Convergence Rates of the DPG Method with Reduced Test Space Degree

Timaeus Bouma
Portland State University

Jay Gopalakrishnan
Portland State University, gjay@pdx.edu

Ammar Harb
Portland State University

CONVERGENCE RATES OF THE DPG METHOD WITH REDUCED TEST SPACE DEGREE

TIMAEUS BOUMA, JAY GOPALAKRISHNAN, AND AMMAR HARB

ABSTRACT. This paper presents a duality theorem of the Aubin-Nitsche type for discontinuous Petrov Galerkin (DPG) methods. This explains the numerically observed higher convergence rates in weaker norms. Considering the specific example of the mild-weak (or primal) DPG method for the Laplace equation, two further results are obtained. First, the DPG method continues to be solvable even when the test space degree is reduced, provided it is odd. Second, a non-conforming method of analysis is developed to explain the numerically observed convergence rates for a test space of reduced degree.

1. INTRODUCTION

The purpose of this note is to provide a theoretical explanation for some numerically observed convergence rates of the discontinuous Petrov-Galerkin (DPG) method. While some aspects of the theory that follows are general, we will use the Laplace equation throughout as the example to illustrate the main points. There are two DPG methods for the Laplace's equation. One is based on an ultra-weak formulation [6] (where constitutive and conservation equations are both integrated by parts) while the other is based on the so-called mild-weak, or primal formulation, developed in [2, 7] (where only the conservation equation is integrated by parts). The example which motivates our study is the latter.

The method will be precisely introduced later. But to outline this study, consider applying the method on a two-dimensional domain Ω meshed by a geometrically conforming finite element mesh of triangles of mesh size h. The method produces an approximation u_h to the solution u of the Laplace's equation in the interior of the mesh elements, as well as an approximation to the flux q on the element interfaces. The first is a polynomial of degree at most k_u on each mesh element and the second is a polynomial of degree at most k_q on each mesh edge. The method uses test functions v that are polynomials of degree at most k_v on each mesh element. It is the interplay between the convergence rates and the degrees k_u, k_q, k_v that we intend to study.

We identify three cases for study. Let $k \geqslant 1$ be an integer. The cases are as shown:

	k_u	k_q	k_v
Case 1:	k	$k-1$	$k+1$,
Case 2:	$k-1$	$k-1$	k,
Case 3:	k	$k-1$	k.

The first case is the standard DPG setting for which error estimates in the energy norm are proven in [7]. The other two cases are motivated by a desire to reduce the test space degree and have not been analyzed previously.

What is the practical importance of reduced order test spaces? We give a three-part answer: First, consider the left hand side matrix of the linear system arising from the DPG method. Its assembly requires computation of the Gram matrix of the test space. Even though this matrix is block diagonal, it is of some practical interest to reduce the block size, especially when operating near the limit of memory bandwidth in multi-core architectures. Second, consider the right hand side computation. In cases where load terms are expensive to evaluate, reduction of test space degree brings significant computational savings. Finally, the third and the most compelling reason that prompted us to investigate this issue, is that there are practical limits on the degree of polynomials one can use in most finite element software. We prefer to hit this practical limiting degree with the trial space, rather than with the test space, because it is the approximation properties of the trial space that determines the final solution quality.

Our numerical experience with a few examples with smooth solutions, one of which is fully reported in Section 4, is summarized in Table 1. We observed that Case 2 is not always stable: It yielded singular stiffness matrices for some even k. However, when k is odd, it converged, albeit at one order less than the standard DPG case displayed in the first row. Keeping k odd and moving to Case 3, we find that the original DPG convergence rates can be recovered, in spite of using a smaller k_v. Finally, we observed that the convergence rate in $L^2(\Omega)$, in all cases, is one order higher than in $H^1(\Omega)$. These observations motivate our ensuing theoretical studies.

TABLE 1. Summary of numerically observed convergence rates

	h-convergence rates of u_h	
	in $H^1(\Omega)$	in $L^2(\Omega)$
Case 1	k	$k+1$
Case 2 (k odd)	$k-1$	k
Case 3 (k odd)	k	$k+1$

We explain the higher convergence rate in $L^2(\Omega)$ by developing a duality argument for DPG methods. The duality theory is general and can be applied beyond the Laplace example. We also give a complete theoretical explanation for the even-odd behavior, including a negative result by counterexample for even k, and a proof of a positive result for odd k. In explaining Case 3, we highlight a connection between the DPG method and a weakly conforming method, and show how to use a nonconforming-type analysis, using the second Strang lemma, in the DPG context.

In the next section, we gather a number of abstract results applicable to any DPG method in a general framework consisting of a trial space of interior and interface variables. In Section 3, we introduce the DPG method for the Dirichlet problem and in distinct subsections, provide explanations for the convergence rates in the above-mentioned three cases. Finally in Section 4, we present details of numerical experiments and discuss the practical importance of lower test order test spaces.

2. General results

Suppose X_0, $\hat{X}$, and Y are Hilbert spaces over $\mathbb{C}$. Solutions are sought in the "trial space" $X = X_0 \times \hat{X}$ and have an "interior" component in X_0 and an "interface" component in $\hat{X}$. Suppose there are continuous sesquilinear forms $\hat{b}(\cdot,\cdot) : \hat{X} \times Y \to \mathbb{C}$ and $b_0(\cdot,\cdot) : X_0 \times Y \to \mathbb{C}$, and let $b(\cdot,\cdot) : X \times Y \to \mathbb{C}$ be set by

$$b(\,(w,\hat{w}),y\,) = b_0(w,y) + \hat{b}(\hat{w},y),$$

for all $(w, \hat{w}) \in X$ and $y \in Y$. Let Y^* denote the space of continuous conjugate-linear functionals on Y. Given any $\ell \in Y^*$ we are interested in approximating an $x \equiv (x_0, \hat{x}) \in X$ satisfying

$$b(x, y) = \ell(y) \qquad \forall y \in Y. \tag{1}$$

Let $X_{h,0} \subseteq X_0$ and $\hat{X}_h \subseteq \hat{X}$ be finite-dimensional subspaces and let $X_h = X_{h,0} \times \hat{X}_h$. Let Y^r denote a finite-dimensional subspace of Y and let $T^r : X \to Y^r$ be defined by $(T^r w, y)_Y = b(w, y)$ for all $y \in Y^r$. Here and throughout $(\cdot, \cdot)_Y$ denotes the inner product in Y. The DPG method for (1) computes $x_h \equiv (x_{h,0}, \hat{x}_h)$ in X_h satisfying

$$b(x_h, y) = \ell(y), \qquad \forall y \in Y_h^r = T^r(X_h). \tag{2}$$

A fundamental quasioptimality result for DPG methods is stated in Theorem 2.3 below. It holds under these assumptions.

Assumption 2.1. Suppose $\{z \in X : b(z, y) = 0, \forall y \in Y\} = \{0\}$ and suppose there exist $C_1, C_2 > 0$ such that

$$C_1 \|y\|_Y \leqslant \sup_{0 \neq z \in X} \frac{|b(z, y)|}{\|z\|_X} \leqslant C_2 \|y\|_Y \qquad \forall y \in Y. \tag{3}$$

Assumption 2.2. There is a linear operator $\Pi : Y \to Y^r$ and a $C_\Pi > 0$ such that for all $w_h \in X_h$ and all $v \in Y$,

$$b(w_h, v - \Pi v) = 0, \qquad \text{and} \qquad \|\Pi v\|_Y \leqslant C_\Pi \|v\|_Y.$$

Theorem 2.3 (see [11]). *Suppose Assumptions 2.1 and 2.2 hold. Then the DPG method* (2) *is uniquely solvable for x_h and*

$$\|x - x_h\|_X \leqslant \frac{C_2 C_\Pi}{C_1} \inf_{z_h \in X_h} \|x - z_h\|_X$$

where x is the unique exact solution of (1).

