COMMUNICATIONS IN COMPUTATIONAL PHYSICS
Vol. 5, No. 2-4, pp. 498-514

Commun. Comput. Phys.
February 2009

Local Discontinuous Galerkin Method with Reduced Stabilization for Diffusion Equations

E. Burman[1] and B. Stamm[2,*]

[1] *Department of Mathematics, Mantell Building, University of Sussex, Falmer, Brighton, BN1 9RF, UK.*
[2] *Institute of Analysis and Scientific Computing, Swiss Institute of Technology, Lausanne, CH-1015, Switzerland.*

Received 2 October 2007; Accepted (in revised version) 8 May 2008

Available online 1 August 2008

Abstract. We extend the results on minimal stabilization of Burman and Stamm [J. Sci. Comp., 33 (2007), pp. 183-208] 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 illustrate the theory.

AMS subject classifications: 65N30, 35F15

Key words: 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 [15], Baker [3], Wheeler [24] and Arnold [1]. Later the discontinuous Galerkin method was applied to the case of elliptic problems written as first order system by Bassi and Rebay [4] and the local discontinuous Galerkin (LDG-) method was proposed by Cockburn and Shu [14]. In the high order framework the LDG-method was analyzed in [10, 11, 13, 20]. 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 mechanisms involved see the papers of Arnold and coworkers [2]. In

*Corresponding author. *Email addresses:* `E.N.Burman@sussex.ac.uk` (E. Burman), `benjamin.stamm@epfl.ch` (B. Stamm)

the high order framework both the first order scalar hyperbolic problem and the diffusion equation were analyzed by Houston and co-workers [18]. Finally the case of elliptic equations in mixed form and hyperbolic equations was given a unified treatment in the framework of Friedrich systems in the papers by Ern and Guermond [16, 17].

Recently it has been discussed how much the methods for elliptic problems in 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 appears 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 analyst is simple curiosity: what are the most basic stability mechanisms of DG-methods?

It was noticed in the paper by Sherwin and coworkers [23] that for certain configurations the discontinuous Galerkin method appears to be stable in the sense that the discrete solution exists even without any stabilization. This phenomenon was also observed and given a detailed analysis by Marazzina in [19] in the case of shape regular quadrilateral meshes. It was shown that it is enough to stabilize the solution on one boundary face. The convergence analysis however was restricted to the case of structured meshes. Cockburn and Dong introduce in [12] an artificial wind to stabilize the scheme using the upwind technique and drop the penalty term. The idea 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 that work it was shown that it is enough to penalize 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 ensure existence of the solution. 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, see also [9]. If these conditions are not met a checkerboard 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 in 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 Preliminary results

2.1 Definitions

Let $\Omega \subset \mathbb{R}^2$ be an open, bounded and convex polygon with boundary $\partial\Omega$. We consider the following diffusion equation with Dirichlet boundary conditions:

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

$$\begin{cases} -\nabla\cdot(\varepsilon\nabla u) = f & \text{in } \Omega, \\ u_{|\partial\Omega} = g & \text{on } \partial\Omega, \end{cases} \tag{2.1}$$

with $\varepsilon \in \mathbb{R}$ s.t. $\varepsilon > 0$, $f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega)$. Problem (2.1) 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

$$\begin{cases} \boldsymbol{\sigma} - \varepsilon^{\frac{1}{2}}\nabla u = \mathbf{0} & \text{in } \Omega, \\ -\nabla\cdot(\varepsilon^{\frac{1}{2}}\boldsymbol{\sigma}) = f & \text{in } \Omega, \\ u_{|\partial\Omega} = g & \text{on } \partial\Omega. \end{cases} \tag{2.2}$$

Let $\mathcal{K}$ be a subdivision of $\Omega \subset \mathbb{R}^2$ into non-overlapping triangles. The triangles κ are closed and their interior denoted by $\overset{\circ}{\kappa}$. For an element $\kappa \in \mathcal{K}$, h_κ denotes its diameter and set $h = \max_{\kappa\in\mathcal{K}} h_\kappa$. Assume that

- $\mathcal{K}$ covers $\overline{\Omega}$ exactly;
- $\mathcal{K}$ does not contain any hanging nodes;
- $\mathcal{K}$ is shape-regular.

