# AN ANALYSIS OF THE PRACTICAL DPG METHOD

J. GOPALAKRISHNAN AND W. QIU

ABSTRACT. We give a complete error analysis of the Discontinuous Petrov Galerkin (DPG) method, accounting for all the approximations made in its practical implementation. Specifically, we consider the DPG method that uses a trial space consisting of polynomials of degree $p$ on each mesh element. Earlier works showed that there is a "trial-to-test" operator $T$, which when applied to the trial space, defines a test space that guarantees stability. In DPG formulations, this operator $T$ is local: it can be applied element-by-element. However, an infinite dimensional problem on each mesh element needed to be solved to apply $T$. In practical computations, $T$ is approximated using polynomials of some degree $r > p$ on each mesh element. We show that this approximation maintains optimal convergence rates, provided that $r \geq p + N$, where $N$ is the space dimension (two or more), for the Laplace equation. We also prove a similar result for the DPG method for linear elasticity. Remarks on the conditioning of the stiffness matrix in DPG methods are also included.

## 1. INTRODUCTION

In this paper we prove error estimates for the discontinuous Petrov-Galerkin (DPG) method applied to the Laplace equation and the equations of linear elasticity. The approach is applicable more generally to other equations as well. An error analysis of an "ideal" DPG method was provided in [5]. Although the ideal method is not practically implementable, a number of important theoretical tools for analysis were developed in [5]. We extend this analysis using a few new lemmas to provide a complete analysis of the fully implementable "practical" DPG method. The distinction between the ideal and practical methods will be clear in the next few paragraphs.

Both methods are easy to describe in a general context. Suppose we want to approximate $\mathscr{u} \in U$ satisfying

$$b(\mathscr{u}, \mathscr{v}) = l(\mathscr{v}), \quad \forall \mathscr{v} \in V. \tag{1.1}$$

Here $U$ is a Hilbert space with norm $\|\cdot\|_U$ and $V$ is a Hilbert space under an inner product $(\cdot,\cdot)_V$ with corresponding norm $\|\cdot\|_V$. (All spaces are over $\mathbb{R}$.) We assume that the bilinear form $b(\cdot,\cdot) : U \times V \mapsto \mathbb{R}$ is continuous and the linear form $l(\cdot) : V \mapsto \mathbb{R}$ is also continuous. Define $T : U \mapsto V$ by

$$(T\mathscr{w}, \mathscr{v})_V = b(\mathscr{w}, \mathscr{v}), \qquad \forall \mathscr{v} \in V. \tag{1.2}$$

Then, the DPG approximation to $\mathscr{u}$, lies in a finite dimensional trial subspace $U_h \subset U$ (where $h$ denotes a parameter determining the finite dimension). It satisfies

$$b(\mathscr{u}_h, \mathscr{v}) = l(\mathscr{v}), \quad \forall \mathscr{v} \in V_h, \tag{1.3}$$

2000 *Mathematics Subject Classification.* 65N30, 65L12.

*Key words and phrases.* discontinuous Galerkin, Petrov-Galerkin, DPG method, ultraweak formulation.

Corresponding author: Weifeng Qiu (qiuxa001@ima.umn.edu).

This work was partly supported by the NSF under grants DMS-1211635 and DMS-1014817. The authors gratefully acknowledge the collaboration opportunities provided by the IMA (Minneapolis) during their 2010-11 program.

where $V_h = T(U_h)$. Since $U_h \neq V_h$ in general, this is a Petrov-Galerkin approximation. The method (1.3) is the *ideal DPG method*. The excellent stability and approximation properties of this method are well known [4, 5].

The main difficulty of the ideal method is that in order to compute $\mathscr{U}_h$, one needs a basis for $V_h$, which must be obtained by applying $T$. This is infeasible, as seen from (1.2), if $V$ is infinite dimensional, unless a solution to (1.2) can be written out in closed form. In certain one-dimensional problems, and in some multi-dimensional problems like the transport equation, the application of $T$ can be exactly written out in closed form (see [3, 4]). But for the vast majority of interesting problems, this is not possible.

Yet, one may approximate $T$ by $T^r$, defined as follows. Let $V^r$ be a finite dimensional subspace of $V$ (where $r$ is a parameter determining the finite dimension). Then $T^r\mathscr{W}$ in $V^r$ is defined by

$$(T^r\mathscr{W}, \mathscr{V})_V = b(\mathscr{W}, \mathscr{V}), \qquad \forall \mathscr{V} \in V^r. \tag{1.4}$$

One can then reconsider the DPG method (1.3) with $V_h^r = T^r(U_h)$ in place of $V_h$, i.e.,

$$b(\mathscr{U}_h^r, \mathscr{V}) = l(\mathscr{V}), \quad \forall \mathscr{V} \in V_h^r. \tag{1.5}$$

This yields an implementable method that is very generally applicable. We refer to this method as the *practical DPG method*.

A serious difficulty still remains when these ideas are applied to standard variational problems. Namely, one application of $T^r$ requires inverting a Gram matrix in the $V$-inner product. This is prohibitively expensive for most standard variational formulations. For instance, if $V$ is $H^1(\varOmega)$, where $\varOmega$ is the computational domain, then inverting the Gram matrix is as expensive as solving the Laplace's equation.

This difficulty can be overcome by hybridization, as shown in the earlier DPG papers [4, 5]. Namely, given a boundary value problem, introducing certain interelement fluxes and traces as new unknowns, we can design an *ultraweak* well-posed variational formulation involving a space $V$ that contains functions *discontinuous* across mesh element interfaces. This then implies that the Gram matrix becomes block diagonal, with one block per mesh element (since $V^r$ may now be chosen to be a DG subspace). The application of $T^r$ is thus reduced to an easy block diagonal inversion, i.e., the action of the operator $T^r$ is *local*.

Such an ultraweak variational formulation has been developed for the Poisson equation in [5], where its wellposedness is also proved. We will heavily rely on such wellposedness results in this paper. An ultraweak formulation for the linear elasticity system is also available now [2]. Both these works analyzed the ideal DPG method (1.3) for the respective ultraweak formulations. The aim of the present paper is to provide an error analysis for the corresponding practical DPG methods (1.5).

In the next section we will present an approach to the analysis of the practical method, continuing in the general context and using the abstract notations introduced above. In Section 3, we perform the error analysis for the practical DPG method for the Laplace equation. We also provide a condition number estimate. In Section 4, we consider a second example of linear elasticity and provide an error analysis.

## 2. The approach to analysis

The purpose of this section is to point out a simple functional analytic route to proving the discrete stability of the practical DPG method (1.5). This discrete stability will follow from a discrete inf-sup condition on the space $V_h^r$. However, in applications, it is often inconvenient to work directly with this space. We prefer to work with $V^r$, which will be some standard polynomial space in most applications (as seen in the examples later). The next theorem shows that the existence of a Fortin operator into this more standard space $V^r$ is a sufficient condition for the convergence of the practical DPG method.

Before we give the result, let us state the assumptions that we shall verify for each of our examples. We assume that

$$\{\mathscr{w} \in U :\ b(\mathscr{w}, \mathscr{v}) = 0,\ \forall \mathscr{v} \in V\} = \{0\} \tag{2.1}$$

and that there is a positive constant $C_1$ such that

$$C_1 \|\mathscr{v}\|_V \le \sup_{\mathscr{w} \in U} \frac{b(\mathscr{w}, \mathscr{v})}{\|\mathscr{w}\|_U}, \qquad \forall \mathscr{v} \in V. \tag{2.2}$$

Above and throughout, we will tacitly assume that the suprema such as the above are taken over nonzero functions. Let $C_2 \ge 0$ be such that

$$b(\mathscr{w}, \mathscr{v}) \le C_2 \|\mathscr{w}\|_U \|\mathscr{v}\|_V, \qquad \forall \mathscr{w} \in U,\ \mathscr{v} \in V. \tag{2.3}$$

Clearly, such a $C_2$ exists due to the continuity of $b(\cdot, \cdot)$. Finally, assume that there exists a linear operator $\Pi : V \mapsto V^r$ such that for all $\mathscr{v} \in V$, we have

$$b(\mathscr{w}, \mathscr{v} - \Pi \mathscr{v}) = 0, \qquad \forall \mathscr{w} \in U_h, \tag{2.4a}$$

$$\|\Pi \mathscr{v}\|_V \le C_\Pi \|\mathscr{v}\|_V. \tag{2.4b}$$

**Theorem 2.1.** *Suppose the assumptions* (2.1)*,* (2.2)*,* (2.3)*, and* (2.4) *hold. Then the problem* (1.1) *is well-posed and*

$$\|\mathscr{u} - \mathscr{u}_h^r\|_U \le \frac{C_2 C_\Pi}{C_1} \inf_{\mathscr{w} \in U_h} \|\mathscr{u} - \mathscr{w}\|_U. \tag{2.5}$$

*Proof.* We apply Babuška's theory [1, 11]. Accordingly, if we prove the discrete inf-sup condition

$$\frac{C_1}{C_\Pi} \|\mathscr{w}\|_U \le \sup_{\mathscr{v} \in V_h^r} \frac{b(\mathscr{w}, \mathscr{v})}{\|\mathscr{v}\|_V}, \qquad \forall \mathscr{w} \in U_h, \tag{2.6}$$

then (2.5) will follow. We prove (2.6) in three steps, the first two of which are fairly standard (but included for readability).

