LOCAL DISCONTINUOUS GALERKIN METHOD FOR DIFFUSION EQUATIONS WITH REDUCED STABILIZATION

E. BURMAN AND B. STAMM

Abstract. We extend the results on minimal stabilization of Burman and Stamm ("*Minimal stabilization of discontinuous Galerkin finite element methods for hyperbolic problems*", J. Sci. Comp., DOI: 10.1007/s10915-007-9149-5) to the case of the local discontinuous Galerkin methods on mixed form. The penalization term on the faces is relaxed to act only on a part of the polynomial spectrum. Stability in the form of a discrete inf-sup condition is proved and optimal convergence follows. Some numerical examples using high order approximation spaces illustrates the theory.

Local discontinuous Galerkin h-FEM, Interior penalty, Diffusion equation.

1. Introduction

Discontinuous Galerkin methods for scalar elliptic problems date back to the pioneering work of Douglas and Dupont (1976) [10], Baker (1977) [3], Wheeler (1978) [18] and Arnold (1982) [1]. Later the discontinuous Galerkin method was applied to the case of elliptic problems written as first order system by Bassi and Rebay (1997) [4] and the local discontinuous Galerkin (LDG-) method was proposed by Cockburn and Shu (1998) [9]. An essential point of a DG-method is that continuity is not imposed by the space and therefore some stabilizing mechanism is needed to impose continuity weakly. A number of approaches have been proposed. For a unified framework for discontinuous Galerkin methods for elliptic problems and a discussion of stabilization mechanism involved see the papers of Arnold and coworkers [2]. In the high order framework both the first order scalar hyperbolic problem and the diffusion equation was analysed by Houston and co-workers [13]. Finally the case of elliptic equations on mixed form and hyperbolic equations was given a unified treatment in the framework of Friedrich systems in the papers by Ern and Guermond [11], [12].

Recently it has been discussed how much the methods for elliptic problems on mixed form really need to be stabilized. Indeed most of the above mentioned references considered sufficient stabilization to obtain stability, however in many cases this is not necessary. There may be many reasons to try to diminish the amount of stabilization added. The computation of stabilization terms is costly and the added stability may perturb the local conservation properties of the scheme. Another reason for the numerical analysist is simple curiosity: what are the most basic stability mechanisms of DG-methods.

It was noticed in the paper by Sherwin and coworkers [17] that for certain configurations the discontinuous Galerkin method was stable in the sense that the discrete solution exists even with any stabilization. This phenomenon was also observed and given a detailed analysis by Marazzina in [14]. It was shown that it is enough to stabilise the solution on one boundary face. The convergence analysis however was restricted to the case of structured meshes. The ideas of minimal stabilization was then applied to the case of first order scalar hyperbolic problems by Burman and Stamm in the case of high order approximation [8]. In this work it was shown that it is enough to penalise the upper two thirds of the polynomial spectrum in order to obtain stability and optimal order graph-norm convergence. As a particular case stabilization of the tangential part of the gradient jump was advocated. The relaxation of the penalty allowed for a local mass conservation property that was independent of the penalty parameter. The same authors then made a detailed analysis of the scalar second order elliptic equation for the case of affine approximation [7]. It was shown in two or three space dimensions that both for the symmetric and the non symmetric formulation a boundary penalty term is sufficient to assure stability. Optimal convergence however requires either that the mesh satisfies a certain macro element property or that the space is enriched with non-conforming quadratic bubbles. If these conditions are not met a checker-board mode can appear that destroys convergence when the mesh is irregular or the data rough. In one space dimension a complete characterization of the stability properties for the symmetric DG-method for scalar elliptic problems was given by Burman and co-workers in [6].

In this note we will revisit the results of [8] and show how the analysis can be extended to the case of the local discontinuous Galerkin method for elliptic problems on mixed form on triangular meshes. Although we add stabilization on all faces it only affects a part of the polynomial spectrum. Since full control of the solution jumps is recovered by an inf-sup argument the method has optimal convergence order. This way the local conservation property of the scheme is independent of the penalty parameter.

2. Technical Results

2.1. **Definitions.** Let $\mathcal{K}$ be a subdivision of $\Omega \subset \mathbb{R}^2$ into non-overlapping triangles. For an element $\kappa \in \mathcal{K}$, h_κ denotes its diameter and set $h = \max_{\kappa\in\mathcal{K}} h_\kappa$. Assume that (i) $\mathcal{K}$ covers $\overline{\Omega}$ exactly, (ii) $\mathcal{K}$ does not contain any hanging nodes, and (iii) $\mathcal{K}$ is locally quasi-uniform in the sense that there exists a constant $\rho > 0$, independent of h, such that $\rho h_\kappa \leq \min_{\kappa'\in\mathcal{N}(\kappa)} h_{\kappa'}$, where $\mathcal{N}(\kappa)$ denotes the set of elements sharing at least one node with κ. Suppose that each $\kappa \in \mathcal{K}$ is an affine image of the reference element $\widehat{\kappa}$. Let $\mathcal{F}_i$ denote the set of interior faces (1-manifolds) of the mesh, i.e., the set of faces that are not included in the boundary $\partial\Omega$. The sets $\mathcal{F}_e$ denote the faces that are included in $\partial\Omega$ and denote $\mathcal{F} = \mathcal{F}_i \cup \mathcal{F}_e$. For $F \in \mathcal{F}$, h_F denotes its diameter. Let us denote $\tilde{h}$ the function defined such that $\tilde{h}|_{\overset{\circ}{\kappa}} = h_\kappa$ for all $\kappa \in \mathcal{K}$ and such that $\tilde{h}|_{\overset{\circ}{F}} = h_F$ for all $F \in \mathcal{F}$.

For a subset $R \subset \Omega$ or $R \subset \mathcal{F}$, $(\cdot,\cdot)_R$ denotes the $L^2(R)$–scalar product, $\|\cdot\|_R = (\cdot,\cdot)_R^{1/2}$ the corresponding norm, and $\|\cdot\|_{s,R}$ the $H^s(R)$–norm. For $s \geq 1$, let $H^s(\mathcal{K})$ be the space of piecewise Sobolev H^s–functions and denote its norm by $\|\cdot\|_{s,\mathcal{K}}$.

For $v \in H^1(\mathcal{K})$, $\boldsymbol{\tau} \in [H^1(\mathcal{K})]^2$ and an interior face $F = \kappa_1 \cap \kappa_2 \in \mathcal{F}_i$, where κ_1 and κ_2 are two distinct elements of $\mathcal{K}$ with respective outer normals $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$, the jump is defined by

$$[v] = v|_{\kappa_1}\boldsymbol{n}_1 + v|_{\kappa_2}\boldsymbol{n}_2, \qquad [\boldsymbol{\tau}] = \boldsymbol{\tau}|_{\kappa_1}\cdot\boldsymbol{n}_1 + \boldsymbol{\tau},|_{\kappa_2}\cdot\boldsymbol{n}_2$$

and the average by

$$\{v\} = \tfrac{1}{2}\left(v|_{\kappa_1} + v|_{\kappa_2}\right), \qquad \{\!\{\boldsymbol{\tau}\}\!\} = \tfrac{1}{2}\left(\boldsymbol{\tau}|_{\kappa_1} + \boldsymbol{\tau}|_{\kappa_2}\right).$$

On outer faces $F = \partial\kappa \cap \partial\Omega \in \mathcal{F}_e$ with outer normal $\boldsymbol{n}$, the jump and the average are defined as $[v] = v|_\kappa \boldsymbol{n}$ and $\{v\} = v|_\kappa$, resp. $[\boldsymbol{\tau}] = \boldsymbol{\tau}|_\kappa \cdot \boldsymbol{n}$ and $\{\!\{\boldsymbol{\tau}\}\!\} = \boldsymbol{\tau}|_\kappa$.