Another well-known result, motivated by [5], is an equivalence of the DPG method with a mixed Bubnov-Galerkin formulation. To state it, we first define the error representation function: let ε^r be the unique element of Y^r satisfying

$$(\varepsilon^r, y)_Y = \ell(y) - b(x_h, y), \forall y \in Y^r. \tag{4}$$

Theorem 2.4. *The following are equivalent statements:*
  i) $x_h \in X_h$ *solves the DPG method* (2).
  ii) $x_h \in X_h$ *and $\varepsilon^r \in Y^r$ solve the mixed formulation*

$$(\varepsilon^r, y)_Y + b(x_h, y) = \ell(y) \qquad\qquad \forall y \in Y^r, \tag{5a}$$
$$b(z_h, \varepsilon^r) = 0 \qquad\qquad \forall z_h \in X_h. \tag{5b}$$

Its simple proof is omitted (see e.g. [9]).

Remark 2.5. The norm of ε^r is bounded by the error: Choosing $y = \varepsilon^r$ in (4), we obtain

$$\|\varepsilon^r\|_Y^2 = (\varepsilon^r, \varepsilon^r)_Y = \ell(\varepsilon^r) - b(x_h, \varepsilon^r) = b(x - x_h, \varepsilon^r).$$

Hence, by Assumption 2.1,

$$\|\varepsilon^r\|_Y \leqslant C_2 \|x - x_h\|_X. \tag{6}$$

This theme is further developed in [3], where $\|\varepsilon^r\|_Y$ is established to be both a reliable and an efficient error estimator.

2.1. **Weakly conforming test space.** Let

$$Y_0^r = \{y \in Y^r : \ \hat{b}(\hat{w}_h, y) = 0, \ \ \forall \hat{w}_h \in \hat{X}_h\} \tag{7}$$

and let $T_0^r : X_0 \to Y_0^r$ be defined by $(T_0^r w, y)_Y = b_0(w, y)$ for all $y \in Y_0^r$. In the examples we have in mind, Y^r is a discontinuous Galerkin (DG) space, and Y_0^r is a subspace with weak interelement continuity constraints, i.e., a weakly conforming space. In such cases, the application of the operator T_0^r requires a global inversion. We then compare these two DPG methods:

$$\text{Find } (x_{h,0}, \hat{x}_h) \in X_h : \quad b(\,(x_{h,0}, \hat{x}_h), y\,) = \ell(y) \qquad\qquad \forall y \in Y_h^r \equiv T^r(X_h). \tag{8a}$$

$$\text{Find } x_{h,0} \in X_{h,0} : \quad b_0(x_{h,0}, y) = \ell(y) \qquad\qquad \forall y \in Y_{h,0}^r \equiv T_0^r(X_{h,0}). \tag{8b}$$

The first is the same as (2), the standard DPG method. We view (8a) as a "hybridized" form of the second method (8b), and the next theorem shows in what sense they are equivalent. The method (8b) is not the preferred for implementation due to the expense of applying T_0^r, but we will use it later for error analysis.

Theorem 2.6. *The test spaces satisfy* $Y_{h,0}^r \subset Y_h^r$. *Hence, if* $(x_{h,0}, \hat{x}_h) \in X_h$ *solves* (8a), *then* $x_{h,0}$ *solves* (8b).

Proof. Let $Y_\perp^r$ be the Y-orthogonal complement of Y_h^r in Y^r. Then we have the orthogonal decomposition

$$Y^r = Y_h^r + Y_\perp^r. \tag{9}$$

Let $y_0 \in Y_{h,0}^r$. Apply (9) to decompose $y_0 = y_h + y_\perp$, with $y_h \in Y_h^r$ and $y_\perp \in Y_\perp^r$.

First, we claim that $y_\perp \in Y_0^r$. This is because

$$\hat{b}(\hat{w}_h, y_\perp) = (T^r(0, \hat{w}_h), y_\perp)_Y = 0 \qquad \forall \hat{w}_h \in \hat{X}_h.$$

The last identity followed from the orthogonality of $y_\perp$ to $T^r(X_h)$.

Next, we claim that $y_\perp = 0$. It suffices to prove that $(y_0, y_\perp)_Y = 0$ since $(y_0, y_\perp)_Y = \|y_\perp\|_Y^2$. Since $y_0 \in Y_{h,0}^r$, there is a $w_h \in X_{h,0}$ such that $y_0 = T_0^r w_h$. Then,

$$\begin{aligned}
(y_0, y_\perp)_Y &= (T_0^r w_h, y_\perp)_Y = b_0(w_h, y_\perp) & \text{as } y_\perp \in Y_0^r \\
&= (T^r(w_h, 0), y_\perp)_Y = 0 & \text{as } T^r(X_h) \perp y_\perp.
\end{aligned}$$

Finally, since $y_\perp = 0$, we have $y_0 = y_h + 0 \in Y_h^r$. Thus $Y_{h,0}^r \subset Y_h^r$. The second statement of the theorem is now obvious by choosing $y \in Y_{h,0}^r$ in (8a). □

2.2. **Injectivity.** Let $B_h : X_h \to (Y^r)^*$ be the operator generated by the form $b(\cdot, \cdot)$, i.e.,

$$(B_h w_h)(y) = b(w_h, y), \qquad \forall w_h \in X_h, \ y \in Y^r.$$

Similarly, let $\hat{B}_h : \hat{X}_h \to (Y^r)^*$ be defined by

$$(\hat{B}_h \hat{z}_h)(y) = \hat{b}(\hat{z}_h, y), \qquad \forall \hat{z}_h \in \hat{X}_h, \ y \in Y^r. \tag{10}$$

The injectivity of B_h yields the unique solvability of the DPG method.

Assumption 2.7. Suppose
  a) $X_{h,0} \subseteq Y^r$,
  b) $\hat{b}(\hat{z}_h, z_0) = 0$ for all $\hat{z}_h \in \hat{X}_h$ and $z_0 \in X_{h,0}$, and
  c) any $z_0 \in X_{h,0}$ satisfying $b_0(z_0, z_0) = 0$ must be zero.

Theorem 2.8. *If* B_h *is injective, then* $\hat{B}_h$ *is injective, and the DPG method* (2) *is uniquely solvable. Conversely, if* $\hat{B}_h$ *is injective, then* B_h *is injective, provided Assumption 2.7 holds.*

Proof. Suppose B_h is injective. The injectivity of $\hat{B}_h$ is obvious from $\hat{B}_h \hat{w}_h = B_h(0, \hat{w}_h)$. We also claim that T^r is injective: Indeed, if $w_h \in X_h$ satisfies $T^r w_h = 0$, then $0 = (T^r w_h, y)_Y = b(w_h, y) = (B_h w_h)(y)$ for all $y \in Y^r$, so $w_h = 0$. The injectivity of T^r implies that $\dim(Y_h^r) = \dim(X_h)$, so the DPG method (2) yields a square system. Moreover, since (2) is the same as

$$(T^r x_h, T^r w_h)_Y = \ell(T^r w_h) \qquad \forall\, w_h \in X_h,$$

the injectivity of T^r also implies that there is a unique solution x_h in X_h.

