ICES REPORT 15-20

October 2015

Discontinuous Petrov-Galerkin (DPG) Method

by

Leszek Demkowicz and Jay Gopalakrishnan

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

Reference: Leszek Demkowicz and Jay Gopalakrishnan, "Discontinuous Petrov-Galerkin (DPG) Method," ICES REPORT 15-20, The Institute for Computational Engineering and Sciences, The University of Texas at Austin, October 2015.

Discontinuous Petrov-Galerkin (DPG) Method

Leszek Demkowicz[a] and Jay Gopalakrishnan[b]
[a]The University of Texas at Austin, Austin, TX 78712, USA,
[b]Portland State University, Portland, OR 97207, USA

Abstract

The article reviews fundamentals of Discontinuous Petrov-Galerkin (DPG) Method with Optimal Test Functions. The main idea admits three different interpretations: a Petrov-Galerkin method with (optimal) test functions that realize the supremum in the inf-sup condition, a minimum-residual method with residual measured in a dual norm, and a mixed formulation where one solves simultaneously for the Riesz representation of the residual. The methodology can be applied to any well-posed variational problem but it is especially effective in context of discontinuous (broken) test spaces. We discuss how one can "break" test functions in any variational formulation, and use the convection-dominated diffusion model problem to illustrate challenges related to the choice of an appropriate test norm.

Keywords: discontinuous test functions, finite elements, adaptivity, Petrov-Galerkin method

1 Introduction. A Model Problem

The name *Discontinuous Petrov Galerkin (DPG) Method* was introduced in a series of papers by Bottasso, Causin, Micheletti and Sacco, see e.g. [5, 15]. The method was built on a variational formulation in which all derivatives are passed to test functions. We named it later the *ultraweak variational formulation* [1] and provided for it a precise functional setting. As the name suggests it, the method uses a Petrov-Galerkin scheme, i.e. the trial and test basis functions are different. The ultraweak formulation uses a non-symmetric functional setting (trial and test spaces are different) and the Petrov-Galerkin scheme is a must. In the work of the Italian colleagues, the word "discontinuous" referred to both trial and test functions. Both sets of functions were predefined a-priori. The main idea behind our version of the (ideal) DPG method [23, 25] consisted in computing the test functions on the fly. More precisely, for each trial function u_h, we compute the corresponding *optimal test function* $v_h = Tu_h$ defined by a *trial-to-test operator* T. In practice, operator T and the optimal test functions have to be approximated, and we talk then about a *practical DPG method*. Critical for the practicality of the method is the use of *discontinuous* or *broken test spaces* which enables the computation of optimal test functions (and their approximation) on the element level. It took us a while to understand that the DPG methodology can be applied to any well-posed variational formulation [14] including those that use continuous trial functions. Consequently, the word "discontinuous" in our DPG method *refers only to test functions*. The ideal DPG method turns out to be equivalent to a *minimum residual method* in which the residual is measured in the dual test norm. In case of the L^2 test space, the DPG method reduces to classical least squares, see e.g. [12, 4] so, it can be also understood as a *generalized least squares method*. The idea of minimizing residuals in a dual norm is also not new, see [6]. Finally, the ideal DPG method is also equivalent to a mixed formulation introduced by Cohen, Dahmen, Schwab and Welper [20, 19] where one simultaneously solves for the approximate solution and the Riesz representation of the residual (we call it the *error representation function*). Element contributions to the norm of the error representation function serve as element error indicators, and provide a basis for adaptivity. We shall discuss these relations in detail.

[1] The name has already been used in the same spirit by J.L. Lions and, more recently, by Cessenat and Despres.

We conclude the opening paragraph by comparing the DPG method (with optimal test functions) with a standard Petrov-Galerkin (PG) approach. Both methodologies use the standard FE technology and differ only in element computations. In the standard PG method, given bilinear and linear forms defining the problem, and trial and test shape functions, we integrate for the corresponding element stiffness matrix and load vector, and return them to the solver. In the DPG method, we enter the element routine only with trial shape functions, but we must also be given a concrete *test inner product* that dictates the dual norm in which the residual is minimized, and provides a basis for computing the optimal test functions. Clearly, for different test inner products, we get different methods and, for that reason, one should talk rather about the *DPG methodology* than a DPG method. Selection of a starting variational formulation (functional setting) and a proper test norm are instrumental in building a DPG method for difficult singularly perturbed problems.

Model problem. The DPG method can be applied to both linear and nonlinear problems. This presentation will focus on linear problems for which a fairly complete theory has now been developed, and we will make only a few informal comments about applications to compressible and incompressible Navier-Stokes equations. To make the discussion more concrete, we will focus on a model diffusion-convection-reaction problem. Given a bounded domain $\Omega \subset \mathbb{R}^N, N = 1, 2, 3$, we want to determine $u = u(x), x \in \Omega$ that satisfies:

$$\left\{ \begin{array}{rll} -\text{div}(a\boldsymbol{\nabla} u - bu) + cu &= f & \text{in } \Omega \\ u &= u_0 & \text{on } \Gamma_u \\ (a\boldsymbol{\nabla} - bu) \cdot n &= \sigma_0 & \text{on } \Gamma_\sigma \end{array} \right. \tag{1}$$

where $a = a_{ij}$ is the diffusion matrix, $b = b_i$ is the advection vector, r is the reaction coefficient, and f is a given source function. Boundary $\Gamma = \partial\Omega$ has been split into two disjoint parts Γ_u, Γ_σ on which the two boundary conditions are satisfied with given data u_0, σ_0. Finally, n stands for the outward normal unit vector, and $\sigma \cdot n$ denotes the dot product of vectors σ and n. Other boundary conditions are possible.

Introducing explicitly flux $\sigma := a\boldsymbol{\nabla} u - bu$, we can rewrite the problem as a system of first order equations:

$$\left\{ \begin{array}{rll} \alpha\sigma - \boldsymbol{\nabla} u + \beta u &= 0 & \text{in } \Omega \\ -\text{div}\sigma + cu &= f & \text{in } \Omega \\ u &= u_0 & \text{on } \Gamma_u \\ \sigma \cdot n &= \sigma_0 & \text{on } \Gamma_\sigma \end{array} \right. \tag{2}$$

where a is assumed to be invertible, and $\alpha := a^{-1}, \beta := a^{-1}b$.

2 Various Variational Formulations

We multiply the constitutive equation with a test function τ, the conservation equation with a test function v, and integrate over the domain Ω. Each of the two equations can then be *relaxed*, i.e. we integrate it by parts, moving derivatives to the test functions, and *we build in* the corresponding boundary condition. For instance, relaxation of the conservation equation leads to:

$$(\sigma, \boldsymbol{\nabla} v) + (cu, v) = (f, v) + \langle \sigma_0, v \rangle \quad v = 0 \text{ on } \Gamma_u$$

where (u, v) denotes the $L^2(\Omega)$ or $(L^2(\Omega)^N$ inner product, and $\langle \sigma_0, v \rangle$ stands for the duality pairing between $H^{1/2}(\Gamma)$ and and its dual $H^{-1/2}(\Gamma)$, generalizing the $L^2(\Gamma)$-inner product. Notice that the term involving the unknown normal flux $\sigma \cdot n$ on Γ_u has been eliminated by requesting the additional condition[1] on test function v.

Depending upon which choice we make, we obtain one of the following four variational formulations.

Trivial (strong) formulation:

$$\begin{cases} \sigma \in H(\text{div}, \Omega), & \sigma \cdot n &= \sigma_0 & \text{on } \Gamma_\sigma \\ u \in H^1(\Omega), & u &= u_0 & \text{on } \Gamma_u \\ (\alpha\sigma, \tau) - (\boldsymbol{\nabla} u, \tau) + (\beta u, \tau) & &= 0 & \tau \in L^2(\Omega)^N \\ -(\text{div}\sigma, v) + (cu, v) & &= (f, v) & v \in L^2(\Omega) \end{cases} \tag{3}$$

Mixed formulation I:

$$\begin{cases} \sigma \in H(\text{div}, \Omega), & \sigma \cdot n &= \sigma_0 & \text{on } \Gamma_\sigma \\ u \in L^2(\Omega) & & & \\ (\alpha\sigma, \tau) + (u, \text{div } \tau) + (\beta u, \tau) & &= 0 & \tau \in H(\text{div}, \Omega) \,:\, \tau \cdot n = 0 \text{ on } \Gamma_\sigma \\ -(\text{div}\sigma, v) + (cu, v) & &= (f, v) & v \in L^2(\Omega) \end{cases} \tag{4}$$

Mixed formulation II:

$$\begin{cases} \sigma \in L^2(\Omega)^N & & & \\ u \in H^1(\Omega), & u &= u_0 & \text{on } \Gamma_u \\ (\alpha\sigma, \tau) - (\boldsymbol{\nabla} u, \tau) + (\beta u, \tau) & &= 0 & \tau \in L^2(\Omega)^N \\ (\sigma, \boldsymbol{\nabla} v) + (cu, v) & &= (f, v) & v \in H^1(\Omega) \,:\, v = 0 \text{ on } \Gamma_u \end{cases} \tag{5}$$

