omega)]^d}\left(\|\varphi_h\|_{V_h} + \|u_h\|_{V_h}\right) \\
&\lesssim \|G(\varphi_h) - \nabla\varphi\|_{[L^2(\Omega)]^d}\left(\sigma_\varphi + \|u_h\|_{V_h}\right),
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality followed by (8), (13) and (12). Since $\|u_h\|_{V_h}$ is bounded, the right hand side of the above inequality tends to zero as $h \to 0$. On the other hand, (11), (13) and (12) yield

$$\begin{aligned}
|j_h(\varphi_h, \varphi_h - u_h)| &\leq [j_h(\varphi_h, \varphi_h)]^{1/2}[j_h(\varphi_h - u_h, \varphi_h - u_h)]^{1/2} \\
&\lesssim [j_h(\varphi_h, \varphi_h)]^{1/2}\left(\|\varphi_h\|_{V_h} + \|u_h\|_{V_h}\right) \\
&\lesssim [j_h(\varphi_h, \varphi_h)]^{1/2}\left(\sigma_\varphi + \|u_h\|_{V_h}\right),
\end{aligned}$$

which, owing to (10) and to the boundedness of $\|u_h\|_{V_h}$, tends to zero as $h \to 0$. In conclusion,

$$\forall\varphi \in C_c^\infty(\Omega), \qquad \lim_{h\to 0}\left[a_h(\varphi_h, \varphi_h) - a_h(\varphi_h, u_h)\right] = (\nu\nabla\varphi, \nabla(\varphi - \widetilde{u}))_{[L^2(\Omega)]^d}. \tag{17}$$

Equations (16) and (17) yield

$$\forall\varphi \in C_c^\infty(\Omega), \qquad \lim_{h\to 0} S_1 = (\nu\nabla\varphi, \nabla(\varphi - \widetilde{u}))_{[L^2(\Omega)]^d} + (f, \widetilde{u} - \varphi)_{L^2(\Omega)}.$$

Using (15) we immediately conclude that, for all $\varphi \in C_c^\infty(\Omega)$, $\lim_{h\to 0} S_2 = 0$. Gathering the above results, for all $\varphi \in C_c^\infty(\Omega)$,

$$\begin{aligned}
&\limsup_{h\to 0}\|G(u_h) - \nabla\widetilde{u}\|^2_{[L^2(\Omega)]^d} \lesssim \\
&\qquad c(\nu\nabla\varphi, \nabla(\varphi - \widetilde{u}))_{[L^2(\Omega)]^d} + (f, \widetilde{u} - \varphi)_{L^2(\Omega)} + \|\nabla\varphi - \nabla\widetilde{u}\|^2_{[L^2(\Omega)]^d}.
\end{aligned}$$

Let now $\{\varphi_m\}_{m\in\mathbb{N}}$ be a sequence converging to $\widetilde{u}$ in $H_0^1(\Omega)$ (the existence of such a sequence follows from the density of $C_c^\infty(\Omega)$ in $H_0^1(\Omega)$). Using the above bound, we conclude that

$$0 \leq \liminf_{h\to 0}\|G(u_h) - \nabla\widetilde{u}\|^2_{[L^2(\Omega)]^d} \leq \limsup_{h\to 0}\|G(u_h) - \nabla\widetilde{u}\|^2_{[L^2(\Omega)]^d} \leq 0,$$

which proves the assertion. □

Remark 2.3. Observe that the passages to the limit for $h \to 0$ and for $m \to \infty$ cannot be exchanged in the proof. Indeed, the estimates from which (14) and (15) are obtained may depend on some norm of φ which does not remain bounded as $m \to \infty$, e.g. the H^2 norm.

Theorem 2.2 (Convergence of the method)**.** *Let $\{\mathcal{T}_h\}_{h\in\mathcal{H}}$ be a family of admissible meshes. Let u_h denote the unique solution of the discrete problem* (4) *on $\mathcal{T}_h$. Then, (i) the sequence $\{r_h^V u_h\}_{h\in\mathcal{H}}$ converges to the solution of* (2)*, say u, in $L^2(\Omega)$ as $h \to 0$; (ii) the sequence $\{G(u_h)\}_{h\in\mathcal{H}}$ converges to ∇u in $[L^2(\Omega)]^d$.*

Proof. Thanks to (12), the sequence $\{u_h\}_{h\in\mathcal{H}}$ is bounded in the $\|\cdot\|_{V_h}$ norm. Theorem 2.1 states that we can extract a subsequence still denoted by $\{r_h^V u_h\}_{h\in\mathcal{H}}$ which converges to an element $\widetilde{u}\in H_0^1(\Omega)$ in $L^2(\Omega)$. Let us focus on the above sub-sequence. According to Lemma 2.3, $\{G(u_h)\}_{h\in\mathcal{H}}$ converges to $\nabla\widetilde{u}$ in $[L^2(\Omega)]^d$. In order to prove the convergence of the method, we have to prove that $\widetilde{u}$ solves (2). Let, now, $\varphi\in C_c^\infty(\Omega)$ and set $\varphi_h \overset{\text{def}}{=} \pi_h^V\varphi$. We have that

$$a_h(u_h,\varphi_h) = (\nu G(u_h), \widetilde{G}(\varphi_h))_{[L^2(\Omega)]^d} + j_h(u_h,\varphi_h).$$

Using Hypothesis 2.4 together with Lemma 2.3 we conclude that

$$\forall\varphi\in C_c^\infty,\qquad \lim_{h\to 0}(\nu G(u_h), \widetilde{G}(\varphi_h))_{[L^2(\Omega)]^d} = (\nu\nabla\widetilde{u},\nabla\varphi)_{[L^2(\Omega)]^d} = a(\widetilde{u},\varphi).$$

On the other hand, (11) together with (12) yield

$$\begin{aligned}|j_h(u_h,\varphi_h)| &\le j_h(\varphi_h,\varphi_h)^{1/2} j_h(u_h,u_h)^{1/2}\\ &\le j_h(\varphi_h,\varphi_h)^{1/2}\|u_h\|_{V_h} \lesssim j_h(\varphi_h,\varphi_h)^{1/2}\|f\|_{L^2(\Omega)},\end{aligned}$$

which tends to 0 as $h\to 0$ by virtue of (10). Moreover,

$$(f, r_h^V\varphi_h)_{L^2(\Omega)} = (f,\varphi)_{L^2(\Omega)} + (f,\varphi - r_h^V\varphi_h)_{L^2(\Omega)},$$

and, using (14),

$$0\le \limsup_{h\to 0}\left|(f,\varphi - r_h^V\varphi_h)_{L^2(\Omega)}\right| \lesssim \limsup_{h\to 0}\|f\|_{L^2(\Omega)}\|\varphi - r_h^V\varphi_h\|_{L^2(\Omega)} = 0,$$

so that, for all $\varphi\in C_c^\infty(\Omega)$, $(f, r_h^V\varphi_h)_{L^2(\Omega)}\to(f,\varphi)_{L^2(\Omega)}$ as $h\to 0$. Thanks to the above results, and since the u_h are solutions of the discrete problem (4), we have that

$$a(\widetilde{u},\varphi) = (f,\varphi)_{L^2(\Omega)}\qquad \forall\varphi\in C_c^\infty(\Omega).$$

Since $C_c^\infty(\Omega)$ is dense in $H_0^1(\Omega)$, $\widetilde{u}=u$ for a.e. $x\in\Omega$. Furthermore, problem (2) has a unique solution, and so the convergence property extends to the whole sequence. The convergence of $\{G(u_h)\}_{h\in\mathcal{H}}$ to ∇u is an immediate consequence of Lemma 2.3 together with the uniqueness of the limit. □

2.5 Symmetric methods

In this section we show how the analysis can be simplified for symmetric methods. The following theorem replaces Lemma 2.3 and Theorem 2.2:

Theorem 2.3 (Convergence of symmetric methods)**.** *Suppose that the bilinear form a_h is symmetric, i.e. $\widetilde{G}=G$ and let $\{\mathcal{T}_h\}_{h\in\mathcal{H}}$ be a family of admissible meshes. Let u_h denote the unique solution of the discrete problem* (4) *on $\mathcal{T}_h$. Then, (i) the sequence $\{r_h^V u_h\}_{h\in\mathcal{H}}$ converges to the solution of* (2)*, say u, in $L^2(\Omega)$ as $h\to 0$; (ii) the sequence $\{G(u_h)\}_{h\in\mathcal{H}}$ converges to ∇u in $[L^2(\Omega)]^d$.*

Proof. Thanks to (12), the sequence $\{u_h\}_{h\in\mathcal{H}}$ is bounded in the $\|\cdot\|_{V_h}$ norm. Theorem 2.1 states that we can extract a subsequence still denoted by $\{r_h^V u_h\}_{h\in\mathcal{H}}$ which converges to an element $\widetilde{u}\in H_0^1(\Omega)$ in $L^2(\Omega)$. Let us focus on the above sub-sequence. Owing to Hypothesis 2.4, $\widetilde{G}(u_h)$ weakly converges to $\nabla\widetilde{u}$ in L^2. Let $\varphi \in C_c^\infty(\Omega)$ and set $\varphi_h \stackrel{\text{def}}{=} \pi_h^V \varphi$. Observe that

$$a_h(u_h,\varphi_h) = (\nu G(u_h), \nabla\varphi)_{[L^2(\Omega)]^d} + \left[(\nu G(u_h), G(u_h) - \nabla\varphi)_{[L^2(\Omega)]^d} + j_h(u_h,\varphi_h)\right],$$

and let S_1 and S_2 the addends in the right hand side. Owing to the weak convergence of $G(u_h)$, $S_1 \to a(\widetilde{u},\varphi)$ as $h \to 0$. Using the Cauchy-Schwarz inequality together with (11) we obtain

$$|S_2| \le \overline{\lambda}\|G(u_h)\|_{[L^2(\Omega)]^d}\|G(\varphi_h) - \nabla\varphi\|_{[L^2(\Omega)]^d} + [j_h(u_h,u_h)]^{1/2}[j_h(\varphi_h,\varphi_h)]^{1/2}.$$

