Symmetric Interior Penalty Galerkin Method for Elliptic Problems

Yekaterina Epshteyn and Béatrice Rivière [1]

Department of Mathematics, University of Pittsburgh, 301 Thackeray, Pittsburgh, PA 15260, U.S.A.

Abstract

This paper presents computable lower bounds of the penalty parameters for stable and convergent symmetric interior penalty Galerkin methods. In particular, we derive the explicit dependence of the coercivity constants with respect to the polynomial degree and the angles of the mesh elements. Numerical examples in all dimensions and for different polynomial degrees are presented. We investigate the numerical effects of loss of coercivity.

Key words: Coercivity, stable v. unstable interior penalty, elliptic problems

1 Introduction

The Symmetric Interior Penalty Galerkin (SIPG) method for elliptic problems was first introduced in the late seventies by Douglas and Dupont [9], Wheeler [21] and Arnold [1,2] and was revived more recently as a popular discontinuous Galerkin method. Some of the general attractive features of the method are the local and high order of approximation, the flexibility due to local mesh refinement and the ability to handle unstructured meshes and discontinuous coefficients. More specific properties include the optimal error estimates in both the H^1 and L^2 norms and the resulting symmetric linear systems easily solved by standard solvers for symmetric matrices (such as conjugate gradient). The analysis and application of SIPG to a wide range of problems can be found in the literature: a non-exhaustive list is given in [4,5,7,11,16,18,19,14] and the references herein.

[1] This research is partially funded by NSF-DMS 0506039.

The SIPG method is obtained by integrating by parts on each mesh element, and summing over all elements. Two stabilization terms are then added: a symmetrizing term corresponding to fluxes obtained after integration by part, and a penalty term imposing a weak continuity of the numerical solution. It is well known that there exists a threshold penalty above which the bilinear form is coercive and the scheme is stable and convergent. Another related discontinuous Galerkin method is the non-symmetric interior penalty Galerkin (NIPG) method [17,12]: this method differs from the SIPG method by only one sign: the symmetrizing term is added instead of being subtracted. On one hand, the loss of symmetry in the scheme gives an immediate coercivity of the bilinear form; the NIPG scheme is stable and convergent for any value of the penalty. On the other hand, optimal error estimates in the L^2 norm cannot be proved via the standard Nitsche lift. As of today, this remains an open problem.

The objective of this work is to derive rigorous computable bounds of the threshold penalty that would yield a stable and convergent SIPG. We consider a general second order elliptic problem on a domain in any dimension, subdivided into simplices. Our main result is an improved coercivity result. In particular, we show that the constant of coercivity depends on the polynomial degree and the smallest $\sin\theta$ over all angles θ in the triangular mesh in 2D or over all dihedral angles θ in the tetrahedral mesh in 3D. We also investigate the effects of the penalty numerically and exhibit unstable oscillatory solutions for penalty values below the threshold penalty. Our results also apply to the incomplete interior penalty Galerkin method [8], that differs from SIPG and NIPG in the fact that the symmetrizing stabilizing term is removed. For this method, the error analysis in the energy norm is identical to the analysis of the SIPG method.

The outline of the paper is as follows: the model problem and scheme are presented in Section 2. Section 3 contains the improved coercivity theorems. Section 4 shows numerical examples in all dimensions that support our theoretical results. Some conclusions follow.

2 Model Problem and Scheme

Let Ω be a domain in $\mathbb{R}^d, d = 1, 2, 3$. Let the boundary of the domain $\partial\Omega$ be the union of two disjoint sets Γ_D and Γ_N. We denote $\mathbf{n}$ the unit normal vector to each edge of $\partial\Omega$ exterior of Ω. For f given in $L^2(\Omega)$, u_D given in $H^{\frac{1}{2}}(\Gamma_D)$ and u_N given in $L^2(\Gamma_N)$, we consider the following elliptic problem:

$$-\nabla \cdot (K\nabla u) + \alpha u = f \text{ in } \Omega, \tag{1}$$
$$u = u_D \text{ on } \Gamma_D, \tag{2}$$
$$K\nabla u \cdot \mathbf{n} = u_N \text{ on } \Gamma_N. \tag{3}$$

Here, the function α is a nonnegative scalar function and K is a matrix-valued function $K = (k_{ij})_{1\leq i,j\leq d}$ that is symmetric positive definite, i.e. there exist two positive constants k_0 and k_1 such that

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad k_0 \mathbf{x}^T\mathbf{x} \leq \mathbf{x}^T K \mathbf{x} \leq k_1 \mathbf{x}^T\mathbf{x}. \tag{4}$$

We can assume that the problem (1)-(3) has a unique solution in $H^1(\Omega)$ when $|\Gamma_D| > 0$ or when $\alpha \neq 0$. On the other hand, when $\partial\Omega = \Gamma_N$ and $\alpha = 0$, problem (1)-(3) has a solution in $H^1(\Omega)$ which is unique up to an additive constant, provided $\int_\Omega f = -\int_{\partial\Omega} g$.

Let $\mathcal{T}_h = \{E\}_E$ be a subdivision of Ω, where E is an interval if $d = 1$, a triangle if $d = 2$, or a tetrahedron if $d = 3$. Let

$$h = \max_{E\in\mathcal{T}_h} h_E,$$

where h_E is the diameter of E.

Let p be a positive integer. Denote by $\mathbb{P}_p(E)$ the space of polynomials of total degree less than p on the element E. The finite element subspace is taken to be

$$\mathcal{D}_p(\mathcal{T}_h) = \{v_h \in L^2(\Omega) : \forall E \in \mathcal{T}_h \quad v_h|_E \in \mathbb{P}_p(E)\}.$$

We note that there are no continuity constraints on the discontinuous finite element spaces. In what follows, we will denote by $\|\cdot\|_{\mathcal{O}}$ the L^2 norm over the domain $\mathcal{O}$.

We now present the scheme. For readibility purposes, we separate the one-dimensional case from the higher dimensional case.

2.1 SIPG in One Dimension

Assuming that $\Omega = (a, b)$, we can write the subdivision:

$$\mathcal{T}_h = \{I_n = (x_n, x_{n+1}) : n = 0, \ldots, N-1\}$$

with $x_0 = a$ and $x_N = b$. We assume that $\Gamma_D = \{a, b\}$ and thus $\Gamma_N = \emptyset$.

If we denote $v(x_n^+) = \lim_{\varepsilon\to 0^+} v(x_n + \varepsilon)$ and $v(x_n^-) = \lim_{\varepsilon\to 0^+} v(x_n - \varepsilon)$, we can define

the jump and average of v at the endpoints of I_n:

$$\forall n = 1, \ldots, N-1, \quad [v(x_n)] = v(x_n^-) - v(x_n^+), \quad \{v(x_n)\} = \frac{1}{2}(v(x_n^-) + v(x_n^+)),$$

$$[v(x_0)] = -v(x_0^+), \quad \{v(x_0)\} = v(x_0^+), \quad [v(x_N)] = v(x_N^-), \quad \{v(x_N)\} = v(x_N^-).$$

The SIPG finite element method for problem (1)-(3) is then : find u_h in $\mathcal{D}_p(\mathcal{T}_h)$ such that :

$$\forall v_h \in \mathcal{D}_p(\mathcal{T}_h), \quad A(u_h, v_h) = L(v_h), \tag{5}$$

where the bilinear form A and linear form L are defined by:

$$\begin{aligned} A(w,v) = &\sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} (K(x)w'(x)v'(x) + \alpha w(x)v(x))\mathrm{dx} + \frac{\sigma_0}{|I_1|}[w(x_0)][v(x_0)] \\ &+ \sum_{n=1}^{N-1} \sigma_n \Big(\frac{1}{2|I_{n+1}|} + \frac{1}{2|I_n|} \Big) [w(x_n)][v(x_n)] + \frac{\sigma_N}{|I_N|}[w(x_N)][v(x_N)] \\ &- \sum_{n=0}^{N} \{K(x_n)w'(x_n)\}[v(x_n)] - \sum_{n=0}^{N} \{K(x_n)v'(x_n)\}[w(x_n)], \end{aligned} \tag{6}$$

$$\begin{aligned} L(v) = &\int_a^b f(x)v(x)\mathrm{dx} + K(a)v'(a)u_D(a) - K(b)v'(b)u_D(b) \\ &+ \frac{\sigma_0}{|I_0|}v(a)u_D(a) + \frac{\sigma_N}{|I_N|}v(b)u_D(b), \end{aligned} \tag{7}$$

where $\{\sigma_n\}_n$ are real positive penalty parameters defined on each subinterval I_n independently. We denote by $\sigma > 0$ the minimum of all σ_n. The energy norm associated to A is:

$$\begin{aligned} \forall v_h \in \mathcal{D}_p(\mathcal{T}_h), \quad \|v_h\|_{\mathcal{E}} = \Big(&\sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} (K(x)(v_h'(x))^2 + \alpha(x)(v_h(x))^2)\mathrm{dx} \\ + \frac{\sigma_0}{|I_1|}[v(x_0)]^2 + &\sum_{n=1}^{N-1} \sigma_n \Big(\frac{1}{2|I_{n+1}|} + \frac{1}{2|I_n|} \Big) [v(x_n)]^2 + \frac{\sigma_N}{|I_N|}[v(x_N)]^2 \Big)^{1/2}. \end{aligned} \tag{8}$$

2.2 SIPG in High Dimensions

Let Γ_h be the set of interior edges in 2D (or faces in 3D) of the subdivision $\mathcal{T}_h$. With each edge (or face) e, we associate a unit normal vector $\mathbf{n}_e$. If e is on the boundary $\partial\Omega$, then $\mathbf{n}_e$ is taken to be the unit outward vector to $\partial\Omega$.