As the first step, we prove that the following inf-sup condition holds:

$$C_1 \|\mathscr{w}\|_U \le \sup_{\mathscr{v} \in V} \frac{|b(\mathscr{w}, \mathscr{v})|}{\|\mathscr{v}\|_V}, \qquad \forall \mathscr{w} \in U. \tag{2.7}$$

This follows from the other inf-sup condition (2.2). Define a linear operator $B : U \to V^*$ by $B\mathscr{w} = b(\mathscr{w}, \cdot) \in V^*$, for all $\mathscr{w} \in U$. It is well known [12] that (2.2) holds if and only if $B^*$ is injective and the range of $B^*$ is closed in $U^*$. Additionally, by (2.1), $B$ is injective. Therefore, by the Closed Range Theorem, $B^*(V) = U^*$, so $(B^*)^{-1} : U^* \to V$ exists. Hence $B^{-1} : V^* \to U$ also exists and is continuous. This proves that problem (1.1) is *well-posed.* We obviously also have $\|B^{-1}\| = \|(B^{-1})^*\| = \|(B^*)^{-1}\|$, i.e.,

$$\inf_{\mathscr{w} \in U} \sup_{\mathscr{v} \in V} \frac{|b(\mathscr{w}, \mathscr{v})|}{\|\mathscr{w}\|_U \|\mathscr{v}\|_V} = \|B^{-1}\|^{-1} = \|(B^*)^{-1}\|^{-1} = \inf_{\mathscr{v} \in V} \sup_{\mathscr{w} \in U} \frac{|b(\mathscr{w}, \mathscr{v})|}{\|\mathscr{w}\|_U \|\mathscr{v}\|_V},$$

which proves (2.7).

As the second step, we prove the following inf-sup condition.

$$\frac{C_1}{C_\Pi}\|\mathscr{W}\|_U \le \sup_{\mathscr{V}\in V^r} \frac{|b(\mathscr{W},\mathscr{V})|}{\|\mathscr{V}\|_V}, \qquad \forall \mathscr{W} \in U_h. \tag{2.8}$$

Note that this differs from (2.6) only in the space in which the supremum is sought. To prove (2.8), we use (2.7) and assumption (2.4) as follows:

$$C_1\|\mathscr{W}\|_U \le \sup_{\mathscr{V}\in V} \frac{b(\mathscr{W},\mathscr{V})}{\|\mathscr{V}\|_V} \le \sup_{\mathscr{V}\in V} \frac{b(\mathscr{W},\Pi\mathscr{V})}{C_\Pi^{-1}\|\Pi\mathscr{V}\|_V}.$$

Now, since $\Pi\mathscr{V}$ is in $V^r$, the last supremum may be bounded by the supremum over all $V^r$, so we obtain (2.8).

As the third and final step, we prove that if $s_1$ is the supremum in (2.8) and $s_2$ is the supremum in (2.6), then $s_1 = s_2$. Obviously, $s_1 \ge s_2$ as $V^r \supseteq V_h^r$. To prove the reverse inequality, observe that $s_1 = \|T^r\mathscr{W}\|_V$, by (1.4). Since $T^r\mathscr{W}$ is in $V_h^r$, we have

$$s_1 = \frac{(T^r\mathscr{W}, T^r\mathscr{W})_V}{\|T^r\mathscr{W}\|_V} \le \sup_{\mathscr{V}\in V_h^r} \frac{(T^r\mathscr{W},\mathscr{V})_V}{\|\mathscr{V}\|_V} = s_2.$$

Therefore, the inf-sup condition (2.6) follows from (2.8). □

*Remark* 2.2 (Test basis). The above proof also shows that under the assumptions of Theorem 2.1, the operator $T^r : U_h \mapsto V^r$ is *injective*: indeed, if $T^r\mathscr{W} = 0$, then $b(\mathscr{W},\mathscr{V}) = 0$ for all $\mathscr{V}$ in $V^r$, so by the inf-sup condition (2.8), we conclude that $\mathscr{W} = 0$. Note that the injectivity of $T^r$ implies that

$$\dim(V_h^r) = \dim(U_h).$$

It also implies that a basis for $V_h^r$ can be computed by applying $T^r$ on any basis for $U_h$.

*Remark* 2.3 (Conditioning). Suppose $\mathscr{B}_i$ is a basis for $U_h$. Then, under the assumptions of Theorem 2.1, $T^r\mathscr{B}_i$ is a basis for $V_h^r$, as seen in Remark 2.2. The $ij$-th entry of the *stiffness matrix* of the DPG method with respect to this basis is given by $S_{ij} = b(\mathscr{B}_j, T^r\mathscr{B}_i) = (T^r\mathscr{B}_j, T^r\mathscr{B}_i)_V$. Clearly, $S$ is symmetric. The above mentioned injectivity of $T^r$ implies that $S$ is also positive definite. To understand the conditioning of $S$, let us first note that

$$\frac{C_1}{C_\Pi}\|\mathscr{W}\|_U \le \|T^r\mathscr{W}\|_V \le C_2\|\mathscr{W}\|_U, \qquad \forall \mathscr{W}\in U_h. \tag{2.9}$$

This follows from the inf-sup condition (2.8) in the proof of Theorem 2.1 and the continuity property (2.3). Next, suppose $\mathscr{X} = \sum_i x_i\mathscr{B}_i$ is the basis expansion of any $\mathscr{X}$ in $U_h$, and $\lambda_0, \lambda_1$ are positive numbers such that

$$\lambda_0\|x\|_{\ell^2}^2 \le \|\mathscr{X}\|_U^2 \le \lambda_1\|x\|_{\ell^2}^2, \qquad \forall \mathscr{X}\in U_h. \tag{2.10}$$

Since $x^tSx = \|T^r\mathscr{X}\|_U^2$, these estimates imply that the Rayleigh quotient $x^tSx/x^tx$ is at most $\lambda_1C_2^2$ and at least $C_1^2\lambda_0/C_\Pi^2$. Hence

$$\kappa(S) \le \frac{\lambda_1}{\lambda_0}\frac{C_2^2C_\Pi^2}{C_1^2}, \tag{2.11}$$

where $\kappa(S)$ is the spectral condition number of $S$. This gives condition numbers comparable to other methods, as we shall see later in our examples.

## 3. First example: Laplace equation

The ideal DPG method for the Laplace equation was developed and analyzed in [5]. In this section, we will set the abstract forms and spaces of the previous section to those from [5] and verify the hypotheses required to apply Theorem 2.1. Roughly speaking, our main result shows that if polynomials of degree $p$ are used to approximate the solution of the Laplace equation, then a sufficient condition for optimal convergence is that $T$ is approximated by polynomials of degree $p+N$, where $N \geq 2$ is the space dimension. In the wording of [5], this means the "enrichment degree" should be chosen to be $N$. Ample numerical evidence, in support of the choice of 2 as enrichment degree, was presented in [5, § 6.1], but all numerical experiments were in the two-dimensional case.

In the remainder of this paper, we let $\Omega$ be a Lipschitz polyhedron in $\mathbb{R}^N$. We denote by $\{\Omega_h\}_{h\in I}$ a family of conforming shape regular simplicial finite element triangulations of $\Omega$. The index $h$ now stands for the maximal diameter of simplexes in $\Omega_h$.