Ultraweak formulation:

$$\begin{cases} \sigma \in L^2(\Omega)^N, u \in L^2(\Omega) & & & \\ (\alpha\sigma, \tau) + (u, \text{div } \tau) + (\beta u, \tau) & &= 0 & \tau \in H(\text{div}, \Omega) \,:\, \tau \cdot n = 0 \text{ on } \Gamma_\sigma \\ (\sigma, \boldsymbol{\nabla} v) + (cu, v) & &= (f, v) & v \in H^1(\Omega) \,:\, v = 0 \text{ on } \Gamma_u \end{cases} \tag{6}$$

The formulations involve two energy spaces: $H^1(\Omega)$ consists of all L^2 functions defined on Ω whose gradient (in the sense of distributions) is also a (vector-valued) function that is square integrable, $H(\text{div}, \Omega)$ consists of all square integrable vector-valued fields on Ω whose divergence (in the sense of distributions) is a function (i.e. a regular distribution) that is square integrable as well. Conforming discretizations of $H^1(\Omega)$ lead to standard continuous (Lagrange) elements, whereas conforming discretizations of $H(\text{div}, \Omega)$ lead to Raviart-Thomas elements with vector-valued functions

[1] Simply speaking, we do not test on Γ_u.

whose normal[1] components must be continuous across interelement boundaries. The boundary conditions are understood in the sense of traces. Notice that only the mixed formulations employ a symmetric functional setting, i.e. the trial and test spaces are the same and, therefore, are eligible for the standard (Bubnov-) Galerkin method.

The non-relaxed equations are equivalent to their strong form and, for both mixed formulations, can be used to eliminate one of the variables to arrive at two *reduced formulations*. Eliminating the flux from the second mixed formulation, we obtain the *classical variational formulation* expressed in terms of u alone:

$$\left\{\begin{array}{l} u \in H^1(\Omega),\ u = u_0 \text{ on } \Gamma_u \\ (a\boldsymbol{\nabla} u, \boldsymbol{\nabla} v) - (bu, \boldsymbol{\nabla} v) + (cu, v) = (f, v) + \langle \sigma_o, v\rangle \\ \qquad\qquad\qquad\qquad v \in H^1(\Omega)\ :\ v = 0 \text{ on } \Gamma_u\,. \end{array}\right. \tag{7}$$

Assuming $c \neq 0$, we can eliminate u from the first mixed formulation, and obtain the second reduced formulation expressed in terms of flux only,

$$\left\{\begin{array}{l} \sigma \in H(\text{div}, \Omega),\ \sigma \cdot n = \sigma_0 \text{ on } \Gamma_\sigma \\ (\alpha\sigma, \tau) + (c^{-1}\text{div}\ \sigma, \text{div}\ \tau) + (\beta c^{-1}\text{div}\ \sigma, \tau) = -(c^{-1}f, \text{div}\ \tau) - (\beta c^{-1} f, \tau) \\ \qquad\qquad\qquad\qquad \tau \in H(\text{div}, \Omega)\ :\ \tau \cdot n = 0 \text{ on } \Gamma_\sigma\,. \end{array}\right. \tag{8}$$

Of course, if c vanishes, the second reduced formulation is not possible.

Under standard assumptions on coefficients a, b, c, the bilinear forms for all all formulations satisfy simultaneously the inf-sup condition (in the corresponding variational setting) with related inf-sup constants [22]. Loosely speaking, all formulations are simultaneously well- or ill-posed. Different variational settings imply of course different regularity assumptions on the load data. Each of the formulations, including those with the non-symmetric functional setting, may serve as a starting point for the DPG method, resulting in the *convergence in a different norm*. One can also mix different formulations in different parts of the domain.

Finally, we mention that there is no need for making the additional assumptions on test functions. Elimination of the side conditions on the test functions (i.e. testing on the whole boundary) leads always to the introduction of additional unknowns on the boundary. For instance, if we test in the classical formulation (7) with all $v \in H^1(\Omega)$, we have to introduce an extra unknown - boundary flux $\hat{\sigma} \cdot n$, a trace of a flux $\sigma \in H(\text{div}, \Omega)$ on Γ_u. It is convenient to assume that $\hat{\sigma} \cdot n$ lives on the whole boundary, and it satisfies the flux boundary conditions on Γ_σ. We obtain,

$$\left\{\begin{array}{l} u \in H^1(\Omega),\ u = u_0 \text{ on } \Gamma_u \\ \hat{\sigma} \cdot n \in H^{-1/2}(\Gamma),\ \hat{\sigma} \cdot n = \sigma_0 \text{ on } \Gamma_\sigma \\ (a\boldsymbol{\nabla} u, \boldsymbol{\nabla} v) - (bu, \boldsymbol{\nabla} v) + (cu, v) - \langle \hat{\sigma} \cdot n, v\rangle = (f, v) \quad v \in H^1(\Omega)\,, \end{array}\right. \tag{9}$$

where $H^{-1/2}(\Gamma)$ is the trace of $H(\text{div}, \Omega)$. Consequently, the extra unknown is discretized with traces of Raviart-Thomas elements. Similar modifications can be made to all other variational formulations resulting in well-posed variational formulations. In classical Finite Element (FE) implementation we prefer to reduce the number of unknowns by imposing the extra boundary conditions on test functions. Once the problem is solved, we come back to the equation above, and test it subsequently with the

[1] In general, the tangential components are discontinuous.

remaining test functions (those that do not vanish on Γ_u) to solve for flux $\hat{\sigma}\cdot n$ on Γ_u. For regular solutions, the additional unknown coincides with the flux corresponding to the primary variable u, and the whole procedure is interpreted as a postprocessing scheme. This is not a must though, we can solve simultaneously for u and $\hat{\sigma}\cdot n$, if needed. The concept of solving for u and the flux simultaneously is critical for understanding the next section.

Comment: Introduction of the additional unknown σ converting the second order problem into the first order system is not unique and has been motivated here by the boundary condition. Also, multiplication of the first equation with the inverse $\alpha = a^{-1}$ is somehow arbitrary. In context of convection-dominated diffusion where $a = \epsilon I$, [9] advocate a multiplicative splitting of the diffusion coefficient and identifying $\sigma = \epsilon^{1/2}\boldsymbol{\nabla} u$ as the additional variable.

3 Breaking Test Spaces

In each of the discussed formulations, we can test with functions coming from larger, *broken test spaces* $H^1(\Omega_h)$ and $H(\mathrm{div},\Omega_h)$. Elements of the new energy spaces live in $H^1(K)$ and $H(\mathrm{div},K)$, for each element K in mesh $\mathcal{T}_h$ but, otherwise, satisfy no global conformity (continuity) conditions. Obviously, discretization of such spaces is much easier compared with the standard spaces that require the enforcement of the global continuity conditions. When testing with discontinuous test functions, we pay the same price as discussed above – we have to introduce additional unknowns that live now on the whole mesh skeleton $\Gamma_h = \bigcup_{K\in\mathcal{T}_h}\partial K$. For instance, "breaking" test functions in the classical formulation (7), we obtain,

$$\left\{\begin{array}{l} u\in H^1(\Omega),\ u=u_0 \text{ on } \Gamma_u \\ \hat{\sigma}\cdot n\in H^{-1/2}(\Gamma_h),\ \hat{\sigma}\cdot n=\sigma_0 \text{ on } \Gamma_\sigma \\ (a\boldsymbol{\nabla} u,\boldsymbol{\nabla} v)-(bu,\boldsymbol{\nabla} v)+(cu,v)-\langle\hat{\sigma}\cdot n,v\rangle=(f,v)\quad v\in H^1(\Omega_h)\,, \end{array}\right. \tag{10}$$

Notice the difference between (9) and (10): the new variable $\hat{\sigma}\cdot n$ is now defined on the whole mesh skeleton Γ_h, and the duality pairing is defined elementwise:

$$\langle\hat{\sigma}\cdot n,v\rangle := \sum_{K\in\mathcal{T}_h}\langle\sigma|_K\cdot n_K, v_K\rangle_{\partial K},\quad v=\{v_K\}_{K\in\mathcal{T}_h}\in H^1(\Omega_h) \tag{11}$$

where $\sigma\in H(\mathrm{div},\Omega)$ is an arbitrary field such that $\sigma\cdot n=\hat{\sigma}\cdot n_K$ on ∂K, with n_K denoting the outward normal unit vector on ∂K. The space of all such restrictions to Γ_h is denoted by $H^{-1/2}(\Gamma_h)$ and equipped with the quotient (minimum energy extension) norm:

$$\|\hat{\sigma}\cdot n\|_{H^{-1/2}(\Gamma_h)} := \inf_{\sigma}\|\sigma\|_{H(\mathrm{div},\Omega)} \tag{12}$$

where the infimum is taken over all fields $\sigma\in H(\mathrm{div},\Omega)$ discussed above.

