A PENALTY FREE NON-SYMMETRIC NITSCHE TYPE METHOD FOR THE WEAK IMPOSITION OF BOUNDARY CONDITIONS

ERIK BURMAN*

Abstract. In this note we show that the non-symmetric version of the classical Nitsche's method for the weak imposition of boundary conditions is stable without penalty term. We prove optimal H^1-error estimates and L^2-error estimates that are suboptimal with half an order in h. Both the pure diffusion and the convection–diffusion problems are discussed.

1. Introduction. In his seminal paper from 1971, [15], Nitsche proposed a consistent penalty method for the weak imposition of boundary conditions. The formulation proposed was symmetric so as to reflect the symmetry of the underlying Poisson problem. Stability was obtained thanks to a penalty term, with a penalty parameter that must satisfy a lower bound to ensure coercivity.

A non-symmetric version of Nitsche's method was later proposed by Freund and Stenberg [9] and it was noted that this method did not need the lower bound for stability. The penalty term however could not be omitted, since coercivity fails, and error estimates degenerate as the penalty parameter goes to zero. The non-symmetric version of Nitsche's method was then proposed as a discontinuous Galerkin (DG) method by Oden et al., [16] and it was proven that the non-symmetric version was stable for polynomial orders $k \geq 2$, by Girault et al. [17] and Larson and Niklasson [14]. In [14] stability for the penalty free case is proved using an inf-sup argument that relies on the important number of degree's of freedom available in high order DG-methods.

To the best of our knowledge no similar results have been proven for the non-symmetric version of Nitsche's method for the imposition of boundary conditions when continuous approximation spaces are used. Indeed in this case the DG-analysis does not work since polynomials may not be chosen independently on different elements because of the continuity constraints. Weak impositition of boundary conditions has been advocated by Hughes et al. for turbulence computations of LES-type in [1]. They showed that the mean flow in the boundary layer was more accurately captured using weakly rather than strongly imposed boundary conditions. They also noted that the non-symmetric Nitsche's method appears stable without penalty [13].

In applications there is interest in reducing the number of free parameters used without increasing the number of degrees of freedom needed for the coupling, see [10] for a discussion. From this point of view a penalty free Nitsche method is a welcome addition to the computational toolbox, in particular for flow problems where the system matrix is non-symmetric anyway, because of the convection terms. It has no penalty parameter and does not make use of Lagrange multipliers.

Numerical evidence also suggests that the unpenalized non-symmetric Nitsche type method has some further interesting properties. When using iterative solution methods in domain decomposition it has been shown to have more favorable convergence properties compared to the symmetric method [8]. For the solution of Cauchy-type inverse problem using steepest descent type algorithms it has been shown numerically to have superior convergence properties in the initial phase of the iterations compared to the symmetric version or strongly imposed conditions, in spite of

*Department of Mathematics, University of Sussex, Brighton, UK–BN1 9QH, United Kingdom; (E.N.Burman@sussex.ac.uk)

the lack of dual consistency.

In view of this the question naturally arises if the penalty free method is sound, or if it can fail under unfortunate circumstances.

In this paper we prove for the Poisson problem that the non-symmetric Nitsche's method is indeed stable and optimally convergent in the H^1-norm for polynomial orders $k \geq 1$. We also show that in this case, the convergence rate of the error in the L^2-norm is suboptimal with only half a power of h. Hence the non-optimality for the non-symmetric Nitsche's method for continuous Galerkin methods is not as important as for DG-methods (see [16] and [11] for numerical evidence of the sub-optimal behavior in this latter case).

We then show how the results may be applied in the case of convection–diffusion equations, considering first the Streamline–diffusion method and then outlining how the results may be extended to the case of the Continuous interior penalty method.

Nitsche's method however has some stabilizing properties of its own, in particular for outflow layers, this phenomenon was analyzed in [18] and illustrated herein with a numerical example. This makes the non-symmetric Nitsche's method an appealing, parameter free, method for flow problems where the system matrix is non-symmetric and the use of stabilized methods usually also results in the loss of half a power of h. It should be noted however that the smallest error in the L^2-norm is obtained with the formulation using penalty on the boundary, as illustrated in the numerical section. So we do not claim that the penalty free method is the most accurate.

We only prove the result in the case of the imposition of boundary conditions but the extensions of the results to the domain decomposition case of [2] or the fictitious domain method of [4] are straightforward using similar techniques as below. Also note that since the main aim of the present paper is the study of weak imposition of boundary conditions, we will assume that the reader has basic understanding of the techniques for analyzing stabilized finite element methods and some arguments are only sketched.

For the sake of clarity, we first prove the main result on the pure diffusion problem and then discuss the extension of our result to the case of convection– diffusion problems. The paper is ended with some numerical examples.

2. The pure diffusion problem. Let Ω be a bounded domain in $\mathbb{R}^2$, with polygonal boundary $\partial\Omega$. Wherever H^2-regularity of the exact solution is needed we also assume that Ω is convex. Let $\{\Gamma_i\}_i$ denote the faces of the polygonal such that $\partial\Omega = \cup_i \Gamma_i$. The Poisson equation that we propose as a model problem is given by

$$\begin{array}{rcll} -\Delta u &=& f & \text{in } \Omega, \\ u &=& g & \text{on } \partial\Omega, \end{array} \tag{2.1}$$

where $f \in L^2(\Omega)$ and $g \in H^{1/2}(\partial\Omega)$ or $g \in H^{3/2}(\partial\Omega)$.

We have the following weak formulation: find $u \in V_g$ such that

$$a(u,v) = (f,v)_\Omega, \quad \forall v \in V_0, \tag{2.2}$$

where $(x,y)_\Omega$ denotes the L^2-scalar product over Ω,

$$V_g := \{v \in H^1(\Omega) : v|_{\partial\Omega} = g\}$$

and

$$a(u,v) := (\nabla u, \nabla v)_\Omega.$$

This problem is well-posed by the Lax-Milgram's lemma, using the standard arguments to account for non-homogeneous boundary conditions. The H^1-stability, $\|u\|_{H^1(\Omega)} \leq C_{R1}(\|f\| + \|g\|_{H^{1/2}(\partial\Omega)})$ holds and under the assumptions on Ω, f and g there holds $\|u\|_{H^2(\Omega)} \leq C_{R2}(\|f\| + \|g\|_{H^{3/2}(\partial\Omega)})$. Here we let $\|x\| := \|x\|_{L^2(\Omega)}$. Below C will be used as a generic constant that may change at each occasion, is independent on h, but not necessarily of the local mesh geometry. We will also use the notation $a \lesssim b$ for $a \leq Cb$.

3. The finite element formulation. Let $\{\mathcal{T}_h\}$ denote a family of quasi uniform and shape regular triangulations fitted to Ω, indexed by the mesh-parameter h. (It is straightforward to lift the quasi uniformity assumption, at the expense of some standard technicalities and readability.) The triangles of $\mathcal{T}_h$ will be denoted K and their diameter $h_K := \operatorname{diam}(K)$. The interior of a set P will be denoted $\mathring{P}$. For a given $\mathcal{T}_h$ the mesh-parameter is determined by $h := \max_{K\in\mathcal{T}_h} h_K$. Shape regularity is expressed by the existence of a constant $c_\rho \in \mathbb{R}$ for the family of triangulations such that, with ρ_K the radius of the largest ball inscribed in an element K, there holds,

$$\frac{h_K}{\rho_K} \leq c_\rho, \forall K \in \mathcal{T}_h.$$

For technical reasons, and to avoid the treatment of special cases, we assume that for all i, Γ_i contains no less than five element faces.

We introduce the standard finite element space of continuous piece wise polynomial functions

$$V_h^k := \{v_h \in H^1(\Omega) : v_h|_K \in \mathbb{P}_k(K), \quad \forall K \in \mathcal{T}_h\},\ k \geq 1,$$

where $\mathbb{P}_k(K)$ denotes the space of polynomials of degree less than or equal to k on the element K. The finite element formulation that we consider then takes the form, find $u_h \in V_h^k$ such that

$$a_h(u_h, v_h) = (f, v_h)_\Omega + \langle g, \nabla v_h \cdot n\rangle_{\partial\Omega} \quad \forall v_h \in V_h^k, \tag{3.1}$$

where $\langle x, y\rangle_{\partial\Omega}$ denotes the L^2-scalar product over the boundary of Ω and

$$a_h(u_h, v_h) := a(u_h, v_h) - \langle \nabla u_h \cdot n, v_h\rangle_{\partial\Omega} + \langle u_h, \nabla v_h \cdot n\rangle_{\partial\Omega}\,. \tag{3.2}$$

Note that in the classical non-symmetric Nitsche's method we also add a penalty term of the form

$$\sum_K \left\langle \gamma h_K^{-1} u_h, v_h\right\rangle_{\partial\Omega\cap\partial K} \tag{3.3}$$

and modify the second term of the right hand side accordingly

$$\sum_K \left\langle g, \gamma h_K^{-1} v_h + \nabla v_h \cdot n\right\rangle_{\partial\Omega\cap\partial K}.$$

The key observation of the present work is that the penalty parameter γ may be chosen to be zero without loss of neither stability nor accuracy.

Inserting the exact solution u into the fomulation (3.1) and integrating by parts immediately leads to the following consistency relation.

Lemma 3.1. *If u is the solution of* (2.1) *and u_h is the solution of* (3.1) *then there holds*

$$a_h(u - u_h, v_h) = 0.$$

For future reference we here recall the classical trace and inverse inequalities satisfied by the spaces V_h^k.