Thanks to (8), (9) and (12), both $\|G(u_h)\|_{[L^2(\Omega)]^d}$ and $[j_h(u_h,u_h)]^{1/2}$ are bounded by $\|f\|_{L^2(\Omega)}$ up o a positive multiplicative constant. Equation (15) together with (10) then yield $|S_2| \to 0$ as $h \to 0$. In conclusion,

$$(f,\varphi)_{L^2(\Omega)} \leftarrow (f,\varphi_h)_{L^2(\Omega)} = a_h(u_h,\varphi_h) \to a(\widetilde{u},\varphi),$$

i.e., $\widetilde{u} = u$ for a.e. $x \in \Omega$ since $C_c^\infty(\Omega)$ is dense in $H_0^1(\Omega)$. The strong convergence of $\{G(u_h)\}_{h\in\mathcal{H}}$ follows. □

2.6 Adjoint methods

Let

$$a_h^*(u_h,v_h) \stackrel{\text{def}}{=} (\nu\widetilde{G}(u_h), G(v_h))_{[L^2(\Omega)]^d} + j_h(u_h,v_h).$$

In this section we investigate the convergence of the adjoint problem: Find $u_h \in V_h$ s.t.

$$a_h^*(u_h^*,v_h) = (f, r_h^V v_h)_{L^2(\Omega)} \qquad \forall v_h \in V_h. \tag{18}$$

Theorem 2.4 (Convergence of adjoint methods)**.** *Let $\{\mathcal{T}_h\}_{h\in\mathcal{H}}$ be a family of admissible meshes. Let u_h^* denote the unique solution of the discrete problem* (18) *on $\mathcal{T}_h$. Then, the sequence $\{r_h^V u_h^*\}_{h\in\mathcal{H}}$ converges to the solution of* (2)*, say u, in $L^2(\Omega)$ as $h \to 0$.*

Proof. Since also a_h^* is coercive, the sequence $\{u_h\}_{h\in\mathcal{H}}$ is bounded in the $\|\cdot\|_{V_h}$ norm. Theorem 2.1 states that we can extract a subsequence still denoted by $\{r_h^V u_h\}_{h\in\mathcal{H}}$ which converges to an element $\widetilde{u} \in H_0^1(\Omega)$ in $L^2(\Omega)$. We shall focus our attention on the above sub-sequence. Let $\varphi \in C_c^\infty(\Omega)$ and set $\varphi \stackrel{\text{def}}{=} \pi_h^V\varphi$. We have

$$\begin{aligned} a_h^*(u_h^*,\varphi_h) &= (\nu\widetilde{G}(u_h^*), \nabla\varphi)_{[L^2(\Omega)]^d} + (\nu\widetilde{G}(u_h^*), G(\varphi_h) - \nabla\varphi)_{[L^2(\Omega)]^d} + j_h(u_h^*,\varphi_h) \\ &\stackrel{\text{def}}{=} S_1 + S_2 + S_3. \end{aligned}$$

Using Hypothesis 2.4 it is clear that $S_1 \to a(\widetilde{u},\varphi)$ as $h \to 0$. For the second term, using (12) we have

$$|S_2| \le \overline{\lambda}\|\widetilde{G}(u_h^*)\|_{[L^2(\Omega)]^d}\|G(\varphi_h) - \nabla\varphi\|_{[L^2(\Omega)]^d},$$

which, owing to (15), tends to zero as $h \to 0$. Similarly, using (10) together with (12), we can prove that $|S_3| \to 0$ as $h \to 0$. We thus have

$$(f,\varphi)_{L^2(\Omega)} \leftarrow (f,\varphi_h)_{L^2(\Omega)} = a_h(u_h^*,\varphi_h) \to a(\widetilde{u},\varphi),$$

i.e., $\widetilde{u} = u$ for a.e. $x \in \Omega$ since $C_c^\infty(\Omega)$ is dense in $H_0^1(\Omega)$. This concludes the proof. □

3 Some examples

In this section we present some examples of conservative dG and FV methods which fit in the abstract framework above. Further examples which are not detailed here include the popular O-method (see, e.g., [1,4]).

3.1 Discontinuous Galerkin methods

In this section we shall present a number of dG methods which fit in the abstract analysis framework above. The weighted averaging techniques introduced in [14] for a domain decomposition method and later extended to dG methods in [18,21] will be used to ensure robust *a priori* estimates with respect to anisotropy and heterogeneity of the diffusion tensor in a suitable energy norm. The asymptotical convergence analysis can be performed following the guidelines of [18] and it is out of the scope of the present work. For all $F \in \mathcal{F}_h$ and for all φ s.t. a (possibly two-valued) trace is defined on F, we introduce the following jump operator:

$$[\![\varphi]\!] \stackrel{\text{def}}{=} \begin{cases} \varphi_{|T_1} - \varphi_{|T_2}, & \text{if } F = \mathcal{F}_h^{T_1} \cap \mathcal{F}_h^{T_2}, \\ \varphi_{|T}, & \text{if } F = \mathcal{F}_h^{T} \cap \mathcal{F}_h^{b}. \end{cases} \tag{19}$$

The space $P_h^{k_V}(\mathcal{T}_h)$, $k_V \geq 1$, will be equipped with the following norm:

$$\|p_h\|_{1,2,h}^2 \stackrel{\text{def}}{=} \|\nabla_h p_h\|_{[L^2(\Omega)]^d}^2 + \sum_{F \in \mathcal{F}_h} \frac{1}{h_F} \|[\![p_h]\!]\|_{L^2(F)}^2 \qquad \forall p_h \in P_h^{k_V}(\mathcal{T}_h), \tag{20}$$

where ∇_h denotes the broken gradient. The proof of Hypothesis 2.2 can be found in [17, §6]. The following assumption need be added to those listed in Definition 1:

Hypothesis 3.1. *Let $\mathcal{H}$ be a countable set and let $\{\mathcal{T}_h\}_{h \in \mathcal{H}}$ denote a family of meshes matching Definition 1. We require that the ratio of the diameter h_T, $T \in \mathcal{T}_h$, to the diameter of the largest ball inscribed in T be bounded from above by a parameter ϱ_3 independent of h.*

Remark 3.1. Hypothesis 3.1 is not needed to prove Lemmata 2.2–2.1 for $k_V \geq 1$, so it is not listed in Definition 1.

For a given $k_V \geq 1$ we let $\mathcal{S}_h = \mathcal{T}_h$ and set

$$V_h \stackrel{\text{def}}{=} P_h^{k_V}(\mathcal{T}_h), \qquad \Sigma_h \stackrel{\text{def}}{=} [P_h^{k_V}(\mathcal{T}_h)]^d.$$

We shall focus on the piecewise constant case $\nu \in [P_h^0(\mathcal{T}_h)]^{d\times d}$. Let $\nu_{|T} = V_T D_T V_T^{-1}$ be the diagonalizaton of ν on $T \in \mathcal{T}_h$, i.e., D_T is a diagonal matrix containing the eigenvalues of ν. Denote with κ the element of $[P_h^0(\mathcal{T}_h)]^{d\times d}$ s.t. $\kappa_{|T} = V_T D_T^{1/2} V_T^{-1}$ for all $T \in \mathcal{T}_h$. The tensor field κ is symmetric, uniformly positive definite and s.t. $\nu = \kappa\kappa$ for a.e. $x \in \Omega$. Let, moreover, $\kappa^{-1} \in [P_h^0(\mathcal{T}_h)]^{d\times d}$ denote the inverse of κ, i.e. $\kappa\kappa^{-1} = I$ for a.e. $x \in \Omega$.

Remark 3.2. The piecewise regular case $\nu \in [C_c^\infty(\mathcal{T}_h)]^{d\times d}$ requires only minor technical modifications in Lemma 3.1 below, which we omit for simplicity of exposition.

Since V_h is a piecewise polynomial space, the reconstruction operator r_h^V can be taken equal to the identity on V_h. For all $F \in \mathcal{F}_h$ and for all φ s.t. a (possibly two-valued) trace is defined on F, we define the following weighted average operator: For a.e. $x \in F$,

$$\{\!\{\varphi\}\!\}_\omega \stackrel{\text{def}}{=} \begin{cases} \omega_2 \varphi_{|T_1} + \omega_1 \varphi_{|T_2}, & \text{if } F = \mathcal{F}_h^{T_1} \cap \mathcal{F}_h^{T_2}, \\ \varphi_{|T}, & \text{if } F = \mathcal{F}_h^{T} \cap \mathcal{F}_h^{b}, \end{cases}$$

where

$$\omega = (\omega_1, \omega_2) \stackrel{\text{def}}{=} \left(\tfrac{\lambda_1}{\lambda_1+\lambda_2}, \tfrac{\lambda_2}{\lambda_1+\lambda_2}\right), \qquad \lambda_i \stackrel{\text{def}}{=} \sqrt{\nu_{|T_i}\mu_F \cdot \mu_F}, \qquad i \in \{1, 2\}.$$

Since $V_h = P_h^{k_V}$, we can take

$$\|v_h\|_{V_h} \stackrel{\text{def}}{=} \|v_h\|_{1,2,h},$$

with $\|\cdot\|_{1,2,h}$ defined as in (20). The following lifting operators will play a crucial role in what follows: For all $F \in \mathcal{F}_h$ and for all $\varphi \in L^2(F)$, let $l \geq 0$ and set

$$(r_{F,\kappa}^l(\varphi), \tau_h)_\Omega \stackrel{\text{def}}{=} (\varphi\mu_F, \{\!\{\kappa\tau_h\}\!\}_\omega)_{[L^2(F)]^d} \qquad \forall \tau_h \in [P_h^l(\mathcal{T}_h)]^d, \tag{21}$$

and define $R_\kappa^l(\varphi) \stackrel{\text{def}}{=} \sum_{F \in \mathcal{F}_h} r_{F,\kappa}^l(\varphi)$. For $l = k_V$ the subscript will be omitted. For all $v_h \in V_h$, the weakly converging gradient is defined as

$$\widetilde{G}(v_h) \stackrel{\text{def}}{=} \nabla_h v_h - \kappa^{-1} R_\kappa(v_h),$$

where ∇_h denotes the broken gradient.

