

Oden Institute REPORT 19-19

December 2019

Discontinuous Petrov-Galerkin Methods in Banach Spaces

by

Jiaqi LI and L. Demkowicz



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

Reference: Jiaqi LI and L. Demkowicz, "Discontinuous Petrov-Galerkin Methods in Banach Spaces," Oden Institute REPORT 19-19, Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, December 2019.

Discontinuous Petrov-Galerkin Methods in Banach Spaces

Jiaqi Li and Leszek Demkowicz

Oden Institute for Computational Engineering and Sciences
The University of Texas at Austin, 201 E 24th St, Austin, TX 78712, USA

Abstract

Following Muga and van der Zee [8], we generalize the standard Discontinuous Petrov-Galerkin method, based on Hilbert spaces, to Banach spaces. Numerical experiments using model 1D convection-dominated diffusion problem are performed and compared with Hilbert setting. It is shown that Banach-based method gives solutions less susceptible to Gibbs phenomenon. h-adaptivity is implemented with the help of the error representation function as error indicator.

1 Introduction

The Discontinuous Petrov-Galerkin (DPG) Method can be applied to any well-posed variational problem [5], and it is best combined with broken test spaces [2] for most efficiency. The DPG method can be interpreted as a minimum-residual method with the residual measured in a dual norm. Consider the abstract problem

$$\begin{cases} \text{Find } u \in \mathcal{U} : \\ Bu = l \quad \text{in } \mathcal{V}' \end{cases} \quad (1)$$

where \mathcal{U}, \mathcal{V} are trial and test spaces (Banach spaces in general), $B : \mathcal{U} \rightarrow \mathcal{V}'$ is a bounded linear operator dictated by the problem and the variational formulation we choose. For a well-posed variational problem, B is bounded below as well.

Given a discrete trial space $\mathcal{U}_h \subset \mathcal{U}$, the ideal DPG method (test space is not discretized yet) solves the minimum residual problem

$$\begin{cases} \text{Find } u_h \in \mathcal{U}_h : \\ \|Bu_h - l\|_{\mathcal{V}'} \quad \text{is minimized.} \end{cases} \quad (2)$$

In Hilbert space setting, both \mathcal{U} and \mathcal{V} are Hilbert spaces, and we can make use of Riesz operator $R_{\mathcal{V}} : \mathcal{V} \rightarrow \mathcal{V}'$, defined by

$$\langle R_{\mathcal{V}}v, \delta v \rangle_{\mathcal{V}', \mathcal{V}} = (v, \delta v)_{\mathcal{V}} \quad \forall \delta v \in \mathcal{V}.$$

By Riesz representation theorem,

$$\|Bu_h - l\|_{\mathcal{V}'} = \|R_{\mathcal{V}}^{-1}(Bu_h - l)\|_{\mathcal{V}}.$$

Consequently, the minimum residual problem (2) can be reformulated as

$$\begin{cases} \text{Find } u_h \in \mathcal{U}_h : \\ \|R_{\mathcal{V}}^{-1}(Bu_h - l)\|_{\mathcal{V}}^2 \text{ is minimized.} \end{cases} \quad (3)$$

The optimality condition translates into

$$(R_{\mathcal{V}}^{-1}(Bu_h - l), R_{\mathcal{V}}^{-1}Bw_h)_{\mathcal{V}} = 0 \quad \forall w_h \in \mathcal{U}_h. \quad (4)$$

Let $\psi \in \mathcal{V}$ be the Riesz representation of the residual, the so called error representation function, i.e.

$$\psi = R_{\mathcal{V}}^{-1}(l - Bu_h). \quad (5)$$

The optimality condition (4) now reads as

$$(\psi, R_{\mathcal{V}}^{-1}Bw_h)_{\mathcal{V}} = \langle Bw_h, \psi \rangle_{\mathcal{V}', \mathcal{V}} = 0 \quad \forall w_h \in \mathcal{U}_h. \quad (6)$$

Finally, we can reformulate the minimum residual problem (3) as a mixed problem

$$\begin{cases} \text{Find } \psi \in \mathcal{V}, u_h \in \mathcal{U}_h : \\ (\psi, v)_{\mathcal{V}} + \langle Bu_h, v \rangle = l(v) & \forall v \in \mathcal{V} \\ \langle Bw_h, \psi \rangle = 0 & \forall w_h \in \mathcal{U}_h \end{cases} \quad (7)$$

where we have omitted subscripts of the duality pairing for simplicity.

The goal of this report is to replace Hilbert spaces with Banach spaces, focusing on Sobolev spaces $W^{1,p}(\Omega)$, $p \geq 2$. Using the Banach analog of Riesz operator, we will introduce error representation function and derive a mixed problem similar to (7). The key difference is that ‘‘Riesz operator’’ for Banach space is no longer linear; hence the need to solve a nonlinear system of equations. Newton’s method with line search is applied. The theory is illustrated with numerical experiments for a 1D convection-diffusion model problem using both classical and ultraweak variational formulations. The experiments include h-adaptivity driven by the error representation function ψ .

In the development of the theory for DPG in Banach spaces, we follow closely the work of Muga and van der Zee [8].

2 Theory: From Hilbert to Banach

2.1 Convection-diffusion problem and variational formulations

To stay focused, we will consider a model convection-diffusion problem. Given a domain $\Omega \subset \mathbb{R}^N$, we want to solve

$$-\nabla \cdot (\epsilon \nabla u - \beta u) = f \quad \text{in } \Omega \quad (8)$$

where ϵ is the diffusion coefficient, β denotes an incompressible advection field, and f is a source term. We assume a flux boundary condition on the inflow boundary,

$$-\epsilon \frac{\partial u}{\partial n} + \beta_n u = \beta_n u_0 \quad \text{on } \Gamma_{\text{in}} \quad (9)$$

where $\beta_n = \beta \cdot n$ and

$$\Gamma_{\text{in}} = \{x \in \Gamma : \beta_n < 0\}.$$

On the remaining, outflow part Γ_{out} of the boundary, homogeneous Dirichlet boundary condition $u = 0$ is imposed. As usual, n denotes the outward normal unit vector on $\Gamma = \partial\Omega$.