Lemma 3.2. *(Trace inequality) There exists $C_T \in \mathbb{R}$ such that for all $v_h \in \mathbb{P}_k(K)$ and for all $K \in \mathcal{T}_h$ there holds*

$$\|v_h\|_{L^2(\partial K)} \leq C_T(h_K^{-\frac{1}{2}}\|v_h\|_{L^2(K)} + h_K^{\frac{1}{2}}\|\nabla v_h\|_{L^2(K)}).$$

Lemma 3.3. *(Inverse inequality) There exists $C_I \in \mathbb{R}$ such that for all $v_h \in \mathbb{P}_k(K)$ and for all $K \in \mathcal{T}_h$ there holds*

$$\|\nabla v_h\|_{L^2(K)} \leq C_I h_K^{-1}\|v_h\|_{L^2(K)}.$$

4. Stability. The non-symmetric Nitsche's method is positive and testing with $v_h = u_h$ immediately gives control of the H^1-seminorm of u_h. In order for the formulation to be well-posed this is not sufficient. Indeed well-posedness is a consequence of the Poincaré inequality that holds provided we have sufficient control of the trace of u_h on $\partial\Omega$. This is the role of the penalty term (3.3), it ensures that the following Poincaré inequality is satisfied

$$\|u_h\| \leq C_P\|u_h\|_{1,h}, \text{ where } \|u_h\|_{1,h}^2 := \|\nabla u_h\|^2 + \|u_h\|_{\frac{1}{2},h,\partial\Omega}^2$$

with

$$\|u_h\|_{\frac{1}{2},h,\partial\Omega}^2 := \sum_K \left\langle h_K^{-1} u_h, u_h\right\rangle_{\partial\Omega\cap\partial K}.$$

Since we have omitted the penalty term, boundary control of u_h is not an immediate consequence of testing with $v_h = u_h$. What we will show below is that control of the boundary term can be recovered by proving an inf-sup condition. Indeed the non-symmetric Nitsche's method can be interpreted as a Lagrange multiplier method where the Lagrange multiplier λ_h has been replaced by the normal gradient of the solution: $\nabla u_h \cdot n$. This interpretation of the Nitsche's method was originally proposed in [20], however without considering the inf-sup condition. In the DG-framework it was considered in [7], where equivalence was shown between a certain Lagrange-multiplier method and a certain DG-method. When Lagrange-multipliers are used to impose continuity, the system has a saddle point structure and the inf-sup condition is the standard way of proving well-posedness. Here we will follow a similar procedure, the only difference is that the solution space and the multiplier space are strongly coupled, since the latter consists simply of the normal gradients of the former. A key result is given in the following lemma where we construct a function in the test space that will allow us to control certain averages of the solution on the boundary. To this end regroup the boundary elements, i.e. the elements with either a face or a vertex on the boundary, in (closed) patches P_j, with boundary ∂P_j, $j = 1...N_P$. Let $F_j := \partial P_j \cap \partial\Omega$. We assume that the P_j are designed such that each F_j has at least four inner nodes (this is strictly necessary only if both end vertices of P_j belong to corner elements with all their vertices on the boundary). Under our assumptions on the mesh, every Γ_i contains at least one patch P_j and there exists c_1, c_2 such that for all j

$$c_1 h \leq \mathrm{meas}(F_j) \leq c_2 h. \tag{4.1}$$

The average value of a function v over F_j will be denoted by $\bar{v}^j$.

LEMMA 4.1. *For any given vector $(r_j)_{j=1}^{N_P} \in \mathbb{R}^{N_P}$ there exists $\varphi_r \in V_h^1$ such that for all $1 \leq j \leq N_P$ there holds*

$$meas(F_j)^{-1} \int_{F_j} \nabla \varphi_r \cdot n \; ds = r_j \tag{4.2}$$

and

$$\|\varphi_r\|_{1,h} \lesssim \left(\sum_{j=1}^{N_P} \|h^{\frac{1}{2}} r_j\|^2_{L^2(F_j)} \right)^{1/2}. \tag{4.3}$$

Proof. We first construct a function $\tilde{\varphi}_j$ taking the value 1 in the interior nodes of $\partial\Omega \cap \partial P_j$ and zero elsewhere. Fix j and let $\tilde{\varphi}_j \in V_h^1$ be defined, in each vertex $x_i \in \mathcal{T}_h$, by

$$\tilde{\varphi}_j(x_i) = \begin{cases} 0 & \text{for } x_i \in K \text{ such that } K \text{ has three vertices on } \partial\Omega; \\ 0 & \text{for } x_i \in \Omega \setminus \overset{\circ}{P}_j; \\ 1 & \text{for } x_i \in \overset{\circ}{F}_j. \end{cases}$$

Let

$$\Xi_j := \mathrm{meas}(F_j)^{-1} \int_{F_j} \nabla \tilde{\varphi}_j \cdot n \; \mathrm{d}s$$

and define the normalised function φ_j by

$$\varphi_j := \Xi_j^{-1} \tilde{\varphi}_j.$$

This quantity is well defined thanks to the following lower bound that holds uniformly in j and h

$$C_\Xi \leq \Xi_j h.$$

The constant C_Ξ only depends on the local geometry of the patches P_j. By definition there holds

$$\mathrm{meas}(F_j)^{-1} \int_{F_j} \nabla \varphi_j \cdot n \; \mathrm{d}s = 1 \tag{4.4}$$

and using the standard inverse inequality (Lemma 3.3)

$$\|\nabla \varphi_j\| \lesssim C_I h^{-1} \Xi_j^{-1} \|\tilde{\varphi}_j\|_{L^2(P_j)} \lesssim C_I h^{-1} \Xi_j^{-1} \mathrm{meas}(P_j)^{1/2} \lesssim C_I C_\Xi^{-1} h. \tag{4.5}$$

Now defining

$$\varphi_r := \sum_{j=1}^{N_P} r_j \varphi_j$$

we immediately see that condition (4.2) is satisfied by equation (4.4). The upper bound (4.3) follows from (4.5), the relation (4.1) and using that

$$\|\varphi_r\|^2_{\frac{1}{2},h,\partial\Omega} := \sum_{j=1}^{N_P} \|h^{-\frac{1}{2}} r_j \varphi_j\|^2_{L^2(F_j)}$$
$$\lesssim \sum_{j=1}^{N_P} h^{-1} r_j^2 \Xi_j^{-2} \|\tilde{\varphi}_j\|^2_{L^2(F_j)} \lesssim C_\Xi^{-2} \sum_{j=1}^{N_P} \|h^{\frac{1}{2}} r_j\|^2_{L^2(F_j)}.$$

□

With the help of this technical lemma it is straightforward to prove the inf-sup condition for the formulation (3.1).

THEOREM 4.2. *There exists $c_s > 0$ such that for all functions $v_h \in V_h^k$ there holds*

$$c_s \|v_h\|_{1,h} \le \sup_{w_h \in V_h^k} \frac{a_h(v_h, w_h)}{\|w_h\|_{1,h}}.$$

Proof. Recall that

$$a_h(v_h, w_h) = (\nabla v_h, \nabla w_h)_\Omega - \langle \nabla v_h \cdot n, w_h \rangle_{\partial\Omega} + \langle v_h, \nabla w_h \cdot n \rangle_{\partial\Omega}.$$

Taking $w_h = v_h$ gives

$$a_h(v_h, v_h) = \|\nabla v_h\|^2.$$

To recover control over the boundary integral we let

$$r_j = h^{-1} \bar{v}^j := h^{-1} \text{meas}(F_j)^{-1} \int_{F_j} v_h \, \mathrm{d}s \tag{4.6}$$

in the construction of φ_r in Lemma 4.1 and note that

$$\langle v_h, \nabla \varphi_r \cdot n \rangle_{\partial\Omega} = \sum_{j=1}^{N_P} \left(\|h^{-1/2} \bar{v}^j\|^2_{L^2(F_j)} + \left\langle (v_h - \bar{v}^j), \nabla \varphi_r \cdot n \right\rangle_{F_j} \right).$$

Using standard approximation,

$$\|v_h - \bar{v}^j\|_{L^2(F_j)} \lesssim h \|\nabla v_h \times n\|_{L^2(F_j)}, \tag{4.7}$$

and by the trace and inverse inequalities of Lemma 3.2 and Lemma 3.3 we have

$$\left\langle (v_h - \bar{v}^j), \nabla \varphi_r \cdot n \right\rangle_{F_j} \lesssim C_T^2 (1 + C_I) \|\nabla v_h\|_{L^2(P_j)} \|\nabla \varphi_r\|_{L^2(P_j)}.$$

Moreover since by Cauchy-Schwarz inequality and the trace inequality,

$$|(\nabla v_h, \nabla w_h)_\Omega - \langle \nabla v_h \cdot n, w_h \rangle_{\partial\Omega}| \lesssim \|\nabla v_h\| \|w_h\|_{1,h}$$

we deduce using the stability (4.3) that

$$
\begin{aligned}
a_h(v_h,\varphi_r) &\geq \sum_{j=1}^{N_P} \|h^{-1/2}\overline{v}^j\|^2_{L^2(F_j)} - C\|\nabla v_h\|\|\varphi_r\|_{1,h} \\
&\geq \sum_{j=1}^{N_P} \|h^{-1/2}\overline{v}^j\|^2_{L^2(F_j)} - C_s\|\nabla v_h\| \left(\sum_{j=1}^{N_P} \|h^{-1/2}\overline{v}^j\|^2_{L^2(F_j)}\right)^{1/2}.
\end{aligned}
$$