Now suppose $\hat{B}_h$ is injective. To prove that B_h is injective, consider a $(w_0, \hat{w}) \in X_h$ satisfying $B_h(w_0, \hat{w}) = 0$. Then

$$
\begin{aligned}
0 &= (B_h(w_0, \hat{w}))(w_0) && \text{by Assumption 2.7(a)} \\
&= b\left((w_0, \hat{w}), w_0\right) = b_0(w_0, w_0) + \hat{b}(\hat{w}, w_0) && \\
&= b_0(w_0, w_0), && \text{by Assumption 2.7(b).}
\end{aligned}
$$

Therefore, by Assumption 2.7(c), $w_0 = 0$. It only remains to show that $\hat{w} = 0$. But $(\hat{B}_h \hat{w})(y) = \hat{b}(\hat{w}, y) = b\left((0, \hat{w}), y\right) = (B_h(w_0, \hat{w}))(y) = 0$ for all $y \in Y^r$. Hence the injectivity of $\hat{B}_h$ implies $\hat{w} = 0$. $\qquad \square$

2.3. **Duality argument for DPG.** By virtue of Theorem 2.4, we may rewrite the DPG method (2) as follows: Find $x_{h,0} \in X_{0,h}$, $\hat{x}_h \in \hat{X}_h$, and $\varepsilon^r \in Y^r$ solving

$$
\begin{aligned}
b_0(w, \varepsilon^r) &= 0 & \forall w \in X_{0,h}, &\qquad (11a) \\
\hat{b}(\hat{w}, \varepsilon^r) &= 0 & \forall \hat{w} \in \hat{X}_h, &\qquad (11b) \\
b_0(x_{h,0}, y) + \hat{b}(\hat{x}_h, y) + (\varepsilon^r, y)_Y &= \ell(y), & \forall y \in Y^r. &\qquad (11c)
\end{aligned}
$$

Defining

$$a(z, \hat{z}, v | w, \hat{w}, y) = \overline{b_0(w, v)} + \overline{\hat{b}(\hat{w}, v)} + b_0(z, y) + \hat{b}(\hat{z}, y) + (v, y)_Y,$$

the mixed system (11) can then be rewritten as

$$a(x_{h,0}, \hat{x}_h, \varepsilon^r | w, \hat{w}, y) = \ell(y), \qquad \forall w \in X_{0,h}, \hat{w} \in \hat{X}_h, y \in Y^r,$$

where the complex conjugate on the first two terms make the form a sesquilinear. Now, observe that with $\varepsilon = 0$, the exact solution $(x_0, \hat{x}, \varepsilon) \in X_0 \times \hat{X} \times Y$ satisfies the same equation for all $w \in X_0, \hat{w} \in \hat{X}, y \in Y$. Hence, we have a 'Galerkin orthogonality' relation

$$a(x_0 - x_{h,0}, \hat{x} - \hat{x}_h, \varepsilon - \varepsilon^r | w, \hat{w}, y) = 0, \qquad (12)$$

for all $w \in X_{0,h}, \hat{w} \in \hat{X}_h, y \in Y^r$. Note also that

$$
\begin{aligned}
|a(z, \hat{z}, v | w, \hat{w}, y)| &\leqslant C_2 \|(z, \hat{z})\|_X \|y\|_Y + C_2 \|(w, \hat{w})\|_X \|v\|_Y + \|v\|_Y \|y\|_Y \\
&\leqslant \left(C_2^2 \|(z, \hat{z})\|_X^2 + 2\|v\|_Y^2\right)^{1/2} \left(C_2^2 \|(w, \hat{w})\|_X^2 + 2\|y\|_Y^2\right)^{1/2} \\
&\leqslant \|a\| \, \|(z, \hat{z}, v)\|_{X_0 \times \hat{X} \times Y} \|(w, \hat{w}, y)\|_{X_0 \times \hat{X} \times Y}
\end{aligned}
$$

where $\|a\|$ is a constant not larger than $\max(C_2^2, 2)$. Under the following assumption, we can extend the Aubin-Nitsche technique [15] to DPG methods, as seen in the next theorem.

Assumption 2.9. Suppose L and Z are Hilbert spaces such that the embeddings $Z \subseteq X_0 \times \hat{X} \times Y$ and $X_0 \subseteq L$ are continuous. Assume that there is a $C_3(h) > 0$ such that for any $g \in L$, there is a $U(g) \in Z$ satisfying

$$a(w, \hat{w}, y | U(g)) = (w, g)_L \qquad (13)$$

for all $(w, \hat{w}, y) \in X_0 \times \hat{X} \times Y$ and

$$\inf_{W \in X_{0,h} \times \hat{X}_h \times Y^r} \|U(g) - W\|_{X_0 \times \hat{X} \times Y} \leqslant C_3(h)\|g\|_L. \tag{14}$$

Theorem 2.10. *Suppose Assumption 2.9 holds. Then,*

$$\|x - x_{h,0}\|_L \leqslant C_3(h)\|a\| \|(x, \hat{x}, \varepsilon) - (x_{h,0}, \hat{x}_h, \varepsilon^r)\|_{X_0 \times \hat{X} \times Y}.$$

Proof. Setting $g = w = x - x_{h,0}$, $\hat{w} = \hat{x} - \hat{x}_h$, and $y = \varepsilon - \varepsilon^r$ in (13),

$$\begin{aligned}
\|x - x_{h,0}\|_L^2 &= a(x - x_{h,0}, \hat{x} - \hat{x}_h, \varepsilon - \varepsilon^r | U(x - x_{h,0})) \\
&= a(x - x_{h,0}, \hat{x} - \hat{x}_h, \varepsilon - \varepsilon^r | U(x - x_{h,0}) - W), \qquad \text{by (12)}, \\
&\leqslant \|a\| \|(x - x_{h,0}, \hat{x} - \hat{x}_h, \varepsilon - \varepsilon^r)\|_{X_0 \times \hat{X} \times Y} \|U(x - x_{h,0}) - W\|_{X_0 \times \hat{X} \times Y}
\end{aligned}$$

for any $W \in X_{0,h} \times \hat{X}_h \times Y^r$. Hence (14) completes the proof. $\square$

Remark 2.11. Let $A : X_0 \times \hat{X} \times Y \to (X_0 \times \hat{X} \times Y)^*$ be the operator generated by $a(\cdot, \cdot)$, i.e., $(A(z, \hat{z}, v))(w, \hat{w}, y) = a(z, \hat{z}, v | w, \hat{w}, y)$ for all $(z, \hat{z}, v), (w, \hat{w}, y) \in X_0 \times \hat{X} \times Y$. If Assumption 2.1 holds, then A is a bijection. (This follows from the Babuška-Brezzi theory [1], applied to the mixed system (5): the "inf-sup condition" follows from (3), and the "coercivity in the kernel condition" is trivial.) Hence, the dual operator of A is also a bijection whereby we conclude that (13) has a unique solution $U(g)$.

Remark 2.12. All results of this section hold for spaces over the real field $\mathbb{R}$ – one only needs to replace $\mathbb{C}$ by $\mathbb{R}$, sesquilinear by bilinear, and conjugate-linear by linear to obtain the corresponding statements for real valued function spaces. The DPG method for the Helmholtz equation [10] provides an example where sesquilinear forms over $\mathbb{C}$ are used. For simplicity, in the remaining sections we will restrict ourselves to real-valued functions.

