Robust DPG Methods for Transient Convection-Diffusion

Truman Ellis, Jesse Chan, Leszek Demkowicz

Abstract

We introduce two robust DPG methods for transient convection-diffusion problems. Once a variational formulation is selected, the choice of test norm critically influences the quality of a particular DPG method. It is desirable that a test norm produce convergence of the solution in a norm equivalent to L^2 while producing optimal test functions that can be accurately computed and maintaining good conditioning of the optimal test function solve on highly adaptive meshes. Two such *robust* norms are introduced and proven to guarantee close to L^2 convergence of the primary solution variable. Numerical experiments demonstrate robust convergence of the two methods.

1 Introduction

The discontinuous Petrov-Galerkin finite element method presents an attractive new framework for developing robust numerical methods for computational mechanics. DPG contains the promise of being an automated scientific computing technology – it provides stability for any variational formulation, optimal convergence rates in a user-defined norm, virtually no pre-asymptotic stability issues on coarse meshes, and a measure of the error residual which can be used to robustly drive adaptivity. The method also delivers Hermitian positive definite stiffness matrices for any problem, weak enforcement of boundary conditions, and several other attractive properties [1, 2, 3]. For the most recent review of DPG, see [4]. The process of developing robust DPG methods for steady convection-diffusion was explored in [5, 6]. We start with an abstract derivation of the DPG framework then define the concept of a robust test norm and specialize to transient convection-diffusion. Two new robust norms are derived and numerical verifications of the theory are presented.

1.1 Space-Time Finite Elements

Most finite element simulations of transient phenomena use a semi-discrete formulation: the PDE is first discretized in space using finite elements and then the leftover system of ordinary differential equations in time is usually solved by a finite difference method. But it is possible to treat time as just another dimension to be discretized with finite elements. Some of the earliest proponents of this approach were Kaczkowski[7], Argyris and Scharpf[8], Fried[9], and Oden[10]. These techniques were built on the underlying concept of Hamilton's principle.

Van der Vegt and van der Ven[11] have advanced a space-time discontinuous Galerkin method for 3D inviscid compressible moving boundary problems. Klaij *et al.* [12] then extended the method to compressible Navier-Stokes while Rhebergen *et al.* [13] developed the method for incompressible Navier-Stokes. Rhebergen and Cockburn[14] also developed a space-time HDG method for incompressible Navier-Stokes. Tezduyar and Behr[15] develop a deforming-spatial-domain/space-time procedure coupled with Galerkin/least-squares to handle incompressible Navier-Stokes flows with moving boundaries and later Aliabadi and Tezduyar[16] apply the procedure to compressible flows. Hughes and Stewart[17] develop a general space-time multiscale framework for deriving stabilized methods for transient phenomena.

It is possible to use the semi-discrete approach to solving transient problems with DPG, but it doesn't appear to be a natural fit with the adaptive nature of DPG. The Courant-Friedrichs-Lewy (CFL) condition is not binding with implicit time integration schemes, but it can be a guiding principle for temporal accuracy. If we are interested in temporally accurate solutions, we are limited by the fact that our smallest mesh elements (which may be orders of magnitude smaller than the largest elements) are constrained to proceed at a much smaller time step than the mesh as a whole. We can either restrict the whole mesh to the smallest time step,

or we can attempt some sort of local time stepping. A space-time DPG formulation presents an attractive choice as we will be able to preserve our natural adaptivity from the steady problems while extending it in time. Thus we achieve an adaptive solution technique for transient problems in a unified framework. This paper expands previous work developing robust DPG methods for steady convection-diffusion to the space-time form.

2 Overview of DPG

2.1 A Generalized Minimum Residual Method

We begin with a well posed variational problem: find $u \in U$ such that

$$b(u, v) = l(v) \quad \forall v \in V$$

where $b(u, v)$ is a bilinear (sesquilinear) form on $U \times V$ and $l \in V'$. Introducing operator $B : U \to V'$ (V' is the dual space to V) defined by $b(u, v) = \langle Bu, v \rangle_{V' \times V}$, we can reformulate the equation in operator form:

$$Bu = l \in V' .$$

We wish to find the element u_h of a finite dimensional subspace which minimizes the residual $Bu - l$ in V':

$$u_h = \arg\min_{w_h \in U_h} \frac{1}{2} \|Bu - l\|_{V'}^2 .$$

This mathematical framework is very natural, but it is not yet practical as the V' norm is not especially amenable to computations. With the assumption that we are working with Hilbert spaces, we can use the Riesz representation theorem to find a complementary object in V rather than V'. Let $R_V : V \ni v \to (v, \cdot) \in V'$ be the Riesz map, which is an isometry. Then the inverse Riesz map lets us represent our residual in V:

$$u_h = \arg\min_{w_h \in U_h} \frac{1}{2} \left\| R_V^{-1}(Bu - l) \right\|_V^2 .$$

Since this is a convex minimization problem, the solution is given by the critical points where the Gâteaux derivative is zero in all directions $\delta u \in U_h$:

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

By definition of the Riesz map this is equivalent to the duality pairing

$$\left\langle Bu_h - l, R_V^{-1} B \delta u_h \right\rangle = 0 \quad \forall \delta u_h \in U_h .$$

We can define an optimal test function $v_{\delta u_h} := R_V^{-1} B \delta u_h$ for each trial function δu_h. This allows us to revert back to our original bilinear form with a finite dimensional set of trial and test functions:

$$b(u_h, v_{\delta u_h}) = l(v_{\delta u_h}).$$

Note that $v_{\delta u_h} \in V$ comes from the auxiliary problem

$$(v_{\delta u_h}, \delta v)_V = \langle R_V v_{\delta u_h}, \delta v \rangle = \langle B \delta u_h, \delta v \rangle = b(\delta u_h, \delta v) \quad \forall \delta v \in V.$$

We might refer to this as an *optimal Petrov-Galerkin* method. We arrive at the same method by realizing the supremum in the inf-sup condition (see [2]), motivating the *optimal* nomenclature. As a minimum residual method, optimal Petrov-Galerkin methods produce Hermitian, positive-definite stiffness matrices since

$$b(u_h, v_{\delta \boldsymbol{u}_h}) = (v_{u_h}, v_{\delta \boldsymbol{u}_h})_V = \overline{(v_{\delta \boldsymbol{u}_h}, v_{u_h})} = \overline{b(\delta u_h, v_{u_h})} .$$

The energy norm of the error is directly related to the residual:

$$\|u_h - u\|_E = \|B(u_h - u)\|_{V'} = \|Bu_h - l\|_{V'} = \left\| R_V^{-1}(Bu_h - l) \right\|_V ,$$

where we designate $R_V^{-1}(Bu_h - l)$ the *error representation function*. This has proven to be a very robust *a-posteriori* error estimator for driving adaptivity [18].

2.2 Transient Convection-Diffusion

2.2.1 Problem Description

In order to better illustrate choice of the U and V spaces, we introduce the transient convection-diffusion problem. Consider spatial domain domain Ω and corresponding space-time domain $Q = \Omega \times [0, T]$ with boundary $\Gamma = \Gamma_- \cup \Gamma_+ \cup \Gamma_0 \cup \Gamma_T$ where Γ_- is the inflow boundary ($\boldsymbol{\beta} \cdot \boldsymbol{n}_x < 0$, where $\boldsymbol{\beta}$ is the convection vector and $\boldsymbol{n}_x$ is the outward spatial normal), Γ_+ is the outflow boundary ($\boldsymbol{\beta} \cdot \boldsymbol{n}_x \geq 0$), Γ_0 is the initial time boundary, and Γ_T is the final time boundary. Let $\Gamma_h := \bigcup \partial K$ denote the entire mesh skeleton, where ∂K denotes the boundary of element K. Γ_{h_x} denotes any parts of the skeleton with a nonzero spatial normal and Γ_{h_t} have a nonzero temporal normal.

The transient convection-diffusion equation is

$$\frac{\partial u}{\partial t} + \nabla \cdot (\boldsymbol{\beta} u) - \epsilon \Delta u = f \,,$$

where u is the quantity of interest, often interpreted to be a concentration of some quantity, ϵ is the diffusion coefficient, and f is the source term.