We now fix $w_h = v_h + \eta\varphi_r$ and note that

$$
\begin{aligned}
a_h(v_h,w_h) &\geq \|\nabla v_h\|^2 + \eta \sum_{j=1}^{N_P} \|h^{-1/2}\overline{v}^j\|^2_{L^2(F_j)} \\
&\quad - C_s\|\nabla v_h\|\eta \left(\sum_{j=1}^{N_P} \|h^{-1/2}\overline{v}^j\|^2_{L^2(F_j)}\right)^{1/2} \\
&\geq (1-\epsilon)\|\nabla v_h\|^2 + \eta(1 - C_s^2\eta/(4\epsilon)) \sum_{j=1}^{N_P} \|h^{-1/2}\overline{v}^j\|^2_{L^2(F_j)}. \qquad (4.8)
\end{aligned}
$$

It follows, using once again the approximation properties of the L^2-projection on the piece wise constants (4.7), that for any $\epsilon < 1$ we may take η sufficiently small so that there exists $c_{\eta,\epsilon}$ such that

$$
c_{\eta,\epsilon}\|v_h\|^2_{1,h} \leq C c_{\eta,\epsilon} \left(\|\nabla v_h\|^2 + \sum_{j=1}^{N_P} \|h^{-1/2}\overline{v}^j\|^2_{L^2(F_j)}\right) \leq a_h(v_h,w_h).
$$

We may conclude by noting that by (4.3), our choice of r_j and the stability of the L^2-projection on piece wise constants there holds

$$
\|w_h\|_{1,h} \leq \|v_h\|_{1,h} + \eta\|\varphi_r\|_{1,h} \leq C_\eta\|v_h\|_{1,h}. \qquad (4.9)
$$

□

5. A priori error estimates. The stability estimate proved in the previous section together with the Galerkin orthogonality of Lemma 3.1 leads to error estimates in the $\|\cdot\|_{1,h}$ norm in a straightforward manner. First we will prove an auxiliary lemma for the continuity of $a_h(\cdot,\cdot)$. To this end we introduce the norm

$$
\|u\|_* := \|u\|_{1,h} + \|h^{\frac{1}{2}}\nabla u\cdot n\|_{L^2(\partial\Omega)}.
$$

LEMMA 5.1. *Let $u \in H^2(\Omega) + V_h^k$ and $v_h \in V_h^k$. Then the bilinear form $a_h(\cdot,\cdot)$ defined by* (3.2) *satisfies*

$$
a_h(u,v_h) \leq C\|u\|_*\|v_h\|_{1,h}.
$$

Proof. The result is immediate by application of the Cauchy-Schwarz inequality and the trace inequality of Lemma 3.2. □

PROPOSITION 5.2. *Let $u \in H^{k+1}(\Omega)$ be the solution of* (2.1) *and u_h the solution of* (3.1)*. Then there holds*

$$\|u - u_h\|_{1,h} \leq Ch^k |u|_{H^{k+1}(\Omega)}.$$

Proof. Let $i_{\mathsf{SZ}}^k u$ denote the Scott-Zhang interpolant of u [19]. Using the approximation properties of the interpolant it is straightforward to show that

$$\|u - i_{\mathsf{SZ}}^k u\|_{1,h} + \|u - i_{\mathsf{SZ}}^k u\|_* \lesssim h^k |u|_{H^{k+1}(\Omega)}.$$

We therefore use the triangle inequality to obtain

$$\|u - u_h\|_{1,h} \leq \|u - i_{\mathsf{SZ}}^k u\|_{1,h} + \|u_h - i_{\mathsf{SZ}}^k u\|_{1,h},$$

where only the second term needs to be bounded. To this end we apply the result of Theorem 4.2 followed by the consistency of Lemma 3.1

$$c_s \|u_h - i_{\mathsf{SZ}}^k u\|_{1,h} \leq \sup_{w_h \in V_h^k} \frac{a_h(u_h - i_{\mathsf{SZ}}^k u, w_h)}{\|w_h\|_{1,h}} = \sup_{w_h \in V_h^k} \frac{a_h(u - i_{\mathsf{SZ}}^k u, w_h)}{\|w_h\|_{1,h}}.$$

By the continuity of Lemma 5.1 and the approximation properties of $i_{\mathsf{SZ}}^k u$ we conclude

$$c_s \|u_h - i_{\mathsf{SZ}}^k u\|_{1,h} \lesssim \|u - i_{\mathsf{SZ}}^k u\|_* \lesssim h^k |u|_{H^{k+1}(\Omega)}.$$

□

For DG-methods it is well-known that the non-symmetric version may suffer from suboptimality in the convergence of the error in the L^2-norm due to the lack of adjoint consistency. This is true also for the non-symmetric Nitsche's method, however since the method is used on the scale of the domain and not of the element the suboptimality may be reduced to $h^{\frac{1}{2}}$ as we prove below.

PROPOSITION 5.3. *Let $u \in H^{k+1}(\Omega)$ be the solution of* (2.1) *and u_h the solution of* (3.1)*. Then*

$$\|u - u_h\| \leq Ch^{k+\frac{1}{2}} |u|_{H^{k+1}(\Omega)}.$$

Proof. Let z satisfy the adjoint problem

$$\left\{ \begin{array}{rcll} -\Delta z &=& u - u_h & \text{in } \Omega, \\ z &=& 0 & \text{on } \partial\Omega. \end{array} \right.$$

Under the assumptions on Ω we know that $\|z\|_{H^2(\Omega)} \leq C_{R2} \|u - u_h\|$. It follows that

$$\begin{aligned} \|u - u_h\|^2 = (u - u_h, -\Delta z)_\Omega = (\nabla(u - u_h), \nabla z)_\Omega - \langle u - u_h, \nabla z \cdot n \rangle_{\partial\Omega} \\ = a_h(u - u_h, z) + 2 \langle u - u_h, \nabla z \cdot n \rangle_{\partial\Omega}. \end{aligned}$$

By Lemma 3.1 and a continuity argument similar to that of Lemma 5.1, using that $(z - i_{\mathsf{SZ}}^1 z)|_{\partial\Omega} \equiv 0$, it follows that

$$\begin{aligned} a_h(u - u_h, z) &= a_h(u - u_h, z - i_{\mathsf{SZ}}^1 z) \\ &= (\nabla(u - u_h), \nabla(z - i_{\mathsf{SZ}}^1 z))_\Omega - \left\langle u - u_h, \nabla(z - i_{\mathsf{SZ}}^1 z) \cdot n \right\rangle_{\partial\Omega} \\ &\lesssim \|u - u_h\|_{1,h} \|z - i_{\mathsf{SZ}}^1 z\|_* \\ &\lesssim h \|u - u_h\|_{1,h} |z|_{H^2(\Omega)}. \end{aligned} \tag{5.1}$$

We also have, using the following global trace inequality

$$\|\nabla z \cdot n\|_{L^2(\partial\Omega)} \lesssim \|z\|_{H^2(\Omega)},$$

that

$$|\langle u - u_h, \nabla z \cdot n\rangle_{\partial\Omega}| \lesssim h^{1/2}\|u - u_h\|_{\frac{1}{2},h,\partial\Omega}\|z\|_{H^2(\Omega)}. \tag{5.2}$$

Collecting the inequalities (5.1) and (5.2) we arrive at the estimate

$$\|u - u_h\|^2 \lesssim (h + h^{1/2})h^k|u|_{H^{k+1}(\Omega)}\|z\|_{H^2(\Omega)}$$

and we conclude by applying the regularity estimate $\|z\|_{H^2(\Omega)} \leq C_{R2}\|u - u_h\|$. □

6. The convection–diffusion problem. Since the method we discuss leads to a non-symmetric system matrix the main interest of the method is for solving flow problems where an advection term makes the problem non-symmetric anyway. Note that there appears to be no analysis that is robust with respect to the Péclet number, even in the case of the non-symmetric discontinuous Galerkin method.

We will therefore now show how the above analysis can be extended to the case of convection–diffusion equations yielding optimal stability and accuracy both in the convection and the diffusion dominated regime. We will consider the following convection–diffusion–reaction equation:

$$\sigma u + \beta \cdot \nabla u - \varepsilon \Delta u = f \text{ in } \Omega, \tag{6.1}$$

and homogeneous Dirichlet boundary conditions. We assume that $\beta \in [W^1_\infty(\Omega)]^2$, $\sigma \in \mathbb{R}$,

$$\sigma - \frac{1}{2}\nabla \cdot \beta \geq c_\sigma \geq 0$$

and $\varepsilon \in \mathbb{R}^+$. In this case the formulation writes: find $u_h \in V_h$ such that

$$\begin{aligned} A_h(u_h, v_h) := (\sigma u_h + \beta \cdot \nabla u_h, v_h)_\Omega - \langle \beta \cdot n, u_h, v_h\rangle_{\partial\Omega^-} \\ + \varepsilon a_h(u_h, v_h) = (f, v_h)_\Omega, \quad \forall v_h \in V_h \end{aligned} \tag{6.2}$$

where $\partial\Omega^\pm := \{x \in \partial\Omega : \pm\beta \cdot n > 0\}$. First note that the positivity of the form now writes

$$A_h(u_h, u_h) \geq \frac{1}{2}\||\beta \cdot n|^{\frac{1}{2}} u_h\|^2_{\partial\Omega} + \|\varepsilon^{\frac{1}{2}}\nabla u_h\|^2, \tag{6.3}$$

hence provided $|\beta \cdot n| > 0$ on some portion of the boundary with non-zero measure the matrix is invertible. In the following we assume that this is the case, but we do not assume that $|\beta \cdot n| > 0$ everywhere on $\partial\Omega$. To prove optimal error estimates in general we require stronger stability results of the type proved above to hold. It appears difficult to prove these stronger results independently of the flow regime. Indeed it is convenient to characterize the flow using the local Péclet number:

$$Pe := \frac{|\beta| h}{\varepsilon}.$$

If $Pe < 1$ the flow is said to be diffusion dominated and if $Pe > 1$ we say that it is convection dominated. We will now treat these two cases separately.