Remark 3.3. To prove the convergence of the method, it is sufficient to work with the lifting operators r_F^0. However, if the exact solution u turns out to be more regular, optimal-order convergence rates can be established in the $\|\cdot\|_{V_h}$-norm when working with the lifting operators $r_F^{k_V-1}$ or $r_F^{k_V}$. The latter choice may be preferable for implementation purposes, especially if non-hierarchical, e.g. nodal-based, basis functions are used. For instance, if u belongs to the broken Sobolev space $H^{k+1}(\mathcal{T}_h)$, the usual *a priori* error analysis techniques can be used to infer a bound of the form $\|u - u_h\|_{V_h} \leq C_u h^k$, with C_u a positive parameter depending on the norm of the exact solution u, on ϱ_i, $i \in \{1 \ldots 3\}$, on k_V and on ν.

Several choices are possible for the consistent gradient G as well as for the bilinear form j_h. Some of the most common methods are presented in Tables 1–2, where we have set

$$\lambda_{\min,F} \stackrel{\text{def}}{=} \begin{cases} \min(\lambda_1, \lambda_2), & \text{if } F = \mathcal{F}_h^{T_1} \cap \mathcal{F}_h^{T_2}, \\ \sqrt{\nu_{|T}\mu_F \cdot \mu_F}, & \text{if } F \in \mathcal{F}_h^{T} \cap \mathcal{F}_h^{b}, \end{cases}$$

and

$$s_h(u_h, v_h) \stackrel{\text{def}}{=} (R_\kappa([\![u_h]\!]), R_\kappa([\![v_h]\!]))_{[L^2(\Omega)]^d}. \tag{22}$$

Remark 3.4. The original formulation of the methods proposed in [5,7,15,16,28] has been modified using the averaging techniques introduced in [18]. Optimal

Table 1: Consistent gradient choices for dG methods. Symmetric methods are marked with a star. The methods are named after the acronyms introduced in [6].

Method	Ref.	$G(u_h)$
SIPG*	Arnold [5]	$\nabla_h u_h - \kappa^{-1}R_\kappa(u_h)$
NIPG	Rivière, Wheeler *et al.* [28]	$\nabla_h u_h + \kappa^{-1}R_\kappa(u_h)$
IPG	Dawson, Sun *et al.* [16]	$\nabla_h u_h$
BR*	Bassi, Rebay *et al.* [7]	$\nabla_h u_h - \kappa^{-1}R_\kappa(u_h)$
LDG*	Cockburn and Shu [15]	$\nabla_h u_h - \kappa^{-1}R_\kappa(u_h)$

asymptotic order estimates which are also robust with respect to anisotropy and heterogeneity can be obtained in the following norm:

$$\|v_h\|^2_{\mathrm{DG},\nu} \stackrel{\mathrm{def}}{=} \|\kappa\nabla_h v_h\|^2_{[L^2(\Omega)]^d} + |v_h|^2_J, \qquad |v_h|^2_J \stackrel{\mathrm{def}}{=} \sum_{F\in\mathcal{F}_h} \frac{1}{h_F}\|\lambda^{1/2}_{\min,F}[\![v_h]\!]\|^2_{L^2(F)}.$$

The above norm is equivalent to $\|\cdot\|_{V_h}$ since, for all $v_h \in V_h$,

$$\underline{\lambda}^{1/2}\|v_h\|_{V_h} \leq \|v_h\|_{\mathrm{DG},\nu} \leq \overline{\lambda}^{1/2}\|v_h\|_{V_h}.$$

The following result was proved in [17]:

Lemma 3.1. *Assume that Hypothesis 3.1 holds. Then, there is* $C_{\mathrm{IP}} > 0$ *depending on* ϱ_i, $i \in \{1\dots3\}$, *on* k_V *but not on* h *s.t., for all* $F \in \mathcal{F}_h$, *for all* $v_h \in V_h$,

$$\|r_{F,\kappa}(v_h)\|^2_{[L^2(\Omega)]^d} \leq C_{\mathrm{IP}}|v_h|^2_J.$$

Furthermore, assume that there is a parameter ϱ_4 *independent of* h *s.t.*

$$h_F|F| \geq \varrho_4|T| \qquad \forall T \in \mathcal{T}_h,\ \forall F \in \mathcal{F}^T_h. \tag{23}$$

Then, for all $F \in \mathcal{F}_h$, *for all* $v_h \in V_h$, *there is* $c_{\mathrm{IP}} > 0$ *depending on* ϱ_i, $i \in \{1\dots4\}$, *on* k_V *but not on* h *s.t.*

$$c_{\mathrm{IP}}|v_h|^2_J \leq \|r_{F,\kappa}(v_h)\|^2_{[L^2(\Omega)]^d}. \tag{24}$$

Remark 3.5. Inequality (24) is only needed to prove the coercivity of the BR method (see Lemma 3.5 below), whereas it is not needed for the other methods listed in Tables 1–2. In what follows we shall therefore tacitly require (23) only when dealing with the BR method.

Lemma 3.2 (Proof of Hypothesis 2.3)**.** *Let the assumptions of Lemma 3.1 hold true. Then, Hypothesis 2.3 holds for all the consistent gradients listed in Table 1.*

Proof. Property (7) is in fact verified with the equal sign. Let us prove (8) for $\widetilde{G}$ (the proof for the gradients listed in Table 1 is similar and will be omitted). For all $v_h \in V_h$,

$$\|\widetilde{G}(v_h)\|^2_{[L^2(\Omega)]^d} \leq 2\|\nabla_h v_h\|^2_{[L^2(\Omega)]^d} + \frac{2}{\underline{\lambda}}\sum_{T\in\mathcal{T}_h}\|R_\kappa([\![v_h]\!])\|^2_{[L^2(T)]^d} \stackrel{\mathrm{def}}{=} S_1 + S_2.$$

According to (21), for all $F \in \mathcal{F}_h$, $r_{F,\kappa}$ is solely supported by the elements which share F. We thus have that $R_\kappa([\![v_h]\!])_{|T} = \sum_{F\in\mathcal{F}_h^T} r_{F,\kappa}([\![v_h]\!])_{|T}$ and, owing to Lemma 3.1,

$$S_2 \leq \frac{2N_\partial}{\underline{\lambda}} \sum_{F\in\mathcal{F}_h} \|r_{F,\kappa}(v_h)\|^2_{[L^2(\Omega)]^d} \leq \frac{2C_{\mathrm{IP}}N_\partial}{\underline{\lambda}}|v_h|^2_J \leq \frac{2C_{\mathrm{IP}}N_\partial\overline{\lambda}}{\underline{\lambda}}\|v_h\|^2_{V_h},$$

which yields $\|\widetilde{G}(v_h)\|^2_{[L^2(\Omega)]^d} \leq 2\left(1+\frac{2C_{\mathrm{IP}}N_\partial\overline{\lambda}}{\underline{\lambda}}\right)\|v_h\|^2_{V_h}$. □

Remark 3.6. The L^2 projector π_h^1 onto the space $P_h^1(\mathcal{T}_h)$ enjoys the following property:

$$\lim_{h\to\infty}\|\varphi - \pi_h^1\varphi\|_{V_h} = 0 \qquad \forall\varphi \in C_c^\infty(\Omega). \tag{25}$$

Lemma 3.3 (Proof of Hypothesis 2.4)**.** *Hypothesis 2.4 holds.*

Proof. Let $\{v_h\}_{h\in\mathcal{H}}$ be a sequence in V_h satisfying the assumptions of Hypothesis 2.4. The sequence $\{\widetilde{G}(v_h)\}_{h\in\mathcal{H}}$ is bounded in $[L^2(\Omega)]^d$, and it converges (up to a subsequence) to some $\tau \in [L^2(\Omega)]^d$. It only remains to prove that $\tau = \nabla v$ for a.e. $x \in \mathbb{R}^d$. Let $\Phi \in [C_c^\infty(\mathbb{R}^d)]^d$, $v_h \in V_h$ and prolong v_h by zero outside Ω. Observe that

$$(\widetilde{G}(v_h), \pi_h^1\Phi)_{[L^2(\mathbb{R}^d)]^d} = -(v_h, \nabla_h\cdot\pi_h^1\Phi)_{L^2(\mathbb{R}^d)} + \sum_{F\in\mathcal{F}_h^i}(\{\!\{v_h\}\!\}_\omega, \mu_F\cdot[\![\pi_h^1\Phi]\!])_{L^2(F)},$$

where $\nabla_h\cdot$ denotes the broken divergence operator. Owing to the regularity of Φ, $[\![\Phi]\!] = 0$ for a.e. $x \in F$, $F \in \mathcal{F}_h$. The above identity then yields

$$\begin{aligned}
&|(v_h, \nabla\cdot\Phi)_{L^2(\Omega)} + (\widetilde{G}(v_h), \pi_h^1\Phi)_{[L^2(\Omega)]^d}| \\
&\quad = |(v_h, \nabla_h\cdot(\Phi - \pi_h^1\Phi))_{L^2(\Omega)} - \sum_{F\in\mathcal{F}_h^i}(\{\!\{v_h\}\!\}_\omega, \mu_F\cdot[\![\Phi - \pi_h^1\Phi]\!])_{L^2(\Omega)}| \\
&\quad \leq \|v_h\|_{V_h}\|\Phi - \pi_h^1\Phi\|_{V_h}.
\end{aligned}$$

Passing to the limit and using (25) and the boundedness of $\{v_h\}_{h\in\mathcal{H}}$ in the $\|\cdot\|_{V_h}$ norm concludes the proof. □

Lemma 3.4 (Proof of Hypothesis 2.5)**.** *Let the stabilization parameters satisfy*

$$\eta_{\mathrm{SIPG}} > N_\partial C_{\mathrm{IP}}, \quad \eta_{\mathrm{NIPG}} > 0, \quad \eta_{\mathrm{IPG}} > N_\partial C_{\mathrm{IP}}/2, \quad \eta_{\mathrm{BR}} > N_\partial, \quad \eta_{\mathrm{LDG}} > 0.$$

Then, Hypothesis 2.5 holds for all the stabilizations of Table 2.

Proof. The continuity of the stabilizations of Table 2 stems from a simple application of the Cauchy-Schwarz inequality. The IFP as well as the LDG stabilizations are clearly positive. Proceeding as in the proof of Lemma 3.2, we have that

$$s_h(v_h, v_h) \leq N_\partial \sum_{F\in\mathcal{F}_h}\|r_{F,\kappa}([\![v_h]\!])\|^2_{[L^2(\Omega)]^d} \leq C_{\mathrm{IP}}N_\partial|v_h|^2_J,$$

which yields the positivity of the SIPG, NIPG and BR stabilization. The term s_h is introduced to reduce the stencil of the above methods to neighbouring elements. 2 immediately follows from the above remark provided the above

Table 2: Consistent stabilization choices for dG methods. Symmetric methods are marked with a star. The bilinear form s_h is defined in (22).