3.1. **Infinite dimensional spaces.** Let $\mathbb{V} = \mathbb{R}^N$. We use $L^2(\Omega, \mathbb{V})$ to denote the set of vector-valued functions whose components are square integrable. We set the trial and test spaces by

$$\begin{aligned} U &= L^2(\Omega;\mathbb{V}) \times L^2(\Omega) \times H_0^{1/2}(\partial\Omega_h) \times H^{-1/2}(\partial\Omega_h),\\ V &= H(\operatorname{div}, \Omega_h) \times H^1(\Omega_h), \end{aligned}$$

where the "broken" Sobolev spaces (admitting interelement discontinuities) are defined by $H(\operatorname{div},\Omega_h) = \{\tau :\ \tau|_K \in H(\operatorname{div}, K),\ \forall K \in \Omega_h\}$ and $H^1(\Omega_h) = \{v :\ v|_K \in H^1(K),\ \forall K \in \Omega_h\}$. They have the natural norms

$$\begin{aligned} \|v\|^2_{H^1(\Omega_h)} &= (v, v)_{\Omega_h} + (\operatorname{grad} v, \operatorname{grad} v)_{\Omega_h},\\ \|q\|^2_{H(\operatorname{div},\Omega_h)} &= (q, q)_{\Omega_h} + (\operatorname{div} q, \operatorname{div} q)_{\Omega_h}. \end{aligned}$$

The derivatives above, and in such notations throughout, are calculated element by element and

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

and $\langle\cdot, \ell\rangle_{1/2,\partial K}$ denotes the action of a functional $\ell$ in $H^{-1/2}(\partial K)$. We will also use $\|r\|_{\Omega_h}$ to denote the norm $(r, r)^{1/2}_{\Omega_h}$. The spaces of traces and fluxes are defined by $H_0^{1/2}(\partial\Omega_h) = \{\eta \in \prod_K H^{1/2}(\partial K) : \exists\, w \in H_0^1(\Omega)$ such that $\eta|_{\partial K} = w|_{\partial K}\ \forall K \in \Omega_h\}$, and $H^{-1/2}(\partial\Omega_h) = \{\eta \in \prod_K H^{-1/2}(\partial K) : \exists\, q \in H(\operatorname{div}, \Omega)$ such that $\eta|_{\partial K} = q \cdot n|_{\partial K}\ \forall K \in \Omega_h\}$, with respective norms

$$\|\hat u\|_{H_0^{1/2}(\partial\Omega_h)} = \inf\big\{\|w\|_{H^1(\Omega)} :\ \forall w \in H_0^1(\Omega) \text{ such that } \hat u|_{\partial K} = w|_{\partial K}\big\}, \tag{3.1}$$

$$\|\hat\sigma_n\|_{H^{-1/2}(\partial\Omega_h)} = \inf\big\{\|q\|_{H(\operatorname{div},\Omega)} :\ \forall q \in H(\operatorname{div}, \Omega) \text{ such that } \hat\sigma_n|_{\partial K} = q\cdot n|_{\partial K}\big\}. \tag{3.2}$$

The spaces $U$ and $V$ are endowed with product norms, i.e.,

$$\begin{aligned} \|(\sigma, u, \hat u, \hat\sigma_n)\|^2_U &= \|\sigma\|^2_\Omega + \|u\|^2_\Omega + \|\hat u\|^2_{H^{1/2}(\partial\Omega_h)} + \|\hat\sigma_n\|^2_{H^{-1/2}(\partial\Omega_h)},\\ \|(\tau, v)\|^2_V &= \|\tau\|^2_{H(\operatorname{div},\Omega_h)} + \|v\|^2_{H^1(\Omega_h)}. \end{aligned}$$

3.2. **Forms.** The ultraweak formulation of the Laplace equation derived in [5] reads as follows: Find $\mathscr{U} \equiv (\sigma, u, \hat{u}, \hat{\sigma}_n) \in U$ satisfying (1.1) for $\mathscr{V} \equiv (\tau, v) \in V$ where the forms $b(\cdot,\cdot)$ and $l(\cdot)$ are set by

$$
\begin{aligned}
b(\mathscr{U},\mathscr{V}) &= (\sigma,\tau)_\Omega - (u, \operatorname{div}\tau)_{\Omega_h} + \langle \hat{u}, \tau\cdot n\rangle_{\partial\Omega_h} - (\sigma, \operatorname{grad} v)_{\Omega_h} + \langle v, \hat{\sigma}_n\rangle_{\partial\Omega_h},\\
l(\mathscr{V}) &= (f, v)_\Omega,
\end{aligned}
$$

for some $f$ in $L^2(\Omega)$. The $u$-component of $\mathscr{U}$ solves the Laplace equation with zero Dirichlet boundary conditions on $\partial\Omega$. For details, consult [5].

3.3. **Discrete spaces.** Let us first establish notation for a few polynomial spaces that we will use here and throughout. Let $P_p(K)$ denote the space of polynomials of degree at most $p$ on a simplex $K$. We write $P_p(K;\mathbb{V})$ for vector valued functions whose components are in $P_p(K)$. Let $\triangle_m(K)$ denote the set of all $m$-dimensional sub-simplices of $K$. Define

$$
\begin{aligned}
\mathring{P}_p(K) &= \{p_p \in P_p(K):\ p_p|_{\partial K} = 0\},\\
P_p(\partial K) &= \{\mu : \mu|_F \in P_p(F),\ \forall F \in \triangle_{N-1}(K)\},\\
\tilde{P}_p(\partial K) &= P_p(\partial K) \cap \mathcal{C}^0(\partial K),
\end{aligned}
$$

where $\mathcal{C}^0(D)$ denotes the set of continuous functions on any domain $D$.

Using these notations, we set the trial approximation space for the DPG method by

$$
\begin{aligned}
U_h = \{(\sigma, u, \hat{u}, \hat{\sigma}_n) \in U\colon\ \ & \sigma|_K \in P_p(K;\mathbb{V}),\ u|_K \in P_p(K),\\
& \hat{u}|_{\partial K} \in \tilde{P}_{p+1}(\partial K),\ \hat{\sigma}_n|_{\partial K} \in P_p(\partial K),\ \ \forall K \in \Omega_h\}.
\end{aligned}
$$

The discrete test space is defined by $V_h^r = T^r(U_h)$, so to complete the prescription of the practical DPG method, we only need to specify $V^r$. Set

$$
V^r = \{(\tau, v) \in V:\ \tau|_K \in P_r(K;\mathbb{V}),\ v|_K \in P_r(K),\ \forall K \in \Omega_h\}. \tag{3.3}
$$

where the degree $r \ge p + N$. Clearly, the application of $T^r$, as defined by (1.4), can proceed locally, element by element, since $V^r$ has no interelement continuity constraints.

3.4. **Verification of the assumptions.** To apply Theorem 2.1 to the above setting, we need to verify its assumptions.

- Assumption (2.1) is verified by [5, Lemma 4.1].
- Assumption (2.2) is verified by [5, Theorem 4.2].
- Assumption (2.3) is easy to verify. For example, to show the continuity of the term $\langle \hat{u}, \tau\cdot n\rangle_{\partial\Omega_h}$, we let $w \in H^1(\Omega)$ be any extension of $\hat{u}$ and observe that

  $$\langle \hat{u}, \tau\cdot n\rangle_{\partial\Omega_h} = (\operatorname{grad} w, \tau)_{\Omega_h} + (w, \operatorname{div}\tau)_{\Omega_h} \le \|w\|_{H^1(\Omega)} \|\tau\|_{H(\operatorname{div},\Omega_h)}.$$

  Taking the infimum over all such extensions $w$, we obtain

  $$\langle \hat{u}, \tau\cdot n\rangle_{\partial\Omega_h} \le \|\hat{u}\|_{H_0^{1/2}(\partial\Omega_h)} \|\tau\|_{H(\operatorname{div},\Omega_h)}.$$

  The other terms in the bilinear form are similar or simpler.
- Assumption (2.4) is verified below.

An operator $\Pi$ satisfying (2.4) will be constructed in the form $\Pi\mathscr{V} = (\Pi_{p+2}^{\operatorname{div}}\tau, \Pi_r^{\operatorname{grad}} v)$. We construct the operator $\Pi_r^{\operatorname{grad}}$ in Lemma 3.2 below, and we construct the operator $\Pi_{p+2}^{\operatorname{div}}$ in Lemma 3.3. But first, we need the following intermediate result. Let $B_r^{\operatorname{grad}}(K) = \{p_r \in P_r(K) : p_r|_E = 0, \forall E \in \triangle_{N-2}(K)\}$ and $h_K = \operatorname{diam}(K)$. Hereon we use $c$ and $C$ to denote generic constants (whose value at different occurrences may differ) independent of $h_K$, but possibly dependent on the shape regularity of $K$ and the polynomial degree $p$. We also let $\langle \cdot, \cdot \rangle_{\partial K}$ denote the $L^2(\partial K)$-inner product.

**Lemma 3.1.** *Let $r = p + N$. Then, for every $v \in H^1(K)$, there is a unique $\varPi_r^0 v \in B_r^{\mathrm{grad}}(K)$ satisfying*

$$(\varPi_r^0 v - v, q_{p-1})_K = 0, \quad \forall q_{p-1} \in P_{p-1}(K), \tag{3.4a}$$

$$\langle \varPi_r^0 v - v, \mu_p \rangle_{\partial K} = 0, \quad \forall \mu_p \in P_p(\partial K), \tag{3.4b}$$

$$\|\varPi_r^0 v\|_K + h_K \|\operatorname{grad} \varPi_r^0 v\|_K \le C \left( \|v\|_K + h_K \|\operatorname{grad} v\|_K \right). \tag{3.4c}$$

*Proof.* First, to see that the number of the equations in (3.4a)-(3.4b) equal $\dim B_r^{\mathrm{grad}}(K)$, observe that

$$\dim B_r^{\mathrm{grad}}(K) = \dim \mathring{P}_r(K) + \sum_{F \in \triangle_{N-1}(K)} \dim \mathring{P}_r(F). \tag{3.5}$$

Let $b_K$ and $b_F$ denote the product of all barycentric coordinates that do not vanish everywhere on $K$ and $F$, resp. Then $\mathring{P}_r(K) = b_K P_{r-N-1}(K)$ and $\mathring{P}_r(F) = b_F P_{r-N}(F)$. Therefore, by our choice of $r$, we have $\dim \mathring{P}_r(K) = \dim P_{p-1}(K)$ and $\dim \mathring{P}_r(F) = \dim P_p(F)$. It then follows from (3.5) that (3.4a)-(3.4b) is a square system for $\varPi_r^0 v$.