In a similar way, we introduce space $H^{1/2}(\Gamma_h)$, the space of restrictions of functions from $H^1(\Omega)$ to mesh skeleton Γ_h,

$$\hat{u}\in H^{1/2}(\Gamma_h)\quad\overset{\text{def}}{\Leftrightarrow}\quad \exists\, u\in H^1(\Omega)\,:\,\hat{u}=u|_{\partial K},\ \forall K\in\mathcal{T}_h\,. \tag{13}$$

The space is again equipped with the minimum energy extension norm. Similarly to pairing (11), elements $\hat{u}\in H^{1/2}(\Gamma_h)$ pair with broken space $H(\mathrm{div},\Omega_h)$,

$$\langle\tau\cdot n,\hat{u}\rangle := \sum_{K\in\mathcal{T}_h}\langle\tau_K\cdot n_K,u\rangle_{\partial K},\quad \tau=\{\tau_K\}_{K\in\mathcal{T}_h}\in H(\mathrm{div},\Omega_h) \tag{14}$$

where $u \in H^1(\Omega)$ is an extension of $\hat{u}$ to Ω. For more discussion on spaces defined on the mesh skeleton, see [24, 31, 14].

"Breaking" test functions in the ultraweak formulation (6), we obtain a new formulation in the form:

$$\begin{cases} \sigma \in L^2(\Omega)^N, u \in L^2(\Omega) & & & \\ \hat{u} \in H^{1/2}(\Gamma_h), & \hat{u} & = u_0 & \text{on } \Gamma_u \\ \hat{\sigma} \cdot n \in H^{-1/2}(\Gamma_h), & \hat{\sigma} \cdot n & = \sigma_0 & \text{on } \Gamma_\sigma \\ (\alpha\sigma, \tau) + (u, \text{div}\, \tau) + (\beta u, \tau) - \langle \tau \cdot n, \hat{u} \rangle & & = 0 & \tau \in H(\text{div}, \Omega_h) \\ (\sigma, \boldsymbol{\nabla} v) + (cu, v) - \langle \hat{\sigma} \cdot n, v \rangle & & = (f, v) & v \in H^1(\Omega_h) \end{cases} \tag{15}$$

Notice that in process of "breaking" test functions we also eliminate boundary conditions on test functions.

In an analogous way, we can "break" test functions in the remaining variational formulations. Only in the case of the strong formulation, the formulation remains unchanged as the L^2 spaces do not present any global conformity assumptions. The main message of the theory presented in [14] is now as follows. Let the original variational formulation be represented in the abstract form,

$$\begin{cases} u \in U \\ b(u, v) = l(v) \quad v \in V \end{cases} \tag{16}$$

where U, V are trial and test spaces, and bilinear (sesquilinear) form $b(u, v)$ and linear (antilinear) form $l(v)$ represent a particular variational formulation. In general, of course, both solution u and test functions v represent *group variables*, e.g. for the discussed ultraweak formulation $u \overset{'}{=} (\sigma, u)$ and $v \overset{'}{=} (\tau, v)$. We make now the following assumptions.

A1: The test norm is of the form:

$$\|v\|_V^2 := \|Cv\|^2 + \|v\|^2 \tag{17}$$

where C is a differential operator of first order, and $\|\cdot\|$ represents L^2-norm. The test space can be replaced with its broken counterpart $V(\Omega_h)$ with the test norm[1] (17) and the corresponding inner product extending naturally to the broken test space,

$$\|v\|_{V(\Omega_h)}^2 := \sum_{K \in \mathcal{T}_h} \left\{ \|Cv_K\|_{L^2(K)}^2 + \|v_K\|_{L^2(K)}^2 \right\}, \quad v = \{v_K\}_{K \in \mathcal{T}_h} \tag{18}$$

with a corresponding natural extension of the bilinear form [2] $b_h(u, v)$, $u \in U, v \in V(\Omega_h)$ that remains continuous,

$$|b_h(u, v)| \leq M \|u\|_U \, \|v\|_{V(\Omega_h)} \,. \tag{19}$$

A2: There exists a corresponding space of Lagrange multipliers $\hat{U}$ defined on the skeleton Γ_h with a pairing $\langle \hat{u}, v \rangle$ satisfying the following two properties[3]:

$$\begin{gathered} v \in V \subset V(\Omega_h) \quad \Leftrightarrow \quad \langle \hat{u}, v \rangle = 0 \quad \forall \hat{u} \in \hat{U}, \quad v \in V(\Omega_h) \\ \langle \hat{u}, v \rangle = 0 \quad \forall v \in V(\Omega_h) \quad \Rightarrow \quad \hat{u} = 0 \end{gathered} \tag{20}$$

[1] We call it a *localizable test norm.*

[2] In practice the differential operators defining $b_h(u, v)$ are defined now element-wise.

[3] We say that that the pairing is *definite*.

A3: The bilinear form in (16) satisfies the inf-sup condition:

$$\sup_{v\in V}\frac{|b(u,,v)|}{\|v\|_V}\geq\gamma\|u\|_U^2\,. \tag{21}$$

Then the corresponding "broken" variational formulation:

$$\begin{cases} u\in U,\ \hat{u}\in\hat{U} \\ \underbrace{b_h(u,v)+\langle\hat{u},v\rangle}_{=:b_{\text{mod}}((u,\hat{u}),v)}=l(v) \quad v\in V(\Omega_h) \end{cases} \tag{22}$$

is well-posed as well, with a new, *mesh-independent* inf-sup constant, and *a mesh-dependent* norm for the Lagrange multiplier,

$$\|\hat{u}\|_{\hat{U}}:=\sup_{v\in V(\Omega_h)}\frac{|\langle\hat{u},v\rangle|}{\|v\|_{V(\Omega_h)}}\,. \tag{23}$$

The result is a straightforward consequence of the assumptions we have made. Indeed, in order to control u, we need only to restrict ourselves to the conforming test functions,

$$\gamma\|u\|_U\leq\sup_{v\in V}\frac{|b(u,v)|}{\|v\|_V}\leq\sup_{v\in V(\Omega_h)}\frac{|b_h(u,v)|}{\|v\|_{V(\Omega_h)}}=\sup_{v\in V(\Omega_h)}\frac{|b_{\text{mod}}(u,0)|}{\|v\|_{V(\Omega_h)}}$$

With u being controlled, we move $b_h(u,v)$ to the right-hand side,

$$\langle\hat{u},v\rangle=b_{\text{mod}}((u,\hat{u}),v)-b_h(u,v)$$

and use (19) and the bound for $\|u\|_U$ to obtain a bound of the Lagrange multiplier[1]

$$\|\hat{u}\|_{\hat{U}}\leq(1+\frac{M}{\gamma})\sup_{v\in V(\Omega_h)}\frac{|b_{\text{mod}}(u,\hat{u})|}{\|v\|_{V(\Omega_h)}}\,.$$

The dual norm for the Lagrange multipliers can be reinterpreted in terms of solution to local Neumann problems. First of all, we have the fundamental algebraic property for the broken test spaces,

$$\left(\sup_{v\in V(\Omega_h)}\frac{|\langle\hat{u},v\rangle|}{\|v\|_{\Omega(h)}}\right)^2=\sum_{K\in\mathcal{T}_h}\left(\sup_{v_K\in V(K)}\frac{|\langle\hat{u},v_K\rangle|}{\|v\|_{V(K)}}\right)^2 \tag{24}$$

Secondly, the Riesz Representation Theorem implies that the element supremum equals the norm of the local Riesz representation of the pairing,

$$\sup_{v_K\in V(K)}\frac{\langle\hat{u},v_K\rangle}{\|v\|_{V(K)}}=\|\hat{U}\|_{V(K)} \tag{25}$$

where $\hat{U}_K$ is the solution of the local variational problem,

$$\begin{cases} \hat{U}_K\in V(K) \\ (C\hat{U}_K,C\delta v)+(\hat{U}_K,\delta v)=\langle\hat{u},\delta v\rangle \quad \delta v\in V(K)\,. \end{cases} \tag{26}$$

[1] Brezzi's argument.