Further let $\boldsymbol{n}_F$ be an arbitray but fixed normal on $F \in \mathcal{F}$ and define $[v]_n = [v] \cdot \boldsymbol{n}_F$.

2.2. **Finite element spaces.** Let $p, \lambda \geq 0$ be two arbitrary integers and let κ be an arbitrary element of the mesh $\mathcal{K}$. Further let $\mathbb{P}_p(\kappa)$ be the space of polynomials of total degree p on κ and introduce the global discontinuous finite element space

$$V_h^p = \{\, v_h \in L^2(\Omega);\, v_h|_\kappa \in \mathbb{P}_p(\kappa),\ \forall \kappa \in \mathcal{K}\,\}. \tag{1}$$

Define the following polynomial space on $\partial\kappa$:

$$\mathbb{P}_\lambda(\partial\kappa) = \{v \in L^2(\partial\kappa) \,:\, v|_F \in \mathbb{P}_\lambda(F),\ \forall F \in \mathcal{F}(\partial\kappa)\},$$

where $\mathbb{P}_\lambda(F)$ is the usual one dimensional polynomial space of total degree λ on F and $\mathcal{F}(\partial\kappa)$ denotes the set of all faces of κ. Observe that there is no continuity required at the vertices of κ. On a global level we define

$$W_h^\lambda = \{v \in L^2(\mathcal{F}) \,:\, v|_F \in \mathbb{P}_\lambda(F),\ \forall F \in \mathcal{F}\,\}. \tag{2}$$

Let us further present some known results.

Lemma 2.1 (Trace inequality)**.** *Let $\boldsymbol{\tau}_h \in [V_h^p]^m$, $m \geq 1$, then there exists a constant $c_T > 0$, independent of the mesh size h, such that*

$$\|\{\!\{\boldsymbol{\tau}_h\}\!\}\|_{\mathcal{F}}^2 + \|[\boldsymbol{\tau}_h]\|_{\mathcal{F}}^2 \leq c_T\, \|\tilde{h}^{-\frac{1}{2}} \boldsymbol{\tau}_h\|_\Omega^2.$$

On the other hand if $\boldsymbol{\tau} \in [H^1(\mathcal{K})]^m$, then there exists a constant $c_T > 0$, independent of the mesh size h, such that

$$\|\{\!\{\boldsymbol{\tau}\}\!\}\|_{\mathcal{F}}^2 + \|[\boldsymbol{\tau}]\|_{\mathcal{F}}^2 \leq c_T\, \left(\|\tilde{h}^{-\frac{1}{2}} \boldsymbol{\tau}\|_\Omega^2 + |\tilde{h}^{\frac{1}{2}} \boldsymbol{\tau}|_{1,\mathcal{K}}^2\right).$$

Lemma 2.2 (Inverse inequality)**.** *Let $v_h \in V_h^p$, then there exists a constant $c > 0$, independent of the mesh size h, such that*

$$\|\nabla v_h\|_\Omega^2 \leq c\, \|\tilde{h}^{-1} v_h\|_\Omega^2.$$

2.3. **Projections.** Let $V_1(\widehat{\kappa}), V_2(\widehat{\kappa}) \subset \mathbb{P}_p(\widehat{\kappa})$, and $V_3(\partial\widehat{\kappa}) \subset \mathbb{P}_p(\partial\widehat{\kappa})$. Then, we address the questions for which spaces $V_1(\widehat{\kappa}), V_2(\widehat{\kappa}), V_3(\partial\widehat{\kappa})$ the following projection exists: Let $v \in L^2(\partial\widehat{\kappa})$ given, then find $\boldsymbol{\pi} \in [V_1(\widehat{\kappa})]^2$ such that

$$\int_{\widehat{\kappa}} \boldsymbol{\pi} \cdot \nabla w_h = 0 \qquad \forall w_h \in V_2(\widehat{\kappa}), \tag{3}$$

$$\int_{\partial\widehat{\kappa}} \boldsymbol{\pi} \cdot \boldsymbol{n} z_h = \int_{\partial\widehat{\kappa}} v z_h \qquad \forall z_h \in V_3(\partial\widehat{\kappa}). \tag{4}$$

Let us remark that the global variants of $V_1(\widehat{\kappa})$, $V_2(\widehat{\kappa})$ will be the spaces in which we will seek for approximations of the flux respectively the primal variable whereas $V_3(\partial\widehat{\kappa})$ defines the part of the spectrum of the jump which may be omitted in the stabilization. Thus, we would like to chose $V_1(\widehat{\kappa}) = V_2(\widehat{\kappa}) = \mathbb{P}_p(\widehat{\kappa})$ in order to ensure full approximability of both variables and have $V_3(\partial\widehat{\kappa})$ as rich as possible to reduce the stabilization to a minimum. Let us discuss several cases:

- $V_1(\widehat{\kappa}) = V_2(\widehat{\kappa}) = \mathbb{P}_p(\widehat{\kappa})$ and $V_3(\partial\widehat{\kappa}) = \mathbb{P}_\lambda(\partial\widehat{\kappa})$: In [8] the theoretical bound for λ of $0 \le \lambda \le \lfloor \frac{p+1}{3} \rfloor - 1$ for $p \ge 2$ has been showed for a scalar projection. It can be further generalized to a vectorial projection by considering componentwise the scalar projection. However this approach may be sub-optimal since in the vectorial case only the normal component of $\boldsymbol{\pi}$ in (3) has to be imposed. Indeed, computations on the reference element $\widehat{\kappa}$ show that the projection is well defined for $0 \le \lambda \le \lfloor \frac{2(p+1)}{3} \rfloor - 1$ and $p \ge 1$, see Appendix. Thus only the upper third of the polynomial spectrum of the jump has to be stabilized to get optimal convergence for the flux and primal variable.
- $V_1(\widehat{\kappa}) = \mathbb{P}_p(\widehat{\kappa})$, $V_2(\widehat{\kappa}) = \mathbb{P}_{p-1}(\widehat{\kappa})$ and $V_3(\partial\widehat{\kappa}) = \mathbb{P}_p(\partial\widehat{\kappa})$: In this case no stabilization is necessary, but optimal convergence for the primary variable can not be achieved. The existence of the projection in this case is assured by the fact that the BDM-space (Brezzi-Douglas-Marini, [5]) is included in $[V_1(\widehat{\kappa})]^2$.

Let $V_1, V_2 \subset V_h^p$ and $V_3 \subset W_h^p$ be the global versions of $V_1(\widehat{\kappa})$, $V_2(\widehat{\kappa})$ and $V_3(\partial\widehat{\kappa})$, i.e.,

$$\begin{aligned} V_i &= \{v_h \in L^2(\Omega) \,:\, v_h|_\kappa \in V_i(\kappa),\ \forall \kappa \in \mathcal{K}\} \quad i = 1,2, \\ V_3 &= \{v_h \in L^2(\mathcal{F}) \,:\, v_h|_{\partial\kappa} \in V_3(\partial\kappa),\ \forall \kappa \in \mathcal{K}\}. \end{aligned}$$

Proposition 2.1 (Global projection)**.** *Let $v \in L^2(\mathcal{F})$, then there exists a projection $\mathbf{\Pi}_h(v) \in [V_1]^2$ such that*

$$\int_\Omega \mathbf{\Pi}_h(v) \cdot \nabla w_h = 0 \qquad \forall\, w_h \in V_2, \tag{5}$$

$$\int_\mathcal{F} (\{\mathbf{\Pi}_h(v)\} \cdot \boldsymbol{n}_F - v) z_h = 0 \qquad \forall\, z_h \in V_3. \tag{6}$$

In addition, the projection satisfies the following stability properties

$$\|\{\mathbf{\Pi}_h(v)\}\|_\mathcal{F}^2 + \|[\mathbf{\Pi}_h(v)]\|_\mathcal{F}^2 \le c\,\|v\|_\mathcal{F}^2, \tag{7}$$

where $c > 0$ is a constant independent of the mesh size h.