We apply flux boundary conditions on the inflow and trace boundary conditions on the outflow

$$\operatorname{tr}(\boldsymbol{\beta} \cdot u - \epsilon \nabla u) \cdot \boldsymbol{n}_x = t_- \quad \text{on} \quad \Gamma_-$$
$$\operatorname{tr}(u) = u_+ \quad \text{on} \quad \Gamma_+$$
$$\operatorname{tr}(u) = u_0 \quad \text{on} \quad \Gamma_0 \,.$$

We note that Dirichlet boundary conditions also induce Dirichlet boundary conditions for the adjoint problem. Since the direction of convection is reversed for the adjoint convection-diffusion problem, this results in boundary layer adjoint solutions, which must be controlled using special weighted norms [19, 5]. However, since the convection-diffusion operator is not self-adjoint, the Cauchy inflow boundary condition induces a Neumann boundary condition for the adjoint problem. As a result, the adjoint solution does not contain boundary layers, simplifying the construction of a robust DPG method.

2.2.2 Relevant Sobolev Spaces

We begin by defining operators $\nabla_{xt} u := \begin{pmatrix} \nabla u \\ \frac{\partial u}{\partial t} \end{pmatrix}$ and $\nabla_{xt} \cdot \boldsymbol{u} := \nabla \cdot \boldsymbol{u}_x + \frac{\partial u_t}{\partial t}$, where $\boldsymbol{u} = (\boldsymbol{u}_x, u_t)$. We will need the following Sobolev spaces defined on our space-time domain.

$$H^1(Q) = \left\{ u \in L^2(Q) \ : \ \nabla u \in \mathbf{L}^2(Q) \right\}$$
$$H^1_{xt}(Q) = \left\{ u \in L^2(Q) \ : \ \nabla_{xt} u \in \mathbf{L}^2(Q) \right\}$$
$$\boldsymbol{H}(\operatorname{div}, Q) = \left\{ \boldsymbol{\sigma} \in \mathbf{L}^2(Q) \ : \ \nabla \cdot \boldsymbol{\sigma} \in L^2(Q) \right\}$$
$$\boldsymbol{H}(\operatorname{div}_{xt}, Q) = \left\{ \boldsymbol{\sigma} \in \mathbf{L}^2(Q) \ : \ \nabla_{xt} \cdot \boldsymbol{\sigma} \in L^2(Q) \right\}$$

We will also need the corresponding broken Sobolev spaces.

$$H^1(Q_h) = \left\{ u \in L^2(Q) \ : \ u|_K \in H^1(K), \ K \in Q_h \right\} \qquad = \prod_{K \in Q_h} H^1(K)$$

$$H^1_{xt}(Q_h) = \left\{ u \in L^2(Q) \ : \ u|_K \in H^1_{xt}(K), \ K \in Q_h \right\} \qquad = \prod_{K \in Q_h} H^1_{xt}(K)$$

$$\boldsymbol{H}(\operatorname{div}, Q_h) = \left\{ \boldsymbol{\sigma} \in \mathbf{L}^2(Q) \ : \ u|_K \in \boldsymbol{H}(\operatorname{div}, K), \ K \in Q_h \right\} \qquad = \prod_{K \in Q_h} \boldsymbol{H}(\operatorname{div}, K)$$

$$\boldsymbol{H}(\operatorname{div}_{xt}, Q_h) = \left\{ \boldsymbol{\sigma} \in \mathbf{L}^2(Q) \ : \ u|_K \in \boldsymbol{H}(\operatorname{div}_{xt}, K), \ K \in Q_h \right\} \qquad = \prod_{K \in Q_h} \boldsymbol{H}(\operatorname{div}_{xt}, K)$$

Consider the following trace operators:

$$\text{tr}^K_{\text{grad}} u = u|_{\partial K_x} \qquad\qquad u \in H^1(K)$$
$$\text{tr}^K_{\text{div}_{xt}} \boldsymbol{\sigma} = \boldsymbol{\sigma}|_{\partial K_{xt}} \cdot \boldsymbol{n}_{K_{xt}} \qquad\qquad \boldsymbol{\sigma} \in \boldsymbol{H}(\text{div}_{xt}, K)$$

where ∂K_x refers to spatial faces of element K, ∂K_{xt} to the full space-time boundary, and $\boldsymbol{n}_{K_{xt}}$ is the unit outward normal on ∂K_{xt}. The operators tr_{grad} and $\text{tr}_{\text{div}_{xt}}$ perform the same operation element by element to produce the linear maps

$$\text{tr}_{\text{grad}} : H^1(Q_h) \to \prod_{K \in Q_h} H^{1/2}(\partial K_x)$$

$$\text{tr}_{\text{div}_{xt}} : \boldsymbol{H}(\text{div}_{xt}, Q_h) \to \prod_{K \in Q_h} H^{-1/2}(\partial K_{xt})$$

Finally, we define spaces of interface functions. In order that our functions be single valued, we use the following definitions.

$$H^{1/2}(\Gamma_{h_x}) = \text{tr}_{\text{grad}} H^1(Q),$$
$$H^{-1/2}_{xt}(\Gamma_h) = \text{tr}_{\text{div}_{xt}} H(\text{div}_{xt}, Q).$$

For more details on broken and trace Sobolev spaces, see [20].

2.2.3 Variational Formulations

There are many possible manipulations that could be performed before arriving at a variational formulation. We begin by reformulating the problem in terms of the first order system:

$$\frac{1}{\epsilon}\boldsymbol{\sigma} - \nabla u = 0$$
$$\nabla_{xt} \cdot \begin{pmatrix} \boldsymbol{\beta} u - \boldsymbol{\sigma} \\ u \end{pmatrix} = f. \tag{1}$$

Multiplying (1) by test functions $\boldsymbol{\tau} \in \boldsymbol{L}^2(Q)$ and $v \in L^2(Q)$, we obtain the following "trivial" variational formulation equivalent to the strong form:

$$
\begin{aligned}
u &\in H^1_{xt}(Q) & u &= u_+ & &\text{on } \Gamma_+ \\
& & u &= u_0 & &\text{on } \Gamma_0 \\
\boldsymbol{\sigma} &\in \boldsymbol{H}(\text{div}, Q) & (\boldsymbol{\beta} u - \epsilon\nabla u) \cdot \boldsymbol{n} &= t_- & &\text{on } \Gamma_- \\
\left(\frac{1}{\epsilon}\boldsymbol{\sigma}, \boldsymbol{\tau}\right) - (\nabla u, \boldsymbol{\tau}) & & &= 0 & &\forall \boldsymbol{\tau} \in \boldsymbol{L}^2(Q) \\
\left(\nabla_{xt} \cdot \begin{pmatrix} \boldsymbol{\beta} u - \boldsymbol{\sigma} \\ u \end{pmatrix}, v\right) & & &= f & &\forall v \in L^2(Q).
\end{aligned}
\tag{2}
$$

We can now choose either to relax (integrate by parts and build in the boundary conditions) or strongly enforce each equation. The steady state case and resulting options are explored and analyzed in further detail in [21] and are termed the trivial formulation (don't relax anything), the classical formulation (relax the second equation), the mixed formulation (relax the first equation), and the ultra-weak formulation (relax both equations). The stability constants for the four formulations are related, but the functional settings and norms of convergence change. Early DPG work emphasized the ultra-weak formulation since in many ways it was the easiest to analyze, though recently the classical formulation has been under very active

consideration. In the interests of simpler analysis, we focus on the ultra-weak formulation in this paper.

$$
\begin{aligned}
&u \in L^2(Q)\,,\ \boldsymbol{\sigma} \in \mathbf{L}^2(Q) \\
&\left(\frac{1}{\epsilon}\boldsymbol{\sigma},\boldsymbol{\tau}\right) + (u, \nabla \cdot \boldsymbol{\tau}) && = 0 \quad \forall \boldsymbol{\tau} \in \boldsymbol{H}(\mathrm{div}, Q) \,:\, \boldsymbol{\tau} \cdot \boldsymbol{n}_x = 0 \text{ on } \Gamma_- \\
&-\left(\left(\begin{array}{c}\boldsymbol{\beta} u - \boldsymbol{\sigma} \\ u\end{array}\right), \nabla_{xt} v\right) && = f \quad \forall v \in H^1_{xt}(Q) \,:\, v = 0 \text{ on } \Gamma_+ \cup \Gamma_0\,,
\end{aligned}
\tag{3}
$$