Classical variational formulation. The standard variational formulation in Hilbert setting [4] is

$$\begin{cases} \text{Find } u \in \mathcal{U} : \\ \int_{\Omega} \epsilon \nabla u \cdot \nabla v - u \beta \cdot \nabla v = \int_{\Omega} f v + \int_{\Gamma_{\text{in}}} \beta_n u_0 v \quad \forall v \in \mathcal{V} \end{cases} \quad (10)$$

where

$$\mathcal{U} = \mathcal{V} = H_{\Gamma_{\text{out}}}^1(\Omega) := \{u \in H^1(\Omega) : u = 0 \text{ on } \Gamma_{\text{out}}\}.$$

Our goal is to replace the Hilbert trial and test spaces by Banach spaces:

$$\begin{aligned} \mathcal{U} &= W_{\Gamma_{\text{out}}}^{1,p'}(\Omega) := \{u \in W^{1,p'}(\Omega) : u = 0 \text{ on } \Gamma_{\text{out}}\} \\ \mathcal{V} &= W_{\Gamma_{\text{out}}}^{1,p}(\Omega) \end{aligned} \quad (11)$$

where $p \geq 2$, $1/p + 1/p' = 1$.

Ultraweak variational formulation. By introducing total flux $\sigma = \epsilon \nabla u - \beta u$, we can rewrite the convection-diffusion problem as a first-order system:

$$\begin{cases} \epsilon^{-1} \sigma - \nabla u + \epsilon^{-1} \beta u = 0 & \text{in } \Omega \\ -\nabla \cdot \sigma = f & \text{in } \Omega \\ -\sigma \cdot n = \beta_n u_0 & \text{on } \Gamma_{\text{in}} \\ u = 0 & \text{on } \Gamma_{\text{out}}. \end{cases} \quad (12)$$

We can now multiply the first equation with a vector-valued test function τ , the second equation with a scalar-valued function v , and integrate over Ω :

$$\begin{cases} (\epsilon^{-1} \sigma, \tau) - (\nabla u, \tau) + (\epsilon^{-1} \beta u, \tau) = 0 \\ -(\nabla \cdot \sigma, v) = (f, v) \end{cases}$$

where (\cdot, \cdot) denotes standard $L^2(\Omega)$ inner product,

$$(u, v) = \int_{\Omega} uv \, dx.$$

For vector-valued functions, we take their inner product and then integrate over Ω . Integrating by parts, and making use of boundary conditions, we finally obtain the ultraweak formulation [4]

$$\begin{cases} \text{Find } \sigma \in (L^2(\Omega))^N, u \in L^2(\Omega) : \\ (\sigma, \epsilon^{-1} \tau) + (u, \nabla \cdot \tau + \epsilon^{-1} \beta \cdot \tau) = 0 & \forall \tau \in H_{\Gamma_{\text{in}}}(\text{div}, \Omega) \\ (\sigma, \nabla v) = (f, v) - \int_{\Gamma_{\text{in}}} \beta_n u_0 v & \forall v \in H_{\Gamma_{\text{out}}}^1(\Omega) \end{cases} \quad (13)$$

where

$$H_{\Gamma_{\text{in}}}(\text{div}, \Omega) := \{\tau \in H(\text{div}, \Omega) : \tau \cdot n = 0 \text{ on } \Gamma_{\text{in}}\}.$$

Consequently,

$$\begin{aligned} \mathcal{U} &= (L^2(\Omega))^N \times L^2(\Omega) \\ \mathcal{V} &= H_{\Gamma_{\text{in}}}(\text{div}, \Omega) \times H_{\Gamma_{\text{out}}}^1(\Omega). \end{aligned} \tag{14}$$

In the Banach setting, we have

$$\begin{aligned} \mathcal{U} &= (L^{p'}(\Omega))^N \times L^{p'}(\Omega) \\ \mathcal{V} &= W_{\Gamma_{\text{in}}}^p(\text{div}, \Omega) \times W_{\Gamma_{\text{out}}}^{1,p}(\Omega) \end{aligned} \tag{15}$$

where $p \geq 2$, $1/p + 1/p' = 1$, and

$$W_{\Gamma_{\text{in}}}^p(\text{div}, \Omega) := \{\tau \in (L^p(\Omega))^N : \text{div } \tau \in L^p(\Omega), \tau \cdot n = 0 \text{ on } \Gamma_{\text{in}}\}.$$

This motivates our study of the analog of Riesz operator for $W^{1,p}(\Omega)$ and $W^p(\text{div}, \Omega) \times W^{1,p}(\Omega)$, which is presented next.

2.2 Representation operator

The DPG method minimizes the error in dual norm. The introduction of Riesz operator is the key to computation of the dual norm. In Banach spaces, we need a similar representation operator that achieves this goal. Consider the test space with norm

$$\begin{aligned} \|v\|_{\mathcal{V}}^p &:= \|v\|_{L^p(\Omega)}^p + \|\nabla v\|_{(L^p(\Omega))^N}^p \\ &= \sum_{|\alpha| \leq 1} \|D^\alpha v\|_{L^p(\Omega)}^p \end{aligned} \tag{16}$$

for the classical variational formulation and

$$\|(\tau, v)\|_{\mathcal{V}}^p := \|\tau\|_{(L^p(\Omega))^N}^p + \|\text{div } \tau\|_{L^p(\Omega)}^p + \|v\|_{L^p(\Omega)}^p + \|\nabla v\|_{(L^p(\Omega))^N}^p \tag{17}$$

for the ultraweak formulation. We shall restrict our attention to $p \in [2, \infty)$.

Let $l \in \mathcal{V}'$. When \mathcal{V} is Hilbert, the Riesz representation of l , $R_{\mathcal{V}}^{-1}l$, minimizes the total ‘‘potential energy’’:

$$R_{\mathcal{V}}^{-1}l = \arg \min_{v \in \mathcal{V}} \frac{1}{2} \|v\|_{\mathcal{V}}^2 - l(v). \tag{18}$$

Analogously, the ‘‘Banach version’’ of the Riesz representation of l , can be defined by considering the p -analog of the energy,

$$R_{\mathcal{V}}^{-1}l := \arg \min_{v \in \mathcal{V}} \frac{1}{p} \|v\|_{\mathcal{V}}^p - l(v). \tag{19}$$

Equivalently,

$$\langle R_{\mathcal{V}}(v), \delta v \rangle := \langle \partial J(v), \delta v \rangle = l(\delta v) \quad \forall \delta v \in \mathcal{V} \tag{20}$$

where $J : \mathcal{V} \rightarrow \mathbb{R}$ is defined by

$$J(v) := \frac{1}{p} \|v\|_{\mathcal{V}}^p. \tag{21}$$

The ‘‘Banach version’’ of Riesz map $R_{\mathcal{V}} : \mathcal{V} \rightarrow \mathcal{V}'$ is given by the Gâteaux derivative of J , $R_{\mathcal{V}} = \partial J$. We can compute the Gâteaux derivative explicitly,

$$\left\{ \begin{array}{l} \langle R_{\mathcal{V}}(v), \delta v \rangle = \sum_{|\alpha| \leq 1} \int_{\Omega} |D^{\alpha} v|^{p-2} D^{\alpha} v D^{\alpha} \delta v \quad \text{for the classical variational formulation} \\ \langle R_{\mathcal{V}}((\tau, v)), (\delta \tau, \delta v) \rangle = \sum_{i=1}^N \int_{\Omega} |\tau_i|^{p-2} \tau_i \delta \tau_i + \int_{\Omega} |\operatorname{div} \tau|^{p-2} \operatorname{div} \tau \operatorname{div} \delta \tau + \sum_{|\alpha| \leq 1} \int_{\Omega} |D^{\alpha} v|^{p-2} D^{\alpha} v D^{\alpha} \delta v \\ \text{for the ultraweak variational formulation.} \end{array} \right. \quad (22)$$

