ICES REPORT 12-21

June 2012

Robust DPG Method for Convection-Dominated Diffusion Problems II: A Natural In Flow Condition

by

Jesse Chan, Norbert Heuer, Tan Bui-Thanh, and Leszek Demkowicz

The Institute for Computational Engineering and Sciences
The University of Texas at Austin
Austin, Texas 78712

Reference: Jesse Chan, Norbert Heuer, Tan Bui-Thanh, and Leszek Demkowicz, Robust DPG Method for Convection-Dominated Diffusion Problems II: A Natural In Flow Condition, ICES REPORT 12-21, The Institute for Computational Engineering and Sciences, The University of Texas at Austin, June 2012.

Robust DPG method for convection-dominated diffusion problems II: a natural inflow condition

Jesse Chan[a], Norbert Heuer[b], Tan Bui-Thanh[a], and Leszek Demkowicz[a]

[a] Institute for Computational Engineering and Sciences,
University of Texas at Austin,
Austin, TX 78712, USA

[b] Facultad de Matemáticas,
Pontificia Universidad Católica de Chile,
Avenida Vicuña Mackenna 4860, Santiago, Chile

1. Introduction

1.1. Singular perturbation problems and robustness

The finite element/Galerkin method has been widely utilized in engineering to solve partial differential equations governing the behavior of physical phenomena in engineering problems. The method relates the solution of a partial differential equation (PDE) to the solution of a corresponding variational problem. The finite element method itself provides several advantages — a framework for systematic mathematical analysis of the behavior of the method, weaker regularity constraints on the solution than implied by the strong form of the equations, and applicability to very general physical domains and geometries.

Historically, the Galerkin method has been very successfully applied to a broad range of problems in solid mechanics, for which the variational problems resulting from the PDE are symmetric and coercive (positive-definite). It is well known that the finite element method produces optimal or near-optimal results for such problems, with the finite element solution matching or coming close to the best approximation of the solution in the finite element space. However, standard Bubnov-Galerkin methods tend to perform poorly for the class of PDEs known as singular perturbation problems. These problems are often characterized by a parameter that may be either very small or very large in the context of physical problems. An additional complication of singular perturbation problems is that very often, in the limiting case of the parameter blowing up or decreasing to zero, the PDE itself will change types (e.g. from elliptic to hyperbolic).

1.1.1. Convection-diffusion

A canonical example of a singularly perturbed problem is the convection-diffusion equation. In 1D, the convection-diffusion equation is

$$\beta u' - \epsilon u'' = f.$$

The equation represents the change in the concentration u of a quantity in a given medium, taking into account both convective and diffusive effects. β represents the speed of convection, while the singular perturbation parameter ϵ represents the diffusivity of the medium. In the limit of an inviscid medium as $\epsilon \to 0$, the equation changes types, from elliptic to hyperbolic, and from second

order to first order. For Dirichlet boundary conditions $u(0) = u_0$ and $u(1) = u_1$, the solution can develop sharp boundary layers of width ϵ near the outflow.

The poor performance of the finite element method for this problem is reflected in the bound on the error in the finite element solution — under the standard Bubnov-Galerkin method with $u \in H^1(0,1)$, we have the bound given in [20]:

$$\|u - u_h\|_\epsilon \leq C \inf_{w_h} \|u - w_h\|_{H^1(0,1)},$$

for $\|u\|_\epsilon^2 \coloneqq \|u\|_{L^2}^2 + \epsilon \|u'\|_{L^2}^2$, with C independent of ϵ. An alternative formulation of the above bound is

$$\|u - u_h\|_{H^1(0,1)} \leq C(\epsilon) \inf_{w_h} \|u - w_h\|_{H^1(0,1)},$$

where $C(\epsilon)$ grows as $\epsilon \to 0$. The dependence of the constant C on ϵ is referred to as a *loss of robustness* — as the singular perturbation parameter ϵ decreases, our finite element error is bounded more and more loosely by the best approximation error. As a consequence, the finite element solution can diverge significantly from the best finite element approximation of the solution for very small values of ϵ. For example, on a coarse mesh, and for small values of ϵ, the Galerkin approximation of the solution to the convection-diffusion equation with a boundary layer develops spurious oscillations everywhere in the domain, even where the best approximation error is small. These oscillations grow in magnitude as $\epsilon \to 0$, eventually polluting the entire solution.

1.1.2. Wave propagation

Another example of a singular perturbation problem which experiences loss of robustness is high frequency wave propagation, in which the singular perturbation parameter is the wavenumber k, where $k \to \infty$. The loss of robustness in this case manifests as "pollution" error, a phenomenon in which the finite element solution degrades over many wavelengths for high wavenumbers (commonly manifesting as a phase error between the FE solution and the exact solution).

1.1.3. Stabilization terms

Traditionally, instability/loss of robustness has been dealt with using residual-based stabilization techniques. Given some variational form, the problem is modified by adding to the bilinear form the strong form of the residual, weighted by a test function and scaled by a stabilization constant τ. The most well-known example of this technique is the streamline-upwind Petrov-Galerkin (SUPG) method, which is a stabilized method for solving the convection-diffusion equation using piecewise linear continuous finite elements [2]. SUPG stabilization not only removes the spurious oscillations from the finite element solution of the convection-diffusion equation, but delivers the best finite element approximation in the H^1 norm. An important difference between residual-based stabilization techniques and other stabilizations is the idea of *consistency* — by adding stabilization terms based on the residual, the exact solution still satisfies the same variational problem (i.e. Galerkin orthogonality still holds).[1]

The addition of residual-based stabilization terms can also be interpreted as a modification of the test functions — in other words, stabilization can be achieved by changing the test space for a given problem. We approach the idea of stabilization through the construction of *optimal test functions* to achieve optimal approximation properties.

[1] Contrast this to an artificial diffusion method, where a specific amount of additional viscosity is added based on the magnitude of the convection and diffusion parameters. The exact solution to the original equation no longer satisfies the new stabilized formulation.

1.2. Discontinuous Petrov-Galerkin methods with optimal test functions

Petrov-Galerkin methods, in which the test space differs from the trial space, have been explored for over 30 years, beginning with the approximate symmetrization method of Barrett and Morton [1]. The idea was continued with the SUPG method of Hughes, and the characteristic Petrov-Galerkin approach of Demkowicz and Oden [11], which introduced the idea of tailoring the test space to change the norm in which a finite element method would converge.

The idea of optimal test functions was introduced by Demkowicz and Gopalakrishnan in [8]. Conceptually, these optimal test functions are the natural result of the minimization of a residual corresponding to the operator form of a variational equation. The connection between stabilization and least squares/minimum residual methods has been observed previously [15]. However, the method in [8] distinguishes itself by measuring the residual of the natural *operator form of the equation*, which is posed in the dual space, and measured with the dual norm, as we now discuss.

Throughout the paper, we assume that the trial space U and test space V are real Hilbert spaces, and denote U' and V' as the respective topological dual spaces. Let $U_h \subset U$ and $V_h \subset V$ be finite dimensional subsets. We are interested in the following problem

$$\begin{cases} \text{Given } l \in V', \text{ find } u_h \in U_h \text{ such that} \\ b(u_h, v_h) = l(v_h), \quad \forall v_h \in V_h, \end{cases} \tag{1}$$

where $b(\cdot,\cdot) : U \times V \to \mathbb{R}$ is a continuous bilinear form. U is chosen to be some trial space of approximating functions, but V_h is as of yet unspecified.

Throughout the paper, we suppose the variational problem (1) to be well-posed. In that case, we can identify a unique operator $B : U \to V'$ such that

$$\langle Bu, v\rangle_V \coloneqq b(u,v), \quad u \in U, v \in V$$

with $\langle\cdot,\cdot\rangle_V$ denoting the duality pairing between V' and V, to obtain the operator form of the continuous variational problem

$$Bu = l \quad \text{in } V'. \tag{2}$$

In other words, we can represent the continuous form of our variational equation (1) equivalently as the operator equation (2) with values in the dual space V'. This motivates us to consider the conditions under which the solution to (1) is the solution to the minimum residual problem in V'

$$u_h = \arg\min_{u_h \in U_h} J(u_h),$$

where $J(w)$ is defined for $w \in U$ as

$$J(w) = \frac{1}{2}\|Bw - l\|_{V'}^2 \coloneqq \frac{1}{2}\sup_{v \in V\setminus\{0\}} \frac{|b(w,v) - l(v)|^2}{\|v\|_V^2}.$$

For convenience in writing, we will abuse the notation $\sup_{v\in V}$ to denote $\sup_{v\in V\setminus\{0\}}$ for the remainder of the paper.

Let us define $R_V : V \to V'$ as the Riesz map, which identifies elements of V with elements of V' by

$$\langle R_V v, \delta v\rangle_V \coloneqq (v, \delta v)_V, \quad \forall \delta v \in V.$$

Here, $(\cdot,\cdot)_V$ denotes the inner product in V. As R_V and its inverse, R_V^{-1}, are both isometries, e.g. $\|f\|_{V'} = \|R_V^{-1} f\|_V, \forall f \in V'$, we have

$$\min_{u_h \in U_h} J(u_h) = \frac{1}{2}\|Bu_h - l\|_{V'}^2 = \frac{1}{2}\left\|R_V^{-1}(Bu_h - l)\right\|_V^2. \tag{3}$$

The first order optimality condition for (3) requires the Gâteaux derivative to be zero in all directions $\delta u \in U_h$, iė;

$$\left(R_V^{-1}(Bu_h - l), R_V^{-1}B\delta u\right)_V = 0, \quad \forall \delta u \in U.$$

We define, for a given $\delta u \in U$, the corresponding *optimal test function* $v_{\delta u}$

$$v_{\delta u} \coloneqq R_V^{-1}B\delta u \quad \text{in } V. \tag{4}$$

The optimality condition then becomes

$$\langle Bu_h - l, v_{\delta u}\rangle_V = 0, \quad \forall \delta u \in U$$

which is exactly the standard variational equation in (1) with $v_{\delta u}$ as the test functions. We can define the optimal test space $V_{\text{opt}} \coloneqq \{v_{\delta u} \text{ s.t. } \delta u \in U\}$. Thus, the solution of the variational problem (1) with test space $V_h = V_{\text{opt}}$ minimizes the residual in the dual norm $\|Bu_h - l\|_{V'}$. This is the key idea behind the concept of optimal test functions.

Since $U_h \subset U$ is spanned by a finite number of basis functions $\{\varphi_i\}_{i=1}^N$, (4) allows us to compute (for each basis function) a corresponding optimal test function v_{φ_i}. The collection $\{v_{\varphi_i}\}_{i=1}^N$ of optimal test functions then forms a basis for the optimal test space. In order to express optimal test functions defined in (4) in a more familiar form, we take $\delta u = \varphi$, a generic basis function in U_h, and rewrite (4) as

$$R_V v_\varphi = B\varphi, \quad \text{in } V',$$

which is, by definition, equivalent to

$$(v_\varphi, \delta v)_V = \langle R_V v_\varphi, \delta v\rangle_V = \langle B\varphi, \delta v\rangle_V = b(\varphi, \delta v), \quad \forall \delta v \in V.$$

As a result, optimal test functions can be determined by solving the auxiliary variational problem

$$(v_\varphi, \delta v)_V = b(\varphi, \delta v), \quad \forall \delta v \in V. \tag{5}$$

However, in general, for standard H^1 and $H(\text{div})$-conforming finite element methods, test functions are continuous over the entire domain, and hence solving variational problem (5) for each optimal test function requires a global operation over the entire mesh, rendering the method impractical. A breakthrough came through the development of discontinous Galerkin (DG) methods, for which basis functions are discontinuous over elements. In particular, the use of discontinuous test functions δv reduces the problem of determining global optimal test functions in (5) to local problems that can be solved in an element-by-element fashion.

We note that solving (5) on each element exactly is infeasible since it amounts to inverting the Riesz map R_V exactly. Instead, optimal test functions are approximated using the standard Bubnov-Galerkin method on an "enriched" subspace $\tilde{V} \subset V$ such that $\dim(\tilde{V}) > \dim(U_h)$ elementwise [6, 8]. In this paper, we assume the error in approximating the optimal test functions is negligible, and refer to the work in [13] for estimating the effects of approximation error on the performance of DPG.

It is now well known that the DPG method delivers the best approximation error in the "energy norm" — that is [4, 8, 21]

$$\|u - u_h\|_{U,E} = \inf_{w \in U_h} \|u - w\|_{U,E}, \tag{6}$$

where the energy norm $\|\cdot\|_{U,E}$ is defined for a function $\varphi \in U$ as

$$\|\varphi\|_{U,E} \coloneqq \sup_{v \in V} \frac{b(\varphi, v)}{\|v\|_V} = \sup_{\|v\|_V = 1} b(\varphi, v) = \sup_{\|v\|_V = 1} \langle B\varphi, v\rangle_V = \|B\varphi\|_{V'} = \|v_\varphi\|_V, \tag{7}$$

where the last equality holds due to the isometry of the Riesz map R_V (or directly from (5) by taking the supremum). An additional consequence of adopting such an energy norm is that, without

knowing the exact solution, the energy error $\|u-u_h\|_{U,E}$ can be determined by computing $\|v_{u-u_h}\|_V$ from the following identity

$$(v_{u-u_h}, \delta v)_V = b(u-u_h, \delta v) = l(\delta v) - b(u_h, \delta v).$$

This is simply a consequence of the least-squares nature of DPG; the energy error is simply the norm of the residual in V'.

Practically speaking, this implies that the DPG method is discretely stable on any mesh. In particular, DPG is unconditionally stable for higher order adaptive meshes, where discrete stability is often an issue.

1.3. Duality between trial and test norms (energy norm pairings)

A clear property of the energy norm defined by (7) is that the trial norm $\|\cdot\|_{U,E}$ is induced by a given test norm. However, the reverse relationship holds as well; for any trial norm, the test norm that induces such a norm is recoverable through duality. We have a result, proved in [4]: assuming, for simplicity, that the bilinear form $b(u,v)$ is definite, given any norm $\|\cdot\|_U$ on the trial space U, for $\varphi \in U$, we can represent $\|\varphi\|_U$ via

$$\|\varphi\|_U = \sup_{v\in V} \frac{b(w,v)}{\|v\|_{V,U}}.$$

where $\|v\|_{V,U}$ is defined through

$$\|v\|_{V,U} = \sup_{w\in U} \frac{b(w,v)}{\|w\|_U}.$$

In particular, given two arbitrary norms $\|\cdot\|_{U,1}$ and $\|\cdot\|_{U,2}$ in U such that $\|\cdot\|_{U,1} \leq c\,\|\cdot\|_{U,2}$ for some constant c, they generate two norms $\|\cdot\|_{V,U,1}$ and $\|\cdot\|_{V,U,2}$ in V defined by

$$\|v\|_{V,U,1} \coloneqq \sup_{w\in U} \frac{b(w,v)}{\|w\|_{U,1}}, \quad \text{and } \|v\|_{V,U,2} \coloneqq \sup_{w\in U} \frac{b(w,v)}{\|w\|_{U,2}},$$

such that $\|\cdot\|_{V,U,1}$ and $\|\cdot\|_{V,U,2}$ induce $\|\cdot\|_{U,1}$ and $\|\cdot\|_{U,2}$ as energy norms in U, respectively. That is,

$$\|\varphi\|_{U,1} = \sup_{v\in V} \frac{b(\varphi,v)}{\|v\|_{V,U,1}}, \quad \text{and } \|\varphi\|_{U,2} = \sup_{v\in V} \frac{b(\varphi,v)}{\|v\|_{V,U,2}}.$$

A question that remains to be addressed is to establish the relationship between $\|\cdot\|_{V,U,1}$ and $\|\cdot\|_{V,U,2}$, given that $\|\cdot\|_{U,1} \leq c\,\|\cdot\|_{U,2}$. But this is straightforward since we have

$$\|v\|_{V,U,2} = \sup_{u\in U} \frac{b(w,v)}{\|w\|_{U,2}} \leq c \sup_{w\in U} \frac{b(w,v)}{\|w\|_{U,1}} = c\|v\|_{V,U,1}.$$

Consequently, a stronger energy norm in U will generate a weaker norm in V and vice versa. In other words, to show that an energy norm $\|\cdot\|_{U,1}$ is weaker than another energy norm $\|\cdot\|_{U,2}$ in U, one simply needs to show the reverse inequality on the corresponding norms in V, that is, $\|\cdot\|_{V,U,1}$ is stronger than $\|\cdot\|_{V,U,2}$.

From now on, unless otherwise stated, we will refer to $\|\cdot\|_{V,U}$ as the test norm that induces a given norm $\|\cdot\|_U$. Likewise, we will refer $\|\cdot\|_{U,V}$ as the trial norm induced by a given test norm $\|\cdot\|_V$. In this paper, for simplicity of exposition, we shall call a pair of norms in U and V that induce each other as an *energy norm pairing*.