We now define the average and the jump for w:

$$\forall e = \partial E_e^1 \cap E_e^2, \quad \{w\} = \frac{1}{2}(w|_{E_e^1}) + \frac{1}{2}(w|_{E_e^2}), \quad [w] = (w|_{E_e^1}) - (w|_{E_e^2}),$$

$$\forall e = \partial E_e^1 \cap \partial\Omega, \quad \{w\} = (w|_{E_e^1}), \quad [w] = (w|_{E_e^1}).$$

The general SIPG variational formulation of problem (1)-(3) is: find u_h in $\mathcal{D}_p(\mathcal{T}_h)$ such that:

$$\forall v_h \in \mathcal{D}_p(\mathcal{T}_h), \quad A(u_h, v_h) = L(v_h), \tag{9}$$

where the bilinear form A and linear form L are defined by:

$$\begin{aligned} A(w,v) = &\sum_{E\in\mathcal{T}_h}\int_E K\nabla w\cdot\nabla v + \int_\Omega \alpha w v + \sum_{e\in\Gamma_h\cup\Gamma_D}\frac{\sigma_e}{|e|^{\beta_0}}\int_e [w][v] \\ &- \sum_{e\in\Gamma_h\cup\Gamma_D}\int_e \{K\nabla w\cdot\mathbf{n}_e\}[v] - \sum_{e\in\Gamma_h\cup\Gamma_D}\int_e \{K\nabla v\cdot\mathbf{n}_e\}[w], \end{aligned} \tag{10}$$

$$L(v) = \int_\Omega fv - \sum_{e\in\Gamma_D}\int_e (K\nabla v\cdot\mathbf{n}_e)u_D + \sum_{e\in\Gamma_D}\int_e \frac{\sigma_e}{|e|^{\beta_0}} v u_D + \sum_{e\in\Gamma_N}\int_e v u_N. \tag{11}$$

The penalty parameter σ_e is a positive constant on each edge (or face) e and we denote by $\sigma > 0$ the minimum of all σ_e. The parameter $\beta_0 > 0$ is a global constant that, in general, is chosen to be one. If $\beta_0 > 1$, then the SIPG method is said to be superpenalized. The energy norm associated to A is:

$$\forall v_h \in \mathcal{D}_p(\mathcal{T}_h), \quad \|v_h\|_{\mathcal{E}} = \Big(\sum_{E\in\mathcal{T}_h}\int_E K(\nabla v_h)^2 + \int_\Omega \alpha v_h^2 + \sum_{e\in\Gamma_h\cup\Gamma_D}\frac{\sigma_e}{|e|^{\beta_0}}\int_e [v_h]^2\Big)^{\frac{1}{2}}. \tag{12}$$

2.3 Error Analysis

We recall the well-known results about the schemes (5) and (9).

Lemma 1 *Consistency. The exact solution of (1)-(3) satisfies the discrete variational problem (5) in one dimension and (9) in two or three dimensions.*

Lemma 2 *Coercivity. There exists a penalty σ^* such that for any $\sigma > \sigma^*$ we have*

$$\forall v_h \in \mathcal{D}_p(\mathcal{T}_h), \quad A(v_h, v_h) \geq C^* \|v_h\|_{\mathcal{E}}^2,$$

for some positive constant C^* independent of h.

Lemma 3 *Continuity. There exists a constant $\tilde{C}$ such that*

$$\forall v_h, w_h \in \mathcal{D}_p(\mathcal{T}_h), \quad A(v_h, w_h) \leq \tilde{C}\|v_h\|_{\mathcal{E}}\|w_h\|_{\mathcal{E}}.$$

Theorem 4 *Error estimates. Let $u \in H^{p+1}(\Omega)$ be the exact solution of (1)-(3). Assume that the coercivity lemma holds true. In addition, assume that*

$\beta_0 \geq 1$. Then, there is a constant C independent of h, but dependent of $\frac{1}{C^*}$, such that

$$\|u - u_h\|_{\mathcal{E}} \leq Ch^p |u|_{H^{p+1}(\Omega)}.$$

These results are proved by using standard trace inequalities [6] and they can be found for example in [2,3,13].

The aim of this work is to determine exactly what is the value σ^* that would guarantee the coercivity. We also obtain a precise expression for both coercivity and continuity constants $C^*, \tilde{C}$. We then show numerically that for penalty values lower than σ^*, unstable solutions could occur.

3 Improved Coercivity and Continuity Lemmas

We will consider each dimension separately as the details of the proofs differ.

3.1 Estimation of σ^* in One Dimension

Theorem 5 *Let $\varepsilon^* = \frac{k_0}{2}$. For any $\varepsilon > 0$, define*

$$\sigma_n^*(\varepsilon) = \begin{cases} \frac{2k_1^2(p+1)^2}{\varepsilon} & \forall n = 1, \ldots, N-1, \\ \frac{3k_1^2(p+1)^2}{4\varepsilon} & n = 0, N. \end{cases} \tag{13}$$

Then for any $0 < \varepsilon < \varepsilon^*$, if $\sigma_n > \sigma_n^*(\varepsilon)$ for all n, there is a constant $0 < C^*(\varepsilon) < 1$, independent of h, such that

$$\forall v_h \in \mathcal{D}_p(\mathcal{T}_h), \quad A(v_h, v_h) \geq C^*(\varepsilon)\|v_h\|_{\mathcal{E}}^2.$$

Moreover, an expression for $C^*(\varepsilon)$ is:

$$C^*(\varepsilon) = \min\{1 - \frac{2\varepsilon}{k_0}, \quad 1 - \frac{3k_1^2(p+1)^2}{4\varepsilon\sigma_0}, \quad 1 - \frac{2k_1^2(p+1)^2}{\varepsilon\sigma_1}, \ldots, 1 - \frac{2k_1^2(p+1)^2}{\varepsilon\sigma_{N-1}}, \quad 1 - \frac{3k_1^2(p+1)^2}{4\varepsilon\sigma_N}\}.$$

Proof Choosing $w = v$ in (6) yields

$$A(v,v) = \sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} K(x)(v'(x))^2 \mathrm{dx} - 2\sum_{n=0}^{N} \{K(x_n)v'(x_n)\}[v(x_n)] + \frac{\sigma_0}{|I_1|}\|[v(x_0)]\|^2 + \sum_{n=1}^{N-1} \sigma_n \Big(\frac{1}{2|I_{n+1}|} + \frac{1}{2|I_n|}\Big)\|[v(x_n)]\|^2 + \frac{\sigma_N}{|I_N|}\|[v(x_N)]\|^2. \tag{14}$$

It suffices to bound the term $-2\sum_{n=0}^{N}\{K(x_n)v'(x_n)\}[v(x_n)]$ and obtain some restrictions on the penalty parameters $\{\sigma_n\}_n$ for the coercivity to hold. Let us first consider one interior point x_n. By definition of the average and the property (4), we have

$$\begin{aligned}|\{K(x_n)v'(x_n)\}| &\leq \frac{1}{2}|K(x_n^-)v'(x_n^-)| + \frac{1}{2}|K(x_n^+)v'(x_n^+)| \\ &\leq \frac{k_1}{2}(|v'(x_n^-)| + |v'(x_n^+)|). \end{aligned} \tag{15}$$

For any interval $I = (s,t)$, the following improved inverse trace inequality holds [20]:

$$\forall v_h \in \mathbb{P}_p(I), \quad |v_h(s)| \leq \frac{p+1}{\sqrt{|I|}}\|v_h\|_I. \tag{16}$$

Hence using (16) we can bound $|v'(x_n^-)|$ and $|v'(x_n^+)|$:

$$|v'(x_n^-)| \leq \frac{p+1}{\sqrt{x_n - x_{n-1}}}\|v'(x)\|_{(x_{n-1},x_n)}, \quad |v'(x_n^+)| \leq \frac{p+1}{\sqrt{x_{n+1}-x_n}}\|v'(x)\|_{(x_n,x_{n+1})}.$$

Using these bounds we obtained for the interior point x_n of the subdivision:

$$\{K(x_n)v'(x_n)\}[v(x_n)] \leq \frac{k_1(p+1)}{2}\left(\frac{\|v'(x)\|_{(x_{n-1},x_n)}}{\sqrt{x_n - x_{n-1}}} + \frac{\|v'(x)\|_{(x_n,x_{n+1})}}{\sqrt{x_{n+1}-x_n}}\right)|[v(x_n)]|. \tag{17}$$

Let us consider now the boundary nodes x_0 and x_N:

$$\begin{aligned}\{K(x_0)v'(x_0)\}[v(x_0)] &\leq |K(x_0)v'(x_0)[v(x_0)]| \\ &\leq k_1(p+1)\frac{\|v'(x)\|_{(x_1,x_0)}}{\sqrt{(x_1-x_0)}}|[v(x_0)]|, \end{aligned} \tag{18}$$

$$\begin{aligned}\{K(x_N)v'(x_N)\}[v(x_N)] &\leq |K(x_N)v'(x_N)[v(x_N)]| \\ &\leq k_1(p+1)\frac{\|v'(x)\|_{(x_{N-1},x_N)}}{\sqrt{(x_N-x_{N-1})}}|[v(x_N)]|. \end{aligned} \tag{19}$$

Combining the bounds above gives:

$$\sum_{n=0}^{N}\{K(x_n)v'(x_n)\}[v(x_n)] \leq \frac{k_1(p+1)}{2}\Big(\|v'(x)\|_{(x_0,x_1)}\frac{|[v(x_0)]|}{\sqrt{x_1-x_0}}$$

$$+\sum_{n=1}^{N}\|v'(x)\|_{(x_{n-1},x_n)}\frac{(|[v(x_{n-1})]| + |[v(x_n)]|)}{\sqrt{x_n-x_{n-1}}} + \|v'(x)\|_{(x_N,x_{N-1})}\frac{|[v(x_N)]|}{\sqrt{x_N-x_{N-1}}}\Big).$$

After application of discrete Cauchy-Shwarz's inequality we have:

$$\sum_{n=0}^{N}\{K(x_n)v'(x_n)\}[v(x_n)] \leq \frac{k_1(p+1)}{2}\Big(2\|v'(x)\|^2_{(x_0,x_1)} + \sum_{n=2}^{N-1}\|v'(x)\|^2_{(x_{n-1},x_n)}$$

$$+2\|v'(x)\|^2_{(x_N,x_{N-1})}\Big)^{\frac{1}{2}}\Big(\frac{\|[v(x_0)]\|^2}{x_1-x_0}+\sum_{n=1}^{N}\frac{(|[v(x_{n-1})]|+[v(x_n)]|)^2}{x_n-x_{n-1}}+\frac{\|[v(x_N)]\|^2}{x_N-x_{N-1}}\Big)^{\frac{1}{2}}.$$