We can remove the conditions on the test functions by introducing trace unknowns

$$
\hat{u} = \mathrm{tr}(u) \qquad\qquad\qquad \text{on } \partial Q_x
$$

$$
\hat{t} = \mathrm{tr}\left(\begin{array}{c}\boldsymbol{\beta} u - \boldsymbol{\sigma} \\ u\end{array}\right) \cdot \boldsymbol{n}_{xt} \qquad\qquad \text{on } \partial Q_{xt}\,.
$$

Our new ultra-weak formulation with conforming test functions is

$$
\begin{aligned}
&u \in L^2(Q)\,,\ \boldsymbol{\sigma} \in \mathbf{L}^2(Q) \\
&\hat{u} \in H^{1/2}(\partial Q_x)\,, && \hat{u} = u_+ \text{ on } \Gamma_+ \\
&\hat{t} \in H^{-1/2}_{xt}(\partial Q)\,, && \hat{t} = t_- \text{ on } \Gamma_-\,, \quad \hat{t} = -u_0 \text{ on } \Gamma_0 \\
&\left(\frac{1}{\epsilon}\boldsymbol{\sigma},\boldsymbol{\tau}\right) + (u, \nabla \cdot \boldsymbol{\tau}) - \langle \hat{u}, \boldsymbol{\tau} \cdot \boldsymbol{n}_x \rangle && = 0 \quad \forall \boldsymbol{\tau} \in \boldsymbol{H}(\mathrm{div}, Q) \\
&-\left(\left(\begin{array}{c}\boldsymbol{\beta} u - \boldsymbol{\sigma} \\ u\end{array}\right), \nabla_{xt} v\right) + \langle \hat{t}, v \rangle && = f \quad \forall v \in H^1_{xt}(Q)\,.
\end{aligned}
\tag{4}
$$

2.2.4 Broken Test Functions

One of the key insights that led to the development of the DPG framework was the process of breaking test functions, that is testing with functions from larger broken Sobolev spaces, replacing $H^1_{xt}(Q)$ with $H^1_{xt}(Q_h)$ and $\boldsymbol{H}(\mathrm{div}, Q)$ with $\boldsymbol{H}(\mathrm{div}, Q_h)$. Discretizing such spaces is much simpler than standard spaces which require enforcement of global continuity conditions. The cost of introducing broken spaces is that we have to extend our interface unknowns $\hat{u}$ and $\hat{t}$ to live on the mesh skeleton. Our ultra-weak formulation with broken test functions looks like

$$
\begin{aligned}
&u \in L^2(Q)\,,\ \boldsymbol{\sigma} \in \mathbf{L}^2(Q) \\
&\hat{u} \in H^{1/2}(\Gamma_{h_x})\,, && \hat{u} = u_+ \text{ on } \Gamma_+ \\
&\hat{t} \in H^{-1/2}_{xt}(\Gamma_h)\,, && \hat{t} = t_- \text{ on } \Gamma_-\,, \quad \hat{t} = -u_0 \text{ on } \Gamma_0 \\
&\left(\frac{1}{\epsilon}\boldsymbol{\sigma},\boldsymbol{\tau}\right) + (u, \nabla \cdot \boldsymbol{\tau}) - \langle \hat{u}, \boldsymbol{\tau} \cdot \boldsymbol{n}_x \rangle && = 0 \quad \forall \boldsymbol{\tau} \in \boldsymbol{H}(\mathrm{div}, Q_h) \\
&-\left(\left(\begin{array}{c}\boldsymbol{\beta} u - \boldsymbol{\sigma} \\ u\end{array}\right), \nabla_{xt} v\right) + \langle \hat{t}, v \rangle && = f \quad \forall v \in H^1_{xt}(Q_h)\,.
\end{aligned}
\tag{5}
$$

The main consequence of breaking test functions is that it reduces the cost of solving for optimal test functions from a global solve to an embarrassingly parallel solve element-by-element. Now that we've derived a suitable variational formulation, we are left with the task of selecting a test norm with which to compute our optimal test functions.

3 Robust Test Norms

The final unresolved choice is what norm to apply to the V space. This is one of the most important factors in designing a robust DPG method as the corresponding Riesz operator needs to be inverted to solve for

the optimal test functions. If the norm produces unresolved boundary layers in the auxiliary problem, then many of the attractive features of DPG may fall apart. This is the primary emphasis of this paper. The problem of constructing stable test norms for steady convection-diffusion was addressed in [5, 6]. In this paper, we extend that work to transient convection-diffusion in space-time.

We define a robust test norm such that the L^2 norm of the solution is bounded by the energy norm of the solution with a constant independent of ϵ. We can rewrite any ultra-weak formulation with broken test functions as the following bilinear form with group variables:

$$b\left((u, \hat{u}), v\right) = (u, A^*v)_{L^2} + \langle \hat{u}, [\![v]\!] \rangle_{\Gamma_h}$$

where A^* represents the adjoint. In the case of convection-diffusion, $u := \{u, \boldsymbol{\sigma}\}$, $\hat{u} := \{\hat{u}, \hat{t}\}$, $v := \{v, \boldsymbol{\tau}\}$.

Note that for conforming v^* satisfying $A^*v^* = u$

$$\|u\|_{L^2}^2 = b(u, v^*) = \frac{b(u, v^*)}{\|v^*\|_V} \|v^*\|_V$$

$$\leq \sup_{v^* \neq 0} \frac{|b(u, v^*)|}{\|v^*\|} \|v^*\| = \|u\|_E \|v^*\|_V \ .$$

This defines a necessary condition for robustness, namely that

$$\|v^*\|_V \lesssim \|u\|_{L^2} \ . \tag{6}$$

If this condition is satisfied, then we get our final result:

$$\|u\|_{L^2} \lesssim \|u\|_E \ .$$

So far, we've assumed that our finite set of optimal test functions are assembled from an infinite dimensional space. In practice, we have found it to be sufficient to use an "enriched" space of higher polynomial dimension than the trial space [22]. This adds an additional requirement when assembling a robust test norm, namely that our optimal test functions should be adequately representable within this enriched space. We illustrate this point by considering three norms which satisfy the above conditions for 1D steady convection-diffusion. The graph norm is $\left(\|A^*v\|_{L^2}^2 + \|v\|_{L^2}^2\right)^{\frac{1}{2}}$:

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

The robust norm was derived in [6]:

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

The case for the coupled robust norm was made in [23]:

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

The argument for the coupled norm was that in certain cases we noticed pollution of u from errors in $\boldsymbol{\sigma}$, for example at singularities in $\boldsymbol{\sigma}$, u also exhibited degraded quality with the robust norm. The coupled robust norm seemed to relax this behavior, i.e. errors in u appear more independent of errors in $\boldsymbol{\sigma}$.

The bilinear form and test norm define a mapping from input trial functions to an optimal test function:

$$T = R_V^{-1}B : U \to V \ .$$

Below, we plot the optimal test functions produced given $\epsilon = 10^{-2}$, a representative trial function $u = x - \frac{1}{2}$, and either the graph norm, the robust norm, or the coupled robust norm. Note that the optimal test functions will be different for any other trial function. In the left column, we see the fully resolved *ideal* optimal test
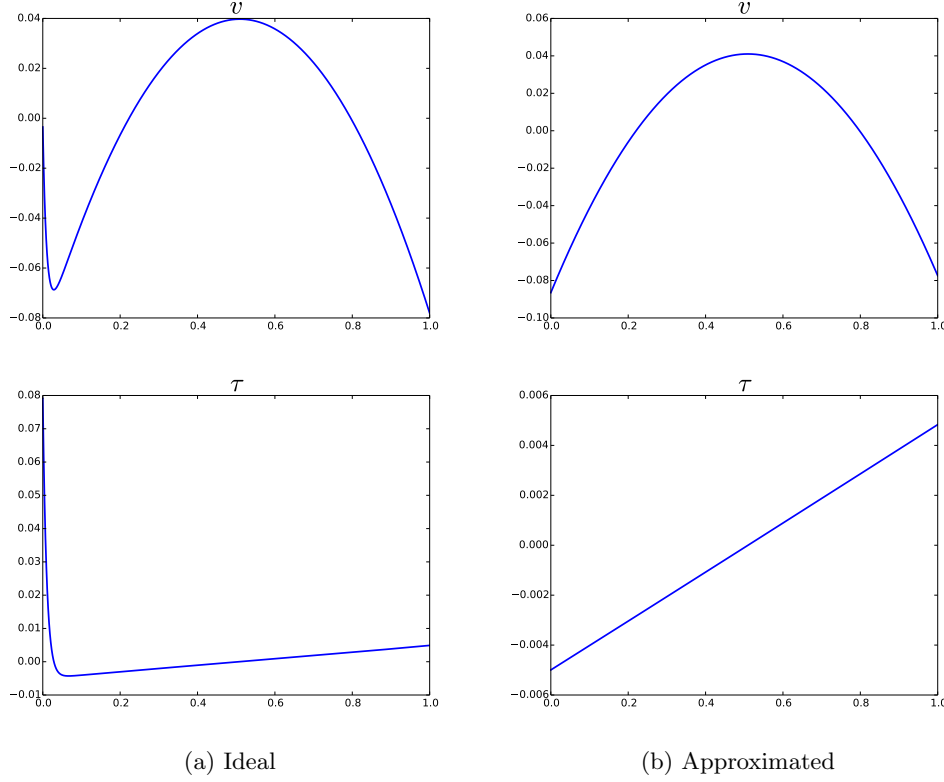
6