1.4. Discontinuous Petrov-Galerkin methods with the ultra-weak formulation

The naming of the discontinuous Petrov-Galerkin method refers to the fact that the method is a Petrov-Galerkin method, and that the test functions are specified to be discontinuous across element boundaries. There is no specification of the regularity of the trial space, and we stress that the idea of DPG is not inherently tied to a single variational formulation [4].

In most of the DPG literature, however, the discontinuous Petrov-Galerkin method refers to the combination of the concept of locally computable optimal test functions in Section 1.2 with the so-called "ultra-weak formulation" [6, 8, 9, 21, 18, 17]. Unlike the previous two sections in which we studied the general equation (1) given by abstract bilinear and linear forms, we now consider a concrete instance of (1) resulting from an ultra-weak formulation for an abstract first-order system of PDEs $Au = f$. Additionally, from this section onwards, we will refer to DPG as the pairing of the ultra-weak variational formulation with the concept of locally computable optimal test functions.

We begin by partitioning the domain of interest Ω into N^{el} non-overlapping elements $K_j, j = 1, \ldots, N^{\text{el}}$ such that $\Omega_h = \cup_{j=1}^{N^{\text{el}}} K_j$ and $\overline{\Omega} = \overline{\Omega}_h$. Here, h is defined as $h = \max_{j \in \{1,\ldots,N^{\text{el}}\}} \operatorname{diam}(K_j)$. We denote the mesh "skeleton" by $\Gamma_h = \cup_{j=1}^{N^{\text{el}}} \partial K_j$; the set of all faces/edges e, each of which comes with a normal vector n_e. The internal skeleton is then defined as $\Gamma_h^0 = \Gamma_h \setminus \partial\Omega$. If a face/edge $e \in \Gamma_h$ is the intersection of ∂K_i and ∂K_j, $i \neq j$, we define the following jumps:

$$[\![v]\!] = \operatorname{sgn}\left(n^-\right) v^- + \operatorname{sgn}\left(n^+\right) v^+, \quad [\![\tau \cdot n]\!] = n^- \cdot \tau^- + n^+ \cdot \tau^+,$$

where

$$\operatorname{sgn}\left(n^\pm\right) = \begin{cases} 1 & \text{if } n^\pm = n_e \\ -1 & \text{if } n^\pm = -n_e \end{cases}.$$

For e belonging to the domain boundary $\partial\Omega$, we define

$$[\![v]\!] = v, \quad [\![\tau \cdot n]\!] = n_e \cdot \tau.$$

Note that we allow arbitrariness in assigning "-" and "+" quantities to the adjacent elements K_i and K_j.

The ultra-weak formulation for $Au = f$ on Ω_h, ignoring boundary conditions for now, reads

$$b\left((u, \widehat{u}), v\right) \coloneqq \langle \widehat{u}, [\![v]\!] \rangle_{\Gamma_h} - (u, A_h^* v)_{\Omega_h} = (f, v)_{\Omega_h}, \tag{8}$$

where we have denoted $\langle \cdot, \cdot \rangle_{\Gamma_h}$ as the duality pairing on Γ_h, $(\cdot, \cdot)_{\Omega_h}$ the L^2-inner product over Ω_h, and A_h^* the formal adjoint resulting from element-wise integration by parts. Occasionally, for simplicity in writing, we will ignore the subscripts in the duality pairing and L^2-inner product if they are Γ_h and Ω_h. Both the inner product and formal adjoint are understood to be taken element-wise. Using the ultra-weak formulation, the regularity requirement on solution variable u is relaxed, that is, u is now square integrable for the ultra-weak formulation (8) to be meaningful, instead of being (weakly) differentiable. The trade-off is that u does not admit a trace on Γ_h even though it did originally. Consequently, we need to introduce an additional new "trace" variable $\widehat{u}$ in (8), that is defined only on Γ_h.

The energy setting is now clear; namely,

$$u \in L^2(\Omega_h) \equiv L^2(\Omega), \quad v \in V = D(A_h^*), \quad \widehat{u} \in \gamma(D(A)),$$

where $D(A_h^*)$ denotes the broken graph space corresponding to A_h^*, and $\gamma(D(A))$ the trace space (assumed to exist) of the graph space of operator A. The first discussion of the well-posedness of DPG with the ultra-weak formulation can be found in [7], where the proof is presented for the Poisson and convection-diffusion equations. A more comprehensive discussion of the abstract setting for DPG with the ultra-weak formulation using the graph space, as well as a more general proof of well-posedness, can be consulted in [3].

1.5. Canonical energy norm pairings for ultra-weak formulation

From the discussion in Section 1.3 of energy norm and test norm pairings, we know that specifying either a test norm or trial norm is sufficient to define an energy pairing. In this section, we derive and discuss two important energy norm pairings, the first of which begins by specifying the canonical norm in U and inducing a test norm on V. The second pairing begins instead by specifying the canonical norm on V under the ultra-weak formulation (8) and inducing an energy norm on the trial space U.

We begin first with the canonical norm in U. Since $\widehat{u} \in \gamma\left(D\left(A\right)\right)$, the standard norm for $\widehat{u}$ is the so-called minimum energy extension norm defined as

$$\|\widehat{u}\| = \inf_{w \in D(A),\, w|_{\Gamma_h} = \widehat{u}} \|w\|_{D(A)}. \tag{9}$$

The canonical norm for the group variable $(u, \widehat{u})$ is then given by

$$\|\left(u, \widehat{u}\right)\|_U^2 = \|u\|_{L^2(\Omega)}^2 + \|\widehat{u}\|^2,$$

Applying the Cauchy-Schwarz inequality, we arrive at

$$b\left(\left(u, \widehat{u}\right), v\right) \leq \|(u, \widehat{u})\|_U \, \|v\|_{V,U}\,,$$

where

$$\|v\|_{V,U}^2 = \|A_h^* v\|_{L^2(\Omega)}^2 + \left(\sup_{\widehat{u} \in \gamma(D(A))} \frac{\langle \widehat{u}, [\![v]\!] \rangle_{\Gamma_h}}{\|\widehat{u}\|} \right)^2.$$

On the other hand, since $v \in D\left(A_h^*\right)$, the canonical norm for v is the broken graph norm:

$$\|v\|_V^2 = \|A_h^* v\|_{L^2(\Omega)}^2 + \|v\|_{L^2(\Omega)}^2\,.$$

Using the Cauchy-Schwarz inequality again, we obtain

$$b\left(\left(u, \widehat{u}\right), v\right) \leq \|(u, \widehat{u})\|_{U,V} \, \|v\|_V\,,$$

where

$$\|(u, \widehat{u})\|_{U,V}^2 = \|u\|_{L^2(\Omega)}^2 + \sup_{v \in D\left(A_h^*\right)} \frac{\langle \widehat{u}, [\![v]\!] \rangle_{\Gamma_h}^2}{\|v\|_V^2}, \tag{10}$$

Using the framework developed in [4], one can show that both pairs $\left(\|(u, \widehat{u})\|_U\,, \|v\|_{V,U}\right)$ and $\left(\|(u, \widehat{u})\|_{U,V}\,, \|v\|_V\right)$ are energy norm pairings in the sense discussed in Section 1.3. That is, the canonical norm $\|(u, \widehat{u})\|_U$ in U induces (generates) the norm $\|v\|_{V,U}$ in V, while the canonical norm $\|v\|_V$ in V induces (generates) the energy norm $\|(u, \widehat{u})\|_{U,V}$ in U. In the DPG literature [21], $\|v\|_{V,U}$ is known as the *optimal test norm*, while $\|v\|_V$ is known as the *quasi-optimal test norm*.

The canonical norm $\|(u, \widehat{u})\|_U$ in U provides a good balance between the standard norms on the field u and the flux $\widehat{u}$ [21]. As a result, if the induced norm $\|v\|_{V,U}$ (namely, the optimal test norm) is used to compute optimal test functions in (5), the finite element error in the canonical norm is the best in the sense of (6).

Unfortunately, the optimal test norm is non-localizable due to the presence of the jump term $[\![v]\!]$.[2] Since the jump terms couple elements together, the evaluation of the jump terms requires

[2] A localizable norm $\|v\|_{V(\Omega_h)}$ can be written in the form

$$\|v\|_{V(\Omega_h)} = \sum_{K \in \Omega_h} \|v\|_{V(K)}\,,$$

where $\|v\|_{V(K)}$ is a norm over the element K.

Trial norm		Test norm
$\boxed{\|u\|^2_{L^2(\Omega)} + \|\widehat{u}\|^2}$	$\Longrightarrow$	$\|A_h^* v\|^2_{L^2(\Omega)} + \left(\sup_{\widehat{u}} \frac{\langle \widehat{u}, [\![v]\!] \rangle_{\Gamma_h}}{\|\widehat{u}\|}\right)^2$
$\|u\|^2_{L^2(\Omega)} + \sup_v \left(\frac{\langle \widehat{u}, [\![v]\!] \rangle_{\Gamma_h}}{\|v\|_V}\right)^2$	$\Longleftarrow$	$\boxed{\|A_h^* v\|^2_{L^2(\Omega)} + \|v\|^2_{L^2(\Omega)}}$

Figure 1: A summary of the derivation of test/trial norm pairings; we begin with the boxed norm on either the trial or test space, and induce the norm on the other space through duality. The optimal *test* norm is naturally derived by beginning with the canonical norm on the trial space, while the quasi-optimal *trial* norm is derived from beginning with the canonical norm on the test space.

contributions from all the elements in the mesh. Consequently, solving for an optimal test function amounts to inverting the Riesz map over the entire mesh Ω_h, making the optimal test norm impractical.

On the other hand, the quasi-optimal test norm $\|v\|_V$, namely the canonical norm in V, *is* localizable, and hence practical. However, it's worth noting the difference between the induced energy norm $\|(u, \widehat{u})\|_{U,V}$ and the canonical norm in U; under the induced norm $\|(u, \widehat{u})\|_{U,V}$ there is no natural interpretation for the norm in which the error in the flux variable $\widehat{u}$ is measured.

Using a variant of the quasi-optimal test norm, numerical results show that the DPG method appears to provide a "pollution-free" method without phase error for the Helmholtz equation [21], and analysis of the pollution-free nature of DPG is currently under investigation. Similar results have also been obtained in the context of elasticity [18] and the linear Stokes equations [19]. On the theoretical side, the quasi-optimal test norm has been shown to yield a well-posed DPG methodology for the Poisson and convection-diffusion equations [7]. More recently, this theory has been generalized to show the well-posedness of DPG for the large class of PDEs of Friedrichs' type [3].

2. Model problem and robustness

The remainder of the paper focuses on a convection-diffusion model problem using the abstract theory that we have discussed so far. In particular, we shall use the DPG method based on the ultra-weak formulation with optimal test functions to solve the model problem and analyze its behavior with respect to ϵ. Our goal is to show the robustness of the method with respect to ϵ, and demonstrate its usefulness as a numerical approach to solving singular-perturbed problems.

We consider the following model convection-diffusion problem on a domain $\Omega \subset \mathbb{R}^d$ with boundary $\partial\Omega \equiv \Gamma$

$$\nabla \cdot (\beta u) - \epsilon \Delta u = f \in L^2(\Omega), \tag{11}$$

which can be cast into the first order form on the group variable (u, σ) as

$$A(u, \sigma) \coloneqq \begin{bmatrix} \nabla \cdot (\beta u - \sigma) \\ \frac{1}{\epsilon}\sigma - \nabla u \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix}. \tag{12}$$

Using the abstract ultra-weak formulation developed in Section 1.4 for the first order system of PDEs (12) we obtain

$$b\left(\left(u, \sigma, \widehat{u}, \widehat{f}_n\right), (v, \tau)\right) = (u, \nabla \cdot \tau - \beta \cdot \nabla v)_{\Omega_h} + \left(\sigma, \epsilon^{-1}\tau + \nabla v\right)_{\Omega_h} - \langle [\![\tau \cdot n]\!], \widehat{u} \rangle_{\Gamma_h} + \left\langle \widehat{f}_n, [\![v]\!] \right\rangle_{\Gamma_h},$$

where (v, τ) is the group test function. It should be pointed out that the divergence and gradient operators are understood to act element-wise on test functions (v, τ) in the broken graph space

$D(A_h^*) \coloneqq H^1(\Omega_h) \times H(\mathrm{div}, \Omega_h)$, but globally as usual on conforming test functions, i.e. $(v, \tau) \in H^1(\Omega) \times H(\mathrm{div}, \Omega)$. It follows that the canonical test norm can be written as

$$\| (v, \tau) \|_V^2 = \| (v, \tau) \|_{H^1(\Omega_h) \times H(\mathrm{div}, \Omega_h)}^2 = \sum_{K \in \Omega_h} \| (v, \tau) \|_{H^1(K) \times H(\mathrm{div}, K)}^2,$$

where

$$\| (v, \tau) \|_{H^1(K) \times H(\mathrm{div}, K)}^2 = \|v\|_{L^2(K)}^2 + \|\nabla v\|_{L^2(K)}^2 + \|\tau\|_{L^2(K)}^2 + \|\nabla \cdot \tau\|_{L^2(K)}^2.$$

In order to define the proper norm on the trial space, boundary conditions need to be specified. We begin by splitting the boundary Γ as follows

$$\begin{aligned} \Gamma_- &\coloneqq \{x \in \Gamma; \beta_n(x) < 0\}, \quad \text{(inflow)} \\ \Gamma_+ &\coloneqq \{x \in \Gamma; \beta_n(x) > 0\}, \quad \text{(outflow)} \\ \Gamma_0 &\coloneqq \{x \in \Gamma; \beta_n(x) = 0\}, \end{aligned}$$

where $\beta_n \coloneqq \beta \cdot n$. Previous work in [10] adopted Dirichlet boundary conditions everywhere on Γ. In this paper, we employ the inflow condition of Hesthaven *et al.* [14], where we set

$$\beta_n u - \sigma_n = u_0, \quad \text{on } \Gamma_-,$$

instead of $\beta_n u = u_0$. The former resembles the latter as ϵ approaches zero[3]; however, the latter induces a more "well-behaved" adjoint problem than the former, which, as we will discuss, affects the performance of DPG.

On the outflow boundary, we apply standard homogeneous Dirichlet boundary conditions

$$u = 0, \quad \text{on } \Gamma_+.$$

This paper is intended to act as an extension of work presented by Heuer and Demkowicz in [10]. The primary focus of the paper is to analyze the DPG method and extend previous results under this new choice of inflow boundary conditions. The difference in the performance of DPG under both new and old boundary conditions is connected to the difference in the adjoint problems induced under each boundary condition. The secondary contribution of this paper will be to analyze the performance of DPG under a new mesh-dependent test norm.

2.1. Norms on U

With the above boundary conditions at hand, the ultra-weak formulation (8) can be fitted in the abstract form (1) as

$$\begin{aligned} b\left(\left(u, \sigma, \widehat{u}, \widehat{f}_n\right), (v, \tau)\right) &= (u, \nabla \cdot \tau - \beta \cdot \nabla v)_{\Omega_h} + \left(\sigma, \epsilon^{-1}\tau + \nabla v\right)_{\Omega_h} \\ &\quad - \langle [\![\tau \cdot n]\!], \widehat{u} \rangle_{\Gamma_h \setminus \Gamma_+} + \left\langle \widehat{f}_n, [\![v]\!] \right\rangle_{\Gamma_h \setminus \Gamma_-} = (f, v) - \langle u_0, v \rangle_{\Gamma_-} = l\left((v, \tau)\right), \end{aligned}$$

which, after using the setting in Section 1.4, suggests the following trial space (see [7, 3] for details):

$$u, \sigma \in L^2(\Omega), \quad \text{and } \left(\widehat{u}, \widehat{f}_n\right) \in \gamma\left(D(A)\right) \subset \gamma\left(H^1(\Omega) \times H(\mathrm{div}, \Omega)\right) = H^{\frac{1}{2}}(\Gamma_h) \times H^{-\frac{1}{2}}(\Gamma_h).$$

The space for u and σ are simply scalar and vector L^2 spaces over Ω, while the space for $\left(\widehat{u}, \widehat{f}_n\right)$ is the trace space of the graph space of the operator A subject to the boundary conditions.

[3] For our model problem, as for many problems of interest in computational fluid dynamics, we expect ∇u to be small near the inflow, and that the solutions to (11) using $\beta_n u - \sigma_n = f_n = u_0$ on Γ_- will converge to that using $u = u_0$ on Γ_- for sufficiently small ϵ.