Application of Young's inequality yields:

$$\sum_{n=0}^{N}\{K(x_n)v'(x_n)\}[v(x_n)] \leq \frac{\varepsilon}{k_0}\|K^{1/2}v'(x)\|^2_{0,I_1} + \sum_{n=2}^{N-1}\frac{\varepsilon}{2k_0}\|K^{1/2}v'(x)\|^2_{0,I_n}$$

$$+\frac{\varepsilon}{k_0}\|K^{1/2}v'(x)\|^2_{0,I_N} + \frac{k_1^2(p+1)^2}{\varepsilon}\Big(\frac{3}{8|I_1|}[v(x_0)]^2 + \sum_{n=1}^{N-1}\Big(\frac{1}{2|I_n|}+\frac{1}{2|I_{n+1}|}\Big)[(v(x_n)]^2 + \frac{3}{8|I_N|}[v(x_N)]^2\Big). \tag{20}$$

Hence using the estimate (20) we obtain a lower bound for the bilinear form (14):

$$A(v,v) \geq \Big(1-\frac{2\varepsilon}{k_0}\Big)\int_{I_1}K(x)v'(x)^2\mathrm{d}x + \sum_{n=2}^{N-1}\Big(1-\frac{\varepsilon}{k_0}\Big)\int_{I_n}K(x)v'(x)^2\mathrm{d}x$$

$$+\Big(1-\frac{2\varepsilon}{k_0}\Big)\int_{I_N}K(x)v'(x)^2\mathrm{d}x + \Big(\sigma_0-\frac{3k_1^2(p+1)^2}{4\varepsilon}\Big)\frac{[v(x_0)]^2}{|I_1|} \tag{21}$$

$$+\sum_{n=1}^{N-1}\Big(\sigma_n-\frac{2k_1^2(p+1)^2}{\varepsilon}\Big)\Big(\frac{1}{2|I_{n+1}|}+\frac{1}{2|I_n|}\Big)[v(x_n)]^2+\Big(\sigma_N-\frac{3k_1^2(p+1)^2}{4\varepsilon}\Big)\frac{[v(x_N)]^2}{|I_N|}.$$

Let us denote $\varepsilon^* = \frac{k_0}{2}$. From (21) the bilinear form (14) is coercive if :

$$\varepsilon < \varepsilon^* \tag{22}$$

and

$$\sigma_n > \begin{cases} \frac{2k_1^2(p+1)^2}{\varepsilon} & \forall n = 1,\dots,N-1, \\ \\ \frac{3k_1^2(p+1)^2}{4\varepsilon} & n = 0, N. \end{cases} \tag{23}$$

This concludes the proof. □

Similarly, one can show the following improved continuity constant.

Lemma 6 *Under the notation of Theorem 5, the continuity constant $\tilde{C}$ of Lemma 3 is given by:*

$$\tilde{C} = \max\{1+\frac{3}{\sigma_0}, \quad 1+\frac{8}{\sigma_1},\dots,1+\frac{8}{\sigma_{N-1}}, \quad 1+\frac{3}{\sigma_N}, 1+\frac{k_1^2(p+1)^2}{2k_0}\}.$$

In this section, we denote $\theta_{\mathcal{T}}$ the angle such that $\sin\theta_{\mathcal{T}}$ has the smallest value over all triangles. In two dimensions, it is to be noted that this angle $\theta_{\mathcal{T}}$ corresponds to the smallest angle over all triangles in the subdivision. We show that the coercivity constant depends on $\theta_{\mathcal{T}}$. In Section 4, we outline a simple algorithm for computing such angle.

Theorem 7 *For any $\varepsilon > 0$, define*

$$\sigma_e^*(\varepsilon) = 5\frac{k_1^2}{\varepsilon}(p+1)(p+2)\cot\theta_{\mathcal{T}}|e|^{\beta_0-1}. \tag{24}$$

Then for any $0 < \varepsilon < k_0$, if $\sigma_e > \sigma_e^*(\varepsilon)$ for all e, there is a constant $0 < C^*(\varepsilon) < 1$, independent of h, such that

$$\forall v_h \in \mathcal{D}_p(\mathcal{T}_h), \quad A(v_h, v_h) \geq C^*(\varepsilon)\|v_h\|_{\mathcal{E}}^2.$$

An expression for C^* is:

$$C^*(\varepsilon) = \min_{e\in\Gamma_h\cup\Gamma_D}\{1-\frac{\varepsilon}{k_0}, \quad 1-\frac{5}{\varepsilon\sigma_e}k_1^2(p+1)(p+2)\cot\theta_{\mathcal{T}}|e|^{\beta_0-1}\}.$$

Proof:

Similarly, as in the one-dimensional case, we choose $w = v$ in (10):

$$\begin{aligned} A(v,v) = &\sum_{E\in\mathcal{T}_h}\int_E K(\nabla v)^2 + \int_\Omega \alpha v^2 \\ -2&\sum_{e\in\Gamma_h\cup\Gamma_D}\int_e\{K\nabla v\cdot\mathbf{n}_e\}[v] + \sum_{e\in\Gamma_h\cup\Gamma_D}\frac{\sigma_e}{|e|^{\beta_0}}\int_e[v]^2. \end{aligned} \tag{25}$$

In order to have coercivity of the bilinear form we need to bound the term $-2\sum_{e\in\Gamma_h\cup\Gamma_D}\int_e\{K\nabla v\cdot\mathbf{n}_e\}[v]$.

Let us first consider one interior edge e shared by two triangles E_1 and E_2. Applying Cauchy-Schwarz inequality we have:

$$\int_e\{K\nabla v\cdot\mathbf{n}_e\}[v] \leq \|\{K\nabla v\cdot\mathbf{n}_e\}\|_e\|[v]\|_e. \tag{26}$$

Using the definition of the average and the property (4), we have

$$\|\{K\nabla v\cdot\mathbf{n}_e\}\|_e \leq \frac{1}{2}\|K\nabla v\cdot\mathbf{n}_e|_{E_1}\|_e + \frac{1}{2}\|K\nabla v\cdot\mathbf{n}_e|_{E_2}\|_e.$$

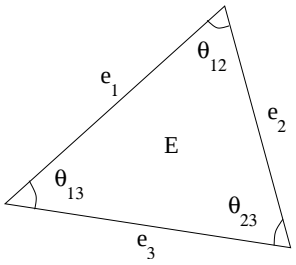

Fig. 1. Angles and edges in a generic triangle.

$$\leq \frac{1}{2}\Big(\|K\nabla v|_{E_1}\|_e + \|K\nabla v|_{E_2}\|_e\Big) \leq \frac{k_1}{2}\Big(\|\nabla v|_{E_1}\|_e + \|\nabla v|_{E_2}\|_e\Big), \tag{27}$$

so we obtain for the interior edge e:

$$\int_e \{K\nabla v\cdot \mathbf{n}_e\}[v] \leq \frac{k_1}{2}\Big(\|\nabla v|_{E_1}\|_e + \|\nabla v|_{E_2}\|_e\Big)\|[v]\|_e. \tag{28}$$

Similarly, for a boundary edge e belonging to the boundary of element E:

$$\int_e \{K\nabla v\cdot \mathbf{n}_e\}[v] \leq k_1\|\nabla v|_E\|_e\|[v]\|_e. \tag{29}$$

We now recall the inverse inequality valid on an edge of a triangle E [20]:

$$\forall v_h \in \mathbb{P}_p(E), \quad \|v_h\|_e \leq \sqrt{\frac{(p+1)(p+2)}{2}\frac{|e|}{|E|}}\|u\|_E. \tag{30}$$

Hence in (30) we need to estimate the ratio $\frac{|e|}{|E|}$, where e is one edge of a triangle E. For this, we consider a triangle with edges e_1, e_2 and e_3. We denote by θ_{ij} the interior angle between edge e_i and edge e_j (see Fig. 1).Without loss of generality, we assume that $e = e_3$.

The area of the triangle E is given by the formula:

$$|E| = \frac{1}{2}|e_i||e_j|\sin\theta_{ij} = \frac{1}{4}|e_3||e_1|\sin\theta_{13} + \frac{1}{4}|e_3||e_2|\sin\theta_{23}$$

The length of the edge e in the triangle E can also be written as :

$$|e| = |e_3| = |e_1|\cos\theta_{13} + |e_2|\cos\theta_{23}.$$

Hence, using the smallest angle $\theta_{\mathcal{T}}$ we have:

$$\frac{|e|}{|E|} = \frac{4}{|e|}\Big(\frac{|e_1|\cos\theta_{13} + |e_2|\cos\theta_{23}}{|e_1|\sin\theta_{13} + |e_2|\sin\theta_{23}}\Big) \leq \frac{4}{|e|}\Big(\frac{|e_1|\cos\theta_{\mathcal{T}} + |e_2|\cos\theta_{\mathcal{T}}}{|e_1|\sin\theta_{\mathcal{T}} + |e_2|\sin\theta_{\mathcal{T}}}\Big).$$

So we obtain the following estimate :

$$\frac{|e|}{|E|} \leq \frac{4\cot\theta_{\mathcal{T}}}{|e|}. \tag{31}$$

Then using inverse inequality (30), and the estimate (31) in (28) and (29) we obtain for the interior edge e of the triangle E_1 and E_2:

$$\int_e \{K\nabla v \cdot \mathbf{n}_e\}[v] \leq k_1 \sqrt{\frac{(p+1)(p+2)}{2|e|}\cot\theta_{\mathcal{T}}}\Big(\|\nabla v\|_{0,E_1} + \|\nabla v\|_{0,E_2}\Big)\|[v]\|_e, \tag{32}$$

and for the boundary edge:

$$\int_e \{K\nabla v \cdot \mathbf{n}_e\}[v] \leq k_1 \sqrt{\frac{2(p+1)(p+2)}{|e|}\cot\theta_{\mathcal{T}}}\|\nabla v\|_E\|[v]\|_e. \tag{33}$$