Theorem 1. (*Representation Theorem for the Dual Space*)

$R_{\mathcal{V}} : \mathcal{V} \rightarrow \mathcal{V}'$ is well defined, and it is bijective. Moreover, we have

$$\|R_{\mathcal{V}}(v)\|_{\mathcal{V}'} = \|v\|_{\mathcal{V}}^{p-1}.$$

The proof is relegated to the appendix. Our representation operator is similar to the duality mapping (section 2.1 in [8]), but it is slightly simpler for numerical implementation.

2.3 Minimization of residual in Banach spaces

We first invoke the best approximation theorem in Banach spaces (See Theorem 2.A in [8], or Section 10.2 in [9]).

Theorem 2. (*Best Approximation*)

Let Y be a Banach space, and $y \in Y$.

- Suppose $M \subset Y$ is a finite-dimensional subspace, then there exists a best approximation $y_0 \in M$ such that

$$\|y - y_0\| = \inf_{z \in M} \|y - z\|.$$

- Suppose $M \subset Y$ is a convex subset of Y , and Y is strictly convex. If there exists a best approximation y_0 to y , then y_0 must be unique.

Proof. Suppose first $M \subset Y$ is a finite-dimensional subspace. Then

$$d := \inf_{z \in M} \|y - z\| \leq \|y - 0\| = \|y\|.$$

By definition of the infimum, there exists a sequence $\{z_j\}$ in M such that $\|y - z_j\| \rightarrow d$. Note that

$$\|z_j\| = \|y - y + z_j\| \leq \|y\| + \|y - z_j\|.$$

Therefore, $\{z_j\}$ is a bounded sequence in M . By Bolzano–Weierstrass theorem, there exists a subsequence, denoted with the same symbol, such that $z_j \rightarrow y_0$ for some $y_0 \in M$. Continuity of norm implies

$$\|y - y_0\| = d = \inf_{z \in M} \|y - z\|.$$

Now let $M \subset Y$ be a convex subset of Y , and Y is strictly convex. Let y_0, y_1 be distinct best approximations to y , such that $\|y - y_0\| = \|y - y_1\| = d$, and $y_0 \neq y_1$. Consider now the convex combination $u = \alpha y_0 + (1 - \alpha)y_1$, for any $\alpha \in (0, 1)$. Due to convexity of M , we know that u is also in M . The strict convexity of Y leads to the inequality

$$\begin{aligned} \|y - u\| &= \|\alpha(y - y_0) + (1 - \alpha)(y - y_1)\| \\ &< \alpha\|y - y_0\| + (1 - \alpha)\|y - y_1\| \\ &= \alpha d + (1 - \alpha)d \\ &= d \end{aligned}$$

which implies u is a better approximation to y in M , a contradiction. \square

Consider the minimum residual problem (2) where $\mathcal{V} = W^{1,p}(\Omega)$. Now $Y = \mathcal{V}'$ is Banach, and $M = BU_h \subset Y$ is a finite-dimensional subspace. Therefore there exists a solution to problem (2). Moreover, as shown in (22), the norm in \mathcal{V} is Gâteaux differentiable, and we have a classical result that \mathcal{V} is reflexive. The reflexivity of space \mathcal{V} combined with Gâteaux differentiability of norm imply the strict convexity of dual space \mathcal{V}' (See [7], Corollary 5.4.18 and Proposition 5.4.7). Therefore the best approximation to l in BU_h is unique. The uniqueness of solution u_h to problem (2) follows from the injectivity of B . Next we derive a necessary condition for Bu_h to be the best approximation to l .

Introduce error representation function

$$\psi = R_{\mathcal{V}}^{-1}(l - Bu_h). \quad (23)$$

For our problem, $B : \mathcal{U} \rightarrow \mathcal{V}'$ is induced by a bilinear form b on $\mathcal{U} \times \mathcal{V}$

$$\langle Bu, v \rangle = b(u, v). \quad (24)$$

We seek to minimize $\|\psi\|_{\mathcal{V}}^{p-1} = \|Bu_h - l\|_{\mathcal{V}'}$, or equivalently, minimize $J(\psi) = \frac{1}{p}\|\psi\|_{\mathcal{V}}^p$ under the constraint

$$\langle R_{\mathcal{V}}(\psi), v \rangle + b(u_h, v) = l(v) \quad \forall v \in \mathcal{V}. \quad (25)$$

It can be shown that $J(\psi)$ is strictly convex in ψ .

Consider the functional $I : \mathcal{V} \times \mathcal{U}_h \rightarrow \mathbb{R}$, defined by

$$I(\psi, u_h) = J(\psi) + b(u_h, \psi) - l(\psi). \quad (26)$$

We shall study the sup-inf and inf-sup problems:

$$\sup_{u_h \in \mathcal{U}_h} \inf_{\psi \in \mathcal{V}} I(\psi, u_h) \quad \text{and} \quad \inf_{\psi \in \mathcal{V}} \sup_{u_h \in \mathcal{U}_h} I(\psi, u_h).$$

Let us look at the sup-inf problem first. For a given $u_h \in \mathcal{U}_h$, $I(\psi, u_h)$ is the sum of a strictly convex functional and a linear functional in ψ ; hence it is strictly convex in ψ . Then $\inf_{\psi \in \mathcal{V}} I(\psi, u_h)$ is achieved for the unique ψ that satisfies (25), i.e., vanishing of Gâteaux derivative. Take the test function $v = \psi$ in (25), we get

$$b(u_h, \psi) - l(\psi) = -\langle R_{\mathcal{V}}(\psi), \psi \rangle. \quad (27)$$

From the expression for $R_{\mathcal{V}}(\psi)$, (22), we have

$$\langle R_{\mathcal{V}}(\psi), \psi \rangle = \|\psi\|_{\mathcal{V}}^p = pJ(\psi). \quad (28)$$