3. APPLICATION TO THE LAPLACE EQUATION

Suppose Ω is a bounded open polygon in $\mathbb{R}^2$ with Lipschitz boundary, meshed by Ω_h, a geometrically conforming shape regular finite element mesh of triangles. Let $h = \max_{K \in \Omega_h} \operatorname{diam} K$. Let $\partial\Omega_h$ denote the collection of all element boundaries ∂K for all elements K in Ω_h. We now study the DPG approximation to the Dirichlet problem

$$\begin{aligned}
-\Delta u &= f && \text{on } \Omega, &&& \text{(15a)} \\
u &= 0 && \text{on } \partial\Omega. &&& \text{(15b)}
\end{aligned}$$

All functions are real-valued in this section.

Omitting a detailed derivation of the method, which can be found in [2, 7], we simply specify how the method can be obtained by setting these within the general framework of section 2:

$$\begin{aligned}
X_0 &= H_0^1(\Omega), \qquad \hat{X} = H^{-1/2}(\partial\Omega_h), \\
Y &= H^1(\Omega_h), \qquad \text{where} \\
H^1(\Omega_h) &= \{v : v|_K \in H^1(K), \ \forall K \in \Omega_h\}, \\
H^{-1/2}(\partial\Omega_h) &= \{\eta \in \prod_K H^{-1/2}(\partial K) : \ \exists \, r \in H(\operatorname{div}, \Omega) \text{ such that} \\
&\qquad \eta|_{\partial K} = r \cdot n|_{\partial K}, \quad \forall K \in \Omega_h\},
\end{aligned}$$

where n denotes the unit outward normals on the boundary of mesh elements. The space $H^{-1/2}(\partial\Omega_h)$ is normed, as in [16], by

$$\|\hat{r}_n\|_{H^{-1/2}(\partial\Omega_h)} = \inf\left\{\|r\|_{H(\mathrm{div},\Omega)} : \ r \in H(\mathrm{div},\Omega) \text{ such that } \hat{r}_n|_{\partial K} = r \cdot n|_{\partial K} \ \forall\, K \in \Omega_h\right\}. \quad (16)$$

The "broken" Sobolev space $H^1(\Omega_h)$ is normed by

$$\|v\|^2_{H^1(\Omega_h)} = (v,v)_{\Omega_h} + (\mathrm{grad}\, v, \mathrm{grad}\, v)_{\Omega_h}. \quad (17)$$

Throughout the rest of the paper, the derivatives are always calculated element by element, and

$$(r,s)_{\Omega_h} = \sum_{K \in \Omega_h} (r,s)_K, \qquad \langle\ell,w\rangle_{\partial\Omega_h} = \sum_{K \in \Omega_h} \langle\ell,w\rangle_{1/2,\partial K},$$

where $(\cdot,\cdot)_K$ denotes the $L^2(K)$-inner product and $\langle\ell,\cdot\rangle_{1/2,\partial K}$ denotes the action of a functional ℓ in $H^{-1/2}(\partial K)$. The bilinear and linear forms of the weak formulation are set by

$$b_0(w,y) = (\mathrm{grad}\, w, \mathrm{grad}\, y)_{\Omega_h}, \qquad \hat{b}(\hat{r}_n,y) = -\langle\hat{r}_n,y\rangle_{\partial\Omega_h}, \qquad \ell(y) = (f,y)_\Omega.$$

Assumption 2.1 was verified for this formulation in [7]. We will denote the exact solution of the resulting weak formulation (1) by $(u,\hat{q}_n) \in X$. Note that $\hat{q}_n|_{\partial K} = \partial_n u|_{\partial K}$ for all $K \in \Omega_h$.

To complete the specification of the method, it only remains to set the discrete spaces. Let $P_k(D)$ denote the set of polynomials of degree at most k on the domain D (with the understanding that the set is trivial when $k < 0$). Let $P_k(\Omega_h) = \{v : v|_K \in P_k(K) \text{ for all } K \in \Omega_h\}$ and let $P_k(\partial\Omega_h)$ denote the set of functions v on $\partial\Omega_h$ having the property $v|_E \in P_k(E)$ for all edges of ∂K and for all $K \in \Omega_h$. Then, recalling the three cases mentioned in section 1, we set, for any integer $k \geqslant 1$,

Case 1	Case 2	Case 3
$X_{h,0} = P_k(\Omega_h) \cap X_0$	$X_{h,0} = P_{k-1}(\Omega_h) \cap X_0$	$X_{h,0} = P_k(\Omega_h) \cap X_0,$
$\hat{X}_h = P_{k-1}(\partial\Omega_h) \cap \hat{X}$	$\hat{X}_h = P_{k-1}(\partial\Omega_h) \cap \hat{X}$	$\hat{X}_h = P_{k-1}(\partial\Omega_h) \cap \hat{X},$
$Y^r = P_{k+1}(\Omega_h)$	$Y^r = P_k(\Omega_h)$	$Y^r = P_k(\Omega_h).$

The discrete solution in each of these cases is denoted by $(u_h, \hat{q}_{n,h}) \in X_h$. We now proceed to study these cases and explain the observations in Table 1.

3.1. **Case 1: Application of the duality argument.** For Case 1, Assumption 2.2 was verified in [7]. This then led to [7, Theorem 4.1], which states that

$$\|u - u_h\|_{H^1(\Omega)} + \|\hat{q}_n - \hat{q}_{n,h}\|_{H^{-1/2}(\partial\Omega_h)} \leqslant C \inf_{(w_h,\hat{r}_{n,h})\in X_h} \left(\|u - w_h\|_{H^1(\Omega)} + \|\hat{q}_n - \hat{r}_{n,h}\|_{H^{-1/2}(\partial\Omega_h)}\right).$$

Here and henceforth, C denotes a generic constant independent of the size of the triangles in Ω_h (but dependent on mesh shape regularity), whose value at different occurrences may vary. As explained in previous papers (see e.g., [6]), applications of the Bramble-Hilbert Lemma in the Lagrange and Raviart-Thomas spaces show that

$$\inf_{w_h\in P_l(\Omega_h)\cap X_0} \|u - w_h\|_{H^1(\Omega)} \leqslant Ch^l|u|_{H^{l+1}(\Omega)}, \qquad\qquad \forall l \geqslant 0, \quad (18a)$$

$$\inf_{\hat{r}_{n,h}\in P_{m-1}(\partial\Omega_h)\cap\hat{X}} \|\hat{q}_n - \hat{r}_{n,h}\|_{H^{-1/2}(\partial\Omega_h)} \leqslant Ch^m \left(|u|_{H^{m+1}(\Omega)} + |f|_{H^m(\Omega)}\right), \quad \forall m \geqslant 1. \quad (18b)$$

Therefore,

$$\|u - u_h\|_{H^1(\Omega)} + \|\hat{q}_n - \hat{q}_{n,h}\|_{H^{-1/2}(\partial\Omega_h)} \leqslant Ch^k \left(|u|_{H^{k+1}(\Omega)} + |f|_{H^k(\Omega)}\right). \quad (19)$$

Hence the $O(h^k)$ convergence of $\|u-u_h\|_{H^1(\Omega)}$ (first entry of Table 1) is completely explained. To explain the $O(h^{k+1})$ convergence of $\|u-u_h\|_{L^2(\Omega)}$, we apply the duality argument of Theorem 2.10. Its hypothesis is verified in the next proof.