Combining the bounds above and using discrete Cauchy-Schwarz's inequality, we obtain:

$$\begin{aligned}
\sum_{e\in\Gamma_h\cup\Gamma_D}\int_e \{K\nabla v \cdot \mathbf{n}_e\}[v] &\leq k_1\sqrt{(p+1)(p+2)\cot\theta_{\mathcal{T}}} \\
\times\Big(\sum_{e\in\Gamma_h}\frac{1}{\sqrt{2}}(\|\nabla v\|_{E_1}+\|\nabla v\|_{E_2})\frac{1}{\sqrt{|e|}}\|[v]\|_e + \sum_{e\in\Gamma_D}\sqrt{2}\|\nabla v\|_E\frac{1}{\sqrt{|e|}}\|[v]\|_e\Big) \\
&\leq k_1\sqrt{(p+1)(p+2)\cot\theta_{\mathcal{T}}} \\
\times\Big(\sum_{e\in\Gamma_h}(\|\nabla v\|^2_{E_1}+\|\nabla v\|^2_{E_2}) + \sum_{e\in\Gamma_D}2\|\nabla v\|^2_E\Big)^{1/2}&\Big(\sum_{e\in\Gamma_h\cup\Gamma_D}\frac{1}{|e|}\|[v]\|^2_e\Big)^{1/2}.
\end{aligned} \tag{34}$$

We now rewrite the first sum over edges as a sum over triangles by decomposing the subdivision into disjoint sets $\mathcal{T}_0, \mathcal{T}_{1D}, \mathcal{T}_{2D}, \mathcal{T}_{1N}, \mathcal{T}_{2N}$ and $\mathcal{T}_{2DN}$. The set $\mathcal{T}_0$ represents the set of triangles with three interior edges. The set $\mathcal{T}_{1D}$ represents the set of triangles with two interior edges and one boundary edge of Dirichlet type. The set $\mathcal{T}_{2D}$ represents the set of triangles with one interior edge and two edges on the Dirichlet boundary. The set $\mathcal{T}_{1N}$ represents the set of triangles with two interior edges and one boundary edge of type Neumann. The set $\mathcal{T}_{2N}$ represents the set of triangles with one interior edge and two boundary edges of type Neumann. Finally, the set $\mathcal{T}_{2DN}$ represents the set of triangles with one interior edge, one Neumann edge and one Dirichlet edge. The sum over edges can then be rewritten as:

$$\begin{aligned}
\sum_{e\in\Gamma_h}(\|\nabla v\|^2_{E_1}+\|\nabla v\|^2_{E_2}) + \sum_{e\in\Gamma_D}2\|\nabla v\|^2_E &= \sum_{E\in\mathcal{T}_0}3\|\nabla v\|^2_E + \sum_{E\in\mathcal{T}_{1D}}4\|\nabla v\|^2_E \\
+\sum_{E\in\mathcal{T}_{2D}}5\|\nabla v\|^2_E + \sum_{E\in\mathcal{T}_{1N}}2\|\nabla v\|^2_E &+ \sum_{E\in\mathcal{T}_{2N}}\|\nabla v\|^2_E + \sum_{E\in\mathcal{T}_{2DN}}3\|\nabla v\|^2_E \\
&\leq 5\sum_{E\in\mathcal{T}_h}\|\nabla v\|^2_E.
\end{aligned}$$

Therefore, by using Young's inequality and the property (4), we have for any positive ε:

$$\sum_{e\in\Gamma_h\cup\Gamma_D}\int_e\{K\nabla v\cdot\mathbf{n}_e\}[v]\le\frac{\varepsilon}{2k_0}\sum_{E\in\mathcal{T}_h}\int_E K(\nabla v)^2$$
$$+\frac{5}{2\varepsilon}\sum_{e\in\Gamma_h\cup\Gamma_D}\frac{k_1^2(p+1)(p+2)\cot\theta_{\mathcal{T}}|e|^{\beta_0-1}}{|e|^{\beta_0}}\int_e[v]^2. \tag{35}$$

Therefore using the estimate (35) we have the following lower bound for the bilinear form (25):

$$A(v,v)\ge\left(1-\frac{\varepsilon}{k_0}\right)\sum_{E\in\mathcal{T}_h}\|K^{\frac{1}{2}}\nabla v\|_E^2$$
$$+\sum_{e\in\Gamma_h\cup\Gamma_D}\frac{\sigma_e-\frac{5}{\varepsilon}k_1^2(p+1)(p+2)\cot\theta_{\mathcal{T}}|e|^{\beta_0-1}}{|e|^{\beta_0}}\|[v]\|_e^2. \tag{36}$$

From (36) the bilinear form (25) is coercive if the following two conditions hold:

$$\varepsilon<k_0, \tag{37}$$

$$\sigma_e>5\frac{k_1^2}{k_0}(p+1)(p+2)\cot\theta_{\mathcal{T}}|e|^{\beta_0-1}. \tag{38}$$

This concludes the proof. □

Corollary 8 *Assume that no super-penalization is used, namely $\beta_0=1$, then the estimate is independent of h:*

$$\forall 0<\varepsilon<k_0,\quad \sigma_e^*(\varepsilon)=5\frac{k_1^2}{\varepsilon}(p+1)(p+2)\cot\theta_{\mathcal{T}}.$$

Lemma 9 *Under the notation of Theorem 7, the continuity constant $\tilde{C}$ of Lemma 3 is given by:*

$$\tilde{C}=\max_{e\in\Gamma_h\cup\Gamma_D}\{1+\frac{|e|^{\beta_0-1}}{\sigma_e^0},\quad 1+\frac{5k_1^2}{k_0}(p+1)(p+2)\cot\theta_{\mathcal{T}}\}.$$

3.3 Estimation of σ^* in Three Dimensions

Theorem 10 *Let $\theta_{\mathcal{T}}$ denote the dihedral angle such that $\sin\theta_{\mathcal{T}}$ has the smallest value over all tetrahedrons in the subdivision. For any $\varepsilon>0$, define*

$$\sigma_e^*(\varepsilon)=\frac{21}{4}\frac{k_1^2}{\varepsilon}(p+1)(p+3)h|\cot\theta_{\mathcal{T}}||e|^{\beta_0-1}. \tag{39}$$

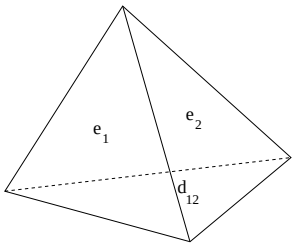

Fig. 2. A tetrahedral element with faces e_i.

Then for any $0 < \varepsilon < k_0$, *if* $\sigma_e > \sigma_e^*(\varepsilon)$ *for all faces* e, *there is a constant* $0 < C^*(\varepsilon) < 1$, *independent of* h, *such that*

$$\forall v_h \in \mathcal{D}_p(\mathcal{T}_h), \quad A(v_h, v_h) \geq C^* \|v_h\|_{\mathcal{E}}^2.$$

An expression for C^* *is*

$$C^*(\varepsilon) = \min_{e \in \Gamma_h \cup \Gamma_D} \{1 - \frac{\varepsilon}{k_0}, \quad 1 - \frac{21}{4\varepsilon\sigma_e} k_1^2 (p+1)(p+3) |\cot \theta_{\mathcal{T}}| |e|^{\beta_0 - 1}\}.$$

Proof: The proof is similar to the one for the two-dimensional case, and thus we will skip some technical details. We first recall the inverse inequality in 3D for a tetrahedral element E with face e [20]:

$$\forall v_h \in \mathbb{P}_p(E), \qquad \|v_h\|_e \leq \sqrt{\frac{(p+1)(p+3)}{3} \frac{|e|}{|E|}} \|v_h\|_E. \tag{40}$$

Here, $|e|$ is the area of the face and $|E|$ is the volume of the tetrahedral element. So as in the case of the triangle we need to estimate the ratio $\frac{|e|}{|E|}$. For this, we fix an element E in $\mathcal{T}_h$ and we denote by $e_i, i = 1, \ldots, 4$ the faces of E and by d_{ij} the common edge to faces e_i and e_j. We will assume that the face e is denoted by e_4. We also denote by θ_{ij} the dihedral angle between faces e_i and e_j. A schematic is given in Fig. 2. The volume of the tetrahedron is given by the formula [15]:

$$|E| = \frac{2}{3|d_{ij}|} |e_i||e_j| \sin \theta_{ij}, \tag{41}$$

therefore we can rewrite the volume as:

$$\begin{aligned} |E| &= \frac{1}{3}\Big(\frac{2}{3|d_{14}|}|e_4||e_1| \sin\theta_{14} + \frac{2}{3|d_{24}|}|e_4||e_2| \sin\theta_{24} + \frac{2}{3|d_{34}|}|e_4||e_3| \sin\theta_{34}\Big) \\ &= \frac{2}{9}|e_4|\Big(\frac{|e_1|}{d_{14}} \sin\theta_{14} + \frac{|e_2|}{d_{24}} \sin\theta_{24} + \frac{|e_3|}{d_{34}} \sin\theta_{34}\Big). \end{aligned} \tag{42}$$

Hence, using the fact that $|d_{ij}| \leq h$, we have :

$$\frac{|e|}{|E|} = \frac{|e_4|}{|E|} = \frac{|e_4|}{\frac{2}{9}|e_4|\Big(\frac{|e_1|}{|d_{14}|} \sin\theta_{14} + \frac{|e_2|}{|d_{24}|} \sin\theta_{24} + \frac{|e_3|}{|d_{34}|} \sin\theta_{34}\Big)}$$

$$\leq \frac{9}{2|e_4|}\frac{|e_4|}{\left(\frac{|e_1|}{h}\sin\theta_{14}+\frac{|e_2|}{h}\sin\theta_{24}+\frac{|e_3|}{h}\sin\theta_{34}\right)}$$

$$\leq \frac{9}{2}\frac{h}{|e_4|}\frac{|e_4|}{\left(|e_1|\sin\theta_{14}+|e_2|\sin\theta_{24}+|e_3|\sin\theta_{34}\right)}. \tag{43}$$