Variational problem (26) is equivalent to a local Neumann problem for operator $C^*C + I$, and it can be shown to be equivalent to a *local Dirichlet problem* for operator $CC^* + I$, see [14]. Consequently, the norm for the (natural) norm for the Lagrange multiplier can be interpreted in terms of minimum energy extension norms $(\|C^*v\|^2 + \|v\|^2)^{1/2}$. This is consistent with the generic norms that we have introduced for traces and fluxes. If we use the standard H^1-test norm in (10), fluxes $\hat{\sigma}\cdot n \in H^{-1/2}(\Gamma_h)$ are measured using the minimum energy extension norm (12), i.e.

$$\|\hat{\sigma}\cdot n\|^2 = \sum_{K\in\mathcal{T}_h} \{\|\text{div } \sigma\|^2 + \|\sigma\|^2\}$$

where σ solve the local Dirichlet problems,

$$\begin{cases} \sigma \in H(\text{div}, K),\ \sigma\cdot n = \hat{\sigma}\cdot n & \text{on } \partial K \\ -\boldsymbol{\nabla}(\text{div } \sigma) + \sigma = 0 & \text{in } K \end{cases}$$

Similarly, if we use the standard $H(\text{div})$-norm for the τ component in (6), the norm for the corresponding Lagrange multiplier, the trace $\hat{u}$ is given by:

$$\|\hat{u}\cdot n\|^2 = \sum_{K\in\mathcal{T}_h} \{\|\boldsymbol{\nabla} u\|^2 + \|u\|^2\}$$

where u solve the local Dirichlet problems,

$$\begin{cases} u \in H^1(K),\ u = \hat{u} & \text{on } \partial K \\ -\text{div } (\boldsymbol{\nabla} u) + u = 0 & \text{in } K \end{cases}$$

If, however, we choose to work with different test norms, the natural norm for the Lagrange multipliers (fluxes and traces) will change accordingly. Notice the philosophical aspect of our discussion: the natural norm for the additional unknowns resulting from breaking test functions is not derived from the trial but rather test norm.

4 Three Hats of the Ideal DPG Method

The methodology discussed in this section applies to any abstract variational problem (16) but, in practice, will be applied to problems (22) employing a broken test space. In order to simplify the exposition, we will use the terminology of problem (16). Obviously, replacing u with group variable $(u, \hat{u})$ and bilinear form $b(u, v)$ with the modified form $b_{\text{mod}}((u, \hat{u}), v)$, we can apply all results of the forthcoming discussion to the modified problem as well.

4.1 Hat 1: a Petrov-Galerkin method with optimal test functions

Given a variational problem (16), we construct its Petrov-Galerkin discretization by selecting a trial space $U_h \subset U$ and a test space $V_h \subset V$, of equal dimension $\dim U_h = \dim V_h < \infty$ and solve the algebraic problem:

$$\begin{cases} u_h \in U_h \\ b(u_h, v_h) = l(v_h) \quad v_h \in V_h\,. \end{cases} \tag{27}$$

The trouble with the discrete problem is that the well-posedness of the continuous problem *does not* imply that the discrete problem is well posed as well. More precisely, the continuous inf-sup condition does not imply its discrete version:

$$\sup_{v\in V}\frac{|b(u_h,v)|}{\|v\|_V}\geq\gamma\|u_h\|_U \quad \not\Longrightarrow \quad \sup_{v_h\in V_h}\frac{|b(u_h,v_h)|}{\|v_h\|_V}\geq\gamma\|u_h\|_U\,. \tag{28}$$

The reason is obvious: the supremum on the continuous level is taken with respect to *all non-zero test functions*, whereas on the discrete level only over the subspace V_h. The situation changes, however, if we do not work with arbitrary selected test functions but only with those that *realize the supremum*. Let $B:U\to V'$ be the operator corresponding to bilinear form $b(u,v)$, i.e. $\langle Bu,v\rangle_{V'\times V}=b(u,v)$, $u\in U, v\in V$, and let $R_V: V\to V'$ denote the Riesz operator corresponding to test inner product $(v,\delta v)_V, \|v\|_V^2=(v,v)_V$. We claim that the supremum in the inf-sup condition *is attained* for $v=R_V^{-1}u_h$. Indeed,

$$\begin{aligned}
\sup_{v\in V}\frac{|b(u_h,v)|}{\|v\|_V} &= \|Bu_h\|_{V'} && \text{(definitions of } B \text{ and dual norm)}\\
&= \|R_V^{-1}Bu_h\|_V && \text{(Riesz operator is an isometry)}\\
&= \frac{(R_V^{-1}Bu_h,R_V^{-1}Bu_h)_V}{\|R_V^{-1}Bu_h\|_V} && (\|v\|_V^2=(v,v)_V)\\
&= \frac{\langle Bu_h,R_V^{-1}Bu_h\rangle}{\|R_V^{-1}Bu_h\|_V} && \text{(definition of Riesz operator)}\\
&= \frac{b(u_h,R_V^{-1}Bu_h)}{\|R_V^{-1}Bu_h\|_V} && \text{(definition of operator } B)
\end{aligned}$$

Consequently, if we introduce *trial to test operator* $T:U\to V$, $T=R_V^{-1}B$, and select the *optimal test space* as $V_h=TU_h$, the problem with discrete stability is solved once and forever. If we test with optimal test functions, the discrete inf-sup constant is always at least as good as the continuous one, $\gamma_h\geq\gamma$. Notice. however, that the inversion of the Riesz operator (determination of the optimal test functions) requires solution of an auxiliary variational problem,

$$\begin{cases} v=Tu_h\in V \\ (v,\delta v)_V=b(u_h,\delta v) \quad \delta v\in V\,. \end{cases} \tag{29}$$

4.2 Hat 2: a minimum residual method

Let us replace the original trial norm $\|u\|_U$ with a special "energy" norm,

$$\|u\|_E=\|Bu\|_{V'}=\|R_V^{-1}Bu\|_V\,. \tag{30}$$

Obviously, with a redefined norm on u, the corresponding continuity and inf-sup constants will change. It takes a second to see that continuity constant M equals one [1]. Now, with optimal test functions,

$$\sup_{v_h\in V_h}\frac{\langle Bu_h,v_h\rangle}{\|v_h\|_V}=\sup_{v\in V}\frac{\langle Bu_h,v\rangle}{\|v\|_V}=\|u_h\|_E\,,$$

[1] With spaces U,V equipped with norms $\|u\|_E$, $\|v\|_V$, bilinear form $b(u,v)$ becomes a *a duality pairing* , comp. [10].

which proves that $\gamma_h \geq 1$ as well. Consequently, Babuška's Theorem [1] implies that

$$\|u - u_h\|_U \leq \underbrace{\frac{M}{\gamma_h}}_{\leq 1} \inf_{w_h \in U_h} \|u - w_h\|_U \,.$$

In other words, the Petrov-Galerkin method delivers an orthogonal projection in the energy norm. But the energy norm is equal to the residual,

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

and, therefore, DPG with optimal test functions is a minimum residual method. Critical is the fact that the residual is measured in the dual norm, consistently with the variational formulation of the problem. It may come as a surprise that *the minimum residual method is the most stable Petrov Galerkin scheme*. In the case of the trivial (strong) variational formulation, the test space is the L^2-space and (with the usual identification of the dual space with L^2 itself) the Riesz operator reduces to identity. With the residual measured in the L^2-norm, the DPG method reduces to classical least squares.

4.3 Hat 3: a mixed method

We can start now from the other end, i.e. the minimum residual method,

$$u_h = \text{arg min}_{w_h \in U_h} J(w_h) \qquad J(w_h) := \frac{1}{2}\|Bw_h - l\|^2_{V'} = \frac{1}{2}\|R_V^{-1}(Bw_h - l)\|^2_V$$

where, again, we use the fact that the Riesz map R_V is an isometry. The minimized functional is a simple quadratic functional, and the minimization problem is equivalent to vanishing of its Gâteaux derivative:

$$(R_V^{-1}(Bu_h - l), R_V^{-1}Bw_h) = 0 \quad \forall w_h \in U_h \,. \tag{31}$$

We can eliminate one of the Riesz operators above by replacing the test inner product with duality pairing in $V' \times V$. If we identify $v_h = R_V^{-1}Bw_h$ as the optimal test function corresponding to w_h and eliminate the first Riesz operator, we recover the Petrov-Gelerkin scheme with optimal test functions,

$$\langle Bu_h, v_h \rangle = \langle l, v_h \rangle \quad v_h := R_V^{-1}Bw_h,\ w_h \in U_h \,.$$

Alternatively, if we identify $\psi := R_V^{-1}(Bu_h - l)$ as a new unknown, we can take the second Riesz operator out to obtain

$$\langle Bw_h, \psi \rangle = 0 \quad w_h \in U_h \,.$$

Rewriting definition of ψ and the equation above in the variational form, we obtain a mixed problem,

$$\begin{cases} \psi \in V,\ u_h \in U_h & & \\ (\psi, v)_V - b(u_h, v) & = -l(v) & v \in V \\ b(w_h, \psi) & = 0 & w_h \in U_h \,. \end{cases} \tag{32}$$

We called ψ *the error representation function* although a better and more informative name would be simply *the Riesz representation of residual*. The mixed problem is equivalent to a constrained minimization problem where one minimizes the functional:

$$\frac{1}{2}\|\psi\|^2_V + l(\psi) \,, \tag{33}$$

over all ψ that satisfy the constraint: $b(w_h, \psi) = 0\,, w_h \in U_h$. It is not a typical mixed problem as it involves an infinite dimensional space V, and a finite dimensional space u_h. The norm $\|\psi\|_V$ equals the residual. For a broken test space,

$$\|\psi\|^2_{V(\Omega_h)} = \sum_{K \in \mathcal{T}_h} \|\psi_K\|^2_{V(K)} \quad \psi = \{\psi_K\}_{K \in \mathcal{T}_h}\,,$$

and the element contributions $\|\psi_K\|^2_{V(K)}$ are perfect *a-posteriori error indicators*. As least squares, DPG method comes with a built-in a-posteriori error estimate.