Method	$j_h(u_h, v_h)$
SIPG*	$\sum_{F\in\mathcal{F}_h}(\eta_{\mathrm{SIPG}}\frac{\lambda_{\min,F}}{h_F}[\![u_h]\!], [\![v_h]\!])_{L^2(F)} - s_h(u_h, v_h)$
NIPG	$\sum_{F\in\mathcal{F}_h}(\eta_{\mathrm{NIPG}}\frac{\lambda_{\min,F}}{h_F}[\![u_h]\!], [\![v_h]\!])_{L^2(F)} + s_h(u_h, v_h)$
IPG	$\sum_{F\in\mathcal{F}_h}(\eta_{\mathrm{IPG}}\frac{\lambda_{\min,F}}{h_F}[\![u_h]\!], [\![v_h]\!])_{L^2(F)}$
BR*	$\sum_{F\in\mathcal{F}_h}(\eta_{\mathrm{BR}} r_{F,\kappa}([\![u_h]\!]), r_{F,\kappa}([\![v_h]\!]))_{[L^2(F)]^d} - s_h(u_h, v_h)$
LDG*	$\sum_{F\in\mathcal{F}_h}(\eta_{\mathrm{LPG}}\frac{\lambda_{\min,F}}{h_F}[\![u_h]\!], [\![v_h]\!])_{L^2(F)}$

assumptions on the stabilization parameters are matched. In order to prove consistency, let $\varphi \in C_c^\infty(\Omega)$. Since $[\![\varphi]\!] = 0$ for a.e. $x \in F$, $F \in \mathcal{F}_h$, the continuity of j_h gives

$$j_h(\pi_h^1\varphi, \pi_h^1\varphi) \lesssim |\varphi_h|_J^2 = |\varphi_h - \varphi|_J^2 \leq \overline{\lambda}\|\pi_h^1\varphi - \varphi\|_{V_h}^2,$$

which, according to (25), tends to zero as $h \to 0$. □

Lemma 3.5 (Proof of Hypothesis 2.6)**.** *Under the assumptions of Lemma 3.4, Hypothesis 2.6 holds true for all the methods of Tables 1–2.*

Proof. For the sake of brevity, the proof will be detailed for the BR and SIPG methods only. For all $v_h \in V_h$, Young's inequality together with Lemma 3.1 yield

$$\begin{aligned}
a_h^{\mathrm{BR}}(v_h, v_h) &= \|\kappa\nabla_h v_h\|_{[L^2(\Omega)]^d}^2 + 2(\kappa\nabla_h v_h, R_\kappa(v_h))_{[L^2(\Omega)]^d} \\
&\quad + \eta_{\mathrm{BR}}\sum_{F\in\mathcal{F}_h}(r_{F,\kappa}([\![v_h]\!]), r_{F,\kappa}([\![v_h]\!]))_{[L^2(\Omega)]^d} \\
&\geq \frac{\epsilon\underline{\lambda}}{1+\epsilon}\|\nabla_h v_h\|_{[L^2(\Omega)]^d}^2 + (\eta_{\mathrm{BR}} - (1+\epsilon)N_\partial)\sum_{F\in\mathcal{F}_h}\|r_{F,\kappa}([\![v_h]\!])\|_{[L^2(\Omega)]^d}^2 \\
&\geq \frac{\epsilon\underline{\lambda}}{1+\epsilon}\|\nabla_h v_h\|_{[L^2(\Omega)]^d}^2 + (\eta_{\mathrm{BR}} - (1+\epsilon)N_\partial)c_{\mathrm{IP}}|v_h|_J^2,
\end{aligned}$$

for all $\epsilon > 0$. Coercivity then holds for $\eta_{\mathrm{BR}} > N_\partial$. Similarly,

$$\begin{aligned}
a_h^{\mathrm{SIPG}}(v_h, v_h) &= \|\kappa\nabla_h v_h\|_{[L^2(\Omega)]^d}^2 + 2(\kappa\nabla_h v_h, R_\kappa(v_h))_{[L^2(\Omega)]^d} \\
&\quad + \eta_{\mathrm{SIPG}}\sum_{F\in\mathcal{F}_h}(\frac{\lambda_{\min,F}}{h_F}[\![v_h]\!], [\![v_h]\!])_{L^2(F)} \\
&\geq \frac{\epsilon\underline{\lambda}}{1+\epsilon}\|\nabla_h v_h\|_{[L^2(\Omega)]^d}^2 \\
&\quad + (\eta_{\mathrm{SIPG}} - (1+\epsilon)N_\partial C_{\mathrm{IP}})\sum_{F\in\mathcal{F}_h}(\frac{\lambda_{\min,F}}{h_F}[\![v_h]\!], [\![v_h]\!])_{L^2(F)},
\end{aligned}$$

yielding coercivity for $\eta_{\mathrm{SIPG}} > N_\partial C_{\mathrm{IP}}$. □

Finally, Hypothesis (2.7) follows from (25)

3.2 A cell-based finite volume method

We consider hereafter a new finite volume method displaying all the ingredients introduced in §2. The salient feature of this method is the increased robustness with respect to standard MPFA methods together with a more compact stencil with respect to the SUSHI method of [24]. The stencil reduction is achieved by renouncing symmetry and using a compact gradient reconstruction for the test function.

Throughout the present and the following section, the following assumption on the mesh need be added to those listed in Definition 1:

Hypothesis 3.2. *Let $\mathcal{H}$ be a countable set and let $\{\mathcal{T}_h\}_{h\in\mathcal{H}}$ denote a family of meshes matching Definition 1. For all $F\in\mathcal{F}_h$ we let $x_F \stackrel{\mathrm{def}}{=} \int_F x/|F|$. Them*

(i) there is a positive parameter ϱ_5 independent of $h\in\mathcal{H}$ s.t.

$$\frac{|x_T - x_F|}{d_{T,F}} \le \varrho_5 \qquad \forall F\in\mathcal{F}_h^T,\ \forall T\in\mathcal{T}_h; \tag{26}$$

(ii) $\mathcal{P}_h$ is a family of points of Ω indexed by the elements of $\mathcal{T}_h$ and $\mathcal{P}_h = \{x_T\}_{T\in\mathcal{T}_h}$ is s.t., for all $T\in\mathcal{T}_h$, $x_T\in T$ and T is star-shaped with respect to x_T, i.e., $[x_T, x]\subset T$ for all $x\in T$;

(iii) there is $\varrho_2 > 0$ s.t., for all $F = \mathcal{F}_h^{T_1}\cap\mathcal{F}_h^{T_2}$, $(T_1,T_2)\in[\mathcal{T}_h]^2$,

$$\varrho_2 \le \frac{d_{T_1,F}}{d_{T_2,F}} \le \frac{1}{\varrho_2},$$

where, for all $T\in\mathcal{T}_h$ and for all $F\in\mathcal{F}_h^T$, we have set $d_{T,F} \stackrel{\mathrm{def}}{=} \mathrm{dist}(x_T, F) > 0$.

For all $T\in\mathcal{T}_h$ and for all $F\in\mathcal{F}_h^T$, we define

$$d_F \stackrel{\mathrm{def}}{=} \begin{cases} d_{T_1,F} + d_{T_2,F}, & \text{if } F = \mathcal{F}_h^{T_1}\cap\mathcal{F}_h^{T_2}, \\ d_{T,F}, & \text{if } F = \mathcal{F}_h^{T}\cap\mathcal{F}_h^{b}. \end{cases}$$

In the present and in the following section, the space $P_h^0(\mathcal{T}_h)$ will be equipped with the the discrete H_0^1 norm:

$$\|p_h\|_{1,2,h}^2 \stackrel{\mathrm{def}}{=} \sum_{F\in\mathcal{F}_h} \frac{1}{d_F}\|[\![p_h]\!]\|_{L^2(F)}^2 \qquad \forall p_h\in P_h^0(\mathcal{T}_h), \tag{27}$$

where the jump operator has been defined in (19). The proof that Hypothesis 2.2 holds for the norm (27) can be found in [24, §5]. Let

$$V_h \stackrel{\mathrm{def}}{=} P_h^0(\mathcal{T}_h), \qquad \Sigma \stackrel{\mathrm{def}}{=} [P_h^0(\mathcal{T}_h)]^d.$$

Since V_h is a piecewise polynomial space, the reconstruction operator r_h^V can be taken equal to the identity on V_h. For all $F\in\mathcal{F}_h$ and for all $v_h\in V_h$ we define the following trace operator $\gamma_F : V_h\to\mathbb{P}^0(F)$:

$$\gamma_F(v_h) \stackrel{\mathrm{def}}{=} \begin{cases} \omega_F^{T_2} v_{h|T_1} + \omega_F^{T_1} v_{h|T_2}, & \forall F = \mathcal{F}_h^{T_1}\cap\mathcal{F}_h^{T_2}, \\ 0, & \forall F = \mathcal{F}_h^{T}\cap\mathcal{F}_h^{b}, \end{cases} \qquad \omega_F^T \stackrel{\mathrm{def}}{=} \frac{d_{T,F}}{d_F} \le 1.$$

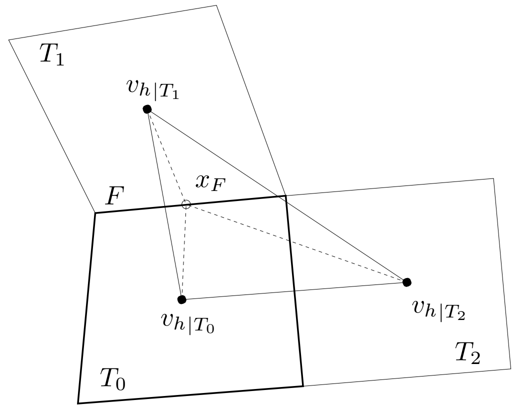

Figure 1: Barycentric interpolation for $d = 2$.