Thus

$$\sup_{u_h \in \mathcal{U}_h} \inf_{\psi \in \mathcal{V}} I(\psi, u_h) = \sup_{u_h \in \mathcal{U}_h} (1-p)J(\psi(u_h)) \quad (29)$$

where ψ is related to u_h by (25). Since we assume $p \leq 2$, we have $1-p < 0$, and minimization of $J(\psi)$ is equivalent to maximization of $(1-p)J(\psi)$. Therefore, our residual minimization problem can be recast as the sup-inf problem

$$\sup_{u_h \in \mathcal{U}_h} \inf_{\psi \in \mathcal{V}} I(\psi, u_h).$$

Next we examine the inf-sup problem. For any given $\psi \in \mathcal{V}$, $I(\psi, u_h)$ is affine in u_h . Its supremum is $+\infty$ unless $b(\delta u_h, \psi) = 0 \quad \forall \delta u_h \in \mathcal{U}_h$. Thus

$$\inf_{\psi \in \mathcal{V}} \sup_{u_h \in \mathcal{U}_h} I(\psi, u_h) = \inf_{\psi \in (BU_h)^\perp} J(\psi) - l(\psi) \quad (30)$$

where

$$(BU_h)^\perp := \{v \in \mathcal{V} \mid \langle B\delta u_h, v \rangle = 0 \quad \forall \delta u_h \in \mathcal{U}_h\}$$

is the common definition of orthogonal complement. The inf-sup problem is now turned into a standard convex minimization problem, and we invoke a classical result from convex analysis (See Proposition 1.2 in Chapter 2, [6]).

Lemma 1. (*Existence and Uniqueness of Solutions*)

Let V be a reflexive Banach space, and \mathcal{C} a non-empty closed convex subset of V . Assume $F : \mathcal{C} \rightarrow \mathbb{R}$ is convex, lower semi-continuous and proper. Moreover, the function F is coercive over \mathcal{C} , i.e.

$$\lim_{\|u\| \rightarrow \infty} F(u) = +\infty \quad \text{for } u \in \mathcal{C}.$$

Then the problem $\inf_{u \in \mathcal{C}} F(u)$ has at least one solution. It has a unique solution if the function F is strictly convex over \mathcal{C} .

For our problem (30), $V = \mathcal{V}$ is reflexive Banach space, and $\mathcal{C} = (BU_h)^\perp \subset \mathcal{V}$ is non-empty, closed, and convex. $F(\psi) = J(\psi) - l(\psi)$ is strictly convex, continuous (thus lower semi-continuous), and proper. By ‘‘proper’’ we mean it nowhere takes the value $-\infty$ and is not identically equal to $+\infty$. F is also coercive, since $J(\psi) = \frac{1}{p}\|\psi\|^p$ while $l(\psi)$ grows only linearly in $\|\psi\|$. Therefore problem (30) has a unique solution. To characterize the unique solution, we invoke another standard result from convex analysis (Proposition 2.1 in Chapter 2, [6]).

Lemma 2. (*Characterization of Solutions*)

We assume that the function F satisfies the condition in Lemma 1, and is Gâteaux differentiable with continuous derivative F' . Then if $u \in \mathcal{C}$, u is a solution of $\inf_{u \in \mathcal{C}} F(u)$ if and only if

$$\langle F'(u), v - u \rangle \geq 0 \quad \forall v \in \mathcal{C}.$$

Denote the unique solution of (30) by ψ^* . Then the lemma requires

$$\langle F'(\psi^*), \psi - \psi^* \rangle \geq 0 \quad \forall \psi \in (BU_h)^\perp. \quad (31)$$

Let $v = \psi - \psi^*$. Then v can take any value in $(BU_h)^\perp$. The optimality condition (31) is true if and only if

$$\langle F'(\psi^*), v \rangle = 0 \quad \forall v \in (BU_h)^\perp. \quad (32)$$

This is equivalent to $F'(\psi^*) \in ((B\mathcal{U}_h)^\perp)^\perp = B\mathcal{U}_h$. The last equality is true since $B\mathcal{U}_h$ is finite dimensional, thus closed. Therefore, there exists $u_h^* \in \mathcal{U}_h$, such that

$$F'(\psi^*) + Bu_h^* = 0 \quad \text{in } \mathcal{V}'. \quad (33)$$

Plugging in the expression for F , and applying the functional on test function $v \in \mathcal{V}$, we get

$$\langle R_{\mathcal{V}}(\psi^*), v \rangle + b(u_h^*, v) = l(v) \quad \forall v \in \mathcal{V}. \quad (34)$$

Moreover, ψ^* need to satisfy the orthogonality condition

$$b(\delta u_h, \psi^*) = 0 \quad \forall \delta u_h \in \mathcal{U}_h \quad (35)$$

which is equivalent to $\psi^* \in (B\mathcal{U}_h)^\perp$. In summary, the unique solution ψ^* to the inf-sup problem (30) can be obtained by solving the mixed system

$$\begin{cases} \text{Find } \psi \in \mathcal{V}, u_h \in \mathcal{U}_h : \\ \langle R_{\mathcal{V}}(\psi), v \rangle + b(u_h, v) = l(v) & \forall v \in \mathcal{V} \\ b(\delta u_h, \psi) = 0 & \forall \delta u_h \in \mathcal{U}_h. \end{cases} \quad (36)$$

On the other hand, the mixed problem admits a unique solution in ψ , and $Bu_h^* = l - R_{\mathcal{V}}(\psi^*)$ is also uniquely determined. Since B is an injection, u_h^* is also unique.

Finally we prove the equivalence of the sup-inf and inf-sup problem. It suffices to prove the existence of a saddle point (See Theorem 7.16-1, [3]). Let (ψ^*, u_h^*) solves the mixed problem (36). We claim that it is a saddle point of $I(\psi, u_h)$, i.e.,

$$\sup_{u_h \in \mathcal{U}_h} I(\psi^*, u_h) = I(\psi^*, u_h^*) = \inf_{\psi \in \mathcal{V}} I(\psi, u_h^*). \quad (37)$$

By direct calculation,

$$I(\psi^*, u_h) - I(\psi^*, u_h^*) = b(u_h - u_h^*, \psi^*) = 0.$$

Thus the first equality in (37) holds. On the other hand,

$$I(\psi^*, u_h^*) = \inf_{\psi \in \mathcal{V}} I(\psi, u_h^*)$$

because ψ^* makes the Gâteaux derivative of a convex functional vanish. We have proved that (ψ^*, u_h^*) is indeed a saddle point of I . Therefore,

$$\sup_{u_h \in \mathcal{U}_h} \inf_{\psi \in \mathcal{V}} I(\psi, u_h) = I(\psi^*, u_h^*) = \inf_{\psi \in \mathcal{V}} \sup_{u_h \in \mathcal{U}_h} I(\psi, u_h). \quad (38)$$

We have managed to show that the minimum residual problem (2) is equivalent to the convex optimization problem (30), or the mixed problem (36). We will utilize this fact in our numerical computations.