The minimum energy extension norm (9) now reads

$$\|\widehat{u}\| = \inf_{w\in H^1(\Omega),\, w|_{\Gamma_+}=0,\, w|_{\Gamma_h\setminus\Gamma_+}=\widehat{u}} \|w\|_{H^1(\Omega)},$$
$$\|\widehat{f_n}\| = \inf_{q\in H(\mathrm{div},\Omega),\, q\cdot n|_{\Gamma_-}=0,\, q\cdot n|_{\Gamma_h\setminus\Gamma_-}=\widehat{f_n}} \|q\|_{H(\mathrm{div},\Omega)}.$$

As a result, the canonical norm on U is given by

$$\left\|\left(u,\sigma,\widehat{u},\widehat{f_n}\right)\right\|_U^2 = \|u\|^2_{L^2(\Omega_h)} + \|\sigma\|^2_{L^2(\Omega_h)} + \|\widehat{u}\|^2 + \|\widehat{f_n}\|^2.$$

2.2. Norms on V

As $\tau \in H(\mathrm{div},\Omega_h)$ and $v \in H^1(\Omega_h)$, we will construct norms on v and τ which are equivalent to the canonical $H^1(K)\times H(\mathrm{div},K)$ norm over a single element

$$\|\,(v,\tau)\,\|^2_{H^1(K)\times H(\mathrm{div},K)} = \|v\|^2_{L^2(K)} + \|\nabla v\|^2_{L^2(K)} + \|\tau\|^2_{L^2(K)} + \|\nabla\cdot\tau\|^2_{L^2(K)}.$$

The squared norm over the entire triangulation Ω_h is defined to be the squared sum of contributions from each element

$$\|\,(v,\tau)\,\|^2_{H^1(\Omega_h)\times H(\mathrm{div},\Omega_h)} = \sum_{K\in\Omega_h} \|\,(v,\tau)\,\|^2_{H^1(K)\times H(\mathrm{div},K)}.$$

The exact norms that we will specify on V will be determined later.

The norms on the skeleton Γ_h for v and τ are defined by duality from the bilinear form

$$\|[\![\tau\cdot n]\!]\| = \|[\![\tau\cdot n]\!]\|_{\Gamma_h\setminus\Gamma_+} \coloneqq \sup_{w\in H^1(\Omega),\, w|_{\Gamma_+}=0} \frac{\langle [\![\tau\cdot n]\!], w\rangle}{\|w\|_{H^1(\Omega)}},$$
$$\|[\![v]\!]\| = \|[\![v]\!]\|_{\Gamma_h^0\cup\Gamma_+} \coloneqq \sup_{\eta\in H(\mathrm{div},\Omega),\, \eta\cdot n|_{\Gamma_-\cup\Gamma_0}=0} \frac{\langle [\![v]\!], \eta\cdot n\rangle}{\|\eta\|_{H(\mathrm{div},\Omega)}}.$$

2.3. Approximability of the quasi-optimal test norm

An obvious choice for the test norm would be the quasi-optimal norm; it is the canonical test norm, and DPG has been shown to be well-posed and robust under such an optimal test norm for a large class of problems [10, 3, 16]. However, computations with the quasi-optimal test norm for convection-diffusion problems turn out to be quite problematic for small diffusion and coarse meshes.

For convection-diffusion, the quasi-optimal test norm is

$$\|\,(v,\tau)\,\|^2_V = \|\nabla\cdot\tau - \beta\cdot\nabla v\|^2_{L^2} + \|\epsilon^{-1}\tau + \nabla v\|^2_{L^2} + \|v\|^2_{L^2} + \|\tau\|^2_{L^2}.$$

Use of this norm for the convection-diffusion problem is difficult — since the problem (5) for optimal test functions is local, we can transform the problem over a single element K to the reference element $\widehat{K}$ and show that it is equivalent to a reaction-diffusion system, with diffusion parameter $\frac{\epsilon}{|K|}$, where $|K|$ is the element measure [17]. We refer to the inverse of this parameter $\frac{|K|}{\epsilon}$ as the element Peclet number Pe. For a coarse mesh and small diffusion parameter ϵ, we will have a large element Peclet number, and optimal test functions under the quasi-optimal test norm will develop strong boundary layers of width Pe, as seen in Figure 2.

In the application of DPG in [6, 8, 9, 21], the approximation of optimal test functions is done using polynomial enrichment. We search for the solution to (5) in the enriched test space $\tilde{V} \approx$

$\prod_K P^{p+\Delta p}(K)$, where p is the polynomial order of the trial space on a given element K.[4] In other words, optimal test functions are approximated element-by-element using polynomials whose order is Δp more than the local order of approximation. Under this scheme, the error in approximation of test functions is tied to the effectiveness of the p-method. Unfortunately, for problems with boundary layers — including the approximation of test functions under the quasi-optimal test norm — the p-method performs very poorly. As a result of this poor approximation, the numerical solutions of the convection-dominated diffusion equation under DPG using the quasi-optimal test norm tend to be of poor quality, and do not exhibit all the proven properties of DPG (for example, the energy error may increase after mesh refinement, even though, by virtue of DPG delivering a best approximation, the energy error for a coarse mesh must be greater than or equal to the energy error for a finer mesh). We conclude that the error in approximation of optimal test functions using simple polynomial enrichment pollutes and ruins the performance of DPG under the quasi-optimal test norm.

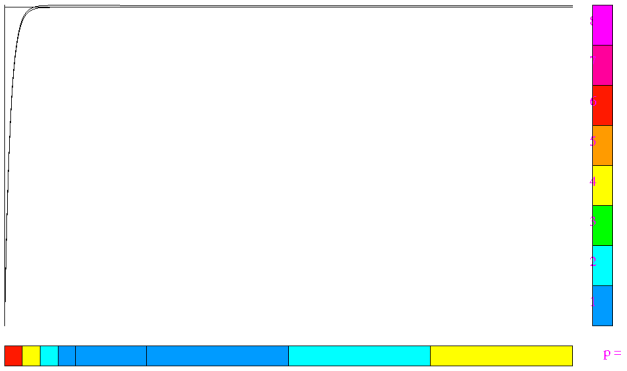


Figure 2: v and τ components of the 1D optimal test functions corresponding to the flux $\widehat{f}_n$ on the right-hand side of a unit element for $\epsilon = 0.01$. The solution has been obtained using automatic hp-adaptivity driven by the test norm with the error tolerance set at 1%.

The difficulty in using the quasi-optimal test norm for convection-diffusion is perplexing at first, considering that the quasi-optimal norm has yielded excellent results for the Helmholtz equation and other wave propagation problems. The difference between the two problems lies in the fact that, for wave propagation problems, the mesh size tends to be on the order of the wavenumber k — the singular perturbation parameter. Transforming the variational problem using the quasi-optimal test norm for wave propagation yields smooth optimal test functions that are approximated much more accurately using only polynomials over the reference element. Typically, the wavenumbers k of physical interest are $O(100)$ with respect to a unit domain. The corresponding finite element problems will typically be solved on meshes containing approximately $O\big(k^d\big)$ elements in $\mathbb{R}^d$, well within the range of a computationally tractable simulation. However, for convection diffusion problems, the relevant range of ϵ for physical problems can be as small as $1e-7$. Solving on under-resolved meshes is thus unavoidable, and the approximability of optimal test functions must be addressed in order to take advantage of the properties of DPG.

Resolving such boundary layers present in test functions under the quasi-optimal test norm has been investigated numerically using specially designed (Shishkin) subgrid meshes by Niemi, Collier, and Calo in [17]. However, even with Shishkin meshes, the approximation of optimal test functions under the quasi-optimal norm is far more expensive and complex to implement than approximation of test functions using a simple p-enriched space for V. We therefore aim instead to design a test

[4] V is only *approximately* equal to the space $\prod_K P^{p+\Delta p}(K)$. In practice, V is constructed using locally H^1-conforming and Raviart-Thomas elements of appropriate order.

norm that does not induce boundary layers, but still delivers good approximation results over a range of ϵ.

3. Analysis of a DPG test norm

We are interested in computing DPG optimal test functions for the convection-diffusion equation with very small values of ϵ; due to the difficulty of approximating optimal test functions, we conclude that the use of the quasi-optimal test norm is infeasible towards this goal.

However, if we naively choose a test norm that does not generate boundary layers, the performance of DPG may be adversely affected. For example, if $\| (v,\tau) \|_V^2 = \|v\|_{H^1(\Omega_h)}^2 + \|\tau\|_{H(\mathrm{div},\Omega_h)}^2$, the $H^1(\Omega_h) \times H(\mathrm{div},\Omega_h)$ norm, then the corresponding test functions will be smooth and free of boundary layers; however, the performance of DPG will provide approximations which worsen in quality as ϵ becomes very small [9, 10].

Our goal is to construct a test norm that compromises between performance of DPG and approximability of test functions. This test norm should not produce boundary layers in the optimal test functions, but still induce an energy norm that yields good approximation properties for small ϵ. We note that, even under the quasi-optimal norm, the norms on the flux and trace variables will likely depend on ϵ. Thus, we aim to construct a test norm for which the DPG method will be robust in ϵ with respect to the *field variables*.

For now, we discuss the steps necessary to analyze the performance of DPG with respect to a non-canonical test norm. We require a priori that the test norm has separable τ and v components — in other words, that there are no terms in the test norm that couple τ and v together. Problem (5) then decouples, such that the components of the vector-valued test function (v,τ) can be solved for independently of each other. The decoupled variational problems are no longer systems but scalar equations in τ and v, for which it is easier to conclude whether or not there are boundary layers in the solutions (the avoidance of boundary layers in the test norm will be discussed in more detail in Section 4, which describes our numerical experiments). **This will ensure that the resulting DPG method does not suffer from approximation errors in the optimal test functions**.

We begin with the following test norm:

$$\|(v,\tau)\|_V^2 \coloneqq \|v\|_{L^2}^2 + \epsilon\|\nabla v\|_{L^2}^2 + \|\beta\cdot\nabla v\|_{L^2}^2 + \frac{1}{\epsilon}\|\tau\|_{L^2}^2 + \|\nabla\cdot\tau\|_{L^2}^2.$$

The use of this norm is problematic for practical computations; we will discuss the reasons why and present a modification of it in Section 3.3.

We can see how this norm will differ from the canonical $H^1(\Omega_h)\times H(\mathrm{div},\Omega_h)$ norm: the clearest difference is the fact that the gradient in the streamline direction is $O(1)$, while the full gradient is $O(\sqrt{\epsilon})$, so that, in our test norm, the streamline gradient of v will be emphasized over the full gradient of v for small ϵ.

The choice of this test norm is implied by the mathematics of the adjoint problem. Roughly speaking, necessary conditions for the performance of DPG to not degenerate as $\epsilon\to 0$ are derived through analysis of specific test functions. For example, if u is the first L^2 component of the solution to the variational problem defined in Section 2, by choosing $(v,\tau)\in H^1(\Omega)\times H(\mathrm{div},\Omega)$ such that

$$\begin{aligned}\nabla\cdot\tau - \beta\cdot\nabla v &= u\\ \frac{1}{\epsilon}\tau - \nabla v &= 0,\end{aligned}$$

we have

$$\|u\|_{L^2}^2 = b\left(\left(u,\sigma,\widehat{u},\widehat{f_n}\right),(v,\tau)\right) \le \left\|\left(u,\sigma,\widehat{u},\widehat{f_n}\right)\right\|_{U,V}\|(v,\tau)\|_V,$$

and we recover the L^2 norm of u from the bilinear form.

Let $\|a\| \lesssim \|b\|$ denote an ϵ-independent bound; specifically, that $\|a\| \leq C\|b\|$ for a constant C independent of ϵ. Consequently, if for any $u \in L^2(\Omega_h)$, $\|(v,\tau)\|_V \lesssim \|u\|_{L^2}$, then dividing through by $\|u\|_{L^2}$ gives the bound

$$\|u\|_{L^2} \lesssim \left\|\left(u, \sigma, \widehat{u}, \widehat{f_n}\right)\right\|_E.$$

In other words, there is the guarantee that the L^2 error in u is at least robustly bounded from above by the energy error. Then, if the energy error (which DPG minimizes) approaches zero, the L^2 error in u will as well. The same exercise can be repeated for the stress σ, as well as the flux variables $\widehat{u}$, $\widehat{f_n}$.

This methodology gives constraints on the quantities found in the test norm; any quantity present in $\|(v,\tau)\|_V$ must be shown to be bounded from above independently of ϵ by the load of the adjoint problem. However, showing this simply amounts to showing *standard energy estimates* for H^1 and $H(\text{div})$-conforming finite elements. A more detailed discussion on the reasoning behind the construction of test norms can be found in [10].

The second step will be to **show the equivalence of the energy norm to explicit norms on** U. Since we do not generally have a closed form expression for the DPG energy norm, we seek to understand the behavior of DPG by finding a norm on U to which the DPG energy norm is equivalent. Since $\left(u, \sigma, \widehat{u}, \widehat{f_n}\right) \in U$ is a group variable from a tensor product space, we construct norms on U through the combination of norms on u, σ, $\widehat{u}$, and $\widehat{f_n}$. Specifically, we use the norm on U

$$\left\|\left(u, \sigma, \widehat{u}, \widehat{f_n}\right)\right\|_U^2 \coloneqq \|u\|^2 + \|\sigma\|^2 + \|\widehat{u}\|^2 + \left\|\widehat{f_n}\right\|^2. \tag{13}$$

For equivalence between norms, two constants are specified. However, since this norm on U is a norm on four separate variables, we can specify not just two but eight equivalence constants.[5] In order to simplify analysis, we phrase this equivalence statement in an alternative form.

Let $\|\cdot\|_E \coloneqq \|\cdot\|_{U,V}$, the energy norm induced by the test norm described above. We seek the bound of $\|\cdot\|_E$ from above and below:

$$\left\|\left(u, \sigma, \widehat{u}, \widehat{f_n}\right)\right\|_{U,1} \lesssim \left\|\left(u, \sigma, \widehat{u}, \widehat{f_n}\right)\right\|_E \lesssim \left\|\left(u, \sigma, \widehat{u}, \widehat{f_n}\right)\right\|_{U,2},$$

where both $\|\cdot\|_{U,1}$ and $\|\cdot\|_{U,2}$ are defined as scaled combinations of the norms on $u, \sigma, \widehat{u}$, and $\widehat{f_n}$

$$\left\|\left(u, \sigma, \widehat{u}, \widehat{f_n}\right)\right\|_{U,i}^2 \coloneqq \left(C_u^i\|u\|\right)^2 + \left(C_\sigma^i\|\sigma\|\right)^2 + \left(C_{\widehat{u}}^i\|\widehat{u}\|\right)^2 + \left(C_{\widehat{f}_n}^i\|\widehat{f_n}\|\right)^2, \quad i = 1,2 \tag{14}$$

Our goal is to explicitly derive the equivalence constants that define the norms $\|\cdot\|_{U,1}$ and $\|\cdot\|_{U,2}$ respectively, taking into account any dependency on ϵ. To do so, we need a relation between trial norms on U and test norms on V.