Remark 2.3. The stability result follows directly from the local construction of the projection and from the equivalence of discrete norms on the reference triangle. We do not address the stability with respect to the polynomial degee p.

Remark 2.4. Another approach consists in directly considering the global projections without constructing the projection locally. This approach can allow a further reduction of the stabilization but goes beyond of the scope of this paper. For details of this approach for second order elliptic problems on scalar form see [7] and [6].

Corollary 2.5 (Inverse trace inequality)**.** *If $v_h \in W_h^p$ is single valued on each face, then there exists a constant $c > 0$, independent of the mesh size h, such that*

$$\|\mathbf{\Pi}_h(v_h)\|_\Omega^2 \le c\,\|\tilde{h}^{\frac{1}{2}} v_h\|_\mathcal{F}^2.$$

3. The discontinuous finite element method

We consider the following diffusion equation with Dirichlet boundary conditions:

Find $u : \Omega \to \mathbb{R}$ such that

$$\begin{cases} -\nabla\cdot(\varepsilon\nabla u) = f & \text{in } \Omega, \\ u_{|\partial\Omega} = g & \text{on } \partial\Omega, \end{cases} \tag{8}$$

with $\varepsilon > 0$, $f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega)$. Problem (8) is equivalent to the following system of first order differential equations:

Find $u : \Omega \to \mathbb{R}$ and $\boldsymbol{\sigma} : \Omega \to \mathbb{R}^d$ such that

$$\begin{cases} \boldsymbol{\sigma} - \varepsilon^{\frac{1}{2}}\nabla u = \mathbf{0} & \text{in } \Omega, \\ -\nabla\cdot(\varepsilon^{\frac{1}{2}}\boldsymbol{\sigma}) = f & \text{in } \Omega, \\ u_{|\partial\Omega} = g & \text{on } \partial\Omega. \end{cases} \tag{9}$$

Define by $P : L^2(\mathcal{F}) \to V_3$ the L^2-projection onto V_3 satisfying

$$\|Pv\|_{\mathcal{F}}^2 \le c_p\|v\|_{\mathcal{F}}^2 \qquad \text{and} \qquad \|(I-P)v\|_{\mathcal{F}}^2 \le c_p\|v\|_{\mathcal{F}}^2, \tag{10}$$

where $c_p > 0$ is a constant independent of h. Then, define the bilinear forms

$$\begin{aligned} a(\boldsymbol{\tau}_h, v_h) &= (\boldsymbol{\tau}_h, \nabla v_h)_\Omega - (\{\boldsymbol{\tau}_h\}, [v_h])_{\mathcal{F}}, \\ j(v_h, w_h) &= \gamma(\tilde{h}^{-1}\varepsilon\,(I-P)[v_h]_n, (I-P)[w_h]_n)_{\mathcal{F}}, \end{aligned}$$

for all $\boldsymbol{\tau}_h \in [V_1]^2$, $v_h, w_h \in V_2$ and where γ is a stabilisation parameter.

Let us define the discontinuous finite element space $\boldsymbol{V}_h = [V_1]^2 \times V_2$ being a discrete subspace of $\boldsymbol{V} = [H^1(\mathcal{K})]^2 \times H^1(\mathcal{K})$. Then, the discrete problem consists of seeking $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{V}_h$ such that

$$A(\boldsymbol{\sigma}_h, u_h; \boldsymbol{\tau}_h, v_h) = F(\boldsymbol{\tau}_h, v_h) \qquad \forall\,(\boldsymbol{\tau}_h, v_h) \in \boldsymbol{V}_h, \tag{11}$$

where

$$\begin{aligned} A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\rho}_h, w_h) &= (\boldsymbol{\tau}_h, \boldsymbol{\rho}_h)_\Omega - a(\varepsilon^{\frac{1}{2}}\boldsymbol{\rho}_h, v_h) + a(\varepsilon^{\frac{1}{2}}\boldsymbol{\tau}_h, w_h) + j(v_h, w_h), \\ F(\boldsymbol{\tau}_h, v_h) &= (f, v_h)_\Omega + (\boldsymbol{\tau}_h, \varepsilon^{\frac{1}{2}} g\boldsymbol{n})_{\mathcal{F}_e} + \gamma(\tilde{h}^{-1}\varepsilon\,(I-P)g, (I-P)v_h)_{\mathcal{F}_e}. \end{aligned}$$

Remark 3.1. Observe that if $W_h^0 \subset V_3$, then the above defined flux variable satisfies the following local mass conservation property, which is independent of the stabilization parameter and the primal variable u_h,

$$\int_{\partial\kappa} \{\boldsymbol{\sigma}_h\}\cdot\boldsymbol{n}_\kappa\, ds = \int_\kappa f\, dx$$

for all interior elements κ and where $\boldsymbol{n}_\kappa$ denote the exterior normal vector of κ.

Remark 3.2. If $W_h^\lambda \subset V_3$ with $\lambda \ge 0$ and using the Bramble-Hilbert lemma one easily shows that the $(I-P)$ operator may be replaced by a differential operator of order $\lambda + 1$ in the tangential directions of the face. In particular if $\lambda = 0$ we get

$$\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}(I-P)[v_h]_n\|_{\mathcal{F}} \le \|\tilde{h}^{\frac{1}{2}}\varepsilon^{\frac{1}{2}}[\nabla v_h]_t\|_{\mathcal{F}},$$

where here $[\nabla v]_t = \nabla v|_{\kappa_1} \times \boldsymbol{n}_1 + \nabla v|_{\kappa_2} \times \boldsymbol{n}_2$ is the tangential jump of the gradient. It follows that an equivalent stabilization term is obtained penalizing the jumps of certain derivatives, leading to a term that is no more complicated or expensive to compute that in the standard case. The following analysis holds in this case also with minor modifications.

Lemma 3.3. *Let* $(\boldsymbol{\tau}, v) \in \boldsymbol{V}$*, then*

$$a(\boldsymbol{\tau}, v) = -(\nabla\cdot\boldsymbol{\tau}, v)_\Omega + ([\boldsymbol{\tau}], \{v\})_{\mathcal{F}_i}.$$

Proof. Straight forward by integration by parts. □

Lemma 3.4 (Coercivity). *Let $(\boldsymbol{\tau}_h, v_h) \in \boldsymbol{V}_h$, then there exists a constant $c_L > 0$ such that*

$$c_L\, A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\tau}_h, v_h) \geq \|\boldsymbol{\tau}_h\|_\Omega^2 + \|\tilde{h}^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} (I-P)[v_h]_n\|_{\mathcal{F}}^2.$$

Proof. The definition of the bilinear form $A(\cdot;\cdot)$ yields

$$A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\tau}_h, v_h) = \|\boldsymbol{\tau}_h\|_\Omega^2 + \gamma \|\tilde{h}^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} (I-P)[v_h]_n\|_{\mathcal{F}}^2,$$

then taking $c_L = 1/\min(1,\gamma)$ completes the proof. □

Lemma 3.5 (Consistency). *Let $u \in H^1(\Omega)$ be the exact solution of problem* (8) *and let $(\boldsymbol{\sigma}_h, u_h)$ be the solution of* (11)*, then*

$$A(\varepsilon^{\frac{1}{2}} \nabla u - \boldsymbol{\sigma}_h, u - u_h; \boldsymbol{\tau}_h, v_h) = 0$$

for all $(\boldsymbol{\tau}_h, v_h) \in \boldsymbol{V}_h$.