For all $T \in \mathcal{T}_h$, for all $F \in \mathcal{F}_h^T$, let $\mathcal{I}_F^T : V_h \to \mathbb{P}^0(F)$ denote a linear interpolation operator s.t.

$$|(\mathcal{I}_F^T \circ \pi_h^0)\varphi - \varphi(x_F)| \leq C_\varphi h_F d_{T,F} \qquad \forall \varphi \in C_c^\infty(\Omega), \tag{28}$$

where $\pi_h^0 \equiv \pi_V$ denotes the L^2 projection onto V_h, x_F is the barycenter of F and C_φ denotes a positive parameter depending on some (bounded) norm of φ.

Remark 3.7. A simple choice for the interpolator $\mathcal{I}_F^T$ is described hereafter. For the sake of simplicity, let $d = 2$. For all $F \in \mathcal{F}_h^T \cap \mathcal{F}_h^b$ we set $\mathcal{I}_F^T v_h = 0$. Let $F \in \mathcal{F}_h^{T_0} \cap \mathcal{F}_h^i$, $T_0 \in \mathcal{T}_h$, and let $T_1 \neq T_2$ be two elements of $\mathcal{T}_h \setminus \{T_0\}$ s.t. their barycenters are not aligned with that of T_0 (see Figure 3.2 for an example). Denote by $\{\alpha_i\}_{i\in\{0\ldots 2\}}$ the barycentric coordinates of x_F with respect to $\{x_{T_i}\}_{i\in\{0\ldots 2\}}$. Then, for all $v_h \in V_h$, we set

$$\mathcal{I}_F^{T_0} v_h \stackrel{\text{def}}{=} \sum_{i=0}^{2} \alpha_i v_{h|T_i}.$$

While the above choice ensures the convergence of the method, it does not yield strong consistency for piecewise linear exact solutions in the presence of heterogeneity. Other choices are possible, but their description lies out of the scope of the present paper. In particular, we refer to [3] for an alternative using the so called L interpolation introduced in [2].

For all $v_h \in V_h$, the gradient reconstructions are defined as follows: For all $T \in \mathcal{T}_h$,

$$\widetilde{G}(v_h)_{|T} \stackrel{\text{def}}{=} \frac{1}{|T|} \sum_{F\in\mathcal{F}_h^T} |F|(\gamma_F v_h - v_{h|T})\mu_F^T,$$

$$G(v_h)_{|T} \stackrel{\text{def}}{=} \frac{1}{|T|} \sum_{F\in\mathcal{F}_h^T} |F|(\mathcal{I}_F^T v_h - v_{h|T})\mu_F^T.$$

The space V_h will be equipped with the following norm:

$$\|v_h\|_{V_h}^2 \stackrel{\text{def}}{=} \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_h^T} \frac{|F|}{d_{T,F}} (\mathcal{I}_F^T v_h - v_{h|T})^2.$$

Remark 3.8. For all $h \in \mathcal{H}$ we have

$$\sum_{F \in \mathcal{F}_h^T} \frac{|F| d_{T,F}}{|T|} = d \qquad \forall T \in \mathcal{T}_h. \tag{29}$$

Lemma 3.6 (Proof of Hypothesis 2.3)**.** *Hypothesis 2.3 holds.*

Proof. Let v_h be a generic element of V_h. The Cauchy-Schwarz inequality gives

$$\frac{[\![v_h]\!]^2}{d_F} \leq \frac{(v_{h|T_1} - \mathcal{I}_F^{T_1} v_h)^2}{d_{T_1,F}} + \frac{(v_{h|T_2} - \mathcal{I}_F^{T_2} v_h)^2}{d_{T_2,F}} \qquad \forall F \in \mathcal{F}_h^{T_1} \cap \mathcal{F}_h^{T_2}.$$

Inequality (7) immediately follows. The Cauchy-Schwarz inequality together with (29) yield

$$\begin{aligned}
\|\widetilde{G}(v_h)\|^2_{[L^2(\Omega)]^d} &= \sum_{T \in \mathcal{T}_h} \frac{1}{|T|} \Big| \sum_{F \in \mathcal{F}_h^T} |F| \omega_F^T [\![v_h]\!] \mu_F \Big|^2 \\
&= \sum_{T \in \mathcal{T}_h} \left| \sum_{F \in \mathcal{F}_h^T} \left(\frac{|F| d_{T_F}}{|T|} \right)^{1/2} \times \left(\frac{|F|}{d_{T,F}}^{1/2} \right) \omega_{F,T} [\![v_h]\!] \mu_F \right| \\
&\leq \sum_{T \in \mathcal{T}_h} \left(\sum_{F \in \mathcal{F}_h^T} \frac{1}{d_{T,F}} \|[\![v_h]\!]\|^2_{L^2(F)} \times \sum_{F \in \mathcal{F}_h^T} \frac{|F| d_{T,F}}{|T|} \right) \\
&\leq d \|v_h\|^2_{1,2,h} \leq d \|v_h\|^2_{V_h}.
\end{aligned} \tag{30}$$

Similarly,

$$\begin{aligned}
\|G(v_h)\|^2_{[L^2(\Omega)]^d} &= \sum_{T \in \mathcal{T}_h} \frac{1}{|T|} \Big| \sum_{F \in \mathcal{F}_h^T} |F| (\mathcal{I}_F^T u_h - u_{h|T}) \mu_F^T \Big|^2 \\
&\leq \sum_{T \in \mathcal{T}_h} \left(\sum_{F \in \mathcal{F}_h^T} \frac{1}{d_{T,F}} \|\mathcal{I}_F^T u_h - u_{h|T}\|^2_{L^2(F)} \times \sum_{F \in \mathcal{F}_h^T} \frac{|F| d_{T,F}}{|T|} \right) \\
&\leq d \|v_h\|^2_{V_h}.
\end{aligned}$$

Observing that $\|v_h\|_{V_h}$ is bounded by assumption whereas the term in brackets tends to 0 as $h \to 0$ concludes the proof of (8). □

Lemma 3.7 (Proof of Hypothesis 2.4)**.** *Hypothesis 2.4 holds.*

Proof. Let $\{v_h\}_{h \in \mathcal{H}}$ be a sequence in V_h satisfying the assumptions of Hypothesis 2.4. The sequence $\{\widetilde{G}(v_h)\}_{h \in \mathcal{H}}$ is bounded, and it converges (up to a subsequence) to some $\tau \in [L^2(\Omega)]^d$. It only remains to prove that $\tau = \nabla v$ for a.e. $x \in \Omega$. Let $\Phi \in [C_c^\infty(\Omega)]^d$ and prolong v_h by zero outside Ω. Define $\Phi_h^T \stackrel{\text{def}}{=} \int_T \Phi / |T| = \pi_h^0 \Phi_{|T}$ for all $T \in \mathcal{T}_h$ and $\Phi_h^F \stackrel{\text{def}}{=} \int_F \Phi / |F|$ for all $F \in \mathcal{F}_h$.

Integration by parts yields

$$
\begin{aligned}
|(\widetilde{G}(v_h), \Phi)_{[L^2(\Omega)]^d} &+ (\nabla{\cdot}\Phi, r_h^V v_h)_{L^2(\Omega)}| \\
&= \Big| \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_h^T} |F|(\gamma_F v_h - v_{h|T})(\Phi_h^F - \Phi_h^T){\cdot}\mu_F^T \Big| \\
&\le \|v_h\|_{V_h} \left(\sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_h^T} |F| d_{T,F} (\Phi_h^F - \Phi_h^T)^2 \right)^{\frac{1}{2}},
\end{aligned}
$$

which proves the assertion. □

Define the stabilization term as follows:

$$
j(u_h, v_h) \overset{\text{def}}{=} \sum_{T\in\mathcal{T}_h} \eta_{\text{CVF}}^T \sum_{F\in\mathcal{F}_h^T} \frac{1}{d_{T,F}} (R_{T,F}(u_h), R_{T,F}(v_h))_{L^2(F)},
$$

where, for all $v_h \in V_h$, we have set $R_{T,F}(v_h) \overset{\text{def}}{=} \mathcal{I}_F^T v_h - v_{h|T} - G(v_h)_{|T}{\cdot}(x_F - x_T)$, and, for all $T \in \mathcal{T}_h$, $0 < \underline{\eta} \le \eta_{\text{CVF}}^T < \overline{\eta} \le \infty$ denotes a positive stabilization parameter.

Lemma 3.8 (Proof of Hypothesis 2.5)**.** *Hypothesis 2.5 holds.*

Proof. The proposed stabilization term is clearly symmetric and positive semidefinite. In order to prove the continuity, observe that, for all $v_h \in V_h$,

$$
j_h(v_h, v_h) \le
$$

$$
2 \sum_{T\in\mathcal{T}_h} \eta_{\text{CVF}}^T \left(\sum_{F\in\mathcal{F}_h^T} \frac{1}{d_{T,F}} \|\mathcal{I}_F^T v_h - v_{h|T}\|_{L^2(F)}^2 + \sum_{F\in\mathcal{F}_h^T} \frac{|F|}{d_{T,F}} (G(v_h){\cdot}(x_F - x_T))^2 \right).
$$

Let S_1^T, S_2^T the addends in brackets. Using (29) together with Hypothesis 3.2 and Lemma 3.6 we have that

$$
\sum_{T\in\mathcal{T}_h} S_2^T \le \varrho_5 \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_h^T} \frac{|F| d_{T,F}}{|T|} |T| |G(v_h)|^2 \le d\varrho_5 \|G(v_h)\|_{[L^2(\Omega)]^d}^2 \le d^2 \varrho_5 \|v_h\|_{V_h}^2,
$$

whence $j_h(v_h, v_h) \le 2\overline{\eta}(1 + d^2\varrho_5)\|v_h\|_{V_h}^2$. Using the above result together with (11) we have

$$
j_h(u_h, v_h) \le j_h(u_h, u_h)^{1/2} j_h(v_h, v_h)^{1/2} \le 2\overline{\eta}(1 + d^2\varrho_5)\|u_h\|_{V_h}\|v_h\|_{V_h}.
$$