Suppose that each $\kappa \in \mathcal{K}$ is an affine image of the reference element $\widehat{\kappa}$, i.e. for each element $\kappa \subset \mathcal{K}$ there exists an affine transformation $\boldsymbol{T}_\kappa : \widehat{\kappa} \to \kappa$. 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 set $\mathcal{F}_e$ denotes 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 by $\tilde{h}$ 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 $F \in \mathcal{F}$.

Denote by Γ the skeleton of the mesh defined by $\Gamma = \{\boldsymbol{x} \in \overline{\Omega} : \exists F \in \mathcal{F} \text{ s.t. } \boldsymbol{x} \in F\}$. For a non-empty subdomain $R \subset \Omega$ or $R \subset \Gamma$, $(\cdot,\cdot)_R$ 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 scalar product, norm and semi-norm respectively by $(\cdot,\cdot)_{s,\mathcal{K}}$, $\|\cdot\|_{s,\mathcal{K}}$ and $|\cdot|_{s,\mathcal{K}}$. For $s=0$ the index s is dropped.

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

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

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 arbitrary but fixed normal on $F \in \mathcal{F}$ and define $[v]_n = [v] \cdot \boldsymbol{n}_F$. Observe that

$$\|[v]\|_{\mathcal{F}} = \|[v]_n\|_{\mathcal{F}}. \tag{2.3}$$

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

$$V_h^p = \{v_h \in L^2(\Omega) : v_h|_\kappa \in \mathbb{P}_p(\kappa), \ \forall \kappa \in \mathcal{K}\}. \tag{2.4}$$

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

$$W_h^\lambda = \{v \in L^2(\Gamma) : v|_F \in \mathbb{P}_\lambda(F), \ \forall F \in \mathcal{F}\}. \tag{2.5}$$

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 holds exists a constant $c_T > 0$, independent of the mesh size h, such that*

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

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

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

Lemma 2.2 (Inverse inequality). *Let $v_h \in V_h^p$. Then there holds*

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

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 question for which spaces $V_1(\widehat{\kappa}),V_2(\widehat{\kappa}),V_3(\partial\widehat{\kappa})$ the following projection exists: Let $v\in L^2(\partial\widehat{\kappa})$ be given, then find $\hat{\boldsymbol{\pi}}\in[V_1(\widehat{\kappa})]^2$ such that

$$\int_{\widehat{\kappa}}\hat{\boldsymbol{\pi}}\cdot\nabla w_h\,d\hat{x}=0 \qquad \forall w_h\in V_2(\widehat{\kappa}), \tag{2.6}$$

$$\int_{\partial\widehat{\kappa}}\hat{\boldsymbol{\pi}}\cdot\boldsymbol{n}_{\widehat{\kappa}}z_h\,d\hat{s}=\int_{\partial\widehat{\kappa}}vz_h\,d\hat{s} \qquad \forall z_h\in V_3(\partial\widehat{\kappa}), \tag{2.7}$$

where $\boldsymbol{n}_{\widehat{\kappa}}$ denotes the outer unit normal of $\widehat{\kappa}$. 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 approximations of the flux and primal variables (σ,u) of problem (2.2), whereas $V_3(\partial\widehat{\kappa})$ defines the part of the spectrum of the jump of u 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. Several choices of $V_1(\widehat{\kappa})$, $V_2(\widehat{\kappa})$ and $V_3(\partial\widehat{\kappa})$ will be discussed in Section 4.1.

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.,

$$V_i=\{v_h\in L^2(\Omega):v_h|_\kappa\circ\boldsymbol{T}_\kappa\in V_i(\widehat{\kappa}),\ \forall\kappa\in\mathcal{K}\} \quad i=1,2,$$
$$V_3=\{v_h\in L^2(\Gamma):v_h|_{\partial\kappa}\circ\boldsymbol{T}_\kappa\in V_3(\partial\widehat{\kappa}),\ \forall\kappa\in\mathcal{K}\}.$$

Proposition 2.1 (Global projection)**.** Assume that the local projection defined by (2.6)-(2.7) is well posed. Let $v\in L^2(\Gamma)$, then there exists a projection $\boldsymbol{\Pi}_h(v)\in[V_1]^2$ such that

$$\int_{\Omega}\boldsymbol{\Pi}_h(v)\cdot\nabla w_h\,dx=0 \qquad \forall w_h\in V_2, \tag{2.8}$$

$$\int_{\mathcal{F}}(\{\boldsymbol{\Pi}_h(v)\}\cdot\boldsymbol{n}_F-v)z_h\,ds=0 \qquad \forall z_h\in V_3. \tag{2.9}$$

