The Dissertation Committee for Brendan Keith
certifies that this is the approved version of the following dissertation:

New ideas in adjoint methods for PDEs: A saddle-point paradigm for finite element analysis and its role in the DPG methodology

Committee:

Leszek F. Demkowicz, Supervisor

George Biros

Thomas J. R. Hughes

J. Tinsley Oden

Nathan V. Roberts

New ideas in adjoint methods for PDEs: A saddle-point paradigm for finite element analysis and its role in the DPG methodology

by

Brendan Keith

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2018

New ideas in adjoint methods for PDEs: A saddle-point paradigm for finite element analysis and its role in the DPG methodology

by

Brendan Keith
The University of Texas at Austin, 2018

Supervisor: Leszek F. Demkowicz

This dissertation presents a novel framework for the construction and analysis of finite element methods with trial and test spaces of unequal dimension. At the heart of this work is a new duality theory suitable for variational formulations with non-symmetric functional settings. The primary application of this theory, in this dissertation, is the development and analysis of discontinuous Petrov–Galerkin (DPG) finite element methods.

This dissertation introduces the DPG* finite element method: the dual to the DPG method. DPG, as a methodology, can be viewed as a practical means to solve overdetermined discretizations of boundary value problems. In a similar way, DPG* delivers a methodology for underdetermined discretizations. Supporting this new finite element method are new results on *a priori* error estimation and *a posteriori* error control. Notably, it is demonstrated that the convergence of a DPG* method is controlled, in part, by a Lagrange multiplier variable which plays the role of the solution variable in DPG methods. An important new result on *a posteriori* error control for DPG methods and comparisons with other related methods are also featured.

The theory developed here is applied to two representative problems coming from linear and nonlinear partial differential equation (PDE) models. To facilitate a thorough mathematical

analysis, Poisson's equation is considered. To demonstrate the utility of the approach in less tractable scenarios, the Oldroyd-B fluid model is also considered. Taken together, the combined analysis of these two models effectively demonstrates the utility of the newly developed paradigm.

Extensive computational experiments support the theoretical work presented in this dissertation. In these experiments, h- and hp-adaptive mesh refinement play a central role. For standard solution-oriented adaptive mesh refinement, local error contributions coming from a global *a posteriori* error estimate are selected to mark individual elements for refinement. For goal-oriented adaptive mesh refinement, local contributions coming from both a primal (DPG) and a dual (or adjoint; DPG*) problem are combined to deliver effective refinement strategies for linear output functionals, also known as *quantities of interest*.

Table of Contents

Abstract **v**

Chapter 1. Introduction **1**

1.1 Objective 1
1.2 Introduction 1
1.3 An illustrative example 3
1.4 Outline 6

Chapter 2. Preliminaries **7**

2.1 Remarks on notation 7
2.2 The Riesz operator 8
2.3 Operator equations 10
2.4 Assumptions 14

Chapter 3. Duality **17**

3.1 Abstract boundary value problems 17
3.2 Duality and the influence function(s) 18
3.3 Duality in the saddle-point setting 19
3.4 Minimum residual principles 20
3.5 Dual minimization principles 21
3.6 The mixed method interpretation 22
3.7 The optimal norm 23
3.8 Nonlinear problems 25

Chapter 4. DPG and DPG* methods **27**

4.1 Practical methods 27
4.2 Ultraweak variational formulations 28
4.3 Solving the primal and dual problems *simultaneously* 35
4.4 Related methods 36
4.4.1 Least-squares and $\mathcal{L}\mathcal{L}^*$ methods 37
4.4.2 Weakly conforming least-squares methods 38
4.4.3 Coercive problems 39

Chapter 5. Examples **41**
5.1 Example 1: Poisson's equation 41
5.2 Example 2: Viscoelastic fluid flow 44

Chapter 6. A priori error estimation **49**
6.1 Mixed methods 49
6.2 The error in a quantity of interest 50
6.3 DPG methods 51
6.4 DPG* methods 52
6.5 Application to Poisson's equation 53

Chapter 7. A posteriori error control **59**
7.1 Abstract stability analysis 59
7.2 The error in a quantity of interest 61
7.3 Reliability and efficiency of a DPG error estimator 62
7.4 Reliability and efficiency of a DPG* error estimator 67
7.4.1 Proofs of Lemmas 7.12 and 7.13 69
7.5 Adaptive mesh refinement 76
7.5.1 Refinement indicators and refinement strategies 77
7.5.2 Marking strategies 79

Chapter 8. Implementation **81**
8.1 Forming the saddle-point linear system 81
8.2 The overdetermined system 82
8.3 The underdetermined system 83
8.4 Solution algorithms for the overdetermined system 84
8.4.1 The normal equation 84
8.4.2 Orthogonal decompositions 85
8.4.3 Generalized least-squares 86

Chapter 9. Numerical experiments **89**
9.1 A DPG* method for Poisson's equation 89
9.1.1 Set-up 90
9.1.2 Pure Dirichlet boundary conditions on a square domain 91
9.1.3 Mixed boundary conditions on an L-shaped domain 94
9.2 A study on goal-oriented adaptive mesh refinement 96

9.2.1 Set-up . . . 96
9.2.2 Experimental design . . . 98
9.2.3 Temperature in a subdomain . . . 99
9.2.4 Flux in a subdomain . . . 101
9.2.5 Temperature on the boundary . . . 103
9.2.6 Flux on the boundary . . . 105
9.2.7 Temperature at a point . . . 107
9.3 A DPG method for viscoelastic fluid flow . . . 109
9.3.1 Set-up . . . 110
9.3.2 Creeping flow with the Oldroyd-B model . . . 112
9.3.3 Effects of inertia in the Oldroyd-B model . . . 121
9.3.4 Creeping flow with the Giesekus model . . . 123
9.3.5 Goal-oriented adaptive mesh refinement . . . 124

Chapter 10. Conclusion **127**
10.1 Discussion . . . 127
10.2 Final remarks . . . 129

Addendum. Other work **131**
Orientation embedded high order shape functions for the exact sequence elements of all shapes . . . 131
The DPG methodology applied to different variational formulations of linear elasticity 132
Coupled variational formulations of linear elasticity and the DPG methodology . . . 134
On perfectly matched layers for discontinuous Petrov–Galerkin methods . . . 135

Bibliography **139**

Chapter 1

Introduction

1.1 Objective

The goal of this dissertation is to present a new saddle-point paradigm suitable for the analysis of discontinuous Petrov–Galerkin (DPG) finite element methods and the associated development of adjoint techniques. The novel framework described here is strongly informed by the duality and optimal control ideas appearing in [10, 77, 110] and is sufficiently general to be applicable to methods outside the purview of the DPG methodology. Although features of the approach taken here have appeared in several contexts in previous research on DPG methods, it was first realized in the present form in [93] and therein led to the discovery of DPG* methods, the duals of this class of methods [48, 87].

1.2 Introduction

In an engineering environment, the prediction of specific quantities of engineering or scientific interest is often the reason for performing a computer simulation. When handling physical models, such *quantities of interest* are usually expressed as functional outputs defined on the space of solution variables. It is well known that such quantities can be expressed using either the solution of a primal problem, coming from the chosen formulation of the model at hand, or using a solution of the complementary adjoint (or dual) problem. Exploiting this observation has led to efficient computational strategies in optimization, uncertainty quantification, and adaptive mesh refinement. The dual problem is also of significant importance in deriving *a priori* and *a posteriori* error estimates in the analysis of finite element methods. In this dissertation, we focus on only the latter three of these aspects.

Given the central role of adjoint methods in engineering, it is important to state the central distinguishing feature of this work. In this dissertation, both the primal and dual problem are embedded in a saddle-point system rather than formulated and discretized directly, as is usually the case. Embedding the underlying equations in the way proposed here introduces auxiliary solution variables. In one case, the embedding delivers a variable which naturally corresponds to the residual, meanwhile, in the other case, the embedding delivers a Lagrange multiplier variable. The first strategy can be identified with a minimum residual problem. The second strategy can be identified with a constrained minimum norm problem.

One of the advantages of this new approach is that it allows the dual problem to be ill-posed. Indeed, many contemporary adjoint methods for PDEs require the physical model to induce an isomorphism between two Hilbert spaces, say U and U'. Instead, the only similar assumption we require is that model be described by an inf-sup stable continuous bilinear form $b : U \times V \to \mathbb{R}$ or, equivalently, a continuous bounded below operator $B : U \to V'$. Let us note here that the space U will not necessarily be the same as V. (*This is known as a non-symmetric functional setting.*) Because the strategy here involves the solution of additional unknowns, it also has additional computation expense. Nevertheless, as we will demonstrate, the majority of this extra expense can be eliminated via the DPG methodology.

The DPG methodology—in its present context, with optimal test functions—was discovered by Demkowicz and Gopalakrishnan in 2009 [44, 46]. In the interim, it has proven its merit in many areas of engineering interest. For instance, consider [30, 31, 65–67, 71–74, 90, 105, 120, 129, 130]. Desirable features of DPG methods include intrinsic numerical stability [34, 46, 52], a positive-definite stiffness matrix for all well-posed boundary value problems [91], flexibility in the choice of variational formulation for the treatment of said boundary value problems [28, 67, 89], and a "built-in" *a posteriori* error estimator which can be used for adaptive mesh refinement [27, 50]. The methodology incorporates a user-defined Riesz operator, which, if chosen well, allows DPG methods to deliver a nearly optimal projection of the solution variable in a prescribed norm [18, 69, 144]. Moreover, by exploiting properties of the underlying variational formulation, some DPG methods also permit highly irregular polytopal meshes [137].

In one sense, DPG methods can be viewed as practical means for solving overdetermined discretizations of partial differential equations (PDEs). In a similar way, DPG* delivers a methodology for underdetermined discretizations. Thus, together, these two methods complete one possible perspective on the construction of finite element methods with trial and test spaces of unequal dimension.

1.3 An illustrative example

The concept of a variational formulation with a non-symmetric functional setting is not included in most standard courses or textbooks on finite element analysis, even when alternative variational formulations are a focus; see, e.g., [15, 112]. One reason for this is simply because most finite element methods use closely related trial and test spaces. This is even the case for many methods with an important Petrov–Galerkin interpretation, such as SUPG [20]. Therefore, this section serves to briefly introduce the concept of a non-symmetric functional setting for unfamiliar readers and to justify the level of generality considered throughout this dissertation.

Setting the stage for the analysis in the subsequent chapters, in this section, several non-symmetric variational formulations are derived from a single toy differential equation. Here, non-uniqueness of the solution to each dual problem is witnessed because of the restrictive boundary conditions in the original primal problem. Many PDEs of general interest are underdetermined like the dual problems described here [5], so it is important to note that non-uniqueness should not be understood as an artifact only arising via an adjoint method. Lastly, before we begin, we note that it is quite ordinary for several well-posed non-symmetric formulations to arise from a single PDE [28, 43, 65, 89]. In this section, we derive two such variational formulations for both the primal and dual problems.

Let $\Omega = (0, 1)$. Consider the following one-dimensional differential equation:

$$\text{(1)} \qquad \begin{cases} -\dfrac{\mathrm{d}}{\mathrm{d}x}u(x) = f(x) \quad \forall\, x \in \Omega\,, \\ \qquad u(0) = u(1) = 0\,. \end{cases}$$

If $f \in L^2(\Omega)$, then clearly $\frac{\mathrm{d}}{\mathrm{d}x}u \in L^2(\Omega)$. Thus, by the J.L. Lions Lemma [36], the solution u

must be a member of the set $H^1(\Omega)$. Multiplying the first equation with any arbitrary function ν and integrating over the unit interval, we arrive at the following variational equation:

$$-\int_0^1 \frac{\mathrm{d}}{\mathrm{d}x}u(x)\nu(x)\,\mathrm{d}x = \int_0^1 f(x)\nu(x)\,\mathrm{d}x \qquad \forall\,\nu\,.$$

Note that in order for these integrals to be well-defined, ν must be a member of $L^2(\Omega)$. Formally integrating by parts, we find

$$\int_0^1 u(x)\frac{\mathrm{d}}{\mathrm{d}x}\nu(x)\,\mathrm{d}x = \int_0^1 f(x)\nu(x)\,\mathrm{d}x \qquad \forall\,\nu\,, \tag{2}$$

which delivers the *compatibility condition* $\int_0^1 f(x)\,\mathrm{d}x = 0$ when $\nu = 1$ is chosen.

Define $L^2_0(\Omega) = \{\phi \in L^2(\Omega) : \int_0^1 \phi(x)\,\mathrm{d}x = 0\}$. It can be shown that the following problem is a well-posed variational formulation of (1), if and only if $f \in L^2_0(\Omega)$:

$$\text{Find } u \in H^1_0(\Omega) \text{ satisfying} \quad -\int_0^1 \frac{\mathrm{d}}{\mathrm{d}x}u(x)\nu(x)\,\mathrm{d}x = \int_0^1 f(x)\nu(x)\,\mathrm{d}x \quad \forall\,\nu \in L^2(\Omega)\,. \tag{3a}$$

Obviously $H^1_0(\Omega) \neq L^2(\Omega)$. This is our first example of a variational formulation with a *non-symmetric* functional setting.

Let g be an arbitrary function in $L^2(\Omega)$. Problem (3a) has a dual which seeks a function $v \in \mathcal{V}$. This is the following:

$$\text{Find } v \in L^2(\Omega) \text{ satisfying} \quad -\int_0^1 \frac{\mathrm{d}}{\mathrm{d}x}\mu(x)v(x)\,\mathrm{d}x = \int_0^1 g(x)\mu(x)\,\mathrm{d}x \quad \forall\,\mu \in H^1_0(\Omega)\,. \tag{3b}$$

Only up to an arbitrary constant does this problem have a unique solution.

Another fundamental concept which will be introduced in this dissertation is the so-called *ultraweak* variational formulation. The derivation of such formulations is simple in the present first-order setting. First, reconsider the formal integration by parts performed in (2). Adjusting the functional setting for u and ν, the (ultra)weak formulation of (3a) is

$$\text{Find } u \in L^2(\Omega) \text{ satisfying} \quad \int_0^1 u(x)\frac{\mathrm{d}}{\mathrm{d}x}\nu(x)\,\mathrm{d}x = \int_0^1 f(x)\nu(x)\,\mathrm{d}x \quad \forall\,\nu \in H^1(\Omega)\,. \tag{4a}$$

Now, by formally integrating by parts, we see that

$$-\int_0^1 \frac{\mathrm{d}}{\mathrm{d}x}u(x)\nu(x)\,\mathrm{d}x + u(0)\nu(0) - u(1)\nu(1) = \int_0^1 f(x)\nu(x)\,\mathrm{d}x \qquad \forall\,\nu\,.$$

This observation can be used to demonstrate that the boundary condition $u(0) = u(1) = 0$ is enforced weakly. The corresponding dual problem is simply

$$\text{(4b)} \qquad \text{Find } v \in H^1(\Omega) \text{ satisfying} \quad \int_0^1 \frac{\mathrm{d}}{\mathrm{d}x} v(x)\mu(x)\,\mathrm{d}x = \int_0^1 g(x)\mu(x)\,\mathrm{d}x \quad \forall\, \mu \in L^2(\Omega)\,.$$

It can be shown that (3a) and (4a) are mutually well-posed and, likewise, (3b) and (4b) are mutually well-posed over the same equivalence class.

When any of the problems above are discretized, it is difficult to find a unifying strategy for forming a discretely stable pair of equal dimension trial and test spaces. This is due, in part, to the fact that the functional setting is not symmetric. To elucidate the present issue, suppose that a stable equal-dimension pair of spaces is chosen for (4b). Now consider (3a). Even though the functional setting is similar in the two problems, this pair cannot simply be modified by just removing trial functions to accommodate the essential boundary conditions appearing in (3a). Indeed, leaving the $L^2(\Omega)$-conforming test space fixed and removing functions from the trial space will result in a trial space with a smaller dimension than the test space. This would be an example of an overdetermined discretization. Likewise, by using the same pair of spaces when discretizing the dual problem (3b), one would uncover an underdetermined discretization.

As mentioned previously, the framework developed here permits both over- and under-determined discretizations, whether or not the underlying functional setting is symmetric. This is made possible by embedding the corresponding variational formulation into a special class of saddle-point systems. This saddle-point perspective leads to a new connection between the primal and dual problems above. Moreover, the embedding process delivers a solution strategy wherein a unique solution of the dual problem will always be selected.

It is important at this point to emphasize that the entrance of the DPG methodology into this general paradigm is secondary. In tandem with the DPG methodology, however, the overall strategy becomes more practical for computation.

1.4 Outline

The dissertation is organized as follows. Chapter 2 introduces notation and motivates several mathematical notions which will be heavily relied upon and further developed in the later chapters. In particular, it presents the general saddle-point problem used here to analyze variational formulations with non-symmetric functional settings. Chapter 3 develops useful intuition for constructing discretizations in this saddle-point framework by considering idealized semi-discrete problems. Chapter 4 introduces DPG and DPG* methods as well as ultraweak variational formulations. The entire machinery of these three chapters is then used to compare with other methods in the literature which have a similar structure. It is at this point that the dissertation narrows upon implications for DPG and DPG* methods. Chapter 5 presents two concrete examples of PDEs to which the preceding theory is later applied. The first example here is the Poisson equation and the second example comes from a common viscoelastic fluid model. Chapter 6 develops the groundwork for *a priori* error estimation with DPG and DPG* methods. Complementing the previous chapter, Chapter 7 introduces several new concepts for *a posteriori* error control with DPG and DPG* methods. Next, Chapter 8 examines solution algorithms for DPG and DPG* methods. Finally, a large number of numerical experiments are consolidated in Chapter 9. Here, adaptive mesh refinement plays a leading role in applying and verifying the preceding theory. Indeed, both classical h- and hp-adaptive mesh refinement are featured here along with an extensive study on goal-oriented adaptive mesh refinement with DPG methods. The dissertation closes with a brief summary and some concluding remarks in Chapter 10.

Remark 1.1. Chapters 2–9 of this dissertation are largely based on material from [48, 87, 90–93]. These central chapters exclude almost any mention of the following additional manuscripts submitted for publication during the author's PhD studies [67, 68, 89, 138]. Therefore, before the references appear, a brief stand-alone addendum is included to briefly summarize these supplementary contributions.

Chapter 2

Preliminaries

In this chapter, several fundamental concepts from functional analysis are introduced. Each will be essential for the theory developed in the subsequent chapters and to the coming analysis of DPG and DPG* methods.

2.1 Remarks on notation

For any Banach space X, let X' denote its topological dual, the set of all continuous linear functionals acting on X. The action of any functional, $E \in X'$, on $x \in X$ is denoted by $\langle E, x\rangle_X$. Here, the bracket $\langle \cdot, \cdot \rangle_X$ can be viewed as the natural duality pairing between X' and X. When the spaces are clear from the context, we will also use $E(x)$ to denote the same number as $\langle E, x\rangle_X$. The norm on X will be denoted $\|\cdot\|_X$. If X is Hilbert, then $\|\cdot\|_X = (\cdot,\cdot)_X^{1/2}$, where $(\cdot,\cdot)_X$ is the associated inner product. All spaces throughout this dissertation will be considered over $\mathbb{R}$, the real field, although generalizations of all statements to the case of the complex field $\mathbb{C}$ are always possible.

Let $d \in \mathbb{N}$. Throughout this dissertation, $\Omega \subseteq \mathbb{R}^d$ will always denote a bounded Lipschitz domain and $\mathcal{T}$, referred to as the *mesh*, will be a finite open disjoint partition of Ω into Lipschitz subdomains $K \in \mathcal{T}$. Specifically, $\mathcal{T}$ is collection of open subsets $K \subseteq \Omega$, $|\mathcal{T}| < \infty$, $\bigcup_{K\in\mathcal{T}} \overline{K} = \overline{\Omega}$, and $K \cap \widetilde{K} = \varnothing$, for all $K \neq \widetilde{K} \in \mathcal{T}$. Here, each $K \in \mathcal{T}$, referred to as an element, is necessarily Lipschitz. In the case $X = L^2(K)$, the equivalence class of square integrable functions on $K \subseteq \Omega$, we will specifically write $\|\cdot\|_K = (\cdot,\cdot)_K^{1/2} = \left(\int_K \cdot^2\right)^{1/2}$ instead of $\|\cdot\|_X = (\cdot,\cdot)_X^{1/2}$.

We will also often deal with bounded linear operators $\mathcal{L} \in B(X,Y)$ between Banach spaces X and Y. From now on, when dealing with quotients, such as in the definition of the norm $\|\mathcal{L}\|_{B(X,Y)} = \sup_{x\in X\setminus\{0\}} \frac{\|\mathcal{L}x\|_Y}{\|x\|_X}$, we will assume that the infima and suprema ignore

zero, wherein such quotients are not defined. Furthermore, because the meaning can often be understood by context, instead of writing out $\|\mathcal{L}\|_{B(X,Y)}$ in totality, at times we will suppress the subscript above and simply write $\|\mathcal{L}\|$.

Another notational convention for norms that we will follow involves vectors $\mathsf{x} \in \mathbb{R}^N$, where $N \in \mathbb{N}$. In this case, for any symmetric positive-definite matrix $A \in \mathbb{R}^{N\times N}$, we will denote $\|\mathsf{x}\|_{\mathsf{A}}^2 = \mathsf{x}^{\mathsf{T}}\mathsf{A}\mathsf{x}$. In the case that A is the identity matrix, we will write $\|\mathsf{x}\|_2^2 = \mathsf{x}^{\mathsf{T}}\mathsf{x}$, which is simply the canonical l^2-norm.

Finally, let $A, B \in \mathbb{R}$ denote mesh-dependent quantities. Occasionally, we will write $A \lesssim B$ or $B \lesssim A$ if there exists a constant $C > 0$, independent of the maximal mesh size, h, or the individual element size, h_K, such that $A \leq CB$ or $B \leq CA$, respectively. Similarly, $A \eqsim B$ is understood to mean that $A \lesssim B$ and $B \lesssim A$.

2.2 The Riesz operator

Let X be a Hilbert space over $\mathbb{R}$. We will use the somewhat unusual notation for the right annihilator of a subset $Y \subseteq X$ and the left annihilator of a subset $Z \subseteq X'$, respectively:

$$Y^{\perp} = \{E \in X' : \langle E, y\rangle_X = 0 \;\forall\, y \in Y\}, \tag{5}$$

$$^{\perp}Z = \{x \in X : \langle E, x\rangle_X = 0 \;\forall\, E \in Z\}. \tag{6}$$

Recall that if $Y \subseteq X$ is a closed subspace, $\overline{Y} = Y$, then $Y^{\perp}$ is isomorphic to $(X/Y)'$. This also holds if H is Banach. Finally, define the orthogonal complement of a subspace $W \subseteq X$ by

$$W_{\perp} = \{x \in X : (x, w)_X = 0 \;\forall\, w \in W\}. \tag{7}$$

Theorem 2.1 (Riesz representation theorem)**.** *Let X be a Hilbert space. Then X is isometrically isomorphic to its topological dual, X'. Moreover, in the real case, the isomorphism $\mathcal{R}_X : X \to X'$ can be defined explicitly as*

$$\langle \mathcal{R}_X\, x, \widetilde{x}\rangle_X = \langle \mathcal{R}_X\, \widetilde{x}, x\rangle_X = (x, \widetilde{x})_X = (\mathcal{R}_X\, x, \mathcal{R}_X\, \widetilde{x})_{X'} \quad \forall\, x, \widetilde{x} \in X\,. \tag{8}$$

We call any operator defined by (8), $\mathcal{R}_X : X \to X'$, a *Riesz operator*. With this operator now defined, we claim the following proposition relating (5) to (7).

Proposition 2.2. *Let X be a Hilbert space and let $W \subseteq X$ be a subspace. Then*

$$W^\perp = \mathcal{R}_X\, W_\perp\,.$$

Proof. Let $f_\perp \in W_\perp$ be arbitrary and define $F = \mathcal{R}_X f_\perp \in X'$. Observe that

$$F(w) = (f_\perp, w)_X = 0 \qquad \forall\, w \in W\,.$$

Therefore, $F \in W^\perp$ and $\mathcal{R}_X\, W_\perp \subseteq W^\perp$.

Similarly, let $F^\perp \in W^\perp$ be arbitrary and define $f = \mathcal{R}_X^{-1}\, F^\perp \in X$. Observe that

$$(f, w)_X = F^\perp(w) = 0 \qquad \forall\, w \in W\,.$$

Therefore, $f \in W_\perp$ and $W^\perp \subseteq \mathcal{R}_X\, W_\perp$. □

Note that $\|\,\mathcal{R}_X\, x\|_{X'} = \|x\|_X$, for all $x \in X$, $\|F\|_{X'} = \|\,\mathcal{R}_X^{-1}\, F\|_X$, for all $F \in X'$, $\mathcal{R}_X = \mathcal{R}_X'$, and $\mathcal{R}_{X'} = \mathcal{R}_X^{-1}$ (under the identification $X \cong X''$).

It is now appropriate to present the following corollary to Theorem 2.1.

Corollary 2.3. *Let $F \in X'$, where X is Hilbert, and let $W \subseteq X$ be a closed subspace. Then*

$$\sup_{x \in X} \frac{|F(x)|^2}{\|x\|_X^2} = \sup_{w \in W} \frac{|F(w)|^2}{\|w\|_X^2} + \sup_{w_\perp \in W_\perp} \frac{|F(w_\perp)|^2}{\|w_\perp\|_X^2}\,. \tag{9}$$

Proof. Let $f = \mathcal{R}_X^{-1}\, F$ and so $(f, x)_X = F(x)$, for all $x \in X$. Moreover, $\|f\|_X = \|F\|_{X'}$. If we orthogonally decompose $f = f_0 + f_\perp$, where $f_0 \in W$ and $f_\perp \in W_\perp$, then, by orthogonality,

$$F(w) = (f_0, w)_X\,, \quad \forall\, w \in W\,, \qquad \text{and} \qquad F(w_\perp) = (f_1, w_\perp)_X\,, \quad \forall\, w_\perp \in W_\perp\,.$$

Therefore, $f_0 = \mathcal{R}_W^{-1}(F|_W)$ and $f_\perp = \mathcal{R}_{W_\perp}^{-1}(F|_{W_\perp})$, where the new Riesz operators, $\mathcal{R}_W : W \to W'$ and $\mathcal{R}_{W_\perp} : W_\perp \to W_\perp'$, are well-defined because W is closed. Finally,

$$\|F\|_{X'}^2 = \|f\|_X^2 = \|f_0\|_W^2 + \|f_\perp\|_{W_\perp}^2 = \|F|_W\|_{W'}^2 + \|F|_{W_\perp}\|_{W_\perp'}^2\,. \tag{10}$$

Identification of (9) with (10) completes the proof. □

2.3 Operator equations

Central to this dissertation are the twin relatives of the operator equation

$$Bu = \ell, \tag{11}$$

given in (12a) and (12b) below. Here, $B : U \to V'$ is a bounded linear operator, U and V are Hilbert spaces, $\ell \in V'$ is given, and $u \in U$ is to be found. The two reformulations are as follows.

(12a) Find $u \in U$ and $\varepsilon \in V$ satisfying
$$\begin{cases} \mathcal{R}_V \varepsilon + Bu = \ell, \\ B'\varepsilon \qquad\quad = 0. \end{cases}$$

(12b) Find $v \in U$ and $\lambda \in V$ satisfying
$$\begin{cases} \mathcal{R}_U v - B'\lambda = 0, \\ Bv \qquad\quad = \ell. \end{cases}$$

Here $\mathcal{R}_V : V \to V'$ is the Riesz operator acting on V and, likewise, $\mathcal{R}_U : U \to U'$ is the Riesz operator acting on U.

It is immediate that if u solves (11), then with $\varepsilon = 0$ it solves (12a), revealing a relationship between (12a) and (11). The relationship between (12b) and (11) is also easy to guess: any solution (v, λ) of (12b) is such that the v component solves (11). We shall see below that, even though related, these formulations are not fully equivalent to (11). The formulation (12a) is the one on which the DPG method is based. The formulation (12b), when discretized, results in the new DPG* method, as we shall see.