It only remains to proof the consistency of j_h. In the rest of the proof, shall assume that (15) holds (a proof is given in Lemma 3.10 below). Let $\varphi \in C_c^\infty(\Omega)$ and set $\varphi_h \overset{\text{def}}{=} \pi_h^0 \varphi$. Observe that

$$
|R_{T,F}(v_h)| \le |\mathcal{I}_F^T \varphi_h - \varphi(x_F)| + |(\nabla\varphi(x_T) - G(\varphi_h)){\cdot}(x_F - x_T)| + c_\varphi |x_T - x_F|^2,
$$

where c_φ denotes a positive parameter depending on a suitable (bounded) norm of φ. Substituting in the expression of j_h and using Hypothesis 3.2 we obtain

$$
\begin{aligned}
\frac{1}{4\overline{\eta}} j_h(\varphi_h, \varphi_h) &\le \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_h^T} \frac{|F|}{d_{T,F}} |\mathcal{I}_F^I \varphi_h - \varphi(x_F)|^2 \\
&\quad + \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_h^T} \frac{|F|}{d_{T,F}} |\nabla\varphi(x_T) - G(\varphi_h)|^2 + |F| \varrho_5 c_\varphi h_T^3.
\end{aligned}
$$

Let S_i, $i \in \{1,2\}$ denote the first two addends in brackets. Using (28) together with (29) we have

$$S_1 \le C_\varphi \sum_{T\in\mathcal{T}_h} |T| \sum_{F\in\mathcal{F}_h^T} \frac{|F| d_{T,F}}{|T|} h_F^2 \le C_\varphi h^2 d|\Omega|,$$

i.e., $S_1 \to 0$ as $h \to 0$. Using Hypothesis 3.2 and (29) we have

$$\begin{aligned} S_2 &\le \sum_{T\in\mathcal{T}_h} |T| |\nabla\varphi(x_T) - G(\varphi_h)|^2 \sum_{F\in\mathcal{F}_h^T} \frac{|F| d_{T,F}}{|T|} \frac{|x_F - x_T|^2}{d_{T,F}^2} \\ &\le d\varrho_5^2 \sum_{T\in\mathcal{T}_h} \|\nabla\varphi(x_T)\|_{[L^2(T)]^d}^2 \\ &\le 2d\varrho_5^2 \sum_{T\in\mathcal{T}_h} \left(\|\nabla\varphi(x_T) - \nabla\varphi\|_{[L^2(T)]^d}^2 + \|\nabla\varphi - G(\varphi_h)\|_{[L^2(T)]^d}^2 \right), \end{aligned}$$

which, since (15) holds, shows that S_2 tends to zero as $h \to 0$. This concludes the proof. □

As the FV method proposed in this section is non-symmetric, it is conditionally coercive. In what follows, we shall provide a computable criterion to check coercivity for a given mesh $\mathcal{T}_h$ and diffusion tensor ν. For the sake of simplicity we shall refer to the interpolator defined in Remark 3.7. For a given $T \in \mathcal{T}_h$ we introduce the bilinear form a_h^T defined as

$$\begin{aligned} a_h^T(u_h, v_h) = (\nu G(u_h)_{|T}, \widetilde{G}(u_h))_{[L^2(T)]^d} \\ + \eta_{\mathrm{CVF}}^T \sum_{F\in\mathcal{F}_h^T} \frac{1}{d_{T,F}} (R_{T,F}(u_h), R_{T,F}(v_h))_{L^2(F)}. \end{aligned}$$

Let $\mathcal{T}_h^T \stackrel{\text{def}}{=} \{T' \in \mathcal{T}_h, \mathcal{F}_h^{T'} \cap \mathcal{F}_h^T \neq \emptyset\}$ denote the set of elements sharing a face with T and set $m^T \stackrel{\text{def}}{=} \mathrm{card}(\mathcal{T}_h^T)$. For brevity of notation, we shall note $\mathcal{T}_h^T = \{T_i\}_{1\le i\le m^T}$ with T_i sharing the internal face F_i with T. Moreover, we define $m^F \stackrel{\text{def}}{=} \mathrm{card}(\mathcal{F}_h^{T_i} \cap \mathcal{F}_h^b)$ and set $\{F_i\}_{m^T+1\le i\le m^T+m^F} \stackrel{\text{def}}{=} \mathcal{F}_h^T \cap \mathcal{F}_h^b$. Define the linear map $X^T : V_h \mapsto \mathbb{R}^{(m^T+m^F)}$ s.t., for all $v_h \in V_h$,

$$X^T(v_h) \stackrel{\text{def}}{=} \{\{v_{h|T_i} - v_{h|T}\}_{1\le i\le m^T}, \{\mathcal{I}_{F_i}^T(v_h) - v_{h|T}\}_{m^T+1\le i\le m^T+m^F}\},$$

and recall that $\mathcal{I}_{F_i}^T(v_h) = 0$ for $m^T + 1 \le i \le m^T + m^F$ (since $\mathcal{I}_{F_i}^T(v_h)$ vanishes on boundary faces). It is a simple matter to verify that for all $T \in \mathcal{T}_h$, there exists a matrix $A_h^T \in \mathbb{R}^{(m^T+m^F)\times(m^T+m^F)}$ s.t., for all $(u_h, v_h) \in [V_h]^2$,

$$a_h^T(u_h, v_h) = (X^T(u_h))^t A^T X^T(v_h).$$

Notice also that, again because $\mathcal{I}_{F_i}^T(v_h) = 0$ for $m^T + 1 \le i \le m^T + m^F$, we can write

$$\mathcal{I}_{F_i}^T v_h = v_{h|T} + \sum_{j=1}^{m^T+m^F} \beta_{ij}^T X^T(v_h)_j, \qquad 1 \le i \le m^T + m^F,$$

where the family of reals $\{\beta_{ij}^T\}_{1\le j\le m^T+m^F}$ verifies $\sum_{j=1}^{m^T+m^F} \beta_{i,j}^T = 1$. Let $B^T \in \mathbb{R}^{(m^T+m^F)\times(m^T+m^F)}$ be the matrix of elements β_{ij}^T and define the norm $\|\cdot\|_T$ as follows: For all $x \in \mathbb{R}^{m^T+m^F}$,

$$\|x\|_T^2 \stackrel{\text{def}}{=} \sum_{i=1}^{m^T+m^F} \frac{|F_i|}{d_{T,F_i}} (B^T x)_i^2. \tag{31}$$

The following result provides a computable local criterion expressed in term of the local matrices $\{A^T\}_{T\in\mathcal{T}_h}$:

Lemma 3.9 (Proof of Hypothesis 2.6)**.** *The bilinear form a_h is coercive if for all $T \in \mathcal{T}_h$, the matrix A^T is uniformly coercive for the norm $\|\cdot\|_T$, i.e. if there is $C > 0$ independent of h s.t., for all $x \in \mathbb{R}^{m^T+m^F}$, $x^t A^T x \ge C\|x\|_T^2$.*

Proof. For all $v_h \in V_h$,

$$\begin{aligned} a_h(v_h, v_h) &= \sum_{T\in\mathcal{T}_h} a_h^T(u_h, u_h) = \sum_{T\in\mathcal{T}_h} (X^T(u_h))^t A^T X^T(u_h) \\ &\ge C \sum_{T\in\mathcal{T}_h} \|X^T(u_h)\|_T^2 = C\|u_h\|_{V_h}^2, \end{aligned}$$

which concludes the proof. □

Lemma 3.10 (Proof of Hypothesis 2.7)**.** *Hypothesis 2.7 holds.*

Proof. Estimates (13)–(14) classically hold for $\pi_V = \pi_h^0$ (see, e.g., [19]). Let now $\varphi \in C_c^\infty(\Omega)$, set $\varphi_h \stackrel{\text{def}}{=} i_h^0\varphi$ and observe that, for all $T \in \mathcal{T}_h$, for a.e. $x \in T$,

$$\begin{aligned} (G(\varphi_h) - \nabla\varphi(x))_{|T} &= \sum_{F\in\mathcal{F}_h^T} \frac{|F|}{|T|} (\mathcal{I}_F^T \varphi_h - \varphi(x_F))\mu_F^T + (\nabla\varphi(\hat{x}_T) - \nabla\varphi(x)) \\ &\quad + \left(\sum_{F\in\mathcal{F}_h^T} \frac{|F|}{|T|} (\varphi(x_F) - \varphi(\hat{x}_T))\mu_F^T - \nabla\varphi(\hat{x}_T) \right) \\ &\stackrel{\text{def}}{=} S_1^T + S_2^T + S_3^T, \end{aligned}$$

where we have used the fact that, owing to assumption (iii) in Definition 1, $\sum_{F\in\mathcal{F}_h^T} \mu_F^T = 0$ for all $T \in \mathcal{T}_h$ to replace $\varphi_{h|T}$ with $\varphi(\hat{x}_T)$ in S_3^T. Clearly, $\|G(\varphi_h) - \nabla\varphi\|_{[L^2(\Omega)]^d}^2 \le 3\sum_{i=1}^3 \sum_{T\in\mathcal{T}_h} \|S_i^T\|_{[L^2(T)]^d}^2$. Estimate (28) together with (29) yields, for all $T \in \mathcal{T}_h$

$$|S_1^T| \le \sum_{F\in\mathcal{F}_h^T} \frac{|F| d_{T,F}}{|T|} \frac{|\mathcal{I}_F^T\varphi_h - \varphi(x_F)|}{d_{T,F}} \le C'_\varphi d h_T,$$

so that $\left(\sum_{T\in\mathcal{T}_h} \|S_1^T\|_{[L^2(\Omega)]^d}^2\right)^{1/2} \le C'_\varphi |\Omega|^{1/2} d h_T$. On the other hand, using classical estimates for π_h^0, we conclude that $\|S_2^T\|_{[L^2(T)]^d} \le C h_T \|\varphi\|_{H^2(T)}$. Finally, thanks to the regularity of φ, there is C''_φ depending on φ and on the mesh regularity s.t. $\|S_3^T\|_{[L^2(T)]^d}^2 \le C''_\varphi |T| h_T^2$, i.e., $\left(\sum_{T\in\mathcal{T}_h} \|S_3^T\|_{[L^2(T)]^d}^2\right)^{1/2} \le C''_\varphi |\Omega|^{1/2} h$. The above estimates yield the desired result. □

Table 3: Convergence results for the FV method of §3.2 with anisotropy ratio of 1.