Hence, to prove that (3.4a)-(3.4b) has a unique solution, it suffices to prove that if $v = 0$, then $\varPi_r^0 v = 0$. Since $\varPi_r^0 v \in B_r^{\mathrm{grad}}(K)$, on any face $F \in \triangle_{N-1}(K)$, we may write $(\varPi_r^0 v)|_F = b_F w_p$ for some $w_p \in P_p(F)$. But then, (3.4b) implies that $\varPi_r^0 v$ must vanish on $\partial K$, so $\varPi_r^0 v = b_K z_{p-1}$ for some $z_{p-1} \in P_{p-1}(K)$. Then (3.4a) implies that $\varPi_r^0 v = 0$ on $K$.

Finally, one can prove (3.4c) using a standard affine mapping argument. □

**Lemma 3.2.** *Let $r = p + N$. Define $\varPi_r^{\mathrm{grad}} v = \varPi_r^0 (v - \overline{v}) + \overline{v}$, where $\overline{v}|_K = |K|^{-1} \int_K v$. Then*

$$(\varPi_r^{\mathrm{grad}} v - v, q_{p-1})_K = 0, \qquad \forall q_{p-1} \in P_{p-1}(K), \tag{3.6a}$$

$$\langle \varPi_r^{\mathrm{grad}} v - v, \mu_p \rangle_{\partial K} = 0, \qquad \forall \mu_p \in P_p(\partial K), \tag{3.6b}$$

$$\|\varPi_r^{\mathrm{grad}} v\|_{H^1(K)} \le C \|v\|_{H^1(K)}, \qquad \forall v \in H^1(K). \tag{3.6c}$$

*Proof.* Obviously, $\varPi_r^{\mathrm{grad}} v - v = (\varPi_r^0 - I)(v - \overline{v})$. Hence, (3.6a) and (3.6b) follows from (3.4a) and (3.4b) of Lemma 3.1. It remains to prove (3.6c). By (3.4c) and the Poincaré-Friedrichs inequality,

$$\begin{aligned}
\|\varPi_r^{\mathrm{grad}} v\|_K &\le \|\overline{v}\|_K + \|\varPi_r^0 (v - \overline{v})\|_K \\
&\le \|\overline{v}\|_K + C \left( \|v - \overline{v}\|_K + h_K \|\operatorname{grad}(v - \overline{v})\|_K \right) \\
&\le C \left( \|v\|_K + h_K \|\operatorname{grad} v\|_K \right), \qquad \text{and} \\
h_K \|\operatorname{grad} \varPi_r^{\mathrm{grad}} v\|_K &= h_K \|\operatorname{grad} \varPi_r^0 (v - \overline{v})\|_K \\
&\le C \left( \|v - \overline{v}\|_K + h_K \|\operatorname{grad}(v - \overline{v})\|_K \right) \\
&\le C h_K \|\operatorname{grad} v\|_K.
\end{aligned}$$

Canceling out $h_K$ and adding, (3.6c) follows. □

**Lemma 3.3.** *There is an operator* $\Pi^{\rm div}_{p+2} : H(\mathrm{div}, K) \mapsto P_{p+2}(K;\mathbb{V})$ *such that for every* $\tau \in H(\mathrm{div}, K)$*, we have*

$$(\Pi^{\rm div}_{p+2}\tau, q_p)_K = (\tau, q_p)_K, \qquad \forall q_p \in P_p(K;\mathbb{V}), \tag{3.7a}$$

$$\langle \Pi^{\rm div}_{p+2}\tau\cdot n, \mu_{p+1}\rangle_{\partial K} = \langle \mu_{p+1}, \tau\cdot n\rangle_{1/2,\partial K} \qquad \forall \mu_{p+1} \in \tilde P_{p+1}(\partial K), \tag{3.7b}$$

$$\|\Pi^{\rm div}_{p+2}\tau\|_{H(\mathrm{div},K)} \le C\|\tau\|_{H(\mathrm{div},K)}. \tag{3.7c}$$

*Proof.* We will first construct the operator on the unit simplex $\hat K$ in $\mathbb{R}^N$. Recalling the notations in §3.3, define $P_p^\perp(\partial\hat K)$ to be the $L^2(\partial\hat K)$-orthogonal complement of $\tilde P_p(\partial \hat K)$ in $P_p(\partial\hat K)$, and

$$B^{\rm div}_{p+2}(\hat K) = \{\hat\tau \in P_{p+1}(\hat K;\mathbb{V}) + \hat x P_{p+1}(\hat K) : \langle \hat p_\perp, \hat\tau\cdot\hat n\rangle_{\partial\hat K} = 0, \quad \forall \hat p_\perp \in P^\perp_{p+1}(\partial\hat K)\}.$$

We construct an operator $\hat\Pi^{\rm div}_{p+2}$ mapping $H(\mathrm{div},\hat K)$ into $B^{\rm div}_{p+2}(\hat K)$ by

$$(\hat\Pi^{\rm div}_{p+2}\hat\tau, \hat q_p)_{\hat K} = (\hat\tau, \hat q_p)_{\hat K}, \qquad \forall \hat q_p \in P_p(\hat K;\mathbb{V}), \tag{3.8a}$$

$$\langle \hat\Pi^{\rm div}_{p+2}\hat\tau\cdot\hat n, \hat\mu_{p+1}\rangle_{\partial\hat K} = \langle \hat\mu_{p+1}, \hat\tau\cdot\hat n\rangle_{1/2,\partial K} \qquad \forall \hat\mu_{p+1} \in \tilde P_{p+1}(\partial\hat K). \tag{3.8b}$$

We claim that (3.8a)–(3.8b) uniquely determine $\hat\Pi^{\rm div}_{p+2}\hat\tau \in B^{\rm div}_{p+2}(\hat K)$. Indeed, if their right hand sides vanish, then since $\hat\Pi^{\rm div}_{p+2}\hat\tau$ is in $B^{\rm div}_{p+2}(\hat K)$, we find that $\hat\Pi^{\rm div}_{p+2}$ is a function in the Raviart-Thomas space whose canonical degrees of freedom vanish (see e.g., [10, Definition 5]), so $\hat\Pi^{\rm div}_{p+2}\hat\tau = 0$. Hence (3.8a)–(3.8b) uniquely defines $\hat\Pi^{\rm div}_{p+2}\hat\tau$.

Now, we define $\Pi^{\rm div}_{p+2}$ on any general simplex $K$ by mapping $\hat\Pi^{\rm div}_{p+2}\hat\tau$ from $\hat K$ using the Piola transform, as follows. Let $G_K$ be the affine homeomorphism from $\hat K$ onto $K$ and let $A$ denote its Fréchet derivative. Given any $\tau \in H(\mathrm{div},K)$, let $\hat\tau(\hat x)$ in $H(\mathrm{div},\hat K)$ be defined by $\tau\circ G_K = (\det A)^{-1}A\hat\tau$. Then, define $\Pi^{\rm div}_{p+2}\tau$ by

$$\Pi^{\rm div}_{p+2}\tau(x) = \frac{A}{\det A}\hat\Pi^{\rm div}_{p+2}\hat\tau(\hat x), \qquad \text{with } x = G_K(\hat x).$$

We will now show that this $\Pi^{\rm div}_{p+2}\tau$ satisfies the three properties in (3.7).

First, observe that (3.8a) and (3.8b) imply the corresponding identities on $K$, namely,

$$(\Pi^{\rm div}_{p+2}\tau - \tau,\, A^{-t}\hat q_p\circ G_K^{-1})_K = 0, \qquad \forall \hat q_p \in P_p(\hat K;\mathbb{V}),$$

$$\langle \Pi^{\rm div}_{p+2}\tau\cdot n,\, \hat\mu_{p+1}\circ G_K^{-1}\rangle_{\partial K} = \langle \hat\mu_{p+1}\circ G_K^{-1}, \tau\cdot n\rangle_{1/2,\partial K} \qquad \forall \hat\mu_{p+1} \in \tilde P_{p+1}(\partial\hat K).$$

This implies (3.7a) and (3.7b).