Theorem 3.1. *Suppose Ω is convex. Then, for Case 1,*

$$\|u - u_h\|_{L^2(\Omega)} \leqslant C h^{k+1} \left(|u|_{H^{k+1}(\Omega)} + |f|_{H^k(\Omega)} \right).$$

Proof. Set

$$Z_1 = H^2(\Omega) \cap X_0, \qquad\qquad L = L^2(\Omega),$$
$$Z_2 = H^2(\Omega) \cap Y, \qquad\qquad Z = Z_1 \times \hat{X} \times Z_2.$$

To verify Assumption 2.9, let $g \in L$. By Remark 2.11, there is a unique $U(g) \equiv (z, \hat{z}_n, d) \in X_0 \times \hat{X} \times Y$ solving (13). Writing out (13) in component form,

$$(d, y)_Y + (\operatorname{grad} z, \operatorname{grad} y)_{\Omega_h} - \langle \hat{z}_n, y \rangle_{\partial \Omega_h} = 0, \qquad \forall y \in Y, \tag{20a}$$
$$(\operatorname{grad} d, \operatorname{grad} w)_{\Omega_h} = (g, w)_{\Omega_h} \qquad \forall w \in X_0, \tag{20b}$$
$$\langle \hat{w}_n, d \rangle_{\partial \Omega_h} = 0 \qquad \forall \hat{w}_n \in \hat{X}. \tag{20c}$$

We need to understand the regularity of solutions of (20). Considering the d component first, we claim that (20c) implies $d \in H_0^1(\Omega)$: Indeed the distributional gradient $\operatorname{grad} d$ acting on a test function $\phi \in \mathcal{D}(\Omega)^2$ satisfies $(\operatorname{grad} d)(\phi) = -(d, \operatorname{div} \phi)_{\Omega_h} = (\operatorname{grad} d, \phi)_{\Omega_h} - \langle d, \phi \cdot n \rangle_{\partial \Omega_h}$ and the last term vanishes by (20c), so the distributional gradient is in $L^2(\Omega)^2$. It is also easy to see that the trace of d vanishes on $\partial \Omega$. Then, (20b) implies that $-\Delta d = g$. Next, consider $z \in H_0^1(\Omega)$. Equation (20a) with $y \in H_0^1(\Omega)$ yields $(\operatorname{grad} z, \operatorname{grad} y) = -(d, y)_{\Omega_h} - (\operatorname{grad} d, \operatorname{grad} y)_{\Omega_h} = -(d, y)_{\Omega_h} + (\Delta d, y)_{\Omega_h}$ which implies $\Delta z = d + g$. Finally, using the equations for z and d in (20a) and integrating by parts, we find $\langle \hat{z}_n, y \rangle_{\partial \Omega_h} = \langle n \cdot \operatorname{grad}(d + z), y \rangle_{\partial \Omega_h}$. Summarizing, the classical form of (20) is

$$-\Delta d = g, \qquad\qquad \text{on } \Omega, \tag{21a}$$
$$d = 0, \qquad\qquad \text{on } \partial\Omega, \tag{21b}$$
$$\Delta z = d + g, \qquad\qquad \text{on } \Omega, \tag{21c}$$
$$z = 0, \qquad\qquad \text{on } \partial\Omega, \tag{21d}$$
$$\hat{z}_n = n \cdot \operatorname{grad}(d + z), \qquad\qquad \text{on } \partial K, \forall K \in \Omega_h. \tag{21e}$$

Thus, by full regularity of the Dirichlet problem on a convex domain [12], d and z are in $H^2(\Omega)$, and moreover,

$$\|d\|_{Z_2} \leqslant C \|g\|_L,$$
$$\|z\|_{Z_1} \leqslant C \left(\|d\|_L + \|g\|_L \right) \leqslant C \|g\|_L,$$
$$\|\hat{z}_n\|_{\hat{X}} \leqslant \|\operatorname{grad}(d + z)\|_{H(\operatorname{div}, \Omega)}$$
$$= \|\operatorname{grad}(d + z)\|_L + \|\Delta(d + z)\|_L$$
$$= \|\operatorname{grad}(d + z)\|_L + \|d\|_L \qquad \text{by (21)},$$
$$\leqslant C \|g\|_L.$$

Hence

$$\|(z, \hat{z}, d)\|_Z \leqslant C \|g\|_L. \tag{22}$$

To complete the verification of Assumption 2.9, we now only need to bound some approximation errors. By the Bramble-Hilbert lemma,

$$\inf_{W \in X_{0,h} \times \hat{X}_h \times Y^r} \|U(g) - W\|_{X_0 \times \hat{X} \times Y}^2$$

$$= \inf_{w_h \in P_k(\Omega_h) \cap X_0} \|z - w_h\|_{H^1(\Omega)}^2 + \inf_{v_h \in P_{k+1}(\partial\Omega_h)} \|d - v_h\|_{H^1(\Omega_h)}^2 + \inf_{\hat{w}_h \in P_{k-1}(\partial\Omega_h) \cap \hat{X}} \|\hat{z}_n - \hat{w}_h\|_{\hat{X}}^2$$

(23)

$$\leqslant Ch^2 \left(|d|_{H^2(\Omega)}^2 + |z|_{H^2(\Omega)}^2 \right) + \inf_{r_h \in R_{k-1}} \|\operatorname{grad}(d + z) - r_h\|_{H(\operatorname{div},\Omega)}^2$$

where R_{k-1} is the Raviart-Thomas subspace [16] of $H(\operatorname{div}, \Omega)$ consisting of all vector functions which when restricted to an element takes the form $x p_1 + p_2$ for some $p_1 \in P_{k-1}(K)$ and some $p_2 \in P_{k-1}(K)^2$. Let Π_{RT}^h denote the Raviart-Thomas projection into R_{k-1}. By its well-known commutativity property with the L^2-projection Π_{k-1}^h onto $P_{k-1}(\Omega_h)$, we have

$$\inf_{r_h \in R_{k-1}} \|\operatorname{grad}(d + z) - r_h\|_{H(\operatorname{div},\Omega)} \leqslant \|(I - \Pi_{\mathrm{RT}}^h) \operatorname{grad}(d + z)\|_{H(\operatorname{div},\Omega)}$$

$$\leqslant \|(I - \Pi_{\mathrm{RT}}^h) \operatorname{grad}(d + z)\|_L + \|(I - \Pi_{k-1}^h)\Delta(d + z)\|_L$$

$$\leqslant \|(I - \Pi_{\mathrm{RT}}^h)\nabla(d + z)\|_L + \|(I - \Pi_{k-1}^h)d\|_L, \qquad \text{by (21),}$$

$$\leqslant Ch|d + z|_{H^2(\Omega)} + Ch|d|_{H^1(\Omega)},$$

where we used the Bramble-Hilbert lemma again in the final step. Hence using the regularity estimate (22),

$$\inf_{W \in X_{0,h} \times \hat{X}_h \times Y^r} \|U(g) - W\|_{X_0 \times \hat{X} \times Y} \leqslant Ch\|g\|_L,$$

thus verifying Assumption 2.9. Now, applying Theorem 2.10,

$$\|u - u_h\|_{L^2(\Omega)} \leqslant Ch \left(\|u - u_h\|_{H^1(\Omega)} + \|\hat{q}_n - \hat{q}_{n,h}\|_{H^{-1/2}(\partial\Omega_h)} + \|\varepsilon - \varepsilon^r\|_{H^1(\Omega_h)} \right)$$

where $\varepsilon = 0$ and ε^r is as in (4). This implies, by virtue of (6) in Remark 2.5,

$$\|u - u_h\|_{L^2(\Omega)} \leqslant Ch \left(\|u - u_h\|_{H^1(\Omega)} + \|\hat{q}_n - \hat{q}_{n,h}\|_{H^{-1/2}(\partial\Omega_h)} \right)$$

so the proof is finished using (19). □

3.2. **Case 2: Explaining the even-odd separation.** This case was not studied in previous works. We must first check if the DPG system is solvable for this case. For this, Theorem 2.8 is useful. Clearly, Assumption 2.7 holds – in fact, it holds for all the three cases: items (a) and (b) are obvious, while (c) follows by the Poincaré inequality. Hence, applying Theorem 2.8, we conclude that the DPG method in Case 2 is uniquely solvable if and only if $\hat{B}_h$ is injective.