3 Numerical Algorithms and Results

3.1 Discretization with broken test spaces

In our analysis above, we have only considered discrete trial space \mathcal{U}_h ; the test space \mathcal{V} is not yet discretized. We will use the standard technique in DPG to discretize \mathcal{V} : broken test spaces ([2], [5]). To stay focused,

we consider the convection-diffusion problem (8). Suppose we discretize the domain Ω with a mesh Ω_h . The collection of element boundaries ∂K for all $K \in \Omega_h$, is denoted by $\partial\Omega_h$. The broken Sobolev space is defined by

$$H^1(\Omega_h) = \{u \in L^2(\Omega) \mid u|_K \in H^1(K), K \in \Omega_h\}. \quad (39)$$

Testing (8) with discontinuous test functions $v \in H^1(\Omega_h)$ and integrating by parts elementwise, we obtain the corresponding DPG formulation

$$\begin{cases} \text{Find } u \in H_{\Gamma_{\text{out}}}^1(\Omega), t \in H^{-1/2}(\partial\Omega_h), t = \beta_n u_0 \text{ on } \Gamma_{\text{in}} : \\ (\epsilon \nabla u, \nabla v)_h - (u, \beta \cdot \nabla v)_h + \langle t, v \rangle = (f, v)_\Omega \quad \forall v \in H^1(\Omega_h) \end{cases} \quad (40)$$

where $(\cdot, \cdot)_h$ denotes L^2 inner product for broken spaces, i.e.

$$(f, g)_h = \sum_{K \in \Omega_h} \int_K fg \, dx \quad (41)$$

$(\cdot, \cdot)_\Omega$ represents $L^2(\Omega)$ inner product, and $\langle \cdot, \cdot \rangle$ stands for the duality pairing of the extra unknown – flux t with broken test functions. Flux t can be identified with the normal trace of a $\sigma \in H(\text{div}, \Omega)$ to element boundaries.

$$\langle t, v \rangle = \sum_{K \in \Omega_h} \langle \sigma|_K \cdot n_K, v|_K \rangle_{\partial K}. \quad (42)$$

When we replace Hilbert spaces with Banach spaces (11), $\mathcal{V} = W^{1,p}(\Omega_h)$, and the space for flux becomes the trace of $W^{p'}(\text{div}, \Omega)$ to element boundaries. In 1D, the flux is just numbers at vertex nodes. By introducing broken test functions, we are effectively replacing our original bilinear form with a new one, $b : \mathcal{U} \times \mathcal{V} \rightarrow \mathbb{R}$,

$$b((u, t), v) = (\epsilon \nabla u, \nabla v)_h - (u, \beta \cdot \nabla v)_h + \langle t, v \rangle. \quad (43)$$

The trial space $\mathcal{U} = \mathcal{U}_0 \times \hat{\mathcal{U}}$, where \mathcal{U}_0 denotes the original trial space when we have continuous test functions, and $\hat{\mathcal{U}}$ denotes the flux space. For ultraweak formulation, we can get (See [5] for details)

$$b((\sigma, u, \hat{\sigma}_n, \hat{u}), (\tau, v)) = (\sigma, \epsilon^{-1} \tau + \nabla v)_h + (u, \nabla \cdot \tau + \epsilon^{-1} \beta \cdot \tau)_h - \langle \tau \cdot n, \hat{u} \rangle - \langle \hat{\sigma}_n, v \rangle. \quad (44)$$

This will give us the operator $B : \mathcal{U} \rightarrow \mathcal{V}'$ through (24). However, now the test space \mathcal{V} is broken, so we need to study how the residual minimization problem is affected.

It suffices to consider $\mathcal{V} = \mathcal{V}_1 \times \mathcal{V}_2$, because the result easily generalizes to finite product $\mathcal{V} = W^{1,p}(\Omega_h) = \prod_{K \in \Omega_h} W^{1,p}(K)$. Suppose $\mathcal{V}_1 = W^{1,p}(\Omega_1)$, $\mathcal{V}_2 = W^{1,p}(\Omega_2)$, $\|(v_1, v_2)\|_{\mathcal{V}} := (\|v_1\|_{\mathcal{V}_1}^p + \|v_2\|_{\mathcal{V}_2}^p)^{1/p}$. We define the representation operator on \mathcal{V} , $R_{\mathcal{V}} : \mathcal{V} \rightarrow \mathcal{V}'$, by

$$\langle R_{\mathcal{V}}(v_1, v_2), (\delta v_1, \delta v_2) \rangle = \langle R_{\mathcal{V}_1}(v_1), \delta v_1 \rangle + \langle R_{\mathcal{V}_2}(v_2), \delta v_2 \rangle \quad (45)$$

where $R_{\mathcal{V}_1}, R_{\mathcal{V}_2}$ is given by (22), with Ω replaced by Ω_1, Ω_2 , respectively. Then we claim the following result.

Theorem 3. (*Representation Theorem for Product Space*)

$R_{\mathcal{V}}$ defined in (45) is well defined and bijective. Moreover,

$$\|R_{\mathcal{V}}(v_1, v_2)\|_{\mathcal{V}'} = \|(v_1, v_2)\|_{\mathcal{V}}^{p-1}.$$

Proof. By Theorem 1, $R_{\mathcal{V}_1}$, $R_{\mathcal{V}_2}$ are both bijective, $\|R_{\mathcal{V}_1}(v_1)\|_{\mathcal{V}'_1} = \|v_1\|^{p-1}$, and $\|R_{\mathcal{V}_2}(v_2)\|_{\mathcal{V}'_2} = \|v_2\|^{p-1}$. Note the isomorphism $i : \mathcal{V}' \rightarrow \mathcal{V}'_1 \times \mathcal{V}'_2$, given by $i(l) = (l_1, l_2)$, where $l_1 \in \mathcal{V}'_1$, $l_2 \in \mathcal{V}'_2$ are defined as

$$l_1(v_1) := l((v_1, 0)), \quad l_2(v_2) := l((0, v_2)).$$

Now $i(R_{\mathcal{V}}(v_1, v_2)) = (R_{\mathcal{V}_1}(v_1), R_{\mathcal{V}_2}(v_2))$. The bijectivity of $R_{\mathcal{V}}(v_1, v_2)$ follows from that of $R_{\mathcal{V}_1}(v_1), R_{\mathcal{V}_2}(v_2)$. Moreover,