Proof. Since $(\boldsymbol{\sigma}_h, u_h)$ is the discrete solution it satisfies

$$A(\boldsymbol{\sigma}_h, u_h; \boldsymbol{\tau}_h, v_h) = F(\boldsymbol{\tau}_h, v_h) \qquad \forall (\boldsymbol{\tau}_h, v_h) \in \boldsymbol{V}_h.$$

On the other hand since $u \in H^1(\Omega)$ we have $[u]|_F = \mathbf{0}$ and $[u]_n|_F = 0\ \forall F \in \mathcal{F}_i$. Additionally applying Lemma 3.3 yields

$$\begin{aligned} & A(\varepsilon^{\frac{1}{2}} \nabla u, u; \boldsymbol{\tau}_h, v_h) \\ = \quad & (\varepsilon^{\frac{1}{2}} \nabla u, \boldsymbol{\tau}_h)_\Omega - (\varepsilon^{\frac{1}{2}} \boldsymbol{\tau}_h, \nabla u)_\Omega + (\{\varepsilon^{\frac{1}{2}} \boldsymbol{\tau}_h\}, [u])_{\mathcal{F}} + a(\varepsilon \nabla u, v_h) + j(u, v_h) \\ = \quad & (-\nabla \cdot (\varepsilon \nabla u), v_h)_\Omega + (\{\boldsymbol{\tau}_h\}, [\varepsilon^{\frac{1}{2}} u])_{\mathcal{F}_e} + \gamma (\tilde{h}^{-1} \varepsilon (I-P)[u]_n, (I-P)[v_h]_n)_{\mathcal{F}_e} \\ = \quad & (f, v_h)_\Omega + (\boldsymbol{\tau}_h, \varepsilon^{\frac{1}{2}} g \boldsymbol{n})_{\mathcal{F}_e} + \gamma (\tilde{h}^{-1} \varepsilon\, (I-P) g, (I-P) v_h)_{\mathcal{F}_e}. \end{aligned}$$

Finally we conclude

$$A(\varepsilon^{\frac{1}{2}} \nabla u, u; \boldsymbol{\tau}_h, v_h) = F(\boldsymbol{\tau}_h, v_h).$$

□

4. Convergence Analysis

We denote by c a generic strictly positive constant independent of the mesh size h that might change at each occurence whereas constants with an index stay fixed. Further the following triple norm is defined for all $(\boldsymbol{\tau}, v) \in \boldsymbol{V}$ by

$$|||\boldsymbol{\tau}, v|||^2 = \|\boldsymbol{\tau}\|_\Omega^2 + \|\varepsilon^{\frac{1}{2}} \nabla v\|_\Omega^2 + \|\tilde{h}^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} [v]\|_{\mathcal{F}}^2.$$

Proposition 4.1 (Inf-Sup Condition). *Assume that the spaces V_1, V_2 and V_3 are chosen such that the projection defined by Proposition 2.1 exists. Then, there exists a constant $c > 0$, independent of the mesh size h, such that*

$$c\, |||\boldsymbol{\tau}_h, v_h||| \leq \sup_{(\boldsymbol{\tau}'_h, v'_h) \in \boldsymbol{V}_h^p} \frac{A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\tau}'_h, v'_h)}{|||\boldsymbol{\tau}'_h, v'_h|||} \qquad \forall (\boldsymbol{\tau}_h, v_h) \in \boldsymbol{V}_h$$

Proof. The proof consists of two lemmas, Lemma 4.1 and 4.2.

Lemma 4.1. *For all $(\boldsymbol{\tau}_h, v_h) \in \boldsymbol{V}_h$ there exists $(\boldsymbol{\tau}'_h, v'_h) \in \boldsymbol{V}_h$ and a constant $c > 0$ independent of the mesh size h such that*

$$c\, |||\boldsymbol{\tau}_h, v_h|||^2 \leq A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\tau}'_h, v'_h).$$

Lemma 4.2. *Fix $(\boldsymbol{\tau}_h, v_h) \in \boldsymbol{V}_h$ and let $(\boldsymbol{\tau}'_h, v'_h) \in \boldsymbol{V}_h$ be the functions defined in Lemma 4.1, then there exists a constant $c > 0$ independent of the mesh size h such that*

$$|||\boldsymbol{\tau}'_h, v'_h||| \leq c\, |||\boldsymbol{\tau}_h, v_h|||.$$

Combining these two lemmas leads to the result. Indeed for all $(\boldsymbol{\tau}_h, v_h) \in \boldsymbol{V}_h$ there exists $(\boldsymbol{\tau}'_h, v'_h) \in \boldsymbol{V}_h$ and $c > 0$ such that

$$A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\tau}'_h, v'_h) \geq c\, |||\boldsymbol{\tau}_h, v_h|||^2 \geq c\, |||\boldsymbol{\tau}_h, v_h|||\, |||\boldsymbol{\tau}'_h, v'_h|||.$$

□

Proof of Lemma 4.1. First fix $(\boldsymbol{\tau}_h, v_h) \in \boldsymbol{V}_h$ and define the vector functions $\boldsymbol{\rho}_h \in [V_2]^2$ and $\boldsymbol{w}_h \in [V_1]^2$ by

$$\boldsymbol{\rho}_h = -\varepsilon^{\frac{1}{2}} \nabla v_h \qquad \text{and} \qquad \boldsymbol{w}_h = \tilde{h}^{-1} \varepsilon^{\frac{1}{2}} \boldsymbol{\Pi}_h (P[v_h]_n)$$

where the projection $\boldsymbol{\Pi}_h$ is defined by Proposition 2.1. We proceed in three steps.
Step 1:
In the first step we show that there exists a constant $c_\rho > 0$ such that

$$\|\varepsilon^{\frac{1}{2}} \nabla v_h\|^2_\Omega \leq A(\boldsymbol{\tau}_h, v_h; 2\boldsymbol{\rho}_h + c_\rho \boldsymbol{\tau}_h, c_\rho v_h) + c_\rho\, \|\tilde{h}^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} P[v_h]_n\|^2_{\mathcal{F}}.$$

The definition of the bilinear form $A(\cdot, \cdot)$ yields

$$\begin{aligned} \|\varepsilon^{\frac{1}{2}} \nabla v_h\|^2_\Omega &= A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\rho}_h, 0) + (\boldsymbol{\tau}_h, \varepsilon \nabla v_h)_\Omega + (\{\varepsilon \nabla v_h\}, [v_h])_{\mathcal{F}} \\ &= A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\rho}_h, 0) + \mathcal{I}_1 + \mathcal{I}_2. \end{aligned}$$

Then using Young's inequality leads to

$$\mathcal{I}_1 \leq c\, \|\boldsymbol{\tau}_h\|^2_\Omega + \tfrac{1}{4} \|\varepsilon^{\frac{1}{2}} \nabla v_h\|^2_\Omega. \tag{12}$$

On the other side, using additionally the trace inequality, Lemma 2.1, yields

$$\mathcal{I}_2 \leq c\, \|\tilde{h}^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} [v_h]\|^2_{\mathcal{F}} + \tfrac{1}{4} \|\varepsilon^{\frac{1}{2}} \nabla v_h\|^2_\Omega. \tag{13}$$

Thus combining (12) and (13) and using coercivity, Lemma 3.4, yields

$$\begin{aligned} \tfrac{1}{2} \|\varepsilon^{\frac{1}{2}} \nabla v_h\|^2_\Omega &\leq A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\rho}_h, 0) + c \left(\|\boldsymbol{\tau}_h\|^2_\Omega + \|\tilde{h}^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} [v_h]_n\|^2_{\mathcal{F}} \right) \\ &\leq A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\rho}_h, 0) + c \left(A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\tau}_h, v_h) + \|\tilde{h}^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} P[v_h]_n\|^2_{\mathcal{F}} \right) \end{aligned}$$

and therefore exists a constant $c_\rho > 0$ such that

$$\|\varepsilon^{\frac{1}{2}} \nabla v_h\|^2_\Omega \leq A(\boldsymbol{\tau}_h, v_h; 2\boldsymbol{\rho}_h + c_\rho \boldsymbol{\tau}_h, c_\rho v_h) + c_\rho\, \|\tilde{h}^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} P[v_h]_n\|^2_{\mathcal{F}}.$$