Example 3.2. We begin with a negative result showing that $\hat{B}_h$ is not injective when $k = 2$. On a mesh consisting of a single element in the xy-plane, namely the unit triangle with vertices $a_0 = (0,0), a_1 = (1,0)$ and $a_2 = (0,1)$, we choose a basis for $\hat{X}_h$: Letting e_i denote the edge opposite to a_i and 1_{e_i} denote the indicator function of e_i, the basis is $(1_{e_2}, x|_{e_2}, 1_{e_1}, y|_{e_1}, 1_{e_0}/\sqrt{2}, x|_{e_0}/\sqrt{2})$. For the trial space Y^r, we choose the polynomial basis $(1, x, y, x^2, xy, y^2)$. The stiffness matrix

of the operator $\hat{B}_h$ with respect to these bases is

$$\begin{pmatrix} 1 & 1/2 & 1 & 1/2 & 1 & 1/2 \\ 1/2 & 1/3 & 0 & 0 & 1/2 & 1/3 \\ 0 & 0 & 1/2 & 1/3 & 1/2 & 1/6 \\ 1/3 & 1/4 & 0 & 0 & 1/3 & 1/4 \\ 0 & 0 & 0 & 0 & 1/6 & 1/12 \\ 0 & 0 & 1/3 & 1/4 & 1/3 & 1/12 \end{pmatrix},$$

whose determinant is zero. Hence, by theorem Theorem 2.8 the DPG method is not uniquely solvable in this example.

This example is closely related to a well-known result [8] that there is a nonzero quadratic function that is zero on the two Gauss-Legendre points (required for an exact integration of a third order polynomial) on each edge of a triangle. Clearly, such a quadratic function is orthogonal to all functions that are linear on each edge of the triangle.

We now show that for odd k, the situation is better.

Lemma 3.3. *Let K be a triangle and $k \geqslant 1$ be an odd integer. Any w in $P_k(K)$ satisfying*

$$\int_E w\, q\, ds = 0 \qquad\qquad \forall\, q \in P_{k-1}(E),\ \forall\ edges\ E \subset \partial K, \tag{24a}$$

$$\int_K w\, r\, dx = 0 \qquad\qquad \forall\, r \in P_{k-3}(K),\ if\ k \geqslant 3, \tag{24b}$$

must vanish on K.

Proof. Equation (24a) implies that $w|_E$ must be a scaled Legendre polynomial of degree exactly k on E. Since k is odd, this implies that the values of w at the endpoints of each edge must have opposite signs. This is impossible unless w vanishes on ∂K. But if $w|_{\partial K} = 0$, then $w \equiv 0$ if $k = 1$. If $k \geqslant 3$, then $w = \lambda_1\lambda_2\lambda_3 s_{k-3}$, for some $s_{k-3} \in P_{k-3}(K)$ where λ_i is the ith barycentric coordinate. Then (24b) implies $w \equiv 0$ on K. $\qquad\square$

Theorem 3.4. *In Case 2, for odd $k \geqslant 3$, these statements hold:*

  i) *The DPG method is uniquely solvable.*
  ii) *The solution $(u_h, \hat{q}_{n,h})$ of the DPG method satisfies*

$$\|u - u_h\|_{H^1(\Omega)} + \|\hat{q}_n - \hat{q}_{n,h}\|_{H^{-1/2}(\partial\Omega_h)} \leqslant Ch^{k-1}\left(|u|_{H^k(\Omega)} + |f|_{H^{k-1}(\Omega)}\right). \tag{25}$$

  iii) *If Ω is convex, then*

$$\|u - u_h\|_{L^2(\Omega)} \leqslant Ch^k\left(|u|_{H^k(\Omega)} + |f|_{H^{k-1}(\Omega)}\right). \tag{26}$$

Proof. By Theorem 2.3, if we verify Assumption 2.2, then the DPG method is uniquely solvable.

To do so, we first claim that there exists a $C_\Pi > 0$ and a unique $\Pi v \in P_k(K)$ for any $v \in H^1(K)$, such that

$$\int_E (v - \Pi v)q\, ds = 0 \qquad\qquad \forall\, q \in P_{k-1}(E),\ \forall\ edges\ E \subset \partial K, \tag{27a}$$

$$\int_K (v - \Pi v)r\, dx = 0 \qquad\qquad \forall\, r \in P_{k-3}(K) \tag{27b}$$

$$\|\Pi v\|_{H^1(K)} \leqslant C_\Pi \|v\|_{H^1(K)} \qquad\qquad \forall\, v \in H^1(K). \tag{27c}$$

It is easy to see that (27a)–(27b) forms a square system for Π, so existence of Πv follows from uniqueness. But uniqueness is already proved by Lemma 3.3. The estimate (27c) follows from a simple scaling argument.

The energy error estimate (25) now follows from Theorem 2.3 and (18). The L^2 error estimate (26) follows from Theorem 2.10: The required verification of Assumption 2.9 proceeds as in the proof of Theorem 3.1 – the only difference is in the degrees of approximation spaces in the first two infimums in (23), a difference that is inconsequential for the rest of the arguments. $\square$

Theorem 3.4 explains all entries in the second row of Table 1. The convergence rate in (25) is suboptimal and limited by the low degree of u_h. This motivates the next case.

3.3. **Case 3: A nonconforming analysis.**
The only difference between Case 2 and Case 3 is that the degree of u_h is increased by one. We analyze Case 3 using a technique of analysis different from the previous subsection, appealing to Theorem 2.6 and the second Strang lemma (see e.g. [4]) in the analyses of nonconforming methods.

Theorem 3.5. *In Case 3, for odd $k \geqslant 1$, these statements hold:*

   i) $\hat{B}_h$ *is injective and the DPG method is uniquely solvable.*

   ii) *The u_h-component of the solution satisfies*

$$\|u - u_h\|_{H^1(\Omega)} \leqslant Ch^k \left(|u|_{H^{k+1}(\Omega)} + |f|_{H^k(\Omega)} \right). \tag{28}$$

   iii) *If Ω is convex, then*

$$\|u - u_h\|_{L^2(\Omega)} \leqslant Ch^{k+1} \left(|u|_{H^{k+1}(\Omega)} + |f|_{H^k(\Omega)} \right). \tag{29}$$

Proof. First, observe that if $k \geqslant 3$, then by the unisolvency of the DPG method in Case 2, namely Theorem 3.4(i), its B_h is injective, which implies by Theorem 2.8 that $\hat{B}_h$ of Case 2 is injective. But since the flux $(\hat{X}_h)$ and test spaces (Y^r) of Case 3 are identical to that of Case 2, both cases have the same $\hat{B}_h$. Hence $\hat{B}_h$ of Case 3 is injective and consequently by Theorem 2.8, B_h of Case 3 is injective. Thus we have proved the first statement of the theorem for $k \geqslant 3$. For $k = 1$, if $(\hat{B}_h \hat{r}_{n,h})(w) = -\langle \hat{r}_{n,h}, w \rangle_{\partial\Omega_h} = 0$ for all $w \in Y^r$, then

$$\int_{\partial K} w\, \hat{r}_{n,h}\, ds = 0, \qquad \forall\, w \in P_k(K).$$

The matrix of this system (for $\hat{r}_{n,h}$) is the transpose of the matrix of (24) (for w), which is invertible by Lemma 3.3. Hence $\hat{r}_{n,h} = 0$, i.e., $\hat{B}_h$ is injective when $k = 1$.