5 A Practical DPG Method

The key point in applying the concept of optimal test functions in practice is the use of broken test spaces. We call the test norm (18) *localizable* as the norm (squared) equals sum of norms (squared) defined over individual element test spaces. With a broken test space and localizable test norm, the inversion of Riesz operator in (29) is done *elementwise*, can be done in parallel, and it does not contribute to global computations. Still, except for very simple problems like convection [23], the element-wise inversion of the Riesz operator can be done only approximately. In all of our work we have pursued the idea of an *enriched test space* [1]. If trial space is, loosely speaking, discretized with elements of order p, we introduce a finite-dimensional *enriched* subspace $V^r \subset V$ corresponding to elements of order $r > p$. Typically, $\Delta p := r - p = 1, 2, 3$. Dimension of the enriched space V^r is *larger* than dimension of the trial space, dim $V^r >>$ dim U_h. The actual optimal test space V_h is then determined by approximating variational problem (29) with the standard Galerkin method, and the enriched space V^r replacing the infinite-dimensional space V,

$$\begin{cases} v^r = T^r u_h \in V^r \subset V \\ (v^r, \delta v)_V = b(u_h, \delta v) \quad \delta v \in V^r\,. \end{cases} \tag{34}$$

Operator T^r is called the *approximate trial-to-test operator*, and $T^r U_h$ is the approximate optimal test space. Replacing test space V with the enriched subspace V^r in the definition of the dual norm and the mixed problem, we realize again that the three formulations of the *practical DPG method* are fully equivalent.

Obviously, replacing the optimal test functions with their approximation or, in another words, approximating the Riesz operator with the enriched space, we loose the optimal properties of the method. The question is: how much? The question was addressed in [29] by introducing the concept of *a Fortin operator* $F^r : V \to V^r$ that satisfies two properties:

$$\begin{aligned} b(w_h, v - F^r v) &= 0 \qquad && \text{(orthogonality)} \\ \|F^r v\|_V \leq C_r \|v\|_V \quad v \in V \qquad && \text{(continuity)} \end{aligned} \tag{35}$$

Continuity constant C_r helps to quantify how much stability we loose by replacing the optimal test functions with their approximations. Let $v = Tu_h$ be the exact optimal test function corresponding to

[1] Broersen and Stevenson [8] use the name of a *search space*.

u_h, and $v^r = T^r u_h$ its approximate counterpart. We have,

$$\begin{aligned}
\frac{|b(u_h, v^r)|}{\|v^r\|} &= \sup_{w \in V^r} \frac{|b(u_h, w)|}{\|w\|} && \text{(def of approximate optimal test functions)} \\
&\geq \frac{|b(u_h, F^r v)|}{\|F^r v\|} && \text{(def of supremum)} \\
&= \frac{|b(u_h, v)|}{\|v\|} \frac{\|v\|}{\|F^r v\|} && \text{(orthogonality property of } F^r\text{)} \\
&\geq \frac{\gamma}{C_r} \|u_h\| && \text{(continuity property of } F^r\text{)}
\end{aligned} \tag{36}$$

Consequently, Babuška's Theorem [1] implies the a-priori error estimate:

$$\|u - u_h\|_U \leq \frac{C_r M}{\gamma} \inf_{w_h \in U_h} \|u - w_h\|_U \,. \tag{37}$$

A family of commuting Fortin operators for Nedéléc simplices of the first type and the exact sequence test spaces has been constructed in [29, 14]. A similar simple argument can be used to assess the damage to the a-posteriori residual error estimate in terms of the Fortin constant C_r, see [13].

The enriched space is defined a-priori by specifying the order increment Δp. In most of our computations we have used $\Delta p = 2$. A common sanity check is to rerun the code with a higher Δp and compare the results. Implicit in the philosophy of the enriched space is the assumption that both the optimal test functions and the error representation functions ψ are rather regular, and can be sufficiently well approximated with the simple p-enrichment scheme. The assumption is satisfied for standard test inner products corresponding to $H^1, H(\text{curl})$ and $H(\text{div})$-spaces and regular load data. It fails to be satisfied for singular perturbation problems and test norms that inherit the (small) perturbation parameter. We will return to this issue momentarily.

Coding the DPG method. We conclude this section with a short comment on how we code the practical DPG method in context of the modified variational formulation (22). We shall use the mixed problem formalism (32) to explain it. The mixed formulation for the "broken problem" reads as follows.

$$\left\{ \begin{array}{lll}
\psi^r \in V^r,\, u_h \in U_h,\, \hat{u}_h \in \hat{U}_h & & \\
(\psi^r, v)_V - b(u_h, v) - \langle \hat{u}_h, v \rangle & = l(v) & v \in V^r \\
b(w_h, \psi) & = 0 & w_h \in U_h \\
\langle \hat{w}_h, \psi \rangle & = 0 & \hat{w}_h \in \hat{U}_h
\end{array} \right. \tag{38}$$

or, in terms of matrices,

$$\left\{ \begin{array}{ll}
G\psi^r - B_1 u_h - B_2 \hat{u}_h & = -l \\
B_1^* \psi^r & = 0 \\
B_2^* \psi^r & = 0
\end{array} \right. \tag{39}$$

With a little abuse of notation, $\psi^r, u_h, \hat{u}_h$ above can be understood as vectors of degrees-of-freedom for the actual unknowns. G is a square $m \times m$ Gram matrix corresponding to discretization of test inner product with m shape functions spanning the enriched test space. B_1 is a rectangular stiffness matrix obtained from discretization of the original bilinear form with n trial shape functions and the m enriched space shape functions, and B_2 is also a rectangular $k \times m$ stiffness matrix coming from the discretization of pairing $\langle \hat{u}_h, v \rangle$ with k trial functions used to approximate $\hat{u}_h$ and the m

enriched space shape functions. Now comes the main point: *with the broken test spaces, the error representation function ψ can be condensed out element-wise*. Solving the first equation for ψ and substituting it into the remaining two equations, we obtain the DPG system,

$$\left(\begin{array}{cc} B_1^* G^{-1} B_1 & B_1^* G^{-1} B_2 \\ B_2^* G^{-1} B_1 & B_2^* G^{-1} B_2 \end{array} \right) \left(\begin{array}{c} u_h \\ \hat{u}_h \end{array} \right) = \left(\begin{array}{c} B_1^* G^{-1} l \\ B_2^* G^{-1} l \end{array} \right) . \tag{40}$$

Note that the static condensation of ψ is equivalent to the determination of approximate optimal test functions and computation of the corresponding DPG element matrices. The rest of the computations follows the standard FE technology. We communicate the element matrices to a solver, and solve for u_h and $\hat{u}_h$. In the backward substitution phase, we compute the element contributions to error representation function ψ and its norm that serves as the a-posteriori error estimate.

The abstract exposition of the subject translates now into concrete a-priori and a-posteriori error estimates for specific problems. For instance, for the broken version of the classical variational formulation (10), we have the a-priori error estimate:

$$\begin{aligned} &\left(\|u - u_h\|^2_{H^1(\Omega)} + \|(\sigma - \hat{\sigma}_h) \cdot n\|^2_{H^{-1/2}(\Gamma_h)} \right)^{1/2} \\ &\quad \leq C \left(\inf_{w_h} \|u - w_h\|^2_{H^1(\Omega)} + \inf_{\hat{\psi}_h} \|(\sigma - \hat{\psi}) \cdot n\|^2_{H^{-1/2}(\Gamma_h)} \right)^{1/2} \end{aligned} \tag{41}$$

where we have used the fact that, for sufficiently regular solution, $\hat{\sigma} \cdot n$ coincides simply with $\sigma \cdot n$. The best approximation error for flux $\sigma \cdot n$, measured in the minimum energy extension norm, can easily be estimated using standard interpolation error estimates. Indeed, let $\Pi_h^{\text{div}} \sigma$ be any well-defined interpolant of σ. Then $\sigma - \Pi_h^{\text{div}} \sigma$ provides an extension for $(\sigma - \Pi_h^{\text{div}} \sigma) \cdot n$ and the best approximation error of $\sigma \cdot n$ can be estimated by the $H(\text{div})$-norm of the interpolation error. This suggests that the order of approximation for u_h and $\hat{\sigma}_h \cdot n$ should be dictated by the exact sequence logic; this way both contributions to the best approximation error are of the same order, and we get the standard error estimate:

$$\left(\|u - u_h\|^2_{H^1(\Omega)} + \|(\sigma - \hat{\sigma}_h) \cdot n\|^2_{H^{-1/2}(\Gamma_h)} \right)^{1/2} \leq C h^{\min\{p,r\}} \left(\|u\|^2_{H^{r+1}(\Omega)} + \|\sigma\|^2_{H^r(\text{div},\Omega)} \right) \tag{42}$$

where p is the order of elements. A code that supports a simultaneous discretization of all energy spaces forming the exact sequence: $H^1, H(\text{curl}), H(\text{div}), L^2$, can naturally be used to implement the DPG method. In the discussed case, we discretize u with H^1-conforming elements, and $\hat{\sigma} \cdot n$ with traces of $H(\text{div})$-conforming elements. An alternative is to build a hybrid code that will support the variables living on the mesh skeleton.