It only remains to prove (3.7c). We do this in two steps. First, we prove an $L^2(K)$ bound using the Piola map's well-known estimates for shape regular meshes, namely

$$c\|\hat\tau\|_{\hat K} \le \|\tau\|_K \frac{|K|^{1/2}}{h_K} \le C\|\hat\tau\|_{\hat K},$$

$$c\|\operatorname{div}\hat\tau\|_{\hat K} \le \|\operatorname{div}\tau\|_K|K|^{1/2} \le C\|\operatorname{div}\hat\tau\|_{\hat K}.$$

Together with the fact that $\hat\Pi^{\rm div}_{p+2}$ is a continuous operator on $H(\mathrm{div},\hat K)$, we obtain

$$\|\Pi^{\rm div}_{p+2}\tau\|_K + h_K\|\operatorname{div}\Pi^{\rm div}_{p+2}\tau\|_K \le C\left(\|\tau\|_K + h_K\|\operatorname{div}\tau\|_K\right). \tag{3.9}$$

In particular, this proves the $L^2(K)$-bound $\|\Pi^{\rm div}_{p+2}\tau\|_K \le C\|\tau\|_{H(\mathrm{div},K)}$.

Next, we prove a better bound on the divergence norm $\| \operatorname{div} \Pi^{\operatorname{div}}_{p+2}\tau\|_K$ by showing that

$$\operatorname{div}(\Pi^{\operatorname{div}}_{p+2}\tau) = \Pi_{p+1} \operatorname{div} \tau$$

where $\Pi_{p+1}$ is the $L^2(K)$-orthogonal projection onto $P_{p+1}(K)$. Indeed, for any $\omega_{p+1} \in P_{p+1}(K)$, we have, due to (3.7a) and (3.7b), that

$$\begin{aligned}(\operatorname{div}(\Pi^{\operatorname{div}}_{p+2}\tau), \omega_{p+1})_K &= -(\Pi^{\operatorname{div}}_{p+2}\tau, \operatorname{grad} \omega_{p+1})_K + \langle (\Pi^{\operatorname{div}}_{p+2}\tau) \cdot n, \omega_{p+1}\rangle_{\partial K} \\ &= -(\tau, \operatorname{grad} \omega_{p+1})_K + \langle \omega_{p+1}, \tau \cdot n\rangle_{1/2,\partial K} \\ &= (\operatorname{div} \tau, \omega_{p+1})_K.\end{aligned}$$

Hence,

$$\| \operatorname{div}(\Pi^{\operatorname{div}}_{p+2}\tau)\|_K = \|\Pi_{p+1} \operatorname{div} \tau\|_K \leq \| \operatorname{div} \tau\|_K. \tag{3.10}$$

Estimates (3.9) and (3.10) prove (3.7c). □

Now we are ready to apply Theorem 2.1 to obtain a convergence result for the practical DPG method for the Laplace equation.

**Theorem 3.4.** *Let $r \geq p + N$. Then the exact and discrete solutions for the DPG method for the Laplace's equation, namely $\mathscr{U} = (\sigma, u, \hat{u}, \hat{\sigma}_n)$ and $\mathscr{U}_h = (\sigma_h, u_h, \hat{u}_h, \hat{\sigma}_{n,h})$, satisfy*

$$\begin{aligned}&\|\sigma - \sigma_h\|_{L^2(\Omega)} + \|u - u_h\|_{L^2(\Omega)} + \|\hat{u} - \hat{u}_h\|_{H^{1/2}_0(\partial\Omega_h)} + \|\hat{\sigma}_n - \hat{\sigma}_{n,h}\|_{H^{-1/2}(\partial\Omega_h)} \\ &\qquad \leq C \inf_{(\rho_h, w_h, \hat{w}_h, \hat{\eta}_h) \in U_h} \Big( \|\sigma - \rho_h\|_{L^2(\Omega)} + \|u - w_h\|_{L^2(\Omega)} \\ &\qquad\qquad + \|\hat{u} - \hat{w}_h\|_{H^{1/2}_0(\partial\Omega_h)} + \|\hat{\sigma}_n - \hat{\eta}_{n,h}\|_{H^{-1/2}(\partial\Omega_h)} \Big).\end{aligned}$$

*Proof.* As already observed, we have verified the first three assumptions of Theorem 2.1. To verify Assumption (2.4), let $\mathscr{V} = (\tau, v)$ and set $\Pi \mathscr{V} = (\Pi^{\operatorname{div}}_{p+2}\tau, \Pi^{\operatorname{grad}}_r v)$. The continuity estimates of $\Pi^{\operatorname{div}}_{p+2}$ and $\Pi^{\operatorname{grad}}_r$ of Lemmas 3.2 and 3.3 (namely (3.6c) and (3.7c)) show that (2.4b) holds. To see that (2.4a) also holds, observe that the identities of these lemmas also imply

$$\begin{aligned}(\rho_h, \tau - \Pi^{\operatorname{div}}_{p+2}\tau)_\Omega &= 0, & (w_h, \operatorname{div}(\tau - \Pi^{\operatorname{div}}_{p+2}\tau))_{\Omega_h} &= 0, \\ \langle \hat{w}_h, (\tau - \Pi^{\operatorname{div}}_{p+2}\tau) \cdot n\rangle_{\partial\Omega_h} &= 0, & (\rho_h, \operatorname{grad}(v - \Pi^{\operatorname{grad}}_r v))_{\Omega_h} &= 0, \\ \langle v - \Pi^{\operatorname{grad}}_r v, \hat{\eta}_h\rangle_{\partial\Omega_h} &= 0, & &\end{aligned}$$

for all $(\rho_h, w_h, \hat{w}_h, \hat{\eta}_h) \in U_h$. While the identities above on the left follow from the identities of (3.6) and (3.7), those on the right are proved by integration by parts. Together these identities imply that $b(\mathscr{W}, \mathscr{V} - \Pi \mathscr{V}) = 0$ for all $\mathscr{W} \in U_h$, so Assumption (2.4) is satisfied. □

*Remark* 3.5 (Enrichment degree). The above arguments point to the potential of choosing different enrichment degrees for the scalar and flux components of the test space. We have in fact proved that if, in place of the $V^r$ set in (3.3), we revise our choice of $V^r$ to

$$V^r = \{(\tau, v) \in V : \ \tau|_K \in P_{p+2}(K; \mathbb{V}), \ v|_K \in P_{p+N}(K), \ \forall K \in \Omega_h\},$$

then, we obtain the *same* convergence result. Obviously, the revised $V^r$ defines a smaller space if $N \geq 3$. The present DPG software packages are set to approximate all components of $T$ by polynomials of the same degree $r$. Our results indicate that this is unnecessary.

As an example of how Theorem 3.4 implies $h$-convergence rates, we state the following.

**Corollary 3.6** (Convergence rates). *Let* $h = \max_{K \in \Omega_h} \operatorname{diam}(K)$, $N = 2$ *or* $3$, *and let the assumptions of Theorem 3.4 hold. Then*

$$\|\sigma - \sigma_h\|_{L^2(\Omega)} + \|u - u_h\|_{L^2(\Omega)} + \|\hat{u} - \hat{u}_h\|_{H_0^{1/2}(\partial\Omega_h)} + \|\hat{\sigma}_n - \hat{\sigma}_{n,h}\|_{H^{-1/2}(\partial\Omega_h)} \\ \le C h^s \big(\|u\|_{H^{s+1}(\Omega)} + \|\sigma\|_{H^{s+1}(\Omega)}\big),$$

*for all* $1/2 < s \le p+1$.

*Proof.* The proof proceeds by bounding the infimum over $(\rho_h, w_h, \hat{w}_h, \hat{\eta}_h) \in U_h$ in Theorem 3.4. It is standard to bound the first two terms in the infimum, so we will only explain how to bound the next two terms. It is well-known (see, e.g., [8, Theorem 8.1]) that there are interpolants $\Pi_{\mathrm{grad}} u \in H_0^1(\Omega)$ and $\Pi_{\mathrm{div}} \sigma \in H(\mathrm{div}, \Omega)$, such that $\Pi_{\mathrm{grad}} u|_K \in P_{p+1}(K)$, $\Pi_{\mathrm{div}} \sigma|_K \in \vec{x} P_p(K) + P_p(K)^3$ for all $K \in \Omega_h$, and the interpolation errors satisfy

$$\|u - \Pi_{\mathrm{grad}} u\|_{H^1(\Omega)} \le C h^s |u|_{H^{s+1}(\Omega)}, \qquad (1/2 < s \le p+1), \tag{3.11a}$$
$$\|\sigma - \Pi_{\mathrm{div}} \sigma\|_{H(\mathrm{div},\Omega)} \le C h^s |\sigma|_{H^{s+1}(\Omega)}, \qquad (0 < s \le p+1). \tag{3.11b}$$

Let $\hat{t}_h$ denote the trace of $\Pi_{\mathrm{grad}} u$ on $\partial\Omega_h$. Then,

$$\inf_{\hat{w}_h} \|\hat{u} - \hat{w}_h\|_{H_0^{1/2}(\partial\Omega_h)} \le \|\hat{u} - \hat{t}_h\|_{H_0^{1/2}(\partial\Omega_h)} \le \|u - \Pi_{\mathrm{grad}} u\|_{H^1(\Omega)}.$$

The last inequality is obtained by observing that $\hat{u}$ is the trace of $u$ on $\partial\Omega_h$ and bounding the infimum in definition (3.1). In a similar fashion, we can estimate the last term in Theorem 3.4 by $\|\sigma - \Pi_{\mathrm{div}} \sigma\|_{H(\mathrm{div},\Omega)}$. The interpolation error estimates (3.11) then finish the proof. ☐