The relation between areas of the faces and dihedral angles in general tetrahedron is given by the formula [15]:

$$|e_k| = \sum_{\substack{i\neq k\\ i=1}}^{4} |e_i|\cos\theta_{ki}. \tag{44}$$

Hence we have using (44) in (43) and using dihedral angle $\theta_{\mathcal{T}}$ defined above:

$$\frac{|e|}{|E|} \leq \frac{9}{2}\frac{h}{|e_4|}\Big(\frac{|e_1|\cos\theta_{14}+|e_2|\cos\theta_{24}+|e_3|\cos\theta_{34}}{|e_1|\sin\theta_{14}+|e_2|\sin\theta_{24}+|e_3|\sin\theta_{34}}\Big)$$

$$\leq \frac{9}{2}\frac{h}{|e_4|}\Big(\frac{|e_1||\cos\theta_{\mathcal{T}}|+|e_2||\cos\theta_{\mathcal{T}}|+|e_3||\cos\theta_{\mathcal{T}}|}{|e_1|\sin\theta_{\mathcal{T}}+|e_2|\sin\theta_{\mathcal{T}}+|e_3|\sin\theta_{\mathcal{T}}}\Big).$$

Therefore we obtain the following estimate for a given face e in tetrahedral element E:

$$\frac{|e|}{|E|} \leq \frac{9}{2}\frac{h|\cot\theta_{\mathcal{T}}|}{|e|}, \tag{45}$$

which is similar to estimate (31). Using a similar argument as in the triangular case, we obtain for the interior face e of the tetrahedral element E_1 and E_2:

$$\int_e \{K\nabla v\cdot\mathbf{n}_e\}[v] \leq k_1\sqrt{\frac{3(p+1)(p+3)}{8|e|}h|\cot\theta_{\mathcal{T}}|}\Big(\|\nabla v\|_{E_1}+\|\nabla v\|_{E_2}\Big)\|[v]\|_e. \tag{46}$$

and for the boundary face we have :

$$\int_e \{K\nabla v\cdot\mathbf{n}_e\}[v] \leq k_1\sqrt{\frac{3(p+1)(p+3)}{2|e|}h|\cot\theta_{\mathcal{T}}|}\|\nabla v\|_E\|[v]\|_e \tag{47}$$

Therefore we can estimate now the term $\sum_{e\in\Gamma_h\cup\Gamma_D}\int_e\{K\nabla v\cdot\mathbf{n}_e\}[v]$. We first apply a discrete Cauchy-Schwarz's inequality, then we decompose the subdivision into disjoint sets $\mathcal{T}_0, \mathcal{T}_{1D}, \mathcal{T}_{2D}, \mathcal{T}_{3D}, \ldots$ as in the 2D case. The greatest coefficient corresponds to the case of a tetrahedron with three Dirichlet boundary faces. Therefore, it is easy to see that we obtain for any $\varepsilon > 0$:

$$\sum_{e\in\Gamma_h\cup\Gamma_D}\int_e\{K\nabla v\cdot\mathbf{n}_e\}[v] \leq \frac{\varepsilon}{2k_0}\sum_{E\in\mathcal{T}_h}\int_E K(\nabla v)^2+$$

$$\frac{21}{8\varepsilon}\sum_{e\in\Gamma_h\cup\Gamma_D}\frac{k_1^2(p+1)(p+3)h|\cot\theta_{\mathcal{T}}||e|^{\beta_0-1}}{|e|^{\beta_0}}\int_e[v]^2. \tag{48}$$

Therefore using the estimate (48) we have the following bound for the bilinear form (25):

$$A(v,v)\geq\left(1-\frac{\varepsilon}{k_0}\right)\sum_{E\in\mathcal{T}_h}\|K^{\frac{1}{2}}(\nabla v)\|_E^2+$$

$$\sum_{e\in\Gamma_h\cup\Gamma_D}\frac{\sigma_e-\frac{21}{4\varepsilon}k_1^2(p+1)(p+3)h|\cot\theta_{\mathcal{T}}||e|^{\beta_0-1}}{|e|^{\beta_0}}\|[v]\|_e^2. \tag{49}$$

Coercivity is then obtained for ε and σ_e satisfying the bounds:

$$\varepsilon<k_0, \tag{50}$$

$$\sigma_e>\frac{21}{4}\frac{k_1^2}{\varepsilon}(p+1)(p+3)h|\cot\theta_{\mathcal{T}}||e|^{\beta_0-1}. \tag{51}$$

This concludes the proof. □

Corollary 11 *Assume that no super-penalization is used, namely $\beta_0=1$, then the estimate becomes:*

$$\sigma_e^*=\frac{21}{4}\frac{k_1^2}{k_0}(p+1)(p+3)h|\cot\theta_{\mathcal{T}}|.$$

Lemma 12 *Under the notation of Theorem 10, the continuity constant $\tilde{C}$ of Lemma 3 is given by:*

$$\tilde{C}=\max_{e\in\Gamma_h\cup\Gamma_D}\{1+\frac{|e|^{\beta_0-1}}{\sigma_e^0},\quad 1+\frac{21}{4}\frac{k_1^2}{k_0}(p+1)(p+2)h|\cot\theta_{\mathcal{T}}|\}.$$

4 Numerical examples

We now present simple computations obtained for the domains $\Omega_1,\Omega_2,\Omega_3$ in 1D, 2D and 3D respectively. The exact solutions are periodic functions defined by:

$$\begin{aligned}
u_1(x)&=\cos(8\pi x)\quad\text{on}\quad\Omega_1=(0,1),\\
u_2(x)&=\cos(8\pi x)+\cos(8\pi y)\quad\text{on}\quad\Omega_2=(0,1)^2,\\
u_3(x)&=\cos(8\pi x)+\cos(8\pi y)+\cos(8\pi z)\quad\text{on}\quad\Omega_3=(0,1)^3.
\end{aligned}$$

The tensor K is the identity tensor. We fix $\beta_0=1$. We vary the number of elements $\mathcal{N}_h$ in the mesh, the polynomial degree and the penalty value (denoted by σ) that is chosen, for simplicity, constant over the whole domain. In each case, we precise the limiting penalty value $\sigma^{**}=\sigma^*(\varepsilon^*)$.

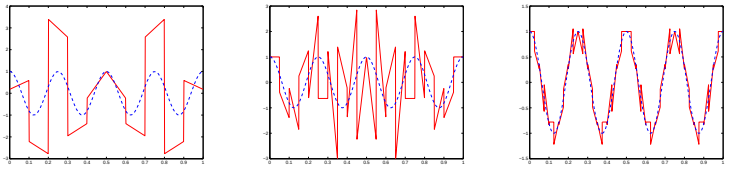

Fig. 3. $p = 1$, $\sigma = 0.5$: $\mathcal{N}_h = 10$ (left), $\mathcal{N}_h = 20$ (center), $\mathcal{N}_h = 40$ (right).

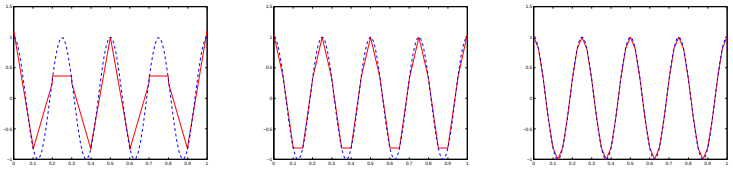

Fig. 4. $p = 1$, $\sigma = 16.5$: $\mathcal{N}_h = 10$ (left), $\mathcal{N}_h = 20$ (center), $\mathcal{N}_h = 40$ (right).

4.1 One-dimensional Problem

We first consider the case of piecewise linears on several meshes containing 10, 20 and 40 intervals respectively. In all figures, the exact solution is drawn as a dashed line whereas the numerical solution is drawn as a solid line. For a penalty value $\sigma = 0.5$ that is smaller than $\sigma^{**} = 16$, oscillations occur for all three meshes (see Fig. 3) and the numerical error is large. When $\sigma > \sigma^{**}$, the numerical solution is accurate (see Fig. 4). The two curves coincide with each other. The errors decrease as the mesh is refined according to the theoretical convergence rate given in Theorem 4.

We repeat the numerical experiments with piecewise quadratics and piecewise cubics. Unstable solutions are obtained for penalty values below the threshold value (see Fig. 5 and Fig. 7). The stable and convergent solutions are shown in Fig. 6 and Fig. 8. It is interesting to point that for the unstable penalty $\sigma = 3.5832$, the solution is accurate for the mesh with 20 elements; however large oscillations occur on meshes with 10 and 40 elements. Finally, Fig. 9 corresponds to a zero penalty on a coarse mesh and a very fine mesh: as expected, refining the mesh is not enough to recover from the loss of coercivity.

A more precise estimate of the accuracy is given in Table 4.1. The absolute L^2 error $\|u - u_h\|_\Omega$ and H^1 error $(\sum_{E \in \mathcal{T}_h} (\|\nabla(u - u_h)\|_E^2 + \|u - u_h\|_E^2))^{1/2}$ are computed for each simulation. We also indicate the limiting penalty values σ_n^{**} for all $n = 0, \ldots, N$. For stable solutions, we choose penalty values that are greater than the limiting value. It is to be noted that when σ is very close to the limiting value σ_n^{**}, the coercivity constant C^* is very close to zero. In that case, numerical oscillations could still occur. This poor coercivity property is discussed in detail in [10].

Table 1
Numerical errors for one-dimensional simulations.