The stability constant includes the Fortin operator continuity constant C_r. For the operators constructed so far, C_r is independent of mesh size h but not the polynomial order p. If C_r were also independent of p, we could claim optimal hp error estimates as well.

In a similar way, we obtain the a-priori error estimate for the broken ultraweak formulation (15):

$$\begin{aligned} &\left(\|u - u_h\|^2_{L^2(\Omega)} + \|\sigma - \sigma_h\|^2_{L^2(\Omega)} + \|u - \hat{u}_h\|^2_{H^{1/2}(\Gamma_h)} + \|(\sigma - \hat{\sigma}) \cdot n\|^2_{H^{-1/2}(\Gamma_h)} \right)^{1/2} \\ &\quad \leq C h^{\min\{p,r\}} \left(\|u\|^2_{H^{r+1}(\Omega)} + \|\sigma\|^2_{H^r(\text{div},\Omega)} \right)^{1/2} \end{aligned} \tag{43}$$

To implement this version of the DPG method, we need $H^1, H(\text{div})$ and L^2-conforming elements. The L^2 elements are used to discretize u and components of σ, traces of H^1-elements are used to discretize "trace" $\hat{u}$, and traces of $H(\text{div})$ elements are needed for the discretization of "flux" $\hat{\sigma}$.

For numerous 2D and 3D numerical experiments for standard model problems including diffusion-convection-reaction, Maxwell, elasticity and Stokes problems, see [24, 7, 27, 26, 31, 13, 14].

Comment: In the standard Galerkin method, the mixed and corresponding reduced formulation deliver the same discrete scheme provided the spaces are selected in such a way that range of $\alpha\sigma_h$ contains the range of $-\boldsymbol{\nabla} u_h + \beta u_h$, a condition that is satisfied e.g. for $a = \epsilon I$, constant b and standard discrete spaces. Simply speaking, the discrete version of the constitutive equation implies then its satisfaction pointwise, similarly to the continuous level. However, this is not the case for the DPG method where the mixed formulation yields different discrete solution than the reduced formulation.

6 Singular Perturbation Problems

By now, the reader should realize that DPG is not a single method but rather a methodology. Dependent upon which variational formulation we choose, the error will be controlled in different norms. More than that, for a fixed functional setting, there are many equivalent test norms we can choose. For each test norm, the method will minimize the residual in the (approximate) dual test norm and deliver the best approximation error in the corresponding "energy norm". Ideally, we would like to start with a *trial norm* of our choice and search for a test norm such that the corresponding energy norm is identical or close to our preselected trial norm.

The choice of an appropriate variational formulation and the test norm emerge as crucial aspects of the DPG methodology in context of singular perturbation problems. In our discussion, we shall focus on the convection-dominated diffusion problem assuming in (1) constant diffusion $a = \epsilon I$, order one advection b, and no reaction, $c = 0$. A discretization is said to be *robust* if stability constant C present in estimates (42), (43) is independent of the perturbation parameter, in our case - the diffusion parameter ϵ. This implies that all involved constants: inf-sup constant γ, continuity constant M and the Fortin operator continuity constant C_r should be independent of ϵ. Can we derive in a systematic way a test norm for which all three conditions are satisfied ?

The moral of the discussion on broken test spaces in Section 3 is twofold: a/ in terms of solution u, the "broken formulation" inherits stability properties from the original variational formulation, b/ the error in the extra unknowns: traces and/or fluxes is controlled (robustly) in a natural norm derived from the test norm. In other words, we may try to derive an optimal test norm to control the original unknown but once we choose it, it implies automatically the norm for the Lagrange multiplier.

We can use for a starting point a very general (and abstract) fact implied by the Closed Range Theorem. Recall that any bilinear form $b(u, v)$ generates two operators $B : U \to V'$, and $B' : V \to U'$. For Hilbert spaces U, V, B' can be identified with the conjugate of operator B. If B' is injective then

$$\|v\|_{opt} := \sup_u \frac{|b(u, v)|}{\|u\|_U} = \|B'v\|_{U'} \tag{44}$$

defines a norm. We call it an *optimal test norm* because, with this test norm, bilinear form $b(u, v)$ becomes a *duality pairing* [36, 10], i.e.

$$\sup_v \frac{|b(u, v)|}{\|v\|_{opt}} = \|u\|_U \,. \tag{45}$$

In other words, the energy norm corresponding to such a test norm *coincides* with the norm in U that we began with. At this point, the ultraweak formulation emerges to be special as we can determine the optimal test norm explicitly,

$$\|v\|_{opt} = \|A^* v\| \,. \tag{46}$$

Unfortunately, in general, the optimal test norm is not *localizable*, i.e. we cannot use it in the broken formulation. We know, however, from the Closed Range Theorem for Closed Operators that operators A and its adjoint A^* are bounded below with the same constant $\alpha > 0$,

$$\|Au\| \geq \alpha \|u\|, \qquad \|A^* v\| \geq \alpha \|v\| \,. \tag{47}$$

Consequently, the ideal test norm and the *scaled adjoint graph norm*,

$$\|v\|_G^2 := \|A^* v\|^2 + c^2 \|v\|^2 \,, \tag{48}$$

are equivalent. Indeed, the optimal test norm is bounded by (48), and

$$\|A^* v\|^2 + c^2 \|v\|^2 \leq \|A^* v\|^2 + \frac{c^2}{\alpha^2} \|A^* v\|^2 = (1 + \frac{c^2}{\alpha^2}) \|A^* v\|^2 \,. \tag{49}$$

The duality argument implies that the energy norm corresponding to the adjoint graph norm is equivalent (with the same equivalence constants) to the original trial L^2-norm. The critical question is whether the equivalence constants are bounded uniformly in perturbation parameter ϵ. If the boundedness below constant is independent of ϵ, we can assume in (48) $c = 1$ (the standard adjoint graph norm), and we obtain a robust DPG method. This is the case for the model problem discussed here [28, 18] and e.g. for a linear acoustics problem, see [36, 27]. In principle, if α depends upon ϵ, we can compensate for α with the scaling constant but the round off error quickly prohibits such practices.

Comment. If we take for a starting point the second-order model problem (1), Broersen and Stevenson [9, 8] noted that the corresponding first order system is not uniquely defined and raised the question of how to select it in an optimal way. The issue concerns the definition of the additional unknown σ (the flux) and scaling of the first equation. Broersen and Stevenson advocated to work with the system,

$$\begin{cases} \sigma - \epsilon^{1/2} \boldsymbol{\nabla} u = 0 \\ -\text{div}(\epsilon^{1/2} u + bu) = f \,, \end{cases} \tag{50}$$

as opposed to

$$\begin{cases} \frac{1}{\epsilon} \sigma - \boldsymbol{\nabla} u = 0 \\ -\text{div}(u + bu) = f \,, \end{cases} \tag{51}$$

used in [28]. Different formulations result in different operators, different adjoints and, consequently, different graph norms. Constants α are different and, more importantly, the natural norms for the Lagrange multipliers (traces and fluxes) are different. Intuitively, we may prefer stronger test norms so the corresponding (dual) norm for fluxes/traces is weaker and it does not dominate the norm for the original unknowns. Additionally, we have to watch for different contributions to $A^* v$ to be of similar order to avoid round off truncations. Use of broken test spaces comes handy here as we can easily rescale dual variable components τ, v element-wise adjusting for element size and ϵ. All these issues are far from trivial and problem dependent.

Comment. For the convection-dominated problem, we do not have to augment (46) with the full group L^2-norm. With the dual variable (τ, v), it is sufficient to add only the L^2-norm of the second component. Also, the standard DPG method does not guarantee element conservation property. We can enforce it by requesting, for each element K, piece-wice constants $(0, 1_K)$ to be in the test space, and minimize the residual in the orthogonal complement of such functions [34]. We do not need then any additional L^2-term to localize the norm. The resulting formulation is robust as well.

Many of the issues discussed here disappear if we avoid breaking test functions all together and work directly with the mixed formulation involving solution u_h and an approximation of the error

representation function ψ. We can work then directly with the optimal test norm (46). This was the original idea by Cohen at all in [19], see also [8].

Comment. With the scaled graph norm and scaling coefficient $c \to 0$, we converge to the optimal test norm which delivers the L^2-projection. The corresponding optimal test functions coincide then with the optimal test functions of Barret and Morton [3] given by another trial-to-test operator $T_{BM} := (B')^{-1} \circ R_U$. The difference between the two trial-to-test operators is explained in the diagram below.

$$\begin{array}{ccc} U & \xrightarrow{B} & V' \\ \downarrow R_U & & \uparrow R_V \\ U' & \xleftarrow{B'} & V \end{array} \quad . \tag{52}$$

Our trial-to-test operator involves inverting Riesz operator R_V, whereas Barret-Morton operator involves inverting conjugate operator B'. Nevertheless, for the ultraweak formulation and the optimal test norm, inverting R_V reduces to the variational problem:

$$(A^* v, A^* \delta v) = (u, A^* \delta v) \quad \forall \delta v \tag{53}$$

which yields the Barret-Morton optimal test function $v = (A^*)^{-1} u$ which, in turn, deliver the L^2-projection. In this context, the DPG method can be seen as a localization technique for the computation of Barret-Morton test functions [17].