(a) Ideal　　　　　　　　　　　　　　(b) Approximated

Figure 1: Graph norm optimal test functions for $u = x - \frac{1}{2}$

function that DPG theory relies on. On the right, we see the approximated optimal test function using a enriched cubic test space.

Mathematically, the graph norm satisfies the necessary condition to be a robust norm, but the ideal optimal test functions contain strong boundary layers which can not be realistically approximated with the provided enriched space. If the approximated optimal test functions can not come sufficiently close to the ideal, then the whole DPG theory falls apart. See [22] for more discussion. This provides an additional condition on a test norm before we can truly call it robust: the ideal test functions must be adequately representable within the provided enriched space. This ultimately comes down to an analysis of the relative magnitudes of individual terms within the test norm, usually attempting to bound reactive or convective terms by diffusive terms. The coupled robust norm satisfies condition (6) and also produces relatively smooth optimal test functions that can be sufficiently approximated with a cubic polynomial space. Niemi *et al.* attempted to approximate boundary layers in optimal shape functions with Shishkin meshes [24, 25].

3.1　Application to Transient Convection-Diffusion

Now we present the analysis leading to two robust norms for transient convection-diffusion. Consider the problem with homogeneous boundary conditions

$$\frac{1}{\epsilon}\boldsymbol{\sigma} - \nabla u = 0$$

$$\frac{\partial u}{\partial t} + \boldsymbol{\beta} \cdot \nabla u - \nabla \cdot \boldsymbol{\sigma} = f$$

$$\beta_n u - \epsilon \frac{\partial u}{\partial n} = 0 \text{ on } \Gamma_-$$

$$u = 0 \text{ on } \Gamma_+$$

$$u = u_0 \text{ on } \Gamma_0.$$

7

(a) Ideal

(b) Approximated

Figure 2: Robust norm optimal test functions for $u = x - \frac{1}{2}$

Let $\tilde{\boldsymbol{\beta}} := \begin{pmatrix} \boldsymbol{\beta} \\ 1 \end{pmatrix}$, then we can rewrite this as

$$\frac{1}{\epsilon}\boldsymbol{\sigma} - \nabla u = 0$$
$$\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} u - \nabla \cdot \boldsymbol{\sigma} = f$$
$$\beta_n u - \epsilon \frac{\partial u}{\partial n} = 0 \text{ on } \Gamma_-$$
$$u = 0 \text{ on } \Gamma_+$$
$$u = u_0 \text{ on } \Gamma_0.$$

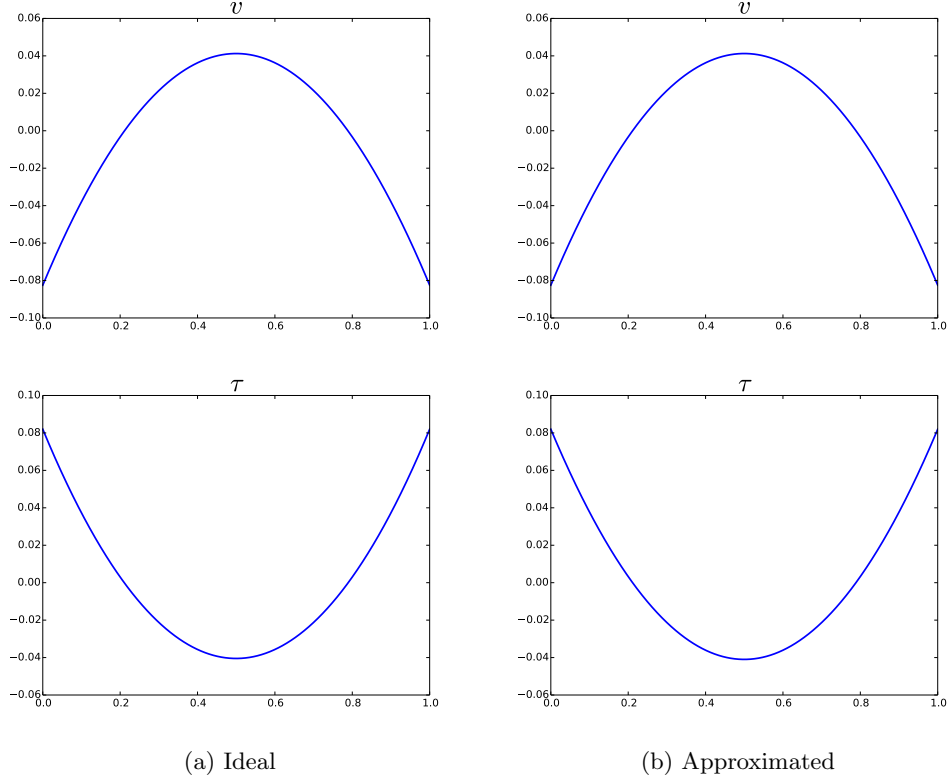The adjoint operator A^* is given by

$$A^*(v, \boldsymbol{\tau}) = \left(\frac{1}{\epsilon}\boldsymbol{\tau} + \nabla v, -\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v + \nabla \cdot \boldsymbol{\tau} \right).$$

We decompose now the continuous adjoint problem

$$A^*(v, \boldsymbol{\tau}) = (\boldsymbol{f}, g)$$
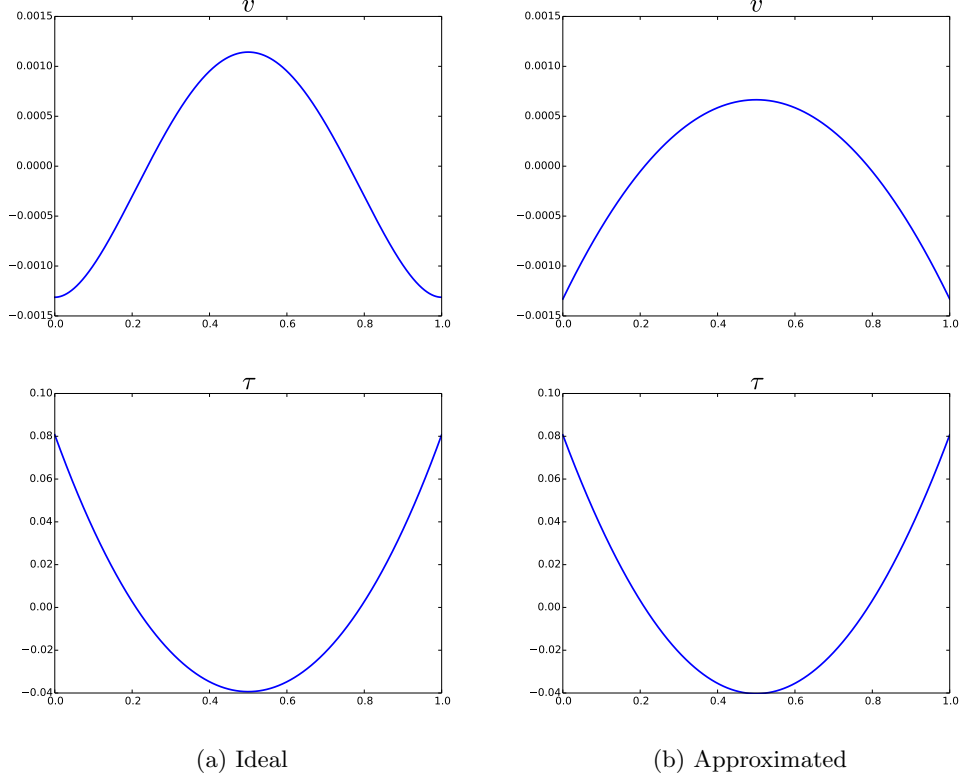
8