$$\begin{aligned} |\langle R_{\mathcal{V}}(v_1, v_2), (\delta v_1, \delta v_2) \rangle| &\leq \|R_{\mathcal{V}_1}(v_1)\|_{\mathcal{V}'_1} \|\delta v_1\|_{\mathcal{V}_1} + \|R_{\mathcal{V}_2}(v_2)\|_{\mathcal{V}'_2} \|\delta v_2\|_{\mathcal{V}_2} \\ &= \|v_1\|_{\mathcal{V}'_1}^{p-1} \|\delta v_1\|_{\mathcal{V}_1} + \|v_2\|_{\mathcal{V}'_2}^{p-1} \|\delta v_2\|_{\mathcal{V}_2} \\ &\leq (\|v_1\|_{\mathcal{V}'_1}^p + \|v_2\|_{\mathcal{V}'_2}^p)^{(p-1)/p} (\|\delta v_1\|_{\mathcal{V}_1}^p + \|\delta v_2\|_{\mathcal{V}_2}^p)^{1/p} \\ &= \|(v_1, v_2)\|_{\mathcal{V}'}^{p-1} \|(\delta v_1, \delta v_2)\|_{\mathcal{V}}. \end{aligned}$$

The equality is achieved when $\delta v_1 = v_1, \delta v_2 = v_2$. Therefore,

$$\|R_{\mathcal{V}}(v_1, v_2)\|_{\mathcal{V}'} = \|(v_1, v_2)\|_{\mathcal{V}'}^{p-1}.$$

□

With the theorem proven, we are now in a position to analyze the minimum residual problem (2) where $\mathcal{V} = \mathcal{V}_1 \times \mathcal{V}_2 = W^{1,p}(\Omega_1) \times W^{1,p}(\Omega_2)$. Since \mathcal{V} is the product of two reflexive Banach spaces, it is a reflexive Banach space. Our argument in Section 2.3 still holds. The minimum residual problem is equivalent to the convex optimization problem (30), or mixed system (36). Finally we discretize $\mathcal{V} = W^{1,p}(\Omega_h)$ with piecewise polynomials, which are globally discontinuous.

3.2 Newton's method with line search

After discretization, we seek to solve the convex optimization problem subject to linear constraints (See (30)):

$$\begin{aligned} &\text{minimize} && f(\psi_h) \\ &\text{subject to} && b(\delta u_h, \psi_h) = 0 \quad \forall \delta u_h \in \mathcal{U}_h \end{aligned} \tag{46}$$

where $f(\psi_h) = J(\psi_h) - l(\psi_h)$, and the domain is $\mathcal{V}_h \subset \mathcal{V}$.

Following standard practice in numerical optimization, we use Newton's method to solve the problem (See Section 10.2 in [1]). Define the stiffness matrix $\mathbf{B}_{ij} := b(e_j, g_i)$, where e_j is the j -th basis function for \mathcal{U}_h , and g_i is the i -th basis function for \mathcal{V}_h . Then the linear constraint can be written as

$$\mathbf{B}\psi_h = \mathbf{0} \tag{47}$$

where ψ_h is the coefficient vector of ψ_h under the basis $\{g_1, g_2, \dots, g_n\}$. For the Newton iteration, We can always start with a feasible ψ_h . In practice, we start with $\psi_h = 0$. The Newton step $\Delta\psi_{nt}$ at feasible ψ_h is characterized by

$$\begin{bmatrix} \nabla^2 f(\psi_h) & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\psi_{nt} \\ \mathbf{u}_h \end{bmatrix} = \begin{bmatrix} -\nabla f(\psi_h) \\ \mathbf{0} \end{bmatrix}. \tag{48}$$

This is similar to what we obtain for DPG in Hilbert space ([5]). With the broken test spaces, the Newton step of representation function $\Delta\psi_{nt}$ can be condensed out element-wise. We assemble and solve the linear

system for u_h ; then we compute $\Delta\psi_{nt}$ locally. After obtaining $\Delta\psi_{nt}$, we do a backtracking line search to ensure the Armijo sufficient decrease condition (See Section 9.2 in [1]):

$$f(\psi_h + t\Delta\psi_{nt}) \leq f(\psi_h) + \alpha t \nabla f(\psi_h)^T \Delta\psi_{nt} \quad (49)$$

where α is some constant in $(0, 1)$. In our computations, we choose $\alpha = 10^{-4}$.

The Newton decrement is defined as

$$\lambda(\psi_h) = (\Delta\psi_{nt}^T \nabla^2 f(\psi_h) \Delta\psi_{nt})^{1/2} \quad (50)$$

and serves as an error indicator for Newton's method. We stop the Newton iteration when λ is small enough. The tolerance is set to 10^{-5} in our numerical experiments.

3.3 Numerical results for 1D problem

As an illustration, we solve the 1D convection-dominated diffusion problem:

$$\begin{cases} -\epsilon u'' + u' &= 0 & \text{in } (0, 1) \\ -\epsilon u' + u &= 1 & \text{at } x = 0 \\ u &= 0 & \text{at } x = 1. \end{cases} \quad (51)$$

We set $\epsilon = 10^{-2}$, and use a uniform mesh consisting of 5 elements. Second-order polynomials are used as basis for \mathcal{U}_h , and for \mathcal{V}_h we use 4-th order polynomials. Figure 1 shows the solution of standard Galerkin formulation for different choices of p . When $p = 2$, our method coincides with DPG method in Hilbert spaces. In this case, the numerical solution exhibits a clear overshoot near the boundary layer at $x = 1$. This is known as Gibbs phenomenon, which occurs when we approximate discontinuous / boundary-layer problems with continuous functions. As we increase p (in the $W^{1,p}$ norm for the test space), the overshoot is visibly reduced. It can also be seen that the solution does not change much if we further increase p over 4.

For ultraweak formulation, $\mathcal{U}_h = L^{p'}(\Omega)$ is discretized with piecewise linear polynomials, in accordance with the exact sequence logic. For \mathcal{V}_h we still use 4-th order polynomials. Figure 2 illustrates behavior of the solution as we increase p . The overshoot is again visibly reduced. We didn't go for higher p like $p = 8$ due to ill-conditioning.

We implement h-adaptivity based on ψ . The following greedy algorithm is used:

- Solve the problem on the current mesh.
- Compute element residual $\|\psi\|^p$, and mark all elements that have residual larger than $\frac{1}{4}\|\psi\|_{\max}^p$.
- Refine each marked element; continue.

Solution of the problem using h-adaptivity is shown in Figure 3. ϵ and polynomial orders are unchanged, and p is set to 4. We start with a mesh consisting of 5 uniformly-spaced elements. The element containing the boundary layer is refined in each iteration step. This verifies that the largest error comes from elements near the boundary layer, as we observe in Figure 1 and 2. After 4 refinements, the numerical solution and the exact one is almost indistinguishable.