$\mathcal{N}_h$	p	σ_n	σ_n^{**} $0<n<N$	σ_n^{**} $n=0,N$	L^2 error	H^1 error
10	1	0	16	12	251.7794	267.3055
160	1	0	16	12	1.5748	2.6545
10	1	0.5	16	12	1.4784	19.2168
10	1	16.5	16	12	0.3167	11.6277
20	1	0.5	16	12	1.1143	40.2165
20	1	16.5	16	12	0.0931	6.2879
40	1	0.5	16	12	0.1334	9.7613
40	1	16.5	16	12	0.0247	3.2048
10	2	1.375	36	27	0.3166	13.8899
10	2	37	36	27	0.0558	3.8357
20	2	1.375	36	27	0.2620	22.1213
20	2	37	36	27	0.0073	1.0252
40	2	1.375	36	27	0.1265	21.1474
40	2	37	36	27	9.1352×10^{-4}	0.2606
10	3	3.5832	64	48	0.1111	9.4335
10	3	65	64	48	0.0082	0.8264
20	3	3.5832	64	48	0.0072	1.2450
20	3	65	64	48	5.5723×10^{-4}	0.1093
40	3	3.5832	64	48	1.3497	467.8908
40	3	65	64	48	3.5977×10^{-5}	0.0138

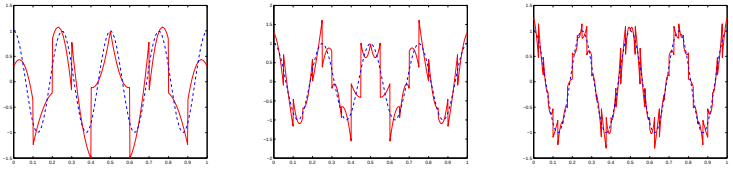

Fig. 5. $p = 2$, $\sigma = 1.375$: $\mathcal{N}_h = 10$ (left), $\mathcal{N}_h = 20$ (center), $\mathcal{N}_h = 40$ (right).

4.2 Two-dimensional Problem

We first explain how to obtain the angle $\theta_{\mathcal{T}}$. This angle will give the largest $\cos\theta$, or the largest $\cot\theta$ over all triangle angles θ. For a given element E, we

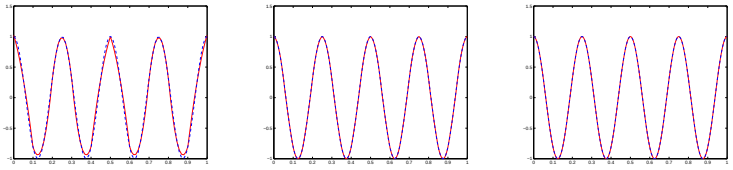

Fig. 6. $p = 2$, $\sigma = 37$: $\mathcal{N}_h = 10$ (left), $\mathcal{N}_h = 20$ (center), $\mathcal{N}_h = 40$ (right).

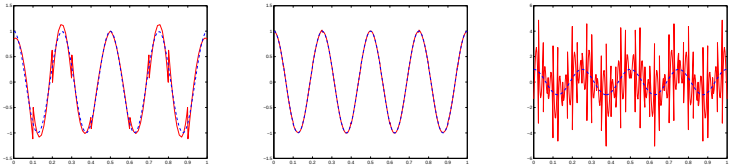

Fig. 7. $p = 3$, $\sigma = 3.5832$: $\mathcal{N}_h = 10$ (left), $\mathcal{N}_h = 20$ (center), $\mathcal{N}_h = 40$ (right).

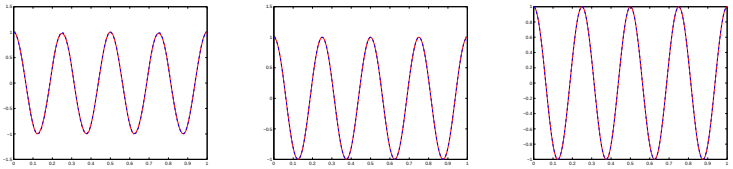

Fig. 8. $p = 3$, $\sigma = 65$: $\mathcal{N}_h = 10$ (left), $\mathcal{N}_h = 20$ (center), $\mathcal{N}_h = 40$ (right).

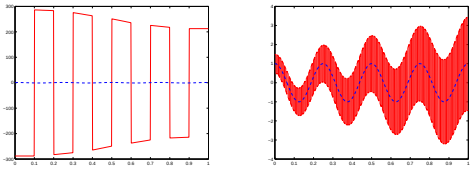

Fig. 9. $p = 1$, $\sigma = 0$: coarse mesh $\mathcal{N}_h = 10$ (left) and refined mesh $\mathcal{N}_h = 160$ (right).

compute a value $\cot\theta_E$ defined by:

(1) Compute the lengths of the edges of E from the vertices coordinates (x_i^E, y_i^E):

$$
\begin{aligned}
|e_1| &= \Big(x_2^E - x_1^E)^2 + (y_2^E - y_1^E)^2\Big)^{1/2} \\
|e_2| &= \Big(x_3^E - x_1^E)^2 + (y_3^E - y_1^E)^2\Big)^{1/2} \\
|e_3| &= \Big(x_2^E - x_3^E)^2 + (y_2^E - y_3^E)^2\Big)^{1/2}
\end{aligned}
$$

(2) Determine the smallest length, say $|e_{i_1}|$. Denote the other two lengths by $|e_{i_2}|$ and $|e_{i_3}|$.

(3) Compute $\cot\theta_E$:

$$
\cos\theta_E = \frac{|e_{i_2}|^2 + |e_{i_3}|^2 - |e_{i_1}|^2}{2|e_{i_1}||e_{i_2}|}, \quad \sin\theta_E = (1-(\cos\theta_E)^2)^{1/2}, \quad \cot\theta_E = \frac{\cos\theta_E}{\sin\theta_E}
$$

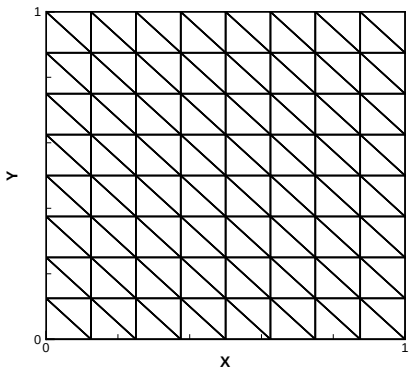

Fig. 10. Structured mesh with 128 elements.

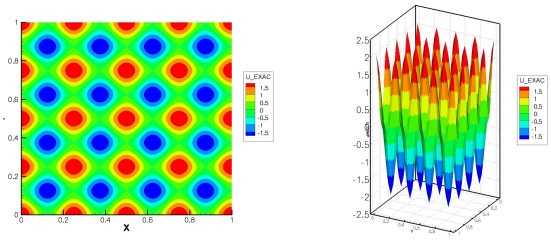

Fig. 11. Exact solution: two-dimensional view (left) and three-dimensional view (right)

The value $\cot\theta_{\mathcal{T}}$ is the maximum of $\cot\theta_E$ over all mesh elements E.

We solve the problem on structured meshes as shown in Fig. 10. For this mesh, the smallest angle is $\theta_{\mathcal{T}} = \frac{\pi}{4}$. The exact solution for reference is shown in Fig. 11. In Fig. 12 and 13, we first consider polynomial degree equal to one on a very fine mesh (2048 elements): the penalty parameter takes the values $0, 3$ and 50. In this case, the limiting value is $\sigma^{**} = 30$. For a penalty value above the limiting value, no oscillations occur whereas for a penalty value below σ^{**}, the solution is unstable. Fig. 14 and 15 show the solution for polynomial degree 2 on a mesh containing 128 elements. We then refine the mesh (512 elements) and obtain Fig. 16 and 17. Finally, for the case of piecewise cubic polynomials, the solutions are shown in Fig. 18 and 19 for a mesh containing 32 elements, and in Fig. 20 and 20 for a mesh containing 128 elements.

We give the error in the L^2 norm for all cases and we also give the limiting value σ^{**} in Table 4.2. For a given penalty, the error decreases as the mesh is refined. Similar conclusions as in the one-dimensional case can be made. For stable method, the error decreases with the right convergence rate. For unstable method, oscillations may occur.

Table 2
Numerical errors for two-dimensional simulations.

$\mathcal{N}_h$	p	σ_e	σ^{**}	L^2 error	H_0^1 error
2048	1	0	30	1.6208681	7.9950783
2048	1	3	30	9.7490787×10^{-1}	1.8162526×10^{2}
2048	1	50	30	5.0342187×10^{-2}	5.5419916
128	2	0	60	$2.7842956e$	1.1398348×10^{2}
128	2	4.5	60	4.9175469	2.8412357×10^{2}
128	2	100	60	1.7501758×10^{-1}	10.840289
512	2	0	60	5.2324755×10^{-2}	4.3847913
512	2	4.5	60	1.7144388×10^{-1}	20.100636
512	2	100	60	1.7976364×10^{-2}	2.1627031
32	3	0	100	4.6476606×10^{-1}	6.6199869
32	3	11	100	52.643677	3.2830936×10^{3}
32	3	150	100	4.7347586×10^{-1}	10.663237
128	3	0	100	7.8099710×10^{-3}	6.0682964×10^{-1}
128	3	11	100	2.1133410×10^{-1}	25.599158
128	3	150	100	6.2219837×10^{-3}	4.8017076×10^{-1}

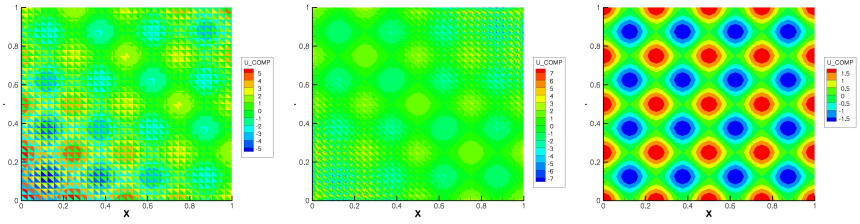

Fig. 12. Two-dimensional view for $p = 1$, $\mathcal{N}_h = 2048$: $\sigma = 0$ (left), $\sigma = 3$ (center), $\sigma = 55$ (right).

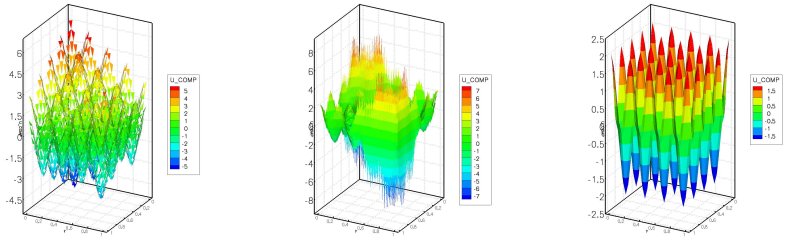

Fig. 13. Three-dimensional view for $p = 1$, $\mathcal{N}_h = 2048$: $\sigma = 0$ (left), $\sigma = 3$ (center), $\sigma = 55$ (right).