<div align="center">(a) Ideal   (b) Approximated</div>

Figure 3: Coupled robust norm optimal test functions for $u = x - \frac{1}{2}$

into two cases a continuous part with forcing term g

$$\frac{1}{\epsilon}\boldsymbol{\tau}_1 + \nabla v_1 = 0$$

$$-\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v_1 + \nabla \cdot \boldsymbol{\tau}_1 = g$$

$$\boldsymbol{\tau}_1 \cdot \boldsymbol{n}_x = 0 \text{ on } \Gamma_-$$

$$v_1 = 0 \text{ on } \Gamma_+$$

$$v_1 = 0 \text{ on } \Gamma_T,$$

and a continuous part with forcing $\boldsymbol{f}$

$$\frac{1}{\epsilon}\boldsymbol{\tau}_2 + \nabla v_2 = \boldsymbol{f}$$

$$-\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v_2 + \nabla \cdot \boldsymbol{\tau}_2 = 0$$

$$\boldsymbol{\tau}_2 \cdot \boldsymbol{n}_x = 0 \text{ on } \Gamma_-$$

$$v_2 = 0 \text{ on } \Gamma_+$$

$$v_2 = 0 \text{ on } \Gamma_T.$$

(The boundary conditions can be derived by taking the ultra-weak formulation and choosing boundary conditions such that the temporal flux and spatial flux terms $\langle \widehat{u}, [\![\tau_n]\!]\rangle_{\Gamma_{out}}$ and $\langle \hat{t}_n, [\![v]\!]\rangle_{\Gamma_{in}}$ are zero.)

We can then derive that the test norms

$$\|(v, \boldsymbol{\tau})\|_{V,K}^2 := \left\|\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v\right\|_K^2 + \epsilon \|\nabla v\|_K^2 + \|v\|_K^2 + \|\nabla \cdot \boldsymbol{\tau}\|_K^2 + \frac{1}{\epsilon}\|\boldsymbol{\tau}\|_K^2 , \tag{7}$$

and

$$\|(v, \boldsymbol{\tau})\|_{V,K}^2 := \left\|\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v\right\|_K^2 + \epsilon \|\nabla v\|_K^2 + \|v\|_K^2 + \left\|\nabla \cdot \boldsymbol{\tau} - \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v\right\|_K^2 + \frac{1}{\epsilon}\|\boldsymbol{\tau}\|_K^2 , \tag{8}$$

<div align="center">9</div>

respectively designated the *robust* test norm and the *coupled robust* test norm, provide the necessary bound $\|v^*\|_V \lesssim \|u\|_{L^2(Q)}$.

In the following lemmas we establish the following bounds:

- Bound on $\|(v_1, \boldsymbol{\tau}_1)\|_V$. Lemma 3.2 gives $\left\|\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v_1\right\| \leq \|g\|$. Since $\nabla \cdot \boldsymbol{\tau}_1 = g + \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v_1$,

$$\|\nabla \cdot \boldsymbol{\tau}_1\| \leq \|g\| + \left\|\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v_1\right\| \leq 2\|g\|.$$

Or, the fact that $\nabla \cdot \boldsymbol{\tau} - \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v_1 = g$ clearly gives

$$\left\|\nabla \cdot \boldsymbol{\tau} - \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v_1\right\| = \|g\|.$$

Lemma 3.1 gives $\|v_1\|^2 + \epsilon \|\nabla v_1\|^2 \leq \|g\|^2$. Since $\epsilon^{1/2} \nabla v_1 = -\epsilon^{-1/2} \boldsymbol{\tau}_1$,

$$\frac{1}{\epsilon} \|\boldsymbol{\tau}_1\|^2 \leq \|g\|^2.$$

Thus, all $(v_1, \boldsymbol{\tau}_1)$ terms in (7) and (8) are accounted for, guaranteeing at least robust control of u.

- Bound on $\|(v_2, \boldsymbol{\tau}_2)\|_V$. The fact that $\nabla \cdot \boldsymbol{\tau} - \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v = 0$ clearly gives

$$\left\|\nabla \cdot \boldsymbol{\tau} - \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v_2\right\| = 0 \leq \|\boldsymbol{f}\|.$$

Lemma 3.1 gives $\|v_2\|^2 + \epsilon \|\nabla v_2\|^2 \leq \epsilon \|\boldsymbol{f}\|^2$. Since $\epsilon^{1/2} \nabla v_2 = \boldsymbol{f} - \epsilon^{-1/2} \boldsymbol{\tau}_2$,

$$\frac{1}{\epsilon} \|\boldsymbol{\tau}_2\|^2 \leq (1 + \epsilon) \|\boldsymbol{f}\|^2.$$

We have not been able to develop bounds on $\left\|\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v_2\right\|$ and $\|\nabla \cdot \boldsymbol{\tau}\|$ which means that we can not guarantee robust control of $\boldsymbol{\sigma}$ with with provided test norms.

We proceed now with the technical estimates.

Lemma 3.1. *For the duration of this lemma, let $v := v_1 + v_2$. Assuming the advection field $\boldsymbol{\beta}$ is incompressible, i.e. $\nabla \cdot \boldsymbol{\beta} = 0$,*
$$\|v\|^2 + \epsilon \|\nabla v\|^2 \leq \|g\|^2 + \epsilon \|\boldsymbol{f}\|^2.$$

Proof. Define $w = e^t v$ and note that $\frac{\partial w}{\partial t} = \left(\frac{\partial v}{\partial t} + v\right) e^t$ while all spatial derivatives go through. Multiplying the adjoint by w and integrating over Q gives

$$-\int_Q \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v w - \epsilon \Delta v w = \int_Q g w - \epsilon \int_Q \nabla \cdot \boldsymbol{f} w$$

or

$$-\int_Q e^t v \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v - \epsilon \int_Q e^t v \Delta v = \int_Q e^t g v - \epsilon \int_Q e^t v \nabla \cdot \boldsymbol{f}$$

Integrating by parts:

$$\int_Q \nabla_{xt} \cdot \left(e^t \tilde{\boldsymbol{\beta}} v\right) v - \int_\Gamma e^t \tilde{\boldsymbol{\beta}} \cdot \boldsymbol{n} v^2 + \epsilon \int_Q e^t \nabla v \cdot \nabla v - \epsilon \int_{\Gamma_x} e^t v \cdot \nabla v \cdot \boldsymbol{n}_x$$
$$= \int_Q e^t g v + \epsilon \int_Q e^t \nabla v \cdot \boldsymbol{f} - \epsilon \int_{\Gamma_x} e^t v \boldsymbol{f} \cdot \boldsymbol{n}_x$$

10

Note that $\nabla_{xt} \cdot e^t v \tilde{\boldsymbol{\beta}} = e^t(\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v + v)$ if $\nabla \cdot \boldsymbol{\beta} = 0$. Moving some terms to the right hand side, we get

$$\int_Q e^t v^2 + \int_Q \epsilon e^t \nabla v \cdot \nabla v$$

$$= \int_Q e^t g v + \epsilon \int_Q e^t \nabla v \cdot \boldsymbol{f} - \epsilon \int_{\Gamma_x} e^t v \boldsymbol{f} \cdot \boldsymbol{n}_x$$

$$- \int_Q e^t \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v v \quad + \int_\Gamma e^t \tilde{\boldsymbol{\beta}} \cdot \boldsymbol{n} v^2 \quad + \epsilon \int_{\Gamma_x} e^t v \cdot \nabla v \cdot \boldsymbol{n}_x$$