Formulations (12a) and (12b) are structurally similar, differing mainly in the placement of the load ℓ. Due to this structural similarity, both formulations can be viewed at once as instantiations of the following general saddle-point problem:

(13) Find $v \in \mathcal{V}$ and $w \in \mathcal{U}$ satisfying
$$\begin{cases} \mathcal{R}_{\mathcal{V}} v + \mathcal{B} w = F\,, \\ \mathcal{B}' v \qquad\quad = G\,, \end{cases}$$

on some Hilbert spaces $\mathcal{U}, \mathcal{V}$, some bounded linear operator $\mathcal{B} : \mathcal{U} \to \mathcal{V}'$, and some given functionals $F \in \mathcal{V}'$ and $G \in \mathcal{U}'$. Indeed, with

$$\mathcal{V} = V,\ \mathcal{U} = U,\ \mathcal{B} = B,\ F = \ell,\ G = 0,$$

we obtain (12a). If instead, we set

$$\mathcal{V} = U,\ \mathcal{U} = V,\ \mathcal{B} = B',\ F = 0,\ G = \ell,$$

then we obtain (12b). Admittedly, the alternate mixed form obtained by exchanging $\mathcal{B}$ and $\mathcal{B}'$ in (13) is more natural for studying the DPG* method and even aligns with the standard notations in mixed method theory [15]. Yet, we have chosen to work with (13) to facilitate comparison with existing DPG literature where the form of (13) is more natural.

We proceed under the assumption that $\mathcal{B}$ is bounded below; i.e., there is a constant $\gamma > 0$ such that

$$\|\mathcal{B}\mu\|_{\mathcal{V}'} \geq \gamma \|\mu\|_{\mathcal{U}} \qquad \forall\, \mu \in \mathcal{U}. \tag{14}$$

Under this assumption, the mixed system (13) has a unique solution for any $F \in \mathcal{V}'$ and $G \in \mathcal{U}'$ (see e.g. [15]). Obviously (14) can also be written out as an inf-sup condition (cf. (32)).

Now consider the mixed system (13) when $G = 0$ and the related problem of finding $w \in \mathcal{U}$ satisfying

$$\mathcal{B}w = F. \tag{15}$$

The regularizing effect of the saddle-point formulation above is already evident: while (13) is always solvable under (14), the related problem (15) is solvable *provided* that F satisfies the compatibility condition $F \in (\operatorname{Null} \mathcal{B}')^{\perp}$. It is advantageous to reinterpret this observation by viewing (15) as an *overdetermined system*. Overdetermined systems are solvable only if they are consistent, i.e., they have compatible data. Irrespective of the load data, what the mixed system (13) solves can be seen by eliminating v (and recalling that $G = 0$):

$$\mathcal{B}' \mathcal{R}_{\mathcal{V}}^{-1} \mathcal{B}w = \mathcal{B}' \mathcal{R}_{\mathcal{V}}^{-1} F. \tag{16}$$

Equation (16) can be immediately identified with what is referred to as a "normal equation" in linear algebra. This is a regularized version of (15). Indeed, whenever (15) has a solution, it must be unique due to (14), and that unique solution is recovered by (16). However, notably (16) has a unique solution even when (15) does not.

Likewise, considering the case $F = 0$, we may argue that the mixed system (13) also helps us solve *underdetermined systems*. Indeed, consider

$$\mathcal{B}'v = G. \tag{17}$$

Assumption (14) implies that $\mathcal{B}'$ is surjective, so (17) is always solvable but its solution need not be unique. Thus, (17) may be viewed as an example of an underdetermined system. Similar to (16), the solution variable v can be readily eliminated from (13) (now recalling that $F = 0$):

$$\mathcal{B}'\,\mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B}w = -G. \tag{18}$$

This equation corresponds to a different normal equation (one of the second type [12]). Notice that the left-hand side operator $\mathcal{B}'\,\mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B} : \mathcal{U} \to \mathcal{U}'$ is the same in both (16) and (18) and that the solution to (18) can be recovered by the relationship $v = -\,\mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B}w$.

To further analyze how the mixed system (13) converts (17) into a uniquely solvable problem, one may decompose any solution of (17) into $\mathcal{V}$-orthogonal components:

$$v = v_0 + v_\perp, \qquad v_0 \in \operatorname{Null}\mathcal{B}', \quad v_\perp \in (\operatorname{Null}\mathcal{B}')_\perp. \tag{19}$$

Recall Proposition 2.2. Therefore,

$$(\operatorname{Null}\mathcal{B}')^\perp = \mathcal{R}_{\mathcal{V}}(\operatorname{Null}\mathcal{B}')_\perp. \tag{20}$$

Since $F = 0$, testing the first equation of (13) with v_0, we find that what (13) selects as its unique solution is in fact simply $v = v_\perp$.

Returning to the case of general F and G, we collect a few identities in the next result. First, note that one may also decompose F into orthogonal components:

$$F = F^0 + F^\perp, \qquad F^0 \in \mathcal{R}_{\mathcal{V}}(\operatorname{Null}\mathcal{B}'), \quad F^\perp \in \mathcal{R}_{\mathcal{V}}(\operatorname{Null}\mathcal{B}')_\perp = (\operatorname{Null}\mathcal{B}')^\perp. \tag{21}$$

Second, note that when (14) holds, $|||\mu|||_{\mathcal{U}} = \|\mathcal{B}\mu\|_{\mathcal{V}'}$ generates an equivalent norm on $\mathcal{U}$ and we may define

$$|||G|||_{\mathcal{U}'} = \sup_{\mu\in\mathcal{U}} \frac{\langle G, \mu\rangle_{\mathcal{U}}}{|||\mu|||_{\mathcal{U}}}.$$

The following proposition presents several important results for systems of the form (13). Identities like (27) have often been referred to by the name Prager–Synge *hypercircle identities* [121, 125] and their use in *a posteriori* error estimation is now standard.

Proposition 2.4. *Suppose $F \in \mathcal{V}'$, $G \in \mathcal{U}'$, $v \in \mathcal{V}$ and $w \in \mathcal{U}$ solve* (13) *and let v_0 and $v_\perp$ be the unique components of the decomposition of v given in* (19). *Similarly, let F^0 and $F^\perp$ be the unique components of the decomposition of F given in* (21). *The following identities then hold:*

$$\|v_0\|_{\mathcal{V}}^2 + \|\mathcal{R}_{\mathcal{V}} v_\perp + \mathcal{B}w\|_{\mathcal{V}'}^2 = \|F\|_{\mathcal{V}'}^2, \tag{22}$$

$$\|v_0\|_{\mathcal{V}}^2 + \|\mathcal{B}w\|_{\mathcal{V}'}^2 = \|F - \mathcal{R}_{\mathcal{V}} v_\perp\|_{\mathcal{V}'}^2. \tag{23}$$

Moreover, $v_0 = \mathcal{R}_{\mathcal{V}}^{-1} F^0$ and

$$\|v_0\|_{\mathcal{V}} = \|F^0\|_{\mathcal{V}'}, \tag{24}$$

$$\|\mathcal{B}w\|_{\mathcal{V}'} = \|F^\perp - \mathcal{R}_{\mathcal{V}} v_\perp\|_{\mathcal{V}'}. \tag{25}$$

If, in addition, (14) *holds, then for any $F \in \mathcal{V}'$, $G \in \mathcal{U}'$, there is a unique $v \in \mathcal{V}$ and $w \in \mathcal{U}$ satisfying* (13) *and the following identities hold:*

$$\|v_\perp\|_{\mathcal{V}} = |||G|||_{\mathcal{U}'}, \tag{26}$$

$$\|v\|_{\mathcal{V}}^2 + |||w|||_{\mathcal{U}}^2 = \|F - \mathcal{R}_{\mathcal{V}} v_\perp\|_{\mathcal{V}'}^2 + |||G|||_{\mathcal{U}'}^2. \tag{27}$$

If in addition, either $F \in (\operatorname{Null} \mathcal{B}')^\perp$ or $\mathcal{B}$ is a bijection, then $v_0 = 0$ and

$$\|v\|_{\mathcal{V}} = |||G|||_{\mathcal{U}'}. \tag{28}$$

Proof. For any $\nu_0 \in \operatorname{Null} \mathcal{B}'$, we have $(\mathcal{R}_{\mathcal{V}}^{-1}\mathcal{B}w, \nu_0)_{\mathcal{V}} = \langle \mathcal{B}w, \nu_0 \rangle_{\mathcal{V}} = \langle \mathcal{B}'\nu_0, w \rangle_{\mathcal{U}} = 0$. Hence $\mathcal{R}_{\mathcal{V}}^{-1}\mathcal{B}w$ is in $(\operatorname{Null} \mathcal{B}')_\perp$. Therefore, when the first equation of (13) is rewritten as

$$v_0 + (v_\perp + \mathcal{R}_{\mathcal{V}}^{-1}\mathcal{B}w) = \mathcal{R}_{\mathcal{V}}^{-1} F, \tag{29}$$

an application of the Pythagorean theorem gives (22). Rewriting (29) as $v_0 + \mathcal{R}_{\mathcal{V}}^{-1}\mathcal{B}w = \mathcal{R}_{\mathcal{V}}^{-1} F - v_\perp$, and applying the Pythagorean theorem again, we obtain (23). Rewriting (29) instead as

$$v_0 - \mathcal{R}_{\mathcal{V}}^{-1} F^0 = \mathcal{R}_{\mathcal{V}}^{-1} F^\perp - (v_\perp + \mathcal{R}_{\mathcal{V}}^{-1}\mathcal{B}w),$$

we note that $v_0 = \mathcal{R}_{\mathcal{V}}^{-1} F^0$ and $\mathcal{R}_{\mathcal{V}}^{-1} F^\perp = v_\perp + \mathcal{R}_{\mathcal{V}}^{-1} \mathcal{B} w$, by orthogonality. Equations (24) and (25) are now obvious.

Next, if (14) holds, then standard mixed theory [15] gives existence of a unique $(v, w) \in \mathcal{V} \times \mathcal{U}$, and $|||\cdot|||_{\mathcal{U}}$ is an equivalent norm on $\mathcal{U}$. To prove (26), we begin by noting that the isometry induced by $\mathcal{R}_{\mathcal{V}}$ implies

$$\|v_\perp\|_{\mathcal{V}} = \sup_{\nu_\perp \in (\operatorname{Null} \mathcal{B}')_\perp} \frac{(\nu_\perp, v_\perp)_{\mathcal{V}}}{\|\nu_\perp\|_{\mathcal{V}}} = \sup_{\nu_\perp \in (\operatorname{Null} \mathcal{B}')_\perp} \frac{\langle \mathcal{R}_{\mathcal{V}} \nu_\perp, v_\perp \rangle_{\mathcal{V}}}{\|\mathcal{R}_{\mathcal{V}} \nu_\perp\|_{\mathcal{V}'}} = \sup_{E^\perp \in \mathcal{R}_{\mathcal{V}} (\operatorname{Null} \mathcal{B}')_\perp} \frac{\langle E^\perp, v_\perp \rangle_{\mathcal{V}}}{\|E^\perp\|_{\mathcal{V}'}}.$$

Here and throughout, supremums over spaces are only taken over nonzero elements of the space. Again, from the identity $\operatorname{Range} \mathcal{B} = (\operatorname{Null} \mathcal{B}')^\perp$ and (20), we conclude that

$$\|v_\perp\|_{\mathcal{V}} = \sup_{E^\perp \in \operatorname{Range} \mathcal{B}} \frac{\langle E^\perp, v_\perp \rangle_{\mathcal{V}}}{\|E^\perp\|_{\mathcal{V}'}} = \sup_{\mu \in \mathcal{U}} \frac{\langle \mathcal{B}\mu, v_\perp \rangle_{\mathcal{V}}}{\|\mathcal{B}\mu\|_{\mathcal{V}'}} = \sup_{\mu \in \mathcal{U}} \frac{\langle \mu, \mathcal{B}' v_\perp \rangle_{\mathcal{V}}}{|||\mu|||_{\mathcal{U}}}.$$

Thus, (26) follows after using the second equation in (13), namely $G = \mathcal{B}'(v_0 + v_\perp) = \mathcal{B}' v_\perp$. Identity (27) now follows by squaring both sides of (26) and adding it to (23).

Finally, when $\mathcal{B}$ is a bijection or $F \in (\operatorname{Null} \mathcal{B}')^\perp$, we conclude that $F^0 = 0$. Therefore, $v_0 = 0$ and (28) follows from (26). □

2.4 Assumptions

During the analysis of finite element methods, it is convenient (cf. [6]) to write operator equations like (11) using a bilinear form defined by

$$b(\mu, \nu) = \langle \mathcal{B}\mu, \nu \rangle_{\mathcal{V}}\,, \tag{30}$$

for all $\mu \in \mathcal{U}$ and $\nu \in \mathcal{V}$. In terms of the bilinear form $b(\cdot, \cdot)$ and the inner product $(\cdot, \cdot)_{\mathcal{V}}$, the mixed problem (13) reduces to finding the unique functions $v \in \mathcal{V}$ and $w \in \mathcal{U}$ satisfying

$$\begin{cases} (v, \nu)_{\mathcal{V}} + b(w, \nu) = F(\nu) & \forall\, \nu \in \mathcal{V}, \\ b(\mu, v) \qquad\qquad\quad = G(\mu) & \forall\, \mu \in \mathcal{U}. \end{cases} \tag{31}$$

Throughout this dissertation, we will require several assumptions on b which are collected here for reference.

Assumption 1. The bilinear form $b : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ is bounded with continuity constant $M = \|\mathcal{B}\| < +\infty$, where

$$\sup_{\mu\in\mathcal{U}} \sup_{\nu\in\mathcal{V}} \frac{b(\mu,\nu)}{\|\mu\|_{\mathcal{U}}\|\nu\|_{\mathcal{V}}} = M\,.$$

Moreover, it satisfies the Banach-Babuška-Nečas inf-sup condition with stability constant $\gamma = \|\mathcal{B}^{-1}\|^{-1} > 0$, where

$$\inf_{\mu\in\mathcal{U}} \sup_{\nu\in\mathcal{V}} \frac{b(\mu,\nu)}{\|\mu\|_{\mathcal{U}}\|\nu\|_{\mathcal{V}}} = \gamma\,. \tag{32}$$

Assumption 2. For finite-dimensional subspaces $\mathcal{U}_h \subseteq \mathcal{U}$ and $\mathcal{V}_h \subseteq \mathcal{V}$, there exists a bounded linear operator $\Pi_h : \mathcal{V} \to \mathcal{V}_h$ such that

$$b(\mu, \nu - \Pi_h\nu) = 0 \qquad \forall\, \mu \in \mathcal{U}_h\,,\ \nu \in \mathcal{V}\,.$$

Assumption 3. In addition to Assumption 2, Π_h is projection, $\Pi_h \circ \Pi_h = \Pi_h$.

Assumption 4. The test space $\mathcal{V}$ is *broken*; that is, $\mathcal{V} = \prod_{K\in\mathcal{T}} \mathcal{V}_K$, where $\mathcal{V}_K = \{\nu|_K : \nu \in \mathcal{V}\}$, for all elements K in the mesh $\mathcal{T}$. Moreover, the corresponding test space norm $\|\cdot\|_{\mathcal{V}}$ is *localizable*; that is, $\|\cdot|_K\|_{\mathcal{V}}$ is also a norm, for all $K \in \mathcal{T}$.

Remark 2.5. Notice that Assumption 1 readily implies that

$$\gamma\|\mu\|_{\mathcal{U}} \le \|\mathcal{B}\mu\|_{\mathcal{V}'} \le M\|\mu\|_{\mathcal{U}} \qquad \forall\, \mu \in \mathcal{U}\,. \tag{33}$$

Remark 2.6. Although only Assumption 2 is actually required for well-posedness of DPG and DPG* methods (see Section 4.1), when Π_h is constructed in practice, it often happens to be a projection [107]. Because this additional structure can be exploited to improve some estimates (see, e.g., Theorems 7.7 and 7.8), Assumption 3 is separately included.

Chapter 3

Duality

This chapter presents the functional analysis framework which is used to formulate the notion of duality in non-symmetric functional settings underpinning this dissertation. We begin by introducing the abstract boundary value problems our construction is built to accommodate. We then continue by developing corresponding *idealized* saddle-point discretizations in the form of (13), where only the single space, $\mathcal{U}_h \subseteq \mathcal{U}$, has been discretized. These idealized semi-discrete problems naturally motivate the design of quasi-optimal graph inner products. The eventual discretization of both spaces $\mathcal{U}$ and $\mathcal{V}$ lead to the DPG and DPG* methods, which appear in the following chapter.

3.1 Abstract boundary value problems

Let $\mathcal{U}$ and $\mathcal{V}$ be Hilbert spaces over $\mathbb{R}$. In the sequel, all variational boundary value problems will be posed using a continuous bilinear form $b : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$. The members of both spaces $\mathcal{U}$ and $\mathcal{V}$ may have many components but are still routinely called *functions*. For a given functional $F \in (\operatorname{Null} \mathcal{B}')^{\perp}$, called the *load*, we define the (*primal*) *solution* to be the unique function $u \in \mathcal{U}$ satisfying

$$b(u, \nu) = F(\nu) \qquad \forall\, \nu \in \mathcal{V}\,. \tag{34}$$

Note that the bilinear form b naturally generates a continuous linear operator $\mathcal{B} : \mathcal{U} \to \mathcal{V}'$ and, taking into account the reflexivity of $\mathcal{V} \sim \mathcal{V}''$, also generates its dual $\mathcal{B}' : \mathcal{V} \to \mathcal{U}'$:

$$\langle \mathcal{B}\mu, \nu \rangle_{\mathcal{V}} = \langle \mathcal{B}'\nu, \mu \rangle_{\mathcal{U}} = b(\mu, \nu) \qquad \forall\, \mu \in \mathcal{U}\,,\ \nu \in \mathcal{V}\,. \tag{35}$$

In most practical scenarios, the observer is not solely interested in all every global feature of the solution of (34), $u = \mathcal{B}^{-1}F$. Instead, they are usually interested in a derived quantity, or

functional output, $G(u)$, called the *quantity of interest* (QOI). In this context, we choose to call the corresponding functional $G \in \mathcal{U}'$ the *goal functional*. This gives rise to the following variational problem:

$$\text{(36a)} \qquad \text{Find } G(u) \in \mathbb{R} \text{ where } u \in \mathcal{U} \text{ satisfies} \qquad b(u,\nu) = F(\nu) \quad \forall\, \nu \in \mathcal{V}\,.$$

This is a problem of central interest in this dissertation. In in next section, we will see that a complementary perspective on (36a) can be taken, using a solution v of the dual problem.

Remark 3.1. In some important scenarios, G is *not* bounded. For instance, G is usually not bounded when it involves the mean normal flux of the gradient of an H^1-variable through a subset of the domain boundary. Indeed, if the goal functional represents the drag or lift coefficient of a blunt or streamlined body (cf. (153)) and the bilinear form b corresponds to a standard continuous Galerkin finite element method for fluid flow, this is a well known concern [10]. Fortunately, in these scenarios, (36a) may be augmented to still accommodate such QOIs [77].

In this dissertation, (36a) will be sufficient to analyze *every* goal functionals we encounter, *including those incorporating boundary fluxes*, as described above. A protracted discussion on this artifact is not appropriate here, but ultimately G will always be bounded due to the use of hybridized $\mathcal{U}$-variables in all DPG and DPG* methods (see, e.g., Section 4.2). In effect, our universal use of hybridized variables in every b we construct can be related to the original technique Giles and Süli used to augment (36a) in [77].

3.2 Duality and the influence function(s)

As stated in (36a), we are interested in the quantity $G(u)$, where $b(u,\nu) = F(\nu)$, for all $\nu \in \mathcal{V}$. Under Assumption 1, the constraint above permits one unique solution $u \in \mathcal{U}$, so evaluating $G(u)$ is obviously equivalent to the problem

$$G(u) = \min\left\{G(u) \,:\, u \in \mathcal{U} \text{ and } b(u,\nu) = F(\nu)\ \forall\, \nu \in \mathcal{V}\right\}.$$

Invoking a Lagrangian $L : \mathcal{U} \times \mathcal{V} \to \overline{\mathbb{R}}$, defined as $L(\mu,\nu) = G(\mu) + F(\nu) - b(\mu,\nu)$, the quantity $G(u)$ can also be characterized as the solution of a saddle-point problem (see [56] for further

details):

$$G(u) = \min_{\mu\in\mathcal{U}} \sup_{\nu\in\mathcal{V}} L(\mu,\nu) = b(u,v) = \max_{\nu\in\mathcal{V}} \inf_{\mu\in\mathcal{U}} L(\mu,\nu) = F(v)\,,$$

where v satisfies $b(\mu, v) = G(\mu)$, for all $\mu \in \mathcal{U}$. Thus, we arrive at the following dual problem:

(36b) Find $F(v) \in \mathbb{R}$ where $v \in \mathcal{V}$ satisfies $\quad b(\mu, v) = G(\mu) \quad \forall\, \mu \in \mathcal{U}\,.$

Presently, notice that v coming from (36b) may not be unique, because we have not assumed that $\mathcal{B}$ is an isomorphism.

For an alternative derivation, define $\mathcal{B} : \mathcal{U} \to \mathcal{V}'$ through the bilinear form as in (35). Likewise, observe that

$$G(u) = \langle G, \mathcal{B}^{-1}F\rangle_{\mathcal{U}} = \langle F, (\mathcal{B}')^{-1}(\{G\})\rangle_{\mathcal{V}} = F(v)\,, \tag{37}$$

where $v \in (\mathcal{B}')^{-1}(\{G\})$ and $(\mathcal{B}')^{-1}(\{G\}) = \{\nu \in \mathcal{V} : \mathcal{B}'\nu = G\}$ is the preimage of $\{G\}$.

Notice from the definitions above that v acts like a generalized Green's function for the functional $G \in \mathcal{U}'$. From now on, we refer to any such v as an *influence function*. Moreover, notice that the QOI can be calculated from the solution of the primal problem (36a) or from a solution of the dual problem (36b).

3.3 Duality in the saddle-point setting

Problems (36a) and (36b) can easily be placed into the saddle-point setting of the previous chapter. Indeed, replicating the bilinear form b used above and introducing the inner product $(\cdot,\cdot)_{\mathcal{V}}$, we immediately arrive at the following primal problem:

(38a) Find $G(u) \in \mathbb{R}$ where $(\varepsilon, u) \in \mathcal{V} \times \mathcal{U}$ satisfies
$$\begin{cases} (\varepsilon,\nu)_{\mathcal{V}} + b(u,\nu) = F(\nu) & \forall\, \nu \in \mathcal{V}\,, \\ b(\mu,\varepsilon) \qquad\qquad = 0 & \forall\, \mu \in \mathcal{U}\,. \end{cases}$$

The corresponding dual saddle-point problem is expressed similarly:

(38b) Find $F(v) \in \mathbb{R}$ where $(v, \lambda) \in \mathcal{V} \times \mathcal{U}$ satisfies
$$\begin{cases} (v,\nu)_{\mathcal{V}} - b(\lambda,\nu) = 0 & \forall\, \nu \in \mathcal{V}\,, \\ b(\mu, v) \qquad\qquad = G(\mu) & \forall\, \mu \in \mathcal{U}\,. \end{cases}$$

Notice that the primal solution component u, coming from (38a), is the same unique function u solving (34). Similarly, the influence function v in (38b) is uniquely determined and solves the dual problem present in (36b): $b(\mu, v) = G(\mu)$, for all $\mu \in \mathcal{U}$. Therefore, the identity $G(u) = F(v)$ still holds. From now on, we are free to refer to *the* influence function v, so long as it is always identified with the unique v coming from (38b).

In Section 6.2, we demonstrate that the two errors coming from discrete versions of (38a) and (38b) inherit a convenient orthogonality property commonly exhibited between primal and dual solutions in conventional finite element methods [10, 110]. In the interim, it is appropriate to demonstrate how the saddle-point construction above informs the DPG methodology. In line with this effort, we proceed by considering idealized semi-discrete versions of (38a) and (38b).

3.4 Minimum residual principles

Let $\mathcal{U}_h \subseteq \mathcal{U}$ be a finite-dimensional subspace of the trial space. It is desirable for us to seek, in some practical sense, the *optimal* (i.e., minimal error) solution $u_h^{\text{opt}} \in \mathcal{U}_h$ to the primal problem (34). Because the exact solution $u = \mathcal{B}^{-1}F$ is inaccessible *a priori*, the optimal solution u_h^{opt} cannot be defined by explicitly invoking u. For illustration, the most convenient (but impractical) notion of "optimal solution" is the so-called best approximation error solution:

$$u_h^{\text{BAE}} = \arg\min_{\mu \in \mathcal{U}_h} \|u - \mu\|_{\mathcal{U}}^2 . \tag{39}$$

Here, the discrete solution u_h^{BAE} naturally corresponds to the orthogonal projection of the exact solution u in the trial space norm $\|\cdot\|_{\mathcal{U}}$. Meanwhile, the minimum value attained, $\|u - u_h^{\text{BAE}}\|_{\mathcal{U}}$, is referred to as the *best approximation error*. Instead of adopting the best approximation error notion (or any similar explicit concept), let us pose optimality implicitly through a minimum residual principle:

$$u_h^{\text{opt}} = \arg\min_{\mu \in \mathcal{U}_h} \|\mathcal{B}\mu - F\|_{\mathcal{V}'}^2 . \tag{40}$$

Define the *condition number* of $\mathcal{B}$ to be

$$\kappa(\mathcal{B}) = \|\mathcal{B}\|\|\mathcal{B}^{-1}\| = \frac{M}{\gamma} \geq 1 . \tag{41}$$

It can be shown that the accuracy of the optimal solution $u_h^{\rm opt}$ depends upon the condition number of $\mathcal{B}$, *viz.*,

$$\|u - u_h^{\rm BAE}\|_{\mathcal{U}} \leq \|u - u_h^{\rm opt}\|_{\mathcal{U}} \leq \kappa(\mathcal{B})\|u - u_h^{\rm BAE}\|_{\mathcal{U}}\,.$$

Observe that $\|\mathcal{B}\mu - F\|_{\mathcal{V}'}^2 = \langle \mathcal{B}\mu - F, \mathcal{R}_{\mathcal{V}}^{-1}(\mathcal{B}\mu - F)\rangle$, for all $\mu \in \mathcal{U}_h$. Therefore, the first-order optimality condition associated with (40) is equivalent to the following variational equation (cf. (16)):

$$\langle \mathcal{B}u_h^{\rm opt}, \mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B}\mu\rangle_{\mathcal{V}} = \langle F, \mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B}\mu\rangle_{\mathcal{V}} \qquad \forall\, \mu \in \mathcal{U}_h\,. \tag{42}$$

Finally, define the residual function corresponding to (40), $\varepsilon_h^{\rm opt} = \mathcal{R}_{\mathcal{V}}^{-1}(F - \mathcal{B}u_h^{\rm opt})$. Observe that the pair $(\varepsilon_h^{\rm opt}, u_h^{\rm opt})$ may be identified with the solution of the following saddle-point problem:

$$\min_{\nu\in\mathcal{V}}\max_{\mu\in\mathcal{U}_h}\left[\|\nu\|_{\mathcal{V}} + \langle \mathcal{B}\mu - F, \nu\rangle_{\mathcal{V}}\right]. \tag{43}$$

This form has been used to construct a PDE-constrained optimization approach to DPG methods in [23].