To conclude this section, we prove that the condition number of the stiffness matrix of the DPG method is no worse than other standard methods – see Remark 2.3 for the definition of the stiffness matrix with respect to a basis $\{\mathscr{B}_i\}$. Consider, for definiteness, the three-dimensional tetrahedral case. We tacitly assume that the basis functions $\mathscr{B}_i$ are local, and obtained, as in usual finite element practice, by mapping from the (reference) unit simplex. For example, a basis for the trial space for the numerical traces is built using a local basis $\{e_j\}$ for $\tilde{P}_{p+1}(\partial K)$, which in turn is obtained by mapping over a basis $\{\hat{e}_j\}$ for $\tilde{P}_{p+1}(\partial \hat{K})$ (where $e_j = \hat{e}_j \circ G_K^{-1}$ and we use the other mapping notations in the proof of Lemma 3.3). Consequently, if $\hat{s} = \sum_j s_j \hat{e}_j$ is the basis expansion for any $\hat{s} \in \tilde{P}_{p+1}(\partial K)$, then by the equivalence of norms in finite dimensional spaces

$$c \sum_j |s_j|^2 \le \inf_{\substack{\hat{e} \in P_{p+1}(\hat{K}), \\ (\hat{e} - \hat{s})|_{\partial \hat{K}} = 0}} \|\hat{e}\|^2_{H^1(\hat{K})} \le C \sum_j |s_j|^2. \tag{3.12}$$

Such arguments will be used in the following proof without further explanation.

**Theorem 3.7** (Conditioning). *Suppose* $\Omega_h$ *is a quasiuniform tetrahedral mesh and the assumptions of Theorem 3.4 hold. Then the spectral condition number of the stiffness matrix* $S$ *of the DPG method satisfies*

$$\kappa(S) \le C h^{-2}.$$

*Proof.* Let us apply (2.11). We have already shown above that $C_1, C_2$ and $C_\Pi$ are independent of $h$. Hence it only suffices to find the dependence of $\lambda_0$ and $\lambda_1$ on $h$ in (2.10).

Let $\mathscr{x} = (\rho, w, \hat{z}, \hat{\eta})$ in $U_h$. As a first step to bound the norm of $\hat{z}$, we recall that the existence of an $H^1(\hat{K})$ polynomial extension [7] implies that for any $S$ in $\tilde{P}_{p+1}(\partial\hat{K})$,

$$\inf_{\substack{\hat{e}_p\in P_{p+1}(\hat{K}),\\ \hat{e}_p|_{\partial\hat{K}}=S}} \|\hat{e}_p\|_{\hat{K}}^2 + \|\operatorname{grad}\hat{e}_p\|_{\hat{K}}^2 \le C \inf_{\substack{\hat{e}\in H^1(\hat{K}),\\ \hat{e}|_{\partial K}=S}} \|\hat{e}\|_{\hat{K}}^2 + \|\operatorname{grad}\hat{e}\|_{\hat{K}}^2.$$

Mapping to $K$ using the affine homeomorphism $G_K$ and scaling both sides by $|K|$, we obtain

$$\inf_{\substack{e_p\in P_{p+1}(K),\\ e_p|_{\partial K}=s}} \|e_p\|_K^2 + h_K^2\|\operatorname{grad} e_p\|_K^2 \le C \inf_{\substack{e\in H^1(K),\\ e|_{\partial K}=s}} \|e\|_K^2 + h_K^2\|\operatorname{grad} e\|_K^2. \tag{3.13}$$

where $s = S\circ G_K^{-1}$. Let us denote the function which achieves the left infimum by $E_p^{\mathrm{grad}} s$. Applying the above inequality, element by element, with $s$ replaced by $\hat{z}$, and using $h_K \le \operatorname{diam}\Omega$, we have proved that

$$\|E_p^{\mathrm{grad}}\hat{z}\|_\Omega^2 \le C\max(1, \operatorname{diam}\Omega)^2 \inf_{\substack{e\in H^1(\Omega),\\ (e-\hat{z})|_{\partial K}=0}} \left(\|e\|_\Omega^2 + \|\operatorname{grad} e\|_\Omega^2\right) \le C\|\hat{z}\|_{H_0^{1/2}(\partial\Omega_h)}^2.$$

Thus,

$$c\|E_p^{\mathrm{grad}}\hat{z}\|_\Omega^2 \le \|\hat{z}\|_{H_0^{1/2}(\partial\Omega_h)}^2 \le \|E_p^{\mathrm{grad}}\hat{z}\|_{H^1(\Omega)}^2,$$

where the upper inequality is obvious from the definition of the $H_0^{1/2}(\partial\Omega_h)$-norm. By an inverse inequality,

$$c\|E_p^{\mathrm{grad}}\hat{z}\|_\Omega^2 \le \|\hat{z}\|_{H_0^{1/2}(\partial\Omega_h)}^2 \le Ch^{-2}\|E_p^{\mathrm{grad}}\hat{z}\|_\Omega^2. \tag{3.14}$$

A similar argument using the $H(\operatorname{div}, \hat{K})$ polynomial extension in [8], gives

$$c\|E_p^{\mathrm{div}}\hat{\eta}\|_\Omega^2 \le \|\hat{\eta}\|_{H^{-1/2}(\partial\Omega_h)}^2 \le Ch^{-2}\|E_p^{\mathrm{div}}\hat{\eta}\|_\Omega^2. \tag{3.15}$$

Combining (3.14) and (3.15), we have

$$c\|\mathscr{x}\|_0^2 \le \|\mathscr{x}\|_U^2 \le Ch^{-2}\|\mathscr{x}\|_0^2, \qquad \forall \mathscr{x}\in U_h, \tag{3.16}$$

where $\|\mathscr{x}\|_0^2 = \|\rho\|_\Omega^2 + \|w\|_\Omega^2 + \|E_p^{\mathrm{grad}}\hat{z}\|_\Omega^2 + \|E_p^{\mathrm{div}}\hat{\eta}\|_\Omega^2$.

To prove (2.10), consider the coefficients in the basis expansion of $\mathscr{x}$. If $z_j$'s denote the coefficients in a basis expansion of the $\hat{z}|_{\partial K}$, then using (3.12) and the minimization property of $E_p^{\mathrm{grad}}\hat{z}$, we obtain

$$c\sum_j |z_j|^2 \le \frac{1}{|K|}\left(\|E_p^{\mathrm{grad}}\hat{z}\|_K^2 + h_K^2\|\operatorname{grad} E_p^{\mathrm{grad}}\hat{z}\|_K^2\right) \le C\sum_j |z_j|^2.$$

By an inverse inequality,

$$c\sum_j |z_j|^2 \le \frac{1}{|K|}\|E_p^{\mathrm{grad}}\hat{z}\|_K^2 \le C\sum_j |z_j|^2.$$

A similar estimate holds for $E_p^{\mathrm{div}}\hat{\eta}$. Combining these with the obvious estimates for the coefficients in the expansion of $\rho$ and $w$, we find that

$$c\|x\|_{\ell^2}^2 \min_{K\in\Omega_h}|K| \le \|\mathscr{x}\|_0^2 \le C\|x\|_{\ell^2}^2 \max_{K\in\Omega_h}|K|. \tag{3.17}$$

Clearly, inequalities (3.17) and (3.16) imply (2.10) with $\lambda_0 = c\min_{K\in\Omega_h}|K|$ and $\lambda_1 = Ch^{-2}\max_{K\in\Omega_h}|K|$, thus completing the proof. $\square$

## 4. Second example: Linear elasticity

Two ideal DPG methods for the linear elasticity equation were developed and analyzed in [2]. The two methods are equivalent for homogeneous isotropic materials. Among their many interesting properties is their robustness with respect to the Poisson ratio, i.e., the method is locking-free. In this section, we will consider the practical version of one of these two methods and prove its optimal convergence. We proceed as in the previous example, by first setting the abstract forms and spaces to those specific to this method then proceed to verify the hypotheses required to apply Theorem 2.1. In this section, we restrict to $N = 2$ or 3. The results and the analysis are similar to those in Section 3, so we will be brief.