$1/h$	nunkw	nnmat	erl2	ocverl2	umin	umax
16	255	3001	$3.98e-03$	–	$7.54e-03$	$9.97e-01$
32	1023	12665	$1.00e-03$	$1.99e+00$	$1.02e-03$	$1.00e-00$
64	4095	51961	$2.71e-04$	$1.89e+00$	$2.79e-04$	$1.00e+00$
128	16383	210425	$6.58e-05$	$2.04e+00$	$9.84e-05$	$1.00e-00$

Table 4: Convergence results for the FV method of §3.2 with anisotropy ratio of 1000.

$1/h$	nunkw	nnmat	erl2	ocverl2	umin	umax
16	255	3001	$2.82e-01$	–	$-2.93e-01$	$1.10e+00$
32	1023	12665	$7.98e-02$	$1.82e+00$	$-1.22e-01$	$1.01e+00$
64	4095	51961	$2.00e-02$	$2.00e+00$	$-1.04e-01$	$1.00e+00$
128	16383	210425	$3.94e-03$	$2.34e+00$	$-8.36e-03$	$1.00e-00$

For the sake of completeness, the order of convergence of the new FV method presented in this section has been numerically evaluated by solving the Dirichlet problem for $d = 2$ with $u = \sin(\pi x)\sin(\pi y)$ ($u_{\min} = 0$, $u_{\max} = 1$), $f = -\triangle u$ and anisotropy ratios of 1 and 1000 on a family of randomly perturbed quadrangular meshes of $(0, 1)^2$. The results are reported in Tables 3.2 and 3.2 and show second order convergence as well as robustness with respect to anisotropy and mesh skewdness. The following indicators are also listed: (i) `nunkw`, the number of unknowns; (ii) `nnmat`, the number of nonzero matrix entries; (iii) `erl2`, the L^2 error; (iv) `ocverl2`, the order of convergence for the L^2 error; (v) `umin` and `umax`, the minimum and maximum value of the discrete solution. A thorough validation of the above method will be the subject of a future work. An asimptotic *a priori* analysis can be performed following the guidelines of [22], but it lies out of the scope of the present work.

3.3 A hybrid finite volume method

The goal of this section is to show that the hybrid finite volume method proposed in [23] fits in the frameword of §2. Hypothesis 3.2 is assumed to hold and $P_h^0(\mathcal{T}_h)$ is again equipped with the norm defined in (27). To this purpose, for each $T \in \mathcal{T}_h$, for all $F \in \mathcal{F}_h^T$ we let $K_{T,F}$ denote the cone defined by F and x_T (see Figure 3.3). Throughout this section, x_F will denote the barycenter of a face $F \in \mathcal{F}_h$. Thanks to Hypothesis 3.2, the cones are well-defined and they satisfy

$$|K_{T,F}| = \frac{|F| d_{T,F}}{d}. \tag{32}$$

(To check the above relation, it suffices to partition F into $d-1$-simplices and obtain $|K_{T,F}|$ as the sum of the volumes of the d-simplices obtained by joining

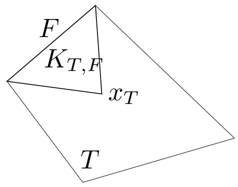

Figure 2: A face based cone for $d = 2$.

x_T to the vertices of each simplex). Define the spaces of hybrid unknowns:

$$H_h \stackrel{\text{def}}{=} \{\{u_h^T\}_{T\in\mathcal{T}_h},\ \{u_h^F\}_{F\in\mathcal{F}_h}\},$$
$$H_h^0 \stackrel{\text{def}}{=} \{v_h \in H_h;\ v_h^F = 0,\ \forall F \in \mathcal{F}_h^b\}.$$

For all $h \in \mathcal{H}$, we let $\mathcal{S}_h \stackrel{\text{def}}{=} \{K_{T,F}\}_{(T\in\mathcal{T}_h, F\in\mathcal{F}_h^T)}$ and set

$$V_h \stackrel{\text{def}}{=} H_h^0, \qquad \Sigma_h \stackrel{\text{def}}{=} [P_h^0(\mathcal{S}_h)]^d.$$

The space V_h is equipped with the following norm:

$$\|v_h\|_{V_h}^2 \stackrel{\text{def}}{=} \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T} \frac{|F|}{d_{T,F}}(v_h^F - v_h^T)^2.$$

The gradient reconstructions are defined as follows: For all $v_h \in V_h$,

$$G(v_h)_{|K_{T,F}} = \widetilde{G}(v_h)_{|K_{T,F}} = G_T(v_h) + R_{T,F}(v_h)\mu_F^T \qquad \forall K_{T,F} \in \mathcal{S}_h,$$

where we have set

$$G_T(v_h) \stackrel{\text{def}}{=} \frac{1}{|T|}\sum_{F\in\mathcal{F}_h^T} |F|(v_h^F - v_h^T)\mu_F^T,$$
$$R_{T,F}(v_h) \stackrel{\text{def}}{=} \frac{d^{1/2}}{d_{T,F}}\left(v_h^F - v_h^T - G_T(v_h)\cdot(x_F - x_T)\right).$$

The reconstruction operator $r_h^V : V_h \to P_h^0(\mathcal{T}_h)$ is defined as follows: For all $v_h \in V_h$, $r_h^V v_h = p_h \in P_h^0(\mathcal{T}_h)$ with $p_{h|T} = v_h^T$, for all $T \in \mathcal{T}_h$. The interpolation operator onto V_h is defined as follows: For all $\varphi \in C_c^\infty(\Omega)$, $\pi_h^V \varphi = \varphi_h \in V_h$ with $\varphi_h^T = \varphi(x_T)$ for all $T \in \mathcal{T}_h$, $\varphi_h^F = \varphi(x_F)$. Observe that φ_h belongs to V_h since φ vanishes on the boundary of Ω.

Remark 3.9. For all $T \in \mathcal{T}_h$ and for all $\hat{x} \in \mathbb{R}^d$, the following relation holds:

$$\sum_{F\in\mathcal{F}_h^T} |F|(\mu_F^T)_i(x_F - \hat{x})_j = \delta_{ij}|T|, \tag{33}$$

where the ith component of a vector quantity was denoted $(\cdot)_i$ and δ_{ij} is the Kronecker symbol.

Proposition 3.1. *For all $v_h \in V_h$ and for all $\sigma_h \in [P_h^0(\mathcal{T}_h)]^d$,*

$$\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T} |K_{T,F}|\sigma_{h|T}\cdot\mu_F^T R_{T,F}(v_h) = 0.$$

Proof. Using the definition of the residual, we obtain

$$\sum_{T\in\mathcal{T}_h} d^{1/2}\sigma_{h|T}\cdot\left(\sum_{F\in\mathcal{F}_h^T}\frac{|K_{T,F}|}{d_{T,F}}(v_h^F-v_h^T)\mu_F^T-\sum_{F\in\mathcal{F}_h^T}\frac{|K_{T,F}|}{d_{T,F}}G_T(v_h)\cdot(x_F-x_T)\mu_F^T\right).$$

Let S_1 and S_2 the addends in brackets. By definition, $S_1=|T|d^{-1}G_T(v_h)$. On the other hand, (32) together with (33) yield

$$S_2=-\frac{1}{d}(G_T(v_h))_i\sum_{F\in\mathcal{F}_h^T}|F|(x_F-x_T)_i\mu_F^T=-\frac{|T|}{d}G_T(v_h),$$

and the desired result follows. □

Lemma 3.11 (Proof of Hypothesis 2.3)**.** *Hypothesis 2.3 holds.*

Proof. The bound (7) can be proved as in Lemma 3.6. In order to prove (8), observe that, owing to Proposition 3.1,

$$\begin{aligned}\|G(v_h)\|^2_{[L^2(\Omega)]^d}&=\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}|K_{T,F}||G(v_h)|^2\\&=\sum_{T\in\mathcal{T}_h}|T||G_T(v_h)|^2+\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}|K_{T,F}||R_{T,F}(v_h)|^2\stackrel{\text{def}}{=}S_1+S_2.\end{aligned}$$

For the first term, using (32) together with (29) we have

$$\begin{aligned}S_1&=\sum_{T\in\mathcal{T}_h}\frac{1}{|T|}\Big|\sum_{F\in\mathcal{F}_h^T}|F|(v_h^F-v_h^T)\mu_F^T\Big|^2\\&\le\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}\frac{|F|d_{T,F}}{|T|}\times\sum_{F\in\mathcal{F}_h^T}\frac{|F|}{d_{T,F}}(v_h^F-v_h^T)^2\le d\|v_h\|^2_{V_h}.\end{aligned}\tag{34}$$

Substituting the expression of $R_{T,F}$ in the second term yields

$$\begin{aligned}S_2&\le 2\left(\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}\frac{|F|}{d_{T,F}}(v_h^F-v_h^T)^2+\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}|K_{T,F}||G_T(v_h)|^2\frac{|x_F-x_T|^2}{d_{T,F}^2}\right)\\&\le 2(1+\varrho_5 d)\|v_h\|^2_{V_h},\end{aligned}$$

which proves the assertion. □

Lemma 3.12 (Proof of Hypothesis 2.4)**.** *Hypothesis 2.4 holds.*

Proof. Let $\{v_h\}_{h\in\mathcal{H}}$ be a sequence in V_h satisfying the assumptions of Hypothesis 2.4. The sequence $\{\widetilde{G}(v_h)\}_{h\in\mathcal{H}}$ is bounded, and it converges (up to a subsequence) to some $\tau\in[L^2(\Omega)]^d$. It only remains to prove that $\tau=\nabla v$ for a.e. $x\in\mathbb{R}^d$. Let $\Phi\in[C_c^\infty(\mathbb{R}^d)]^d$ and let $\Phi_h=\pi_h^V\Phi$. We have

$$(\widetilde{G}(v_h),\Phi)_{[L^2(\mathbb{R}^d)]^d}=\sum_{T\in\mathcal{T}_h}(G_T(v_h),\Phi)_{[L^2(T)]^d}+\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}(R_{T,F}(v_h)\mu_F^T,\Phi)_{[L^2(K_{T,F})]^d}$$