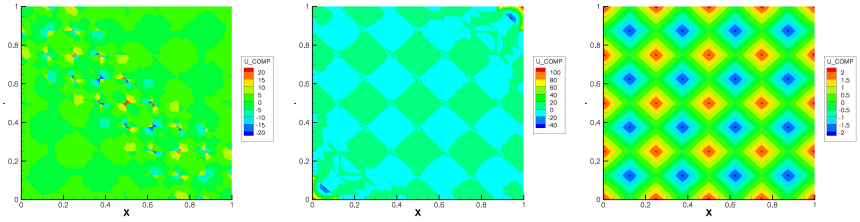

Fig. 14. Two-dimensional view for $p = 2$, $\mathcal{N}_h = 128$: $\sigma = 0$ (left), $\sigma = 4.5$ (center), $\sigma = 100$ (right).

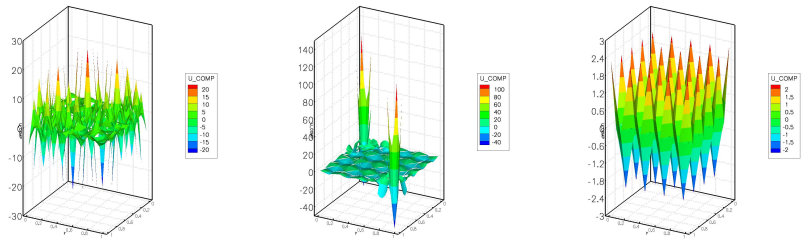

Fig. 15. Three-dimensional view for $p = 2$, $\mathcal{N}_h = 128$: $\sigma = 0$ (left), $\sigma = 4.5$ (center), $\sigma = 100$ (right).

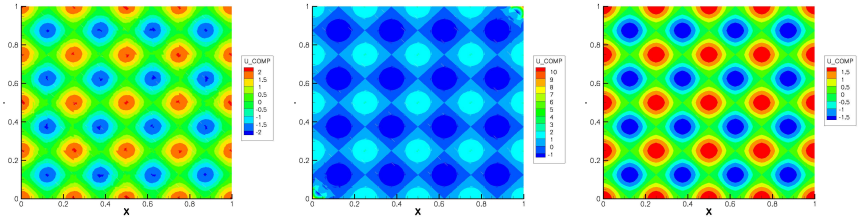

Fig. 16. Two-dimensional view for $p = 2$, $\mathcal{N}_h = 512$: $\sigma = 0$ (left), $\sigma = 4.5$ (center), $\sigma = 100$ (right).

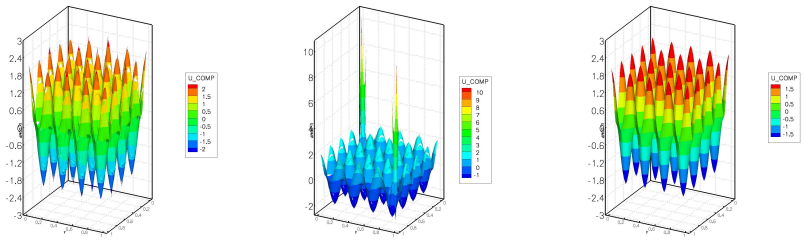

Fig. 17. Three-dimensional view for $p = 2$, $\mathcal{N}_h = 512$: $\sigma = 0$ (left), $\sigma = 4.5$ (center), $\sigma = 100$ (right).

4.3 Unstructured 2D mesh

We consider an unstructured coarse mesh shown in Fig. 22. This mesh contains 219 triangles and 876 triangles after uniform refinement. We only present some results for the case of piecewise quadratic approximations. As before we vary the penalty parameters $\sigma = 0, 7, 5, 150$. The limiting penalty value is

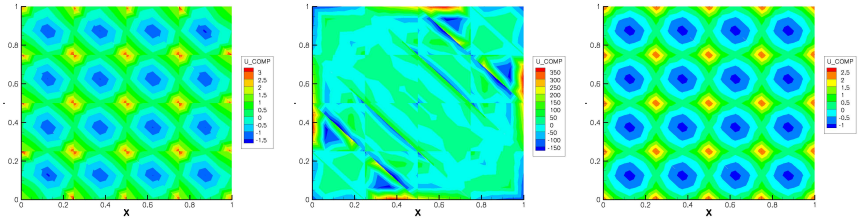

Fig. 18. Two-dimensional view for $p = 3$, $\mathcal{N}_h = 32$: $\sigma = 0$ (left), $\sigma = 11$ (center), $\sigma = 150$ (right).

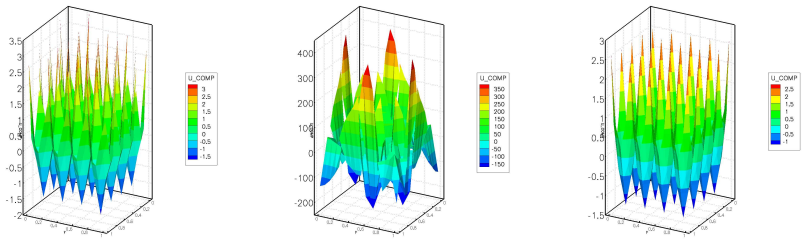

Fig. 19. Three-dimensional view for $p = 3$, $\mathcal{N}_h = 32$: $\sigma = 0$ (left), $\sigma = 11$ (center), $\sigma = 150$ (right).

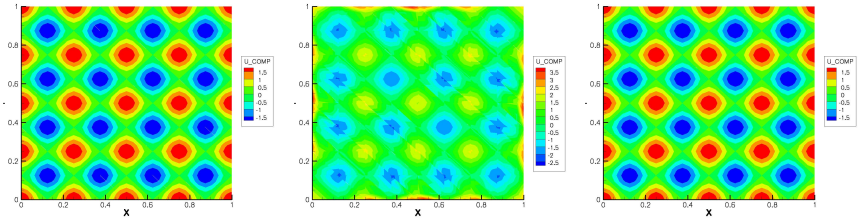

Fig. 20. Two-dimensional view for $p = 3$, $\mathcal{N}_h = 128$: $\sigma = 0$ (left), $\sigma = 11$ (center), $\sigma = 150$ (right).

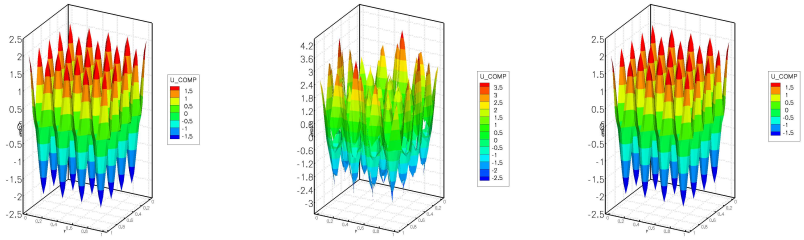

Fig. 21. Three-dimensional view for $p = 3$, $\mathcal{N}_h = 128$: $\sigma = 0$ (left), $\sigma = 11$ (center), $\sigma = 150$ (right).

$\sigma^{**} = 129.4676$. The solutions on the coarse mesh are shown in Fig. 23 and 24 whereas the solutions on a refined mesh are shown in Fig. 25 and 26.

We give the error in the L^2 and H_0^1 norms for all cases in Table 4.3.

We present in Fig. 27 the numerical convergence of the SIPG solution for a "good" penalty value (larger than σ^{**}) and a "bad" penalty value (smaller

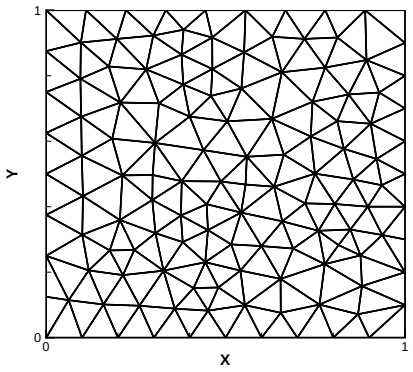

Fig. 22. Unstructured mesh with 219 elements.

Table 3
Numerical errors for two-dimensional unstructured mesh simulations.

$\mathcal{N}_h$	p	σ_e	σ^{**}	L^2 error	H_0^1 error
219	2	0	129.4676	1.0262113	52.510991
219	2	7.5	129.4676	6.3221136×10^{-1}	66.159341
219	2	150	129.4676	5.9683013×10^{-2}	4.9992556
876	2	0	129.4676	5.5677943×10^{-2}	5.8047835
876	2	7.5	129.4676	2.2284393×10^{-2}	4.3895847
876	2	150	129.4676	8.0261140×10^{-3}	1.2883024

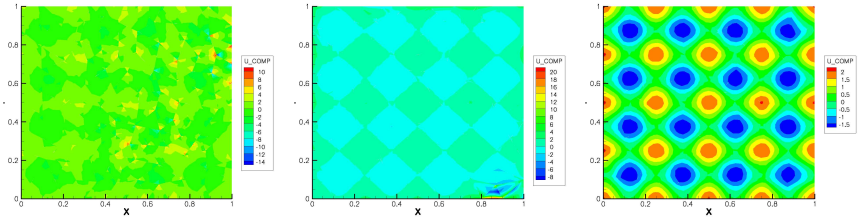

Fig. 23. Two-dimensional view for unstructured mesh and $p = 2$, $\mathcal{N}_h = 219$: $\sigma = 0$ (left), $\sigma = 7.5$ (center), $\sigma = 150$ (right).