In view of the equality (6.3) we introduce the following strengthened norm

$$\|v_h\|_{1,h,\beta}^2 := \varepsilon \|v_h\|_{1,h}^2 + \frac{1}{2} \||\beta\cdot n|^{\frac{1}{2}} v_h\|_{\partial\Omega}^2.$$

This norm is suitable in the diffusion dominated regime, but will be modified by the introduction of stabilization when the convection dominated regime is considered.

6.1. Diffusion dominated regime $Pe < 1$**.** In this case we may prove an inf-sup condition similar to that of Theorem 4.2. For simplicity we assume that $\sigma = 0$.

PROPOSITION 6.1. *(Inf-sup for convection–diffusion, $Pe < 1$.) For all functions $v_h \in V_h^k$ there holds*

$$c_s \|v_h\|_{1,h,\beta} \le \sup_{w_h \in V_h^k} \frac{A_h(v_h, w_h)}{\|w_h\|_{1,h,\beta}}. \tag{6.4}$$

Clearly, compared to the proof of Theorem 4.2 we only need to show how to handle the term

$$(\beta \cdot \nabla v_h, \varphi_r)_\Omega - \langle \beta \cdot n\, v_h, \varphi_r \rangle_{\partial\Omega^-}.$$

The necessary bound on this term is given in the following Lemma.

LEMMA 6.2. *Let φ_r be the function of Lemma 4.1 with r chosen as in* (4.6)*. Then for $Pe < 1$ there holds for all $\mu > 0$*

$$\begin{aligned}(\beta \cdot \nabla v_h, \eta\varphi_r)_\Omega &- \langle \beta \cdot n\, v_h, \eta\varphi_r \rangle_{\partial\Omega^-} \\ &\le \mu(\varepsilon \|\nabla v_h\|^2 + \||\beta\cdot n|^{\frac{1}{2}} v_h\|_{L^2(\partial\Omega)}^2) + C_\partial^2 (4\mu)^{-1} \eta^2 \varepsilon \|v_h\|_{\frac{1}{2},h,\partial\Omega}^2.\end{aligned}$$

Proof. Let

$$(\beta \cdot \nabla v_h, \eta\varphi_r)_\Omega - \langle \beta \cdot n\, v_h, \eta\varphi_r \rangle_{\partial\Omega^-} = T_1 + T_2.$$

By the definition of the Péclet number and the Cauchy-Schwarz inequality, we have

$$T_1 \le Pe\,\varepsilon^{\frac{1}{2}} \|\nabla v_h\| \eta \varepsilon^{\frac{1}{2}} \|h^{-1}\varphi_r\|.$$

From the construction of φ_r, a scaling argument, the stability (4.3) and the choice of r (4.6) we deduce that

$$\|h^{-1}\varphi_r\| \lesssim \|\nabla \varphi_r\| \le C_\partial \|v_h\|_{\frac{1}{2},h,\partial\Omega}.$$

Using the arithmetic-geometric inequality we have

$$T_1 \le \mu\varepsilon \|\nabla v_h\|^2 + C_\partial^2 (4\mu)^{-1} Pe^2 \eta^2 \varepsilon \|v_h\|_{\frac{1}{2},h,\partial\Omega}^2.$$

For T_2 we have using a Cauchy-Schwarz inequality, the definition of the Péclet number and the stability (4.3)

$$T_2 \le \||\beta\cdot n|^{\frac{1}{2}} v_h\|_{L^2(\partial\Omega)} Pe^{\frac{1}{2}} \eta \varepsilon^{\frac{1}{2}} \|\varphi_r\|_{\frac{1}{2},h,\partial\Omega} \le C_\partial \||\beta\cdot n|^{\frac{1}{2}} v_h\|_{L^2(\partial\Omega)} \eta \varepsilon^{\frac{1}{2}} \|v_h\|_{\frac{1}{2},h,\partial\Omega}.$$

We apply the arithmetic-geometric inequality once again to conclude. □

Proof. (Proposition 6.1) The inf-sup stability (6.4) now follows by taking $w_h := v_h + \eta\varphi_r$ and proceeding as in equation (4.8) using (6.3) and Lemma 6.2 in the following fashion

$$A_h(v_h, v_h + \eta\varphi_r) \geq (1 - \epsilon - \mu)\varepsilon\|\nabla v_h\|^2 + (\frac{1}{2} - \mu)\||\beta \cdot n|^{\frac{1}{2}} v_h\|^2_{L^2(\Omega)} \\ + \eta(1 - C_s^2\eta/(4\epsilon) - C_\partial^2\eta/(4\mu))\varepsilon\|v_h\|^2_{\frac{1}{2},h,\partial\Omega}.$$

We may now choose $\epsilon = 1/4$ and $\mu = 1/4$ and then η small enough so that positivity is ensured. Then

$$A_h(v_h, v_h + \eta\varphi_r) \geq C_\eta\|v_h\|^2_{1,h,\beta}.$$

We conclude as in Theorem 4.2, but now using the norm $\|\cdot\|_{1,h,\beta}$,

$$\|w_h\|_{1,h,\beta} \leq \|v_h\|_{1,h,\beta} + \eta\|\varphi_r\|_{1,h,\beta} \leq \|v_h\|_{1,h,\beta} + \eta C\|v_h\|_{1,h,\beta} + \eta\||\beta \cdot n|^{\frac{1}{2}}\varphi_r\|_{L^2(\partial\Omega)} \\ \leq C\|v_h\|_{1,h,\beta} + Pe^{\frac{1}{2}}\eta\varepsilon^{\frac{1}{2}}\|\varphi_r\|_{1,h} \leq C_{Pe,\eta}\|v_h\|_{1,h,\beta}.$$

$\square$

Proceeding as in Proposition 5.2, this leads to optimal a priori estimates in the norm $\|\cdot\|_{1,h}$ for $Pe < 1$.

PROPOSITION 6.3. *Let $u \in H^{k+1}(\Omega)$ be the solution of* (6.1) *and u_h the solution of* (6.2) *and assume that $Pe < 1$. Then*

$$\|u - u_h\|_{1,h} \leq Ch^k|u|_{H^{k+1}(\Omega)}.$$

Proof. As in the proof of Proposition 5.2 we arrive at the following representation of the discrete error

$$c_s\|u_h - i^k_{\mathsf{SZ}}u\|_{1,h,\beta} \leq \sup_{w_h \in V_h^k} \frac{A_h(u_h - i^k_{\mathsf{SZ}}u, w_h)}{\|w_h\|_{1,h,\beta}} = \sup_{w_h \in V_h^k} \frac{A_h(u - i^k_{\mathsf{SZ}}u, w_h)}{\|w_h\|_{1,h,\beta}}.$$

By the continuity of Lemma 5.1 and an integration by parts in the convective term we obtain

$$A_h(u_h - i^k_{\mathsf{SZ}}u, w_h) \lesssim \varepsilon\|u - i^k_{\mathsf{SZ}}u\|_*\|w_h\|_{1,h} \\ + (u - i^k_{\mathsf{SZ}}u, \beta \cdot \nabla w_h)_\Omega + \langle \beta \cdot n(u - i^k_{\mathsf{SZ}}u), w_h\rangle_{\partial\Omega^+}) \\ \lesssim \varepsilon(\|u - i^k_{\mathsf{SZ}}u\|_* + Pe\|h^{-1}(u - i^k_{\mathsf{SZ}}u)\| + Pe\|u - i^k_{\mathsf{SZ}}u\|_{\frac{1}{2},h,\partial\Omega})\|w_h\|_{1,h,\beta}.$$

As a consequence

$$\varepsilon\|u_h - i^k_{\mathsf{SZ}}u\|_{1,h} \leq \|u_h - i^k_{\mathsf{SZ}}u\|_{1,h,\beta} \\ \lesssim c_s^{-1}\varepsilon(\|u - i^k_{\mathsf{SZ}}u\|_* + Pe\|h^{-1}(u - i^k_{\mathsf{SZ}}u)\| + Pe\|u - i^k_{\mathsf{SZ}}u\|_{\frac{1}{2},h,\partial\Omega}).$$

The claim follows by dividing through by ε, using approximation and the assumption $Pe < 1$. $\square$

6.2. Convection dominated regime: the Streamline–diffusion mehod. In the convection dominated regime, when $Pe > 1$, we need to add some stabilization in order to obtain a robust scheme. We will here first consider the simple case of Streamline-diffusion (SD) stabilization and assuming $\sigma = 0$. In the next section the results will be extended to include the Continuous interior penalty (CIP) method.