3.5 Dual minimization principles

Let us again return to (36). These problems defining u and v can be rewritten $\mathcal{B}u = F$ and $\mathcal{B}'v = G$, respectively. Applying $\mathcal{B}'\mathcal{R}_{\mathcal{V}}^{-1}$, to both sides of the former equation delivers (16). Since $F \in (\mathrm{Null}\,\mathcal{B}')^{\perp}$, we may identify w in (16) with u above. Moreover, comparison with (42) immediately demonstrates that $u_h^{\rm opt} = u$ when $\mathcal{U}_h = \mathcal{U}$. Meanwhile, if we express $v = \mathcal{R}_{\mathcal{V}}\mathcal{B}\lambda$, the latter equation delivers $\mathcal{B}'\,\mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B}\lambda = G$, which can be identified with (18) for $\lambda = -w$. With little ambiguity, we shall also refer to the Lagrange multiplier $\lambda \in \mathcal{U}$ as the (trial space) influence function. Its distinction from the (test space) influence function $v \in \mathcal{V}$ should always be clear from the context.

Define $\mathcal{A} : \mathcal{U} \to \mathcal{U}'$ by $\mathcal{A} = \mathcal{B}'\,\mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B}$ and note that $\mathcal{A} = \mathcal{A}'$ is an isomorphism. Define an inner product $a(\cdot,\cdot) : \mathcal{U}\times\mathcal{U} \to \mathbb{R}$ by $a(\mu,\widetilde{\mu}) = \langle \mathcal{A}\mu, \widetilde{\mu}\rangle_{\mathcal{U}}$, for all $\mu, \widetilde{\mu} \in \mathcal{U}$. A Bubnov-Galerkin approximation, $\widetilde{\lambda}_h$, of the influence function λ above can be characterized readily:

$$a(\widetilde{\lambda}_h, \mu) = G(\mu) \qquad \forall\, \mu \in \mathcal{U}_h\,.$$

The discrete solution above, $\widetilde{\lambda}_h$, can be seen to come from the Ritz method following from the quadratic energy principle

$$\min_{\mu\in\mathcal{U}_h}\left[\,\frac{1}{2}\,a(\mu,\mu)-G(\mu)\,\right].$$

Equivalently, the approximation can be characterized as $\widetilde{\lambda}_h=\lambda_h^{\rm opt}$, where $\lambda_h^{\rm opt}$ is defined as the optimal solution coming from the minimum residual problem

$$\lambda_h^{\rm opt}=\arg\min_{\mu\in\mathcal{U}_h}|||\mathcal{A}\mu-G|||^2_{\mathcal{U}'}\,. \tag{44}$$

Indeed, observe that $|||E|||^2_{\mathcal{U}'}=\langle E,\mathcal{A}^{-1}E\rangle_{\mathcal{U}}$, for all $E\in\mathcal{U}'$, so the first-order optimality condition for (44) delivers

$$\langle\mathcal{A}\lambda_h^{\rm opt}-G,\mathcal{A}^{-1}\mathcal{A}\mu\rangle_{\mathcal{U}}=\langle\mathcal{A}\lambda_h^{\rm opt}-G,\mu\rangle_{\mathcal{U}}=0\qquad\forall\,\mu\in\mathcal{U}_h\,.$$

Because this second perspective will be the most insightful, we will denote the approximation $\widetilde{\lambda}_h$ as $\lambda_h^{\rm opt}$ from now on.

Observe that an approximation to some influence function $v\in\mathcal{V}$, denoted $v_h^{\rm opt}$, where

$$b(\mu,v_h^{\rm opt})=G(\mu)\qquad\forall\,\mu\in\mathcal{U}_h\,, \tag{45}$$

can be recovered by post-processing $\lambda_h^{\rm opt}$. Indeed, $v_h^{\rm opt}=\mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B}\lambda_h^{\rm opt}$. Moreover,

$$v_h^{\rm opt}=\arg\min_{\nu\in\mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B}(\mathcal{U}_h)}|||\mathcal{B}'\nu-G|||_{\mathcal{U}'}\,.$$

Finally, observe that the pair $(v_h^{\rm opt},\lambda_h^{\rm opt})$ can be identified with the solution of the following saddle-point problem:

$$\min_{\nu\in\mathcal{V}}\max_{\mu\in\mathcal{U}_h}\left[\,\|\nu\|_{\mathcal{V}}-\langle\mathcal{B}'\nu-G,\mu\rangle_{\mathcal{U}}\right]. \tag{46}$$

3.6 The mixed method interpretation

Equations (43) and (46) deliver the following semi-discrete variational problems:

(47a) Find $u_h^{\rm opt}\in\mathcal{U}_h$ and $\varepsilon_h^{\rm opt}\in\mathcal{V}$ satisfying

$$\begin{cases}(\varepsilon_h^{\rm opt},\nu)_{\mathcal{V}}+b(u_h^{\rm opt},\nu)=F(\nu) & \forall\,\nu\in\mathcal{V}\,,\\ b(\mu,\varepsilon_h^{\rm opt}) \qquad\qquad\qquad\quad =0 & \forall\,\mu\in\mathcal{U}_h\,,\end{cases}$$

(47b) Find $\lambda_h^{\mathrm{opt}} \in \mathcal{U}_h$ and $v_h^{\mathrm{opt}} \in \mathcal{V}$ satisfying
$$\begin{cases} (v_h^{\mathrm{opt}}, \nu)_{\mathcal{V}} - b(\lambda_h^{\mathrm{opt}}, \nu) = 0 & \forall\, \nu \in \mathcal{V}\,, \\ b(\mu, v_h^{\mathrm{opt}}) = G(\mu) & \forall\, \mu \in \mathcal{U}_h\,. \end{cases}$$

Obviously, these idealized problems are closely associated to (38a) and (38b) above. Clearly, because $\mathcal{B}'$ may have a non-trivial null space, an important distinction between (45) and (47b) is that the latter always has a unique solution. In the following chapter, discretizing $\mathcal{V}_h$ in (47) eventually leads to "practical" (i.e., fully discrete) DPG and DPG* methods. Before moving on, however, we shortly detour to briefly consider the optimal inner product $(\cdot,\cdot)_{\mathcal{V}}$ to use in the saddle-point problems above and strategies for nonlinear problems.

3.7 The optimal norm

Let $F \in (\operatorname{Null}\mathcal{B}')^{\perp}$ be arbitrary and $u = \mathcal{B}^{-1}F$. Suppose that a norm $|||\cdot|||_{\mathcal{V}}$ exists such that

$$\|u-\mu\|_{\mathcal{U}} = |||\mathcal{B}\mu - F|||_{\mathcal{V}'} \qquad \forall\, \mu \in \mathcal{U}\,, \tag{48}$$

where

$$|||F|||_{\mathcal{V}'} = \sup_{\nu\in\mathcal{U}} \frac{\langle F, \nu\rangle_{\mathcal{V}}}{|||\nu|||_{\mathcal{V}}}\,.$$

With this norm, which we will call the *optimal norm*, both (39) and (40) coincide and so $u_h^{\mathrm{BAE}} \overset{\mathrm{opt}}{=} u_h^{\mathrm{opt}}$. Here, the equality with the opt-notation indicates to the reader it holds only in this special setting.

Setting $\mu = 0$ and rewriting $F = \mathcal{B}u$, (48) demonstrates that $|||\mathcal{B}\cdot|||_{\mathcal{V}'} = \|\cdot\|_{\mathcal{U}}$. Moreover,

$$\langle \mathcal{R}_{\mathcal{U}}\,\mu, \widetilde{\mu}\rangle_{\mathcal{U}} = (\mu,\widetilde{\mu})_{\mathcal{U}} \overset{\mathrm{opt}}{=} (\mathcal{B}\mu, \mathcal{B}\widetilde{\mu})_{\mathcal{V}'} = \langle \mathcal{B}'\,\mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B}\mu, \widetilde{\mu}\rangle_{\mathcal{U}} \qquad \forall\, \mu, \widetilde{\mu} \in \mathcal{U}\,. \tag{49}$$

Now, clearly, $\mathcal{R}_{\mathcal{U}} \overset{\mathrm{opt}}{=} \mathcal{B}'\,\mathcal{R}_{\mathcal{V}}^{-1}\,\mathcal{B} = \mathcal{A}$. In the case that $\mathcal{B}$ is an isomorphism, the Riesz operator for the optimal test norm is clearly $\mathcal{R}_{\mathcal{V}} \overset{\mathrm{opt}}{=} \mathcal{B}\,\mathcal{R}_{\mathcal{U}}^{-1}\,\mathcal{B}'$. Here, the definitions of these special Riesz operators are easy to remember by observing that they are the unique isometries where the following diagram commutes:

$$\begin{array}{ccc} \mathcal{U}' & \xleftarrow{\ \mathcal{B}'\ } & \mathcal{V} \\ {\scriptstyle \mathcal{R}_{\mathcal{U}}}\uparrow & & \downarrow{\scriptstyle \mathcal{R}_{\mathcal{V}}} \\ \mathcal{U} & \xrightarrow{\ \mathcal{B}\ } & \mathcal{V}' \end{array}\,.$$

Note that if $\|\cdot\|_{\mathcal{V}} \overset{\text{opt}}{=} |||\cdot|||_{\mathcal{V}}$, then $\|\mathcal{B}\| \overset{\text{opt}}{=} \gamma \overset{\text{opt}}{=} 1$ in (33), $\kappa(\mathcal{B}) \overset{\text{opt}}{=} 1$ in (41), and $b(\mu,\nu)$ becomes a duality pairing [21].

Alternatively, if $(\operatorname{Null}\mathcal{B}')^{\perp} \neq \{0\}$, then $|||\cdot|||_{\mathcal{V}}$ may be treated as a quotient norm or endowed with a Riesz map defined $\mathcal{R}_{\mathcal{V}} \overset{\text{opt}}{=} \mathcal{B}\,\mathcal{R}_{\mathcal{U}}^{-1}\,\mathcal{B}' + \mathcal{P}$, where $\mathcal{P} : \mathcal{V} \to \operatorname{Null}\mathcal{B}'$ is an orthogonal projection. The general setting we wish to consider is summarized by the following proposition.

Proposition 3.2. *Let Assumption 1 hold. Let $\mathcal{P} : \mathcal{V} \to \operatorname{Null}\mathcal{B}'$ be the orthogonal projection. Define*

$$|||\mu|||_{\mathcal{U}} = \|\mathcal{B}\mu\|_{\mathcal{V}'} \qquad \text{and} \qquad |||\nu|||_{\mathcal{V}}^2 = \|\mathcal{B}'\nu\|_{\mathcal{U}'}^2 + \|\mathcal{P}\nu\|_{\mathcal{V}}^2\,,$$

for all $\mu \in \mathcal{U}$ and $\nu \in \mathcal{V}$. Then

$$\|\mu\|_{\mathcal{U}} = |||\mathcal{B}\mu|||_{\mathcal{V}'} \qquad \text{and} \qquad \|\nu\|_{\mathcal{V}}^2 = |||\mathcal{B}'\nu|||_{\mathcal{U}'}^2 + \|\mathcal{P}\nu\|_{\mathcal{V}}^2\,, \tag{50}$$

for all $\mu \in \mathcal{U}$ and $\nu \in \mathcal{V}$.

Proof. The first relationship in (50) immediately follows from (49). To prove the second, let $\nu = \nu_0 + \nu_\perp$, where $\nu_0 \in \operatorname{Null}\mathcal{B}'$ and $\nu_\perp \in (\operatorname{Null}\mathcal{B}')_\perp$. Clearly, $\nu_0 = \mathcal{P}\nu$. By the Closed Range Theorem, we find $(\operatorname{Null}\mathcal{B}')^{\perp} = \operatorname{Range}\mathcal{B}$ and both $(\operatorname{Null}\mathcal{B}')$ and $(\operatorname{Null}\mathcal{B}')^{\perp}$ are closed. Therefore, $\|\nu\|_{\mathcal{V}}^2 = \|\mathcal{P}\nu\|_{\mathcal{V}}^2 + \|\nu_\perp\|_{\mathcal{V}}^2$, where

$$\|\nu_\perp\|_{\mathcal{V}} = \sup_{\widetilde{\nu}_\perp \in (\operatorname{Null}\mathcal{B}')_\perp} \frac{(\nu,\widetilde{\nu}_\perp)_{\mathcal{V}}}{\|\widetilde{\nu}_\perp\|_{\mathcal{V}}} = \sup_{E^\perp \in (\operatorname{Null}\mathcal{B}')^\perp} \frac{\langle E^\perp,\nu\rangle_{\mathcal{V}}}{\|E^\perp\|_{\mathcal{V}'}} = \sup_{\mu\in\mathcal{U}} \frac{\langle \mathcal{B}\mu,\nu\rangle_{\mathcal{U}}}{\|\mathcal{B}\mu\|_{\mathcal{V}'}} = |||\mathcal{B}'\nu|||_{\mathcal{U}'}\,.$$

□

Corollary 3.3. *Let Assumption 1 hold. Let $F \in \operatorname{Range}\mathcal{B}$ and $G \in \mathcal{U}'$. Then, for $u = \mathcal{B}^{-1}F$ and any $v \in (\mathcal{B}')^{-1}(\{G\})$,*

$$\|u-\mu\|_{\mathcal{U}} = |||\mathcal{B}\mu - F|||_{\mathcal{V}'} \qquad \text{and} \qquad \|v-\nu\|_{\mathcal{V}}^2 = \|\mathcal{P}(v-\nu)\|_{\mathcal{V}}^2 + |||\mathcal{B}'\nu - G|||_{\mathcal{U}'}^2\,,$$

for all $\mu \in \mathcal{U}$ and $\nu \in \mathcal{V}$.

Proof. This immediately follows from (50). □

3.8 Nonlinear problems

The majority of the theory presented in this chapter relies upon that fact the operator $\mathcal{B}$ is linear. Let us now consider a continuous nonlinear operator from $\mathcal{U}$ to $\mathcal{V}'$, $\mathcal{B}^{\text{nl.}} : \mathcal{U} \to \mathcal{V}'$, where $\langle \mathcal{B}^{\text{nl.}}\mu, \nu\rangle_{\mathcal{V}} = b^{\text{nl.}}(\mu, \nu)$. If all the operators $\mathcal{B}$ above are substituted by $\mathcal{B}^{\text{lin.}}[u_0]$, a well-defined linearization of $\mathcal{B}^{\text{nl.}}$ at the point u_0, then the corresponding saddle-point problems (43) and (46) would deliver an optimal discrete solution u_h^{opt} for the given linearization.

Let $F \in \mathcal{V}'$. In the case that $\mathcal{B}^{\text{lin.}}[u_0] = D_u\mathcal{B}^{\text{nl.}}[u_0]$, the Fréchet derivative of $\mathcal{B}^{\text{nl.}}$ at u_0, then the optimization problem (39) can be used to generate an idealized Gauss-Newton algorithm for the equation $\mathcal{B}^{\text{nl.}}u = F$:

$$u_h^{\text{opt}} \mapsto u_h^{\text{opt}} + \underset{\mu\in\mathcal{U}_h}{\arg\min} \left\|\mathcal{B}^{\text{lin.}}[u_0]\mu - F^{\text{nl.}}[u_0]\right\|_{\mathcal{V}'} . \tag{51}$$

Here, the load $F^{\text{nl.}}[u_0]$ at each iteration is simply the residual $F^{\text{nl.}}[u_0](\nu) = F(\nu) - b^{\text{nl.}}(u_0, \nu)$. Note that the optimal test norm in (51) must be updated at each iteration *along with* the solution increment, u_0.

In the next chapter, we finally introduce the DPG and DPG* methods. Our DPG strategy for nonlinear problems corresponds to the Gauss-Newton algorithm (51) and follows a well established trajectory explored in many papers [30, 32, 90, 130]. In place of (51), a similar Picard-type algorithm has also been experimented with in [106]. Developing a unified nonlinear theory for DPG methods, as well as similar nonlinear residual minimization strategies for non-Hilbert norms, is a topic of contemporary research. Recent advances in nonlinear DPG methods can be found in [25, 26] and for residual minimization in Banach spaces in [104, 105].

Chapter 4

DPG and DPG* methods

This chapter introduces DPG and DPG* methods. Both these methods are fully discrete versions of (47), where the test space is *broken*. Explicit examples of DPG and DPG* methods are collected in the following chapter. This chapter also defines so-called *ultraweak variational formulations*. Ultraweak formulations are used in constructing each of the representative methods analyzed in the later chapters. This chapter closes with a comparison of DPG and DPG* and other well-known methods, along with some other important remarks.

4.1 Practical methods

Suppose a bilinear form $b : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ arises from a weak formulation of a PDE on a domain $\Omega \subseteq \mathbb{R}^d$ which is partitioned into a mesh Ω_h. In this scenario, if there are Hilbert spaces $\mathcal{V}(K)$ on each mesh element K such that

$$\mathcal{V} = \prod_{K \in \Omega_h} \mathcal{V}(K), \tag{52}$$

then the system (31), in the case $G = 0$, is called a *DPG formulation*. In the case $F = 0$, it is called a *DPG* formulation*. Spaces of the form (52) are called *broken spaces* [28].

DPG and DPG* methods arise as special fully discrete versions of (47) where the test space is *broken*. The standard construction is as follows. Fix the same space $\mathcal{U}_h \subseteq \mathcal{U}$ introduced above and select another finite-dimensional space $\mathcal{V}_h \subseteq \mathcal{V}$, generally with higher dimension. Now, consider the discrete problem of finding $v_h \in \mathcal{V}_h$ and $w_h \in \mathcal{U}_h$ satisfying

$$\begin{cases} (v_h, \nu)_{\mathcal{V}} + b(w_h, \nu) = F(\nu) & \forall\, \nu \in \mathcal{V}_h, \\ b(\mu, v_h) \qquad\qquad\quad = G(\mu) & \forall\, \mu \in \mathcal{U}_h. \end{cases} \tag{53}$$

When $\mathcal{V}$ is a broken space of the form (52), $\mathcal{V}_h$ can be chosen to consist of functions with no continuity constraints across mesh element interfaces. In this case, $G = 0$ delivers *DPG methods* and $F = 0$ delivers *DPG* methods*. To parallel (12) and (47), the two methods described above are, in turn:

(54a) Find $u_h \in \mathcal{U}_h$ and $\varepsilon_h \in \mathcal{V}_h$ satisfying
$$\begin{cases} (\varepsilon_h, \nu)_{\mathcal{V}} + b(u_h, \nu) = F(\nu) & \forall\, \nu \in \mathcal{V}_h\,, \\ b(\mu, \varepsilon_h) = 0 & \forall\, \mu \in \mathcal{U}_h\,, \end{cases}$$

(54b) Find $\lambda_h \in \mathcal{U}_h$ and $v_h \in \mathcal{V}_h$ satisfying
$$\begin{cases} (v_h, \nu)_{\mathcal{V}} - b(\lambda_h, \nu) = 0 & \forall\, \nu \in \mathcal{V}_h\,, \\ b(\mu, v_h) = G(\mu) & \forall\, \mu \in \mathcal{U}_h\,. \end{cases}$$

In both cases, we typically must find $\mathcal{V}_h$ with $\dim(\mathcal{V}_h) > \dim(\mathcal{U}_h)$ large enough to assert discrete stability.

A key feature of (53) is that the the top left form, $(\cdot,\cdot)_{\mathcal{V}}$, being an inner product, is always coercive. Hence, the discrete stability of (53) is guaranteed solely by a single discrete inf-sup condition. This condition is often *easy to obtain* in practice since we can always increase $\dim(\mathcal{V}_h)$ without violating the coercivity of the top left term. This inf-sup condition has been analytically established for various DPG methods through the construction of local [28, 82, 107] or global [29] Fortin operators [15] on generously large test spaces (cf. Section 6.1). The same inf-sup condition also confirms the stability of the corresponding DPG* methods. An alternative characterization of the methods above can be found in a Petrov–Galerkin form in [93, Section 4.1].

4.2 Ultraweak variational formulations

Many PDEs originate in the following strong form:

$$\mathcal{L}u = f\,,$$

where $\mathcal{L}$ is a linear differential operator and f is a prescribed function. It is possible to give many general DPG and DPG* formulations for such operator equations using the framework of [49, Appendix A] (which generalizes the Friedrichs systems framework developed in [22, 59, 142]).

Let $d, k, m \geq 1$ be integers and let $\Omega \subseteq \mathbb{R}^d$ be a bounded open set. We use multiindices $\alpha = (\alpha_1, \dots \alpha_d)$ of length $|\alpha| = \alpha_1 + \cdots + \alpha_d \leq k$. Suppose we are given functions $a_{ij\alpha} : \Omega \to \mathbb{R}$ for each $i = 1, \dots, l$, $j = 1, \dots, m$, and each $|\alpha| \leq k$. Let $\mathcal{L}$ be the differential operator acting on functions $u : \Omega \to \mathbb{R}^m$ such that

$$[\mathcal{L}u]_i = \sum_{j=1}^{m} \sum_{|\alpha| \leq k} \partial^\alpha (a_{ij\alpha} u_j), \qquad i \in \{1, \dots, l\}. \tag{55}$$

Wherever appropriate, let L^2 denote *either* the l- or m-fold Cartesian product of $L^2(\Omega)$. Likewise, let $\mathcal{D}$ denote *either* the l- or m-fold Cartesian product of $\mathcal{D}(\Omega)$, where $\mathcal{D}(\Omega)$ is the space of infinitely differentiable functions that are compactly supported on Ω (and accordingly, $\mathcal{D}'$ denotes distributional vector fields). Let $\mathcal{L}^*$ be the formal adjoint differential operator of $\mathcal{L}$; i.e., $\mathcal{L}^*$ satisfies $(\mathcal{L}\phi, \psi)_{L^2} = (\phi, \mathcal{L}^*\psi)_{L^2}$ for all $\phi, \psi \in \mathcal{D}$. From now on, we will simply denote all such L^2-inner products on Ω as $(\cdot, \cdot)_\Omega = (\cdot, \cdot)_{L^2}$. Likewise, all L^2-inner products restricted to a measurable subset $K \subseteq \Omega$ will be denoted $(\cdot, \cdot)_K$.

The action of $\mathcal{L}^*$ on $v : \Omega \to \mathbb{R}^l$ is given by

$$[\mathcal{L}^* v]_j = \sum_{i=1}^{l} \sum_{|\alpha| \leq k} (-1)^{|\alpha|} a_{ij\alpha} \, \partial^\alpha v_i, \qquad j \in \{1, \dots, m\}. \tag{56}$$

We assume that the coefficients $a_{ij\alpha}$ are such that both $\mathcal{L}u$ and $\mathcal{L}^*v$ are well-defined distributions for all $u, v \in L^2$, i.e.,

$$\mathcal{L}u \text{ and } \mathcal{L}^*v \text{ are in } \mathcal{D}' \text{ for all } u, v \in L^2. \tag{57a}$$

(This holds, e.g., if $a_{ij\alpha}$ are constants.)

We may now define Sobolev-like graph spaces by virtue of (57a). On any nonempty open subset $K \subseteq \Omega$, define the Hilbert spaces $H(\mathcal{L}, K) = \{u \in L^2(K)^m : \mathcal{L}u \in L^2(K)^l\}$ and, likewise, $H(\mathcal{L}^*, K) = \{v \in L^2(K)^l : \mathcal{L}^*v \in L^2(K)^m\}$. Pause now and reflect that if we let $\mathcal{L} = \operatorname{grad}$, the canonical gradient operator, then $\mathcal{L}^* = -\operatorname{div}$ and $H(\mathcal{L}, K) = H(\operatorname{grad}, K) = H^1(K)$ and $H(\mathcal{L}^*, K) = H(\operatorname{div}, K)$. To simplify notation, we use the abbreviations $H(\mathcal{L}) = H(\mathcal{L}, \Omega)$ and $H(\mathcal{L}^*) = H(\mathcal{L}^*, \Omega)$. Also define linear operators $\mathscr{D} : H(\mathcal{L}) \to H(\mathcal{L}^*)'$ and $\mathscr{D}^* : H(\mathcal{L}^*) \to H(\mathcal{L})'$

such that

$$\langle \mathscr{D} u, v\rangle_{H(\mathcal{L}^*)} = (\mathcal{L}u, v)_\Omega - (u, \mathcal{L}^* v)_\Omega \,, \qquad \langle \mathscr{D}^* v, u\rangle_{H(\mathcal{L})} = (\mathcal{L}^* v, u)_\Omega - (v, \mathcal{L}u)_\Omega \,,$$

for all $u \in H(\mathcal{L})$ and $v \in H(\mathcal{L}^*)$. Note that $\mathscr{D}^* = -\mathscr{D}'$, by these definitions. These graph spaces are equipped with natural *graph norms*:

$$\|u\|^2_{H(\mathcal{L})} = \|\mathcal{L}u\|^2_{L^2} + \|u\|^2_{L^2} \,, \qquad \|v\|^2_{H(\mathcal{L}^*)} = \|\mathcal{L}^* v\|^2_{L^2} + \|v\|^2_{L^2} \,. \tag{57b}$$

With these norms, notice that both $\mathscr{D}$ and $\mathscr{D}^*$ are bounded. Indeed, $|\langle \mathscr{D} u, v\rangle_{H(\mathcal{L}^*)}| \leq \|\mathcal{L}u\|_{L^2}\|v\|_{L^2} + \|u\|_{L^2}\|\mathcal{L}^* v\|_{L^2} \leq \|u\|_{H(\mathcal{L})}\|v\|_{H(\mathcal{L}^*)}$.

Finally, we may incorporate homogeneous boundary conditions. Recall the definition of the left annihilator in (6). Define $H_0(\mathcal{L}) \subseteq H(\mathcal{L})$ and $H_0(\mathcal{L}^*) \subseteq H(\mathcal{L}^*)$ to be two subspaces satisfying

$$H_0(\mathcal{L}) = {}^\perp\mathscr{D}^*(H_0(\mathcal{L}^*)) \,, \qquad H_0(\mathcal{L}^*) = {}^\perp\mathscr{D}(H_0(\mathcal{L})) \,. \tag{57c}$$

Observe that (57c) does not uniquely characterize either $H_0(\mathcal{L})$ or $H_0(\mathcal{L}^*)$. These definitions permit many different so-called "mixed" homogeneous boundary conditions.

We will consider two boundary value problems: given $f, g \in L^2$,

$$\text{find } u \in H_0(\mathcal{L}) \text{ satisfying } \mathcal{L}u = f \tag{58a}$$

and

$$\text{find } v \in H_0(\mathcal{L}^*) \text{ satisfying } \mathcal{L}^* v = g. \tag{58b}$$

To derive a broken "ultraweak formulation" for (58a) and (58b), we focus on the scenario where Ω is partitioned into a mesh Ω_h of finitely many open disjoint elements K such that $\overline{\Omega}$ is the union of closures of all mesh elements K in Ω_h. For functions u and v, we denote by $\mathcal{L}_h u$ and $\mathcal{L}_h^* v$ the functions obtained by applying $\mathcal{L}$ and $\mathcal{L}^*$ to $u|_K$ and $v|_K$, respectively, element by element, for all $K \in \Omega_h$.