4.1. **The spaces.** We set the trial and test spaces by

$$\begin{aligned} U &= L^2(\Omega;\mathbb{M}) \times L^2(\Omega;\mathbb{V}) \times H_0^{1/2}(\partial\Omega_h;\mathbb{V}) \times H^{-1/2}(\partial\Omega_h;\mathbb{V}) \times \mathbb{R}, \\ V &= H(\mathrm{div},\Omega_h;\mathbb{S}) \times H^1(\Omega_h;\mathbb{V}) \times L^2(\Omega;\mathbb{K}) \times \mathbb{R}, \end{aligned}$$

where $\mathbb{M} = \mathbb{R}^{N\times N}$, $\mathbb{S}$ consists of symmetric matrices in $\mathbb{M}$, and $\mathbb{K}$ consists of skew-symmetric matrices in $\mathbb{M}$. The trial and test spaces are normed by

$$\begin{aligned} \|(\sigma,u,\hat u,\hat\sigma_n,\alpha)\|_U^2 &= \|\sigma\|_\Omega^2 + \|u\|_\Omega^2 + \|\hat u\|^2_{H^{1/2}(\partial\Omega_h)} + \|\hat\sigma_n\|^2_{H^{-1/2}(\partial\Omega_h)} + |\alpha|^2, \\ \|(\tau,v,q,\beta)\|_V^2 &= \|\tau\|^2_{H(\mathrm{div},\Omega_h)} + \|v\|^2_{H^1(\Omega_h)} + \|q\|_\Omega^2 + |\beta|^2. \end{aligned}$$

4.2. **The forms.** The (second) ultraweak formulation derived in [2] reads as follows: Find $\mathscr{U} \equiv (\sigma,u,\hat u,\hat\sigma_n,\alpha) \in U$ satisfying (1.1) for all $\mathscr{V} \equiv (\tau,v,q,\beta) \in V$ where the forms $b(\cdot,\cdot)$ and $l(\cdot)$ are set by

$$\begin{aligned} b(\mathscr{U},\mathscr{V}) = {}& (\mathcal{A}\sigma,\tau)_{\Omega_h} + (u,\mathrm{div}\,\tau)_{\Omega_h} - \langle \hat u, \tau\, n\rangle_{\partial\Omega_h} + Q_0^{-1}(\alpha I, \mathcal{A}\tau)_\Omega \\ & + (\sigma,\mathrm{grad}\, v)_{\Omega_h} + (\sigma,q)_\Omega - \langle v,\hat\sigma_n\rangle_{\partial\Omega_h} \\ & + Q_0^{-1}(\mathcal{A}\sigma,\beta I)_\Omega \\ l(\mathscr{V}) = {}& (f,v)_\Omega, \end{aligned}$$

for some $f$ in $L^2(\Omega;\mathbb{V})$. Here, $\mathcal{A}$ is the generalized compliance tensor (see e.g., [2, Remark 2.1]) and $Q_0$ is the essential infimum of the trace of the matrix $\mathcal{A}(x)I$ over $x \in \Omega$. Throughout, we assume that $\mathcal{A}$ is element-wise constant. We note that above and throughout, the inner products of matrix-valued functions, such as $(\sigma,\tau)_K$, are computed by integrating the Frobenius product of the two matrices.

It is easy to see that the resulting $\sigma$ and $u$ satisfies $\mathcal{A}\sigma = \varepsilon(u)$, where $\varepsilon(u) = (\mathrm{grad}\, u + (\mathrm{grad}\, u)')/2$, and $\mathrm{div}\,\sigma = f$ on $\Omega$, and $u = 0$ on $\partial\Omega$, and $\alpha = 0$. For details, consult [2].

4.3. **Discrete spaces.** Symmetric, skew-symmetric, and general matrix-valued functions whose entries are in $P_p(K)$ are denoted by $P_p(K;\mathbb{S}), P_p(K;\mathbb{K})$, and $P_p(K;\mathbb{M})$, resp. Using these notations, we set the trial approximation space for the DPG method by

$$\begin{aligned} U_h = \{(\sigma,u,\hat u,\hat\sigma_n,\alpha) \in U:\quad & \sigma|_K \in P_p(K;\mathbb{M}),\ u|_K \in P_p(K;\mathbb{V}), \\ & \hat u|_{\partial K} \in \tilde P_{p+1}(\partial K;\mathbb{V}),\ \hat\sigma_n|_{\partial K} \in P_p(\partial K;\mathbb{V}),\ \alpha \in \mathbb{R},\ \ \forall K \in \Omega_h\}. \end{aligned}$$

The discrete test space is defined by $V_h^r = T^r(U_h)$, so to complete the prescription of the practical DPG method, we only need to specify $V^r$. Set

$$V^r = \{(\tau, v, q, \beta) \in V : \quad \tau|_K \in P_r(K;\mathbb{S}),\ v|_K \in P_r(K;\mathbb{V}), \\ q|_K \in P_p(K;\mathbb{K}),\ \beta \in \mathbb{R},\ \forall K \in \Omega_h\},$$

for some integer $r \geq p + N$.

4.4. **Verification of the assumptions.** To apply Theorem 2.1 to the above setting, we need to verify its assumptions.

- Assumption (2.1) is verified by [2, Lemma 5.2].
- Assumption (2.2) is verified by [2, Lemma 5.3].
- Assumption (2.3) can be easily verified, as in the case of the Laplace equation.
- Assumption (2.4) is verified next.

Let $\mathscr{v} = (\tau, v, q, \beta) \in V$. The operator $\Pi$ satisfying (2.4) will take the form

$$\Pi \mathscr{v} = (\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau, \Pi_r^{\mathrm{grad}} v, \Pi_p^{\mathbb{K}} q, \beta). \tag{4.1}$$

We set $\Pi_r^{\mathrm{grad}} v$ to be the one defined in Lemma 3.2, but applied component by component, to the vector valued function $v$. The operator $\Pi_p^{\mathbb{K}}$ is simply the $L^2$-orthogonal projection onto $\{q \in L^2(\Omega;\mathbb{K}) : q|_K \in P_p(K;\mathbb{K}), \forall K \in \Omega_h\}$. It remains to construct the operator $\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}$. We do so, based on a set of degrees of freedom given in [9], in the next lemma.

**Lemma 4.1.** *There is an operator $\Pi_{p+2}^{(\mathrm{div},\mathbb{S})} : H(\mathrm{div}, K;\mathbb{S}) \to P_{p+2}(K;\mathbb{S})$ such that for every $\tau \in H(\mathrm{div}, K;\mathbb{S})$, we have*

$$(\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau, q_p)_K = (\tau, q_p)_K, \qquad \forall q_p \in P_p(K;\mathbb{S}), \tag{4.2a}$$

$$\langle \Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau \cdot n, \mu_{p+1}\rangle_{\partial K} = \langle \mu_{p+1}, \tau\cdot n\rangle_{1/2,\partial K}, \qquad \forall \mu_{p+1} \in \tilde{P}_{p+1}(\partial K;\mathbb{V}), \tag{4.2b}$$

$$\|\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau\|_{H(\mathrm{div},K)} \leq C\|\tau\|_{H(\mathrm{div},K)}. \tag{4.2c}$$

*Proof.* We only give the proof for $N = 3$ as the proof for $N = 2$ is similar. As in the proof of Lemma 3.3, we will first construct the operator on the unit simplex $\hat{K}$ in $\mathbb{R}^N$. Define $P_{p+1}^{\perp}(\partial\hat{K};\mathbb{V}) = L^2(\partial\hat{K};\mathbb{V})$-orthogonal complement of $\tilde{P}_{p+1}(\partial\hat{K};\mathbb{V})$ in $P_{p+1}(\partial\hat{K};\mathbb{V})$ and set $P_{p+2}^0(\hat{K};\mathbb{S}) = \{\hat{\tau} \in P_{p+2}(\hat{K};\mathbb{S}) : \ \langle \hat{s}, \hat{\tau}\hat{n}^- \cdot \hat{n}^+\rangle_{\hat{e}} = 0,\ \forall \hat{s} \in P_{p+2}(\hat{e}),\ \forall \hat{e} \in \triangle_1(\hat{K})\}$ where, for each edge $\hat{e} \in \triangle_1(\hat{K})$, $\hat{n}^+$ and $\hat{n}^-$ are the normal vectors of the two faces sharing $\hat{e}$. Let $B_{p+2}^{\mathrm{div}}(\hat{K};\mathbb{S}) = \{\hat{\tau} \in P_{p+2}^0(\hat{K};\mathbb{S}) : \langle \hat{v}, \hat{\tau}\hat{n}\rangle_{\partial\hat{K}} = 0$ for all $\hat{v} \in P_{p+1}^{\perp}(\partial\hat{K};\mathbb{V})\}$. We define $\hat{\Pi}_{p+2}^{(\mathrm{div},\mathbb{S})} :\ H(\mathrm{div},\hat{K};\mathbb{S}) \mapsto B_{p+2}^{\mathrm{div}}(\hat{K};\mathbb{S})$ by

$$(\hat{\Pi}_{p+2}^{(\mathrm{div},\mathbb{S})}\hat{\tau}, \hat{q}_p)_{\hat{K}} = (\hat{\tau}, \hat{q}_p)_{\hat{K}}, \qquad \forall \hat{q}_p \in P_p(\hat{K};\mathbb{S}), \tag{4.3a}$$

$$\langle \hat{\Pi}_{p+2}^{(\mathrm{div},\mathbb{S})}\hat{\tau}\cdot\hat{n}, \hat{\mu}_{p+1}\rangle_{\partial\hat{K}} = \langle \hat{\mu}_{p+1}, \hat{\tau}\cdot\hat{n}\rangle_{1/2,\partial K}, \qquad \forall \hat{\mu}_{p+1} \in \tilde{P}_{p+1}(\partial\hat{K};\mathbb{V}). \tag{4.3b}$$

By [9, Theorem 2.1], these equations are uniquely solvable, so $\hat{\Pi}_{p+2}^{(\mathrm{div},\mathbb{S})}$ is well defined.