The formulation now takes the form: find $u_h \in V_h^k$ such that

$$A_{SD}(u_h, v_h) := (\beta \cdot \nabla u_h, v_h + \delta\beta \cdot \nabla v_h)_\Omega - \sum_K (\varepsilon \Delta u_h, \delta\beta \cdot \nabla v_h)_K - \langle \beta \cdot n\, u_h, v_h \rangle_{\partial\Omega^-} + \varepsilon a_h(u_h, v_h) = (f, v_h + \delta\beta \cdot \nabla v_h)_\Omega, \quad \forall v_h \in V_h^k, \quad (6.5)$$

where $\delta = \gamma_{SD} h/|\beta|$ when $Pe > 1$ and $\delta = 0$ otherwise. At high Péclet numbers, the enhanced robustness of the stabilized method allows us to work in the stronger norm $|||u_h|||_{h,\delta}$ defined by

$$|||u_h|||_{h,\delta}^2 := \|\delta^{\frac{1}{2}} \beta \cdot \nabla u_h\|^2 + \frac{1}{2} \||\beta \cdot n|^{\frac{1}{2}} u_h\|_{L^2(\partial\Omega)}^2 + \varepsilon \|\nabla u_h\|^2. \quad (6.6)$$

We will also use the weaker form $|||u_h|||_{h,0}^2$ defined by (6.6) with $\delta = 0$ and for the convergence analysis we introduce the norm

$$|||u_h|||_*^2 := \|\delta^{-\frac{1}{2}} u_h\|^2 + \varepsilon \|\nabla u_h \cdot n\|_{-\frac{1}{2},h,\partial\Omega}^2 + \sum_K \|\delta^{\frac{1}{2}} \varepsilon \Delta u_h\|_{L^2(K)}^2 + \varepsilon \|u_h\|_{\frac{1}{2},h,\partial\Omega}^2 + |||u_h|||_{h,\delta}^2.$$

Testing the formulation (6.5) with $v_h = u_h$ yields the positivity

$$c|||u_h|||_{h,\delta}^2 \leq A_{SD}(u_h, u_h) \quad (6.7)$$

in the standard way using an element wise inverse inequality to absorb the second order term, i.e.

$$\sum_K (\varepsilon \Delta u_h, \delta\beta \cdot \nabla u_h)_K \leq \frac{1}{2} C_I^2 \gamma_{SD} Pe^{-1/2} \|\varepsilon^{\frac{1}{2}} \nabla u_h\|^2 + \frac{1}{2} \|\delta^{\frac{1}{2}} \beta \cdot \nabla u_h\|^2.$$

Clearly for $\gamma_{SD} < 1/(C_I^2)$ stability holds for $Pe > 1$.

Unfortunately the norms proposed above seem too weak to allow for optimal error estimates. Indeed, since we do not control all of $\|u_h\|_{1,h}$, for general $u \in H^2 + V_h^1$, $v_h \in V_h^1$ there does not hold $A_{SD}(u, v_h) \leq |||u|||_* |||v_h|||_{h,\delta}$, (c.f. Lemma 5.1) unless an assumption on the boundary velocity such as $|\beta \cdot n| h > \varepsilon$ is made. It also appears to be difficult to obtain an inf-sup condition similar to (6.4) in the high Péclet regime.

We therefore use another technique to prove optimal convergence directly. The idea is to construct an interpolation operator $\pi_\partial u$, such that the *interpolation error* $u - \pi_\partial u$ satisfies the continuity estimate:

$$A_{SD}(u - \pi_\partial u, v_h) \leq |||u - \pi_\partial u|||_* |||v_h|||_{h,\delta}. \quad (6.8)$$

Assume that we have an interpolation operator $\pi_\partial : H^1(\Omega) \mapsto V_h^1$ such that the following hypothesis are satisfied.

(**H1**) Approximation,

$$\|\pi_\partial u - u\| + h\|\nabla(\pi_\partial u - u)\| \leq C h^{k+1} |u|_{H^{k+1}(\Omega)}. \quad (6.9)$$

(**H2**) Normal gradient,

$$\int_{F_i} \nabla(\pi_\partial u - u)\cdot n \, \mathrm{d}s = 0, \quad i = 1 \dots N_P, \tag{6.10}$$

where F_i are the boundary segments introduced in Section 4.

Under assumptions (**H1**) and (**H2**), we may prove the optimal convergence of the SD-method.

PROPOSITION 6.4. *Let $u \in H^{k+1}(\Omega)$ be the solution of* (6.1) *and u_h the solution of* (6.5). *Assume that there exists $\pi_\partial u \in V_h^k$ satisfying (***H1***) and (***H2***). Then*

$$|||u - u_h|||_{h,\delta} \lesssim h^{k+\frac{1}{2}}(1 + Pe^{-\frac{1}{2}})|u|_{H^{k+1}(\Omega)}.$$

Proof. It follows from the approximation properties of π_∂ that

$$|||u - \pi_\partial u|||_* \lesssim \|\beta\|_\infty^{\frac{1}{2}} h^{k+\frac{1}{2}}(1 + Pe^{-1/2})|u|_{H^{k+1}(\Omega)}.$$

We now need to prove the continuity (6.8). Note that

$$A_{SD}(u - \pi_\partial u, v_h) = (\delta^{\frac{1}{2}}\beta\cdot\nabla(u - \pi_\partial u) + \delta^{-\frac{1}{2}}(u - \pi_\partial u), \delta^{\frac{1}{2}}\beta\cdot\nabla v_h)_K$$

$$- \sum_K (\delta^{\frac{1}{2}}\varepsilon\Delta(u - \pi_\partial u), \delta^{\frac{1}{2}}\beta\cdot\nabla v_h)_K + \langle \beta\cdot n\,(u - \pi_\partial u), v_h\rangle_{\partial\Omega^+} + \varepsilon a_h(u - \pi_\partial u, v_h)$$

$$\lesssim |||u - \pi_\partial u|||_* |||v_h|||_{h,\delta} + \underbrace{\varepsilon a_h(u - \pi_\partial u, v_h)}_{I_1}.$$

Consider now the term I_1. We will prove the continuity

$$\varepsilon a_h(u - \pi_\partial u, v_h) \le \varepsilon^{\frac{1}{2}}\|u - \pi_\partial u\|_* \varepsilon^{\frac{1}{2}}\|v_h\|_{1,h} \le |||u - \pi_\partial u|||_* |||v_h|||_{h,\delta} \tag{6.11}$$

Using Cauchy-Schwarz inequality and a trace inequality we show the continuity of the first and last term.

$$I_1 = \varepsilon(\nabla(u - \pi_\partial u), \nabla v_h)_\Omega - \varepsilon\,\langle \nabla(u - \pi_\partial u)\cdot n, v_h\rangle_{\partial\Omega} + \varepsilon\,\langle \nabla v_h\cdot n, (u - \pi_\partial u)\rangle_{\partial\Omega}$$

$$\le \varepsilon^{\frac{1}{2}}\|u - \pi_\partial u\|_* |||v_h|||_{h,0} - \varepsilon\,\langle \nabla(u - \pi_\partial u)\cdot n, v_h\rangle_{\partial\Omega}.$$

For the remaining term we must exploit the orthogonality property (6.10) of $\pi_\partial u$ on the boundary. Indeed by decomposing the boundary integral on the N_P subdomains F_i we have, denoting by $\bar{v}_h^i$ the average of v_h over the boundary segment F_i.

$$\begin{aligned}
\varepsilon\,\langle \nabla(u - \pi_\partial u)\cdot n, v_h\rangle_{\partial\Omega} &= \varepsilon \sum_{i=1}^{N_P} \left\langle \nabla(u - \pi_\partial u)\cdot n, v_h - \bar{v}_h^i\right\rangle_{F_i} \\
&\le \varepsilon \sum_{i=1}^{N_P} \|\nabla(u - \pi_\partial u)\cdot n\|_{L^2(F_i)} \|v_h - \bar{v}_h^i\|_{L^2(F_i)} \\
&\lesssim \varepsilon^{\frac{1}{2}}\|\nabla(u - \pi_\partial u)\cdot n\|_{-\frac{1}{2},h,\partial\Omega}\varepsilon^{\frac{1}{2}}\|\nabla v_h\| \\
&\lesssim \varepsilon^{\frac{1}{2}}\|u - \pi_\partial u\|_* \varepsilon^{\frac{1}{2}}\|v_h\|_{1,h}.
\end{aligned}$$

Where we used the approximation properties of the local average and a trace inequality. Collecting the above estimates and noting that

$$\varepsilon^{\frac{1}{2}}\|u-\pi_\partial u\|_* \le |||u-\pi_\partial u|||_*, \quad \varepsilon^{\frac{1}{2}}\|v_h\|_{1,h} \le |||v_h|||_{h,0}$$

concludes the proof of (6.8).

Using the positivity (6.7), and the consistency of the method we have, setting $e_h := u_h - \pi_\partial u$, and using that $Pe > 1$

$$\begin{aligned}|||e_h|||^2_{h,\delta} = A_{SD}(e_h,e_h) = A_{SD}(u-\pi_\partial u, e_h) \lesssim |||u-\pi_\partial u|||_* |||e_h|||_{h,\delta} \\ \lesssim h^{k+\frac{1}{2}}\|\beta\|^{\frac{1}{2}}_\infty (1+Pe^{-\frac{1}{2}})|u|_{H^{k+1}(\Omega)}|||e_h|||_{h,\delta}.\end{aligned}$$

□

We end this section by the following Lemma establishing the existence of the interpolation π_∂ with the required properties.