With this in mind, define the broken spaces

$$H(\mathcal{L}_h) = \prod_{K\in\Omega_h} H(\mathcal{L},K)\,, \qquad\qquad H(\mathcal{L}_h^*) = \prod_{K\in\Omega_h} H(\mathcal{L}^*,K)\,,$$

which naturally conform to (52). Clearly, $H(\mathcal{L}_h)$ and $H(\mathcal{L}_h^*)$ are inner product product spaces with corresponding graph norms,

$$\|u\|^2_{H(\mathcal{L}_h)} = \sum_{K\in\mathcal{T}} \|\mathcal{L}u\|^2_{L^2(K)} + \|u\|^2_{L^2}\,, \qquad \|v\|^2_{H(\mathcal{L}_h^*)} = \sum_{K\in\mathcal{T}} \|\mathcal{L}^*v\|^2_{L^2(K)} + \|v\|^2_{L^2}\,. \tag{59}$$

The natural inner products on these spaces, induced by these graph norms, are defined by

$$(u,\widetilde{u})_{H(\mathcal{L}_h)} = (\mathcal{L}_h u, \mathcal{L}_h\widetilde{u})_\Omega + (u,\widetilde{u})_\Omega, \qquad (v,\widetilde{v})_{H(\mathcal{L}_h^*)} = (\mathcal{L}_h^* v, \mathcal{L}_h^*\widetilde{v})_\Omega + (v,\widetilde{v})_\Omega, \tag{60}$$

for all $u,\widetilde{u}\in H(\mathcal{L}_h), v,\widetilde{v}\in H(\mathcal{L}_h^*)$. Now define the corresponding bounded linear operators $\mathscr{D}_h : H(\mathcal{L}_h)\to H(\mathcal{L}_h^*)'$ and $\mathscr{D}_h^* : H(\mathcal{L}_h^*)\to H(\mathcal{L}_h)'$ by

$$\langle \mathscr{D}_h u, v\rangle_{H(\mathcal{L}_h^*)} = (\mathcal{L}_h u, v)_\Omega - (u, \mathcal{L}_h^* v)_\Omega, \qquad \langle \mathscr{D}_h^* v, u\rangle_{H(\mathcal{L}_h)} = (\mathcal{L}_h^* v, u)_\Omega - (v, \mathcal{L}_h u)_\Omega,$$

for all $u\in H(\mathcal{L}_h)$, and all $v\in H(\mathcal{L}_h^*)$. From now on, when using the operators $\mathscr{D}_h$ and $\mathscr{D}_h^*$, we simply denote $\langle \mathscr{D}_h\cdot,\cdot\rangle_h = \langle \mathscr{D}_h\cdot,\cdot\rangle_{H(\mathcal{L}_h^*)}$ or, likewise, $\langle \mathscr{D}_h^*\cdot,\cdot\rangle_h = \langle \mathscr{D}_h^*\cdot,\cdot\rangle_{H(\mathcal{L}_h)}$, since the meaning can easily be deduced from context. Finally, let

$$\begin{aligned} Q(\mathcal{L}_h) &= \{p\in H(\mathcal{L}_h)' :\ \text{there is a } v\in H_0(\mathcal{L}^*) \text{ such that } p = \mathscr{D}_h^* v\}\,,\\ Q(\mathcal{L}_h^*) &= \{q\in H(\mathcal{L}_h^*)' :\ \text{there is a } u\in H_0(\mathcal{L}) \text{ such that } q = \mathscr{D}_h u\}\,. \end{aligned}$$

These are Hilbert spaces when normed by the so-called *minimum energy extension* (quotient) norm, i.e., $\|q\|_{Q(\mathcal{L}_h^*)} = \inf\{\|u\|_{H(\mathcal{L})} : u\in H(\mathcal{L}) \text{ satisfying } \mathscr{D}_h u = q\}$.

Multiplying (58a) by a function $\nu\in H(\mathcal{L}_h^*)$ and applying the definition of $\mathscr{D}_h$, we get $(u,\mathcal{L}_h^*\nu)_\Omega + \langle \mathscr{D}_h u,\nu\rangle_h = (f,\nu)_\Omega$ for all ν in $H(\mathcal{L}_h^*)$. Setting $\mathscr{D}_h u$ to q, a new unknown in $Q(\mathcal{L}_h^*)$, we obtain the following *ultraweak formulation* with $F(\nu) = (f,\nu)_\Omega$. Given any $F\in H(\mathcal{L}_h^*)'$, find $u\in L^2$ and $q\in Q(\mathcal{L}_h^*)$ such that

$$(u,\mathcal{L}_h^*\nu)_\Omega + \langle q,\nu\rangle_h = F(\nu) \qquad \forall\,\nu\in H(\mathcal{L}_h^*). \tag{61a}$$

Similarly proceeding with (58b) and setting $F(\nu) = (g, \nu)_\Omega$, we obtain an ultraweak formulation of the dual problem: Given any $F \in H(\mathcal{L}_h)'$, find $u \in L^2$ and $p \in Q(\mathcal{L}_h)$ such that

$$(v, \mathcal{L}_h \nu)_\Omega + \langle p, \nu \rangle_h = F(\nu) \qquad \forall\, \nu \in H(\mathcal{L}_h). \tag{61b}$$

The next result shows that (61a) is uniquely solvable whenever (58a) is, as well as a similar connection between (61b) and (58b).

Theorem 4.1 (Well-posedness of the broken formulations)**.** *Suppose* (57) *holds. Then*

1. *Whenever $\mathcal{L} : H_0(\mathcal{L}) \to L^2$ is a bijection, problem* (61a) *is well-posed. Moreover, if $F(\nu) = (f, \nu)_\Omega$ for some $f \in L^2$, then the unique solution u of* (61a) *is in $H_0(\mathcal{L})$, solves* (58a)*, and satisfies $q = \mathscr{D}_h u$.*
2. *Whenever $\mathcal{L}^* : H_0(\mathcal{L}^*) \to L^2$ is a bijection, problem* (61b) *is well-posed. Moreover, if $F(\nu) = (g, \nu)_\Omega$ for some $g \in L^2$, then the unique solution v of* (61b) *is in $H_0(\mathcal{L}^*)$, solves* (58b)*, and satisfies $p = \mathscr{D}_h^* u$.*

Proof. The first statement is exactly the statement of [49, Theorem A.5]. The second statement also follows from [49, Theorem A.5] when $\mathcal{L}$ is replaced by $\mathcal{L}^*$. □

Naturally, formulations (61a) and (61b) also have adjoints. For instance, the adjoint of the ultraweak formulation (61a) is the following: Given any $G \in (L^2 \times Q(\mathcal{L}_h^*))'$, find $v \in H(\mathcal{L}_h^*)$ such that

$$(\mu, \mathcal{L}_h^* v)_\Omega + \langle \rho, v \rangle_h = G(\mu, \rho) \qquad \forall\, \mu \in L^2,\, \rho \in Q(\mathcal{L}_h^*). \tag{62a}$$

Similarly, the adjoint of (61b) is: Given any $G \in (L^2 \times Q(\mathcal{L}_h))'$, find $u \in H(\mathcal{L}_h)$ such that

$$(\mu, \mathcal{L}_h u)_\Omega + \langle \rho, u \rangle_h = G(\mu, \rho) \qquad \forall\, \mu \in L^2,\, \rho \in Q(\mathcal{L}_h). \tag{62b}$$

Under similar conditions to Theorem 4.1, these variational formulations are also well-posed, as the following theorem demonstrates.

Theorem 4.2 (Well-posedness of the adjoint problems)**.** *Suppose* (57) *holds. Then*

1. *Whenever* $\mathcal{L} : H_0(\mathcal{L}) \to L^2$ *is a bijection, problem* (62a) *is well-posed. Moreover, if* $G(\mu, \rho) = (g, \mu)_\Omega$ *for some* $g \in L^2$*, then the unique solution* v *of* (62a) *is in* $H_0(\mathcal{L}^*)$ *and solves* (58b).
2. *Whenever* $\mathcal{L}^* : H_0(\mathcal{L}^*) \to L^2$ *is a bijection, problem* (62b) *is well-posed. Moreover, if* $G(\mu, \rho) = (f, \mu)_\Omega$ *for some* $f \in L^2$*, then the unique solution* u *of* (62b) *is in* $H_0(\mathcal{L})$ *and solves* (58a).

Proof. Both claims are closely related and follow similarly from Theorem 4.1. Therefore, we prove only the first statement.

Let the operator $\mathcal{B} : L^2 \times Q(\mathcal{L}_h^*) \to H(\mathcal{L}_h^*)'$ be defined $\langle \mathcal{B}(\mu, \rho), \nu \rangle_{H(\mathcal{L}_h^*)} = (\mu, \mathcal{L}_h^* \nu)_\Omega + \langle \rho, \nu \rangle_h$, for all $\nu \in H(\mathcal{L}_h)$ and $(\mu, \rho) \in L^2 \times Q(\mathcal{L}_h^*)$. Recall that $F \in H(\mathcal{L}_h^*)' = \operatorname{Range} \mathcal{B}$ in (61a) was arbitrary. Therefore, as a consequence of the first statement in Theorem 4.1, we conclude that $\mathcal{B}$ is both bounded below (cf. (14)) *and* surjective. That is, $\mathcal{B}$ is a bijection and, by the Closed Range Theorem, $(\operatorname{Null} \mathcal{B}')_\perp = \{0\}$. Hence, we conclude that (62a) is well-posed.

Next, suppose $G((\mu, \rho)) = (g, \mu)_\Omega$. Then (62a) yields

$$(\mu, \mathcal{L}_h^* v)_\Omega = (g, \mu)_\Omega, \tag{63}$$

$$\langle \rho, v \rangle_h = 0, \tag{64}$$

for all $\mu \in L^2$ and $\rho \in Q(\mathcal{L}_h^*)$. Equation (63) yields $\mathcal{L}_h^* v = g$ since $H(\mathcal{L}_h^*)$ is continuously embedded in L^2. It remains to show that v is in $H_0(\mathcal{L}^*)$. Note that for all $\phi \in \mathcal{D}$, by the distributional definition of $\mathcal{L}$ and the definition of $\mathscr{D}_h$,

$$\langle \mathcal{L}^* v, \phi \rangle_{\mathcal{D}} = (\mathcal{L}\phi, v)_\Omega = (\mathcal{L}_h^* v, \phi)_\Omega + \langle \mathscr{D}_h \phi, v \rangle_h.$$

Since $\mathcal{D}$ is contained in $H_0(\mathcal{L})$, $\mathscr{D}_h \phi$ is in $Q(\mathcal{L}_h^*)$ and the last term vanishes by virtue of (64). Morover, since $\mathcal{D}$ is densely contained in L^2, this shows that $\mathcal{L}^* v = \mathcal{L}_h^* v = g$. Thus $v \in H(\mathcal{L}^*)$. Using (64) again, observe (cf. [49, Lemma A.3]) that

$$0 = \langle \rho, v \rangle_h = \langle \mathscr{D}_h \mu, v \rangle_h = \langle \mathscr{D} \mu, v \rangle_{H(\mathcal{L}^*)},$$

for all $\rho = \mathscr{D}_h\mu \in Q(\mathcal{L}_h^*)$, where $\mu \in H_0(\mathcal{L})$. Therefore, $v \in {}^{\perp}\mathscr{D}(H_0(\mathcal{L}))$. Finally, v is in $H_0(\mathcal{L}^*)$ simply by (57c). □

Evidently, Theorems 4.1 and 4.2 give a class of examples where DPG and DPG* methods can be formulated.[1] Letting $\mathcal{U} = L^2 \times Q(\mathcal{L}_h)$ and $\mathcal{V} = H_0(\mathcal{L}_h)$, define the bilinear form $b : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ as follows

$$b((\mu,\rho),\nu) = (\mu, \mathcal{L}_h\nu)_\Omega + \langle \rho, \nu\rangle_h \qquad \forall\, (\mu,\rho) \in \mathcal{U},\ \nu \in \mathcal{V}. \tag{65}$$

We may now consider the DPG* formulations of (58a). Treatment of the dual problem (58b) is obviously similar.

Theorem 4.3 (Ultraweak DPG* formulation of (58a))**.** *Let $(\cdot,\cdot)_\mathcal{V}$ be any inner product on $H(\mathcal{L}_h)$ equivalent to $(\cdot,\cdot)_{H(\mathcal{L}_h)}$. Suppose* (57) *holds, $\mathcal{L}^* : H_0(\mathcal{L}^*) \to L^2$ is a bijection, and b is as in* (65)*. Then, given a $G \in (L^2 \times Q(\mathcal{L}_h))'$, the problem of finding a function $u \in H(\mathcal{L}_h)$ satisfying*

$$\begin{cases} (u,\nu)_\mathcal{V} \ - \ b((\lambda,\sigma),\nu) & = 0 \qquad\qquad \forall\, \nu \in H(\mathcal{L}_h), \\ b((\mu,\rho),u) & = G((\mu,\rho)) \qquad \forall\, (\mu,\rho) \in L^2 \times Q(\mathcal{L}_h), \end{cases} \tag{66}$$

is well-posed. Moreover, if $G((\mu,\rho)) = (f,\mu)_\Omega$ for some $f \in L^2$, then the unique solution u is in $H_0(\mathcal{L})$ and satisfies $\mathcal{L}u = f$, i.e., u solves (58a)*.*

Proof. As in the proof of Theorem 4.2, the operator $\mathcal{B} : L^2 \times Q(\mathcal{L}_h) \to H(\mathcal{L}_h)'$, defined by $\langle \mathcal{B}(\mu,\rho), \nu\rangle_{H(\mathcal{L}_h)} = b((\mu,\rho),\nu)$ for all $(\mu,\rho) \in L^2 \times Q(\mathcal{L}_h)$ and $\nu \in H(\mathcal{L}_h)$, is a bijection. Hence, (66) is a problem of form (13) with $\mathcal{B}$ satisfying (14) and we conclude that (66) has a unique solution (see, e.g., Proposition 2.4). The remaining statements immediately follow from Theorem 4.2. □

[1] Unlike for other types of broken variational formulations (cf. [28, 65, 89]), these results for broken ultraweak formulations circumvent the requirement of having a complete theory for traces in order to prove well-posedness.

4.3 Solving the primal and dual problems *simultaneously*

One interesting feature of ultraweak variational formulations is the capacity to solve both a primal and a dual problem *simultaneously* to a tunable accuracy. Begin by reflecting on the loads $F \in \mathcal{V}'$ and $G \in \mathcal{U}'$ appearing in (13):

$$\begin{cases} \mathcal{R}_{\mathcal{V}} v + \mathcal{B} w = F\,, \\ \mathcal{B}' v \qquad\quad = G\,. \end{cases} \tag{67}$$

Let $\mathcal{B}$ to be an isomorphism and define $F = \mathcal{R}_{\mathcal{V}} (\mathcal{B}')^{-1} G + \ell$, for some fixed $\ell \in (\operatorname{Null} \mathcal{B}')^{\perp} = \mathcal{V}'$.[2] Noting that $v = (\mathcal{B}')^{-1} G$, by the second equation in (67), it is readily seen that $\mathcal{B} w = \ell$. Therefore, with this choice of loads, $w = u$ solves the primal problem (11), $\mathcal{B} u = \ell$, and v solves the dual problem (17), $\mathcal{B}' v = G$, simultaneously.

Introducing the load F as proposed above involves the inversion the linear operator $\mathcal{B}'$. In practice, this is usually not feasible and, therefore, precludes the construction of any such load in most circumstances. Nevertheless, consider the following system of equations:

$$\begin{cases} (\mathcal{L}^* v, \mathcal{L}^* \nu)_{\Omega} + (w, \mathcal{L}^* \nu)_{\Omega} = (g, \mathcal{L}^* \nu)_{\Omega} + (f, \nu)_{\Omega} & \forall\, \nu \in H(\mathcal{L}^*)\,, \\ (\mathcal{L}^* v, \mu)_{\Omega} \qquad\qquad\qquad\quad = (g, \mu)_{\Omega} & \forall\, \mu \in L^2\,. \end{cases} \tag{68}$$

This corresponds to a system like (67) with $\mathcal{V} = H(\mathcal{L}^*)$ and $\mathcal{U} = L^2$. In (68), v clearly satisfies $(\mathcal{L}^* v, \mu)_{\Omega} = (g, \mu)_{\Omega}$. That is, v solves the dual problem (58b), $\mathcal{L}^* v = g$, in a strong sense. Substituting $\mu = \mathcal{L}^* \nu$ into (68) and canceling terms in the first equation, we immediately find that $(w, \mathcal{L}^* \nu)_{\Omega} = (f, \nu)_{\Omega}$. That is, $w = u$ solves the primal problem (58a), $\mathcal{L} u = f$, in the ultraweak sense. Since (68) only involves the inversion of $\mathcal{B}'$ implicitly, it can be used to render a practical finite element method.

To avoid solving this new mixed problem for both v and w simultaneously, upon discretization, broken test spaces spaces can be used. In this setting, we must consider the following

[2] In the case of an injective but *not* surjective $\mathcal{B}$, consider $F = \mathcal{B}(\mathcal{B}'\, \mathcal{R}_{\mathcal{V}}^{-1}\, \mathcal{B})^{-1} G + \ell \in (\operatorname{Null} \mathcal{B}')^{\perp} \subsetneq \mathcal{V}'$.

related system with solution $(w,\sigma)\in\mathcal{U}=L^2\times Q(\mathcal{L}_h^*)$ and $v\in\mathcal{V}=H(\mathcal{L}_h^*)$:

$$\text{(69)}\qquad \begin{cases} (\mathcal{L}_h^* v,\mathcal{L}_h^*\nu)_\Omega+\alpha(v,\nu)_\Omega+(w,\mathcal{L}_h^*\nu)_\Omega+\langle\sigma,\nu\rangle_h=(g,\mathcal{L}_h^*\nu)_\Omega+(f,\nu)_\Omega & \forall\,\nu\in H(\mathcal{L}_h^*),\\ (\mathcal{L}_h^* v,\mu)_\Omega=(g,\mu)_\Omega & \forall\,\mu\in L^2,\\ \langle v,\rho\rangle_h=0 & \forall\,\rho\in Q(\mathcal{L}_h^*). \end{cases}$$

Here, the parameter $\alpha>0$ has been introduced to ensure that the inner product $(v,\nu)_{\mathcal{V}}=(\mathcal{L}_h^* v,\mathcal{L}_h^*\nu)_\Omega+\alpha(v,\nu)_\Omega$ is coercive over the new broken space.

The consequent manipulations are inspired by [69, Lemma 7]. First, notice that the last two equations in (69) uniquely determine v. Therefore, after testing the middle equation with $\mu=\mathcal{L}_h^*\nu$, observe that the first equation can be rewritten

$$(w,\mathcal{L}_h^*\nu)_\Omega+\langle\sigma,\nu\rangle_h=(f,\nu)_\Omega-\alpha(v,\nu)\,.$$

By linearity, $(w,\sigma)_\Omega=(u,q)_\Omega+\alpha(e,r)$, where $(u,q)\in\mathcal{U}$ solves the ultraweak primal problem $(u,\mathcal{L}_h^*\nu)_\Omega+\langle q,\nu\rangle_h=(f,\nu)_\Omega$ (cf. (58a)) and $(e,r)=(e(v),r(v))$ is a *pollution term* defined by the equation $(e,\mathcal{L}_h^*\nu)_\Omega+\langle r,\nu\rangle_h=-(v,\nu)$. Clearly, $w\to u$ as $\alpha\to 0$.

4.4 Related methods

Let $\mathcal{V}=H(\mathcal{L}_h^*)$. For any $F\in H(\mathcal{L}_h^*)'$, the ultraweak DPG formulation defined by (66) can be restated as the following system of variational equations:

$$\text{(70a)}\qquad \begin{cases} (\varepsilon,\nu)_{\mathcal{V}}\;+\;(u,\mathcal{L}_h^*\nu)_\Omega+\langle p,\nu\rangle_h=F(\nu) & \forall\,\nu\in H(\mathcal{L}_h^*),\\ (\mu,\mathcal{L}_h^*\varepsilon)_\Omega=0 & \forall\,\mu\in L^2,\\ \langle\rho,\varepsilon\rangle_h=0 & \forall\,\rho\in Q(\mathcal{L}_h^*). \end{cases}$$

Likewise, letting $\mathcal{V}=H(\mathcal{L}_h)$, an ultraweak DPG formulation corresponding to (66) may be defined for any $G=G_\Omega\times G_h\in(L^2\times Q(\mathcal{L}_h))'$:

$$\text{(70b)}\qquad \begin{cases} (v,\nu)_{\mathcal{V}}\;-\;(\lambda,\mathcal{L}_h\nu)_\Omega-\langle\sigma,\nu\rangle_h=0 & \forall\,\nu\in H(\mathcal{L}_h),\\ (\mu,\mathcal{L}_h v)_\Omega=G_\Omega(\mu) & \forall\,\mu\in L^2,\\ \langle\rho,v\rangle_h=G_h(\rho) & \forall\,\rho\in Q(\mathcal{L}_h). \end{cases}$$

Both of the formulations defined above relate to the primal problem (58a) with $u = v$. Clearly, the role of $\mathcal{L}_h$ and $\mathcal{L}_h^*$ can be interchanged if a solution of the dual problem (58b) is of interest.

The link between DPG and least-squares methods is well established in the literature (see e.g. [91]). DPG* methods, as it turns out, can be readily identified with the category of so-called $\mathcal{L}\mathcal{L}^*$ methods [24]. In this section, we briefly illustrate this and a couple other notable relationships in the context of the mixed problems introduced in Section 2.3.

4.4.1 Least-squares and $\mathcal{L}\mathcal{L}^*$ methods

Let $\mathcal{V} = L^2$. It is well-known that least-squares finite element methods [14] follow from the following saddle-point formulation (cf. (12a) and (70a)):

$$\begin{cases} (\varepsilon, \nu)_\Omega + (\mathcal{L}u, \nu)_\Omega = F(\nu) & \forall\, \nu \in L^2\,, \\ (\mathcal{L}\mu, \varepsilon)_\Omega \qquad\quad = 0 & \forall\, \mu \in H_0(\mathcal{L})\,. \end{cases} \tag{71}$$

This may be identified with a mixed problem akin to (12a) using the strong formulation of (58a), rather than the ultraweak formulation, as in (70a). Indeed, let $\mathcal{R} = \mathcal{R}_{L^2}$ be the L^2 Riesz operator in (71) and recall identity (16), where $(\mathcal{B}\mu)(\cdot) = (\mathcal{L}\mu, \cdot)_\Omega$. Then observe that $\mathcal{B}'\,\mathcal{R}_\mathcal{V}^{-1}\,\mathcal{B} = (\mathcal{R}\,\mathcal{L})'\,\mathcal{R}^{-1}(\mathcal{R}\,\mathcal{L}) = \mathcal{L}'\,\mathcal{R}\,\mathcal{L}$ and $\mathcal{B}'\,\mathcal{R}_\mathcal{V}^{-1}\,F = \mathcal{L}'F$. That is,

$$\langle \mathcal{B}'\,\mathcal{R}_\mathcal{V}^{-1}\,\mathcal{B}u, \mu\rangle = \langle \mathcal{B}'\,\mathcal{R}_\mathcal{V}^{-1}\,F, \mu\rangle \qquad \Longleftrightarrow \qquad (\mathcal{L}u, \mathcal{L}\mu)_\Omega = F(\mathcal{L}\mu)\,.$$

In the case $F(\cdot) = (f, \cdot)_\Omega$, $\mathcal{B}'\,\mathcal{R}_\mathcal{V}^{-1}\,F = \mathcal{L}'\,\mathcal{R}\,f$ and so $F(\mathcal{L}\nu) = (f, \mathcal{L}\nu)_\Omega$. Therefore, the variational equation above can be readily identified with the first-order optimality condition for the functional $\mathcal{J} : u \mapsto \|\mathcal{L}u - f\|_{L^2}^2$, $\partial_u \mathcal{J} = 0$.

Contrary to (71), so-called $\mathcal{L}\mathcal{L}^*$ methods [24] relate to the following system (cf. (12b) and (70b)):

$$\begin{cases} (v, \nu)_\Omega - (\mathcal{L}^*\lambda, \nu)_\Omega = 0 & \forall\, \nu \in L^2\,, \\ (\mathcal{L}^*\mu, v)_\Omega \qquad\quad = G(\mu) & \forall\, \mu \in H_0(\mathcal{L}^*)\,. \end{cases}$$

Likewise, consider (18), where $(\mathcal{B}\mu)(\cdot) = (\mathcal{L}^*\mu, \cdot)_\Omega$ and $G(\cdot) = (f, \cdot)_\Omega$. In this case, we see that $\mathcal{L}\mathcal{L}^*$ formulations may again be identified with (58a), in this case using a saddle-point

expression akin to (12b). Indeed, observe that

$$\langle \mathcal{B}' \mathcal{R}_{\mathcal{V}}^{-1} \mathcal{B}\lambda, \mu \rangle = \langle G, \mu \rangle \qquad \Longleftrightarrow \qquad (\mathcal{L}^*\lambda, \mathcal{L}^*\mu)_\Omega = (f, \mu)_\Omega \,.$$

The variational equation above indicates, in a weak sense, that $\mathcal{L}\mathcal{L}^*\lambda = f$. Recalling that the solution is determined by the transformation $v = \mathcal{R}_{\mathcal{V}}^{-1}\mathcal{B}\lambda = \mathcal{L}^*\lambda$, we have $\mathcal{L}v = f$ weakly, as well.

4.4.2 Weakly conforming least-squares methods

A weakly conforming least squares method [60] for the primal problem (58a) seeks a minimizer of the least squares functional

$$w \mapsto \|\mathcal{L}w - f\|_{L^2}^2 \,,$$

under the conformity constraint

$$\langle w, \rho \rangle_h = 0 \quad \forall\, \rho \in Q(\mathcal{L}_h) \,.$$

Here, of course, the operator $\mathcal{L}$ is understood element-wise so we may saliently replace it by $\mathcal{L}_h$. This leads to the following saddle-point problem for the two solution components w and σ:

$$\begin{cases} (\mathcal{L}_h w, \mathcal{L}_h \nu)_\Omega + \langle \sigma, \nu \rangle_h = (f, \mathcal{L}_h \nu)_\Omega & \forall\, \nu \in H(\mathcal{L}_h) \,, \\ \langle w, \rho \rangle_h = 0 & \forall\, \rho \in Q(\mathcal{L}_h) \,. \end{cases} \tag{72}$$

If we use an ultraweak DPG* formulation (70b) with its corresponding graph inner product (60), scaled by an arbitrary constant $\alpha > 0$, we arrive at

$$\begin{cases} (\mathcal{L}_h v, \mathcal{L}_h \nu)_\Omega + \alpha (v, \nu)_\Omega - (\lambda, \mathcal{L}_h \nu)_\Omega - \langle \sigma, \nu \rangle_h = 0 & \forall\, \nu \in H(\mathcal{L}_h) \,, \\ (\mu, \mathcal{L}_h v)_\Omega = (f, \mu)_\Omega & \forall\, \mu \in L^2 \,, \\ \langle v, \rho \rangle_h = 0 & \forall\, \rho \in Q(\mathcal{L}_h) \,. \end{cases} \tag{73}$$

From the second equation in (73), observe that $f = \mathcal{L}_h v$. Therefore, the first equation can be rewritten as

$$(f - \lambda, \mathcal{L}_h \nu)_\Omega + \alpha (v, \nu)_\Omega = \langle \sigma, \nu \rangle_h \quad \forall\, \nu \in H(\mathcal{L}_h) \,.$$

Testing only with $\nu \in H(\mathcal{L})$, so that the term $\langle \sigma, \nu \rangle_h$ vanishes, it can now be seen that $\lambda \to f$ as $\alpha \to 0$. Consequently, this particular DPG* formulation can be viewed as a regularization of the weakly conforming least-squares formulation (72).

4.4.3 Coercive problems

Each of the examples presented thus far have only considered strong or ultraweak formulations of the PDE (58a) and (58b). A broad broken space theory for DPG methods in more traditional weak formulations exists and has been applied to several different boundary value problems in [28, 47, 65, 89]. For brevity, we will not expand on the necessary details here, but instead only allude to the generality of the present theory by remarking on the coercive and unbroken setting.

Let $\mathcal{U} = \mathcal{V}$. Consider any coercive bilinear form $a(\cdot,\cdot) : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ with coercivity constant $\alpha > 0$, $\alpha\|\nu\|_{\mathcal{V}} \leq a(\nu,\nu)$ for all $\nu \in \mathcal{V}$, and define the bilinear form $b = a$. For the problem of finding $u = v \in \mathcal{V}$ such that $a(u,\nu) = \ell(\nu)$, for all $\nu \in \mathcal{V}$, we may consider the following two mixed formulations:

$$\begin{cases} (\varepsilon,\nu)_{\mathcal{V}} + a(u,\nu) = \ell(\nu) & \forall\, \nu \in \mathcal{V}\,, \\ a(\varepsilon,\mu) = 0 & \forall\, \mu \in \mathcal{V}\,, \end{cases} \qquad \begin{cases} (v,\nu)_{\mathcal{V}} - a(\lambda,\nu) = 0 & \forall\, \nu \in \mathcal{V}\,, \\ a(v,\mu) = \ell(\nu) & \forall\, \mu \in \mathcal{V}\,. \end{cases}$$

The broken analogs of these and other types of variational formulations can used in DPG and DPG* methods.