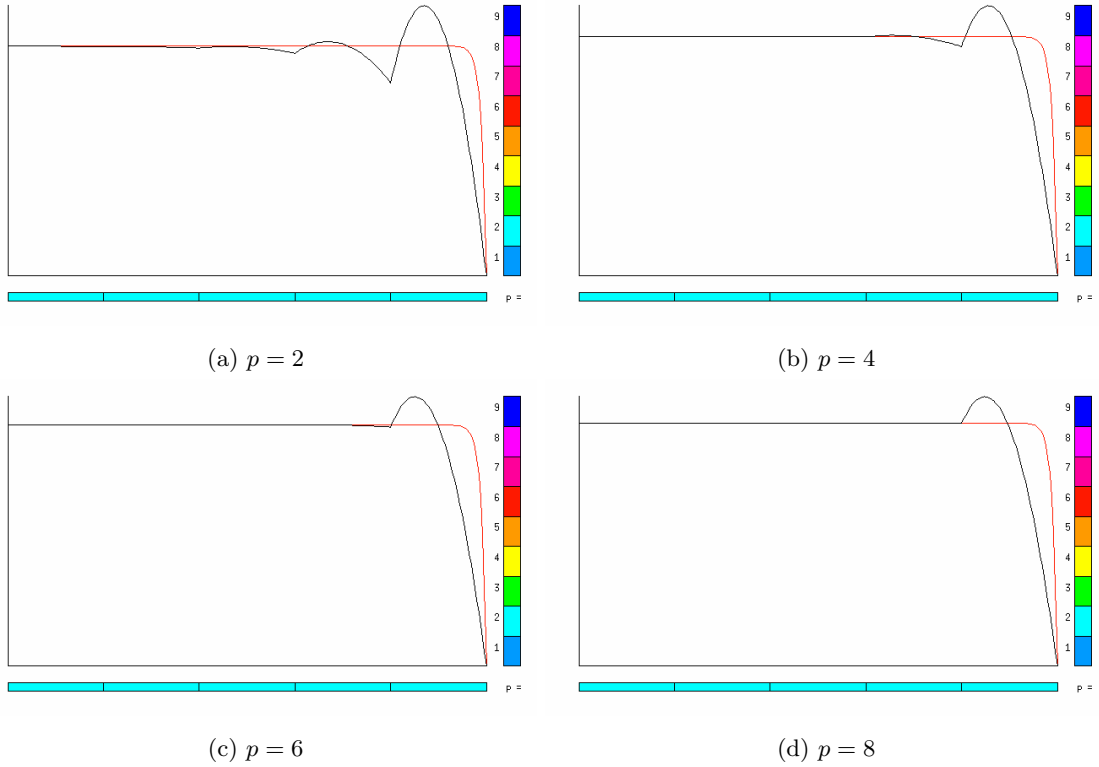


Figure 1: Solution of standard Galerkin formulation for different p . The black line represents numerical solution, and the red line stands for the exact one.

4 Conclusion

We have presented a generalization of the DPG method to the case of reflexive Banach test spaces $W^{1,p}(\Omega)$ with $p \geq 2$. Our generalization is based on minimum residual method. The relation between the residual minimization problem and the mixed problem is explored, with the introduction of a Lagrangian $I(\psi, u_h)$.

With broken test spaces and Newton's method, we get a linear system similar to what we have for DPG in Hilbert space. Error representation function can be condensed out elementwise, and we solve for u_h first; then ψ is computed in each element. We have performed numerical experiments by solving 1D convection-dominated diffusion. It is demonstrated that the DPG method in Banach space indeed reduces the overshoot. h-adaptivity based on the error representation function works well. Future work is to apply our method to 2D and 3D problems. In the meanwhile, new theory regarding traces of Banach spaces may need to be developed.

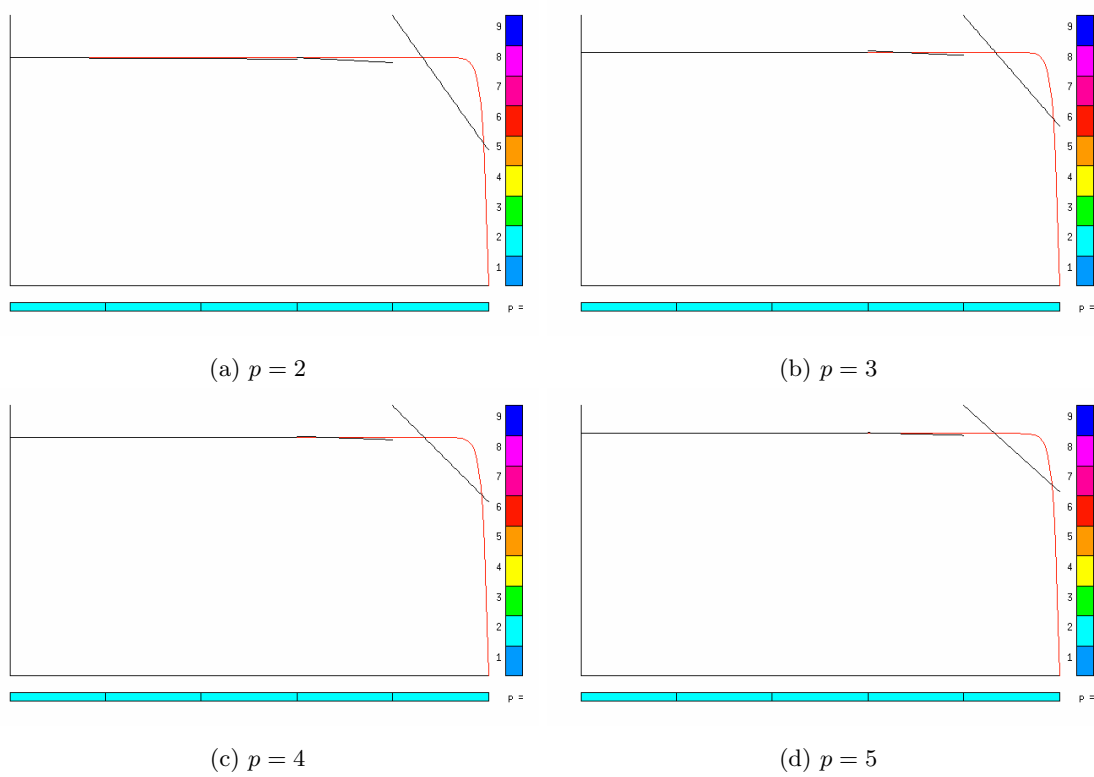


Figure 2: Solution of ultraweak formulation for different p . The black line represents numerical solution, and the red line stands for the exact one.