Step 2:
In the second step we show that there exists a constant $c_w > 0$ such that

$$\|\tilde{h}^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} P[v_h]_n\|^2_{\mathcal{F}} \leq A(\boldsymbol{\tau}_h, v_h; 2\boldsymbol{w}_h + c_w \boldsymbol{\tau}_h, c_w v_h).$$

Firstly observe that by the definitions of the bilinear form $A(\cdot, \cdot)$ and of the projection $\boldsymbol{\Pi}_h$ we have

$$\begin{aligned} A(\boldsymbol{\tau}_h, v_h; \boldsymbol{w}_h, 0) &= (\boldsymbol{\tau}_h, \boldsymbol{w}_h)_\Omega - (\varepsilon^{\frac{1}{2}} \boldsymbol{w}_h, \nabla v_h) + (\{\varepsilon^{\frac{1}{2}} \boldsymbol{w}_h\}, [v_h])_{\mathcal{F}} \\ &= (\boldsymbol{\tau}_h, \boldsymbol{w}_h)_\Omega + (\{\varepsilon^{\frac{1}{2}} \boldsymbol{w}_h\} \cdot \boldsymbol{n}_F, [v_h]_n)_{\mathcal{F}} \end{aligned}$$

since $v_h \in V_2$. Secondly, again by the definition of the projection $\boldsymbol{\Pi}_h$ we may write

$$(\{\varepsilon^{\frac{1}{2}} \boldsymbol{w}_h\} \cdot \boldsymbol{n}_F, [v_h]_n)_{\mathcal{F}} = \|\tilde{h}^{-\frac{1}{2}} \varepsilon^{\frac{1}{2}} P[v_h]_n\|^2_{\mathcal{F}} + (\{\varepsilon^{\frac{1}{2}} \boldsymbol{w}_h\} \cdot \boldsymbol{n}_F, (I - P)[v_h]_n)_{\mathcal{F}}.$$

Therefore we have

$$\begin{aligned}\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}P[v_h]_n\|_{\mathcal{F}}^2 &= A(\boldsymbol{\tau}_h, v_h; \boldsymbol{w}_h, 0) - (\boldsymbol{\tau}_h, \boldsymbol{w}_h)_\Omega - (\{\varepsilon^{\frac{1}{2}}\boldsymbol{w}_h\}\cdot\boldsymbol{n}_F, (I-P)[v_h]_n)_{\mathcal{F}} \\ &= A(\boldsymbol{\tau}_h, v_h; \boldsymbol{w}_h, 0) - \mathcal{I}_1 - \mathcal{I}_2.\end{aligned}$$

Using Young's inequality and the inverse trace inequality, Corollary 2.5, leads to

$$|\mathcal{I}_1| \leq c\,\|\boldsymbol{\tau}_h\|_\Omega^2 + \tfrac{1}{4}\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}P[v_h]_n\|_{\mathcal{F}}^2. \tag{14}$$

On the other hand applying Young's inequality and the stability property of the projection $\boldsymbol{\Pi}_h$, (7), yields

$$|\mathcal{I}_2| \leq c\,\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}(I-P)[v_h]_n\|_{\mathcal{F}}^2 + \tfrac{1}{4}\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}P[v_h]_n\|_{\mathcal{F}}^2. \tag{15}$$

Thus, combining (14) and (15) and using coercivity, Lemma 3.4, yields

$$\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}P[v_h]_n\|_{\mathcal{F}}^2 \leq A(\boldsymbol{\tau}_h, v_h; 2\boldsymbol{w}_h + c_w\boldsymbol{\tau}_h, c_w v_h).$$

Step 3:

Now it only remains to combine coercivity and the results of **Step 1** and **Step 2**:

$$\begin{aligned}|||\boldsymbol{\tau}_h, v_h|||^2 &= \|\boldsymbol{\tau}_h\|_\Omega^2 + \|\varepsilon^{\frac{1}{2}}\nabla v_h\|_\Omega^2 + \|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}[v_h]_n\|_{\mathcal{F}}^2 \\ &\leq A(\boldsymbol{\tau}_h, v_h; c_L\boldsymbol{\tau}_h, c_L v_h) + \|\varepsilon^{\frac{1}{2}}\nabla v_h\|_\Omega^2 + \|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}P[v_h]_n\|_{\mathcal{F}}^2 \\ &\leq A(\boldsymbol{\tau}_h, v_h; (c_L+c_\rho)\boldsymbol{\tau}_h + 2\boldsymbol{\rho}_h, (c_L+c_\rho)v_h) + (1+c_\rho)\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}P[v_h]_n\|_{\mathcal{F}}^2 \\ &\leq A(\boldsymbol{\tau}_h, v_h; \boldsymbol{\tau}_h', v_h')\end{aligned}$$

where $\boldsymbol{\tau}_h' = (c_L + c_\rho + (1+c_\rho)c_w)\boldsymbol{\tau}_h + 2\boldsymbol{\rho}_h + 2(1+c_\rho)\boldsymbol{w}_h = c_1\boldsymbol{\tau}_h + 2\boldsymbol{\rho}_h + 2c_2\boldsymbol{w}_h$ and $v_h' = (c_L + c_\rho + (1+c_\rho)c_w)v_h = c_1 v_h$. □

Proof of Lemma 4.2. By definition of the triple norm:

$$|||\boldsymbol{\tau}_h', v_h'|||^2 = \|\boldsymbol{\tau}_h'\|_\Omega^2 + c_1^2\,\|\varepsilon^{\frac{1}{2}}\nabla v_h\|_\Omega^2 + c_1^2\,\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}[v_h]\|_{\mathcal{F}}^2.$$

For the first term use Corollary 2.5 and (10)

$$\begin{aligned}\|\boldsymbol{\tau}_h'\|_\Omega^2 &\leq c_1^2\|\boldsymbol{\tau}_h\|_\Omega^2 + 4\|\boldsymbol{\rho}_h\|_\Omega^2 + 4c_2^2\|\boldsymbol{w}_h\|_\Omega^2 \\ &\leq c_1^2\|\boldsymbol{\tau}_h\|_\Omega^2 + 4\|\varepsilon^{\frac{1}{2}}\nabla v_h\|_\Omega^2 + 4c_2^2 c_{IT}\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}P[v_h]_n\|_{\mathcal{F}}^2 \\ &\leq c_1^2\|\boldsymbol{\tau}_h\|_\Omega^2 + 4\|\varepsilon^{\frac{1}{2}}\nabla v_h\|_\Omega^2 + 4c_2^2 c_{IT} c_p\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}[v_h]\|_{\mathcal{F}}^2 \\ &\leq \max(c_1^2, 4, 4c_2^2 c_{IT} c_p)|||\boldsymbol{\tau}_h, v_h|||.\end{aligned}$$

Thus there exists a constant $c > 0$ such that

$$|||\boldsymbol{\tau}_h', v_h'||| \leq c\,|||\boldsymbol{\tau}_h, v_h|||.$$

□

Let us denote by $\boldsymbol{\pi}_h$ the piecewise vectorial L^2-projection $\boldsymbol{\pi}_h : [L^2(\Omega)]^2 \to [V_1]^2$ and by π_h its scalar version $\pi_h : L^2(\Omega) \to V_2$ satisfying the following approximation results

$$\|\boldsymbol{\pi}_h\boldsymbol{\tau} - \boldsymbol{\tau}\|_{k,\mathcal{K}} \leq ch^{s_1-k}|\boldsymbol{\tau}|_{s_1,\mathcal{K}} \quad k = 0,1 \tag{16}$$

$$\|\pi_h v - v\|_{k,\mathcal{K}} \leq ch^{s_2-k}|v|_{s_2,\mathcal{K}} \quad k = 0,1 \tag{17}$$

for all $\boldsymbol{\tau} \in [H^{r_1}(\mathcal{K})]^2$, $v \in H^{r_2}(\mathcal{K})$ and with $s_i = \min(p_i + 1, r_i)$ for some space specific p_i. Further let $\boldsymbol{\sigma}$ and u denote the exact solution of (9) and let $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{V}_h$ be the solution of (11), then define

$$\begin{aligned} \boldsymbol{\eta}_\sigma &= \boldsymbol{\sigma} - \boldsymbol{\pi}_h(\boldsymbol{\sigma}), \\ \eta_u &= u - \pi_h u, \end{aligned} \quad \text{and} \quad \begin{aligned} \boldsymbol{\xi}_\sigma &= \boldsymbol{\sigma}_h - \boldsymbol{\pi}_h(\boldsymbol{\sigma}), \\ \xi_u &= u_h - \pi_h u. \end{aligned} \tag{18}$$