Chapter 5

Examples

This chapter presents explicit examples of ultraweak DPG and DPG* methods. For brevity, only the two specific PDEs which are later accompanied by numerical experiments (see Chapter 9) are included here. The first example is the Poisson equation. This derivation is presented in the context of both DPG and DPG* methods using the rigorous procedure outlined in the previous chapter. The second example pertains to a class of well established viscoelastic fluid models. In contrast to Section 5.1, the derivation in this latter example is physically motivated and we also only treat the DPG scenario. Here, we note that the corresponding dual problem, realized as a DPG* method, must still be solved with the goal-oriented adaptive mesh refinement strategy featured in Section 9.3.5. In both examples, the solution variables u, v, ε, and λ from the previous chapters are replaced by related group variables.

5.1 Example 1: Poisson's equation

In this example, we derive the ultraweak bilinear forms which appear in DPG and DPG* methods for Poisson's equation (54). We then introduce two corresponding inner products, $(\cdot,\cdot)_{\mathcal{V}}$, which will be used with these methods in later analysis. Here, $\vec{u} = (u, \vec{q}, \widehat{u}, \widehat{q})$ and $\vec{v} = (v, \vec{p})$ will denote the DPG and DPG* solution variables, respectively. Similarly, $\vec{\varepsilon} = (\varepsilon, \vec{\xi})$ and $\vec{\lambda} = (\lambda, \vec{\zeta}, \widehat{\lambda}, \widehat{\zeta})$ will denote the associated residual and Lagrange multiplier.

Recall (55) and (56). On a bounded open set $\Omega \subseteq \mathbb{R}^d$ with connected Lipschitz boundary, set $m = d + 1$ and define

$$\mathcal{L}\,\vec{\mu} = (-\operatorname{div}\vec{\sigma}, \vec{\sigma} - \operatorname{grad}\mu), \tag{74}$$

where $\vec{\mu} = (\mu, \vec{\sigma})$. Here, $\vec{\sigma} : \Omega \to \mathbb{R}^d$ represents a flux variable and $\mu : \Omega \to \mathbb{R}$ represents a

field variable. Note that the equation $\mathcal{L}(u,\vec{q}) = (f,\vec{0})$, after elimination of $\vec{q}$, results in the well-known Poisson equation $-\Delta u = f$.

Let us now derive the corresponding ultraweak bilinear forms. Begin by observing that $\mathcal{L}^*$, given by (56), can be written as

$$\mathcal{L}^*\,\vec{\nu} = (\operatorname{div}\vec{\tau}, \vec{\tau} + \operatorname{grad}\nu)\,,$$

where $\vec{\nu} = (\nu,\vec{\tau})$. Obviously (57a) is satisfied. By the triangle inequality, we immediately see that both $H(\mathcal{L})$ and $H(\mathcal{L}^*)$ coincide with $H(\operatorname{div},\Omega)\times H^1(\Omega)$. Now, using integration by parts, observe that

$$\langle \mathscr{D}^*(\mu,\vec{\sigma}),(\nu,\vec{\tau})\rangle_h = -\langle \mathscr{D}(\nu,\vec{\tau}),(\mu,\vec{\sigma})\rangle_h = \langle \vec{\tau}\cdot\vec{n},\mu\rangle_{H^{1/2}(\partial\Omega)} + \langle \vec{\sigma}\cdot\vec{n},\nu\rangle_{H^{1/2}(\partial\Omega)}\,, \tag{75}$$

where $\vec{n}$ denotes the unit outward normal on $\partial\Omega$.

For boundary conditions, consider

$$H_0(\mathcal{L}) = H_0(\mathcal{L}^*) = H_0^1(\Omega)\times H(\operatorname{div},\Omega)\,. \tag{76}$$

This choice corresponds to the Dirichlet problem, $u = 0$ on $\partial\Omega$, as we shall demonstrate. From (75), it follows that (57c) holds. Along the lines of (75), we also have

$$\langle \mathscr{D}_h(\mu,\vec{\sigma}),(\nu,\vec{\tau})\rangle_h = -\sum_{K\in\Omega_h}\Big[\langle \vec{\tau}\cdot\vec{n},\mu\rangle_{H^{1/2}(\partial K)} + \langle \vec{\sigma}\cdot\vec{n},\nu\rangle_{H^{1/2}(\partial K)}\Big].$$

Therefore, $Q(\mathcal{L}_h)$, being the range of $\mathscr{D}_h^*|_{H_0(\mathcal{L}^*)}$, can be characterized as $Q(\mathcal{L}_h) = H_0^{1/2}(\partial\Omega_h)\times H^{-1/2}(\partial\Omega_h)$, where

$$H_0^{1/2}(\partial\Omega_h) = \Big\{\widehat{\mu}\in\prod_{K\in\Omega_h} H^{1/2}(\partial K) : \exists\,\mu\in H_0^1(\Omega) \text{ such that } \widehat{\mu}|_{\partial K} = \mu|_{\partial K}\Big\}, \tag{77a}$$

$$H^{-1/2}(\partial\Omega_h) = \Big\{\widehat{\sigma}\in\prod_{K\in\Omega_h} H^{-1/2}(\partial K) : \exists\,\vec{\sigma}\in H(\operatorname{div},\Omega) \text{ such that } \widehat{\sigma}|_{\partial K} = \vec{\sigma}\cdot\vec{n}|_{\partial K}\Big\}. \tag{77b}$$

Moreover, observe that $H(\mathcal{L}_h) = H^1(\Omega_h)\times H(\operatorname{div},\Omega_h)$, where

$$H^1(\Omega_h) = \prod_{K\in\Omega_h} H^1(K), \qquad\qquad H(\operatorname{div},\Omega_h) = \prod_{K\in\Omega_h} H(\operatorname{div},K).$$

Evidently, $Q(\mathcal{L}_h) = Q(\mathcal{L}_h^*)$ and $H(\mathcal{L}_h) = H(\mathcal{L}_h^*)$. Putting these deductions together, the broken ultraweak bilinear form in (61a) becomes

$$\begin{aligned} b_1((\mu,\vec{\sigma},\widehat{\mu},\widehat{\sigma}),(\nu,\vec{\tau})) = &\sum_{K\in\Omega_h}\Big[(\vec{\sigma},\vec{\tau}+\operatorname{grad}_h\nu)_K + (\mu,\operatorname{div}_h\vec{\tau})_K\Big] \\ &- \sum_{K\in\Omega_h}\Big[\langle\widehat{\sigma},\nu\rangle_{H^{1/2}(\partial K)} + \langle\vec{\tau}\cdot\vec{n},\widehat{\mu}\rangle_{H^{1/2}(\partial K)}\Big]. \end{aligned} \tag{78a}$$

where $(\mu,\vec{\sigma}) \in L^2$, $(\widehat{\mu},\widehat{\sigma}) \in Q(\mathcal{L}_h^*)$, and $(\nu,\vec{\tau}) \in H(\mathcal{L}_h^*)$. Here, recall that $(\cdot,\cdot)_K$ denotes the inner product in $L^2(K)$ or its Cartesian products. Likewise, the corresponding broken ultraweak bilinear form in (61b) is

$$\begin{aligned} b_2((\mu,\vec{\sigma},\widehat{\mu},\widehat{\sigma}),(\nu,\vec{\tau})) = &\sum_{K\in\Omega_h}\Big[(\vec{\sigma},\vec{\tau}-\operatorname{grad}_h\nu)_K - (\mu,\operatorname{div}_h\vec{\tau})_K\Big] \\ &+ \sum_{K\in\Omega_h}\Big[\langle\widehat{\sigma},\nu\rangle_{H^{1/2}(\partial K)} + \langle\vec{\tau}\cdot\vec{n},\widehat{\mu}\rangle_{H^{1/2}(\partial K)}\Big], \end{aligned} \tag{78b}$$

The bijectivity of $\mathcal{L},\mathcal{L}^* : H_0(\mathcal{L}) \to L^2$ can be proved by standard techniques (see e.g. [45]). For the test space, $H(\mathcal{L}_h) = H(\mathcal{L}_h) = H^1(\Omega_h) \times H(\operatorname{div},\Omega_h)$, clearly any norms equivalent to the graph norms (59) may be used in the mixed problems (54). In the coming chapters, we consider both the graph norms $\|(\nu,\vec{\tau})\|^2_{H(\mathcal{L}_h)}$ and $\|(\nu,\vec{\tau})\|^2_{H(\mathcal{L}_h^*)}$ as candidates, as well as the norm

$$\|(\nu,\vec{\tau})\|^2_{H^1(\Omega_h)\times H(\operatorname{div},\Omega_h)} = \sum_{K\in\mathcal{T}}\Big(\|\operatorname{grad}\nu\|^2_{L^2(K)} + \|\operatorname{div}\vec{\tau}\|^2_{L^2(K)}\Big) + \|\nu\|^2_{L^2} + \|\vec{\tau}\|^2_{L^2}\,.$$

Remark 5.1. Let $K \subseteq \Omega$ be a connected subdomain with Lipschitz boundary and define the domain-dependent trace operators $\operatorname{tr}^K : H^1(K) \to H^{1/2}(\partial K)$ and $\operatorname{tr}_n^K : H(\operatorname{div},K) \to H^{-1/2}(\partial K)$, where $\operatorname{tr}^K u = u|_{\partial K}$ and $\operatorname{tr}_n^K \vec{\sigma} = \vec{\sigma}|_{\partial K}\cdot\boldsymbol{n}$ for smooth functions. Using these definitions, construct the corresponding mesh-dependent trace operators $\operatorname{tr} = \prod_{K\in\mathcal{T}}\operatorname{tr}^K$ and $\operatorname{tr}_n = \prod_{K\in\mathcal{T}}\operatorname{tr}_n^K$. Using the divergence theorem, these operators may be identified with $\mathscr{D}_h$ and $\mathscr{D}_h^*$ coming from the simple differential operator $\mathcal{L} = \operatorname{grad}$ [28, 49]. Now, because the bijective operator $\mathcal{L}$ in (74) is defined simply using grad and its (negative) dual div, the bilinear form in (78b) can be expressed in another convenient form. Namely, also adopting the convenient overloaded notation,

$$\langle\operatorname{tr}\mu,\vec{\sigma}\cdot\vec{n}\rangle_h = \langle\operatorname{tr}_n\vec{\sigma},\mu\rangle_h = (\operatorname{grad}_h\mu,\vec{\sigma})_\Omega + (\mu,\operatorname{div}_h\vec{\sigma})_\Omega\,, \tag{79}$$

allows the bilinear in (78b) to be rewritten in the following simplified form, which will be useful in some of the proceeding analysis:

$$b((\mu,\vec{\sigma},\widehat{\mu},\widehat{\sigma}),(\nu,\vec{\tau})) = (\vec{\sigma},\vec{\tau}-\operatorname{grad}_h\nu)_\Omega - (\mu,\operatorname{div}_h\vec{\tau})_\Omega + \langle\widehat{\mu},\vec{\tau}\cdot\vec{n}\rangle_h + \langle\widehat{\sigma},\nu\rangle_h\,. \tag{80}$$

5.2 Example 2: Viscoelastic fluid flow

The notation in this section is distinguished from the previous section by relying on symbols in bold font for vector- and matrix-valued functions (e.g. tensors) and by using the symbols ∇ and $\nabla\cdot$ to denote the gradient and divergence operators, respectively. Additionally, the symbols $\boldsymbol{\nabla}$ and $\boldsymbol{\nabla}\cdot$ are used here to denoted the row-wise gradient and divergence operators, respectively. When necessary, the element-wise definitions of these distributional operators are again labeled with the h-subscript.

In this section, we consider continua in which the constitutive equation for the Cauchy stress tensor $\boldsymbol{\sigma}$ is defined by the following viscoelastic fluid law:

$$\boldsymbol{\sigma} = -p\,\mathbf{I} + \eta_{\mathrm{S}}\big(\boldsymbol{\nabla}\,\mathbf{u} + \boldsymbol{\nabla}\,\mathbf{u}^{\mathsf{T}}\big) + \mathbf{T}\,. \tag{81}$$

Here, p is the *pressure*, $\mathbf{u}$ is the *fluid velocity*, and η_{S} is the *solvent viscosity*. The non-Newtonian term, the *extra stress tensor* $\mathbf{T}$, is an additional variable governed by an ancillary relationship, weakly coupled to the underlying kinematic conservation laws. For instance, the Giesekus model [76], with *mobility factor* $\alpha\in[0,1]$, characterizes the advection and decay of $\mathbf{T}$ along the streamlines of the fluid by the rule

$$\mathbf{T} + \lambda\,\mathfrak{L}_{\mathbf{u}}\mathbf{T} + \alpha\frac{\lambda}{\eta_{\mathrm{P}}}\mathbf{T}^2 = \eta_{\mathrm{P}}\big(\boldsymbol{\nabla}\,\mathbf{u} + \boldsymbol{\nabla}\,\mathbf{u}^{\mathsf{T}}\big)\,. \tag{82}$$

Here, $\lambda>0$ is the *relaxation time*, η_{P} is the *polymeric viscosity*, and $\mathfrak{L}$ is the *Lie derivative* operator—in this situation, acting on $\mathbf{T}$ in the direction $\mathbf{u}$ and, in this case, often called the *upper-convected Maxwell derivative* [100]—*viz.*,

$$\mathfrak{L}_{\mathbf{u}}\mathbf{T} = \frac{\partial\mathbf{T}}{\partial t} + (\mathbf{u}\cdot\boldsymbol{\nabla})\mathbf{T} - (\boldsymbol{\nabla}\,\mathbf{u})\mathbf{T} - \mathbf{T}(\boldsymbol{\nabla}\,\mathbf{u})^{\mathsf{T}}\,. \tag{83}$$

Notably, when $\alpha = 0$ the Giesekus model reduces to the Oldroyd-B model [113]. For a more elaborate description of the physical nature of this fluid model, consult [90, 115].

Arguably, the mathematically rigorous derivation of variational formulations for Poisson's equation in the previous section was considerably tedious. Alternative to that approach, a more physically motivated derivation of ultraweak variational formulations can also be be considered. This is very useful when handling much more complicated PDE models. For illustration, consider the non-inertial conservation of momentum which appears in Stokes' equations:

$$-\boldsymbol{\nabla}\cdot\boldsymbol{\sigma} = \rho\,\mathbf{f}\,.$$

Here, ρ is the *mass density*, $\boldsymbol{\sigma}$ is again the Cauchy stress, and $\mathbf{f}$ is a *body force density*. Multiplying this equation by a test velocity $\mathbf{v}$, integrating over a single element K, and then integrating by parts, we obtain

$$(\boldsymbol{\sigma}, \boldsymbol{\nabla}\,\mathbf{v})_K - \langle \boldsymbol{\sigma}\cdot\boldsymbol{n}_K, \mathbf{v}\rangle_{\partial K} = (\rho\,\mathbf{f}, \mathbf{v})_K\,. \tag{84}$$

Here, assuming that the arguments themselves are sufficiently smooth, we may simply understand the boundary term $\langle\cdot,\cdot\rangle_{\partial K}$ to denote the integral over the element boundary of the scalar product of its two arguments.

As appeared rigorously in (78), let us simply require that the interior values of the stress variable $\boldsymbol{\sigma}$ in (84) inside the element be disassociated from the *traction* on the boundary, $\boldsymbol{\sigma}|_{\partial K}\cdot\boldsymbol{n}_K$. To do this, simply introduce a new unknown flux variable $\widehat{\mathbf{t}}$, defined only on the mesh skeleton, to replace the element traction term. Then, summing over each element in the mesh, we arrive at the equation

$$(\boldsymbol{\sigma}, \boldsymbol{\nabla}_h\,\mathbf{v})_\Omega - \langle\widehat{\mathbf{t}}, \mathbf{v}\rangle_h = (\rho\,\mathbf{f}, \mathbf{v})_\Omega\,.$$

Similar to the previous section, $\boldsymbol{\nabla}_h$ here indicates that the gradient is intended element-wise and $\langle\cdot,\cdot\rangle_h$ indicates the accumulation of all related boundary terms.

We now extend a similar formal derivation to the following system of non-transient equations coming from the constitutive law (82) with the *mobility factor* $\alpha = 0$:

$$\rho(\mathbf{L}\mathbf{u}) + \nabla p - \eta_{\mathrm{S}}\,\boldsymbol{\nabla}\cdot\mathbf{L} - \boldsymbol{\nabla}\cdot\mathbf{T} = \rho\,\mathbf{f}\,, \tag{85a}$$

$$\mathbf{L} - \boldsymbol{\nabla}\,\mathbf{u} = \mathbf{0}\,, \tag{85b}$$

$$\nabla\cdot\ \mathbf{u} = 0\,, \tag{85c}$$

$$\mathbf{T} + \lambda\,\mathcal{L}_{\mathbf{u}}\,\mathbf{T} - \eta_{\mathrm{P}}(\mathbf{L} + \mathbf{L}^{\mathsf{T}}) = \mathbf{0}\,. \tag{85d}$$

Recall that the variable $\mathbf{T}$ is the extra stress tensor and notice that $\mathbf{L}$ is simply the *velocity gradient*. Because of the advective term $\rho(\mathbf{L}\mathbf{u})$ in the momentum equation (85a), this describes an inertial Oldroyd-B fluid model.

In the system above (85), equations (85a) and (85b), together, impose conservation of linear momentum in the fluid. Meanwhile, (85c) is simply the conservation of mass under the assumption of incompressibility or, equivalently, the conservation of volume. Finally, (85d) is simply a restatement of (83) for a steady flow. Here, we have also introduced the *autonomous Lie derivative* [101, Section 1.6] of $\mathbf{T}$ in the direction of $\mathbf{u}$, $\mathcal{L}_{\mathbf{u}}\,\mathbf{T} = \mathfrak{L}_{\mathbf{u}}\,\mathbf{T} - \frac{\partial\mathbf{T}}{\partial t}$. Ultimately, each of the solution variables $\mathbf{u}$, p, $\mathbf{L}$, and $\mathbf{T}$ found in (85) will be involved in our ultraweak formulation.

Now, formally test (85a) with smooth vector fields, $\mathbf{v}$, (85b) with smooth tensors, $\mathbf{M}$, (85c) with smooth functions, q, and (85d) with smooth *symmetric* tensors, $\mathbf{S}$. If we assume that these test functions vanish on the domain boundary $\partial\Omega$, then, after integration by parts over each element and forming the sum of the resulting equations over all elements $K \in \mathcal{T}$, we arrive at the following ultraweak nonlinear form:

$$\begin{aligned} b_0^{\mathrm{nl.}}((\mathbf{u},p,\mathbf{L},\mathbf{T}),(\mathbf{v},q,\mathbf{M},\mathbf{S})) = {}& \rho(\mathbf{L}\mathbf{u},\mathbf{v})_{\Omega} - (p,\nabla_h\cdot\mathbf{v})_{\Omega} + \eta_{\mathrm{S}}(\mathbf{L},\boldsymbol{\nabla}_h\,\mathbf{v})_{\Omega} + (\mathbf{T},\boldsymbol{\nabla}_h\,\mathbf{v})_{\Omega} \\ & + (\mathbf{L},\mathbf{M})_{\Omega} + (\mathbf{u},\boldsymbol{\nabla}_h\cdot\mathbf{M})_{\Omega} - (\mathbf{u},\nabla_h\,q)_{\Omega} \\ & + (\mathbf{T},\mathbf{S})_{\Omega} - \lambda(\mathbf{T}\otimes\mathbf{u},\boldsymbol{\nabla}_h\,\mathbf{S})_{\Omega} - 2\lambda(\mathbf{L}\,\mathbf{T},\mathbf{S})_{\Omega} - 2\eta_{\mathrm{P}}(\mathbf{L},\mathbf{S})_{\Omega}\,. \end{aligned}$$

If, however, the test functions are unrestricted on the boundary of the domain and are allowed to be discontinuous across element interfaces then the additional contributions readily collect

together to form the interface terms

$$\widehat{b}((\widehat{\mathbf{u}},\widehat{\mathbf{t}},\widehat{\mathbf{j}}),(\mathbf{v},q,\mathbf{M},\mathbf{S})) = -\langle\widehat{\mathbf{t}},\mathbf{v}\rangle_h - \langle\widehat{\mathbf{u}},\mathbf{M}\boldsymbol{n}\rangle_h + \langle\widehat{\mathbf{u}}\cdot\boldsymbol{n},q\rangle_h + \lambda\langle\widehat{\mathbf{j}},\mathbf{S}\rangle_h\,.$$

If we assume a smooth solution near the boundary of each element of the mesh, $K \in \mathcal{T}$, the new interface solution variables can be identified with the restriction of the velocity field, $\widehat{\mathbf{u}} = \mathbf{u}|_{\partial K}$, the normal flux of the Cauchy stress, $\widehat{\mathbf{t}} = \boldsymbol{\sigma}|_{\partial K}\boldsymbol{n}_K$, and the normal flux in the hybrid variable $\mathbf{T}\otimes\mathbf{u}$, $\widehat{\mathbf{j}} = (\mathbf{u}|_{\partial K}\cdot\boldsymbol{n}_K)\mathbf{T}|_{\partial K}$. In this case, the entire broken ultraweak nonlinear form becomes

$$b^{\text{nl.}}((\mathbf{u},p,\mathbf{L},\mathbf{T},\widehat{\mathbf{u}},\widehat{\mathbf{t}},\widehat{\mathbf{j}}),(\mathbf{v},q,\mathbf{M},\mathbf{S})) = b_0^{\text{nl.}}((\mathbf{u},p,\mathbf{L},\mathbf{T}),(\mathbf{v},q,\mathbf{M},\mathbf{S})) + \widehat{b}((\widehat{\mathbf{u}},\widehat{\mathbf{t}},\widehat{\mathbf{j}}),(\mathbf{v},q,\mathbf{M},\mathbf{S}))\,.$$

Following the Gauss–Newton strategy described in Section 3.8 requires that we define a linearization of the form $b^{\text{nl.}}$. Notice that $\widehat{b}(\cdot,\cdot)$ is linear in each of its solution arguments, $\widehat{\mathbf{u}},\widehat{\mathbf{t}}$ and $\widehat{\mathbf{j}}$. Meanwhile, $b_0^{\text{nl.}}(\cdot,\cdot)$ is only linear in the pressure variable, p (and, of course, in each test variable).

One may verify that $b_0^{\text{lin.}}[\mathbf{u}_0,\mathbf{L}_0,\mathbf{T}_0]$, given below, is the first variation of $b_0^{\text{nl.}}$ at the arbitrary point $\mathfrak{u}_0 = (\mathbf{u}_0,p_0,\mathbf{L}_0,\mathbf{T}_0)$:

$$\begin{aligned}
&b_0^{\text{lin.}}[\mathbf{u}_0,\mathbf{L}_0,\mathbf{T}_0]((\mathbf{u},p,\mathbf{L},\mathbf{T}),(\mathbf{v},q,\mathbf{M},\mathbf{S}))\\
&\quad= D_{\mathrm{u}}b_0^{\text{nl.}}[\mathbf{u}_0,0,\mathbf{L}_0,\mathbf{T}_0]((\mathbf{u},p,\mathbf{L},\mathbf{T}),(\mathbf{v},q,\mathbf{M},\mathbf{S}))\\
&\quad= \rho(\mathbf{L}_0\mathbf{u}+\mathbf{L}\mathbf{u}_0,\mathbf{v}) - (p,\nabla_h\cdot\mathbf{v}) + \eta_{\mathrm{S}}(\mathbf{L},\boldsymbol{\nabla}_h\mathbf{v}) + (\mathbf{T},\boldsymbol{\nabla}_h\mathbf{v}) + (\mathbf{L},\mathbf{M}) + (\mathbf{u},\boldsymbol{\nabla}\cdot\mathbf{M})\\
&\quad\quad - (\mathbf{u},\nabla_h q) + (\mathbf{T},\mathbf{S}) - \lambda(\mathbf{T}_0\otimes\mathbf{u}+\mathbf{T}\otimes\mathbf{u}_0,\boldsymbol{\nabla}_h\mathbf{S}) - 2\lambda(\mathbf{L}_0\,\mathbf{T}+\mathbf{L}\,\mathbf{T}_0,\mathbf{S}) - 2\eta_{\mathrm{P}}(\mathbf{L},\mathbf{S})\\
&\quad= (\mathbf{u},\rho\mathbf{L}_0^{\mathsf{T}}\mathbf{v} - \nabla_h q + \boldsymbol{\nabla}\cdot\mathbf{M} - \lambda\,\boldsymbol{\nabla}_h\mathbf{S}:\mathbf{T}_0)\\
&\quad\quad + (\mathbf{L},\eta_{\mathrm{S}}\,\boldsymbol{\nabla}_h\mathbf{v} + \rho\mathbf{v}\otimes\mathbf{u}_0 + \mathbf{M} - 2\eta_{\mathrm{P}}\mathbf{S} - 2\lambda\mathbf{S}\,\mathbf{T}_0)\\
&\quad\quad - (p,\nabla_h\cdot\mathbf{v}) + (\mathbf{T},\boldsymbol{\nabla}_h\mathbf{v} + \mathbf{S} - \lambda(\mathbf{u}_0\cdot\boldsymbol{\nabla})\mathbf{S} - 2\lambda\mathbf{L}_0^{\mathsf{T}}\mathbf{S})\,,
\end{aligned} \tag{86}$$

where $\boldsymbol{\nabla}_h\mathbf{S}:\mathbf{T}_0 = \sum_{i,j,k}(\partial_k\mathbf{S}_{ij})(\mathbf{T}_0)_{ij}\mathbf{e}_k$. Moreover, because of linearity in the arguments $p,\widehat{\mathbf{u}},\widehat{\mathbf{t}}$, and $\widehat{\mathbf{j}}$, we may use

$$\begin{aligned}
F^{\text{nl.}}[\mathbf{u}_0,\mathbf{L}_0,\mathbf{T}_0](\mathbf{v},q,\mathbf{M},\mathbf{S}) &= \rho(\mathbf{f},\mathbf{v}) - b^{\text{nl.}}((\mathbf{u}_0,0,\mathbf{L}_0,\mathbf{T}_0,\mathbf{0},\mathbf{0},\mathbf{0}),(\mathbf{v},q,\mathbf{M},\mathbf{S}))\\
&= \rho(\mathbf{f},\mathbf{v}) - b_0^{\text{nl.}}((\mathbf{u}_0,0,\mathbf{L}_0,\mathbf{T}_0),(\mathbf{v},q,\mathbf{M},\mathbf{S}))
\end{aligned}$$

in (51). The associated graph norm (57b) can now be readily computed from the final expression given in (86). Additional details are provided in [90].

Chapter 6

A priori error estimation

This chapter presents the fundamentals of analysis for *a priori* error estimation in the saddle-point paradigm developed above. In particular, it contains new theory for DPG* methods developed in [48].

6.1 Mixed methods

Having by now explained the connections between (53) and the mixed formulation (31), it should not be a surprise that the error analysis of these methods reduces to standard mixed method theory. To put the results in this chapter in context, recall the functions $v \in \mathcal{V}$ and $w \in \mathcal{U}$ from the equivalent saddle-point problems (13) and (31):

$$\begin{cases} (v,\nu)_{\mathcal{V}} + b(w,\nu) = F(\nu) & \forall\, \nu \in \mathcal{V}, \\ b(\mu, v) \qquad\qquad\quad = G(\mu) & \forall\, \mu \in \mathcal{U}. \end{cases} \tag{87}$$

The central assumption we need for the upcoming error analysis is the existence of a Fortin operator [82]. Namely, assume that there is a continuous linear operator $\Pi_h : \mathcal{V} \to \mathcal{V}_h$ such that

$$b(\mu, \nu - \Pi_h \nu) = 0 \qquad \forall\, \mu \in \mathcal{U}_h,\ \nu \in \mathcal{V}. \tag{88}$$

Under this assumption, the standard theory of mixed methods [15] yields the following *a priori* estimate for the solutions $v_h \in \mathcal{V}_h$ and $w_h \in \mathcal{U}_h$ coming from the corresponding discrete problem (53).