In addition, the projection satisfies the following stability properties

$$\|\{\boldsymbol{\Pi}_h(v)\}\|_{\mathcal{F}}^2+\|[\boldsymbol{\Pi}_h(v)]\|_{\mathcal{F}}^2\le c\|v\|_{\mathcal{F}}^2, \tag{2.10}$$

and

$$\|\boldsymbol{\Pi}_h(v_h)\|_{\mathcal{K}}^2\le c_{IT}\|\tilde{h}^{\frac{1}{2}}v_h\|_{\mathcal{F}}^2. \tag{2.11}$$

Remark 2.1. 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 degree p.

Remark 2.2. 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 the scope of this paper. For details of this approach see [12] and for second order elliptic problems in scalar form see [6,7,9].

3 The discontinuous finite element method

Define by $P: L^2(\Gamma) \to V_3$ the L^2-projection onto V_3 satisfying

$$\|Pv\|_{\mathcal{F}}^2 \le \|v\|_{\mathcal{F}}^2 \qquad \text{and} \qquad \|(I-P)v\|_{\mathcal{F}}^2 \le \|v\|_{\mathcal{F}}^2, \tag{3.1}$$

for all $v \in L^2(\Gamma)$. Then, define the bilinear forms

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

for all $\boldsymbol{\tau}_h \in [V_1]^2$, $v_h, w_h \in V_2$ and where $\gamma > 0$ is a stabilization parameter independent of h.

Let us define the discontinuous finite element space $\boldsymbol{V}_h = [V_1]^2 \times V_2$ as a finite dimensional 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

$$A(\boldsymbol{\sigma}_h, u_h; \boldsymbol{\tau}_h, v_h) = F(\boldsymbol{\tau}_h, v_h) \qquad \forall (\boldsymbol{\tau}_h, v_h) \in \boldsymbol{V}_h, \tag{3.2}$$

where

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

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

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,

$$\int_{\partial\kappa} \{\varepsilon^{\frac{1}{2}}\boldsymbol{\sigma}_h\} \cdot \boldsymbol{n}_\kappa \, ds = \int_\kappa f \, dx$$

for all interior elements κ and where $\boldsymbol{n}_\kappa$ denotes 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 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

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

where here $[\nabla v_h]_t = \nabla v_h|_{\kappa_1} \times \boldsymbol{n}_1 + \nabla v_h|_{\kappa_2} \times \boldsymbol{n}_2$ 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 than in the standard case. The following analysis holds in this case also with minor modifications.

Lemma 3.1. *Let* $(\boldsymbol{\tau},v)\in \boldsymbol{V}$*. Then*

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

Proof. Straightforward by integration by parts. □

Lemma 3.2 (Coercivity)**.** *Let* $(\boldsymbol{\tau}_h,v_h)\in \boldsymbol{V}_h$*, then there exists a constant* $c_L>0$ *independent of the mesh size* h *such that*

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

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

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

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

Lemma 3.3 (Consistency)**.** *Let* $u\in H^2(\Omega)$ *be the exact solution of problem* (2.1) *and let* $(\boldsymbol{\sigma}_h,u_h)$ *be the solution of* (3.2)*. 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 $u\in H^2(\Omega)$ we have $[u]|_F=\boldsymbol{0}$ for all $F\in\mathcal{F}_i$. Additionally applying Lemma 3.1 yields for all $(\boldsymbol{\tau}_h,v_h)\in \boldsymbol{V}_h$

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

Finally we conclude that

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

This completes the proof of the lemma. □

4 Convergence analysis

We denote by c a generic strictly positive constant independent of the mesh size h (but possibly dependent of the diffusion coefficient ε) that may change at each occurrence whereas constants with an index stay fixed. Further we define the following triple norm for all $(\boldsymbol{\tau},v)\in\boldsymbol{V}$:

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

Proposition 4.1 (Inf-Sup Condition)**.** Assume that the spaces V_1, V_2 and V_3 are chosen such that the projection defined in Proposition 2.1 exists and that $\nabla V_2\subseteq[V_1]^2$. Then, there exists a constant $c>0$, independent of the mesh size h, such that