Next we prove (28). Recall that Y^r is defined in (7) and $Y^r_{h,0}$ in (8b). By Theorem 2.6, $u_h \in X_{h,0}$ satisfies (8b), i.e.,

$$b_0(u_h, y) = (f, y)_\Omega, \qquad \forall y \in Y^r_{h,0}. \tag{30}$$

We proceed by viewing this as a nonconforming Petrov-Galerkin discretization of

$$b_0(u, y) = (f, y)_\Omega, \qquad \forall y \in H^1_0(\Omega)$$

and bounding the consistency error in an argument akin to the second Strang lemma. Let C_p denote the constant, derived from Poincaré inequality, such that $\|w\|_{H^1(\Omega)} \leqslant C_p \|\operatorname{grad} w\|_{L^2(\Omega)}$

for all $w \in H_0^1(\Omega)$. Then, for any $w_h \in X_{h,0}$

$$\|u_h - w_h\|_{H^1(\Omega)} \leqslant C_p \sup_{0 \neq z_h \in X_{h,0}} \frac{(\mathrm{grad}(u_h - w_h), \mathrm{grad}\, z_h)_\Omega}{\|\mathrm{grad}\, z_h\|_{L^2(\Omega)}} \leqslant C_p^2 \sup_{0 \neq z_h \in X_{h,0}} \frac{b_0(u_h - w_h, z_h)}{\|z_h\|_{H^1(\Omega)}}$$

$$\leqslant C_p^2 \sup_{0 \neq y \in Y_0^r} \frac{b_0(u_h - w_h, y)}{\|y\|_Y} = C_p^2 \|T_0^r(u_h - w_h)\|_Y = C_p^2 \sup_{0 \neq y \in Y_{h,0}^r} \frac{b_0(u_h - w_h, y)}{\|y\|_Y}$$

$$= C_p^2 \sup_{0 \neq y \in Y_{h,0}^r} \frac{b_0(u_h - u, y) + b_0(u - w_h, y)}{\|y\|_Y}$$

$$= C_p^2 \sup_{0 \neq y \in Y_{h,0}^r} \frac{(f, y)_\Omega - b_0(u, y) + b_0(u - w_h, y)}{\|y\|_Y}, \tag{31}$$

where we have used (30). Since $b((u, \hat{q}_n), y) = (f, y)_\Omega$ for all $y \in Y$, the term representing the consistency error in (31) can be written as $(f, y)_\Omega - b_0(u, y) = \hat{b}(\hat{q}_n, y)$. By the definition of Y_0^r (see (7)), we also have $\hat{b}(\hat{q}_n, y) = \hat{b}(\hat{q}_n - \hat{r}_{n,h}, y)$ for any $\hat{r}_{n,h} \in \hat{X}_h$ and $y \in Y_0^r$. Therefore,

$$\|u_h - w_h\|_{H^1(\Omega)} \leqslant C_p^2 \sup_{0 \neq y \in Y_{h,0}^r} \frac{b((u - w_h, \hat{q}_n - \hat{r}_{n,h}), y)}{\|y\|_Y} \leqslant C_p^2 C_2 C \left(\|\hat{q}_n - \hat{r}_{n,h}\|_{\hat{X}} + \|u - w_h\|_{H^1(\Omega)} \right).$$

Since $\hat{r}_{n,h}$ and $\hat{q}_n$ are element-by-element traces of an r_h in R_{k-1} and $q = \mathrm{grad}\, u$, respectively,

$$\|\hat{r}_{n,h} - \hat{q}_n\|_{\hat{X}} \leqslant \|r_h - \mathrm{grad}\, u\|_{H(\mathrm{div},\Omega)},$$

so

$$\|u_h - w_h\|_{H^1(\Omega)} \leqslant C \left(\inf_{r_h \in R_{k-1}} \|r_h - \mathrm{grad}\, u\|_{H(\mathrm{div},\Omega)} + \|u - w_h\|_{H^1(\Omega)} \right).$$

Finally, by the triangle inequality,

$$\|u - u_h\|_{H^1(\Omega)} \leqslant \|u - w_h\|_{H^1(\Omega)} + \|u_h - w_h\|_{H^1(\Omega)}$$

$$\leqslant C \left(\|u - w_h\|_{H^1(\Omega)} + h^k (|u|_{H^{k+1}(\Omega)} + |f|_{H^k(\Omega)}) \right)$$

for any $w_h \in X_{h,0}$. Choosing w_h to be an appropriate interpolant, the proof of (28) is finished.

The final estimate (29) is proved by verifying Assumption 2.9 (along the lines of the proof of Theorem 3.1) and applying Theorem 2.10.   □

The final row of Table 1 is now completely explained by Theorem 3.5.

4. Numerical Results

In this section, we report results from a numerical experiment. The presented DPG method for the Laplace equation was used to solve the Dirichlet problem with Ω set to the unit square. The function f was chosen so that the exact solution is $u = sin(\pi x)sin(\pi y)$. We construct an $n \times n$ uniform mesh by dividing Ω into n^2 congruent squares and further subdividing each square into two triangles by connecting the diagonal of positive slope. Its mesh size is $h = \sqrt{2}/n$. The method is applied on a sequence of such meshes with geometrically increasing n. The implementation of the method is done using FEniCS [13, 14]. Computed discretization errors in Cases 1, 2, and 3 are reported.

A baseline is provided by Case 1, reported in Table 2. The last column reports the rate of convergence in $L^2(\Omega)$, approximately calculated using two successive rows by $\log_2(\|u - u_h\|_{L^2(\Omega)}/\|u - u_{h/2}\|_{L^2(\Omega)})$. The $H^1(\Omega)$-convergence rate is computed similarly. We observe from the table that the $L^2(\Omega)$-rate is one order higher than the $H^1(\Omega)$-rate, as expected from Theorem 3.1.

TABLE 2. Case 1: $(k_u, k_q, k_v) = (k, k-1, k+1)$

n	$\|u - u_h\|_{H^1(\Omega)}$	rate	$\|u - u_h\|_{L^2(\Omega)}$	rate
		$k = 1$		
2	1.53E+00	0.86	2.61E-01	1.65
4	8.43E-01	0.96	8.33E-02	1.90
8	4.32E-01	0.99	2.23E-02	1.97
16	2.18E-01	1.00	5.67E-03	1.99
32	1.09E-01	1.00	1.42E-03	2.00
64	5.45E-02		3.57E-04	
		$k = 2$		
2	4.67E-01	1.85	3.24E-02	2.91
4	1.29E-01	1.95	4.31E-03	2.98
8	3.34E-02	1.99	5.47E-04	2.99
16	8.42E-03	2.00	6.87E-05	3.00
32	2.11E-03	2.00	8.60E-06	3.00
64	5.28E-04		1.08E-06	
		$k = 3$		
2	1.01E-01	2.94	5.52E-03	4.04
4	1.32E-02	3.00	3.36E-04	4.07
8	1.65E-03	3.01	2.00E-05	4.04
16	2.06E-04	3.00	1.22E-06	4.02
32	2.57E-05		7.50E-08	