Recall from Section 1.3 that every test norm induces a corresponding trial norm, and vice versa. Let $\|\cdot\|_{U,1} \simeq \|\cdot\|_{U,2}$ mean that the norms $\|\cdot\|_{U,1}$ and $\|\cdot\|_{U,2}$ are equivalent, with equivalence constants independent of ϵ. By equivalence of finite dimensional norms and the discussion in Section 1.3 on the duality between test norms/energy norms, the norms (14) on U induce the equivalent test

[5] Sharper estimates are attainable if these constants are allowed to vary over the mesh Ω_h. See Section 3.4 for a discussion.

norms on $(v,\tau) \in H^1(\Omega_h) \times H(\mathrm{div}, \Omega_h)$

$$\begin{aligned}
\|(v,\tau)\|_{V,U,i} &\simeq \sup_{(u,\sigma,\widehat{u},\widehat{f}_n)\in U} \frac{b\left(\left(u,\sigma,\widehat{u},\widehat{f}_n\right),(v,\tau)\right)}{C_u^i\|u\| + C_\sigma^i\|\sigma\| + C_{\widehat{u}}^i\|\widehat{u}\| + C_{\widehat{f}_n}^i\|\widehat{f}_n\|} \\
&= \sup_{(u,\sigma,\widehat{u},\widehat{f}_n)\in U} \frac{(u, \nabla\cdot\tau - \beta\cdot\nabla v) + \left(\sigma, \epsilon^{-1}\tau + \nabla v\right) - \langle [\![\tau_n]\!], \widehat{u}\rangle_{\Gamma_-\cup\Gamma_h^0} + \langle \widehat{f}_n, [\![v]\!]\rangle_{\Gamma_+\cup\Gamma_h^0}}{C_u^i\|u\| + C_\sigma^i\|\sigma\| + C_{\widehat{u}}^i\|\widehat{u}\| + C_{\widehat{f}_n}^i\|\widehat{f}_n\|} \\
&\simeq \frac{1}{C_u^i}\|g\| + \frac{1}{C_\sigma^i}\|f\| + \frac{1}{C_{\widehat{u}}^i}\sup_{\widehat{u}\neq 0,\, \widehat{u}|_{\Gamma_+}=0} \frac{\langle [\![\tau\cdot n]\!], \widehat{u}\rangle}{\|\widehat{u}\|} + \frac{1}{C_{\widehat{f}_n}^i}\sup_{\widehat{f}_n\neq 0,\, \widehat{f}_n|_{\Gamma_-}=0} \frac{\langle \widehat{f}_n, [\![v]\!]\rangle}{\|\widehat{f}_n\|},
\end{aligned}$$

where f and g are defined element-wise over Ω_h as

$$\begin{aligned}
g &\coloneqq \nabla\cdot\tau - \beta\cdot\nabla v \\
f &\coloneqq \epsilon^{-1}\tau + \nabla v.
\end{aligned}$$

By definition of the norms on the quantities defined on the skeleton Γ_h, this gives the characterization of the induced test norm

$$\|\,(v,\tau)\,\|_{V,U,i} \simeq \frac{1}{C_u^i}\|g\| + \frac{1}{C_\sigma^i}\|f\| + \frac{1}{C_{\widehat{u}}^i}\|\,[\![\tau\cdot n]\!]\,\| + \frac{1}{C_{\widehat{f}_n}^i}\|\,[\![v]\!]\,\|, \quad i = 1,2.$$

We can now use this relation to compare different norms on U by comparing their induced norms on V (recall that showing a robust inequality between two norms on U is equivalent to showing the robust *reverse* inequality in the induced norms on V). Namely, we can show the bound of $\|\cdot\|_{U,1} \lesssim \|\cdot\|_E$ by showing the bound $\|(v,\tau)\|_{V,U,1} \gtrsim \|(v,\tau)\|_V$, and likewise for $\|\cdot\|_E \lesssim \|\cdot\|_{U,2}$.

Since the techniques used to show such bounds are more involved, we break the procedure up into two steps:

1. Decompose test functions (v,τ) into three separate, more easily analyzable components (Section 3.1).
2. Derive adjoint estimates (Section 3.2).

3.1. Decomposition into analyzable components

Having reduced the problem of comparing norms on U to the comparison of norms on V, we break the analysis of $(v,\tau) \in V$ into the analysis of three subproblems. Define the decomposition

$$(v,\tau) = (v_0,\tau_0) + (v_1,\tau_1) + (v_2,\tau_2)\,,$$

where (v_1,τ_1) satisfies

$$\begin{aligned}
\epsilon^{-1}\tau_1 + \nabla v_1 &= 0, \\
\nabla\cdot\tau_1 - \beta\cdot\nabla v_1 &= \nabla\cdot\tau - \beta\cdot\nabla v = g,
\end{aligned}$$

and (v_2,τ_2) satisfies

$$\begin{aligned}
\epsilon^{-1}\tau_2 + \nabla v_2 &= \epsilon^{-1}\tau + \nabla v = f, \\
\nabla\cdot\tau_2 - \beta\cdot\nabla v_2 &= 0.
\end{aligned}$$

Both $(v_1,\tau_1), (v_2,\tau_2) \in H(\mathrm{div};\Omega)\times H^1(\Omega)$ are understood to satisfy these relations in a conforming sense over the domain Ω; however, the divergence of τ and gradient of v on the right hand side are still understood to be taken in an element-wise fashion.

We will additionally require both $(v_1, \tau_1), (v_2, \tau_2)$ to satisfy the adjoint homogeneous boundary conditions

$$\tau_i \cdot n = 0, \quad \text{on } \Gamma_- \tag{15}$$
$$v_i = 0, \quad \text{on } \Gamma_+ \tag{16}$$

for $i = 1, 2$. The selection of $H(\text{div}, \Omega) \times H^1(\Omega)$ conforming test functions satisfying the specific boundary conditions above removes the contribution of the jump terms over the skeleton Γ_h in the bilinear form, allowing us to analyze field terms in the induced test norms separately from the boundary/jump terms.

Finally, by construction, $(v_0, \tau_0) \in H^1(\Omega_h) \times H(\text{div}, \Omega_h)$ must satisfy

$$\epsilon^{-1}\tau_0 + \nabla v_0 = 0$$
$$\nabla \cdot \tau_0 - \beta \cdot \nabla v_0 = 0$$

with jumps

$$[\![v_0]\!] = [\![v]\!], \quad \text{on } \Gamma_h^0$$
$$[\![\tau_0 \cdot n]\!] = [\![\tau \cdot n]\!], \quad \text{on } \Gamma_h^0.$$

and boundary conditions

$$v_0 = v, \quad \text{on } \Gamma_+$$
$$\tau_0 \cdot n = \tau \cdot n, \quad \text{on } \Gamma_- \cup \Gamma_0.$$

Notice that the evaluation the bilinear form $b\left(\left(u, \sigma, \widehat{u}, \widehat{f}_n\right), (v, \tau)\right)$ with each specific test functions returns only one part of the bilinear form. Furthermore, by choosing the proper loads $g = u$ and $f = \sigma$, we can recover from the bilinear form the norms of u and σ (as described in Section 3), as well as the norms on $\widehat{u}$, and $\widehat{f}_n$.[6]

We have now decomposed an arbitrary test function (τ, v) into a discontinuous contribution and two continuous contributions. Recall that our goal is to show the robust bound from above and below of the DPG energy norm by $\|\cdot\|_{U,1}$ and $\|\cdot\|_{U,2}$:

$$\left\|\left(u, \sigma, \widehat{u}, \widehat{f}_n\right)\right\|_{U,1} \lesssim \left\|\left(u, \sigma, \widehat{u}, \widehat{f}_n\right)\right\|_E \lesssim \left\|\left(u, \sigma, \widehat{u}, \widehat{f}_n\right)\right\|_{U,2}.$$

Under the duality of trial and test norms and the decomposition of test functions $(\tau, v) \in V$ into $(\tau_0, v_0), (\tau_1, v_1)$, and (τ_2, v_2), the above bound is equivalent to bounding each component

$$\|(v, \tau)\|_{V,U,1} \gtrsim \sum_{i=0}^{2} \|(v_i, \tau_i)\|_V \gtrsim \|(v, \tau)\|_{V,U,2}.$$

Bounding $\|(v_0, \tau_0)\|$ requires the use of techniques first developed in [7] and adapted to convection-diffusion in [7] and [10]. However, since $(\tau, v) \in H(\text{div}, \Omega) \times H^1(\Omega)$, the bound from above of test functions $\|(v_1, \tau_1)\|_V$ and $\|(v_2, \tau_2)\|_V$ is reduced to proving classical error estimates for the adjoint equations

$$\epsilon^{-1}\tau_1 + \nabla v_1 = 0$$
$$\nabla \cdot \tau_1 - \beta \cdot \nabla v_1 = g,$$
$$\tau_1 \cdot n|_{\Gamma_-} = 0,$$
$$v_1|_{\Gamma_+} = 0.$$

[6] To recover the norms on $\widehat{u}$, and $\widehat{f}_n$, the loads f, and g must be zero, and the jumps of the test function (v, τ) must be chosen specifically.

and

$$\begin{aligned}\epsilon^{-1}\tau_2 + \nabla v_2 &= f\\ \nabla\cdot\tau_2 - \beta\cdot\nabla v_2 &= 0,\\ \tau_2\cdot n|_{\Gamma_-} &= 0,\\ v_2|_{\Gamma_+} &= 0.\end{aligned}$$

More generally, we can analyze the adjoint equations

$$\epsilon^{-1}\tau + \nabla v = f \tag{17}$$
$$\nabla\cdot\tau - \beta\cdot\nabla v = g, \tag{18}$$

for arbitrary data $f, g \in L^2(\Omega)$ and boundary conditions $[\![\tau\cdot n]\!]_{\Gamma_-} = 0$ and $[\![v]\!]_{\Gamma_+} = 0$. In other words, we want to analyze the stability properties of the adjoint equations by deriving bounds of the form $\|\,(v_1,\tau_1)\,\|_V \lesssim \|g\|_{L^2}$ and $\|\,(v_2,\tau_2)\,\|_V \lesssim \|f\|_{L^2}$.

3.2. Adjoint estimates

The final step to estimating the induced norm on U by a selected localizable test norm on V is to derive adjoint stability estimates on τ and v in terms of localizable normed quantities. We will construct complete test norms on V through combinations of these normed quantities.

We introduce first the bounds derived; the proofs will be given later. For this analysis, it will be necessary to assume certain technical conditions on β. For each proof, we require $\beta \in C^2(\bar{\Omega})$ and $\beta, \nabla\cdot\beta = O(1)$. Additionally, we will assume that some or all of the following assumptions hold:

$$\nabla\times\beta = 0, \quad 0 < C \le |\beta|^2 + \frac{1}{2}\nabla\cdot\beta, \quad C = O(1), \tag{19}$$
$$\nabla\beta + \nabla\beta^T - \nabla\cdot\beta I = O(1), \tag{20}$$
$$\nabla\cdot\beta = 0. \tag{21}$$

Under proper assumptions on β, we have the robust bounds, which are proved in the Appendix.

- **Lemma 2**: *For β satisfying* (19) *and* (20)*, and $v_1 \in H^1(\Omega)$, satisfying equations* (17) *and* (18) *with $f = 0$, and with boundary conditions* (15) *and* (16),

  $$\|\beta\cdot\nabla v_1\| \lesssim \|g\|.$$

  Similarly, from $\nabla\cdot\tau_1 - \beta\cdot\nabla v_1 = g$, we get $\|\nabla\cdot\tau_1\| \lesssim \|g\|$ as well.

- **Lemma 3**: *For β satisfying* (19)*, and $v \in H^1(\Omega)$ satisfying equations* (17) *and* (18) *and boundary conditions* (15) *and* (16)*, and for sufficiently small ϵ,*

  $$\epsilon\|\nabla v\|^2 + \|v\|^2 \lesssim \|g\|^2 + \epsilon\|f\|^2.$$

  We can characterize both v_1 and v_2 in the above decompositions using this theorem by setting either $f = 0$ or $g = 0$.

- **Lemma 4**: *For β satisfying* (19)*,* (21)*, and solutions $v_0 \in H^1(\Omega_h)$ and $\tau_0 \in H(\mathrm{div}, \Omega_h)$ of equations* (17) *and* (18) *with $f = g = 0$,*

  $$\|\nabla v_0\| = \frac{1}{\epsilon}\|\tau_0\| \lesssim \frac{1}{\epsilon}\|\,[\![\tau_0\cdot n]\!]\,\|_{\Gamma_h\setminus\Gamma_+} + \frac{1}{\sqrt{\epsilon}}\|\,[\![v_0]\!]\,\|_{\Gamma_h^0\cup\Gamma_+}.$$

We are interested in showing the equivalence of the DPG energy norm with norms $\|\cdot\|_{U,1}$ and $\|\cdot\|_{U,2}$, respectively. We will show this by bounding $\|\cdot\|_V$ from below by $\|\cdot\|_{V,U,1}$ and from above by $\|\cdot\|_{V,U,2}$ (the induced test norms for $\|\cdot\|_{U,1}$ and $\|\cdot\|_{U,2}$, respectively).

3.3. A mesh-dependent test norm

Ideally, we would be interested in the use of the test norm

$$\| (v,\tau) \|_V^2 = \|v\|^2 + \epsilon\|\nabla v\|^2 + \|\beta\cdot\nabla v\|^2 + \|\nabla\cdot\tau\|^2 + \frac{1}{\epsilon}\|\tau\|^2$$

for practical computations. However, the presence of the term $\|v\|$ together with $\sqrt{\epsilon}\|\nabla v\|$ (and similarly $\|\nabla\cdot\tau\|$ and $\frac{1}{\sqrt{\epsilon}}\|\tau\|$ terms) induces boundary layers in the optimal test functions for under-resolved meshes. We can see this by recovering the strong form of the variational problem defining test functions. We first note that the variational problems for the v and τ components of optimal test functions decouple from each other under this test norm. Then, examining the variational problem for the v component only of an optimal test function, and assuming $\nabla\cdot\beta = 0$ for illustrative purposes, we have

$$\begin{aligned}((v,0),(\delta v,\delta\tau))_V &= (v,\delta v) + \epsilon\,(\nabla v,\nabla\delta v) + (\beta\cdot\nabla v,\beta\cdot\nabla\delta v)\\ &= (v - \epsilon\Delta v - \nabla\cdot((\beta\otimes\beta)\,\nabla v)\,,\delta v)_{L^2} + \langle\epsilon\nabla v\cdot n,\delta v\rangle + \langle n\cdot(\beta\otimes\beta)\,\nabla v,\delta v\rangle.\end{aligned}$$

After integration by parts, we recover the strong form of the operator L inducing such a variational problem

$$Lv \coloneqq v - \epsilon\Delta v - \nabla\cdot((\beta\otimes\beta)\,\nabla v)\,,$$

where we neglect the resulting boundary terms from integration by parts for now.

The streamline direction β induces an anisotropic diffusion, while the $\sqrt{\epsilon}\|\nabla v\|_{L^2}$ term induces a small isotropic diffusion contribution everywhere. Since any vector in the cross-stream direction is in the null space of the anisotropic diffusion tensor, in the cross-stream directions, the optimal test function is governed only by the cross-stream part of the operator L

$$L_{\beta^\perp} \coloneqq v - \epsilon\Delta v,$$

and can develop boundary layers in those directions. The presence of boundary layers has been verified through numerical computation as well; using an H^1-conforming finite element code with *hp*-adaptivity [5], the solution to the variational problem defining the optimal test function under the above test norm was computed. Figure 3 shows the result of such a computation for the v component of an optimal test function under the above test norm. To avoid boundary layers in

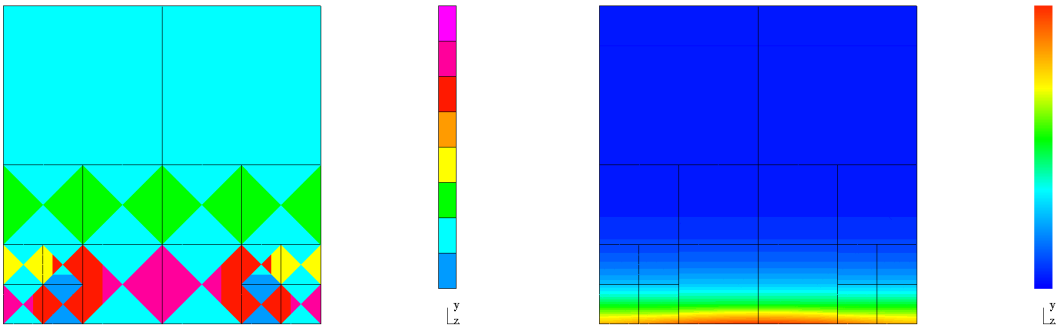


Figure 3: The v component of the optimal test function corresponding to flux $\widehat{u} = x(1-x)$ on the bottom side of a unit element for $\epsilon = 0.01$. The corresponding *hp*-mesh used to compute the solution is displayed to the left.

the optimal test functions, we follow [10] in scaling the L^2 contributions of v by $C_v(K)$, such that, when transformed to the reference element, both $C_v(K)\|v\|^2$ and $\epsilon\|\nabla v\|^2$ are of the same magnitude.

Similarly, we scale the L^2 contributions of τ by $C_\tau(K)$ such that $\frac{C_\tau(K)}{\epsilon}\|\tau\|^2$ and $\|\nabla\cdot\tau\|^2$ are of the same magnitude as well. For this paper, we consider only isotropic refinements on quadrilateral elements in 2D.

Our test norm, as defined over a single element K, is now

$$\|(v,\tau)\|^2_{V,K} = \min\left\{\frac{\epsilon}{|K|},1\right\}\|v\|^2 + \epsilon\|\nabla v\|^2 + \|\beta\cdot\nabla v\|^2 + \|\nabla\cdot\tau\|^2 + \min\left\{\frac{1}{\epsilon},\frac{1}{|K|}\right\}\|\tau\|^2.$$

This modified test norm avoids boundary layers in the locally computed optimal test functions, but for adaptive meshes, provides additional stability in areas of heavy refinement, where the best approximation error tends to be large and stronger robustness is most necessary. This leads to a test norm which produces easily approximable optimal test functions, but still provides *asymptotically* the strongest test norm and tightest robustness results in the areas of highest error.