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

The proof consists of two lemmas, Lemmas 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 $\boldsymbol{\rho}_h,\boldsymbol{w}_h\in[V_1]^2$ 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

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

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

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

Then using Young's inequality leads to

$$\mathcal{I}_1 \le c\|\boldsymbol{\tau}_h\|_{\mathcal{K}}^2 + \tfrac{1}{4}\|\varepsilon^{\frac{1}{2}}\nabla v_h\|_{\mathcal{K}}^2. \tag{4.1}$$

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

$$\mathcal{I}_2 \le c\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}[v_h]\|_{\mathcal{F}}^2 + \tfrac{1}{4}\|\varepsilon^{\frac{1}{2}}\nabla v_h\|_{\mathcal{K}}^2. \tag{4.2}$$

Thus combining (4.1) and (4.2) and using coercivity, Lemma 3.2, yield

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

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

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

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

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

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

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

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

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

by (2.3). Therefore we have

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

Using Young's inequality and the inverse trace inequality (2.11) leads to

$$|\mathcal{I}_1| \le c\|\boldsymbol{\tau}_h\|_{\mathcal{K}}^2 + \tfrac{1}{4}\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}P[v_h]\|_{\mathcal{F}}^2. \tag{4.3}$$

On the other hand applying Young's inequality and the stability property of the projection $\boldsymbol{\Pi}_h$, (2.10), yields

$$|\mathcal{I}_2| \le c\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}(I-P)[v_h]\|_{\mathcal{F}}^2 + \tfrac{1}{4}\|\tilde{h}^{-\frac{1}{2}}\varepsilon^{\frac{1}{2}}P[v_h]\|_{\mathcal{F}}^2. \tag{4.4}$$

Thus, combining (4.3) and (4.4) and using coercivity, Lemma 3.2, yield

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

Step 3: Now it only remains to combine coercivity and the results of **Step 1** and **Step 2**:

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

where

$$\begin{aligned}
&\boldsymbol{\tau}'_h = (c_L+c_\rho+(1+c_\rho)c_w)\boldsymbol{\tau}_h + 2\boldsymbol{\rho}_h + 2(1+c_\rho)\boldsymbol{w}_h = c_1\boldsymbol{\tau}_h + 2\boldsymbol{\rho}_h + 2c_2\boldsymbol{w}_h, \\
&v'_h = (c_L+c_\rho+(1+c_\rho)c_w)v_h = c_1 v_h.
\end{aligned}$$

This completes the proof of Lemma 4.1. □

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

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

For the first term use (2.11) and (3.1)

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

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

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

This completes the proof of Lemma 4.2. □

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

$$\|\boldsymbol{\pi}_h\boldsymbol{\tau}-\boldsymbol{\tau}\|_{k,\mathcal{K}} \le c|\tilde{h}^{s_1-k}\boldsymbol{\tau}|_{s_1,\mathcal{K}}, \quad k=0,1, \tag{4.5}$$

$$\|\pi_h v-v\|_{k,\mathcal{K}} \le c|\tilde{h}^{s_2-k}v|_{s_2,\mathcal{K}}, \quad k=0,1 \tag{4.6}$$

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 specifying the polynomial order of the discrete space. Further let $\boldsymbol{\sigma}$ and u denote the exact solution of (2.2) and let $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{V}_h$ be the solution of (3.2), then define

$$\begin{aligned} \boldsymbol{\eta}_\sigma &= \boldsymbol{\sigma} - \boldsymbol{\pi}_h \boldsymbol{\sigma}, \\ \eta_u &= u - \pi_h u, \end{aligned} \qquad \text{and} \qquad \begin{aligned} \boldsymbol{\xi}_\sigma &= \boldsymbol{\sigma}_h - \boldsymbol{\pi}_h \boldsymbol{\sigma}, \\ \xi_u &= u_h - \pi_h u. \end{aligned} \tag{4.7}$$

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 (4.7). Then

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

Proof. Develop

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

and apply the Cauchy-Schwarz inequality for the first term

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

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 (3.1) is used. For the remaining terms similar arguments are used. The trace inequality, Lemma 2.1, yields