Despite fantastic stability properties, the adjoint graph norm is not easy to work with. The perturbation parameter ϵ has been moved to the test norm, and the resolution of the optimal test functions and/or error representation function ψ may be almost as difficult as the solution of the original problem. The original method proposed by Cohen, Dahmen and Welper [19] uses a double adaptive algorithm. Given a trial space, an additional a-posteriori error estimate for ψ is used to determine adaptively an "enriched test spaces" to secure a sufficient resolution of ψ. Once the error in ψ is of order of norm of ψ (see [19] for a precise criterion), the norm of ψ serves as an error indicator for refining the trial space, see also [21].

A different strategy was used in [30] where the authors used special Shishkin subelement meshes to resolve the optimal test functions.

Finally, a fundamentally different strategy was pursued in [28, 18]. The bilinear form for the ultraweak variational formulation with broken test spaces can be written concisely in the abstract form:

$$b((u, \hat{u}), v) = (u, A^* v) + \langle \hat{u}, v \rangle \tag{54}$$

where $u \in U = L^2(\Omega)^N$, test functions come from the broken counterpart of $V = D(A^*)$, and $\hat{u} \in \hat{U}$, where $\hat{U}$ is the space of Lagrange multipliers corresponding to the broken test space. Pick $u \in L^2(\Omega)^N$ and consider the corresponding solution v_u of the adjoint problem:

$$\begin{cases} v_u \in V := D(A^*) \\ A^* v_u = u \, . \end{cases} \tag{55}$$

Let $\|v\|_V$ denote an arbitrary test norm. Then,

$$\begin{aligned}
\|u\|^2 &= (u, A^* v_u) && (A^* v_u = u) \\
&= b((u,\hat{u}), v_u) && (v \text{ is globally conforming}) \\
&= \frac{|b((u,\hat{u}), v_u)|}{\|v_u\|_V} \|v_u\|_V \\
&\leq \sup_v \frac{|b((u,\hat{u}), v)|}{\|v\|_V} \|v_u\|_V && (\text{definition of supremum})\,.
\end{aligned}$$

Thus, if we can select the test norm in such a way that the solution of the adjoint problem (55) can be controlled robustly by the L^2-norm of the right-hand side,

$$\|v_u\|_V \leq C\|u\| \qquad (C \text{ independent of } \epsilon) \tag{56}$$

then the L^2-norm of u is controlled robustly by the energy norm corresponding to the test norm,

$$\|u\| \leq C \sup_v \frac{|b((u,\hat{u}), v)|}{\|v\|_V} \tag{57}$$

In other words, we try to select the test norm in such a way that at least the inf-sup constant for the ultraweak formulation is ϵ-independent. One obvious choice is the adjoint graph norm but other choices are possible as well. The search for an appropriate test norm leads thus to the stability analysis of the adjoint problem (on the continuous level).

For example, if we select in our model problem (1) for Γ_u the outflow boundary and for Γ_σ the inflow boundary,

$$\begin{aligned}
\Gamma_u &= \Gamma_{\text{out}} := \{x \in \Gamma \,:\, b(x)\cdot n > 0\} \\
\Gamma_\sigma &= \Gamma_{\text{in}} := \{x \in \Gamma \,:\, b(x)\cdot n \leq 0\}\,,
\end{aligned} \tag{58}$$

we learn that, under mild regularity assumption on advection b [18], the following quantities are controlled robustly in ϵ,

$$\|v\|,\ \epsilon^{1/2}\|v\|,\ \|b\cdot\nabla v\|,\ \|\text{div}\ \tau\|,\ \frac{1}{\epsilon^{1/2}}\|\tau\|\,. \tag{59}$$

Thus, if we put these Lego blocks together, we obtain a candidate for a test norm satisfying condition (56),

$$\|(\tau, v)\|_V^2 := \|v\|^2 + \epsilon\|v\|^2 + \|b\cdot\nabla v\|^2 + \|\text{div}\ \tau\|^2 + \frac{1}{\epsilon}\|\tau\|^2\,. \tag{60}$$

Note that, contrary to the adjoint graph norm, the terms with τ and v are now separated so the inversion of the Riesz operator decouples into two separate problems for τ and v. Unfortunately, for coarse meshes, the diffusion terms above are still dominated by reaction terms and, consequently, the corresponding optimal test functions develop boundary layers and are difficult to resolve. A simple remedy to the problem is to replace the coefficients in front of the zero order terms with *mesh dependent* terms. The modified test norm for an element K takes the form:

$$\|(\tau, v)\|_{V(K)}^2 := \min\{\frac{\epsilon}{h_K}, 1\}\|v\|^2 + \epsilon\|\nabla v\|^2 + \|b\cdot\nabla v\|^2 + \|\text{div}\ \tau\|^2 + \min\{\frac{1}{\epsilon}, \frac{1}{h_K}\}\|\tau\|^2 \tag{61}$$

where h_K is the element size. We called it *the robust norm*. Notice that norm (61) is smaller than norm (60) so the condition (56) is still satisfied. Fig. 1 compares optimal test functions for a 1D

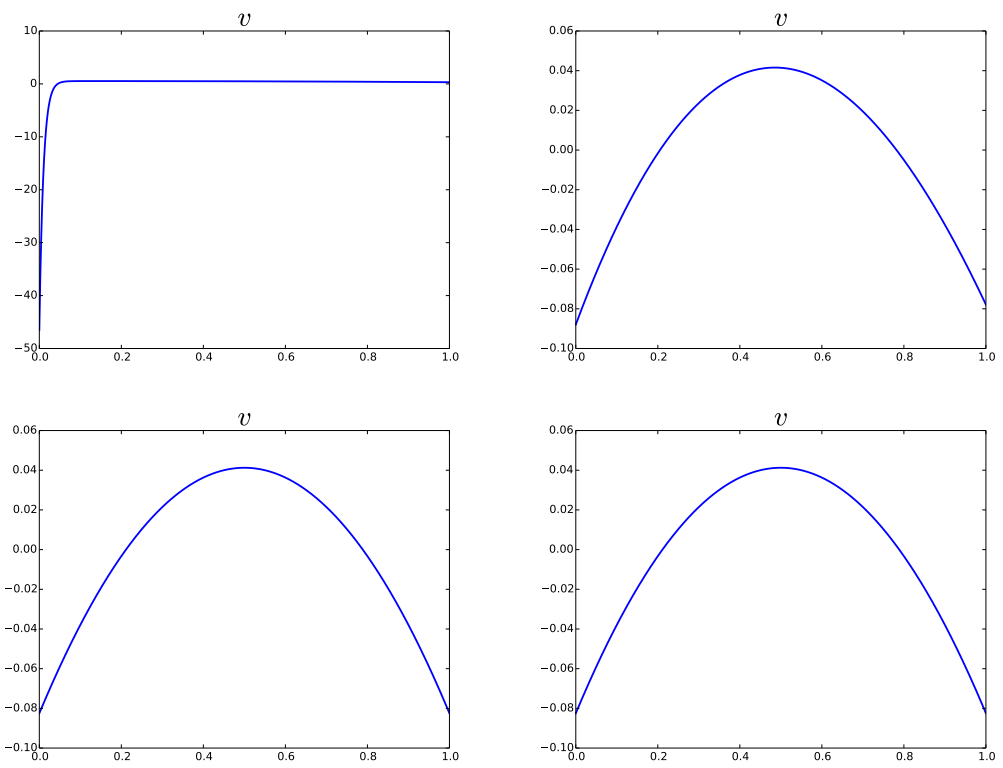

Figure 1: 1D convection-dominated diffusion problem with $\epsilon = 10^{-2}$ and $b = 1$. v-component of optimal test functions corresponding to a trial function $u(x) = x - \frac{1}{2}$ (left) and their resolution with polynomials of order $r = 3$ (right) for different test norms; top: adjoint graph norm, bottom: robust test norm.

version of our model problem for the adjoint graph and robust test norms. The optimal test functions corresponding to the robust test norm dot not develop anymore boundary layers and can be easily resolved with the enriched test space strategy.

Returning to the abstract notation, we can claim now the following error estimate:

$$\|u - u_h\| \leq C\|(u, \hat{u}) - (u_h, \hat{u}_h)\|_E = \inf_{(w_h, \hat{w}_h) \in U_h \times \hat{U}_h} \|(u, \hat{u}) - (w_h, \hat{w}_h)\|_E \tag{62}$$

In other words, we control *robustly* the L^2-norm of u by the minimized residual. We cannot claim, however, that the method is robust in the mathematical sense as the continuity constant M does depend upon ϵ.