Lemma 6.5. *The interpolation operator $\pi_\partial : H^1(\Omega) \mapsto V_h^1$ satisfying the properties (**H1**) and (**H2**) exists.*

Proof. Let $\pi_\partial u := i^k_{\mathsf{SZ}} u + \varphi_r$ where φ_r is the function of Lemma 4.1 with the r_j chosen such that

$$r_j = \overline{\nabla u \cdot n}^j - \overline{\nabla i^k_{\mathsf{SZ}} u \cdot n}^j.$$

Clearly by construction there holds

$$\begin{aligned}\int_{F_i} (\nabla \pi_\partial u \cdot n - \nabla u \cdot n)\ \mathrm{d}s &= \int_{F_i} (\nabla i^k_{\mathsf{SZ}} u \cdot n + \nabla \varphi_r \cdot n - \nabla u \cdot n)\ \mathrm{d}s \\ &= \int_{F_i} (\nabla i^k_{\mathsf{SZ}} u \cdot n + r_i - \nabla u \cdot n)\ \mathrm{d}s = 0.\end{aligned}$$

To prove the approximation results we decompose the error

$$\|u - \pi_\partial u\| \le \|u - i^k_{\mathsf{SZ}} u\| + \|i^k_{\mathsf{SZ}} u - \pi_\partial u\| \le C h^{k+1}|u|_{H^{k+1}(\Omega)} + \|\varphi_r\|.$$

Using local Poincaré inequalities and the stability (4.3) of φ_r we get

$$\begin{aligned}\|\varphi_r\| \lesssim \|h\nabla\varphi_r\| \lesssim h^{\frac{3}{2}} \left(\sum_{i=1}^{N_P} \|r_i\|^2_{L^2(F_i)}\right)^{\frac{1}{2}} \\ = h^{\frac{3}{2}} \left(\sum_{i=1}^{N_P} \|\overline{\nabla u \cdot n}^i - \overline{\nabla i^k_{\mathsf{SZ}} u \cdot n}^i\|^2_{L^2(F_i)}\right)^{\frac{1}{2}}.\end{aligned}$$

Using the stability of the projection onto piece wise constants, element wise trace inequalities and finally approximation, we conclude

$$\begin{aligned}&\|\overline{\nabla u \cdot n}^i - \overline{\nabla i^k_{\mathsf{SZ}} u \cdot n}^i\|^2_{L^2(F_i)} \le \|\nabla u \cdot n - \nabla i^k_{\mathsf{SZ}} u \cdot n\|^2_{L^2(F_i)} \\ &\le 2C_T^2 (h^{-1}\|\nabla(u - i^k_{\mathsf{SZ}} u)\|^2_{L^2(P_i)} + h \sum_{K\in P_i} \|D^2(u - i^k_{\mathsf{SZ}} u)\|^2_{L^2(K)}) \lesssim h^{2k-1}|u|^2_{H^{k+1}(P_i)}\end{aligned}$$

where D^2u is the standard multi-index notation for all the second derivatives of u. We conclude that

$$\|\varphi_r\| \lesssim h^{\frac{3}{2}}(\sum_{i=1}^{N_P} \|\nabla u \cdot n - \nabla i^k_{\mathsf{SZ}} u \cdot n\|^2_{L^2(F_i)})^{\frac{1}{2}} \lesssim h^{k+1}|u|_{H^{k+1}(\Omega)}$$

The estimate on the gradient is immediate by

$$\begin{aligned}\|\nabla(u - \pi_\partial u)\| &\leq \|\nabla(u - i^k_{\mathsf{SZ}} u)\| + \|\nabla(i^k_{\mathsf{SZ}} u - \pi_\partial u)\| \\ &\leq \|\nabla(u - i^k_{\mathsf{SZ}} u)\| + C_I h^{-1}\|i^k_{\mathsf{SZ}} u - \pi_\partial u\| \lesssim h^k |u|_{H^{k+1}(\Omega)}.\end{aligned}$$

□

6.2.1. Convection dominated regime: the Continuous Interior Penalty method. In this section we will sketch how the above results extend to symmetric stabilization methods assuming that $c_\sigma > 0$. To reduce technicalities we also assume that $\beta \in \mathbb{R}^2$. We give a full proof only in the case of piecewise affine finite elements. Recall that the CIP method is obtained by adding a penalty term on the jump of the gradient over element faces to the finite element formulation (6.2). The formulation then writes: find $u_h \in V_h^k$ such that

$$A_h(u_h, v_h) + J_h(u_h, v_h) = (f, v_h)_\Omega, \quad \forall v_h \in V_h^k, \tag{6.12}$$

where

$$J_h(u_h, v_h) := \gamma_{CIP} \sum_{K \in \mathcal{T}_h} \sum_{F \in \partial K \setminus \partial\Omega} \int_F h_F^2 |\beta \cdot n_F| [\nabla u_h \cdot n_F][\nabla v_h \cdot n_F] \, \mathrm{d}s,$$

with $[x]$ denoting the jump of the quantity x over the face F and n_F the normal to F, the orientation is arbitrary but fixed in both cases.

The analysis once again depends on the construction of a special interpolant $\pi_{CIP} u \in V_h^k$. This time $\pi_{CIP} u$ must satisfy both the optimal approximation error estimates of (6.9), the property (6.10) on the normal gradient, and the additional design condition:

$$(u - \pi_{CIP} u, \beta \cdot \nabla v_h) \leq \|h^{-\frac{1}{2}}|\beta|^{\frac{1}{2}}(u - \pi_{CIP} u)\| \gamma_{CIP}^{-\frac{1}{2}} J_h(v_h, v_h)^{\frac{1}{2}}, \; \forall v_h \in V_h^k. \tag{6.13}$$

Once such an interpolant has been proven to exist, the technique of [3], combined with the analysis above, may be used to prove quasi-optimal L^2-convergence for $c_\sigma > 0$. Using a similarly designed interpolation operator, an inf-sup condition can be used to prove stability and error estimates in the norm $|||\cdot|||_{h,\delta}$ following [6, 5]. Here we will first the error estimate in the L^2-norm, assuming the existence of $\pi_{CIP} u$ and then show how to construct the interpolant in the special case $k = 1$.

PROPOSITION 6.6. *Assume that* $\pi_{CIP} u \in V_h^k$ *satisfying,* (6.9), (6.10) *and* (6.13) *exists. Let* $u \in H^{k+1}(\Omega)$ *be the solution to* (6.1), *with* $c_\sigma > 0$, *and* u_h *be the solution to* (6.12). *Then*

$$\|u - u_h\| \lesssim (c_\sigma)^{-1}(\sigma^{\frac{1}{2}} h^{\frac{1}{2}} + |\beta|^{\frac{1}{2}}(1 + Pe^{-\frac{1}{2}}))h^{k+\frac{1}{2}}|u|_{H^{k+1}(\Omega)}.$$

Proof. Let $e_h := u_h - \pi_{CIP} u$. There holds with $c_\sigma > 0$,

$$c_\sigma \|e_h\|^2 + |||e_h|||^2_{h,0} + J_h(e_h, e_h) \leq A_h(e_h, e_h) + J_h(e_h, e_h).$$

By the consistency of the method we have

$$c_\sigma \|e_h\|^2 + |||e_h|||^2_{h,0} + J_h(e_h, e_h) \leq A_h(u - \pi_{CIP} u, e_h) - J_h(\pi_{CIP} u, e_h).$$

Finally by the continuity (6.11), that holds thanks to property (6.10), we have

$$\begin{aligned}
&A_h(u - \pi_{CIP} u, e_h) - J_h(\pi_{CIP} u, e_h) \\
&\quad = (\sigma(u - \pi_{CIP} u), e_h) + (u - \pi_{CIP} u, \beta \cdot \nabla e_h) - \int_{\partial\Omega} \beta \cdot n(u - \pi_{CIP} u) e_h \, \mathrm{d}s \\
&\qquad\qquad + \varepsilon a_h(u - \pi_{CIP} u, e_h) + J_h(\pi_{CIP} u, e_h) \\
&\qquad \leq ((\sigma^{\frac{1}{2}} h^{\frac{1}{2}} + C|\beta|^{\frac{1}{2}} \gamma_{CIP}^{-\frac{1}{2}}) \|h^{-\frac{1}{2}}(u - \pi_{CIP} u)\| + \|u - \pi_{CIP} u\|_{1,h,\beta} \\
&\qquad\qquad + \varepsilon^{\frac{1}{2}} \|u - \pi_{CIP} u\|_* + J_h(\pi_{CIP} u, \pi_{CIP} u)^{\frac{1}{2}}) \\
&\qquad\qquad\qquad \times (\sigma^{\frac{1}{2}} \|e_h\|^2 + \|e_h\|^2_{1,h,\beta} + J_h(e_h, e_h))^{\frac{1}{2}} \qquad (6.14)
\end{aligned}$$

and we end the proof by applying approximation estimates. ∎

We will now prove the existence of the interpolant $\pi_{CIP} u$ in the case of piecewise affine continuous finite element approximation.

Lemma 6.7. *The function $\pi_{CIP} u \in V_h^1$, satisfying* (6.9), (6.10) *and* (6.13)*, is well defined and satisfies the approximation estimate*

$$\|u - \pi_{CIP} u\| + h\|\nabla(u - \pi_{CIP} u)\| \lesssim h^2 |u|_{H^2(\Omega)}.$$

Proof. We write $\pi_{CIP} u := \pi_h u + \varphi_{CIP}$ where $\pi_h u$ denotes the L^2-projection on V_h^1 and $\varphi_{CIP} \in V_h^1$ is a function defined on patches P_i that satisfy the inequalities (4.2) and (4.3), but also has the property

$$\int_{P_i} \varphi_{CIP} \, \mathrm{d}x = 0, \; i = 1, \ldots, N_P.$$

Clearly for this to hold we must modify the definition of the patches on the faces F_i to include interior nodes in the domain. For simplicity we assume that any element containing a node that connects to two nodes in the boundary segment $\bar{F}_i$ (through edges that may be associated to other elements) is included in the patch P_i (see Figure 6.1). Define two functions w_I and w_F on P_i (also illustrated in Figure 6.1) such that

$$w_I := \begin{cases} 1 \text{ in all nodes } x \in \overset{\circ}{P}_i \\ 0 \text{ in all nodes } x \in \Omega \setminus \overset{\circ}{P}_i \end{cases}, \quad w_F := \begin{cases} 1 \text{ in all nodes } x \in \overset{\circ}{F}_i \\ 0 \text{ in all nodes } x \in \bar{\Omega} \setminus \overset{\circ}{F}_i \end{cases}.$$

We must now show that there exists a function $\varphi_i = a w_I + b w_F$ satisfying the two constraints

$$\int_{P_i} \varphi_i \, \mathrm{d}x = 0, \quad \overline{\nabla \varphi_i \cdot n}^i = r_i. \qquad (6.15)$$

The construction of $\pi_{CIP} u$ is obtained by choosing $r_i = \overline{\nabla u \cdot n}^i - \overline{\nabla \pi_h u \cdot n}^i$ in the system (6.15) above and then defining $\varphi_{CIP}|_{P_i} := \varphi_i$.