To disburden the continuity proof for the bilinear form $A(\cdot,\cdot;\cdot,\cdot)$ we define a well scaled auxiliary norm:

$$]|\boldsymbol{\eta}_\sigma, \eta_u|[^2 = |||\boldsymbol{\eta}_\sigma, \eta_u|||^2 + \|\tilde{h}^{\frac{1}{2}}\{\boldsymbol{\eta}_\sigma\}\|_{\mathcal{F}}^2.$$

Proposition 4.2 (Continuity)**.** *Let $\boldsymbol{\eta}_\sigma$, η_u, $\boldsymbol{\xi}_\sigma$ and ξ_u be defined by* (18)*. Then*

$$A(\boldsymbol{\eta}_\sigma, \eta_u; \boldsymbol{\xi}_\sigma, \xi_u) \le]|\boldsymbol{\eta}_\sigma, \eta_u|[\; |||\boldsymbol{\xi}_\sigma, \xi_u|||.$$

Proof. Develop

$$A(\boldsymbol{\eta}_\sigma, \eta_u; \boldsymbol{\xi}_\sigma, \xi_u) = (\boldsymbol{\eta}_\sigma, \boldsymbol{\xi}_\sigma)_\Omega - a(\varepsilon^{\frac{1}{2}}\boldsymbol{\xi}_\sigma, \eta_u) + a(\varepsilon^{\frac{1}{2}}\boldsymbol{\eta}_\sigma, \xi_u) + j(\eta_u, \xi_u),$$

and apply the Cauchy-Schwarz inequality for the first term

$$(\boldsymbol{\eta}_\sigma, \boldsymbol{\xi}_\sigma)_\Omega \le \|\boldsymbol{\eta}_\sigma\|_\Omega \; \|\boldsymbol{\xi}_\sigma\|_\Omega \le]|\boldsymbol{\eta}_\sigma, \eta_u|[\; |||\boldsymbol{\xi}_\sigma, \xi_u|||.$$

Use the same argument for the last term

$$j(\eta_u, \xi_u) \le j(\eta_u, \eta_u)^{\frac{1}{2}} \; j(\xi_u, \xi_u)^{\frac{1}{2}} \le c \,]|\boldsymbol{\eta}_\sigma, \eta_u|[\; |||\boldsymbol{\xi}_\sigma, \xi_u|||,$$

where additionally the stability result (10) is used. For the remaining terms similar arguments are used. The trace inequality, Lemma 2.1, yields

$$\begin{aligned} -a(\varepsilon^{\frac{1}{2}}\boldsymbol{\xi}_\sigma, \eta_u) &= -(\boldsymbol{\xi}_\sigma, \varepsilon^{\frac{1}{2}}\nabla\eta_u)_\Omega + (\{\boldsymbol{\xi}_\sigma\}, \varepsilon^{\frac{1}{2}}[\eta_u])_{\mathcal{F}} \\ &\le \|\boldsymbol{\xi}_\sigma\|_\Omega \|\varepsilon^{\frac{1}{2}}\nabla\eta_u\|_\Omega + \|\tilde{h}^{\frac{1}{2}}\{\boldsymbol{\xi}_\sigma\}\|_{\mathcal{F}} \|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}[\eta_u]\|_{\mathcal{F}} \\ &\le \|\boldsymbol{\xi}_\sigma\|_\Omega \|\varepsilon^{\frac{1}{2}}\nabla\eta_u\|_\Omega + c\,\|\boldsymbol{\xi}_\sigma\|_\Omega \|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}[\eta_u]\|_{\mathcal{F}} \\ &\le c\,]|\boldsymbol{\eta}_\sigma, \eta_u|[\; |||\boldsymbol{\xi}_\sigma, \xi_u|||. \end{aligned}$$

In the same manner we develop

$$a(\varepsilon^{\frac{1}{2}}\boldsymbol{\eta}_\sigma, \xi_u) \le c\,]|\boldsymbol{\eta}_\sigma, \eta_u|[\; |||\boldsymbol{\xi}_\sigma, \xi_u|||,$$

and respecting all bounds yields

$$A(\boldsymbol{\eta}_\sigma, \eta_u; \boldsymbol{\xi}_\sigma, \xi_u) \le c\,]|\boldsymbol{\eta}_\sigma, \eta_u|[\; |||\boldsymbol{\xi}_\sigma, \xi_u|||.$$

□

Proposition 4.3 (Approximability)**.** *Let $\boldsymbol{\eta}_\sigma$, η_u, $\boldsymbol{\xi}_\sigma$ and ξ_u be defined by* (18) *and let V_1, V_2 such that the approximation results* (16), (17) *hold for some p_1, p_2. Assume that $u \in H^r(\mathcal{K})$, then for all $0 \le s_\sigma \le \min(p_1+1, r-1)$ and $0 \le s_u \le \min(p_2+1, r)$:*

$$\begin{aligned} |||\boldsymbol{\eta}_\sigma, \eta_u||| &\le c\left(h^{s_\sigma}|u|_{s_\sigma+1,\mathcal{K}} + h^{s_u-1}|u|_{s_u,\mathcal{K}}\right), \\ ]|\boldsymbol{\eta}_\sigma, \eta_u|[&\le c\left(h^{s_\sigma}|u|_{s_\sigma+1,\mathcal{K}} + h^{s_u-1}|u|_{s_u,\mathcal{K}}\right). \end{aligned}$$

Proof. Since $u \in H^r(\mathcal{K})$ it follows that $\boldsymbol{\sigma} \in [H^{r-1}(\mathcal{K})]^2$. Using the standard approximation properties of the L^2-projection, (16), (17), yields

$$\begin{aligned} \|\boldsymbol{\eta}_\sigma\|_\Omega &\le c\,h^{s_\sigma}|\boldsymbol{\sigma}|_{s_\sigma,\mathcal{K}} = c\,h^{s_\sigma}|u|_{s_\sigma+1,\mathcal{K}}, \\ |\boldsymbol{\eta}_\sigma|_{1,\mathcal{K}} &\le c\,h^{s_\sigma-1}|\boldsymbol{\sigma}|_{s_\sigma,\mathcal{K}} = c\,h^{s_\sigma-1}|u|_{s_\sigma+1,\mathcal{K}}, \end{aligned}$$

since $\boldsymbol{\sigma} = \varepsilon^{\frac{1}{2}} \nabla u$. In addition,

$$\begin{aligned} \|\eta_u\|_\Omega &\leq ch^{s_u}|u|_{s_u,\mathcal{K}}, \\ |\eta_u|_{1,\mathcal{K}} &= \leq ch^{s_u-1}|u|_{s_u,\mathcal{K}}. \end{aligned}$$

For the boundary terms, the trace inequality, Lemma 2.1, is applied:

$$\|\tilde{h}^{-\frac{1}{2}}[\eta_u]\|_{\mathcal{F}} \leq c\left(\|\tilde{h}^{-1}\eta_u\|_\Omega + |\eta_u|_{1,\mathcal{K}}\right) \leq c\, h^{s_u-1}|u|_{s_u,\mathcal{K}}.$$