Note that $1 \le \|e^t\|_\infty = e^T$. Then

$$\|v\|^2 + \epsilon \|\nabla v\|^2$$

$$\le e^T \left(\int_Q g v + \epsilon \int_Q \nabla v \cdot \boldsymbol{f} - \epsilon \int_{\Gamma_-} v \underbrace{\boldsymbol{f} \cdot \boldsymbol{n}_x}_{= \tau_n + \frac{\partial v}{\partial \boldsymbol{n}_x}} - \epsilon \int_{\Gamma_+} \underbrace{v}_{=0} \boldsymbol{f} \cdot \boldsymbol{n}_x \right.$$

$$\left. - \int_Q \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v v \quad + \int_\Gamma \tilde{\boldsymbol{\beta}} \cdot \boldsymbol{n} v^2 \quad + \epsilon \int_{\Gamma_-} v \cdot \nabla v \cdot \boldsymbol{n}_x \quad + \epsilon \int_{\Gamma_+} \underbrace{v}_{=0} \frac{\partial v}{\partial \boldsymbol{n}_x} \right)$$

$$= e^T \left(\int_Q g v + \epsilon \int_Q \nabla v \cdot \boldsymbol{f} - \epsilon \cancel{\int_{\Gamma_-} v \frac{\partial v}{\partial \boldsymbol{n}_x}} + \epsilon \cancel{\int_{\Gamma_x} v \frac{\partial v}{\partial \boldsymbol{n}_x}} - \frac{1}{2} \int_Q \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v^2 \quad + \int_\Gamma \tilde{\boldsymbol{\beta}} \cdot \boldsymbol{n} v^2 \right)$$

$$= e^T \left(\int_Q g v + \epsilon \int_Q \nabla v \cdot \boldsymbol{f} + \frac{1}{2} \int_Q \overset{0}{\cancel{\nabla_{xt} \cdot \tilde{\boldsymbol{\beta}} v^2}} - \frac{1}{2} \int_\Gamma \tilde{\boldsymbol{\beta}} \cdot \boldsymbol{n} v^2 \quad + \int_\Gamma \tilde{\boldsymbol{\beta}} \cdot \boldsymbol{n} v^2 \right)$$

$$= e^T \left(\int_Q g v + \epsilon \int_Q \nabla v \cdot \boldsymbol{f} + \frac{1}{2} \left(\int_{\Gamma_0} \underbrace{-v^2}_{\le 0} + \int_{\Gamma_T} \overset{0}{\cancel{v^2}} + \int_{\Gamma_-} \underbrace{\boldsymbol{\beta} \cdot \boldsymbol{n}_x v^2}_{\le 0} + \int_{\Gamma_+} \boldsymbol{\beta} \cdot \boldsymbol{n}_x \overset{0}{\cancel{v^2}} \right) \right)$$

$$\le e^T \left(\int_Q g v + \epsilon \int_Q \nabla v \cdot \boldsymbol{f} \right)$$

$$\le e^T \left(\frac{\|g\|^2}{2} + \epsilon \frac{\|\boldsymbol{f}\|^2}{2} + \frac{\|v\|^2}{2} + \epsilon \frac{\|\nabla v\|^2}{2} \right)$$

$\square$

Lemma 3.2. *If $\left\| \nabla \boldsymbol{\beta} - \frac{1}{2} \nabla \cdot \boldsymbol{\beta} \boldsymbol{I} \right\|_{L^\infty} \le C_{\boldsymbol{\beta}}$ we can bound*

$$\left\| \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v_1 \right\| \lesssim \|g\|.$$

Proof. Multiply $-\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v = g - \nabla \cdot \boldsymbol{\tau}$ by $-\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v$ and integrate over Q to get

$$\left\| \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \right\|^2 = - \int_Q g \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v + \int_Q \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \nabla \cdot \boldsymbol{\tau}. \tag{9}$$

Note that

$$\frac{1}{\epsilon} \int_Q \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \nabla \cdot \boldsymbol{\tau} = - \int_Q \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \nabla \cdot \nabla v$$

$$= - \int_{\Gamma_x} \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \nabla v \cdot \boldsymbol{n}_x + \int_Q \nabla(\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v) \cdot \nabla v$$

$$= - \int_{\Gamma_x} \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \nabla v \cdot \boldsymbol{n}_x + \int_Q (\nabla\tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v) \cdot \nabla v + \int_Q \tilde{\boldsymbol{\beta}} \cdot \nabla\nabla_{xt} v \cdot \nabla v$$

$$= - \int_{\Gamma_x} \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \nabla v \cdot \boldsymbol{n}_x + \int_Q (\nabla\boldsymbol{\beta} \cdot \nabla v) \cdot \nabla v + \frac{1}{2} \int_Q \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt}(\nabla v \cdot \nabla v)$$

$$= - \int_{\Gamma_x} \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \nabla v \cdot \boldsymbol{n}_x + \int_Q (\nabla\boldsymbol{\beta} \cdot \nabla v) \cdot \nabla v + \frac{1}{2} \int_\Gamma \tilde{\boldsymbol{\beta}} \cdot \boldsymbol{n}(\nabla v \cdot \nabla v) - \frac{1}{2} \int_Q \nabla_{xt} \cdot \tilde{\boldsymbol{\beta}}(\nabla v \cdot \nabla v)$$

$$= - \int_{\Gamma_x} \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \nabla v \cdot \boldsymbol{n}_x + \int_Q (\nabla\boldsymbol{\beta} \cdot \nabla v) \cdot \nabla v + \frac{1}{2} \int_\Gamma \tilde{\boldsymbol{\beta}} \cdot \boldsymbol{n}(\nabla v \cdot \nabla v) - \frac{1}{2} \int_Q \nabla \cdot \boldsymbol{\beta}(\nabla v \cdot \nabla v)$$

$$= - \int_{\Gamma_x} \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \nabla v \cdot \boldsymbol{n}_x + \frac{1}{2} \int_\Gamma \tilde{\boldsymbol{\beta}} \cdot \boldsymbol{n}(\nabla v \cdot \nabla v) + \int_Q \nabla v(\nabla\boldsymbol{\beta} - \frac{1}{2}\nabla \cdot \boldsymbol{\beta}\boldsymbol{I})\nabla v$$

Plugging this into (9), we get

$$\left\|\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v\right\|^2 = -\int_Q g\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v + \epsilon\int_Q \nabla v(\nabla\boldsymbol{\beta} - \frac{1}{2}\nabla\cdot\boldsymbol{\beta I})\nabla v$$

$$-\epsilon\int_{\Gamma_x}\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v\nabla v\cdot\boldsymbol{n}_x + \frac{\epsilon}{2}\int_\Gamma \tilde{\boldsymbol{\beta}}\cdot\boldsymbol{n}(\nabla v\cdot\nabla v)$$

$$= -\int_Q g\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v + \epsilon\int_Q \nabla v(\nabla\boldsymbol{\beta} - \frac{1}{2}\nabla\cdot\boldsymbol{\beta I})\nabla v$$

$$-\epsilon\int_{\Gamma_-}\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v\underbrace{\nabla v\cdot\boldsymbol{n}_x}_{=0} -\epsilon\int_{\Gamma_+}\left(\underbrace{\frac{\partial v}{\partial t}}_{=0} +\boldsymbol{\beta}\cdot\nabla v\right)\nabla v\cdot\boldsymbol{n}_x$$

$$+\frac{\epsilon}{2}\int_{\Gamma_-}\underbrace{\boldsymbol{\beta}\cdot\boldsymbol{n}_x}_{<0}(\nabla v\cdot\nabla v) + \frac{\epsilon}{2}\int_{\Gamma_+}\boldsymbol{\beta}\cdot\boldsymbol{n}_x(\nabla v\cdot\nabla v)$$

$$+\frac{\epsilon}{2}\int_{\Gamma_0}\underbrace{n_t}_{<0}(\nabla v\cdot\nabla v) + \frac{\epsilon}{2}\int_{\Gamma_T}n_t\underbrace{(\nabla v\cdot\nabla v)}_{=0}$$

$$\leq -\int_Q g\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v + \epsilon\int_Q \nabla v(\nabla\boldsymbol{\beta} - \frac{1}{2}\nabla\cdot\boldsymbol{\beta I})\nabla v$$

$$+\epsilon\int_{\Gamma_+}\left(-\frac{\partial v}{\partial\boldsymbol{n}_x}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n}_x\nabla v\right)\cdot\nabla v$$

$$= -\int_Q g\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v + \epsilon\int_Q \nabla v(\nabla\boldsymbol{\beta} - \frac{1}{2}\nabla\cdot\boldsymbol{\beta I})\nabla v$$