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

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

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

This ends the proof. $\square$

Proposition 4.3 (Approximability)**.** Let $\boldsymbol{\eta}_\sigma$, η_u, $\boldsymbol{\xi}_\sigma$ and ξ_u be defined by (4.7) and let V_1, V_2 such that the approximation results (4.5), (4.6) hold for some p_1, p_2. Assume that $u \in H^r(\mathcal{K})$ with some $r \geq 2$. Then for all $0 \leq s_\sigma \leq \min(p_1+1, r-1)$ and $0 \leq s_u \leq \min(p_2+1, r)$:

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

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, (4.5), (4.6), yields

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

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

$$\|\eta_u\|_{\mathcal{K}} \leq c |\tilde{h}^{s_u} u|_{s_u,\mathcal{K}}, \quad |\eta_u|_{1,\mathcal{K}} \leq c |\tilde{h}^{s_u-1} u|_{s_u,\mathcal{K}}.$$

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

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

In the same manner we develop

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

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

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

which gives the desired inequality. $\square$

Theorem 4.1 (Convergence)**.** *Assume that the spaces V_1, V_2 and V_3 are chosen such that the projection defined by Proposition 2.1 exists, that $\nabla V_2 \subseteq [V_1]^2$ and that the approximation results* (4.5), (4.6) *hold for some p_1, p_2. Let $\boldsymbol{\sigma}$ and u denote the exact solution of* (2.2) *and let $\boldsymbol{\sigma}_h$ and u_h be the solution of* (3.2). *Assume that $u \in H^r(\mathcal{K}) \cap H^2(\Omega)$ with $r \geq 2$; then for all $0 \leq s_\sigma \leq$* $\min(p_1+1, r-1)$ *and* $0 \leq s_u \leq \min(p_2+1, r)$

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

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

Remark 4.1. 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

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

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

Remark 4.2. Note that in the case of $V_1 = V_h^p$, $V_2 = V_h^{p-1}$, $V_3 = W_h^{p-1}$, the convergence is suboptimal for smooth problems. Indeed if $p \leq r-2$ it follows that $s_u = p$ and $s_\sigma = p+1$. Thus

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

Proof of Theorem 4.1. Let $\boldsymbol{\eta}_\sigma$, η_u, $\boldsymbol{\xi}_\sigma$ and ξ_u be defined by (4.7). 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

$$|||\boldsymbol{\eta}_\sigma, \eta_u||| \leq c \left(|\tilde{h}^{s_\sigma} u|_{s_\sigma+1,\mathcal{K}} + |\tilde{h}^{s_u-1} u|_{s_u,\mathcal{K}} \right), \tag{4.8}$$

for all $0 \leq s_\sigma \leq \min(p_1+1, r-1)$ and $0 \leq s_u \leq \min(p_2+1, r)$. For the second term use the inf-sup condition, the consistency and the continuity result, Proposition 4.1, Lemma 3.3 and Proposition 4.2,

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

This completes the proof of Theorem 4.1. $\square$

4.1 Existence of local projections

The above analysis relies on the existence of the local projection defined by (2.6)-(2.7). Let us present some cases where the projection exists.

• $V_1(\widehat{\kappa})=V_2(\widehat{\kappa})=\mathbb{P}_p(\widehat{\kappa})$ 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 $p\ge2$ has been shown for a scalar projection. It can be further generalized to a vectorial projection by considering the scalar projection componentwise. However this approach may be suboptimal since in the vectorial case only the normal component of $\boldsymbol{\pi}$ in (2.6) 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 $p\ge1$. 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

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.

• $V_1(\widehat{\kappa})=V_2(\widehat{\kappa})=\mathbb{P}_p(\widehat{\kappa})$ and $V_3(\partial\widehat{\kappa})=\mathbb{P}_p(\partial\widehat{\kappa})\backslash\mathbb{P}_\lambda(\partial\widehat{\kappa})$: This approach consists of stabilizing the lower modes of the polynomial spectrum of the jump. 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 $\lfloor\frac{p}{2}\rfloor+1$.

• $V_1(\widehat{\kappa})=\mathbb{P}_p(\widehat{\kappa})$, $V_2(\widehat{\kappa})=\mathbb{P}_{p-1}(\widehat{\kappa})$ and $V_3(\partial\widehat{\kappa})=\mathbb{P}_{p-1}(\partial\widehat{\kappa})$: In this case no stabilization is necessary, but optimal convergence for the primary variable is not obtained without reconstruction. The existence of the projection in this case is proven in [5, Lemma 2.1].