Theorem 6.1. *Suppose* (14) *and* (88) *hold. Let $F \in \mathcal{V}'$ and $G \in \mathcal{U}'$. Then there is a constant C such that the complete solution of* (31)*, $(v,w) \in \mathcal{V} \times \mathcal{U}$, satisfies the error estimate*

$$\|v - v_h\|_{\mathcal{V}} + \|w - w_h\|_{\mathcal{U}} \le C \left[\inf_{\nu \in \mathcal{V}_h} \|v - \nu\|_{\mathcal{V}} + \inf_{\mu \in \mathcal{U}_h} \|w - \mu\|_{\mathcal{U}} \right]. \tag{89}$$

6.2 The error in a quantity of interest

Here, it is advantageous to make use of the so-called discrete *trial-to-test* operator $\Theta_h : \mathcal{U}_h \to \mathcal{V}_h$, defined

$$(\Theta_h \mu, \nu)_{\mathcal{V}} = b(\mu, \nu) \qquad \forall\, \mu \in \mathcal{U}_h\,,\ \nu \in \mathcal{V}_h\,.$$

Generally, because $\dim \mathcal{V}_h > \dim \mathcal{U}_h$, the range of this operator is a proper closed subspace of the test space, $\Theta_h(\mathcal{U}_h) \subsetneq \mathcal{V}_h$. Let $F \in (\operatorname{Null} \mathcal{B}')^{\perp}$ and $G \in \mathcal{U}'$. Now, consider (87). In the DPG context, $G = 0$, so (87) reduces to the system

$$\begin{cases} (\varepsilon, \nu)_{\mathcal{V}} + b(u, \nu) = F(\nu) & \forall\, \nu \in \mathcal{V}, \\ b(\mu, \varepsilon) = 0 & \forall\, \mu \in \mathcal{U}. \end{cases} \tag{90a}$$

Likewise, in DPG* context, $F = 0$ and (87) reduces to

$$\begin{cases} (v, \nu)_{\mathcal{V}} - b(\lambda, \nu) = 0 & \forall\, \nu \in \mathcal{V}, \\ b(\mu, v) = G(\mu) & \forall\, \mu \in \mathcal{U}. \end{cases} \tag{90b}$$

The following fundamental result demonstrates that the duality seen in (37) also holds for the errors in these general saddle-point methods (including DPG and DPG*), even though these two discrete solutions exist in spaces of *different* dimensions. This may be viewed as a generalization of [110, Lemma 3.1] or [10, Equation 1.8].

Theorem 6.2. *Let u_h be the $\mathcal{U}_h$-solution of* (54a) *and v_h be the $\mathcal{V}_h$-solution of* (54b)*. Likewise, let u and v be the exact solutions of the continuous problems* (90a) *and* (90b)*, respectively. Then the following identity holds for all $\mu \in \mathcal{U}_h$:*

$$G(u - u_h) = b(u - u_h, v - \Theta_h \mu) = b(u - \mu, v - v_h) = F(v - v_h)\,. \tag{91}$$

Proof. Let ε_h be the $\mathcal{V}_h$-solution of (54a) and let ε be the corresponding exact solution. Due to Galerkin orthogonality (cf. (54a) and (90a)),

$$\begin{cases} (\varepsilon - \varepsilon_h, \nu)_{\mathcal{V}} + b(u - u_h, \nu) = 0 & \forall\, \nu \in \mathcal{V}_h\,, \\ b(\mu, \varepsilon - \varepsilon_h) = 0 & \forall\, \mu \in \mathcal{U}_h\,, \end{cases}$$

Substituting the first equation above into the second, we find that $b(u-u_h,\Theta_h\mu)=0$, for any $\mu\in\mathcal{U}_h$. Moreover, inspecting (54b), it is clear that $v_h\in\Theta_h(\mathcal{U}_h)$. Therefore,

$$G(u-u_h)=b(u-u_h,v)=b(u-u_h,v-\Theta_h\mu)=b(u-u_h,v-v_h)\,.$$

Similarly, due to Galerkin orthogonality (cf. (54b) and (90b)), $b(\mu,v-v_h)=0$, for any $\mu\in\mathcal{U}_h$. Finally, note that $\varepsilon=0$, since $F\in(\operatorname{Null}\mathcal{B}')^\perp$, and

$$F(v-v_h)=b(u,v-v_h)=b(u-\mu,v-v_h)=b(u-u_h,v-v_h)\,. \qquad \square$$

With the notation of Theorem 6.2, define the error in the quantity of interest to be $e_{\mathrm{QOI}}=b(u-u_h,v-v_h)$. By (91), $|e_{\mathrm{QOI}}|\le|||u-u_h|||_{\mathcal{U}}\|v-\Theta_h\mu\|_{\mathcal{V}}$, for all $\mu\in\mathcal{U}_h$. Similarly, $|e_{\mathrm{QOI}}|\le|||u-\mu|||_{\mathcal{U}}\|v-v_h\|_{\mathcal{V}}$, for all $\mu\in\mathcal{U}_h$. Invoking (33), we now arrive at the following corollary (cf. [77, Corollary 3.2]).

Corollary 6.3. *The following crude upper bounds hold:*

$$|e_{\mathrm{QOI}}|\le\|\mathcal{B}\|\cdot\begin{cases}\|u-u_h\|_{\mathcal{U}}\ \inf\limits_{\mu\in\mathcal{U}_h}\|v-\Theta_h\mu\|_{\mathcal{V}}\,,\\ \|v-v_h\|_{\mathcal{V}}\ \inf\limits_{\mu\in\mathcal{U}_h}\|u-\mu\|_{\mathcal{U}}\,.\end{cases}\tag{92}$$

6.3 DPG methods

Consider (54a) and (90a). Due to the structure of these equations, Theorem 6.1 can be significantly simplified.

Theorem 6.4. *Suppose* (14) *and* (88) *hold. Let* $F\in(\operatorname{Null}\mathcal{B}')^\perp$. *Then there is a constant* C *such that the complete DPG solution of* (90a), $(\varepsilon,u)\in\mathcal{V}\times\mathcal{U}$, *satisfies the error estimate*

$$\|\varepsilon_h\|_{\mathcal{V}}+\|u-u_h\|_{\mathcal{U}}\le C\ \inf_{\mu\in\mathcal{U}_h}\|u-\mu\|_{\mathcal{U}}\,.$$

Proof. Substitute v, w, v_h, and w_h in (89) with ε, u, ε_h, and u_h, respectively. By (24) and (26), we see that $\varepsilon=0$. The result follows immediately. $\square$

At times, it is possible to get an improvement on this bound using the Aubin–Nitsche duality argument. Indeed, suppose F is a functional in $\mathcal{V}'$ and we are interested in bounding $G(u-u_h)$, a functional of the error $u-u_h$. Thus, consider the functions $v \in \mathcal{V}$ and $\lambda \in \mathcal{U}$ solving (90b). To conduct the duality argument, we suppose that there is a positive $c_0(h)$ that goes to 0 as $h \to 0$ satisfying

$$\inf_{\mu \in \mathcal{U}_h} \|v - \Theta_h \mu\|_{\mathcal{V}} \leq \; c_0(h) \, \|G\|_{\mathcal{U}'} \, . \tag{93}$$

This usually holds when the solution of (90b) has sufficient regularity [69, 70]. In this case, we may state the following theorem.

Theorem 6.5. *Suppose* (93) *holds in addition to the assumptions of Theorem 6.4. Let* $F \in (\operatorname{Null} \mathcal{B}')^{\perp}$. *Then the error in the DPG* solution component* u_h *satisfies*

$$G(u-u_h) \leq \; c_0(h) \, \|G\|_{\mathcal{U}'} \, \|\mathcal{B}\| \inf_{\mu \in \mathcal{U}_h} \|u - \mu\|_{\mathcal{U}} \, .$$

Proof. From Corollary 6.3 and (93), we see that

$$G(u-u_h) \leq \|\mathcal{B}\| \, \|u-u_h\|_{\mathcal{U}} \inf_{\mu \in \mathcal{U}_h} \|v - \Theta_h \mu\|_{\mathcal{V}} \leq c_0(h) \, \|G\|_{\mathcal{U}'} \|\mathcal{B}\| \, \|u-u_h\|_{\mathcal{U}} \, .$$

The proof is completed by applying Theorem 6.4. □

Remark 6.6. Inequalities like (93) often arise due to elliptic regularity results which depend on the domain and, clearly, depend on the PDE under consideration. Due to the presence of the trial-to-test operator in (93), however, in DPG methods and in other methods in this framework, such inequalities will even depend upon the chosen norm. Exploiting duality in DPG methods, as evidenced above, as well by employing Bramble–Hilbert arguments, optimal convergence rates have been proven in the literature for many DPG methods. The interested reader is encouraged to consult [16, 69, 70] for further details.

6.4 DPG* methods

Consider (54b) and (90b). In this case, Theorem 6.1 cannot be simplified and we are generally still left with the following standard mixed method estimate.

Theorem 6.7. *Suppose* (14) *and* (88) *hold. Then there is a constant C such that the complete DPG* solution of* (90b)*, $(v, \lambda) \in \mathcal{V} \times \mathcal{U}$, satisfies the error estimate*

$$\|v - v_h\|_{\mathcal{V}} + \|\lambda - \lambda_h\|_{\mathcal{U}} \leq C \left[\inf_{\nu \in \mathcal{V}_h} \|v - \nu\|_{\mathcal{V}} + \inf_{\mu \in \mathcal{U}_h} \|\lambda - \mu\|_{\mathcal{U}} \right].$$

Again, it is still possible to get an improvement using the Aubin–Nitsche duality argument. In this case, suppose F is a functional in $(\operatorname{Null} \mathcal{B}')^{\perp}$ and suppose that we are instead interested in bounding $F(v - v_h)$, a functional of the error $v - v_h$. Now, consider the solution $u \in \mathcal{U}$ coming from (90a). In this argument, we must suppose that there is a positive function $c_1(h)$ that goes to 0 as $h \to 0$ satisfying

$$\inf_{\mu \in \mathcal{U}_h} \|u - \mu\|_{\mathcal{U}} \leq \; c_1(h) \, \|F\|_{\mathcal{V}'}. \tag{94}$$

Again, this usually holds when the solution of (90a) has sufficient regularity.

Theorem 6.8. *Suppose* (94) *holds in addition to the assumptions of Theorem 6.7. Then the error in the DPG* solution component v_h satisfies*

$$F(v - v_h) \leq \; c_1(h) \, \|F\|_{\mathcal{V}'} \, \|\mathcal{B}\| \left[\inf_{\nu \in \mathcal{V}_h} \|v - \nu\|_{\mathcal{V}}^2 + \inf_{\mu \in \mathcal{U}_h} \|\lambda - \mu\|_{\mathcal{U}}^2 \right]^{1/2}.$$

Proof. From Corollary 6.3 and (94), we see that

$$F(v - v_h) \leq \|\mathcal{B}\| \, \|v - v_h\|_{\mathcal{V}} \inf_{\mu \in \mathcal{U}_h} \|u - \mu\|_{\mathcal{U}} \leq c_1(h) \, \|F\|_{\mathcal{V}'} \, \|\mathcal{B}\| \, \|v - v_h\|_{\mathcal{V}} \, .$$

The proof is completed by applying Theorem 6.7. □

6.5 Application to Poisson's equation

At this point, an essential difficulty in DPG* methods (that was not present in DPG methods) becomes clear. Consider using a DPG* form $b(\cdot, \cdot)$ given by Theorem 4.3 for solving the primal problem $\mathcal{L}v = f$. Then the error in v_h computed by the DPG* method not only depends on the regularity of the solution v, but also on the regularity of an extraneous Lagrange multiplier λ. This is evident from the best approximation error bounds appearing in Theorems 6.7 and 6.8. The analysis of the Poisson equation carried out in this section clarifies this observation further.

Given $f \in L^2(\Omega)$, consider approximating the Dirichlet solution v

$$-\Delta v = f \quad \text{in } \Omega, \qquad v = 0 \quad \text{on } \partial\Omega, \tag{95}$$

by the DPG* method. We follow the setting of Section 5.1. Accordingly, we set

$$\mathcal{U} = L^2(\Omega) \times L^2(\Omega)^d \times H_0^{1/2}(\partial\Omega_h) \times H^{-1/2}(\partial\Omega_h), \qquad \mathcal{V} = H^1(\Omega_h) \times H(\mathrm{div}, \Omega_h),$$

where $H_0^{1/2}(\partial\Omega_h)$ and $H^{-1/2}(\partial\Omega_h)$ are defined in (77). Recall that the DPG* formulation of (95) characterizes two variables,

$$\vec{v} = (v, \vec{p}) \in \mathcal{V}, \qquad \vec{\lambda} = (\lambda, \vec{\zeta}, \widehat{\lambda}, \widehat{\zeta}) \in \mathcal{U},$$

satisfying

$$((v, \vec{p}), (\nu, \vec{\tau}))_{\mathcal{V}} - b_2((\lambda, \vec{\zeta}, \widehat{\lambda}, \widehat{\zeta}), (\nu, \vec{\tau})) = 0, \tag{96a}$$

$$b_2((\mu, \vec{\sigma}, \widehat{\mu}, \widehat{\sigma}), (v, \vec{p})) = (f, \mu)_\Omega, \tag{96b}$$

for all $\vec{\nu} = (\nu, \vec{\tau})$ in $\mathcal{V}$ and all $\vec{\mu} = (\mu, \vec{\sigma}, \widehat{\mu}, \widehat{\sigma})$ in $\mathcal{U}$. Here, $b_2(\cdot, \cdot)$ is given by (78b) and, as before, $(\cdot, \cdot)_\Omega$ denotes the inner product in $L^2(\Omega)$ (or its Cartesian products).

Define the operator $\mathcal{B}_2$ corresponding to the bilinear form b_2 as in (30). By Theorem 4.3, $\mathcal{B}_2$ is a bijection, so obviously (14) holds. Let Ω_h be a shape-regular mesh of simplices and let $P_p(K)$ denote the space of polynomials of degree at most p on a simplex K. Define $P^p(\partial K) = \{\mu : \mu|_F \in P(E)\ \forall$ codimension one sub-simplices E of $K\}$ and $\widetilde{P}^p(\partial K) = P^p(\partial K) \cap C^0(\partial K)$, where $C^0(D)$ denotes the set of all continuous functions on a domain D. A Fortin operator satisfying (88) for the case

$$\mathcal{U}_h = \{(\vec{\sigma}, \mu, \widehat{\sigma}, \widehat{\mu}) \in \mathcal{U} : \vec{\sigma}|_K \in P^p(K)^d,\ \mu|_K \in P^p(K),\ \widehat{\sigma} \in P^p(\partial K),\ \widehat{\mu} \in \widetilde{P}^{p+1}(\partial K)\},$$
$$\mathcal{V}_h = \{(\vec{\tau}, \nu) \in \mathcal{V} : \vec{\tau}|_K \in P^{p+d}(K)^d,\ \nu|_K \in P^{p+d}(K)\}.$$

was constructed in [82].

To understand the practical convergence rates in this DPG* method, we must ascertain the regularity of the corresponding Lagrange multiplier $\vec{\lambda}$. One way to do this is to write out

the boundary value problem that $\vec{\lambda}$ satisfies, as done in [16, 70]. An alternate technique can be seen in [69], which directly manipulates the variational equation (96a) using the information in (96b). We follow the latter approach in the proof of the following proposition.

Proposition 6.9. *The solution components $\lambda, \vec{\zeta}, \widehat{\lambda}, \widehat{\zeta}$ of the system* (96) *can be characterized using the remaining solution components $v, \vec{p}$ and f as*

$$\begin{aligned} \lambda &= f + e, & \widehat{\lambda} &= e, \\ \vec{\zeta} &= \vec{p} + \vec{r}, & \widehat{\zeta} &= 2\vec{p}\cdot\vec{n} + \vec{r}\cdot\vec{n}, \end{aligned} \tag{97}$$

where $(e, \vec{r})$ is in the space $H_0(\mathcal{L})$ defined in (76) *and satisfies the Dirichlet problem $\mathcal{L}(e, \vec{r}) = (v+2f, \vec{0})$ where $\mathcal{L}$ is as in* (74)*. Specifically, $e \in H_0^1(\Omega)$ satisfies $-\Delta e = v+2f$ and $\vec{r} = -\operatorname{grad} e$.*

Proof. Recall Theorem 4.3. We now know that (78b) and (96b) implies that $(v, \vec{p})$ satisfies $\mathcal{L}(v, \vec{p}) = (f, \vec{0})$, i.e.,

$$\vec{p} - \operatorname{grad} v = \vec{0}, \qquad -\operatorname{div} \vec{p} = f. \tag{98}$$

Next, we manipulate the first term of (96a) as follows:

$$\begin{aligned} ((v, \vec{p}), (\nu, \vec{\tau}))_{\mathcal{V}} &= (\vec{p}, \vec{\tau})_\Omega + (\operatorname{div} \vec{p}, \operatorname{div} \vec{\tau})_\Omega + (v, \nu)_\Omega + (\operatorname{grad} v, \operatorname{grad} \nu)_\Omega \\ &= (\vec{p}, \vec{\tau} - \operatorname{grad} \nu)_\Omega + (\operatorname{div} \vec{p}, \operatorname{div} \vec{\tau})_\Omega + (v, \nu)_\Omega + 2(\operatorname{grad} v, \operatorname{grad} \nu)_\Omega \\ &= (\vec{p}, \vec{\tau} - \operatorname{grad} \nu)_\Omega + (f, -\operatorname{div} \vec{\tau})_\Omega + (v, \nu)_\Omega \\ &\quad + 2 \sum_{K \in \Omega_h} \left[\langle \vec{n} \cdot \operatorname{grad} v, \nu \rangle_{H^{1/2}(\partial K)} - (\Delta v, \nu)_K \right] \\ &= b_2((f, \vec{p}, 0, 2\vec{p}\cdot\vec{n}), (\nu, \vec{\tau})) + (v + 2f, \nu)_\Omega, \end{aligned}$$

where we have used (98) twice. Now, let $e \in L^2(\Omega)$, $\vec{r} \in L^2(\Omega)^d$, and $(\widehat{e}, \widehat{r}) \in Q(\mathcal{L}_h)$ satisfy

$$b_2((e, \vec{r}, \widehat{e}, \widehat{r}), (\nu, \vec{\tau})) = (v + 2f, \nu)_\Omega$$

for all $(\nu, \vec{\tau}) \in H(\mathcal{L}_h) = \mathcal{V}$. This is a variational equation of the form (61a). Hence, by the first item of Theorem 4.1, both e and $\vec{r}$ are unique. Moreover, $(\vec{r}, e) \in H_0(\mathcal{L}^*)$ satisfies

$\mathcal{L}^*(\vec{r}, e) = (0, v + 2f)$, on Ω, and $e|_{\partial K} = \widehat{e}|_{\partial K}$ and $\vec{r} \cdot \vec{n}|_{\partial K} = \widehat{r}|_{\partial K}$, on all mesh element boundaries. Thus,

$$((v, \vec{p}), (\nu, \vec{\tau}))_{\mathcal{V}} = b_2((f + e, \vec{p} + \vec{r}, e, (2\vec{p} + \vec{r}) \cdot \vec{n}), (\nu, \vec{\tau})).$$

Comparing this with (96a), the result follows. □

We may now apply Theorem 6.7 along with standard Bramble-Hilbert arguments (see [82, Corollary 3.6] for details) to obtain convergence rates dictated by the following corollary to Theorem 6.7.

Corollary 6.10. *Let $h = \max_{K \in \Omega_h} \operatorname{diam}(K)$, $d = 2, 3$, and let the assumptions of Theorem 6.7 and Proposition 6.9 hold. Let $\vec{v}_h = (v_h, \vec{p}_h) \in \mathcal{V}_h$ and $\vec{\lambda}_h = (\lambda_h, \vec{\zeta}_h, \widehat{\lambda}_h, \widehat{\zeta}_h) \in \mathcal{U}_h$ be the DPG* solutions to* (54b)*, with $G(\vec{\mu}) = (f, \mu)$. Let $e \in H_0^1(\Omega)$ satisfy $-\Delta e = v + 2f$. Then*

$$\|\vec{v} - \vec{v}_h\|_{\mathcal{V}} + \|\vec{\lambda} - \vec{\lambda}_h\|_{\mathcal{U}} \leq Ch^s\big(\|v\|_{H^{s+2}(\Omega)} + \|e\|_{H^{s+2}(\Omega)}\big)$$

for all $1/2 < s < p + 1$.

Proof. Note that the left-hand side of the inequalities in Theorem 6.7 and Corollary 6.10 coincide. We now consider the right-hand side of the inequality in Theorem 6.7. The following inequality,

$$\inf_{\vec{\nu} \in \mathcal{V}_h} \|\vec{v} - \vec{\nu}\|_{\mathcal{V}}^2 \leq Ch^s\big(\|v\|_{H^{s+1}(\Omega)} + \|\vec{p}\|_{H^{s+1}(\Omega)}\big),$$

is immediate by substituting the function $\vec{\nu} = \prod_{K \in \mathcal{T}} (\Pi_{\text{grad}}^K v, \Pi_{\text{div}}^K \vec{p})$, where $\Pi_{\text{grad}}^K : H^1(K) \to P^{p+1}(K)$ and $\Pi_{\text{div}}^K : H(\text{div}, K) \to \vec{x}P^p(K) + P^p(K)^d$ are local nodal and Raviart-Thomas interpolation operators for each element $K \in \mathcal{T}$. It is well known (cf. [51]) that there also exist global interpolants $\Pi_{\text{grad}} v|_K \in H_0^1(\Omega)$, $\Pi_{\text{div}} \vec{p}|_K \in H(\text{div}, \Omega)$, and $\Pi\lambda|_K \in L^2(\Omega)$ such that $\Pi_{\text{grad}} v|_K \in P^{p+1}(K)$, $\Pi_{\text{div}} \vec{p}|_K \in \vec{x}P^p(K) + P^p(K)^d$, and $\Pi\lambda|_K \in P_p(K)$, for all $K \in \Omega$. Moreover, there exist constants C, depending only on the polynomial degree p, such that

$$\|v - \Pi_{\text{grad}} v\|_{H^1(\Omega)} \leq Ch^s |v|_{H^{1+s}(\Omega)}, \qquad (1/2 < s \leq p + 1), \tag{99a}$$

$$\|\vec{p} - \Pi_{\text{div}} \vec{p}\|_{H(\text{div},\Omega)} \leq Ch^s |\vec{p}|_{H^{1+s}(\Omega)}, \qquad (0 < s \leq p + 1), \tag{99b}$$

$$\|\lambda - \Pi\lambda\|_{L^2(\Omega)} \leq Ch^s |\lambda|_{H^s(\Omega)}, \qquad (0 < s \leq p + 1). \tag{99c}$$

Notice that $\|\vec{\lambda}-\vec{\lambda}_h\|_{\mathcal{U}}^2 = \|\lambda-\lambda_h\|_{L^2(\Omega)}^2 + \|\vec{\zeta}-\vec{\zeta}_h\|_{L^2(\Omega)}^2 + \|\widehat{\lambda}-\widehat{\lambda}_h\|_{H^{-1/2}(\partial\mathcal{T})} + \|\widehat{\zeta}-\widehat{\zeta}_h\|_{H^{1/2}(\partial\mathcal{T})}$. Consider first $\|\vec{\zeta}-\vec{\zeta}_h\|_{L^2(\Omega)} \leq h^s|\vec{\zeta}|_{H^s(\Omega)}$ and $\|\lambda-\lambda_h\|_{L^2(\Omega)} \leq h^s|\lambda|_{H^s(\Omega)}$ by (99c). Now, by Proposition 6.9, $|\vec{\zeta}|_{H^s(\Omega)} \leq |\vec{p}|_{H^s(\Omega)} + |\vec{r}|_{H^s(\Omega)}$. Similarly, however, also invoking the identity $f = -\Delta v$, we have $|\lambda|_{H^s(\Omega)} \leq 2|f|_{H^s(\Omega)} + |e|_{H^s(\Omega)} \leq C|v|_{H^{s+2}(\Omega)} + |e|_{H^s(\Omega)}$. Next, letting $\widehat{\sigma} = \operatorname{tr}_n \vec{\sigma}$ denote the normal trace of any $\vec{\sigma} \in H(\operatorname{div},\mathcal{T})$, $\|\widehat{\sigma}\|_{H^{1/2}(\partial\mathcal{T})} \leq \|\vec{\sigma}\|_{H(\operatorname{div},\mathcal{T})}$, by continuity. Therefore, by (128b), we see $\|\widehat{\zeta}-\widehat{\zeta}_h\|_{H^{1/2}(\partial\mathcal{T})} \leq Ch^s(2|\vec{p}|_{H^{1+s}(\Omega)} + |\vec{r}|_{H^{1+s}(\Omega)})$. Likewise, using the trace theorem and (99a), $\|\widehat{\lambda}-\widehat{\lambda}_h\|_{H^{-1/2}(\partial\mathcal{T})} \leq Ch^s|e|_{H^{1+s}(\Omega)}$. Recalling that $\vec{p} = \operatorname{grad} v$ and $\vec{r} = -\operatorname{grad} e$ completes the proof. □

The conclusion from Corollary 6.10 is that even if the solution v has high regularity throughout the entire domain and up to the boundary, the convergence rate of the DPG* method is also controlled by a pollution variable e, which may happen to be less regular. Indeed, by elliptic regularity [61], e, which satisfies $-\Delta e = v + 2f$, will be at least as regular as v in the interior of the domain, but may not be as regular up to the boundary.

To illustrate how to get higher order convergence rates using duality, we want to apply Theorem 6.8. To this end, we require sufficient regularity in the solution of the dual problem. Consider the case of full regularity, namely, for any $g \in L^2(\Omega)$, the solution $u \in H_0^1(\Omega)$ of the Dirichlet problem $-\Delta u = g$ satisfies

$$\|u\|_{H^2(\Omega)} \leq C\|g\|_{L^2(\Omega)}. \tag{100}$$

The inequality above is well known to hold on convex polygonal domains. In this case, we apply Theorem 6.8 with $F \in \mathcal{V}'$ defined

$$F((\vec{\tau},\nu)) = (v-v_h,\nu)_\Omega. \tag{101}$$

Note that F only sees the error in the first solution component of $\vec{v}_h = (v_h,\vec{p}_h)$ and that

$$\|F\|_{\mathcal{V}'} \leq \|v-v_h\|_{L^2(\Omega)}. \tag{102}$$

We now need to verify (94), so let us consider the present analog of (90a), with the functional F in (101):

$$\begin{cases} (\vec{\varepsilon},\vec{\nu})_{\mathcal{V}} + b_2(\vec{u},\vec{\nu}) = F(\vec{\nu}) & \forall\, \vec{\nu}\in\mathcal{V}, \\ b_2(\vec{\mu},\vec{\varepsilon}) = 0 & \forall\, \vec{\mu}\in\mathcal{U}. \end{cases} \tag{103}$$

First, observe that Theorem 4.1 implies $\mathcal{B}_2$ is a bijection, so $\vec{\varepsilon}=0$. Therefore the first line in (90a) reduces to $b_2(\vec{u},\vec{\nu}) = F(\vec{\nu})$ for all $\vec{\nu}\in\mathcal{V}$. This is an equation of the form (61b). Hence, the second item of Theorem 4.1 implies that $\mathcal{L}^*\vec{u} = (v-v_h,\vec{0})$; i.e., we may write $\vec{u}=(u,-\operatorname{grad} u)\in H_0(\mathcal{L}^*) = H^1_0(\Omega)\times H(\operatorname{div},\Omega)$ such that $u\in H^1_0(\Omega)$ satisfies $-\Delta u = v-v_h$. Now, due to the full regularity estimate (100) applied to u, we have $\|u\|_{H^2(\Omega)}\le C\|v-v_h\|_{L^2(\Omega)}$.