3.4. Equivalence of energy norm with $\|\cdot\|_U$

The main theoretical result of this paper can now be given:

Lemma 1. *Under the mesh-dependent test norm*

$$\|(v,\tau)\|^2_{V,\Omega_h} = \|C_v v\|^2 + \epsilon\|\nabla v\|^2 + \|\beta\cdot\nabla v\|^2 + \|\nabla\cdot\tau\|^2 + \|C_\tau\tau\|^2,$$

where $C_v, C_\tau \in L^2(\Omega)$ are defined elementwise through

$$C_v|_K = \min\left\{\sqrt{\frac{\epsilon}{|K|}},1\right\}$$
$$C_\tau|_K = \min\left\{\frac{1}{\sqrt{\epsilon}},\frac{1}{\sqrt{|K|}}\right\}.$$

If β satisfies (19)*,* (20)*, and* (21)*, the DPG energy norm $\|\cdot\|_E$ satisfies the following equivalence relations*

$$\|u\|_{L^2} + \|\sigma\|_{L^2} + \epsilon\,\|\widehat{u}\| + \sqrt{\epsilon}\left\|\widehat{f}_n\right\| \lesssim \left\|\left(u,\sigma,\widehat{u},\widehat{f}_n\right)\right\|_E$$
$$\left\|\left(u,\sigma,\widehat{u},\widehat{f}_n\right)\right\|_E \lesssim \|u\|_{L^2} + \left\|\frac{1}{\epsilon C_\tau}\sigma\right\|_{L^2} + \frac{1}{\sqrt{\epsilon}}\left(\|\widehat{u}\| + \left\|\widehat{f}_n\right\|\right).$$

Proof. We begin by proving the bound from below. As a consequence of the duality of norms discussed in Section 1.3, we know that the norm $\|u\|_{U,1}$ is induced by a specific test norm $\|v\|_{V,U,1}$. To bound $\|\cdot\|_E$ robustly from above or below by a given norm $\|u\|_{U,2}$ on U now only requires the robust bound in the opposite direction of $\|v\|_{V,U,1}$ by $\|v\|_{V,U,2}$.

For f and g defined in (17) and (18),

$$f = \epsilon^{-1}\tau + \nabla v$$
$$g = \nabla\cdot\tau - \beta\cdot\nabla v,$$

we can characterize the test norm for

$$\left\|\left(u,\sigma,\widehat{u},\widehat{f}_n\right)\right\|^2_{U,1} = \|u\|^2 + \|\sigma\|^2 + \epsilon\|\widehat{u}\|^2 + \sqrt{\epsilon}\|\widehat{u}\|^2$$

through the equivalence relation

$$\|(v,\tau)\|_{V,U,1} \simeq \sup_{u,\sigma,\widehat{u},\widehat{f}_n}\frac{b\left(\left(u,\sigma,\widehat{u},\widehat{f}_n\right),(\tau,v)\right)}{\|u\| + \|\sigma\| + \epsilon\|\widehat{u}\| + \sqrt{\epsilon}\|\widehat{u}\|}$$
$$\simeq \|g\| + \|f\| + \frac{1}{\epsilon}\sup_{\widehat{u}\neq 0,\,\widehat{u}|_{\Gamma_+}=0}\frac{\langle[\![\tau\cdot n]\!],\widehat{u}\rangle}{\|\widehat{u}\|} + \frac{1}{\sqrt{\epsilon}}\sup_{\widehat{f}_n\neq 0,\,\widehat{f}_n|_{\Gamma_-}=0}\frac{\langle\widehat{f}_n,[\![v]\!]\rangle}{\|\widehat{f}\|},$$

which, by definition of the boundary norms, is

$$\| (v,\tau) \|_{V,U,1} \simeq \|g\| + \|f\| + \frac{1}{\epsilon}\| [\![\tau \cdot n]\!] \| + \frac{1}{\sqrt{\epsilon}}\| [\![v]\!] \|.$$

We wish to show the bound

$$\|(v,\tau)\|_{V,\Omega_h} \lesssim \|g\| + \|f\| + \frac{1}{\epsilon}\| [\![\tau \cdot n]\!] \| + \frac{1}{\sqrt{\epsilon}}\| [\![v]\!] \|.$$

By noting that both

$$\|C_v v_0\| \leq \|v_0\|,$$
$$\|C_\tau \tau_0\| \leq \frac{1}{\sqrt{\epsilon}} \|\tau_0\|,$$

we have that $\|(v,\tau)\|_{V,\Omega_h} \leq \|(v,\tau)\|_V$, so it suffices to prove the bound for the mesh-independent test norm

$$\| (v,\tau) \|_V^2 = \|v\|^2 + \epsilon\|\nabla v\|^2 + \|\beta \cdot \nabla v\|^2 + \|\nabla \cdot \tau\|^2 + \frac{1}{\epsilon}\|\tau\|^2.$$

We will bound $\| (v,\tau) \|_V$ for all (v,τ) by decomposing $(v,\tau) = (v_0,\tau_0) + (v_1,\tau_1) + (v_2,\tau_2)$ as described in Section 3.1.

By the triangle inequality, robustly bounding $\| (v,\tau) \|_V$ from above reduces to robustly bounding each component

$$\| (v_0,\tau_0) \|_V, \| (v_1,\tau_1) \|_V, \| (v_2,\tau_2) \|_V \lesssim \|g\| + \|f\| + \frac{1}{\epsilon}\| [\![\tau \cdot n]\!] \| + \frac{1}{\sqrt{\epsilon}}\| [\![v]\!] \|.$$

- **Bound on** $\| (v_0,\tau_0) \|_V$

  Lemma 4 gives control over $\sqrt{\epsilon}\|\nabla v_0\| + \frac{1}{\epsilon}\|\tau_0\|$ through

  $$\|\nabla v_0\| = \frac{1}{\epsilon}\|\tau_0\| \lesssim \frac{1}{\epsilon} \|[\![\tau_0 \cdot n]\!]\|_{\Gamma_h \setminus \Gamma_+} + \frac{1}{\sqrt{\epsilon}} \|[\![v_0]\!]\|_{\Gamma_h^0 \cup \Gamma_+} = \frac{1}{\epsilon} \|[\![\tau \cdot n]\!]\|_{\Gamma_h \setminus \Gamma_+} + \frac{1}{\sqrt{\epsilon}} \|[\![v]\!]\|_{\Gamma_h^0 \cup \Gamma_+}.$$

  Lemma 4.2 of [7] gives us the Poincare inequality for discontinuous functions

  $$\|v_0\| \lesssim \|\nabla v_0\| + \| [\![v]\!] \|.$$

  Since $g = 0$, $\|\nabla \cdot \tau_0\| = \|\beta \cdot \nabla v_0\| \lesssim \|\nabla v_0\|$, which we now have control over as well.

- **Bound on** $\| (v_1,\tau_1) \|_V$

  With $f = 0$, Lemma 2 provides the bound

  $$\|\beta \cdot \nabla v_1\| \lesssim \|g\|.$$

  Noting that $\nabla \cdot \tau_1 = g + \beta \cdot \nabla v_1$ gives $\|\nabla \cdot \tau_1\| \lesssim \|g\|$ as well. Lemma 3 gives

  $$\epsilon\|\nabla v_1\|^2 + \|v_1\|^2 \lesssim \|g\|^2,$$

  and noting that $\epsilon^{-1/2}\tau_1 = \epsilon^{1/2}\nabla v_1$ gives $\epsilon\|\nabla v_1\|^2 = \epsilon^{-1}\|\tau_1\|^2 \lesssim \|g\|^2$ as well.

- **Bound on** $\| (v_2,\tau_2) \|_V$

  Lemma 3 provides, for ϵ sufficiently small,

  $$\epsilon\|\nabla v_2\|^2 + \|v_2\|^2 \lesssim \epsilon\|f\|^2 \leq \|f\|^2.$$

  We have $\epsilon^{-1}\tau_2 = f - \nabla v_2$, so $\epsilon^{-1}\|\tau_2\| \lesssim \|f\| + \|\nabla v_2\|$. Lemma 3 implies $\|\nabla v_2\|^2 \lesssim \|f\|^2$, so for $\epsilon \leq 1$, we have $\epsilon^{-1/2}\|\tau_2\| \leq \epsilon^{-1}\|\tau_2\| \lesssim \|f\|$. The remaining terms can be bounded by noting that, with $g = 0$, $\|\nabla \cdot \tau_2\| = \|\beta \cdot \nabla v_2\| \lesssim \|\nabla v_2\| \lesssim \|f\|$.

We have shown the robust bound of the norm $\|\cdot\|_{U,1}$ on U by the energy norm; for a full equivalence statement, we require a bound from above on the energy norm by the norm $\|\cdot\|_{U,2}$ on U. By the duality of the energy and test norm, this is equivalent to bounding the test norm from below by the test norm induced by $\|\cdot\|_{U,2}$. For a norm on U of the form

$$\left\|\left(u,\sigma,\widehat{u},\widehat{f}_n\right)\right\|_{U,2}^2 = \|u\|^2 + \|C_\sigma\sigma\|^2 + \frac{1}{\epsilon}\left(\|\widehat{u}\|^2 + \|\widehat{f}_n\|^2\right),$$

the induced test norm is equivalent to

$$\begin{aligned}
\|\,(\tau,v)\,\|_{V,U,2} &\simeq \sup_{(u,\sigma,\widehat{u},\widehat{f}_n)\in U\setminus\{0\}} \frac{b\left(\left(u,\sigma,\widehat{u},\widehat{f}_n\right),(\tau,v)\right)}{\left\|\left(u,\sigma,\widehat{u},\widehat{f}_n\right)\right\|_E} \\
&\simeq \sup_{(u,\sigma,\widehat{u},\widehat{f}_n)\in U\setminus\{0\}} \frac{(u,\nabla\cdot\tau-\beta\cdot\nabla v)+\left(\sigma,\epsilon^{-1}\tau+\nabla v\right)-\langle[\![\tau_n]\!],\widehat{u}\rangle+\langle\widehat{f}_n,[\![v]\!]\rangle}{\|u\|+\left\|(\epsilon C_\tau)^{-1}\,\sigma\right\|+\frac{1}{\sqrt{\epsilon}}\left(\|\widehat{u}\|+\|\widehat{f}_n\|\right)} \\
&\simeq \|g\|+\|\epsilon C_\tau f\|+\sqrt{\epsilon}\left(\sup_{\widehat{u},\widehat{f}_n\neq 0}\frac{\langle[\![\tau_n]\!],\widehat{u}\rangle+\langle\widehat{f}_n,[\![v]\!]\rangle}{\|\widehat{u}\|+\|\widehat{f}_n\|}\right),
\end{aligned}$$

where f and g are

$$\begin{aligned}
f &= \frac{1}{\epsilon}\tau+\nabla v \\
g &= \nabla\cdot\tau-\beta\cdot\nabla v,
\end{aligned}$$

the loads of the adjoint problem defined in (17), (18).

Note that $\epsilon C_\tau \leq \sqrt{\epsilon}$. Then, by the triangle inequality, we have the bounds

$$\begin{aligned}
\|\epsilon C_\tau f\| &\leq C_\tau\|\tau\|+\epsilon C_\tau\|\nabla v\| \lesssim \|\,(\tau,v)\,\|_{V,\Omega_h} \\
\|g\| &\leq \|\nabla\cdot\tau\|+\|\beta\cdot\nabla v\| \lesssim \|\,(\tau,v)\,\|_{V,\Omega_h}
\end{aligned}$$

We estimate the supremum on the jumps of (τ,v) by following [10]; we begin by choosing $\eta\in H(\mathrm{div};\Omega)$, $w\in H^1(\Omega)$, such that $(\eta-\beta w)\cdot n|_{\Gamma_+}=0$ and $w|_{\Gamma_-\cup\Gamma_0}=0$, and integrating the boundary pairing by parts to get

$$\begin{aligned}
\langle[\![\tau\cdot n]\!],w\rangle+\langle[\![v]\!],(\eta-\beta w)\cdot n\rangle &= (\tau,\nabla w)+(\nabla\cdot\tau,w)+(\eta-\beta w,\nabla v)+(\nabla\cdot(\eta-\beta w),v) \\
&\lesssim \|C_\tau\tau\|\left\|\frac{1}{C_\tau}\nabla w\right\|+\|\nabla\cdot\tau\|\|w\| \\
&\quad+\sqrt{\epsilon}\|\nabla v\|\frac{1}{\sqrt{\epsilon}}\|\eta\|+\|\beta\cdot\nabla v\|\|w\| \\
&\quad+\|C_v v\|\left\|\frac{1}{C_v}\nabla\cdot\eta\right\|+\|C_v v\|\left\|\frac{1}{C_v}w\right\| \\
&\quad+\|C_v v\|\left\|\frac{1}{C_v}\nabla w\right\|,
\end{aligned}$$

where we have used that $\epsilon<1$, $\nabla\cdot\beta=O(1)$, and that $\|\beta\cdot\nabla w\|\lesssim\|\nabla w\|$.

Without loss of generality, assume the problem is scaled such that $\max_{K\in\Omega_h}|K|\leq 1$. Then, $\frac{1}{C_\tau^2}\leq\frac{1}{C_v^2}\leq\frac{1}{\epsilon}$, and an application of discrete Cauchy-Schwarz gives us

$$\begin{aligned}
\langle[\![\tau\cdot n]\!],w\rangle+\langle[\![v]\!],(\eta-\beta w)\cdot n\rangle &\lesssim \|\,(\tau,v)\,\|_{V,\Omega_h}\frac{1}{\sqrt{\epsilon}}\left(\|\eta\|_{H(\mathrm{div},\Omega)}+\|w\|_{H^1(\Omega)}\right), \\
&\lesssim \|\,(\tau,v)\,\|_{V,\Omega_h}\frac{1}{\sqrt{\epsilon}}\left(\|\eta-\beta w\|_{H(\mathrm{div},\Omega)}+\|w\|_{H^1(\Omega)}\right),
\end{aligned}$$

since $\|\eta\|_{H(\mathrm{div},\Omega)} = \|\eta - \beta w + \beta w\|_{H(\mathrm{div},\Omega)} \leq \|\eta - \beta w\|_{H(\mathrm{div},\Omega)} + \|\beta w\|_{H(\mathrm{div},\Omega)} \lesssim \|\eta - \beta w\|_{H(\mathrm{div},\Omega)} + \|w\|_{H^1(\Omega)}$. Dividing through and taking the supremum gives

$$\sup_{w,\eta\neq 0} \frac{\langle [\![\tau \cdot n]\!], w\rangle + \langle [\![v]\!], (\eta - \beta w)\cdot n\rangle}{\left(\|\eta - \beta w\|_{H(\mathrm{div},\Omega)} + \|w\|_{H^1(\Omega)}\right)} \lesssim \|\,(\tau, v)\,\|_{V,\Omega_h} \frac{1}{\sqrt{\epsilon}}.$$

To finish the proof, define $\rho \in H^{1/2}(\Gamma_h)$ and $\phi \in H^{-1/2}(\Gamma_h)$ such that $\rho = w|_{\Gamma_h}$ and $\phi = (\eta - \beta w)\cdot n|_{\Gamma_h}$, and note that, from [7], by the definition of the trace norms on $[\![\tau \cdot n]\!]$ and $[\![v]\!]$

$$\sup_{\rho,\phi\neq 0} \frac{\langle [\![\tau \cdot n]\!], \rho\rangle + \langle [\![v]\!], \phi\rangle}{\|\rho\|_{H^{1/2}(\Gamma_h)} + \|\phi\|_{H^{-1/2}(\Gamma_h)}} = \sup_{w,\eta\neq 0} \frac{\langle [\![\tau \cdot n]\!], w\rangle + \langle [\![v]\!], (\eta - \beta w)\cdot n\rangle}{\|w\|_{H^1(\Omega)} + \|\eta - \beta w\|_{H(\mathrm{div},\Omega)}}.$$

Together, the bounds on the jump terms and the bounds on $\|g\|$ and $\|f\|$ imply $\left\|\left(u, \sigma, \widehat{u}, \widehat{f}_n\right)\right\|_E \lesssim \left\|\left(u, \sigma, \widehat{u}, \widehat{f}_n\right)\right\|_{U,2}$. □

3.5. Comparison of boundary conditions

It is worth addressing the effect of boundary conditions on stability. Specifically, a test norm that provides stability for one set of boundary conditions may perform poorly for another set. Take, for example, the test norm defined in Section 3.4 and the convection-diffusion problem with Dirichlet boundary conditions.