In the same manner we develop

$$\|\tilde{h}^{\frac{1}{2}}\{\boldsymbol{\eta}_\sigma\}\|_{\mathcal{F}} \leq c\left(\|\boldsymbol{\eta}_\sigma\|_\Omega + |\tilde{h}\boldsymbol{\eta}_\sigma|_{1,\mathcal{K}}\right) \leq c\, h^{s_\sigma}|u|_{s_\sigma+1,\mathcal{K}}.$$

Recalling the definitions of the triple norm and the auxiliary norm yields

$$\begin{aligned} |||\boldsymbol{\eta}_\sigma, \eta_u||| &\leq c\left(h^{s_\sigma}|u|_{s_\sigma+1,\mathcal{K}} + h^{s_u-1}|u|_{s_u,\mathcal{K}}\right), \\ |]\boldsymbol{\eta}_\sigma, \eta_u[| &\leq c\left(h^{s_\sigma}|u|_{s_\sigma+1,\mathcal{K}} + h^{s_u-1}|u|_{s_u,\mathcal{K}}\right). \end{aligned}$$

□

Theorem 4.3 (Convergence)**.** *Assume that the spaces V_1, V_2 and V_3 are chosen such that the projection defined by Proposition 2.1 exists and that the approximation results* (16), (17) *hold for some p_1, p_2. Let $\boldsymbol{\sigma}$ and u denote the exact solution of* (9) *and let $\boldsymbol{\sigma}_h$ and u_h be the solution of* (11). *Assume that $u \in H^r(\mathcal{K}) \cap H^1(\Omega)$ with $r \geq 1$; then for all $0 \leq s_\sigma \leq \min(p_1+1, r-1)$ and $0 \leq s_u \leq \min(p_2+1, r)$*

$$|||\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, u - u_h||| \leq c\left(h^{s_\sigma}|u|_{s_\sigma+1,\mathcal{K}} + h^{s_u-1}|u|_{s_u,\mathcal{K}}\right)$$

where $c > 0$ is independent of the mesh size h.

Remark 4.4. If $V_1 = V_2 = V_h^p$, then choose $s = s_u = s_\sigma + 1$. Indeed, observe that if $p + 1 \geq r$, then

$$\min(p+1, r) = \min(p+1, r-1) + 1$$

and thus the largest admissible s_σ, s_u are the choice of $s = s_u = s_\sigma + 1$. On the other hand if $p + 1 \leq r - 1$, then

$$\min(p+1, r) = \min(p+1, r-1).$$

Thus $0 \leq s \leq \min(p+1, r)$ implies that $0 \leq s-1 \leq \min(p+1, r-1)$. As a consequence

$$|||\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, u - u_h||| \leq c\, h^{s-1}|u|_{s,\mathcal{K}}$$

for all $0 \leq s \leq \min(p+1, r)$.

Remark 4.5. Note that in the case of Remark 2.4, i.e. $V_1 = V_h^p$, $V_2 = V_h^{p-1}$, $V_3 = W_h^p$, the convergence of the primal variable is suboptimal for smooth problems. Indeed if $p \leq r - 2$ it follows that $s_u = p$ and $s_\sigma = p + 1$. Thus

$$|||\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, u - u_h||| \leq c\left(h^{p+1}|u|_{p+2,\mathcal{K}} + h^{p-1}|u|_{p,\mathcal{K}}\right) \leq c\, h^{p-1}.$$

Proof. Let $\boldsymbol{\eta}_\sigma$, η_u, $\boldsymbol{\xi}_\sigma$ and ξ_u be defined by (18). Use the triangle inequality

$$|||\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, u - u_h||| \leq |||\boldsymbol{\eta}_\sigma, \eta_u||| + |||\boldsymbol{\xi}_\sigma, \xi_u|||,$$

and by Proposition 4.3 the first term is bounded by

$$|||\boldsymbol{\eta}_\sigma, \eta_u||| \leq c\left(h^{s_\sigma}|u|_{s_\sigma+1,\mathcal{K}} + h^{s_u-1}|u|_{s_u,\mathcal{K}}\right), \tag{19}$$

for all $0 \le s_\sigma \le \min(p_1+1, r-1)$ and $0 \le s_u \le \min(p_2+1, r)$. For the second term use the inf-sup condition, the consistency and the continuity result, Proposition 4.1, Lemma 3.5 and Proposition 4.2,

$$\begin{aligned} |||\boldsymbol{\xi}_\sigma, \xi_u||| &\le c \sup_{(\boldsymbol{\tau}_h, v_h)\in \boldsymbol{V}_h^p} \frac{A(\boldsymbol{\xi}_\sigma, \xi_u; \boldsymbol{\tau}_h, v_h)}{|||\boldsymbol{\tau}_h, v_h|||} = c \sup_{(\boldsymbol{\tau}_h, v_h)\in \boldsymbol{V}_h^p} \frac{A(\boldsymbol{\eta}_\sigma, \eta_u; \boldsymbol{\tau}_h, v_h)}{|||\boldsymbol{\tau}_h, v_h|||} \\ &\le c \sup_{(\boldsymbol{\tau}_h, v_h)\in \boldsymbol{V}_h^p} \frac{|]\boldsymbol{\eta}_\sigma, \eta_u[|\ |||\boldsymbol{\tau}_h, v_h|||}{|||\boldsymbol{\tau}_h, v_h|||} = c\,|]\boldsymbol{\eta}_\sigma, \eta_u[| \\ &\le c\left(h^{s_\sigma}|u|_{s_\sigma+1,\mathcal{K}} + h^{s_u-1}|u|_{s_u,\mathcal{K}}\right). \end{aligned}$$

□

5. Numerical results

In this section we report some basic numerical results for the method with $V_1 = V_2 = V_h^p$, $V_3 = W_h^0$ and a stabilization term consisting of the jump of the tangential part of the gradient as presented in Remark 3.2. We compare our method to the existing local discontinuous Galerkin (LDG-) method for the problem (8) with smooth solution, i.e. we consider a domain $\Omega = (0,1)^2$ with $\varepsilon = 1$,

$$f(x,y) = 40\left(1 - \frac{(x-0.25)^2 + (y-0.25)^2}{0.1}\right)\exp\left(-\frac{(x-0.25)^2 + (y-0.25)^2}{0.1}\right)$$

and corresponding Dirichlet boundary condition such that the solution consists of

$$u(x,y) = \exp\left(-\frac{(x-0.25)^2 + (y-0.25)^2}{0.1}\right) \in C^\infty(\bar{\Omega}).$$

We considerer sequences of unstructured meshes for polynomial degrees $p = 1, \ldots, 7$. For the computations the C++ library *life*, a unified C++ implementation of the finite and spectral element methods in 1D, 2D and 3D, is used, see [15, 16].

Figure 1 shows the behavior of the approximations u_h and $\boldsymbol{\sigma}_h$ for h-refinement and fixed polynomial degree p. It shows similar behavior of the solutions of the here presented method and the LDG method.

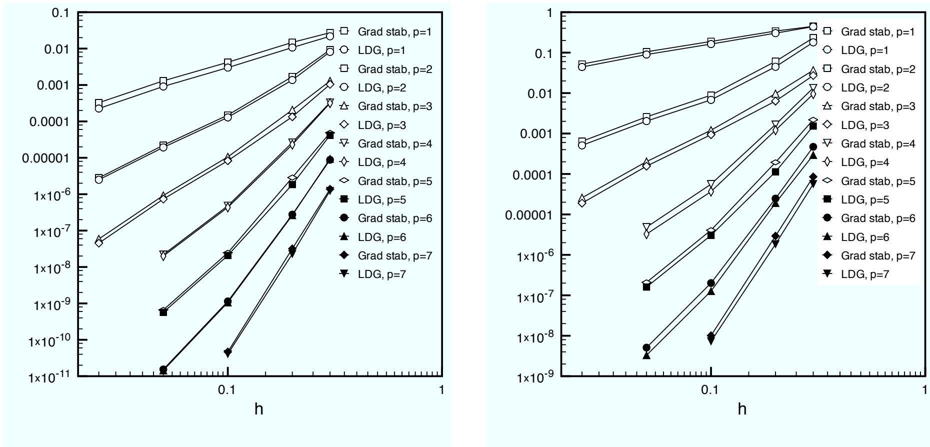
Figure 2 shows the behavior of the approximations u_h and $\boldsymbol{\sigma}_h$ for p-refinement and fixed mesh size h. Observe the exponential decay of the error for both methods.