We may now invoke the complement of Corollary 6.10, [69, Theorem 6]:

Theorem 6.11. *Let $p\in\mathbb{N}_0$, let $\vec{u}$ be the solution to* (103) *for some arbitrary $F\in\mathcal{V}'$. Let $\vec{u}_h$ be the corresponding DPG solution. Then there exists a constant depending on p and the shape regularity of $\mathcal{T}$ such that*

$$\inf_{\mu\in\mathcal{U}_h}\|u-\mu\|_{\mathcal{U}} \le Ch^{p+1}\big(\|u\|_{H^{p+2}(\Omega)} + \|\vec{q}\|_{H^{p+1}(\mathcal{T})}\big).$$

Using the fact that $\vec{q}=-\operatorname{grad} u$, we now have the estimate

$$\inf_{\mu\in\mathcal{U}_h}\|u-\mu\|_{\mathcal{U}} \le Ch\|u\|_{H^2(\Omega)} \le Ch\|v-v_h\|_{L^2(\Omega)},$$

which is of the same form as (94). Finally, it is clear that assumption (94) holds with $c_1(h)=Ch$. Then, (102) and Theorem 6.8 imply

$$\|v-v_h\|^2_{L^2(\Omega)} = F(v-v_h) \le Ch\|v-v_h\|_{L^2(\Omega)}\left[\inf_{\nu\in\mathcal{V}_h}\|\vec{v}-\vec{\nu}\|^2_{\mathcal{V}} + \inf_{\vec{\mu}\in\mathcal{U}_h}\|\vec{\lambda}-\vec{\mu}\|^2_{\mathcal{U}}\right]^{1/2},$$

which provides one order of convergence higher in the L^2 norm for the solution component v_h.

Ultimately, the upshot of this analysis is that poor *a priori* convergence rates are possible with this method, even for infinitely smooth solutions v, due to the Lagrange multiplier λ, which may not be as smooth (cf. Section 9.1.2). Thus, we require an adaptive algorithm that helps one capture irregular solutions. This is one reason that we proceed in the next chapter to study *a posteriori* error control.

Chapter 7

A posteriori error control

This chapter presents contemporary *a posteriori* error estimation theory for both DPG and DPG* methods and a complementary discussion on adaptive mesh refinement (AMR) strategies. The new results found here on *a posteriori* error estimation with DPG methods are adapted from the author's contributions in [93] and build on the original theory presented in [27]. The *a posteriori* error estimation theory here for DPG* methods corresponds the same two-dimensional setting considered [48]. Notably, similar 3D results for DPG* methods can be found in [93], but are not documented here.

7.1 Abstract stability analysis

The simple fact supporting *a posteriori* error estimation in computing for engineering and science is that the discretization error present in all simulations must be carefully controlled in order to instill confidence in the computed results. All simulations have errors which affect the quality of the simulation. Some errors naturally arise from modeling assumptions—which may also be estimated *a posteriori* [111]—but those which arise from a finite element discretization are the concern of the analysis here.

The field of *a posteriori* error estimation in finite element methods began with the Ph.D. work of Ladevèze [96] and the research of Babuška et al. [7, 8]. These early ideas have prospered throughout the intervening decades and have been applied to a plethora of methods for problems of engineering interest. A review of the first few decades of work on *a posteriori* error estimation is given in [2].

The vast majority of early work on *a posteriori* error estimation was focused on global estimates of the solution error. After approximately two decades, a new optimal control-based

error estimation theory devoted to the error in functional outputs of finite element methods—so-called quantities of interest (QOI)—was developed by Oden and Prudhomme [110, 122], Becker and Rannacher [10], Patera and Peraire [119], and Giles and Süli [77]. The work here builds on many of the founding principles developed by these authors.

The DPG* method was discovered when studying *a posteriori* goal-oriented error estimation and adaptivity [93]. As both the original and dual problems are just special mixed methods, Brezzi's general theory applies in the analysis of their stability [15, 19]. However, these problems have special structure, so ultimately the general theory does not yield optimal estimates. Therefore, we begin this chapter with the following proposition.

Proposition 7.1. *Suppose $F \in \mathcal{V}'$ and $G \in \mathcal{U}'$, $v \in \mathcal{V}$ and $w \in \mathcal{U}$ solve* (13)*, and $\mathcal{B}$ is bounded below, as in* (14)*. Let F^0 and $F^\perp$ be the unique components of the decomposition of F in* (21)*. Then the following identities hold:*

$$\|v\|_{\mathcal{V}}^2 = \|F^0\|_{\mathcal{V}'}^2 + |||G|||_{\mathcal{U}'}^2, \tag{104}$$

$$|||w|||_{\mathcal{U}} \leq \|F^\perp\|_{\mathcal{V}'} + |||G|||_{\mathcal{U}'}, \tag{105}$$

If, in addition, $\nu \in \mathcal{V}$ and $\mu \in \mathcal{U}$ are arbitrary, then

$$\|v - \nu\|_{\mathcal{V}}^2 = \inf_{\mu \in \mathcal{U}} \|F - \mathcal{R}_{\mathcal{V}}\,\nu - \mathcal{B}\mu\|_{\mathcal{V}'}^2 + |||G - \mathcal{B}'\nu|||_{\mathcal{U}'}^2, \tag{106}$$

$$|||w - \mu|||_{\mathcal{U}} \leq \inf_{\nu_\perp \in (\operatorname{Null}\mathcal{B}')_\perp} \big(\|F^\perp - \mathcal{R}_{\mathcal{V}}\,\nu_\perp - \mathcal{B}\mu\|_{\mathcal{V}'} + |||G - \mathcal{B}'\nu_\perp|||_{\mathcal{U}'}\big). \tag{107}$$

If, in addition, $\mathcal{B}$ is an isomorphism, then

$$\|v - \nu\|_{\mathcal{V}} = |||G - \mathcal{B}'\nu|||_{\mathcal{U}'}\,. \tag{108}$$

Proof. Each of these results follow from Proposition 2.4. To arrive at (104), simply invoke (24) and (26):

$$\|v\|_{\mathcal{V}}^2 = \|v_0\|_{\mathcal{V}}^2 + \|v_\perp\|_{\mathcal{V}}^2 = \|F^0\|_{\mathcal{V}'}^2 + |||G|||_{\mathcal{U}'}^2.$$

To arrive at (105), recall only (25):

$$|||w|||_{\mathcal{U}} = \|F^\perp - \mathcal{R}_{\mathcal{V}} v_\perp\|_{\mathcal{V}'} \leq \|F^\perp\|_{\mathcal{V}'} + \|\mathcal{R}_{\mathcal{V}} v_\perp\|_{\mathcal{V}'}\,.$$

Taking into account $\|\mathcal{R}_{\mathcal{V}} v_\perp\|_{\mathcal{V}'} = \|v_\perp\|_{\mathcal{V}} = |||G|||_{\mathcal{U}'}$, the result follows.

Next, we use the standard identity $(\operatorname{Null} \mathcal{B}')^\perp = \operatorname{Range} \mathcal{B}$ to conclude that

$$\|F^0\|_{\mathcal{V}'} = \inf_{E^\perp \in (\operatorname{Null} \mathcal{B}')^\perp} \|F - E^\perp\|_{\mathcal{V}} = \inf_{\mu \in \mathcal{U}} \|F - \mathcal{B}\mu\|_{\mathcal{V}'}. \tag{109}$$

Equations (106) and (107) now readily follow from (104) and (105). Indeed, apply the saddle-point operator in (13) to the variables $(v - \nu, w - \mu) \in \mathcal{V} \times \mathcal{U}$:

$$\begin{cases} \mathcal{R}_{\mathcal{V}}(v - \nu) + \mathcal{B}(w - \mu) = F - \mathcal{R}_{\mathcal{V}}\, \nu - \mathcal{B}\mu\,, \\ \mathcal{B}'(v - \nu) \qquad\qquad\qquad\quad = G - \mathcal{B}'\nu\,. \end{cases}$$

This induces new problems of the same form as (13), but with new loads and new unique solutions $(v - \nu, w - \mu) \in \mathcal{V} \times \mathcal{U}$. Equation (106) is immediate from (105) and (109). To arrive at (107), invoke (105) noting that $(F - \mathcal{R}_{\mathcal{V}}\, \nu)^\perp = F^\perp - \mathcal{R}_{\mathcal{V}}\, \nu_\perp$:

$$|||w - \mu|||_{\mathcal{U}} \leq \|F^\perp - \mathcal{R}_{\mathcal{V}}\, \nu_\perp - \mathcal{B}\mu\|_{\mathcal{V}'} + |||G - \mathcal{B}'\nu|||_{\mathcal{U}'}.$$

Because $\mathcal{B}'\nu = \mathcal{B}'\nu_\perp$, taking the infimum over all $\nu_\perp \in (\operatorname{Null} \mathcal{B}')_\perp$ finishes the argument. Finally, (108) follows immediately from (106) since $\mathcal{R}_{\mathcal{V}} v_h \in \operatorname{Range} \mathcal{B}$ and so $\inf_{\mu \in \mathcal{U}} \|F - \mathcal{R}_{\mathcal{V}}\, v_h - \mathcal{B}\mu\|_{\mathcal{V}'} = 0$. □

7.2 The error in a quantity of interest

Recall Corollary 6.3. This result provided crude upper bounds on the error in a QOI, in terms of approximation errors, which may be used to predict its convergence rate. In this subsection, we provide a similar upper bound in terms of the residuals of the DPG and DPG* problems, which may be viewed as a motivating factor for the goal-oriented adaptive mesh refinement strategies introduced at the end of this chapter. The main result in this section, Corollary 7.3, follows immediately from Theorem 7.2. Note that the bound in (110) does not involve either of the auxiliary functions ε_h and λ_h from (54). This is an unusual feature when compared to analogous upper bounds for other mixed methods (cf. [109]).

Theorem 7.2. *Let $\mu \in \mathcal{U}$ and $\nu \in \mathcal{V}$ be arbitrary. Then the following upper bound holds:*

$$|b(u-\mu, v-\nu)| \leq \|\mathcal{B}^{-1}\| \|\mathcal{B}\mu - F\|_{\mathcal{V}'} \|\mathcal{B}'\nu - G\|_{\mathcal{U}'} .$$

Proof. Observe that $|b(u-\mu, v-\nu)| = |b(u-\mu, v-\nu-\nu_0)| \leq |||u-\mu|||_{\mathcal{U}} \|v-\nu-\nu_0\|_{\mathcal{V}}$ for all $\nu_0 \in \operatorname{Null} \mathcal{B}$. Therefore, by Proposition 3.2,

$$|b(u-\mu, v-\nu)| \leq \|\mathcal{B}\mu - F\|_{\mathcal{V}'} (\|\mathcal{P}(v-\nu-\nu_0)\|_{\mathcal{V}}^2 + |||\mathcal{B}'\nu - G|||_{\mathcal{U}'}^2)^{1/2} \quad \forall\, \nu_0 \in \operatorname{Null} \mathcal{B} .$$

Recall that $v \in (\operatorname{Null} \mathcal{B})_\perp$ by Proposition 2.4. Setting $\nu_0 = \mathcal{P}\nu \in \operatorname{Null} \mathcal{B}$ and observing that $|||\cdot|||_{\mathcal{U}'} \leq \gamma^{-1} \|\cdot\|_{\mathcal{U}'} = \|\mathcal{B}^{-1}\| \|\cdot\|_{\mathcal{U}'}$ delivers the required result. □

Corollary 7.3. *Let u_h be the $\mathcal{U}_h$-solution of* (54a) *and v_h be the $\mathcal{V}_h$-solution of* (54b). *Define $e_{\mathrm{QOI}} = b(u-u_h, v-v_h)$. Then the following crude upper bound holds:*

$$|e_{\mathrm{QOI}}| \leq \|\mathcal{B}^{-1}\| \|\mathcal{B}u_h - F\|_{\mathcal{V}'} \|\mathcal{B}'v_h - G\|_{\mathcal{U}'} . \tag{110}$$

7.3 Reliability and efficiency of a DPG error estimator

In this section, we consider a specific well-studied implicit estimator for the (energy norm) error

$$|||u-u_h|||_{\mathcal{U}} = \|\mathcal{B}u_h - F\|_{\mathcal{V}'} .$$

Namely, we consider $\eta(u_h)$, where

$$\eta(\mu) = \sup_{\nu \in \mathcal{V}_h} \frac{b(\mu,\nu) - F(\nu)}{\|\nu\|_{\mathcal{V}}} \quad \forall\, \mu \in \mathcal{U} . \tag{111}$$

From now on, we may denote the expression for this error estimator simply as $\eta(\mu) = \|\mathcal{B}\mu - F\|_{\mathcal{V}_h'}$.

The DPG error estimator (111) has been well-studied and analyzed in the literature. For expanded discussions on it, we refer the interested reader to [27, 50]. Before we present the main result of this section, we summarize the most important properties of $\eta(u)$.

Lemma 7.4. *Suppose that $\mathcal{B}$ is bounded below, as in* (14). *Let $\mathcal{V}_h$ be any closed subspace of $\mathcal{V}$. Define $\varepsilon_h \in \mathcal{V}_h$ to be the unique solution of $\langle \mathcal{R}_{\mathcal{V}} \varepsilon_h, \nu \rangle = \langle F - \mathcal{B}u_h, \nu \rangle$, for all $\nu \in \mathcal{V}_h$. Then*

$$\|\mathcal{B}(u-u_h)\|_{\mathcal{V}'}^2 = \|\mathcal{R}_{\mathcal{V}} \varepsilon_h + \mathcal{B}u_h - F\|_{\mathcal{V}'}^2 + \|\varepsilon_h\|_{\mathcal{V}}^2 .$$

Proof. Simply observe that

$$\begin{aligned}\|\mathcal{B}(u-u_h)\|_{\mathcal{V}'}^2 &= \|(\mathcal{R}_{\mathcal{V}}\,\varepsilon_h - \mathcal{R}_{\mathcal{V}}\,\varepsilon_h) + \mathcal{B}u_h - F\|_{\mathcal{V}'}^2 \\ &= \|\,\mathcal{R}_{\mathcal{V}}\,\varepsilon_h + \mathcal{B}u_h - F\|_{\mathcal{V}'}^2 - 2\langle \mathcal{R}_{\mathcal{V}}\,\varepsilon_h + \mathcal{B}u_h - F, \varepsilon_h\rangle_{\mathcal{V}} + \|\,\mathcal{R}_{\mathcal{V}}\,\varepsilon_h\|_{\mathcal{V}'}^2 \\ &= \|\,\mathcal{R}_{\mathcal{V}}\,\varepsilon_h + \mathcal{B}u_h - F\|_{\mathcal{V}'}^2 + \|\varepsilon_h\|_{\mathcal{V}}^2\,. \end{aligned}$$

□

Recall (88). The primary result in this section is an improvement on the main theorem in [27] in the case that the Fortin operator $\Pi_h : \mathcal{V} \to \mathcal{V}_h$ is a bounded projection. To prove the result, we will require Theorem 7.5 and Theorem 7.6. The first has become reasonably well-known in the literature and many different proofs for it are given in [135]. The second is perhaps much less known.

Theorem 7.5 (Complementary projections)**.** *Let Π be any bounded projection on a Hilbert space $\mathcal{W}$, $\Pi \circ \Pi = \Pi$. Then*

$$\|\Pi\| = \|1 - \Pi\|\,.$$

Theorem 7.6 (Pythagoras)**.** *Let $\mathcal{W}$ be a Hilbert space and $\mathcal{W}_0 \subseteq \mathcal{W}$ be a nontrivial closed subspace. Let $\mathcal{P} : \mathcal{W} \to \mathcal{W}_0$ be the orthogonal projection onto $\mathcal{W}_0$ and let $\Pi : \mathcal{W} \to \mathcal{W}_0$ be any other bounded projection onto $\mathcal{W}_0 = \Pi(\mathcal{W})$. Then*

$$\|\Pi - \mathcal{P}\|^2 + 1 = \|\Pi\|^2\,.$$

Proof. Throughout this proof, the subscript-$\mathcal{W}$ in norms and inner products is suppressed.

If $\Pi = \mathcal{P}$, we are done. Assume that $\Pi \neq \mathcal{P}$ and so $\big\{w \in \mathcal{W} : \|(\mathcal{P} - \Pi)w\| > 0\big\} \neq \varnothing$. Define $\mathcal{W}_\perp = (\mathcal{W}_0)_\perp$. By Corollary 2.3, observe that

$$\sup_{\widetilde{w}\in\mathcal{W}} \frac{(w, \Pi w)^2}{\|\widetilde{w}\|^2} = \sup_{w_0\in\mathcal{W}_0} \frac{(w, \Pi w_0)^2}{\|w_0\|^2} + \sup_{w_\perp\in\mathcal{W}_\perp} \frac{(w, \Pi w_\perp)^2}{\|w_\perp\|^2} \qquad \forall\, w \in \mathcal{W}\,. \tag{112}$$

Note that $\Pi w_0 = w_0$ and $\Pi w_\perp = \Pi(1-\mathcal{P})w_\perp = (\Pi - \mathcal{P})w_\perp$ for any $w_0 \in \mathcal{W}_0$ and $w_\perp \in \mathcal{W}_\perp$,

respectively. Therefore,

$$\begin{aligned}\|\Pi\|^2 = \sup_{w\in\mathcal{W}} \frac{\|\Pi w\|^2}{\|w\|^2} &= \sup_{\widetilde{w},w\in\mathcal{W}} \frac{(\widetilde{w},\Pi w)^2}{\|\widetilde{w}\|^2\|w\|^2} \\ &= \sup_{\widetilde{w}\in\mathcal{W}} \left(\sup_{w_0\in\mathcal{W}_0} \frac{(\widetilde{w},\Pi w_0)^2}{\|\widetilde{w}\|^2\|w_0\|^2} + \sup_{w_\perp\in\mathcal{W}_\perp} \frac{(\widetilde{w},\Pi w_\perp)^2}{\|\widetilde{w}\|^2\|w_\perp\|^2} \right) \\ &= \sup_{\widetilde{w}\in\mathcal{W}} \left(\sup_{w_0\in\mathcal{W}_0} \frac{(\widetilde{w},w_0)^2}{\|\widetilde{w}\|^2\|w_0\|^2} + \sup_{w_\perp\in\mathcal{W}_\perp} \frac{(\widetilde{w},(\Pi-\mathcal{P})w_\perp)^2}{\|\widetilde{w}\|^2\|w_\perp\|^2} \right), \end{aligned} \tag{113}$$

where the third equality follows from (112). Clearly,

$$\|\Pi\|^2 \le \sup_{\widetilde{w},w\in\mathcal{W}} \frac{(\widetilde{w},w)^2}{\|\widetilde{w}\|^2\|w\|^2} + \sup_{\widetilde{w},w\in\mathcal{W}} \frac{(\widetilde{w},(\Pi-\mathcal{P})w)^2}{\|\widetilde{w}\|^2\|w\|^2} = 1 + \|\Pi-\mathcal{P}\|^2 .$$

Now, define $w_\Pi = \arg\max_{\|w\|=1} \|(\Pi-\mathcal{P})w\| \neq 0$ and then $w_{\Pi,0} = (\Pi-\mathcal{P})w_\Pi \neq 0$. Consider $\widetilde{w} = w_{\Pi,0}$ in (113) and observe that

$$\|\Pi\|^2 \ge \sup_{w_0\in\mathcal{W}_0} \frac{(w_{\Pi,0},w_0)^2}{\|w_{\Pi,0}\|^2\|w_0\|^2} + \sup_{w_\perp\in\mathcal{W}_\perp} \frac{(w_{\Pi,0},(\Pi-\mathcal{P})w_\perp)^2}{\|w_{\Pi,0}\|^2\|w_\perp\|^2}.$$

Note that $w_{\Pi,0}\in\mathcal{W}_0$, $w_\Pi\in\mathcal{W}_\perp$, $\|w_{\Pi,0}\| = \|\Pi-\mathcal{P}\|$, and $\|w_\Pi\| = 1$. Finally, consider $w_0 = w_{\Pi,0}$ and $w_\perp = w_\Pi$ and observe that

$$\|\Pi\|^2 \ge \frac{(w_{\Pi,0},w_{\Pi,0})^2}{\|w_{\Pi,0}\|^4} + \frac{((\Pi-\mathcal{P})w_\Pi,(\Pi-\mathcal{P})w_\Pi)^2}{\|\Pi-\mathcal{P}\|^2} = 1 + \|\Pi-\mathcal{P}\|^2. \qquad \square$$

Theorem 7.7 (Improved *a posteriori* error estimates for DPG)**.** *Let* (14) *hold. Furthermore, let* Π_h *be a Fortin operator, as in* (88)*, where* $\Pi_h\circ\Pi_h = \Pi_h$*. Let* $F\in\operatorname{Range}\mathcal{B}$*,* $u = \mathcal{B}^{-1}F$*, and* $u_h\in\mathcal{U}_h$ *be arbitrary. Denote the best approximation of* u *in* $\mathcal{U}_h$ *as*

$$u_h^{\text{BAE}} = \arg\min_{\mu\in\mathcal{U}_h} \|u-\mu\|_{\mathcal{U}}.$$

Then the computable residual $\eta(\mu) = \|F-\mathcal{B}\mu\|_{\mathcal{V}_h'}$ *and the data approximation error* $\operatorname{osc}(F) = \|F\circ(1-\Pi_h)\|_{\mathcal{V}'}$ *satisfy*

$$\gamma^2\|u-u_h\|_{\mathcal{U}}^2 \le \eta(u_h)^2 + \left(\eta(u_h)\sqrt{\|\Pi_h\|^2-1} + \operatorname{osc}(F)\right)^2, \tag{114}$$

$$\eta(u_h) \le \|\mathcal{B}\|\|u-u_h\|_{\mathcal{U}}\,, \tag{115}$$

$$\operatorname{osc}(F) \le \|\mathcal{B}\|\|\Pi_h\|\|u-u_h^{\text{BAE}}\|_{\mathcal{U}}\,. \tag{116}$$

Replacing the trial norm $\|\cdot\|_{\mathcal{U}}$ with the energy norm $|||\cdot|||_{\mathcal{U}}$, we have the following complementary theorem. The proof of Theorem 7.8 is similar to the proof of Theorem 7.7, so it is left for the reader.

Theorem 7.8 (Practical a posteriori error estimates, in the energy norm, for DPG)**.** *Let* (14) *hold. Furthermore, let Π_h be a Fortin operator, as in* (88)*, where $\Pi_h \circ \Pi_h = \Pi_h$. Let $F \in \operatorname{Range} \mathcal{B}$, $u = \mathcal{B}^{-1}F$, and $u_h \in \mathcal{U}_h$ be arbitrary. Then the computable residual $\eta(\mu) = \|\mathcal{B}\mu - F\|_{\mathcal{V}_h'}$ and the data approximation error $\operatorname{osc}(F) = \|F \circ (1 - \Pi_h)\|_{\mathcal{V}'}$ satisfy*

$$
\begin{aligned}
|||u - u_h|||_{\mathcal{U}}^2 &\leq \eta(u_h)^2 + \left(\eta(u_h)\sqrt{\|\Pi_h\|^2 - 1} + \operatorname{osc}(F)\right)^2. \\
\eta(u_h) &\leq |||u - u_h|||_{\mathcal{U}}^2 \,, \\
\operatorname{osc}(F) &\leq \|\Pi_h\| |||u - u_h^{\mathrm{opt}}|||_{\mathcal{U}}^2 \,.
\end{aligned}
$$

Remark 7.9. The reliability bound in (114) is a new version of that found in [27], with the additional assumption that $\Pi_h \circ \Pi_h = \Pi_h$. Note that if $\operatorname{osc}(F) = 0$ then

$$
\gamma \|u - u_h\|_{\mathcal{U}} \leq |||u - u_h|||_{\mathcal{U}} \leq \|\Pi_h\| \eta(u_h) \,.
$$

Moreover, if $\|\Pi_h\| = 1$—that is, if Π_h is an orthogonal projection—then

$$
\gamma^2 \|u - u_h\|_{\mathcal{U}}^2 \leq |||u - u_h|||_{\mathcal{U}}^2 \leq \eta(u_h)^2 + \operatorname{osc}(F)^2.
$$

Remark 7.10. Reproducing the remarks in [27, Therorem 2.1], each of the inequalities in Theorem 7.7 can be demonstrated to be sharp.

Proof of Theorem 7.7. To arrive at (115), simply observe that

$$
\eta(u_h) = \|\mathcal{B}(u - u_h)\|_{\mathcal{V}_h'} \leq \|\mathcal{B}(u - u_h)\|_{\mathcal{V}'} \leq \|\mathcal{B}\| \|u - u_h\|_{\mathcal{U}} \,.
$$

To arrive at (116), first let $\nu \in \mathcal{V}$ where $\|\nu\|_{\mathcal{V}} = 1$ be arbitrary. Then, by (88) and Galerkin orthogonality,

$$
F(\nu - \Pi_h \nu) = b(u, \nu - \Pi_h \nu) = b(u - \mu, \nu - \Pi_h \nu) \leq \|\mathcal{B}\| \|1 - \Pi_h\| \|u - \mu\|_{\mathcal{U}} \,,
$$

for every $\mu \in \mathcal{U}_h$. The result then follows from Theorem 7.5: $\|1 - \Pi_h\| = \|\Pi_h\|$.

The reliability bound is much more subtle. Begin by defining $\widetilde{\mathcal{V}}_h = \text{Range}(\Pi_h) \subseteq \mathcal{V}_h$. Now, define $\widetilde{\varepsilon}_h \in \widetilde{\mathcal{V}}_h$ to be the unique solution of $\langle \mathcal{R}_{\mathcal{V}} \widetilde{\varepsilon}_h, \widetilde{\nu} \rangle_{\mathcal{V}} = \langle F - \mathcal{B}u_h, \widetilde{\nu} \rangle_{\mathcal{V}}$, for all $\widetilde{\nu} \in \widetilde{\mathcal{V}}_h$. Recall that $\gamma \|\mu\|_{\mathcal{U}} \leq \|\mathcal{B}\mu\|_{\mathcal{V}'}$, for all $\mu \in \mathcal{U}$. Therefore, because $\widetilde{\mathcal{V}}_h$ is a closed subspace of $\mathcal{V}$, note that

$$\gamma^2 \|u - u_h\|_{\mathcal{U}}^2 \leq \|\mathcal{B}(u - u_h)\|_{\mathcal{V}'}^2 = \| \mathcal{R}_{\mathcal{V}} \widetilde{\varepsilon}_h + \mathcal{B}u_h - F\|_{\mathcal{V}'}^2 + \|\widetilde{\varepsilon}_h\|_{\mathcal{V}}^2 \,, \tag{117}$$

by Lemma 7.4.