The bilinear form for the case of Dirichlet boundary conditions is

$$b\left((u, \sigma, \widehat{u}, \widehat{\sigma}_n), (v, \tau)\right) = (u, \nabla\cdot\tau - \beta\cdot\nabla v) + \left(\sigma, \epsilon^{-1}\tau + \nabla v\right) + \langle \widehat{u}, [\![\tau\cdot n]\!]\rangle_{\Gamma_h^0} + \langle \widehat{f}_n, [\![v]\!]\rangle_{\Gamma_h}.$$

Notice that the boundary terms in the final bilinear form are different; hence, the adjoint problems associated with Section 3.2 will now carry different boundary conditions as well. Likewise, the stability properties proven previously will not hold under a different set of boundary conditions.

As it turns out, the robust bounds given in Section 3.4 hold in $\mathbb{R}^d$ for arbitrary d; however, we can show that for the case of Dirichlet boundary conditions, the same results do not hold, even in 1D. Consider now the 1D analogue of the estimate given by Lemma 2. In 1D, $\|\beta\cdot\nabla v_1\| \lesssim \|g\|$ reduces to the inequality

$$\|\beta v_1'\| \lesssim \|g\|, \quad g \in L^2(\Omega_h).$$

Without this inequality, we are unable to prove the robust bound on the L^2 error $\|u - u_h\|_{L^2} \lesssim \left\|(u, \sigma, \widehat{u}, \widehat{f}_n) - (u_h, \sigma_h, \widehat{u}_h, \widehat{f}_{n,h})\right\|_E$.

The adjoint problem corresponding to Lemma 2 in Section 3.2 is likewise reduced in 1D to the scalar equation

$$\epsilon v_1'' + \beta v_1' = -g \tag{22}$$

with $v_1 \in H_0^1((0,1))$. After multiplying this equation by $\beta v_1'$ and integrate by parts over Ω_h, we can apply Young's inequality to get

$$\frac{\epsilon}{2}\beta v_1'^2\Big|_0^1 + \|\beta v_1'\|_{L^2}^2 \leq \frac{1}{2}\|g\|^2 + \frac{1}{2}\|\beta v_1'\|^2,$$

implying that

$$\|\beta v_1'\|_{L^2}^2 \lesssim \|g\|^2 + \beta\epsilon v_1'(0)^2.$$

Let us restrict ourselves to the cases where v_1 is sufficiently smooth for $v'(0)$ to be well defined. Taking $g = 1$ (corresponding to a piecewise constant approximation) we can solve (22) exactly. The

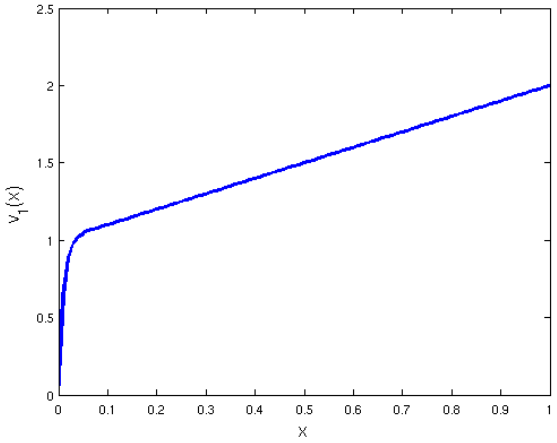


Figure 4: $v_1(x) = \frac{e^{-\frac{x}{\epsilon}}}{e^{\frac{1}{\epsilon}}-1}\left(e^{\frac{1}{\epsilon}}\left(e^{\frac{x}{\epsilon}}-1\right)+\left(e^{\frac{1}{\epsilon}}-1\right)e^{\frac{x}{\epsilon}}x\right)$, the solution to the adjoint equation for $f = 0$ and constant β and load g for $\epsilon = .01$.

solution v_1 is plotted in Figure 4, where we can see that $v_1(x)$ develops strong boundary layers of width ϵ near the inflow boundary $x = 0$. Consequently, $\frac{\epsilon}{2}v_1'(0)^2 \approx \epsilon^{-1}$. Thus, we cannot conclude $\|\beta v'\| \lesssim \|g\|$ when g is a constant,[7] and as a consequence cannot conclude that the robust error bound $\|u-u_h\|_{L^2} \lesssim \|(u,\sigma,\widehat{u},\widehat{f_n})-(u_h,\sigma_h,\widehat{u}_h,\widehat{f}_{n,h})\|_E$ holds for the solution u_h. More detailed 1D error bounds for Dirichlet boundary conditions are provided in [9], and indicate the same lack of robustness under the test norm derived in this paper.[8]

In higher dimensions, the adjoint problem is of the same form as the primal problem with the direction of convection reversed. However, the primal problem determines adjoint boundary conditions on Γ_- and Γ_+. Thus, wheareas for the primal problem, data is convected from the inflow to the outflow, in the adjoint problem, data is convected from the outflow to the inflow boundary instead.

We can intuitively explain the loss of robustness under our derived test norm by the presence of the Dirichlet boundary condition on v at the inflow boundary. Since the direction of convection is reversed in the adjoint equation, we can interpret the adjoint as representing the convection of a concentration v from the outflow to the inflow boundary. In the presence of a Dirichlet boundary condition at the inflow, v can develop strong boundary layers at the inflow. As a consequence, the quantities $\|\beta\cdot\nabla v\|$ and $\sqrt{\epsilon}\|\nabla v\|$ are no longer robustly bounded by $\|f\|$ and $\|g\|$, and we can no longer derive robust bounds on the error $\|u-u_h\|_{L^2}$ by the error in the energy norm.

Recall our strategy for analysis was to decompose of (v,τ) into continuous and discontinuous portions. Mathematically speaking, the use of Dirichlet boundary conditions on the primal problem introduces strong boundary layers into the solution v of the adjoint equation — in other words, boundary layers are introduced into the continuous portions of our decomposition of (v,τ).[9] The new inflow boundary condition on the primal problem relaxes the wall boundary condition induced

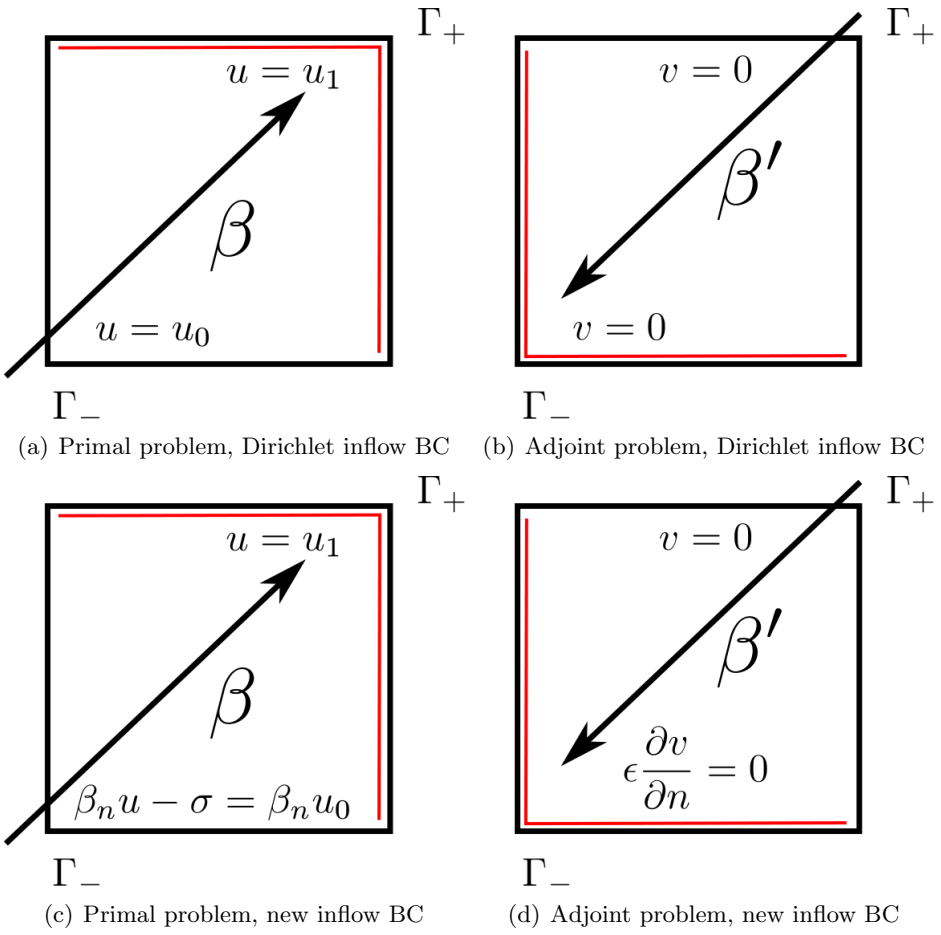
[7] Unlike the case of Dirichlet boundary conditions, the inflow condition on $\widehat{f_n} = u(0) - \epsilon u'(0)$ induces an adjoint boundary condition $\tau(0) = 0$, or equivalently $v'(0) = 0$, removing the non-robust term from the estimate.

[8] Demkowicz and Heuer proved in [10] that for Dirichlet boundary conditions, robustness as $\epsilon \to 0$ is achieved by the test norm

$$\|(\tau,v)\|_{V,w}^2 = \|v\| + \epsilon\|\nabla v\| + \|\beta\cdot\nabla v\|_{w+\epsilon} + \|\nabla\cdot\tau\|_{w+\epsilon} + \frac{1}{\epsilon}\|\tau\|_{w+\epsilon}$$

where $\|\cdot\|_{w+\epsilon}$ is a weighted L^2 norm, where the weight $w \in (0,1)$ is required to vanish on Γ_- and satisfy $\nabla w = O(1)$. The need for this weight is necessary to account for the loss of robustness at the inflow.

[9] The boundary conditions do not introduce boundary layers into the actual computed test functions. However, an interesting phenomenon observed is that, for small ϵ, a lack of robustness can manifest itself during numerical experiments as additional refinements near the inflow boundary, precisely where the continuous parts of the decomposition of (v,τ) develop boundary layers.



(a) Primal problem, Dirichlet inflow BC (b) Adjoint problem, Dirichlet inflow BC

(c) Primal problem, new inflow BC (d) Adjoint problem, new inflow BC

Figure 5: Comparison of primal and adjoint problems under both the standard Dirichlet and the new inflow boundary condition. The outflow boundary for each problem is denoted in red. For the standard Dirichlet inflow condition, the solution to the adjoint problem can develop strong boundary layers at the outflow of the adjoint problem. Notice, under the new inflow conditions, the relaxation of a wall-stop boundary condition with a zero-stress condition at the outflow boundary of the adjoint problem.

on the adjoint/dual problem with a boundary condition that does not generate boundary layers, resulting in stronger stability estimates for the adjoint, and a better result for the primal problem.

4. Numerical experiments

In each numerical experiment, we vary $\epsilon = .01, .001, .0001$ in order to demonstrate robustness over a range of ϵ. This is intended to mirror the experience with roundoff effects in numerical experiments [10]; for "worst-case" linear solvers, such as LU decomposition without pivoting, the effect of roundoff error becomes evident in the solving of optimal test functions for $\epsilon \leq O(1e-5)$. The roundoff itself comes from the conditioning of the Gram matrix under certain test norms; for example, if the weighted $H(\mathrm{div};\Omega) \times H^1(\Omega)$ norm is used for the test norm $\| (\tau, v) \|_V$ (as was done in [8]), for an element of size h, $\|v\|^2_{L^2} = O(h)$, while $\|\nabla v\|^2_{L^2} = O(h^{-1})$. As $h \to 0$, the seminorm portion of the test norm dominates the Gram matrix, leading to a near-singular and ill-conditioned system.

The effect of roundoff error is often characterized by an increase in the energy error, which (assuming negligible error in the approximation of test functions) is proven to decrease for any series of refined meshes. These roundoff effects are dependent primarily on the mesh, appearing

when trying to fully resolve very thin boundary layers by introducing elements of size ϵ through adaptivity. The effects of roundoff error were successfully treated in [9] by dynamically rescaling the test norms based on element size, a practical remedy not covered yet by the present analysis.

4.1. Eriksson-Johnson model problem

To confirm our theoretical results, we adopt a modification of a problem first proposed by Eriksson and Johnson in [12]. For the choice of $\Omega = (0,1)^2$, $f = 0$, and $\beta = (1,0)^T$, the convection diffusion equation reduces to

$$\frac{\partial u}{\partial x} - \epsilon \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) = 0,$$

which has an exact solution by separation of variables, allowing us to analyze convergence of DPG for a wide range of ϵ. For boundary conditions, we impose $u = 0$ on Γ_+ and $\beta_n u - \sigma_n$ on Γ_-, which reduces to

$$\begin{aligned} u - \sigma_x &= u_0 - \sigma_{x,0}, \quad x = 0, \\ \sigma_y &= 0, \quad y = 0, 1, \\ u &= 0, \quad x = 1. \end{aligned}$$

In this case, our exact solution is the series

$$u(x,y) = C_0 + \sum_{n=1}^{\infty} C_n \frac{\exp(r_2(x-1) - \exp(r_1(x-1)))}{r_1 \exp(-r2) - r_2 \exp(-r1)} \cos(n\pi y),$$

where

$$\begin{aligned} r_{1,2} &= \frac{1 \pm \sqrt{1 + 4\epsilon\lambda_n}}{2\epsilon}, \\ \lambda_n &= n^2\pi^2\epsilon. \end{aligned}$$

The constants C_n depend on a given inflow condition u_0 at $x = 0$ via the formula

$$C_n = \int_0^1 u_0(y) \cos(n\pi y).$$

All computations have been done using the adaptive DPG code Camellia, built on the Sandia toolbox Trilinos [19].

4.1.1. Solution with $C_1 = 1, C_{n\neq 1} = 0$

We begin with the solution taken to be the first non-constant term of the above series. We set the inflow boundary condition to be exactly the value of $u - \sigma_x$ corresponding to the exact solution.

In each case, we begin with a square 4 by 4 mesh of quadrilateral elements with order $p = 3$. We choose $\Delta p = 5$, though we note that the behavior of DPG is nearly identical for any $\Delta p \leq 3$, and qualitatively the same for $\Delta p = 2$. h-refinements are executed using a greedy refinement algorithm, where element energy error e_K^2 is computed for all elements K, and elements such that $e_K^2 \leq \alpha \max_K e_K^2$ are refined. We make the arbitrary choice of taking $\alpha = .2$ for each of these experiments.

We are especially interested in the ratio of energy error and total L^2 error in both σ and u, which we denote as $\|u - u_h\|_{L^2}$. The bounds on $\|\cdot\|_E$ presented in Section 3.4 imply that, using the above test norm, $\|u - u_h\|_{L^2} / \|u - u_h\|_E \leq C$ independent of ϵ. Figure 8, which plots the ratio of L^2 to energy error, seems to imply that (at least for this model problem) $C = O(1)$. Additionally,

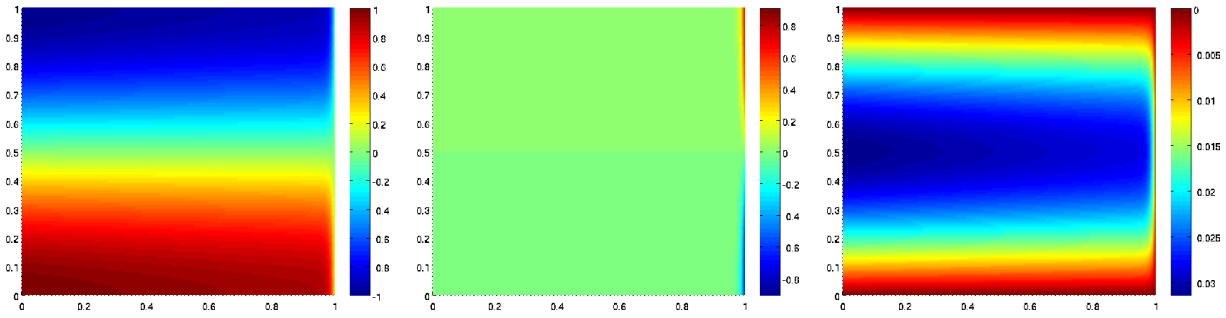


Figure 6: Solution for u, σ_x, and σ_y for $\epsilon = .01$, $C_1 = 1$, $C_n = 0$, $n \neq 1$

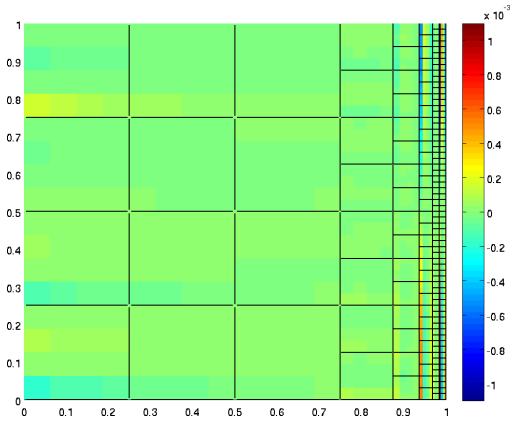


Figure 7: Adapted mesh and pointwise error for $\epsilon = .01$

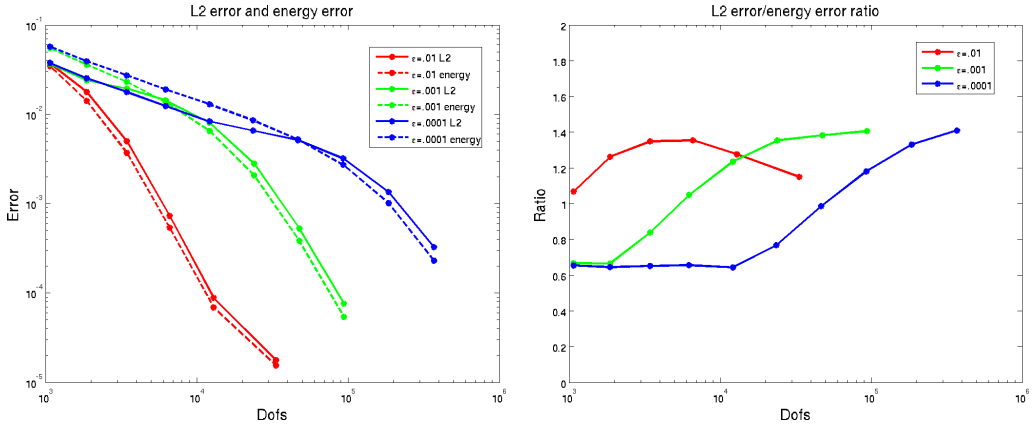


Figure 8: L^2 and energy errors, and their ratio for $\epsilon = .01$, $\epsilon = .001$, $\epsilon = .0001$

while we do not have a robust lower bound ($\|u - u_h\|_{L^2}/\|u - u_h\|_E$ can approach 0 as $\epsilon \to 0$), our numerical results appear to indicate the existence of an ϵ-independent lower bound.