To study φ_i, first map the patch P_i to the reference patch $\hat{P}_i$. Consider the linear system for $v := (a,b)^T \in \mathbb{R}^2$ of the form:

$$\mathcal{A}v := \begin{bmatrix} \int_{\hat{P}_i} \hat{w}_I \ \mathrm{d}\hat{x} & \int_{\hat{P}_i} \hat{w}_F \ \mathrm{d}\hat{x} \\ \int_{\hat{F}_i} \nabla \hat{w}_I \cdot \hat{n} \ \mathrm{d}\hat{s} & \int_{\hat{F}_i} \nabla \hat{w}_F \cdot \hat{n} \ \mathrm{d}\hat{s} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$
$$= \begin{bmatrix} 0 \\ \int_{\hat{F}_i} \nabla(\hat{u} - \pi_h \hat{u}) \cdot \hat{n} \ \mathrm{d}\hat{s} \end{bmatrix} =: \hat{f}.$$

We must prove that the matrix $\mathcal{A}$ is invertible, but this is immediate noting that the two coefficients in the first line of the matrix both are strictly positive, whereas in the second line the coefficient in the first column is negative by construction and that in the right column is positive. The stability estimate (4.3) now follows from a scaling argument back to the physical patch P_i. Indeed since the matrix $\mathcal{A}$ is invertible we have

$$|v| \lesssim \sup_{w \in \mathbb{R}^2} \frac{w^T \mathcal{A} v}{|w|} = \sup_{w \in \mathbb{R}^2} \frac{w^T \hat{f}}{|w|} = |\hat{f}|.$$

By norm equivalence we have

$$\|\hat{\varphi}_i\|_{\hat{P}_i} \lesssim \|\nabla \hat{\varphi}_i\|_{\hat{P}_i} \lesssim |v| \lesssim |\hat{f}|.$$

After scaling back to the physical element we get

$$h^{-1} \|\varphi_i\|_{P_i} \lesssim \|\nabla \varphi_i\|_{P_i} \lesssim |f| \lesssim \|h^{\frac{1}{2}} \nabla(u - \pi_h u) \cdot n\|_{F_i}, \tag{6.16}$$

which proves (4.3).

The approximation error estimates are proven in the same way as in Lemma 6.5. Indeed by a similar decomposition of the error we have for this case

$$\|u - \pi_{CIP} u\| \leq \|u - \pi_h u\| + \|\pi_h u - \pi_{CIP} u\| \lesssim h^2 |u|_{H^2(\Omega)} + \|\varphi_{CIP}\|$$

and for φ_{CIP} we may conclude using the proof of Lemma 6.5, using (6.16).

It remains to prove the continuity (6.13). This follows from

$$(u - \pi_{CIP} u, \beta \cdot \nabla v_h) = (u - \pi_h u, \beta \cdot \nabla v_h) + \sum_{i=1}^{N_P} (\varphi_i, \beta \cdot \nabla v_h)$$
$$= (u - \pi_h u, \beta \cdot \nabla v_h - I_{CIP} \beta \cdot \nabla v_h) + \sum_{i=1}^{N_P} (\varphi_i, (\beta \cdot \nabla v_h - \pi_{0,P_i} \beta \cdot \nabla v_h)).$$

Here I_{CIP} denotes a particular quasi interpolation operator defined using averages of $\beta \cdot \nabla v_h$ in each node (see [3]) and π_{0,P_i} denotes the projection on piecewise constant functions on P_i. Using norm equivalence on discrete spaces and mapping from the reference patch, we observe that

$$\|h^{\frac{1}{2}} |\beta|^{-\frac{1}{2}} (\beta \cdot \nabla v_h - I_{CIP} \beta \cdot \nabla v_h)\|^2 \lesssim \gamma_{CIP}^{-1} J_h(v_h, v_h)$$

and

$$\sum_{i=1}^{N_P} \|h^{\frac{1}{2}} |\beta|^{-\frac{1}{2}} (\beta \cdot \nabla v_h - \pi_{0,P_i} \beta \cdot \nabla v_h)\|_{P_i}^2 \lesssim \gamma_{CIP}^{-1} J_h(v_h, v_h).$$

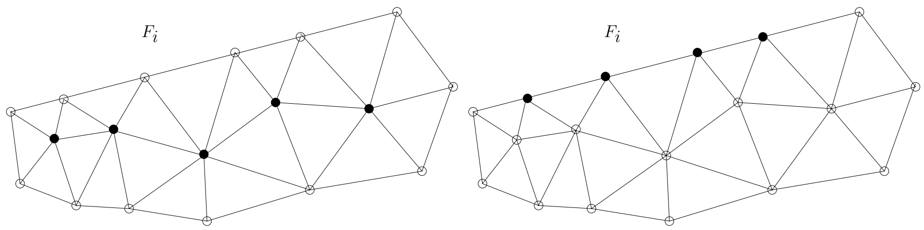

FIG. 6.1. *Example of a boundary patch P_i, with the functions w_I (left) and w_F (right). The functions take the value 1 in filled nodes and zero in the other nodes.*

N	Nitsche H^1	strong H^1	Nitsche L^2	strong L^2
10	7.0E-1 (—)	6.7E-1 (—)	2.4E-2 (—)	2.0E-2 (—)
20	3.5E-1 (1.0)	3.5E-1 (0.94)	5.5E-3 (2.1)	5.5E-3 (1.9)
40	1.7E-1 (1.0)	1.7E-1 (1.0)	1.3E-3 (2.1)	1.3E-3 (2.1)
80	8.2E-2 (1.1)	8.2E-2 (1.1)	3.3E-4 (2.0)	3.1E-4 (2.1)

TABLE 7.1

Comparison of errors between the non-symmetric Nitsche method and standard strongly imposed boundary conditions, using piece wise affine approximation on unstructured meshes.

The first claim was proved in [3] and the second holds since $\beta \cdot \nabla v_h$ is constant on each element. □

Remark 1. For high order element the construction of the interpolant $\pi_{CIP} u$ is much more technical and beyond the scope of the present work. Indeed it is no longer sufficient to prove orthogonality of φ_i against a constant on P_i, but it must be shown to be orthogonal to the continuous finite element space of order $k-1$, on P_i. On the other hand the patches P_i can be chosen freely, provided $diam(P_i) = O(h)$.

7. Numerical examples. We study two different numerical examples, both have been computed using the package FreeFem++ [12]. First we consider a simple problem with smooth exact solution, then we consider a convection-diffusion problem and show the stabilizing effect of the Nitsche type weak boundary condition for convection dominated flow.

7.1. Problem with smooth solution. We consider equation (2.1) in the unit square, with $f = 5\pi^2 \sin(\pi x)\sin(2\pi y)$ and $g = 0$. The mesh is unstructured with $N = 10, 20, 40, 80$ elements per side. The exact solution is then given by $u = \sin(\pi x)\sin(2\pi y)$. We give the convergence in both the L^2-norm and the H^1-norm for piece wise affine approximation in Table 7.1. The case of quadratic approximation is considered in Table 7.2. The order p in $O(h^p)$ is given in parenthesis next to the error. We have not managed to construct an example exhibiting the suboptimal convergence order of the Nitsche method. Some cases with non-homogeneous boundary conditions, not reported here, were computed both with affine and quadratic elements. They all had optimal convergence on the finer meshes. The theoretical results do not extend to the symmetric version of Nitsche's method and stability is unlikely to hold on general meshes. Applying the symmetric method to the proposed numerical example yields a solution with clear boundary oscillations on the coarse meshes see Figure 7.1. On finer meshes these oscillations vanish and the performance is similar to that of the non-symmetric method. Note that although the convergence of the Nitsche method is optimal in this case, the error constant of the non-symmetric

N	Nitsche H^1	strong H^1	Nitsche L^2	strong L^2
10	5.3E-2 (—)	5.1E-2 (—)	1.7E-3 (—)	6.5E-4 (—)
20	1.4E-2 (1.9)	1.4E-2 (1.9)	2.2E-4 (2.9)	9.6E-5 (2.8)
40	3.5E-3 (2.0)	3.5E-3 (2.0)	2.1E-5 (3.4)	1.1E-5 (3.1)
80	8.6E-4 (2.0)	8.6E-4 (2.0)	2.5E-6 (3.1)	1.4E-6 (3.0)

TABLE 7.2

Comparison of errors between the non-symmetric Nitsche method and standard strongly imposed boundary conditions, using piece wise quadratic approximation on unstructured meshes.

error norm	$\gamma = 0$	$\gamma = 10$	$\gamma = 20$	$\gamma = 40$	$\gamma = 80$
$\|u - u_h\|_{L^2}$	3.3E-4	2.9E-4	3.0E-4	3.0E-4	3.0E-4
$\|u - u_h\|_{H^1}$	8.2E-2	8.2E-2	8.2E-2	8.2E-2	8.2E-2

TABLE 7.3

Study of the dependence of the accuracy on the penalty parameter, piece wise affine approximation, unstructured mesh, $N = 80$

method in the L^2-norm is a factor two larger than that of the strongly imposed boundary conditions for piece wise quadratic approximation. The same computations were made on structured meshes (not reported here) and this effect was slightly larger in this case, with a factor two in the affine case and four in the quadratic case. The errors in the H^1-norm on the other hand are of comparable size for the two methods.

This motivates a study of how the error depends on the penalty parameter γ in (3.3). We therefore run a series of computations with $\gamma = 0, 10, 20, 40, 80$. In Table 7.3 we report the results for piece wise affine approximation and in Table 7.4 the results for piece wise quadratic approximation. We note that there is a visible, but negligible, effect on the error measured in the L^2-norm, but no effect on the error in the H^1-norm.