$$+\epsilon\int_{\Gamma_+}\left(-\frac{\partial v}{\partial\boldsymbol{n}_x}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n}_x\frac{\partial v}{\partial\boldsymbol{n}_x}\boldsymbol{n}_x\right)\cdot\frac{\partial v}{\partial\boldsymbol{n}_x}\boldsymbol{n}_x$$

$$= -\int_Q g\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v + \epsilon\int_Q \nabla v(\nabla\boldsymbol{\beta} - \frac{1}{2}\nabla\cdot\boldsymbol{\beta I})\nabla v$$

$$\underbrace{-\frac{\epsilon}{2}\int_{\Gamma_+}\left(\frac{\partial v}{\partial\boldsymbol{n}_x}\right)^2\boldsymbol{\beta}\cdot\boldsymbol{n}_x}_{<0}$$

$$\leq -\int_Q g\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v + \epsilon\int_Q \nabla v(\nabla\boldsymbol{\beta} - \frac{1}{2}\nabla\cdot\boldsymbol{\beta I})\nabla v$$

$$\leq \frac{\|g\|^2}{2} + \frac{\left\|\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v\right\|^2}{2} + \epsilon\int_Q \nabla v(\nabla\boldsymbol{\beta} - \frac{1}{2}\nabla\cdot\boldsymbol{\beta I})\nabla v$$

$$\leq \frac{\|g\|^2}{2} + \frac{\left\|\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v\right\|^2}{2} + \epsilon C_{\boldsymbol{\beta}}\|\nabla v\|^2$$

$$\leq \left(\frac{1}{2} + C_{\boldsymbol{\beta}}\right)\|g\|^2 + \frac{\left\|\tilde{\boldsymbol{\beta}}\cdot\nabla_{xt}v\right\|^2}{2}$$

$$\square$$

4   Numerical Tests

The norms given in (7) and (8) are robust, but the reaction (0th order) terms still dominate the diffusion terms which produces boundary layers in optimal test functions and prohibits their resolution with a simple

enrichment strategy. We can mitigate this by introducing mesh-dependent norms:

$$\|(v, \boldsymbol{\tau})\|_{V,K}^2 := \left\| \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \right\|_K^2 + \epsilon \|\nabla v\|_K^2 + \min\left(\frac{\epsilon}{h^2}, 1\right) \|v\|_K^2 + \|\nabla \cdot \boldsymbol{\tau}\|_K^2 + \min\left(\frac{1}{\epsilon}, \frac{1}{h^2}\right) \|\boldsymbol{\tau}\|_K^2 , \quad (10)$$

and

$$\|(v, \boldsymbol{\tau})\|_{V,K}^2 := \left\| \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \right\|_K^2 + \epsilon \|\nabla v\|_K^2 + \min\left(\frac{\epsilon}{h^2}, 1\right) \|v\|_K^2 + \left\| \nabla \cdot \boldsymbol{\tau} - \tilde{\boldsymbol{\beta}} \cdot \nabla_{xt} v \right\|_K^2 + \min\left(\frac{1}{\epsilon}, \frac{1}{h^2}\right) \|\boldsymbol{\tau}\|_K^2 . \quad (11)$$

Note that any version of (7) and (8) with smaller coefficients also satisfies the criteria for robustness. The mesh dependent coefficients were chosen in an attempt to balance the relative size of "reaction" terms like $\|v\|$ which scale like h^d with "diffusive" terms like $\epsilon \|\nabla v\|$ which scale like h^{d-2}. This is also the mechanism by which we avoid creating sharp boundary layers in our optimal test functions – by correctly balancing reactive and diffusive terms. In the following numerical experiments, we compute with these mesh dependent norms.

We verify robust convergence of our transient coupled robust norm on an analytical solution (shown in Figure 4) that decays to a steady state Eriksson-Johnson problem:

$$u = \exp(-lt) \left[\exp(\lambda_1 x) - \exp(\lambda_2 x)\right] + \cos(\pi y) \frac{\exp(s_1 x) - \exp(r_1 x)}{\exp(-s_1) - \exp(-r_1)} ,$$

where $l = 4$, $\lambda_{1,2} = \frac{-1 \pm \sqrt{1-4\epsilon l}}{-2\epsilon}$, $r_1 = \frac{1+\sqrt{1+4\pi^2\epsilon^2}}{2\epsilon}$, and $s_1 = \frac{1-\sqrt{1+4\pi^2\epsilon^2}}{2\epsilon}$. The problem domain is $[-1, 0] \times [-0.5, 0.5]$ and $\boldsymbol{\beta} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. We show robustness for $\epsilon = 10^{-2}$, 10^{-4}, 10^{-6}, 10^{-8} for linear, quadratic, and quartic polynomial trial functions. Flux boundary conditions were applied based on the exact solution at $x = -1$ and $t = 0$ while trace boundary conditions were set at $y = -0.5$, $y = 0.5$, and $x = 0$. An adaptive solve was undertaken using a greedy refinement strategy in which any elements with at least 20% of the energy error of highest energy error element were refined at each step. See [18] for details on adaptivity within the DPG context.

In the plot legends, L^2 indicates $\left(\|u - u_{\text{exact}}\|_L^2 + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\text{exact}}\|_{L^2} \right)^{\frac{1}{2}}$ while V^* indicates the energy error reported by the method. Despite a lack of guaranteed control $\boldsymbol{\sigma}$ by norms (10) and (11), $\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\text{exact}}\|_{L^2}$ is included in the L^2 error computation and does appear to be under control in the problems considered here. When plotted in isolation, the L^2 error in $\boldsymbol{\sigma}$ was usually orders of magnitude smaller than $\|u - u_{\text{exact}}\|_{L^2}$.
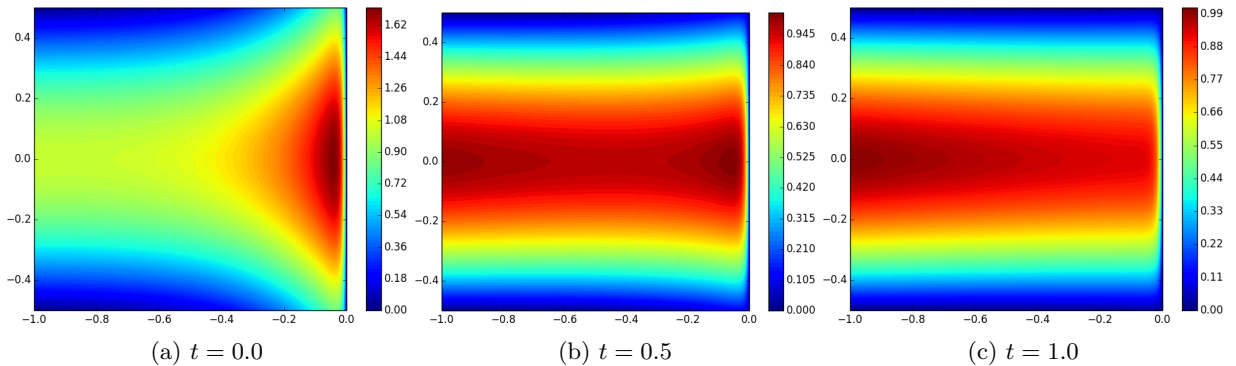