The effect of a mesh dependent scalings on the $\|v\|^2$ and $\|\tau\|^2$ terms in the test norm can be seen in the ratios of L^2 to energy error; as the mesh is refined, the constants in front of the L^2 terms for v and τ converge to stationary values (providing the full robustness implied by our adjoint energy estimates), and the ratio of L^2 to energy error transitions from a smaller to a larger value. The transition point happens later for smaller ϵ, which we expect, since the transition of the ratio corresponds to the introduction of elements whose size is of order ϵ through mesh refinement.

We examined how small ϵ needed to be in order to encounter roundoff effects as well. In [10], the smallest resolvable ϵ using only double precision arithmetic was $1e-4$. The solution of optimal test functions is now done using both pivoting and equilibration, improving conditioning. Roundoff

effects still appear, but at smaller values of ϵ.

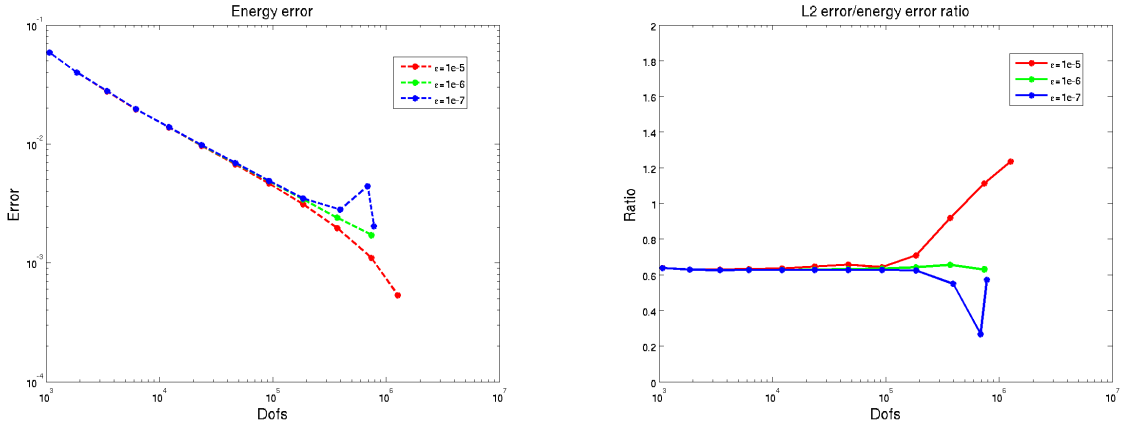


Figure 9: Energy error and L^2/energy error ratio for $\epsilon = 1e-5$, $\epsilon = 1e-6$, $\epsilon = 1e-7$. Non-monotonic behavior of the energy error indicates conditioning issues and roundoff effects.

Without anisotropic refinements, it still becomes computationally difficult to fully resolve the solution for ϵ smaller than $1e-5$. Regardless, for all ranges of ϵ, DPG does not lose robustness, as indicated by the rates and ratio between L^2 and energy error in Figure 9 remaining bounded from both above and below. For $\epsilon = 1e-5$, we observe that the ratio of L^2 error increases, corresponding to the scaling of the test norm with mesh size (the transition in test norm occurs after 8 refinements, which, for an initial 4×4 mesh, implies a minimum element size of about $1.5e-05$. At this point, rescaled test norm allows us to take advantage of the full magnitude of the L^2 term for $\|v\|$ and $\|\tau\|$ implied by our adjoint estimates). By analogy, for smaller $\epsilon = 1e-6, 1e-7$, the transition period should begin near the 10th and 11th refinement iterations; however, we do not observe such behavior, possibly due to roundoff effects. For $\epsilon = 1e-6$, the ratio simply remains constant, but for $\epsilon = 1e-7$, we observe definite roundoff effects, as the energy error increases at the 11th refinement. Since DPG is optimal in the energy norm for a mesh-independent test norm[10], we expect monotonic decrease of the energy error with mesh refinement. Non-monotonic behavior indicates either approximation or roundoff error, and as we observed no qualitative difference between using $\Delta p = 5$ and $\Delta p = 6$ for these experiments, we expect that the approximation error is negligible and conclude roundoff effects are at play when these phenomena are observed.

It is worth noting that for $\epsilon \leq 1e-5$, we do not perform enough refinements to completely resolve the boundary layer, so $|K| \geq \epsilon$ for all $K \in \Omega_h$. Thus, any roundoff effects observed are not due to the conditioning issues associated with the differing scales of the $\|v\|_{L^2(K)}$ and $\|\nabla v\|_{L^2(K)}$ terms discussed previously.

4.1.2. Neglecting σ_n

In practice, we will not have prior knowledge of σ_n at the inflow, and will have to set $\beta_n u - \sigma_n = u_0$, ignoring the viscous contribution to the boundary condition. The hope is that for small ϵ, this omission will be negligible. Figure 10 indicates that, between $\epsilon = .005$ and $\epsilon = .001$, the omission of σ_n in the boundary condition becomes negligible, and both our error rates and ratios of L^2 to energy error become identical to the case where σ_n is explicitly accounted for in the inflow condition. For large $\epsilon = .01$, the L^2 error stagnates around $1e-3$, or about 7% relative error.

[10] While the test norm changes with the mesh, it increases monotonically. A strictly stronger test norm implies $\frac{b(u,v)}{\|v\|_1} \geq \frac{b(u,v)}{\|v\|_2}$ for any $\|v\|_1 \leq \|v\|_2$

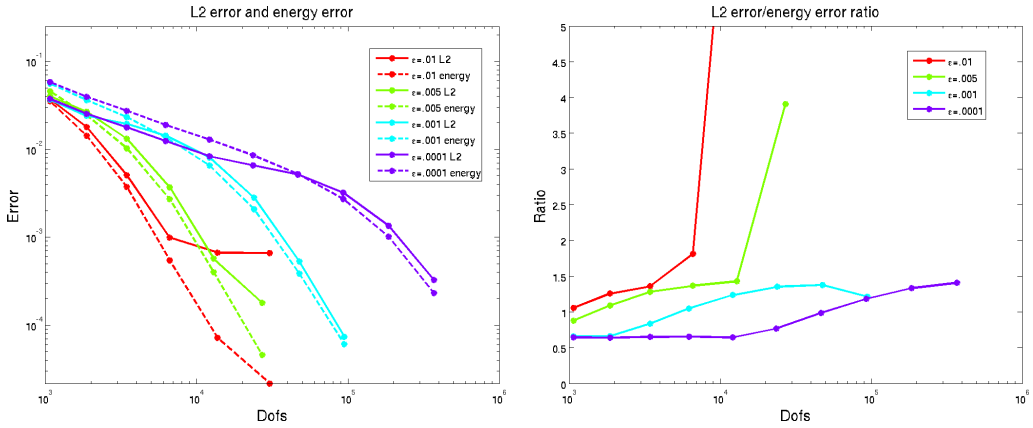


Figure 10: L^2 and energy errors and their ratio when neglecting σ_n at the inflow.

4.1.3. Discontinuous inflow data

We note also that an additional advantage of selecting this new boundary condition is a relaxation of regularity requirements: as $\widehat{f}_n \in H^{-1/2}(\Gamma_h)$, strictly discontinuous inflow boundary conditions are no longer "variational crimes". We consider the discontinuous inflow condition

$$u_0(y) = \begin{cases} (y-1)^2, & y > .5 \\ -y^2, & y \leq .5 \end{cases}$$

as an example of a more difficult test case.

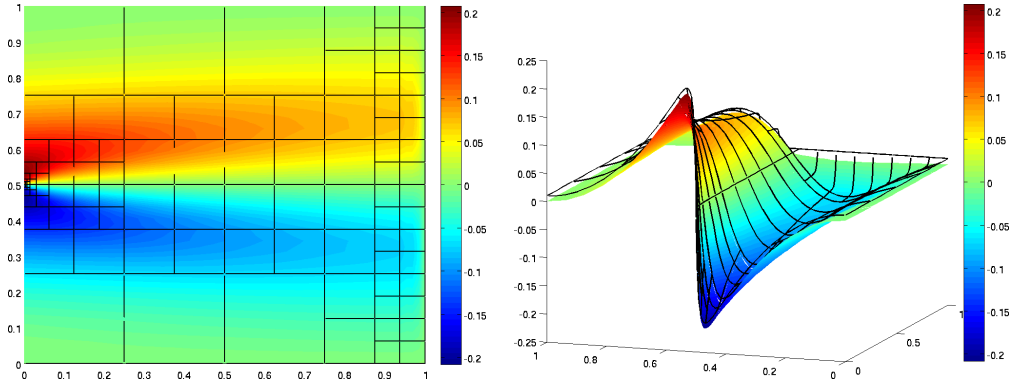


Figure 11: Solution variables u and $\widehat{u}$ with discontinuous inflow data u_0 for $\epsilon = .01$.

Figure 11 shows the solution u and overlaid trace variable $\widehat{u}$, which both demonstrate the regularizing effect of viscosity on the discontinuous boundary condition at $x = 0$. However, we do not have a closed-form solution with which to compare results for a stricly discontinuous u_0. In order to analyze convergence, we approximate u_0 with 20 terms of a Fourier series, giving a near-discontinuity for u_0.

The ratios of L^2 to energy error are now less predictable than for the previous example, in part due to the difficulty in approximating highly oscillatory boundary conditions. The numerical experiments were originally performed by applying boundary conditions via interpolation; the result was that the highly oscillatory inflow boundary condition was not sampled enough to be properly resolved, causing the solution to converge to a solution different than the exact solution. The experiments were repeated using the penalty method to enforce inflow conditions; however, we note that the proper way to do so is to use an L^2 projection at the boundary. Even when using the penalty method, however, the ratios still remain bounded and close to 1 for ϵ varying over two orders of magnitude, as predicted by theory.

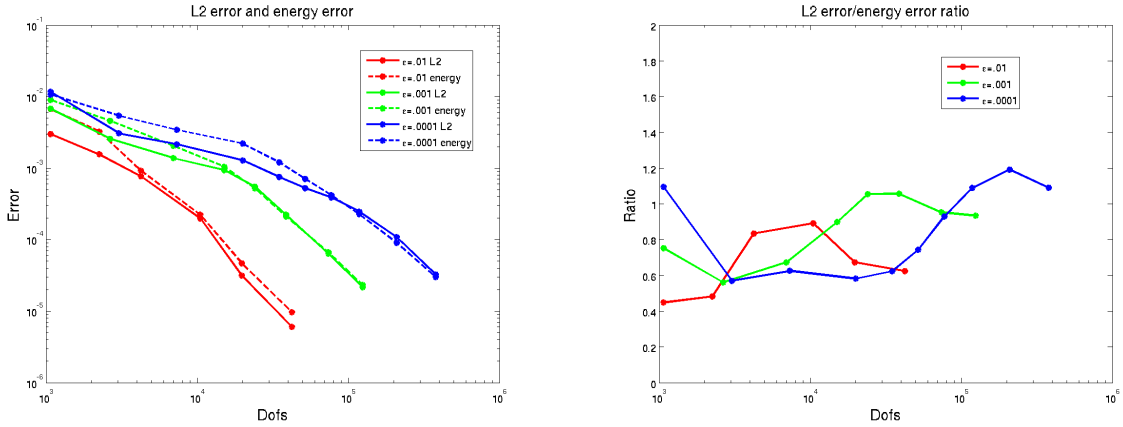


Figure 12: L^2 and energy errors, and their ratio for $\epsilon = .01$, $\epsilon = .001$, $\epsilon = .0001$, with discontinuous u_0 approximated by a Fourier expansion.

5. Conclusions

We have presented in this paper the analysis of a non-canonical test norm and its corresponding DPG energy norm for the convection-diffusion equation in the small-diffusion limit. Additionally, we have introduced a non-standard inflow boundary condition, and have explored the difference between between this and the standard Dirichlet inflow boundary condition.

Numerical results are presented in order to verify the results derived in this paper. However, at least for our model problem, numerical experiments appear to demonstrate results that are stronger than our proofs indicate, delivering solutions for u and σ that are extremely close to their best L^2 projections.

The theory presented in this paper have been successfully extrapolated to nonlinear singular perturbation problems and systems of equations, and has been applied in context of the Burgers and Navier-Stokes equations. These results will be presented in an upcoming paper.

6. Acknowledgements

The authors would like to acknowledge the help of Nathan Roberts with software development and Dr Robert Moser for fruitful discussions and perspective. Chan and Demkowicz were supported by the Department of Energy [National Nuclear Security Administration] under Award Number [DE-FC52-08NA28615]. Heuer was supported through the project "Fondecyt-Chile 1110324" and by the "J Tinsley Oden Faculty Research Fellowship" of the Institute of Computational Engineering and Sciences (ICES).

References

[1] J. Barrett and K. Morton. Optimal Petrov—Galerkin methods through approximate symmetrization. *IMA Journal of Numerical Analysis*, 1(4):439–468, 1981.

[2] A. Brooks and T. Hughes. Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Comp. Meth. Appl. Mech. Engr*, 32:199–259, 1982.

[3] T. Bui-Thanh, L. Demkowicz, and O. Ghattas. A unified discontinuous Petrov-Galerkin method and its analysis for Friedrichs' systems. *Submitted to SIAM J. Numer. Anal.*, 2011. Also ICES report ICES-11-34, November 2011.

[4] T. Bui-Thanh, Leszek Demkowicz, and Omar Ghattas. Constructively well-posed approximation method with unity inf-sup and continuity constants for partial differential equations. *Mathematics of Computation*, 2011. To appear.

[5] L. Demkowicz. *Computing With hp-adaptive Finite Elements: One and two dimensional elliptic and Maxwell problems*. Chapman & Hall/CRC Applied Mathematics and Nonlinear Science Series. Chapman & Hall/CRC, 2006.

[6] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. Part I: The transport equation. *Comput. Methods Appl. Mech. Engrg.*, 2009. accepted, see also ICES Report 2009-12.

[7] L. Demkowicz and J. Gopalakrishnan. Analysis of the DPG method for the Poisson equation. *SIAM J. Numer. Anal.*, 49(5):1788–1809, September 2011.

[8] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. ii. Optimal test functions. *Num. Meth. for Partial Diff. Eq*, 27:70–105, 2011.

[9] L. Demkowicz, J. Gopalakrishnan, and A. Niemi. A class of discontinuous Petrov-Galerkin methods. III. Adaptivity. Technical Report 10-01, ICES, 2010.

[10] L. Demkowicz and N. Heuer. Robust DPG method for convection-dominated diffusion problems. Technical Report 11-33, ICES, 2011.

[11] L. Demkowicz and J.T Oden. An adaptive characteristic Petrov-Galerkin finite element method for convection-dominated linear and nonlinear parabolic problems in one space variable. *Journal of Computational Physics*, 67(1):188 – 213, 1986.