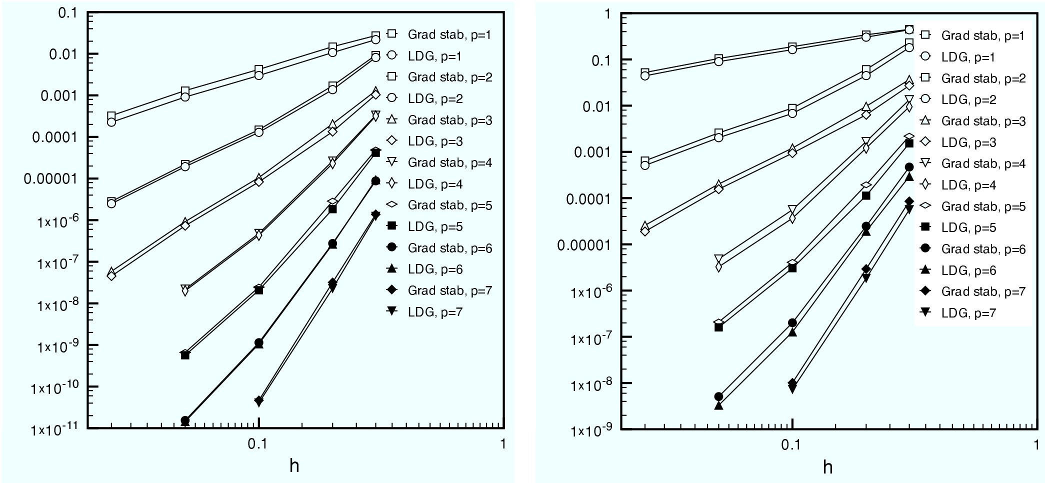
5 Numerical results

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 local discontinuous Galerkin (LDG-) method of [14] for the problem (2.1) with smooth solution, i.e. we consider a domain $\Omega=(0,1)^2$ with $\varepsilon=1$,

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

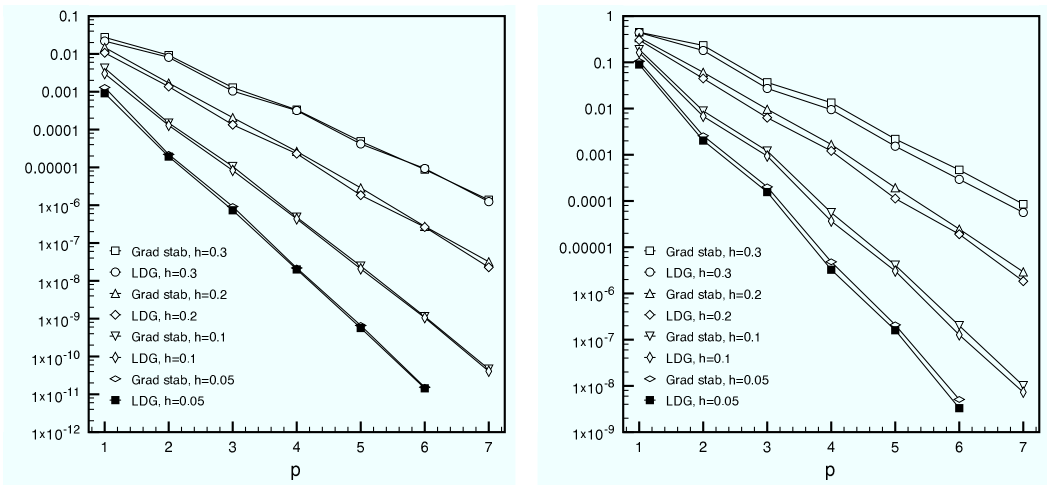
and corresponding Dirichlet boundary condition such that the solution consists of

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



(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.

We consider sequences of unstructured meshes for polynomial degrees $p=1,\cdots,7$. 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 [21,22].

Fig. 1 shows the behavior of the approximations u_h and σ_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.

Fig. 2 shows the behavior of the approximations u_h and σ_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 the anonymous reviewers for their comments that helped improve the manuscript. This project received financial support from 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(4):742–760, 1982.

[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(5):1749–1779 (electronic), 2001/02.