Denote by S_1 and S_2 the addends in the right hand side. Integration by part yields

$$
\begin{aligned}
|(G_T(v_h),\Phi)_{[L^2(\Omega)]^d} + (\nabla{\cdot}\Phi, r_h^V v_h)_{L^2(\Omega)}| &= \Big|\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T} |F|(v_h^F - v_h^T)(\Phi_h^F - \Phi_h^T){\cdot}\mu_F^T\Big| \\
&\le \|v_h\|_{V_h}\left(\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}|F|d_{T,F}(\Phi_h^F-\Phi_h^T)^2\right)^{\frac{1}{2}},
\end{aligned}
$$

which proves that $S_1 \to -(v,\nabla{\cdot}\Phi)_{L^2(\mathbb{R}^d)}$ as $h\to 0$ since the sequence $\{v_h\}_{h\in\mathcal{H}}$ is bounded in the $\|\cdot\|_{V_h}$ norm. Let us now consider the second term. Owing to the regularity of Φ, there exists $C_\Phi > 0$ only depending on Φ s.t. $|\int_{K_{T,F}}(\Phi-\Phi_h^T)| \le C_\Phi|K_{T,F}|h$. Using Proposition 3.1 with $\sigma_h = \Phi_h$ and (34), the Cauchy-Schwarz inequality yields

$$
\begin{aligned}
S_2 &= \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}(R_{T,F}(v_h)\mu_F^T, \Phi-\Phi_h)_{[L^2(K_{T,F})]^d} \\
&\le \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}|R_{T,F}(v_h)|\Big|\int_{K_{T,F}}(\Phi-\Phi_h^T)\Big| \\
&\le \sqrt{2}C_\Phi h|\Omega|^{\frac{1}{2}}\left(\|v_h\|_{V_h}^2 + d\varrho_5^2\|G_T(v_h)\|_{[L^2(\Omega)]^d}^2\right)^{\frac{1}{2}} \\
&\le \sqrt{2}C_\Phi h|\Omega|^{\frac{1}{2}}(1+d\varrho_5)\|v_h\|_{V_h},
\end{aligned}
$$

which proves that $S_2 \to 0$ as $h\to 0$. □

Since residual terms are incorporated in the gradient reconstruction, the above method can be shown to be stable without further penalization. We thus take $j_h(u_h,v_h)=0$, which trivially satisfies Hypothesis 2.5.

Let $\nu_h \in [P_h^0(\mathcal{T}_h)]^{d\times d}$ be s.t., for all $T\in\mathcal{T}_h$, $\nu_{h|T} = \int_T \nu/|T|$.

Lemma 3.13 (Proof of Hypothesis 2.6)**.** *Hypothesis 2.6 holds.*

Proof. Let $v_h \in V_h$. Using Proposition 3.1 with σ_h s.t. $\sigma_{h|T} = G_T(v_h)$ for all $T\in\mathcal{T}_h$,

$$
a_h(v_h,v_h) = \sum_{T\in\mathcal{T}_h}|T|\nu_{h|T}G_T(v_h){\cdot}G_T(v_h) + \sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}|K_{T,F}|\nu_{h|T}\mu_F^T{\cdot}\mu_F^T R_{T,F}(v_h)^2.
$$

Denote by S_1 and S_2 the addends in the right hand side. Clearly, $S_1 \ge \underline{\lambda}\|G_T(v_h)\|_{[L^2(\Omega)]^d}^2$. For $\epsilon > 0$ and for all $(a,b)\in\mathbb{R}^2$, there holds $(a-b)^2 \ge \frac{\epsilon}{1+\epsilon}a^2 - \epsilon b^2$. Applying the above inequality with $a = v_h^F - v_h^T$ and $b = G_T(v_h){\cdot}(x_F - x_T)$ yields:

$$
\begin{aligned}
S_2 &\ge \frac{\epsilon\underline{\lambda}}{1+\epsilon}\|v_h\|_{V_h}^2 - \epsilon d\overline{\lambda}\sum_{T\in\mathcal{T}_h}\sum_{F\in\mathcal{F}_h^T}|K_{T,F}||G_T(v_h)|^2\frac{|x_F-x_T|^2}{d_{T,F}^2} \\
&\ge \frac{\epsilon\underline{\lambda}}{1+\epsilon}\|v_h\|_{V_h}^2 - \epsilon d\varrho_5^2\overline{\lambda}\sum_{T\in\mathcal{T}_h}|T||G_T(v_h)|^2 \\
&= \frac{\epsilon\underline{\lambda}}{1+\epsilon}\|v_h\|_{V_h}^2 - \epsilon d\varrho_5^2\overline{\lambda}\|G_T(v_h)\|_{[L^2(\Omega)]^d}^2.
\end{aligned}
$$

Coercivity thus holds for $\epsilon \le \underline{\lambda}/(d\varrho_5^2\overline{\lambda})$. □

Remark 3.10. A coercivity constant independent of the anisotropy ratio $\underline{\lambda}/\overline{\lambda}$ could be derived proceeding as in [24]. We have preferred this shorter proof for brevity of presentation.

Lemma 3.14 (Proof of Hypothesis 2.7)**.** *Hypothesis 2.7 holds.*

Proof. Let $\varphi \in C_c^\infty(\Omega)$ and set $\varphi_h \stackrel{\text{def}}{=} \pi_h^V \varphi$. Observe that

$$\|\varphi_h\|_{V_h}^2 = \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_h^T} |K_{T,F}| \frac{d}{d_{T,F}^2} (\varphi_h^F - \varphi_h^T)^2$$
$$\leq d \|\nabla\varphi\|_{[L^\infty(\Omega)]^d}^2 \sum_{T\in\mathcal{T}_h} \sum_{F\in\mathcal{F}_h^T} |K_{T,F}| \frac{|x_F - x_T|^2}{d_{T,F}^2} \leq d\varrho_5^2 \|\nabla\varphi\|_{[L^\infty(\Omega)]^d}^2 |\Omega|,$$

i.e., (13) is verified with $\sigma_\varphi = (d|\Omega|)^{1/2} \varrho_5 \|\nabla\varphi\|_{[L^\infty(\Omega)]^d}$. The proof of (14) is classical and will be omitted (see e.g. [19]). It has been proved in [24, Lemma 4.3] that $\|G(v_h) - \nabla\varphi\|_{[L^\infty(\Omega)]^d} \leq C_\varphi h$, where $C_\varphi > 0$ is a parameter depeding on φ, on d and on the mesh regularity parameters ϱ_i, $i \in \{1\ldots 3, 5\}$. As a consequence, $\|G(v_h) - \nabla\varphi\|_{[L^2(\Omega)]^d} \leq |\Omega|^{1/2} \|G(v_h) - \nabla\varphi\|_{[L^\infty(\Omega)]^d}$ tends to zero as $h \to 0$, which concludes the proof. □

References

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

[2] 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.

[3] 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.

[4] L. Agélas and R. Masson, *Convergence of the finite volume MPFA O scheme for heterogenesous anisotropic diffusion problems on general meshes*, Finite Volumes for Complex Applications V (R. Eymard and J.-M. Hérard, eds.), John Wiley & Sons, 2008, pp. 145–152.

[5] D. N. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal. **19** (1982), 742–760.

[6] D. N. Arnold, F. Brezzi, B. Cockburn, and D. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal. **39** (2002), no. 5, 1749–1779.

[7] F. Bassi, S. Rebay, G. Mariotti, S. Pedinotti, and M. Savini, *A high-order accurate discontinuous finite element method for inviscid and viscous turbomachinery flows*, Proceedings of the 2nd European Conference on Turbomachinery Fluid Dynamics and Thermodynamics (R. Decuypere and G. Dibelius, eds.), 1997, pp. 99–109.

[8] S. Brenner, *Poincaré-Friedrichs inequalities for piecewise H^1 functions*, SIAM J. Numer. Anal. **41** (2003), no. 1, 306–324.

[9] H. Brezis, *Analyse fonctionnelle*, 2005 ed., Dunod, Paris, 1983.

[10] 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.

[11] ______, *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.

[12] 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.

[13] A. Buffa and C. Ortner, *Compact embeddings of broken Sobolev spaces and applications*, IMA J. Num. Anal. **29** (2009), 827–855.

[14] E. Burman and P. Zunino, *A domain decomposition method based on weighted interior penalties for advection-diffusion-reaction problems*, SIAM J. Numer. Anal. **44** (2006), no. 2, 1612–1638.

[15] B. Cockburn and C.-W. Shu, *The local discontinuous Galerkin finite element method for convection-diffusion systems*, SIAM J. Numer. Anal. **35** (1998), 2440–2463.

[16] C. Dawson, S. Sun, and M. F. Wheeler, *Compatible algorithms for coupled flow and transport*, Comput. Methods Appl. Mech. Engrg. **193** (2004), 2565–2580.

[17] D. A. Di Pietro and A. Ern, *Discrete functional analysis tools for discontinuous Galerkin methods with application to the incompressible Navier-Stokes equations*, Math. Comp. (2010), To appear. Preprint available at `http://hal.archives-ouvertes.fr/hal-00278925/fr/`.

[18] D. A. Di Pietro, A. Ern, and J.-L. Guermond, *Discontinuous Galerkin methods for anisotropic semi-definite diffusion with advection*, SIAM J. Numer. Anal. **46** (2008), no. 2, 805–831.

[19] A. Ern and J.-L. Guermond, *Theory and practice of finite elements*, Applied Mathematical Sciences, vol. 159, Springer-Verlag, New York, NY, 2004.

[20] A. Ern and J.-L. Guermond, *Discontinuous Galerkin methods for Friedrichs' systems. I. General theory*, SIAM J. Numer. Anal. **44** (2006), no. 2, 753–778.

[21] A Ern, A. F. Stephansen, and P. Zunino, *A discontinuous Galerkin method with weighted averages for advection-diffusion equations with locally small and anisotropic diffusivity*, IMA J. Num. Anal. **29** (2009), no. 2, 235–256.

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

[23] R. Eymard, T. Gallouët, and R. Herbin, *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 nonconforming meshes SUSHI: a scheme using stabilization and hybrid interfaces*, IMA J. Num. Anal. (2010), To appear, doi:10.1093/imanum/drn084.

[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] A. Lasis and E. Süli, *Poincaré-type inequalities for broken Sobolev spaces*, Tech. Report 03/10, Oxford University Computing Laboratory, Oxford, England, 2003.

[27] ______, *hp-version discontinuous Galerkin finite element method for semilinear parabolic problems*, SIAM J. Numer. Anal. **45** (2007), no. 4, 1544–1569.

[28] B. Rivière, M. F. Wheeler, and V. Girault, *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. Part I*, Comput. Geosci. **8** (1999), 337–360. MR 2001d:65145