[12] K. Eriksson and C. Johnson. Adaptive streamline diffusion finite element methods for stationary convection-diffusion problems. *Mathematics of Computation*, 60(201):pp. 167–188, 1993.

[13] J. Gopalakrishnan and W. Qiu. An analysis of the practical DPG method. Technical report, IMA, 2011. submitted.

[14] J. Hesthaven. A stable penalty method for the compressible Navier-Stokes equations. iii. multi dimensional domain decomposition schemes. *SIAM J. Sci. Comput*, 17:579–612, 1996.

[15] T. Hughes, L. Franca, and G. Hulbert. A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advective-diffusive equations. *Comp. Meth. Appl. Mech. Engr*, 73:173–189, 1989.

[16] N. Roberts and T. Bui-Thanh and L. Demkowicz. The DPG method for the Stokes problem. In preparation, June 2012.

[17] A. Niemi, N. Collier, and V. Calo. Discontinuous Petrov-Galerkin method based on the optimal test space norm for one-dimensional transport problems. *Procedia CS*, 4:1862–1869, 2011.

[18] Antti H. Niemi, Jamie A. Bramwell, and Leszek F. Demkowicz. Discontinuous Petrov–Galerkin method with optimal test functions for thin-body problems in solid mechanics. *Computer Methods in Applied Mechanics and Engineering*, 200(9–12):1291 – 1300, 2011.

[19] N. Roberts, D. Ridzal, P. Bochev, and L. Demkowicz. A Toolbox for a Class of Discontinuous Petrov-Galerkin Methods Using Trilinos. Technical Report SAND2011-6678, Sandia National Laboratories, 2011.

[20] H. Roos, M. Stynes, and L. Tobiska. *Robust numerical methods for singularly perturbed differential equations: convection-diffusion-reaction and flow problems*. Springer series in computational mathematics. Springer, 2008.

[21] J. Zitelli, I. Muga, L. Demkowicz, J. Gopalakrishnan, D. Pardo, and V.M. Calo. A class of discontinuous Petrov–Galerkin methods. Part IV: The optimal test norm and time-harmonic wave propagation in 1D. *Journal of Computational Physics*, 230(7):2406 – 2432, 2011.

A. Proof of lemmas/stability of the adjoint problem

We present now the proofs of the three lemmas used in this paper to show the equivalence of the DPG energy norm to norms on U. We reduce the adjoint problem to the scalar second order equation

$$-\epsilon\Delta v - \beta\cdot\nabla v = g - \epsilon\nabla\cdot f \tag{23}$$

with boundary conditions

$$-\epsilon\nabla v\cdot n = f\cdot n, \quad x\in\Gamma_- \tag{24}$$

$$v = 0, \quad x\in\Gamma_+ \tag{25}$$

and treat the cases $f=0$, $g=0$ separately. The above boundary conditions are the reduced form of boundary conditions (15) and (16) corresponding to $\tau\cdot n|_{\Gamma_-}=0$ and $v|_{\Gamma_+}=0$. Additionally, the $\nabla\cdot$ operator is understood now in the weak sense, as the dual operator of $-\nabla : H^1_0(\Omega)\to L^2(\Omega)$, such that $\nabla\cdot f\in\left(H^1_0(\Omega)\right)'$.

The normal trace of $f\cdot n$ is treated using a density argument — for $f\in C^\infty(\Omega)$, we derive inequalities that are independent of $f\cdot n$ and $\nabla\cdot f$. We extend these inequalities to $f\in L^2(\Omega)$ by taking f to be the limit of smooth functions.

Lemma 2. *Assume v satisfies* (23)*, with boundary conditions* (15) *and* (16)*, and β satisfies* (19) *and* (20)*. If $\nabla\cdot f=0$ and ϵ is sufficiently small, then*

$$\|\beta\cdot\nabla v\|\lesssim\|g\|.$$

Proof. Define $v_\beta=\beta\cdot\nabla v$. Multiplying the adjoint equation (23) by v_β and integrating over Ω gives

$$\|v_\beta\|^2 = -\int_\Omega g v_\beta - \epsilon\int_\Omega \Delta v v_\beta.$$

Note that

$$-\int_\Omega \beta\cdot\nabla v\Delta v = -\int_\Omega \beta\cdot\nabla v\nabla\cdot\nabla v.$$

Integrating this by parts, we get

$$-\int_\Omega \beta\cdot\nabla v\nabla\cdot\nabla v = \int_\Omega \nabla(\beta\cdot\nabla v)\cdot\nabla v - \int_\Gamma n\cdot\nabla v\beta\cdot\nabla v.$$

Since $\nabla(\beta\cdot\nabla v)=\nabla\beta\cdot\nabla v+\beta\cdot\nabla\nabla v$, where $\nabla\beta$ and $\nabla\nabla v$ are understood to be tensors,

$$\int_\Omega \nabla(\beta\cdot\nabla v)\cdot\nabla v = \int_\Omega(\nabla\beta\cdot\nabla v)\cdot\nabla v + \int_\Omega \beta\cdot\nabla\nabla v\cdot\nabla v$$

If we integrate by parts again and use that $\nabla v\cdot\nabla\nabla v=\nabla\frac{1}{2}(\nabla v\cdot\nabla v)$, we get

$$\begin{aligned}-\int_\Omega \Delta v v_\beta &= -\int_\Gamma n\cdot\nabla v\beta\cdot\nabla v + \frac{1}{2}\int_\Gamma \beta_n(\nabla v\cdot\nabla v) - \frac{1}{2}\int_\Omega \nabla\cdot\beta(\nabla v\cdot\nabla v) + \int_\Omega(\nabla\beta\cdot\nabla v)\cdot\nabla v\\ &= -\int_\Gamma n\cdot\nabla v\beta\cdot\nabla v + \frac{1}{2}\int_\Gamma\beta_n(\nabla v\cdot\nabla v) + \int_\Omega \nabla v\left(\nabla\beta - \frac{1}{2}\nabla\cdot\beta I\right)\nabla v\end{aligned}$$

Finally, substituting this into our adjoint equation multiplied by v_β, we get

$$\|v_\beta\|^2 = -\int_\Omega g\beta\cdot\nabla v + \epsilon\int_\Gamma\left(-n\cdot\nabla v\beta + \frac{1}{2}\beta_n\nabla v\right)\cdot\nabla v + \epsilon\int_\Omega \nabla v\left(\nabla\beta - \frac{1}{2}\nabla\cdot\beta I\right)\nabla v$$

The last term can be bounded by our assumption on $\|\nabla\beta - \frac{1}{2}\nabla\cdot\beta I\|^2 \le C$:

$$\epsilon\int_\Omega \nabla v\left(\nabla\beta - \frac{1}{2}\nabla\cdot\beta I\right)\nabla v \le C\frac{\epsilon}{2}\|\nabla v\|^2.$$

For the boundary terms, on Γ_-, $\nabla v\cdot n = 0$, reducing the integrand over the boundary to $\beta_n|\nabla v|^2 \le 0$. On Γ_+, $v = 0$ implies $\nabla v\cdot\tau = 0$, where τ is any tangential direction. An orthogonal decomposition in the normal and tangential directions yields $\nabla v = (\nabla v\cdot n)n$, reducing the above to

$$\epsilon\int_\Gamma -\frac{1}{2}|\beta_n|(\nabla v\cdot n)^2 \le 0.$$

Applying these inequalities to our expression for $\|v_\beta\|^2$ leaves us with the estimate

$$\|v_\beta\|^2 \le -\int_\Omega g\beta\cdot\nabla v + C\frac{\epsilon}{2}\|\nabla v\|^2.$$

Since $C = O(1)$, an application of Young's inequality and Lemma 3 complete the estimate. □

Lemma 3. *Assume β satisfies* (19)*. Then, for v satisfying equation* (23) *with boundary conditions* (15) *and* (16) *and sufficiently small ϵ,*

$$\epsilon\|\nabla v\|^2 + \|v\|^2 \lesssim \|g\|^2 + \epsilon\|f\|^2$$

Proof. Since $\nabla\times\beta = 0$, and Ω is simply connected, there exists a scalar potential ψ, $\nabla\psi = \beta$ by properties of the exact sequence. The potential is non-unique up to a constant, and we choose the constant such that $e^\psi = O(1)$. Take the transformed function $w = e^\psi v$; following (2.26) in [10], we substitute w into the the left hand side of equation (23), arriving at the relation

$$-\epsilon\Delta w - (1-2\epsilon)\beta\cdot\nabla w + \left((1-\epsilon)|\beta|^2 + \epsilon\nabla\cdot\beta\right)w = e^\psi(g - \epsilon\nabla\cdot f)$$

Multiplying by w and integrating over Ω gives

$$-\epsilon\int_\Omega \Delta w w - (1-2\epsilon)\int_\Omega \beta\cdot\nabla w w + \int_\Omega\left((1-\epsilon)|\beta|^2 + \epsilon\nabla\cdot\beta\right)w^2 = \int_\Omega e^\psi(g-\epsilon\nabla\cdot f)w$$

Integrating by parts gives

$$-\epsilon\int_\Omega \Delta w w - (1-2\epsilon)\int_\Omega \beta\cdot\nabla w w = \epsilon\left(\int_\Omega|\nabla w|^2 - \int_\Gamma w\nabla w\cdot n\right) + \frac{(1-2\epsilon)}{2}\left(\int_\Omega\nabla\cdot\beta w^2 - \int_\Gamma\beta_n w^2\right)$$

Note that $w = 0$ on Γ_+ reduces the boundary integrals over Γ to just the inflow Γ_-. Furthermore, we have $\nabla w = e^\psi(\nabla v + \beta v)$. Applying the above and boundary conditions on Γ_-, the first boundary integral becomes

$$\int_{\Gamma_-} w\nabla w\cdot n = \int_{\Gamma_-} we^\psi(\nabla v + \beta v)\cdot n = \int_{\Gamma_-} we^\psi(f\cdot n + \beta_n v)$$

Noting $\int_{\Gamma_-}\beta_n w^2 \le 0$ through $\beta_n < 0$ on the inflow gives

$$\epsilon\int_\Omega|\nabla w|^2 + \int_\Omega\left((1-\epsilon)|\beta|^2 + \frac{1}{2}\nabla\cdot\beta\right)w^2 - \epsilon\int_{\Gamma_-} we^\psi f\cdot n \le \int_\Omega e^\psi(g-\epsilon\nabla\cdot f)w$$

assuming ϵ is sufficiently small. Our assumptions on β imply $\left((1-\epsilon)|\beta|^2 + \frac{1}{2}\nabla\cdot\beta\right) \lesssim 1$ and $e^{\psi} = O(1)$. We can then bound from below:

$$\epsilon\|\nabla w\|^2 + \|w\|^2 - \epsilon\int_{\Gamma_-} we^{\psi} f\cdot n \lesssim \epsilon\int_{\Omega}|\nabla w|^2 + \int_{\Omega}\left((1-\epsilon)|\beta|^2 + \frac{1}{2}\nabla\cdot\beta\right)w^2 - \epsilon\int_{\Gamma_-} we^{\psi} f\cdot n$$

Interpreting $\nabla\cdot f$ as a functional, the right hand gives

$$\int_{\Omega} e^{\psi}(g - \epsilon\nabla\cdot f)w = \int_{\Omega} e^{\psi} g + \int_{\Omega}\epsilon f\cdot\nabla(e^{\psi}w) - \int_{\Gamma}\epsilon f\cdot n e^{\psi} w$$

The boundary integral on Γ reduces to Γ_-, which is then nullified by the same term on the left hand side, leaving us with

$$\epsilon\|\nabla w\|^2 + \|w\|^2 \lesssim \int_{\Omega} e^{\psi} g + \int_{\Omega}\epsilon f\cdot\nabla(e^{\psi}w) = \int_{\Omega} e^{\psi} g + \int_{\Omega}\epsilon f\cdot(\beta w + \nabla w)$$

From here, the proof is identical to the final lines of the proof of Lemma 1 in [10]; an application of Young's inequality (with δ) to the right hand side and bounds on $\|v\|, \|\nabla v\|$ by $\|w\|, \|\nabla w\|$ complete the estimate. □

Lemma 4. *Let β satisfy conditions* (19) *and* (21)*, and let $v \in H^1(\Omega_h)$, $\tau \in H(\mathrm{div}, \Omega_h)$ satisfy equations* (17) *and* (18) *with $f = g = 0$. Then*

$$\|\nabla v\| = \frac{1}{\epsilon}\|\tau\| \lesssim \frac{1}{\epsilon}\|\,[\![\tau\cdot n]\!]\,\|_{\Gamma_h\setminus\Gamma_+} + \frac{1}{\sqrt{\epsilon}}\|\,[\![v]\!]\,\|_{\Gamma_h^0\cup\Gamma_+}$$

Proof. We begin by choosing ψ as the unique solution to the following problem

$$\begin{aligned} -\epsilon\Delta\psi + \nabla\cdot(\beta\psi) &= -\nabla\cdot\tau \\ \epsilon\nabla\psi\cdot n - \beta_n\psi - \tau\cdot n = 0, &\quad x\in\Gamma_- \\ \psi = 0, &\quad x\in\Gamma_+. \end{aligned}$$

Since $\nabla\cdot\beta = 0$, we can conclude that the bilinear form is coercive and the problem is well posed [10]. The well-posedness of the above problem directly implies that $\nabla\cdot(\tau - (\epsilon\nabla\psi - \beta\psi)) = 0$ in a distributional sense, and thus there exists a $z \in H(\mathrm{curl}, \Omega)$ such that

$$\tau = (\epsilon\nabla\psi - \beta\psi) + \nabla\times z$$

Since $\nabla\cdot\beta = 0$, we satisfy condition (19). Noting that the sign on β is opposite now of the sign on $\epsilon\Delta\psi$, the problem for ψ matches the adjoint problem for $f = \frac{1}{\epsilon}\tau$. Given the boundary conditions on ψ, we can use a trivial modification of the proof of Lemma 3 to bound

$$\epsilon\|\nabla\psi\|_{L^2}^2 + \|\psi\|_{L^2}^2 \lesssim \frac{1}{\epsilon}\|\tau\|_{L^2}^2.$$

By the above bound and the triangle inequality,

$$\|\nabla\times z\|_{L^2} \le \epsilon\|\nabla\psi\|_{L^2} + \|\beta\psi\|_{L^2} + \|\tau\|_{L^2} \lesssim \frac{1}{\sqrt{\epsilon}}\|\tau\|_{L^2}.$$

On the other hand, using the decomposition and boundary conditions directly, we can integrate by parts over Ω_h to arrive at

$$\begin{aligned} \|\tau\|_{L^2}^2 &= (\tau, \epsilon\nabla\psi - \beta\psi + \nabla\times z)_{\Omega_h} = (\tau, \epsilon\nabla\psi) - (\tau, \beta\psi) + (\tau, \nabla\times z) \\ &= (\tau, \epsilon\nabla\psi) + \epsilon(\nabla v, \beta\psi) - \epsilon(\nabla v, \nabla\times z) \\ &= \epsilon\langle[\tau\cdot n], \psi\rangle - \epsilon\langle n\cdot\nabla\times z, [\![v]\!]\rangle - \epsilon(\nabla\cdot\tau, \psi) + \epsilon(\nabla\cdot(\beta v), \psi). \end{aligned}$$

Note that $\nabla \cdot (\beta v) - \nabla \cdot \tau = 0$ removes the contribution of the pairings on the domain and leaves us with only boundary pairings. By definition of the boundary norms on $[\![\tau \cdot n]\!]$ and $[\![v]\!]$ and the fact that $\nabla \times z$ is trivially in $H(\mathrm{div}, \Omega)$,

$$\begin{aligned}\|\tau\|_{L^2}^2 &= \epsilon\langle [\tau \cdot n], \psi\rangle - \epsilon\langle n \cdot \nabla \times z, [\![v]\!]\rangle = \epsilon\langle [\tau \cdot n], \psi\rangle_{\Gamma_h \setminus \Gamma_+} - \epsilon\langle n \cdot \nabla \times z, [\![v]\!]\rangle_{\Gamma_h \setminus (\Gamma_- \cup \Gamma_0)} \\ &\lesssim \epsilon \|[\tau \cdot n]\| \|\psi\|_{H^1(\Omega)} + \epsilon \| [\![v]\!] \| \|\nabla \times z\|_{L^2}.\end{aligned}$$

Applying the bounds $\|\psi\|_{H^1(\Omega)} \leq \frac{1}{\epsilon}\|\tau\|_{L^2}$ and $\|\nabla \times z\|_{L^2} \lesssim \frac{1}{\sqrt{\epsilon}}\|\tau\|_{L^2}$, and noting that $\|\nabla v\| = \frac{1}{\epsilon}\|\tau\|$ completes the proof. □