7.2. Problem with outflow layer. For this case we only compare the solutions qualitatively. We consider the problem with a convection term (6.1). To create an outflow layer we have chosen $f := 1$, $\beta := (0.5, 1)$, $\sigma := 0$ in Ω. We discretized Ω with a structured mesh having 80 piece wise affine elements on each side. The contour-plots for $\varepsilon = 0.1, 0.001, 0.00001$ are reported in Figure 7.2 for Nitsche's method and in Figure 7.3 for the strongly imposed boundary conditions. Note that no stabilization has been added in either case. This computation illustrates the strong stabilizing effect of the weakly imposed boundary condition. A theoretical explanation of this phenomenon was given in [18]. Finally we consider the effect of adding stabilization to the computation. In this case we take $N = 80$ with piece wise quadratic approximation. We report the results of a computation without stabilization, with the SD-method ($\gamma_{SD} = 0.2$) and with the CIP-method ($\gamma_{CIP} = 0.005$) in Figure 7.4. Note that the stabilized methods clean up the remaining spurious oscillations in both cases.

Acknowledgements This note would not have been written without Professor Tom Hughes who told me that the non-symmetric Nitsche's method appeared to be stable without penalty in large-eddy simulations and pointed me to the reference [13]. I would also like to thank Professor Rolf Stenberg for interesting discussions on the subject of Nitsche's method.

error norm	$\gamma = 0$	$\gamma = 10$	$\gamma = 20$	$\gamma = 40$	$\gamma = 80$
$\|u - u_h\|_{L^2}$	2.1E-5	1.3E-5	1.2E-5	1.2E-5	1.2E-5
$\|u - u_h\|_{H^1}$	3.5E-3	3.5E-3	3.5E-3	3.5E-3	3.5E-3

TABLE 7.4

Study of the dependence of the accuracy on the penalty parameter, piece wise quadratic approximation, unstructured mesh, $N = 40$

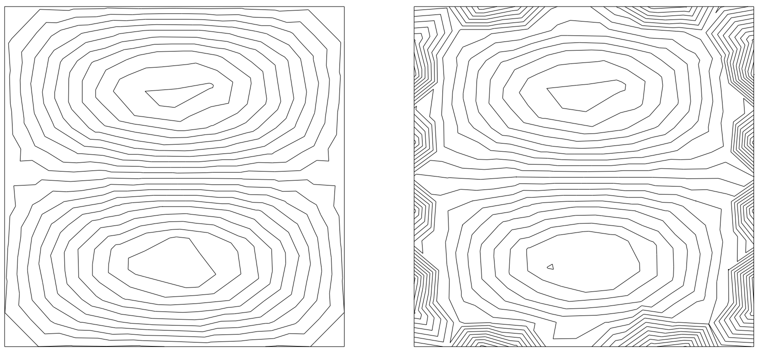

FIG. 7.1. *Comparison of the contourplots of the unstabilized non-symmetric method (left) and symmetric (right) method, piece wise affine approximation, N=10.*

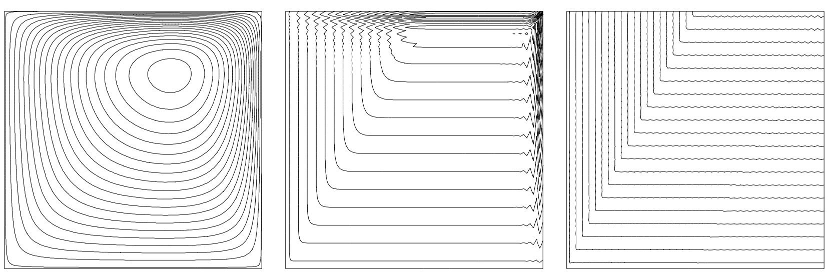

FIG. 7.2. *Convection-diffusion equation discretized using the non-symmetric Nitsche boundary condition, $N = 80$, piece wise affine approximation, from left to right: $\varepsilon = 0.1$, $\varepsilon = 0.001$, $\varepsilon = 0.00001$.*

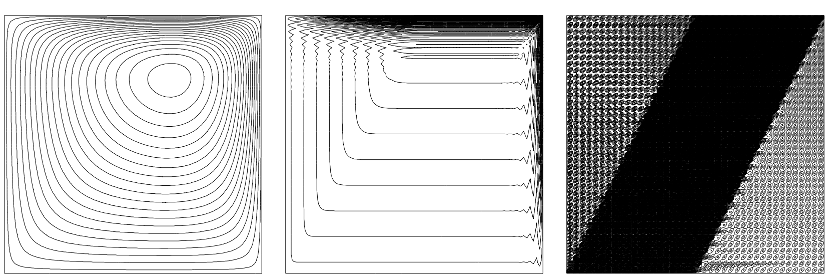

FIG. 7.3. *Convection-diffusion equation discretized using strongly imposed boundary condition, $N = 80$, piece wise affine approximation, from left to right: $\varepsilon = 0.1$, $\varepsilon = 0.001$, $\varepsilon = 0.00001$.*

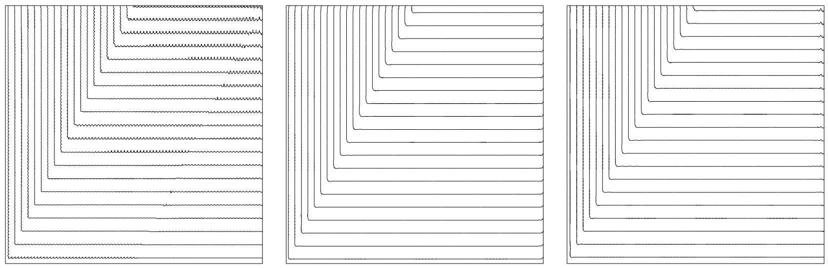

FIG. 7.4. *Convection-diffusion equation discretized using the non-symmetric Nitsche boundary condition, $N = 80$, $\varepsilon = 0.00001$, piece wise quadratic approximation, from left to right: no stabilization, SD-stabilization ($\gamma_{SD} = 0.5$), CIP-stabilization ($\gamma_{CIP} = 0.005$).*

REFERENCES

[1] Y. Bazilevs and T. J. R. Hughes. Weak imposition of Dirichlet boundary conditions in fluid mechanics. *Comput. & Fluids*, 36(1):12–26, 2007.

[2] R. Becker, P. Hansbo, and R. Stenberg. A finite element method for domain decomposition with non-matching grids. *M2AN Math. Model. Numer. Anal.*, 37(2):209–225, 2003.

[3] E. Burman. A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty. *SIAM J. Numer. Anal.*, 43(5):2012–2033 (electronic), 2005.

[4] E. Burman. Ghost penalty. *C. R. Math. Acad. Sci. Paris*, 348(21-22):1217–1220, 2010.

[5] E. Burman and A. Ern. Continuous interior penalty *hp*-finite element methods for advection and advection-diffusion equations. *Math. Comp.*, 76:1119–1140, 2007.

[6] E. Burman and P. Hansbo. Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 193(15-16):1437–1453, 2004.

[7] E. Burman and B. Stamm. Bubble stabilized discontinuous Galerkin method for parabolic and elliptic problems. *Numerische Mathematik*, 116:213–241, 2010. 10.1007/s00211-010-0304-9.

[8] E. Burman and P. Zunino. A domain decomposition method based on weighted interior penalties for advection-diffusion-reaction problems. *SIAM J. Numer. Anal.*, 44(4):1612–1638 (electronic), 2006.

[9] J. Freund and R. Stenberg. On weakly imposed boundary conditions for second order problems. In Morandi Cecchi et al., editor, *Proceedings of the Ninth Int. Conf. Finite Elements in Fluids*, pages 327–336, 1995.

[10] A. Gerstenberger and W. A. Wall. An embedded Dirichlet formulation for 3D continua. *Internat. J. Numer. Methods Engrg.*, 82(5):537–563, 2010.

[11] J. Guzmán and B. Rivière. Sub-optimal convergence of non-symmetric discontinuous Galerkin methods for odd polynomial approximations. *J. Sci. Comput.*, 40(1-3):273–280, 2009.

[12] F. Hecht, O. Pironneau, A. Le Hyaric, and K. Ohtsuka. *FreeFem++ v. 2.11. User's Manual.* University of Paris 6.

[13] T.J.R. Hughes, G. Engel, L. Mazzei, and M. Larson. Comparison of discontinuous and continuous Galerkin methods based on error estimates, conservation, robustness and efficiency. In G. Karniadakis et al., editor, *Discontinuous Galerkin Methods*. Springer, Berlin, 2000.

[14] M. G. Larson and A. J. Niklasson. Analysis of a nonsymmetric discontinuous Galerkin method for elliptic problems: stability and energy error estimates. *SIAM J. Numer. Anal.*, 42(1):252–264 (electronic), 2004.

[15] J. Nitsche. Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abh. Math. Sem. Univ. Hamburg*, 36:9–15, 1971. Collection of articles dedicated to Lothar Collatz on his sixtieth birthday.

[16] J. T. Oden, I. Babuška, and C. E. Baumann. A discontinuous *hp* finite element method for diffusion problems. *J. Comput. Phys.*, 146(2):491–519, 1998.

[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 J. Numer. Anal.*, 39(3):902–931, 2001.

[18] F. Schieweck. On the role of boundary conditions for CIP stabilization of higher order finite elements. *Electron. Trans. Numer. Anal.*, 32:1–16, 2008.

[19] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[20] R. Stenberg. On some techniques for approximating boundary conditions in the finite element method. *J. Comput. Appl. Math.*, 63(1-3):139–148, 1995. International Symposium on Mathematical Modelling and Computational Methods Modelling 94 (Prague, 1994).