[3] G. A. Baker. Finite element methods for elliptic equations using nonconforming elements. *Math. Comp.*, 31(137):45–59, 1977.

[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(2):267–279, 1997.

[5] F. Brezzi, J. Douglas, Jr., and L. D. Marini. Two families of mixed finite elements for second order elliptic problems. *Numer. Math.*, 47(2):217–235, 1985.

[6] E. Burman, A. Ern, I. Mozolevski, and B. Stamm. 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.*, 354(10):599–602, 2007.

[7] E. Burman and B. Stamm. Low order discontinuous Galerkin methods for second order elliptic problems. Technical report, EPFL-IACS report 04, 2007.

[8] E. Burman and B. Stamm. Minimal stabilization for discontinuous Galerkin finite element methods for hyperbolic problems. *J. Sci. Comput.*, 33(2):183–208, 2007.

[9] E. Burman and B. Stamm. Symmetric and non-symmetric discontinuous Galerkin methods stabilized using bubble enrichment. *C. R. Math. Acad. Sci. Paris*, 346(1-2):103–106, 2008.

[10] P. Castillo, B. Cockburn, I. Perugia, and D. Schötzau. An a priori error analysis of the local discontinuous Galerkin method for elliptic problems. *SIAM J. Numer. Anal.*, 38(5):1676–1706 (electronic), 2000.

[11] P. Castillo, B. Cockburn, D. Schötzau, and C. Schwab. Optimal a priori error estimates for the *hp*-version of the local discontinuous Galerkin method for convection-diffusion problems. *Math. Comp.*, 71(238):455–478 (electronic), 2002.

[12] B. Cockburn and B. Dong. An analysis of the minimal dissipation local discontinuous Galerkin method for convection-diffusion problems. *J. Sci. Comput.*, 32(2):233–262, 2007.

[13] B. Cockburn, G. Kanschat, I. Perugia, and D. Schötzau. Superconvergence of the local discontinuous Galerkin method for elliptic problems on Cartesian grids. *SIAM J. Numer. Anal.*, 39(1):264–285 (electronic), 2001.

[14] B. Cockburn and C.-W. Shu. The local discontinuous Galerkin method for time-dependent convection-diffusion systems. *SIAM J. Numer. Anal.*, 35(6):2440–2463 (electronic), 1998.

[15] J. Douglas and T. Dupont. Interior penalty procedures for elliptic and parabolic Galerkin methods. In *Computing methods in applied sciences (Second Internat. Sympos., Versailles, 1975)*, pages 207–216. Lecture Notes in Phys., Vol. 58. Springer, Berlin, 1976.

[16] A. Ern and J.-L. Guermond. Discontinuous Galerkin methods for Friedrichs' systems. I. General theory. *SIAM J. Numer. Anal.*, 44(2):753–778 (electronic), 2006.
[17] A. Ern and J.-L. Guermond. Discontinuous Galerkin methods for Friedrichs' systems. II. Second-order elliptic PDEs. *SIAM J. Numer. Anal.*, 44(6):2363–2388 (electronic), 2006.
[18] P. Houston, C. Schwab, and E. Süli. Discontinuous hp-finite element methods for advection-diffusion-reaction problems. *SIAM J. Numer. Anal.*, 39(6):2133–2163 (electronic), 2002.
[19] D. Marazzina. Mixed discontinuous Galerkin methods with minimal stabilization. In *Numerical Mathematics and Advanced Applications*, pages 448–456. Springer, Berlin, 2006.
[20] I. Perugia and D. Schötzau. An hp-analysis of the local discontinuous Galerkin method for diffusion problems. *J. Sci. Comput.*, 17(1-4):561–571, 2002.
[21] C. Prud'homme. A domain specific embedded language in C++ for automatic differentiation, projection, integration and variational formulations. *Scientific Programming*, 14(2):81–110, 2006.
[22] C. Prud'homme. Life: Overview of a unified C++ implementation of the finite and spectral element methods in 1d, 2d and 3d. In *Workshop On State-Of-The-Art In Scientific And Parallel Computing*, volume 4699 of *Lecture Notes in Computer Science*, pages 712–721. Springer-Verlag, 2007.
[23] 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(5):752–784, 2006.
[24] M. F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM J. Numer. Anal.*, 15(1):152–161, 1978.