(a) $t = 0.0$ (b) $t = 0.5$ (c) $t = 1.0$

Figure 4: Transient Eriksson-Johnson solution

5   Conclusions

As expected, convergence of the energy error appears to be a reliable predictor of convergence of the L^2 error. This relation is especially tight for moderate values of ϵ. We've developed two robust test norms for
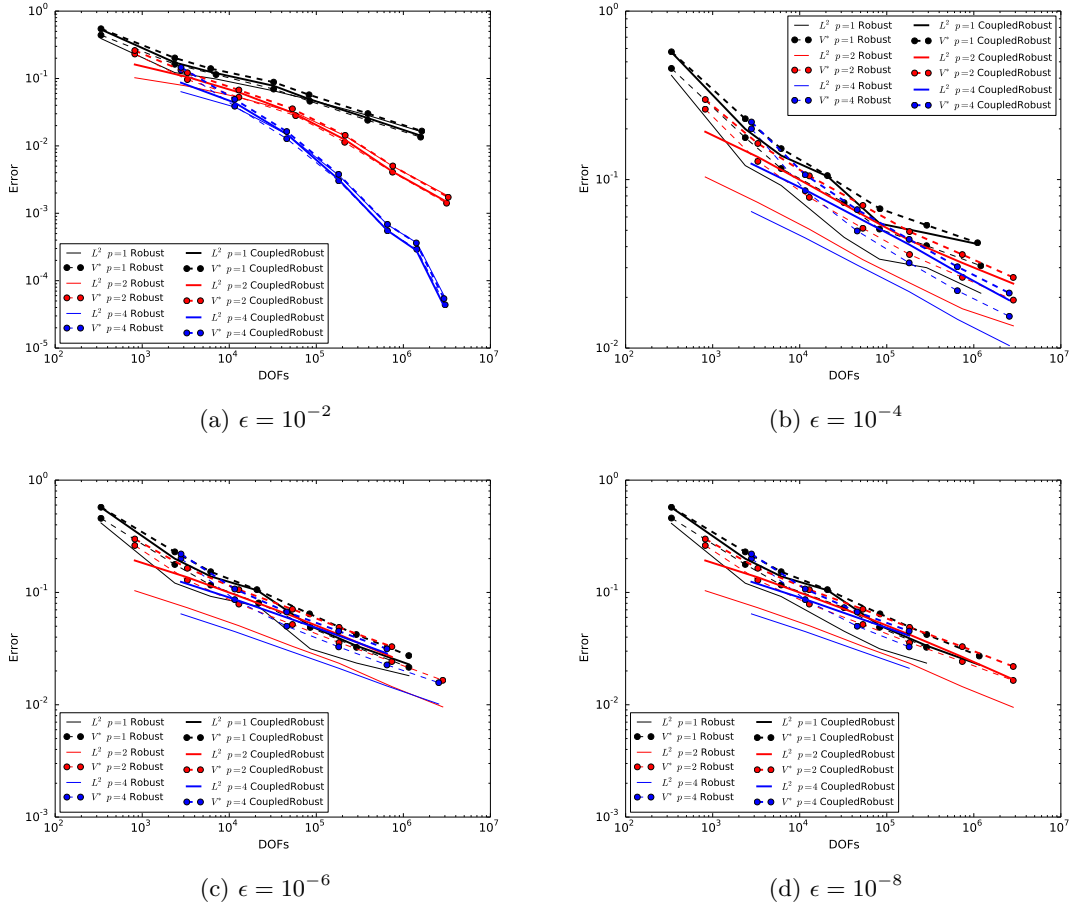
(a) $\epsilon = 10^{-2}$ (b) $\epsilon = 10^{-4}$

(c) $\epsilon = 10^{-6}$ (d) $\epsilon = 10^{-8}$

Figure 5: Convergence to analytical solution

transient convection-diffusion, though neither one guarantees robust control over $\boldsymbol{\sigma}$ as we had with their steady analogs.

References

[1] L. Demkowicz, J. Gopalakrishnan, A class of discontinuous Petrov-Galerkin methods. Part II: Optimal test functions, Numer. Meth. Part. D. E.

[2] L. Demkowicz, J. Gopalakrishnan, Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations (eds. X. Feng, O. Karakashian, Y. Xing), Vol. 157, IMA Volumes in Mathematics and its Applications, 2014, Ch. An Overview of the DPG Method, pp. 149–180.

[3] W. Dahmen, C. Huang, C. Schwab, G. Welper, Adaptive Petrov-Galerkin methods for first order transport equations, SIAM J. Numer. Anal. 50 (5) (2012) 2420–2445. arXiv:http://epubs.siam.org/doi/pdf/10.1137/110823158, doi:10.1137/110823158.

[4] L. Demkowicz, J. Gopalakrishnan, Discontinuous Petrov-Galerkin (DPG) method, Tech. Rep. 15-20, ICES (Dec. 2015).

[5] L. Demkowicz, N. Heuer, Robust DPG method for convection-dominated diffusion problems, SIAM J. Numer. Anal. 51 (5) (2013) 1514–2537.

[6] J. Chan, N. Heuer, T. Bui-Thanh, L. Demkowicz, A robust DPG method for convection-dominated diffusion problems II: Adjoint boundary conditions and mesh-dependent test norms, Comp. Math. Appl. 67 (4) (2014) 771 – 795, high-order Finite Element Approximation for Partial Differential Equations.

[7] Z. Kaczkowski, The method of finite space-time elements in dynamics of structures, J. Tech. Phys. 16 (1) (1975) 69 – 84.

[8] J. Argyris, D. Scharpf, Finite elements in time and space, Nucl. Eng. Des. 10 (4) (1969) 456 – 464. doi:http://dx.doi.org/10.1016/0029-5493(69)90081-8.

[9] I. Fried, Finite-element analysis of time-dependent phenomena, AIAA J. 7 (6) (1969) 1170–1173. doi:10.2514/3.5299.

[10] J. Oden, A general theory of finite elements. II. Applications, Int. J. Numer. Meth. Eng. 1 (3) (1969) 247–259. doi:10.1002/nme.1620010304.

[11] J. van der Vegt, H. van der Ven, Space-time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows: I. general formulation, J. Comp. Phys. 182 (2) (2002) 546 – 585.

[12] C. Klaij, J. van der Vegt, H. van der Ven, Spacetime discontinuous Galerkin method for the compressible NavierStokes equations, J. Comp. Phys. 217 (2) (2006) 589 – 611. doi:http://dx.doi.org/10.1016/j.jcp.2006.01.018.

[13] S. Rhebergen, B. Cockburn, J. Van Der Vegt, A space-time discontinuous Galerkin method for the incompressible Navier-Stokes equations, J. Comput. Phys. 233 (2013) 339–358. doi:10.1016/j.jcp.2012.08.052.

[14] S. Rhebergen, B. Cockburn, A spacetime hybridizable discontinuous Galerkin method for incompressible flows on deforming domains, J. Comp. Phys. 231 (11) (2012) 4185 – 4204. doi:http://dx.doi.org/10.1016/j.jcp.2012.02.011.

[15] T. Tezduyar, M. Behr, J. Liou, A new strategy for finite element computations involving moving boundaries and interfaces – The deforming-spatial-domain/space-time procedure: I. The concept and the preliminary numerical tests, Comput. Methods in Appl. Mech. Eng. 94 (3) (1992) 339 – 351.

[16] S. Aliabadi, T. Tezduyar, Space-time finite element computation of compressible flows involving moving boundaries and interfaces, Comput. Methods in Appl. Mech. Eng. 107 (12) (1993) 209 – 223.

[17] T. Hughes, J. Stewart, A space-time formulation for multiscale phenomena, J. Comput. Appl. Math. 74 (12) (1996) 217 – 229. doi:http://dx.doi.org/10.1016/0377-0427(96)00025-8.

[18] L. Demkowicz, J. Gopalakrishnan, A. Niemi, A class of discontinuous Petrov-Galerkin methods. Part III: Adaptivity, Appl. Numer. Math. 62 (4) (2012) 396–427. doi:10.1016/j.apnum.2011.09.002.

[19] H. G. Roos, M. Stynes, L. Tobiska, Robust Numerical Methods for Singularly Perturbed Differential Equations, 2nd Edition, Vol. 24 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 2008.

[20] C. Carstensen, L. D. J. Gopalakrishnan, Breaking spaces and forms for the DPG method and applications including Maxwell equations, Tech. Rep. 15-18, ICES (2015).

[21] L. Demkowicz, Various variational formulations and closed range theorem, Tech. Rep. 15-03, ICES (Jan. 2015).

[22] J. Gopalakrishnan, W. Qiu, An analysis of the practical DPG method, Math. Comp. 83 (286) (2014) 537–552.

[23] J. Chan, A DPG method for convection-diffusion problems, Ph.D. thesis, University of Texas at Austin (2013).

[24] A. H. Niemi, N. O. Collier, V. M. Calo, Automatically stable discontinuous Petrov-Galerkin methods for stationary transport problems: Quasi-optimal test space norm, Comput. Math. Appl. 66 (10) (2013) 2096–2113.

[25] A. H. Niemi, N. O. Collier, V. M. Calo, Discontinuous Petrov-Galerkin method based on the optimal test space norm for steady transport problems in one space dimension, Journal of Computational Science 4 (3) (2013) 157–163.