Acknowledgments

The authors thank Thomas P. Wihler for discussions on the stabilization of LDG-methods and acknowledges the financial support provided through the Swiss National Science Foundation under grant 200021 − 113304.

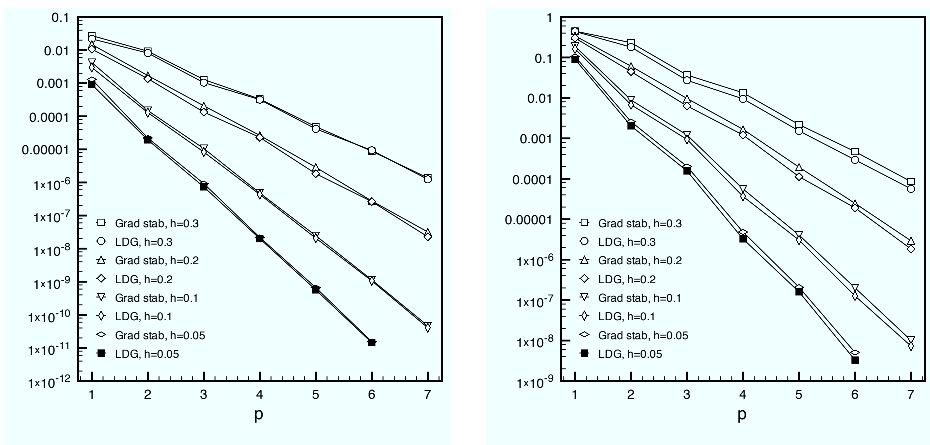
References

1. D. N. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal. **19** (1982), no. 4, 742–760. MR MR664882 (83f:65173)
2. D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal. **39** (2001/02), no. 5, 1749–1779 (electronic). MR MR1885715 (2002k:65183)
3. G. A. Baker, *Finite element methods for elliptic equations using nonconforming elements*, Math. Comp. **31** (1977), no. 137, 45–59. MR MR0431742 (55 #4737)
4. F. Bassi and S. Rebay, *A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations*, J. Comput. Phys. **131** (1997), no. 2, 267–279. MR MR1433934 (97m:76078)



(a) $\|u - u_h\|_{L^2(\Omega)}$ (b) $\|\nabla u - \sigma_h\|_{L^2(\Omega)}$

FIGURE 1. Accuracy for h-refinement and different polynomial orders p.



(a) $\|u - u_h\|_{L^2(\Omega)}$ (b) $\|\nabla u - \sigma_h\|_{L^2(\Omega)}$

FIGURE 2. Accuracy for p-refinement and different mesh sizes h.

5. F. Brezzi, J. Douglas, Jr., and L. D. Marini, *Two families of mixed finite elements for second order elliptic problems*, Numer. Math. **47** (1985), no. 2, 217–235. MR MR799685 (87g:65133)
6. E. Burman, Ern A., Mozolevski I., and Stamm B., *The symmetric discontinuous Galerkin method does not need stabilization in 1d for polynomial orders* $p \geq 2$, C. R. Acad. Sci. Paris Sér. I Math. (submitted).
7. E. Burman and B. Stamm, *Low order discontinuous Galerkin methods for second order elliptic problems.*, Tech. report, EPFL-IACS report 04, 2007.
8. ———, *Minimal stabilization of discontinuous Galerkin finite element methods for hyperbolic problems*, J. Sci. Comput. (DOI: 10.1007/s10915-007-9149-5).

9. B. Cockburn and C.-W. Shu, *The local discontinuous Galerkin method for time-dependent convection-diffusion systems*, SIAM J. Numer. Anal. **35** (1998), no. 6, 2440–2463 (electronic). MR MR1655854 (99j:65163)
10. J. Douglas and T. Dupont, *Interior penalty procedures for elliptic and parabolic Galerkin methods*, Computing methods in applied sciences (Second Internat. Sympos., Versailles, 1975), Springer, Berlin, 1976, pp. 207–216. Lecture Notes in Phys., Vol. 58.
11. 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 (electronic). MR MR2218968 (2007g:65107)
12. ———, *Discontinuous Galerkin methods for Friedrichs' systems. II. Second-order elliptic PDEs*, SIAM J. Numer. Anal. **44** (2006), no. 6, 2363–2388 (electronic). MR MR2272598 (2007k:65179)
13. P. Houston, Ch. Schwab, and E. Süli, *Discontinuous hp-finite element methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal. **39** (2002), no. 6, 2133–2163 (electronic). MR MR1897953 (2003d:65108)
14. D. Marazzina, *Mixed discontinuous Galerkin methods with minimal stabilization*, Numerical mathematics and advanced applications, Springer, Berlin, 2006, pp. 448–456. MR MR2303672
15. C. Prud'homme, *A domain specific embedded language in C++ for automatic differentiation, projection, integration and variational formulations*, Scientific Programming **14** (2006), no. 2, 81–110.
16. ———, *Life: Overview of a unified C++ implementation of the finite and spectral element methods in 1d, 2d and 3d*, Workshop On State-Of-The-Art In Scientific And Parallel Computing, Lecture Notes in Computer Science, Springer-Verlag, 2006, Accepted, p. 10.
17. S. J. Sherwin, R. M. Kirby, J. Peiró, R. L. Taylor, and O. C. Zienkiewicz, *On 2D elliptic discontinuous Galerkin methods*, Internat. J. Numer. Methods Engrg. **65** (2006), no. 5, 752–784. MR MR2195978 (2007b:65127)
18. M. F. Wheeler, *An elliptic collocation-finite element method with interior penalties*, SIAM J. Numer. Anal. **15** (1978), no. 1, 152–161. MR MR0471383 (57 #11117)

Appendix

The matrix that defines the projection in the case of $V_1(\widehat{\kappa}) = V_2(\widehat{\kappa}) = \mathbb{P}_p(\widehat{\kappa})$ and $V_3(\partial\widehat{\kappa}) = \mathbb{P}_\lambda(\partial\widehat{\kappa})$ has been computed using a Matlab code. In order for the projection to exist two criterias has to be satisfied. Firstly the number of columns should be at least the number of lines in order to have more degrees of freedom than conditions and the rank of the matrix has to be equal to the number of lines in order to ensure existence of at least one projection. The following table shows the largest possible λ for each p such that the projection exists, noted as $\lambda^\star$:

p	1	2	3	4	5	6	7	8
$\lambda^\star$	0	1	1	2	3	3	4	5

Observe that $\lambda^\star$ behaves as $\lfloor \frac{2(p+1)}{3} \rfloor - 1$. Another approach consists of stabilizing the lower modes of the polynomial spectrum of the jump which implies that $V_3(\partial\widehat{\kappa}) = \mathbb{P}_p(\partial\widehat{\kappa}) \backslash \mathbb{P}_\lambda(\partial\widehat{\kappa})$. The following table shows the smallest possible λ for each p such that the projection exists:

p	1	2	3	4	5	6	7	8
$\lambda^\star$	1	2	2	3	3	4	4	5

Observe that $\lambda^\star$ behaves as $\lfloor \frac{p}{2} \rfloor + 1$.

Institute of Analysis and Scientific Computing, Swiss Institute of Technology, Lausanne, CH-1015, Switzerland
E-mail address: erik.burman@epfl.ch

Institute of Analysis and Scientific Computing, Swiss Institute of Technology, Lausanne, CH-1015, Switzerland
E-mail address: benjamin.stamm@epfl.ch