Next, we define $\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}$ on any general simplex $K$ by mapping $\hat{\Pi}_{p+2}^{(\mathrm{div},\mathbb{S})}$ from $\hat{K}$ using the Piola transform for symmetric matrix-valued functions. Recalling the mapping $G_K$ from $\hat{K}$ onto $K$ and its derivative $A$, we define

$$\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau(x) = \frac{1}{\det A} A\hat{\Pi}_{p+2}^{(\mathrm{div},\mathbb{S})}\hat{\tau}(\hat{x})A^t,$$

for any $\tau \in H(\mathrm{div}, K; \mathbb{S})$. Here, given $\tau$ on $K$, the function $\hat\tau$ on $\hat K$ is defined by $(\det A)\tau(x) = A\hat\tau(\hat x)A^t$, with $x = G_K(\hat x)$. As in the proof of Lemma 3.3, it is now easy to see that $\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau$ satisfies (4.2a) and (4.2b).

Next, we observe that the commutativity property

$$\mathrm{div}\, \Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau = \Pi_{p+1}\,\mathrm{div}\,\tau, \tag{4.4}$$

holds, where $\Pi_{p+1}$ denotes the $L^2(K;\mathbb{V})$-orthogonal projection onto $P_{p+1}(K,\mathbb{V})$. Let $\omega_{p+1} \in P_{p+1}(K;\mathbb{V})$. Then

$$\begin{aligned}
(\mathrm{div}(\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau), \omega_{p+1})_K &= -(\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau, \mathrm{grad}\,\omega_{p+1})_K + \langle(\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau)\cdot n, \omega_{p+1}\rangle_{\partial K}\\
&= -(\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau, \varepsilon(\omega_{p+1}))_K + \langle(\Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau)\cdot n, \omega_{p+1}\rangle_{\partial K}\\
&= -(\tau, \varepsilon(\omega_{p+1}))_K + \langle\omega_{p+1}, \tau\cdot n\rangle_{1/2,\partial K}, \qquad \text{(by (4.1))},\\
&= -(\tau, \mathrm{grad}\,\omega_{p+1})_K + \langle\omega_{p+1}, \tau\cdot n\rangle_{1/2,\partial K},\\
&= (\mathrm{div}\,\tau, \omega_{p+1})_K.
\end{aligned}$$

which proves (4.4).

It only remains to prove the estimate of (4.2c). This can now be done as in the proof of the estimate (3.7c) of Lemma 3.3, in two steps, using (4.4) in place of the commutativity property used there. ☐

The main result of this section is the following.

**Theorem 4.2.** *Suppose that $r \ge p + N$ and suppose that the compliance tensor $\mathcal{A}$ is element-wise constant. Then, the difference between the discrete solution of the practical DPG method, $\mathscr{U}_h = (\sigma_h, u_h, \hat u_h, \hat\sigma_{n,h}, \alpha_h)$, and the exact solution $\mathscr{U} = (\sigma, u, \hat u, \hat\sigma_n, \alpha)$ satisfies*

$$\begin{aligned}
&\|\sigma - \sigma_h\|_{L^2(\Omega)} + \|u - u_h\|_{L^2(\Omega)} + \|\hat u - \hat u_h\|_{H_0^{1/2}(\partial\Omega_h)} + \|\hat\sigma_n - \hat\sigma_{n,h}\|_{H^{-1/2}(\partial\Omega_h)} + |\alpha - \alpha_h|\\
&\qquad \le C \inf_{(\rho_h, w_h, \hat w_h, \hat\eta_h,)\in U_h} \Big( \|\sigma - \rho_h\|_{L^2(\Omega)} + \|u - w_h\|_{L^2(\Omega)}\\
&\qquad\qquad + \|\hat u - \hat u_h\|_{H_0^{1/2}(\partial\Omega_h)} + \|\hat\sigma_n - \hat\sigma_{n,h}\|_{H^{-1/2}(\partial\Omega_h)}\Big).
\end{aligned}$$

*Proof.* As mentioned above, we only need to verify Assumption (2.4) for the $\Pi$ in (4.1) and apply Theorem 2.1. By the inequalities of previous lemmas, the estimate (2.4b) is obvious. To prove (2.4a), namely $b(\mathscr{W}, \mathscr{V} - \Pi\mathscr{V}) = 0$ for all $\mathscr{W} \in U_h$, it suffices to prove the following eight identities

$$\begin{aligned}
(\mathcal{A}\rho_h, \tau - \Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau)_\Omega &= 0, & (w_h, \mathrm{div}(\tau - \Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau))_{\Omega_h} &= 0,\\
\langle\hat w_h, (\tau - \Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau)\cdot n\rangle_{\partial\Omega_h} &= 0, & (\rho_h, \mathrm{grad}(v - \Pi_r^{\mathrm{grad}} v))_{\Omega_h} &= 0,\\
\langle v - \Pi_r^{\mathrm{grad}} v, \hat\eta_h\rangle_{\partial\Omega_h} &= 0, & Q_0^{-1}(\gamma_h I, \mathcal{A}(\tau - \Pi_{p+2}^{(\mathrm{div},\mathbb{S})}\tau))_\Omega &= 0\\
(\rho_h, q - \Pi_p^{\mathbb{K}} q)_\Omega &= 0, & Q_0^{-1}(\mathcal{A}\rho_h, (\beta - \beta)I)_\Omega &= 0,
\end{aligned}$$

for all $\mathscr{W} \equiv (\rho_h, w_h, \hat w_h, \hat\eta_h, \gamma_h) \in U_h$. The first five are proved exactly as in the proof of Theorem 3.4 but using the new lemma. The sixth is obvious from (4.2a). To see the seventh, denoting by $\mathrm{skw}\,\rho_h$ the skew-symmetric part of $\rho_h$, observe that $(\rho_h, q - \Pi_p^{\mathbb{K}} q)_\Omega = (\mathrm{skw}(\rho_h), q - \Pi_p^{\mathbb{K}} q)_\Omega = 0$, by the definition of $\Pi_p^{\mathbb{K}}$.

Therefore, applying Theorem 2.1, we obtain a quasioptimality estimate. This yields the estimate of the theorem after observing that in the infimum over $\mathscr{w} \equiv (\rho_h, w_h, \hat{w}_h, \hat{\eta}_h, \gamma_h)$ in $U_h$, we may choose $\gamma_h = \alpha$. □

We conclude by noting that results similar to Corollary 3.6 and Theorem 3.7 can be established for this example as well. The arguments are very similar.

## References


[1] I. Babuška, *Error-bounds for finite element method*, Numer. Math., 16 (1970/1971), pp. 322–333.

[2] J. Bramwell, L. Demkowicz, J. Gopalakrishnan and W. Qiu, *A locking-free hp DPG method for linear elasticity with symmetric stresses*, to appear in Numerische Mathematik, 2012.

[3] L. Demkowicz and J. Gopalakrishnan, *A class of discontinuous Petrov-Galerkin methods. Part I: The transport equation*, Computer Methods in Applied Mechanics and Engineering, 199 (2010), pp. 1558–1572.

[4] ———, *A class of discontinuous Petrov-Galerkin methods. Part II: Optimal test functions*, Numerical Methods for Partial Differential Equations, 27 (2011), pp. 70–105.

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

[6] L. Demkowicz, J. Gopalakrishnan, and A. Niemi, *A class of discontinuous Petrov-Galerkin methods. Part III: Adaptivity*, Applied Numerical Mathematics, 62:396–427, 2012.

[7] L. Demkowicz, J. Gopalakrishnan, and J. Schöberl. *Polynomial extension operators. Part I.* SIAM J. Numer. Anal., 46(6):3006–3031, 2008.

[8] ———, *Polynomial extension operators. Part III.* Math. Comp., 81(279):1289–1326, 2012.

[9] J. Gopalakrishnan and J. Guzmán, *Symmetric non-conforming mixed finite elements for linear elasticity*, *SIAM J Numer. Anal.*, 49(4):1504–1520, 2011.

[10] J.-C. Nédélec, *Mixed Finite Elements in* $\mathbb{R}^3$, Numer. Math., 35 (1980), pp. 315–341.

[11] J. Xu and L. Zikatanov, *Some observations on Babuška and Brezzi theories*, Numer. Math., 94 (2003), pp. 195–202.

[12] K. Yosida, *Functional analysis*, Classics in Mathematics, Springer-Verlag, Berlin, 1995. Reprint of the sixth (1980) edition.



Portland State University, PO Box 751, Portland, OR 97207-0751
*E-mail address*: gjay@pdx.edu

The Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, Minnesota 55455
*E-mail address*: qiuxa001@ima.umn.edu