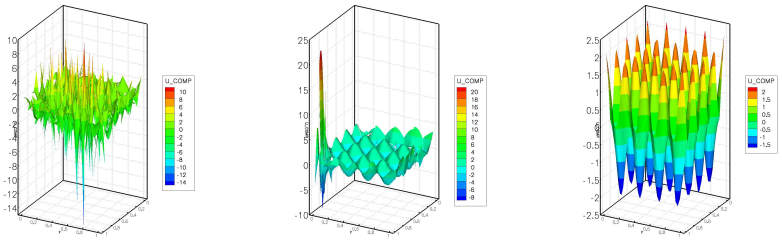

Fig. 24. Three-dimensional view for unstructured mesh and $p = 2$, $\mathcal{N}_h = 219$: $\sigma = 0$ (left), $\sigma = 7.5$ (center), $\sigma = 150$ (right).

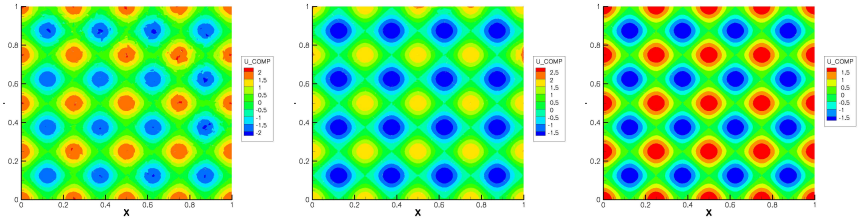

Fig. 25. Two-dimensional view for unstructured mesh and $p = 2$, $\mathcal{N}_h = 876$: $\sigma = 0$ (left), $\sigma = 7.5$ (center), $\sigma = 150$ (right).

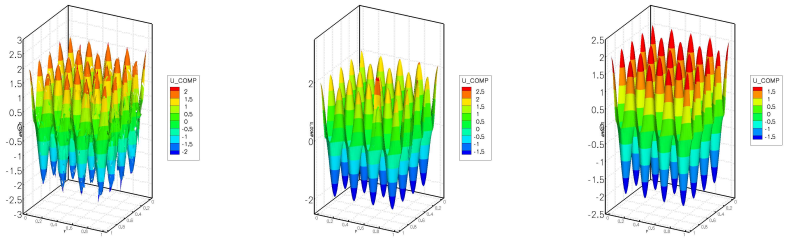

Fig. 26. Three-dimensional view for unstructured mesh and $p = 2$, $\mathcal{N}_h = 876$: $\sigma = 0$ (left), $\sigma = 7.5$ (center), $\sigma = 150$ (right).

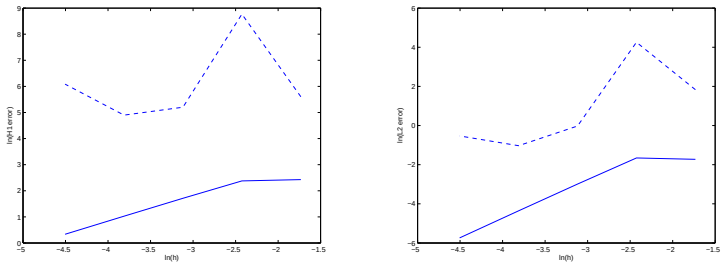

Fig. 27. Numerical convergence rates for the case $\sigma = 3$ (dashed line) and $\sigma = 50$ (solid line): H_0^1 errors (left) and L^2 errors (right).

than σ^{**}). Piecewise linear approximation is used. The stable solution converges with the expected convergence rate ($\mathcal{O}(h^2)$ for the L^2 error) whereas the unstable solution does not converge as the mesh size decreases.

4.4 Three-dimensional Problem

We first explain how to obtain the angle $\theta_{\mathcal{T}}$. The value $|\cot\theta_{\mathcal{T}}|$ is the maximum of $|\cot\theta_E|$ over all mesh elements E. For a given element E, the angle θ_E is the one that yields the smallest $\sin\theta_{E,\xi}$ over all edges ξ of the tetrahedron. We now explain how to obtain $\theta_{E,\xi}$ for given E and ξ.

(1) Compute the equations of the planes corresponding to the two faces of

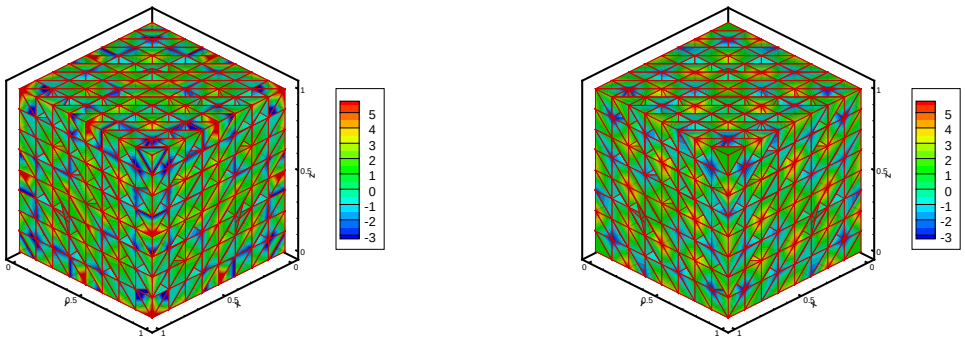

Fig. 28. Solution on tetrahedral mesh: $\sigma = 5$ (left) and $\sigma = 80$ (right).

E that share the common edge ξ.

$$\forall i = 1, 2, \quad a^i_{E,\xi} x + b^i_{E,\xi} y + c^i_{E,\xi} z + d^i_{E,\xi} = 0.$$

(2) The normal vectors to those two faces are

$$i = 1, 2, \quad \mathbf{n}_{e_i} = (a^i_{E,\xi}, b^i_{E,\xi}, c^i_{E,\xi}).$$

(3) Compute $\cos\theta_{E,\xi}$ and $\sin\theta_{E,\xi}$:

$$\cos\theta_{E,\xi} = \mathbf{n}_{e_1} \cdot \mathbf{n}_{e_2}, \quad \sin\theta_{E,\xi} = (1 - (\cos\theta_{E,\xi})^2)^{1/2}.$$

The mesh contains 8640 tetrahedral elements. Piecewise quadratic approximation is used. In Fig. 28, we show the numerical solution with penalty values $\sigma = 5$ and $\sigma = 80$. The limiting penalty value for these simulations is $\sigma^{**} = 78.75$. The absolute L^2 error is 1.5074340 for $\sigma = 5$ and 0.32264918 for $\sigma = 80$. The absolute H^1_0 error is 112.83199 for $\sigma = 5$ and 19.834390 for $\sigma = 80$.

5 Conclusions

By presenting lower bounds of the penalty parameter useful for practical computations, this paper removes one known disadvantage of the symmetric interior penalty methods, namely the fact that stability of the method is obtained for an unknown large enough penalty value. Even though we focused on the elliptic problems, our improved coercivity and continuity results can be applied to the analysis of the SIPG method for time-dependent problems.

References

[1] D.N. Arnold. *An interior penalty finite element method with discontinuous elements*. PhD thesis, The University of Chicago, 1979.

[2] D.N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM Journal on Numerical Analysis*, 19:742–760, 1982.

[3] D.N. Arnold, F. Brezzi, B. Cockburn, and L.D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM Journal on Numerical Analysis*, 39(5):1749–1779, 2002.

[4] G.A. Baker, W.N. Jureidini, and O.A. Karakashian. Piecewise solenoidal vector fields and the Stokes problem. *SIAM Journal on Numerical Analysis*, 27:1466–1485, 1990.

[5] R. Becker, P. Hansbo, and R. Stenberg. A finite element method for domain decomposition with non-matching grids. *M2AN Math. Model. Numer. Anal.*, 37:209–225, 2003.

[6] P. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, 1978.

[7] B. Cockburn, G.E. Karniadakis, and C.-W. Shu, editors. *First International Symposium on Discontinuous Galerkin Methods*, volume 11 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, 2000.

[8] C. Dawson, S. Sun, and M.F. Wheeler. Compatible algorithms for coupled flow and transport. *Comput. Meth. Appl. Mech. Engng*, 193:2565–2580, 2004.

[9] J. Douglas and T. Dupont. *Lecture Notes in Physics*, volume 58, chapter Interior penalty procedures for elliptic and parabolic Galerkin methods. Springer-Verlag, 1976.

[10] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*. Number 159 in Applied Mathematical Sciences. Springer, 2003.

[11] V. Girault, B. Rivière, and M.F. Wheeler. A discontinuous Galerkin method with non-overlapping domain decomposition for the Stokes and Navier-Stokes problems. *Mathematics of Computation*, 74:53–84, 2005.

[12] P. Houston, C. Schwab, and E. Süli. Discontinuous hp-finite element methods for advection-diffusion reaction problems. *SIAM Journal on Numerical Analysis*, 39(6):2133–2163, 2002.

[13] O.A. Karakashian and W. Jureidini. A nonconforming finite element method for the stationary Navier-Stokes equations. *SIAM Journal on Numerical Analysis*, 35:93–120, 1998.

[14] S. Kaya and B. Rivière. A discontinuous subgrid eddy viscosity method for the time-dependent Navier-Stokes equations. *SIAM Journal on Numerical Analysis*, 43(4):1572–1595, 2005.

[15] J.R. Lee. The law of cosines in a tetrahedron. *J. Korea Soc. Math. Ed. Ser. B: Pure Appl. Math.*, 4:1–6, 1997.

[16] B. Rivière and V. Girault. Discontinuous finite element methods for incompressible flows on subdomains with non-matching interfaces. *Computer Methods in Applied Mechanics and Engineering*, 2005. to appear.

[17] B. Rivière, M.F. Wheeler, and V. Girault. A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems. *SIAM Journal on Numerical Analysis*, 39(3):902–931, 2001.

[18] B. Rivière and I. Yotov. Locally conservative coupling of Stokes and Darcy flow. *SIAM Journal on Numerical Analysis*, 42(5):1959–1977, 2005.

[19] S. Sun and M.F. Wheeler. Symmetric and nonsymmetric discontinuous galerkin methods for reactive transport in porous media. *SIAM Journal on Numerical Analysis*, 43(1):195–219, 2005.

[20] T. Warburton and J.S. Hesthaven. On the constants in hp-finite element trace inverse inequalities. *Comput. Methods Appl. Mech. Engrg.*, 192:2765–2773, 2003.

[21] M.F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM Journal on Numerical Analysis*, 15(1):152–161, 1978.