TABLE 3. Case 2: $(k_u, k_q, k_v) = (k-1, k-1, k)$

n	$\|u - u_h\|_{H^1(\Omega)}$	rate	$\|u - u_h\|_{L^2(\Omega)}$	rate
		$k = 3$		
2	4.67E-01	1.85	3.24E-02	2.91
4	1.29E-01	1.95	4.31E-03	2.98
8	3.34E-02	1.99	5.47E-04	2.99
16	8.42E-03	2.00	6.87E-05	3.00
32	2.11E-03	2.00	8.60E-06	3.00
64	5.28E-04		1.08E-06	
		$k = 5$		
2	1.70E-02	3.92	7.24E-04	4.90
4	1.13E-03	3.98	2.43E-05	4.97
8	7.14E-05	4.00	7.76E-07	4.99
16	4.48E-06	4.00	2.44E-08	5.00
32	2.80E-07		7.64E-10	

Next, we consider Case 2, reported in Table 4. The table is computed similarly to Case 1, however only odd k are considered since the problem in Case 2 is not well posed for even k – see Example 3.2. We observe that the $H^1(\Omega)$-convergence is $O(h^{k-1})$, confirming the first theoretical estimate of Theorem 3.4. The rate of convergence is increased by one in the next column in accordance with the second estimate of Theorem 3.4.

Results from Case 3 are reported in Table 4. We observe that the $H^1(\Omega)$-convergence rate is $k+1$, the same as in Case 1, even though the test space is of a lesser degree. These observations illustrate and confirm the theoretical results of Theorem 3.5.

Other possibilities exist besides the three cases investigated, so, as a caveat, we present observations of suboptimal convergence in the case $(k_u, k_q, k_v) = (3, 0, 3)$. The DPG method is

TABLE 4. Case 3: $(k_u, k_q, k_v) = (k, k-1, k)$

n	$\|u - u_h\|_{H^1(\Omega)}$	rate	$\|u - u_h\|_{L^2(\Omega)}$	rate
	$k = 1$			
2	1.59E+00	0.87	3.08E-01	1.38
4	8.71E-01	0.99	1.18E-01	1.82
8	4.37E-01	1.00	3.34E-02	1.95
16	2.18E-01	1.00	8.63E-03	1.99
32	1.09E-01	1.00	2.18E-03	2.00
64	5.45E-02		5.45E-04	
	$k = 3$			
2	1.01E-01	2.94	5.38E-03	3.93
4	1.32E-02	3.00	3.53E-04	4.02
8	1.66E-03	3.01	2.18E-05	4.02
16	2.06E-04	3.00	1.34E-06	4.01
32	2.57E-05	3.00	8.32E-08	4.00
64	3.21E-06		5.19E-09	
	$k = 5$			
2	2.45E-03	4.94	8.82E-05	5.89
4	7.94E-05	5.00	1.49E-06	5.98
8	2.49E-06	5.00	2.36E-08	6.00
16	7.77E-08	5.00	3.69E-10	6.01
32	2.42E-09		5.71E-12	

TABLE 5. Poor $H^1(\Omega)$ and $L^2(\Omega)$ convergence for the case $(k_u, k_q, k_v) = (k, k-3, k)$

n	$\|u - u_h\|_{H^1(\Omega)}$	rate	$\|u - u_h\|_{L^2(\Omega)}$	rate
	$k = 3$			
2	1.02E-01	2.85	6.68E-03	2.50
4	1.42E-02	2.70	1.18E-03	1.99
8	2.18E-03	2.38	3.14E-04	1.96
16	4.21E-04	2.13	8.06E-05	1.99
32	9.59E-05		2.03E-05	

uniquely solvable in this case: This would follow from Theorem 2.8 once we prove that $\hat{B}_h$ is injective. If $\hat{B}_h \hat{z}_n = 0$, then by definition (10), $\hat{b}(\hat{z}_n, v) = 0$ for all $v \in P_3(\Omega_h)$, so in particular,

$$\hat{z}_n \in P_0(\partial \Omega_h): \quad \hat{b}(\hat{z}_n, v) = 0, \quad \forall\, v \in P_2(\Omega_h).$$

This implies, by the already known unisolvency of Case 1 with $k = 1$, i.e., $(k_u, k_q, k_v) = (1, 0, 2)$, and Theorem 2.8, that $\hat{z}_n = 0$. Therefore, the method is well-defined for the $(k_u, k_q, k_v) = (3, 0, 3)$ case. Yet, the theory we presented does not guarantee optimal convergence rates in this case. The numerical results reported in Table 5 show that the practically observed convergence rates in $H^1(\Omega)$ and $L^2(\Omega)$ are indeed suboptimal in this case. In fact, we observe second order convergence in $L^2(\Omega)$ as in case 1 with $k = 1$. An error analysis that proceeds exactly like the error analysis of case 3 will predict this suboptimal rate (the rate being limited by the order k_q of $\hat{X}_h$). However, the practically observed $H^1(\Omega)$ rates are higher than what the same analysis would predict.

References

[1] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Number 15 in Springer Series in Computational Mathematics. Springer-Verlag, New York, 1991.

[2] D. Broersen and R. Stevenson. A Petrov-Galerkin discretization with optimal test space of a mild-weak formulation of convection-diffusion equations in mixed form. *Preprint*, 2013.

[3] C. Carstensen, L. Demkowicz, and J. Gopalakrishnan. *A posteriori* error control for DPG methods. *Preprint*, 2013.

[4] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland Publishing Company, Amsterdam, 1978.

[5] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. *SIAM J Numer. Anal.*, 50(5):2420–2445, 2012.

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

[7] L. Demkowicz and J. Gopalakrishnan. A primal DPG method without a first-order reformulation. *Computers and Mathematics with Applications*, 66(6):1058–1064, 2013.

[8] M. Fortin and M. Soulie. A non-conforming piecewise quadratic finite element on triangles. *International Journal for Numerical Methods in Engineering*, 19(4):505–520, doi: 10.1002/nme.1620190405, 1983.

[9] J. Gopalakrishnan. Five lectures on DPG methods. Available as *arXiV preprint* 1306.0557, 2013.

[10] J. Gopalakrishnan, I. Muga, and N. Olivares. Dispersive and dissipative errors in the DPG method with scaled norms for the Helmholtz equation. *SIAM J. Sci. Comput.*, 36 (2014), pp. A20–A39.

[11] J. Gopalakrishnan and W. Qiu. An analysis of the practical DPG method. *Math. Comp*, 83(286) (2014 (electronically appeared 2013)), pp. 537–552.

[12] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Number 24 in Monographs and Studies in Mathematics. Pitman Advanced Publishing Program, Marshfield, Massachusetts, 1985.

[13] A. Logg, K. -A. Mardal, G. N. Wells, et al. *Automated Solution of Differential Equations by the Finite Element Method* Springer 978-3-642-23098-1 doi: 10.1007/978-3-642-23099-8, 2012.

[14] A. Logg and G. N. Wells, et al. DOLFIN: Automated Finite Element Computing. *ACM Transactions on Mathematical Software*, 37(2) Available as arXiV preprint 1103.6248, doi: 10.1145/1731022.1731030, 2010.

[15] J. Nitsche. Ein Kriterium für die Quasi-Optimalität des Ritzschen Verfahrens. *Numer. Math.*, 11:346–348, 1968.

[16] P.-A. Raviart and J. M. Thomas. Primal hybrid finite element methods for 2nd order elliptic equations. *Math. Comp.*, 31(138):391–413, 1977.

Address for all authors: PO Box 751, Portland State University, Portland, OR 97207-0751.