7 Nonlinear Problems and Other Work

Nonlinear problems. The concept of the robust test norm developed systematically for the model convection-dominated diffusion problem has been formally extrapolated to both compressible and incompressible steady-state Navier-Stokes equations [16, 32]. At present, there is no systematic theory for nonlinear problems. We linearize the nonlinear equations and apply the linear DPG to the

linearized problem. If we freeze the test norm, this can be interpreted as a Newton-Gauss method applied to minimize the non-linear residual. In practice, however, the test norm is *not* fixed as it evolves with the background solution. Direct minimization of nonlinear residual has been investigated in [11].

Preconditioning , solvers and other related work. DPG delivers a positive-definite hermitian stiffness matrix suggesting the use of Conjugate Gradient (CG) method. An additive Schwartz preconditioner has been studied in [2]. Wieners and Wohlmut [35] proposed a preconditioner for the skeleton problem resulting from static condensation of all element local degrees-of-freedom, and initiated a study on multigrid methods. Convergence in weaker norms and first duality arguments were studied in [9]. A general framework for a fast implementation of ultraweak DPG methods for linear and nonlinear problems has been developed in [33].

The DPG method is a very young technology and the work on the method has barely started. The methodology offers a number of very attractive features: choice of different variational formulations (functional settings), choice of a specific norm, positive-definite and hermitian stiffness matrix, a-posteriori error estimate built-in, to mention a few. The work on a systematic treatment of singular perturbation problems is far from finished, and we expect to see new developments coming soon. Understanding of minimum residual methods with residual measured in dual norms for non-linear methods is minimal. The method is computationally expensive on the element level, and we need to develop new techniques, both algorithmic and purely implementational (use of multiple CPUs and GPUs) to accelerate the element computations.

We hope that this short exposition stimulates a further research on the method.

8 Acknowledgments

The work has been supported with grants by AFOSR (FA9550-12-1-0484) and National Science Foundation (DMS-1418822).

References

[1] I. Babuška. Error-bounds for finite element method. *Numer. Math*, 16, 1970/1971.

[2] A.T. Barker, S.C. Brenner, E.-H. Park, and L.-Y. Sung. A one-level additive Schwarz preconditioner for a discontinuous Petrov–Galerkin method. Technical report, Dept. of Math., Luisiana State University, 2013. http://arxiv.org/abs/1212.2645.

[3] J.W. Barret and K.W. Morton. Approximate symmetrization and Petrov-Galerkin methods for diffusion-convection problems. *Comput. Methods Appl. Mech. Engrg.*, 46:97–122, 1984.

[4] P. Bochev and M.D. Gunzburger. *Least-Squares Finite Element Methods*, volume 166 of *Applied Mathematical Sciences*. Springer Verlag, 2009.

[5] C.L. Bottasso, S. Micheletti, and R. Sacco. The discontinuous Petrov-Galerkin method for elliptic problems. *Comput. Methods Appl. Mech. Engrg.*, 191:3391–3409, 2002.

[6] J.H. Bramble, R.D. Lazarov, and J.E. Pasciak. A least-squares approach based on a discrete minus one inner product for first order systems. *Math. Comp*, 66, 1997.

[7] J. Bramwell, L. Demkowicz, J. Gopalakrishnan, and W. Qiu. A locking-free hp DPG method for linear elasticity with symmetric stresses. *Numer. Math.*, 122(4):671–707, 2012.

[8] D. Broersen and R. A. Stevenson. A robust Petrov-Galerkin discretisation of convection-diffusion equations. *Comput. Math. Appl.*, 68(11):1605–1618, 2014.

[9] D. Broersen and R. P. Stevenson. A petrov-galerkin discretization with optimal test space of a mild-weak formulation of convection-diffusion equations in mixed form. *IMA J. Numer. Anal.*, 35(1):39–73, 2015.

[10] T. Bui-Thanh, L. Demkowicz, and O. Ghattas. Constructively well-posed approximation methods with unity inf-sup and continuity. *Math. Comp.*, 82(284):1923–1952, 2013.

[11] T. Bui-Thanh and O. Ghattas. A PDE-constrained optimization approach to the discontinuous Petrov-Galerkin method with a trust region inexact Newton-CG solver. *Comput. Methods Appl. Mech. Engrg.*, 278:20–40, 2014.

[12] R. Cai, Z.and Lazarov, T.A. Manteuffel, and S.F. McCormick. First-order system least squares for second-order partial differential equations. I. *SIAM J. Numer. Anal.*, 31:1785–1799, 1994.

[13] C. Carstensen, L. Demkowicz, and J. Gopalakrishnan. A posteriori error control for DPG methods. *SIAM J. Numer. Anal.*, 52(3):1335–1353, 2014.

[14] C. Carstensen, L. Demkowicz, and J. Gopalakrishnan. Breaking spaces and forms for the DPG method and applications including Maxwell equations. *Num. Math.*, 2015. submitted.

[15] P. Causin and R. Sacco. A discontinuous Petrov-Galerkin method with Lagrangian multipliers for second order elliptic problems. *SIAM J. Numer. Anal.*, 43, 2005.

[16] J. Chan, L. Demkowicz, and R. Moser. A DPG method for steady viscous compressible flow. *Computers and Fluids*, 98, 2014.

[17] J. Chan, J. Gopalakrishnan, and L. Demkowicz. Global properties of DPG test spaces for convection-diffusion problems. Technical Report 5, ICES, 2013.

[18] J. Chan, N. Heuer, Tan Bui-Thanh B., and L. Demkowicz. A robust DPG method for convection-dominated diffusion problems II: Adjoint boundary conditions and mesh-dependent test norms. *Comput. Math. Appl.*, 67(4):771–795, 2014.

[19] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and variational stabilization for convection-diffusion equations. *ESAIM Math. Model. Numer. Anal.*, 46(5):1247–1273, 2012. see also Technical Report 2011/323, Institut fuer Geometrie und Praktische Mathematik.

[20] W. Dahmen, Ch. Huang, Ch. Schwab, and G. Welper. Adaptive Petrov Galerkin methods for first order transport equations. *SIAM J. Num. Anal.*, 50(5), 2012.

[21] W. Dahmen, Ch. Plesken, and G. Welper. Double greedy algorithms: Reduced basis methods for transport dominated problems. *ESAIM Math. Model. Numer. Anal.*, 48(3):623–663, 2014.

[22] L. Demkowicz. Various variational formulations and Closed Range Theorem. Technical report, ICES, January 15–03.

[23] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. Part I: The transport equation. *Comput. Methods Appl. Mech. Engrg.*, 199(23-24):1558–1572, 2010.

[24] L. Demkowicz and J. Gopalakrishnan. Analysis of the DPG method for the Poisson problem. *SIAM J. Num. Anal.*, 49(5):1788–1809, 2011. see also ICES Report 2010/37.

[25] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. Part II: Optimal test functions. *Numer. Meth. Part. D. E.*, pages 70–105, 2011. see also ICES Report 9/16.

[26] L. Demkowicz and J. Gopalakrishnan. A primal DPG method without a first order reformulation. *Comput. Math. Appl.*, 66(6):1058–1064, 2013.

[27] L. Demkowicz, J. Gopalakrishnan, I. Muga, and J. Zitelli. Wavenumber explicit analysis for a DPG method for the multidimensional Helmholtz equation. *Comput. Methods Appl. Mech. Engrg.*, 213-216:126–138, 2012.

[28] L. Demkowicz and N. Heuer. Robust DPG method for convection-dominated diffusion problems. *SIAM J. Num. Anal*, 51:2514–2537, 2013. see also ICES Report 2011/13.

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

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

[31] N. Roberts, Tan Bui-Thanh B., and L. Demkowicz. The DPG method for the Stokes problem. *Comput. Math. Appl.*, 67(4):966–995, 2014.

[32] N. Roberts, L. Demkowicz, and R. Moser. A discontinuous Petrov-Galerkin methodology for adaptive solutions to the incompressible Navier-Stokes equations. *J. Comp. Phys.*, 2015. accepted.

[33] N. V. Roberts. Camellia: A software framework for Discontinuous Petrov-Galerkin methods. *Comput. Math. Appl.*, 68:1581–1604, 2014.

[34] Ellis T., L. Demkowicz, and J. Chan. Locally conservative discontinuous Petrov-Galerkin finite elements for fluid problems. *Comput. Math. Appl.*, 68:1530–1549, 2014. Special Issue on Least Squares and DPG Methods.

[35] Ch. Wieners and B. Wohlmuth. Robust operator estimates and the application to substructuring methods for first–order systems. *ESAIM Math. Mod. Num. Anal.*, 2014.

[36] J. Zitelli, I. Muga, L. Demkowicz, J. Gopalakrishnan, D. Pardo, and V. Calo. A class of discontinuous Petrov-Galerkin methods. Part IV: Wave propagation problems. *J. Comp. Phys.*, 230:2406–2432, 2011. see also ICES Report 2010/17.