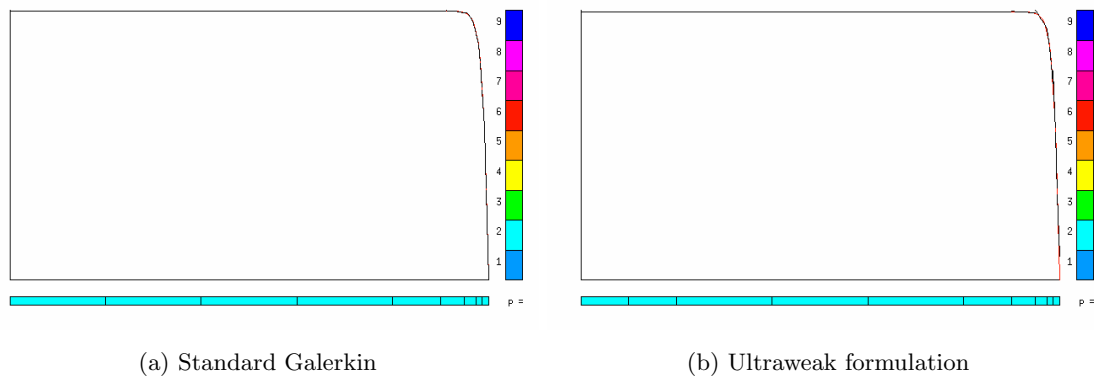


Figure 3: h-adaptivity result after 4 refinements. Starting mesh consists of 5 elements and is uniformly spaced. $p = 4$. The black line represents numerical solution, and the red line stands for the exact one.

Appendix. Proof of representation theorem for Banach space

In the appendix, we present proof of Theorem 1. First consider $\mathcal{V} = W^{1,p}(\Omega)$. From the expression of $R_{\mathcal{V}}$, (22), we have

$$\begin{aligned}
 \left| \sum_{|\alpha| \leq 1} \int_{\Omega} |D^{\alpha} v|^{p-2} D^{\alpha} v D^{\alpha} \delta v \right| &\leq \sum_{|\alpha| \leq 1} \int_{\Omega} |D^{\alpha} v|^{p-1} |D^{\alpha} \delta v| \\
 &\leq \sum_{|\alpha| \leq 1} \left(\int_{\Omega} |D^{\alpha} v|^p \right)^{\frac{p-1}{p}} \left(\int_{\Omega} |D^{\alpha} \delta v|^p \right)^{1/p} \\
 &= \sum_{|\alpha| \leq 1} \|D^{\alpha} v\|_{L^p(\Omega)}^{p-1} \|D^{\alpha} \delta v\|_{L^p(\Omega)} \\
 &\leq \left(\sum_{|\alpha| \leq 1} \|D^{\alpha} v\|_{L^p(\Omega)}^p \right)^{\frac{p-1}{p}} \left(\sum_{|\alpha| \leq 1} \|D^{\alpha} \delta v\|_{L^p(\Omega)}^p \right)^{1/p} \\
 &= \|v\|_{\mathcal{V}}^{p-1} \|\delta v\|_{\mathcal{V}}.
 \end{aligned} \tag{52}$$

Equality is achieved for $\delta v = v$. Therefore $R_{\mathcal{V}}(v) \in \mathcal{V}'$, and

$$\|R_{\mathcal{V}}(v)\|_{\mathcal{V}'} = \|v\|_{\mathcal{V}}^{p-1}. \tag{53}$$

It remains for us to prove $R_{\mathcal{V}} : \mathcal{V} \rightarrow \mathcal{V}'$ is one to one and onto. Note that $R_{\mathcal{V}}$ is the Gâteaux derivative of $J : \mathcal{V} \rightarrow \mathbb{R}$, and J is strictly convex. Thus $R_{\mathcal{V}}$ is strictly monotonic, i.e.

$$\langle R_{\mathcal{V}}(u) - R_{\mathcal{V}}(v), u - v \rangle > 0 \quad \text{for } u \neq v. \tag{54}$$

For $u \neq v$, we have $R_{\mathcal{V}}(u) - R_{\mathcal{V}}(v) \neq 0$, hence the injectivity of $R_{\mathcal{V}}$. Finally, to prove that $R_{\mathcal{V}}$ is surjective, consider the variational problem: given $l \in \mathcal{V}'$,

$$\begin{cases} \text{Find } v \in \mathcal{V} : \\ \langle R_{\mathcal{V}}(v), \delta v \rangle = l(\delta v) \quad \forall \delta v \in \mathcal{V} \end{cases} \tag{55}$$

Due to the strict convexity of J , the variational problem (55) is equivalent to the minimization problem

$$v = \arg \min_{w \in \mathcal{V}} J(w) - l(w) \tag{56}$$

which has a unique solution by Lemma 1. The arguments for $\mathcal{V} = W^p(\text{div}, \Omega) \times W^{1,p}(\Omega)$ are identical.

Acknowledgments. J. Li and L. Demkowicz were partially supported with NSF grant No. 1819101.

References

- [1] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, 2004.
- [2] C. CARSTENSEN, L. DEMKOWICZ, AND J. GOPALAKRISHNAN, *Breaking Spaces and Forms for the DPG Method and Applications including Maxwell Equations*, *Comput. Math. Appl.*, 72 (2016), pp. 494–522.
- [3] P. G. CIARLET, *Linear and Nonlinear Functional Analysis with Applications*, vol. 130, SIAM, 2013.
- [4] L. DEMKOWICZ, *Various Variational Formulations and Closed Range Theorem*, ICES Report, 15-03 (2015).
- [5] L. DEMKOWICZ AND J. GOPALAKRISHNAN, *Discontinuous Petrov-Galerkin (DPG) Method*, ICES Report, 15-20 (2015).
- [6] I. EKELAND AND R. TEMAM, *Convex Analysis and Variational Problems*, vol. 28, SIAM, 1999.
- [7] R. E. MEGGINSON, *An Introduction to Banach Space Theory*, Graduate Texts in Mathematics, Springer, 1 ed., 1998.
- [8] I. MUGA AND K. G. VAN DER ZEE, *Discretization of Linear Problems in Banach Spaces: Residual Minimization, Nonlinear Petrov-Galerkin, and Monotone Mixed Methods*, arXiv e-prints, (2015), p. arXiv:1511.04400.
- [9] I. STAKGOLD AND M. J. HOLST, *Green's Functions and Boundary Value Problems*, vol. 99, John Wiley & Sons, 2011.