Define $\mathcal{P}_h : \mathcal{V} \to \widetilde{\mathcal{V}}_h$ to be the orthogonal projection onto the range of Π_h. Because $\widetilde{\varepsilon}_h \in \widetilde{\mathcal{V}}_h$, notice that

$$(\widetilde{\varepsilon}_h, \nu - \Pi_h \nu)_{\mathcal{V}} = (\widetilde{\varepsilon}_h, \mathcal{P}_h(\nu - \Pi_h \nu))_{\mathcal{V}} = (\widetilde{\varepsilon}_h, \mathcal{P}_h \nu - \Pi_h \nu)_{\mathcal{V}} \quad \forall\, \nu \in \mathcal{V} \,. \tag{118}$$

Now, observe that

$$\begin{aligned}
\| \mathcal{R}_{\mathcal{V}} \widetilde{\varepsilon}_h + \mathcal{B}u_h - F\|_{\mathcal{V}'} &= \sup_{\nu \in \mathcal{V}} \frac{(\widetilde{\varepsilon}_h, \nu)_{\mathcal{V}} + b(u_h, \nu) - F(\nu)}{\|\nu\|_{\mathcal{V}}} \\
&= \sup_{\nu \in \mathcal{V}} \frac{(\widetilde{\varepsilon}_h, \nu - \Pi_h \nu)_{\mathcal{V}} + b(u_h, \nu - \Pi_h \nu) - F(\nu - \Pi_h \nu)}{\|\nu\|_{\mathcal{V}}} \\
&= \sup_{\nu \in \mathcal{V}} \frac{(\widetilde{\varepsilon}_h, \mathcal{P}_h \nu - \Pi_h \nu)_{\mathcal{V}} - F(\nu - \Pi_h \nu)}{\|\nu\|_{\mathcal{V}}} \\
&\leq \|\widetilde{\varepsilon}_h\|_{\mathcal{V}} \|\mathcal{P}_h - \Pi_h\| + \|F \circ (1 - \Pi_h)\|_{\mathcal{V}'} \\
&= \|\widetilde{\varepsilon}_h\|_{\mathcal{V}} \sqrt{\|\Pi_h\|^2 - 1} + \text{osc}(F) \,. \qquad (119)
\end{aligned}$$

In the third line we have used (118) and (88), meanwhile, in the final line we have used Theorem 7.6. Finally, observe that

$$\|\widetilde{\varepsilon}_h\|_{\mathcal{V}} = \sup_{\widetilde{\nu} \in \widetilde{\mathcal{V}}_h} \frac{b(u_h, \widetilde{\nu}) - F(\widetilde{\nu})}{\|\widetilde{\nu}\|_{\mathcal{V}}} \leq \sup_{\nu \in \mathcal{V}_h} \frac{b(u_h, \nu) - F(\nu)}{\|\nu\|_{\mathcal{V}}} = \eta(u_h) \,.$$

With this observation in hand, (117) and (119) complete the proof. □

7.4 Reliability and efficiency of a DPG* error estimator

Recall the variational formulation derived in Section 5.1 for Poisson's equation. Using (74) and the trace operators in (79), this variational formulation can also be expressed as

$$b(\vec{\mu},\vec{\nu}) = (\vec{\mu}_\Omega, \mathcal{L}_h\vec{\nu})_\Omega + \langle\widehat{\mu},\vec{\tau}\cdot\vec{n}\rangle_h + \langle\widehat{\sigma},\nu\rangle_h\,, \tag{120}$$

where $\vec{\mu} = (\mu,\vec{\sigma},\widehat{\mu},\widehat{\sigma})$, $\vec{\mu}_\Omega = (\mu,\vec{\sigma})$, $\vec{\nu} = (\nu,\vec{\tau})$, and $\mathcal{L}_h(\nu,\vec{\tau}) = (-\operatorname{div}_h\vec{\tau},\ \vec{\tau} - \operatorname{grad}_h\nu)$. Many other physical models deliver well-posed broken variational formulations in a similar setting. Important examples include Stokes flow [126,129], linear elasticity [18,89], and acoustics [81,120].

Let Ω be a two-dimensional domain with Lipschitz continuous boundary. Define $\mathcal{E}_{\text{int}} \subseteq \mathcal{E}$ to be the set of all interior edges in $\mathcal{T}$ and define h_E to be the length of any edge $E \in \mathcal{T}$. Let an interior edge $E = \partial K^+ \cap \partial K^- \in \mathcal{E}_{\text{int}}$. Then define the following jump operations on any two-valued vector function $\vec{\tau}$:

$$[\![\vec{\tau}]\!]_E = \tau_{K^+}\cdot\vec{n}_{K^+} + \tau_{K^-}\cdot\vec{n}_{K^-}\,, \qquad [\![\vec{\tau}]\!]_E^\perp = \vec{n}_{K^+}^\perp\cdot\tau_{K^+} + \vec{n}_{K^-}^\perp\cdot\tau_{K^-}\,.$$

Here, $\vec{n}_K^\perp$ is the tangential unit vector; i.e., if $\vec{n}_K = (n_1,n_2)$ then $\vec{n}_K^\perp = (-n_2,n_1)$. If $E \in \mathcal{E}\setminus\mathcal{E}_{\text{int}}$ is an exterior edge on the boundary of an element K, then simply set

$$[\![\vec{\tau}]\!]_E = \tau_K\cdot\vec{n}_K\,, \qquad [\![\vec{\tau}]\!]_E^\perp = \vec{n}_K^\perp\cdot\tau_K\,.$$

Similarly, define the following jump operation on any two-valued *scalar* function ν:

$$[\![\nu]\!]_E = \nu_{K^+}\vec{n}_{K^+} + \nu_{K^-}\vec{n}_{K^-}\,,$$

if $E \in \mathcal{E}_{\text{int}}$, and $[\![\nu]\!]_E = \nu_K\vec{n}_K$ otherwise. For simplicity, we use $[\![\tau]\!]$ and $[\![\nu]\!]$ to denote the corresponding single-valued functions on the entire set of edges $\mathcal{E}$, which agrees with $[\![\tau]\!]_E$ and $[\![\nu]\!]_E$, respectively, for each $E \in \mathcal{E}$. Similarly define $[\![\tau]\!]^\perp$.

We continue with three crucial lemmas; the proof of the first can be found in [28, 45], while the proofs of the second and third are given at the end of this section. Here and throughout, define $\widetilde{P}^1(\partial\mathcal{T})$ to be the set of first-order, continuous piecewise polynomials on the mesh skeleton and define $\widetilde{P}_0^1(\partial\mathcal{T}) = \widetilde{P}^1(\partial\mathcal{T}) \cap \operatorname{tr}(H_0^1(\Omega))$.

Lemma 7.11. *For all $\vec{p} \in H(\mathrm{div}, \mathcal{T})$ and all $v \in H^1(\mathcal{T})$, the following identities hold:*

$$\sup_{\widehat{\mu} \in H_0^{1/2}(\partial\mathcal{T})} \frac{\langle \widehat{\mu}, \vec{p} \cdot \vec{n} \rangle_h}{\|\widehat{\mu}\|_{H^{1/2}(\partial\mathcal{T})}} = \sup_{\mu \in H_0^1(\Omega)} \frac{\langle \mathrm{tr}\, \mu, \vec{p} \cdot \vec{n} \rangle_h}{\|\mu\|_{H^1(\Omega)}},$$
$$\sup_{\widehat{\sigma} \in H^{-1/2}(\partial\mathcal{T})} \frac{\langle \widehat{\sigma}, v \rangle_h}{\|\widehat{\sigma}\|_{H^{-1/2}(\partial\mathcal{T})}} = \sup_{\vec{\sigma} \in H(\mathrm{div}, \Omega)} \frac{\langle \mathrm{tr}_n\, \vec{\sigma}, v \rangle_h}{\|\vec{\sigma}\|_{H(\mathrm{div}, \Omega)}}.$$

Lemma 7.12. *Fix $p \in \mathbb{N}$ and let $\vec{p}_h \in \prod_{K \in \mathcal{T}} (P^p(K))^d$. If $\langle \widehat{u}_h, \vec{p}_h \cdot \vec{n} \rangle_h = 0$ for all $\widehat{u}_h \in \widetilde{P}_0^1(\partial\mathcal{T})$ then*

$$\sup_{\mu \in H_0^1(\Omega)} \frac{\langle \mathrm{tr}\, \mu, \vec{p}_h \cdot \vec{n} \rangle_h^2}{\|\mu\|_{H^1(\Omega)}^2} \eqsim \sum_{E \in \mathcal{E}_{\mathrm{int}}} h_E \big\| [\![\vec{p}_h]\!]_E \big\|_{L^2(E)}^2 . \tag{121}$$

Lemma 7.13. *Fix $p \in \mathbb{N}$ and let $v_h \in \prod_{K \in \mathcal{T}} P^p(K)$. If $\int_E [\![v_h]\!]_E = 0$ for all $E \in \mathcal{E}$ then*

$$\sup_{\vec{\sigma} \in H(\mathrm{div}, \Omega)} \frac{\langle \mathrm{tr}_n(\vec{\sigma}), v_h \rangle_h^2}{\|\vec{\sigma}\|_{H(\mathrm{div}, \Omega)}^2} \eqsim \sum_{E \in \mathcal{E}} h_E \big\| [\![v_h]\!]_E \big\|_{H^1(E)}^2 \eqsim \sum_{E \in \mathcal{E}} h_E^{-1} \big\| [\![v_h]\!]_E \big\|_{L^2(E)}^2 . \tag{122}$$

Theorem 7.14. *Assume that $\Omega \subseteq \mathbb{R}^2$ is a bounded Lipschitz domain and let $\mathcal{T}$ be a regular subdivision of Ω into triangles. Finally, for all $\vec{\mu} = (\vec{\mu}_\Omega, \widehat{\mu}, \widehat{\sigma}) \in \mathcal{U}$ and $\vec{\nu} = (\nu, \vec{\tau}) \in \mathcal{V}$, where $\mathcal{U} = L^2(\Omega)^{d+1} \times H_0^{1/2}(\partial\mathcal{T}) \times H^{-1/2}(\partial\mathcal{T})$ and $\mathcal{V} = H^1(\mathcal{T}) \times H(\mathrm{div}, \mathcal{T})$, let*

$$b(\vec{\mu}, \vec{\nu}) = (\vec{\mu}_\Omega, \mathcal{L}_h \vec{\nu})_\Omega + \langle \widehat{\mu}, \boldsymbol{\tau} \cdot \vec{n} \rangle_h + \langle \widehat{\sigma}, \nu \rangle_h \qquad \text{and} \qquad G(\vec{\mu}) = (f, \vec{\mu}_\Omega)_\Omega .$$

Assume that $\mathcal{L}_h : \mathcal{V} \to (L^2(\Omega)))^{d+1}$ and $f \in (L^2(\Omega)))^{d+1}$. Let $Q \subseteq \prod_{K \in \mathcal{T}} (P^p(K))^{d+1}$ be a finite-dimensional polynomial subspace of L^2 and define $\mathcal{U}_h = \{(\vec{\mu}_\Omega, \widehat{\sigma}, \widehat{\mu}) \in \mathcal{U} : \vec{\mu}_\Omega \in Q,\, \widehat{\sigma}|_{\partial K} \in P^p(\partial K),\, \widehat{\mu}|_{\partial K} \in \widetilde{P}^{p+1}(\partial K),\, \forall K \in \mathcal{T}\}$. Then if $\mathcal{V}_h$ is a finite-dimensional polynomial subspace of $\mathcal{V}$ and $\vec{v}_h = (\vec{p}_h, v_h) \in \mathcal{V}_h$, satisfies

$$b(\vec{\mu}, \vec{v}_h) = G(\vec{\mu}), \quad \forall\, \vec{\mu} \in \mathcal{U}_h ,$$

there exist constants $C_2 > C_1 > 0$, independent of the maximum h_E, such that

$$C_1 \eta_i^*(\vec{v}_h) \le \|\vec{v} - \vec{v}_h\|_{\mathcal{V}} \le C_2 \eta_i^*(\vec{v}_h), \tag{123}$$

where $i \in \{1, 2\}$,

$$\eta_1^*(\vec{v}_h)^2 = \big\| \mathcal{L}\vec{v}_h - f \big\|_\Omega^2 + \sum_{E \in \mathcal{E}_{\mathrm{int}}} h_E \big\| [\![\vec{p}_h]\!]_E \big\|_{L^2(E)}^2 + \sum_{E \in \mathcal{E}} h_E \big\| [\![v_h]\!]_E \big\|_{H^1(E)}^2 , \tag{124a}$$

and

$$\eta_2^*(\vec{v}_h)^2 = \left\|\mathcal{L}\vec{v}_h - f\right\|_\Omega^2 + \sum_{E\in\mathcal{E}_{\rm int}} h_E \left\|[\![\vec{p}_h]\!]_E\right\|_{L^2(E)}^2 + \sum_{E\in\mathcal{E}} h_E^{-1}\left\|[\![v_h]\!]_E\right\|_{L^2(E)}^2 . \tag{124b}$$

Proof of Theorem 7.14. Define $(\mathcal{B}\vec{\mu})(\cdot) = b(\vec{\mu},\cdot)$ for all $\vec{\mu}\in\mathcal{U}$. Recall that $\gamma\|\vec{\mu}\|_{\mathcal{U}} \le \|\mathcal{B}\vec{\mu}\|_{\mathcal{V}'} \le M\|\vec{\mu}\|_{\mathcal{U}}$, where $M>0$ and $\gamma>0$ are the continuity constant and inf–sup stability constant for the bilinear form b, respectively. Therefore, by (108),

$$M^{-1}\sup_{\vec{\mu}\in\mathcal{U}}\frac{b(\vec{\mu},\vec{v}_h)-G(\vec{\mu})}{\|\vec{\mu}\|_{\mathcal{U}}} \le \|\vec{v}-\vec{v}_h\|_{\mathcal{V}} \le \gamma^{-1}\sup_{\vec{\mu}\in\mathcal{U}}\frac{b(\vec{\mu},\vec{v}_h)-G(\vec{\mu})}{\|\vec{\mu}\|_{\mathcal{U}}} .$$

It is readily observed that $\sup_{u\in L^2(\Omega)}\frac{|(u,\mathcal{L}\vec{v}_h-f)_\Omega|}{\|u\|_\Omega} = \left\|\mathcal{L}\vec{v}_h-f\right\|_\Omega$. Therefore,

$$\sup_{\vec{\mu}\in\mathcal{U}}\frac{(b(\vec{\mu},\vec{v}_h)-G(\vec{\mu}))^2}{\|\vec{\mu}\|_{\mathcal{U}}^2} = \left\|\mathcal{L}\vec{v}_h-f\right\|_\Omega^2 + \sup_{\widehat{u}\in H_0^{1/2}(\partial\mathcal{T})}\frac{\langle\widehat{u},\vec{p}_h\cdot\vec{n}\rangle_h^2}{\|\widehat{u}\|_{H^{1/2}(\partial\mathcal{T})}^2} + \sup_{\widehat{\sigma}\in H^{-1/2}(\partial\mathcal{T})}\frac{\langle\widehat{\sigma},v_h\rangle_h^2}{\|\widehat{\sigma}\|_{H^{-1/2}(\partial\mathcal{T})}^2},$$

and the bounds in (123) follow immediately from Lemmas 7.11–7.13. □

7.4.1 Proofs of Lemmas 7.12 and 7.13

Definition 7.15. For all $\widehat{w}\in H^{1/2}(\partial\mathcal{T})$, there exists a H^1-norm minimum energy extension, $\mathscr{E}(\widehat{w})\in H^1(\Omega)$:

$$\mathscr{E}(\widehat{w})\big|_K = \operatorname*{arg\,min}_{\substack{w\in H^1(K)\\ \mathrm{tr}^K(w)=\widehat{w}|_{\partial K}}} \|w\|_{H^1(K)}, \quad \forall\, K\in\mathcal{T}.$$

Similarly, for all $\widehat{w}\in H^{-1/2}(\partial\mathcal{T})$ there exists an $H(\mathrm{div})$-norm minimum energy extension, $\mathscr{E}_n(\widehat{w})\in H(\mathrm{div},\Omega)$:

$$\mathscr{E}_n(\widehat{w})\big|_K = \operatorname*{arg\,min}_{\substack{\vec{w}\in H(\mathrm{div},K)\\ \mathrm{tr}_n^K(\vec{w})=\widehat{w}|_{\partial K}}} \|\vec{w}\|_{H(\mathrm{div},K)}, \quad \forall\, K\in\mathcal{T}.$$

It is known that both $\mathscr{E}: H^{1/2}(\partial\mathcal{T})\to H^1(\Omega)$ and $\mathscr{E}_n: H^{-1/2}(\partial\mathcal{T})\to H(\mathrm{div},\Omega)$ are bounded linear operators.

For any fixed edge $E\in\mathcal{E}$, consider a single-valued function on the mesh skeleton $\widehat{w}_E\in H^{1/2}(\partial\mathcal{T})$ such that $\widehat{w}_E|_{E'}=0$ for all edges $E'\neq E$. Let Ω_E denote the patch of elements

associated with E, i.e., $\Omega_E = \bigcup\{K \in \mathcal{T} : \operatorname{meas}(\partial K \cap E) \neq \emptyset\}$. Observe that $\mathscr{E}(\widehat{w}_E)|_{K'} = 0$ for all elements $K' \not\subseteq \Omega_E$ since $\operatorname{tr}^{K'} \widehat{w}_E = 0$. Similarly, for any function $\widehat{w}_E \in H^{-1/2}(\partial\mathcal{T})$ such that $\widehat{w}_E|_{E'} = 0$ for all edges $E' \neq E$, $\mathscr{E}_n(\widehat{w}_E)|_{K'} = \vec{0}$ for all elements $K' \not\subseteq \Omega_E$.

For a concrete example, consider $\widehat{b}_E \in C^0(\partial\mathcal{T})$ such that $\widehat{b}_E|_{E'} = 0$ for all edges $E' \neq E$ and, specifically, $\widehat{b}_E|_E$ is parameterized by the positive parabola $\gamma_E(s) = 4s(1-s)$, where s is the relative distance from one vertex of E to the next. In this case, we may construct a so-called H^1 *edge bubble function* $b_E = \mathscr{E}(\widehat{b}_E)$. Likewise, we may construct an $H(\operatorname{div})$ *edge bubble function* $\vec{b}_E = \mathscr{E}_n(\widehat{b}_E)$. Notice that both b_E and $\vec{b}_E$ vanish outside Ω_E.

For the function $\widehat{b}_E$ given above, we also have the following crucial lemma which follows from the equivalence of all norms on finite-dimensional vector spaces and a standard scaling argument.

Theorem 7.16 (Verfürth [140])**.** *Let $E \in \mathcal{E}$ be an arbitrary edge and consider the edge bubble function b_E associated with E. Then, for any function $\widehat{w}|_E \in P^p(E)$ and $\widehat{w}|_{E'} = \{0\}$ for all edges $E' \neq E$,*

$$\|\widehat{b}_E\widehat{w}\|_{L^2(E)} \leq \|\widehat{w}\|_{L^2(E)} \lesssim \big\|\widehat{b}_E^{1/2}\widehat{w}\big\|_{L^2(E)}\,, \tag{125a}$$

$$|\mathscr{E}(\widehat{b}_E\widehat{w})|_{H^1(\Omega_E)} \lesssim h_E^{-1/2}\|\widehat{w}\|_{L^2(E)}\,. \tag{125b}$$

Remark 7.17. Note that a standard scaling argument also shows that

$$h_E^2\|\mathscr{E}_n(w)\|^2_{H(\operatorname{div},\Omega_E)} \lesssim \|\mathscr{E}_n(w)\|^2_{L^2(\Omega_E)} + h_E^2\|\operatorname{div}\mathscr{E}_n(w)\|^2_{L^2(\Omega_E)} \lesssim h_E\|w\|^2_{L^2(E)}\,. \tag{125c}$$

Assume that $\mathcal{T}$ is constrained such that each $K \in \mathcal{T}$ has at most a fixed number of edges, say N_{e}, and each edge $E \in \mathcal{E}$ has at most a fixed number of hanging nodes, say N_{n}, then

$$\|\mu\|_{H^1(\Omega)} \leq \sum_{E\in\mathcal{E}_{\mathrm{int}}} \|\mu\|_{H^1(\Omega_E)} \leq N_{\mathrm{e}}N_{\mathrm{n}}\|\mu\|_{H^1(\Omega)}\,, \quad \forall\,\mu \in H^1_0(\Omega)\,. \tag{126}$$

Similarly, $\|\vec{\sigma}\|_{H(\operatorname{div},\Omega)} \eqsim \sum_{E\in\mathcal{E}} \|\vec{\sigma}\|_{H(\operatorname{div},\Omega_E)}$ for all $\vec{\sigma} \in H(\operatorname{div},\Omega)$.

Slightly more subtle, however, is the inequality in (127a) which requires the following space of edge bubble functions:

$$H^{1/2}_{\mathrm{bubb.}}(\mathcal{E}_{\mathrm{int}}) = \prod_{E\in\mathcal{E}_{\mathrm{int}}} \{\widehat{w}_E \in H^{1/2}_0(\partial\mathcal{T}) : \widehat{w}_E|_{E'} = 0\ \forall E' \neq E\}\,.$$

With this definition, let $\{\widehat{w}_E\}_{E\in\mathcal{E}_{\mathrm{int}}} \in H^{1/2}_{\mathrm{bubb.}}(\mathcal{E}_{\mathrm{int}})$ be arbitrary and define $\widehat{w} = \sum_{E\in\mathcal{E}_{\mathrm{int}}} \widehat{w}_E \in H^{1/2}_0(\partial\mathcal{T})$. Then

$$\|\mathscr{E}(\widehat{w})\|^2_{H^1(\Omega)} \lesssim \sum_{E\in\mathcal{E}_{\mathrm{int}}} \|\mathscr{E}(\widehat{w}_E)\|^2_{H^1(\Omega_E)}\,. \tag{127a}$$

This can be seen beginning from the linearity of $\mathscr{E}$:

$$\begin{aligned}
\|\mathscr{E}(\widehat{w})\|^2_{H^1(\Omega)} = \Big\| \sum_{E\in\mathcal{E}_{\mathrm{int}}} \mathscr{E}(\widehat{w}_E)\Big\|^2_{H^1(\Omega)} &= \sum_{K\in\mathcal{T}} \Big\| \sum_{E\in\mathcal{E}_{\mathrm{int}}} \mathscr{E}(\widehat{w}_E)\Big\|^2_{H^1(K)} \\
&= \sum_{K\in\mathcal{T}} \Big\| \sum_{\substack{E\in\mathcal{E}_{\mathrm{int}}\\ \mathrm{meas}(E\cap\partial K)\neq 0}} \mathscr{E}(\widehat{w}_E)\Big\|^2_{H^1(K)} \\
&\le \sum_{K\in\mathcal{T}} N_e N_n \sum_{\substack{E\in\mathcal{E}_{\mathrm{int}}\\ \mathrm{meas}(E\cap\partial K)\neq 0}} \|\mathscr{E}(\widehat{w}_E)\|^2_{H^1(K)} \\
&= N_e N_n \sum_{E\in\mathcal{E}_{\mathrm{int}}} \|\mathscr{E}(\widehat{w}_E)\|^2_{H^1(\Omega)} \\
&= N_e N_n \sum_{E\in\mathcal{E}_{\mathrm{int}}} \|\mathscr{E}(\widehat{w}_E)\|^2_{H^1(\Omega_E)}\,.
\end{aligned}$$

For arbitrary $\widehat{w} \in H^{-1/2}(\partial\mathcal{T})$, define the zero extension of $\widehat{w}|_E$ throughout the whole mesh skeleton, $\widehat{w}_E$ by $\widehat{w}_E|_E = \widehat{w}|_E$ and $\widehat{w}_E|_{E'} = 0$ for all edges $E' \neq E$. When decomposing $\widehat{w} = \sum_{E\in\mathcal{E}} \widehat{w}_E$ in this way, it likewise holds that

$$\|\mathscr{E}_n(\widehat{w})\|^2_{H(\mathrm{div},\Omega)} \lesssim \sum_{E\in\mathcal{E}} \|\mathscr{E}_n(\widehat{w}_E)\|^2_{H(\mathrm{div},\Omega_E)}\,, \quad \forall\, \widehat{w} \in H^{-1/2}(\partial\mathcal{T})\,. \tag{127b}$$

Recall two results from the literature. Theorem 7.18 is a special case of [54, Lemma 5] for $n = 2$ and $k = 1$. Theorem 7.19 is a generalization of the Clément interpolant [38] taken from [54], which builds upon ideas originally established in the $\mathbb{R}^3$–$H(\mathrm{curl})$ setting by Schöberl [132]. We remark that the local bounded commuting cochain projectors developed in [62] also appear appropriate to establish the bounds stated in Theorem 7.19.

Theorem 7.18. *Given any $\vec{\sigma} \in H(\mathrm{div},\Omega)$, there exist $\varphi \in H^1(\Omega)$ and $\vec{\psi} \in H^1(\Omega)$ such that $\vec{\sigma} = \mathrm{curl}(\varphi) + \vec{\psi}$ and*

$$\|\varphi\|_{H^1(\Omega)} + \|\vec{\psi}\|_{H^1(\Omega)} \lesssim \|\vec{\sigma}\|_{H(\mathrm{div},\Omega)}\,.$$

In this situation, $\vec{\sigma} = \mathrm{curl}(\varphi) + \vec{\psi}$ is called a regular decomposition *of $\vec{\sigma}$.*

Theorem 7.19 (Demlow, Hirani, Schöberl [54,132]). *Let $E \in \mathcal{E}$, $\mu \in H^1(\Omega)$, and $\vec{\sigma} \in H(\mathrm{div}, \Omega)$ be arbitrary. Let h_E denote the length of the edge E and let $\varphi \in H^1(\Omega)$ and $\vec{\psi} \in H^1(\Omega)$ belong to a regular decomposition $\vec{\sigma} = \mathrm{curl}(\varphi) + \vec{\psi}$. There exist commuting quasi-interpolation operators $\mathcal{I} : H^1(\Omega) \to \widetilde{P}^1(\mathcal{T})$ and $\vec{\mathcal{I}} : H(\mathrm{div}, \Omega) \to \mathcal{RT}^0(\mathcal{T})$, such that $\mathrm{curl} \circ \mathcal{I} = \vec{\mathcal{I}} \circ \mathrm{curl}$ and the following inequalities hold:*

$$\|\mu - \mathcal{I}\mu\|^2_{L^2(E)} \lesssim h_E \|\mu\|^2_{H^1(\Omega_E)}\,, \tag{128a}$$

$$\|\varphi - \mathcal{I}\varphi\|^2_{L^2(E)} + \|(\vec{\psi} - \vec{\mathcal{I}}\vec{\psi}) \cdot \vec{n}\|^2_{L^2(E)} \lesssim h_E \|\vec{\sigma}\|^2_{H(\mathrm{div}, \Omega_E)}\,. \tag{128b}$$

Each of the inequalities above also hold with $H^1(\Omega)$ replaced by $H^1_0(\Omega)$ and $H(\mathrm{div}, \Omega)$ replaced by $H_0(\mathrm{div}, \Omega)$. In this case, $\mathcal{I} : H^1_0(\Omega) \to \widetilde{P}^1_0(\mathcal{T})$ and $\vec{\mathcal{I}} : H_0(\mathrm{div}, \Omega) \to \mathcal{RT}^0_0(\mathcal{T})$.

Proof of Lemma 7.12. For smooth enough functions $\widehat{w}$ defined on the mesh skeleton,

$$\langle \widehat{w}, \vec{p}_h \cdot \vec{n} \rangle_h = \sum_{K \in \mathcal{T}} (\widehat{w}, \vec{p}_h \cdot \vec{n}_K)_{\partial K} = \sum_{E \in \mathcal{E}_{\mathrm{int}}} (\widehat{w}, [\![\vec{p}_h]\!])_E\,. \tag{129}$$

Let $\mu \in H^1_0(\Omega)$ be arbitrary and set $\widehat{w} = \mathrm{tr}(\mu - \mathcal{I}\mu) \in H^{1/2}(\partial\mathcal{T})$. Because $\langle \mathrm{tr}\, \mathcal{I}\mu, \vec{p}_h \cdot \vec{n} \rangle_h = 0$ and $H^{1/2}(\partial\mathcal{T}) \subsetneq L^2(\partial\mathcal{T})$, it follows that $\langle \mathrm{tr}\, \mu, \vec{p}_h \cdot \vec{n} \rangle_h = \langle \widehat{w}, \vec{p}_h \cdot \vec{n} \rangle_h = \sum_{E \in \mathcal{E}_{\mathrm{int}}} (\widehat{w}, [\![\vec{p}_h]\!])_E$. Clearly, by (128a),

$$(\widehat{w}, [\![\vec{p}_h]\!])_E \le \|\mu - \mathcal{I}\mu\|_{L^2(E)} \|[\![\vec{p}_h]\!]\|_{L^2(E)} \lesssim h_E^{1/2} \|\mu\|_{H^1(\Omega_E)} \|[\![\vec{p}_h]\!]\|_{L^2(E)}\,,$$

for each edge $E \in \mathcal{E}$. Therefore,

$$\langle \mathrm{tr}\, \mu, \vec{p}_h \cdot \vec{n} \rangle_h \lesssim \Big(\sum_{E \in \mathcal{E}_{\mathrm{int}}} \|\mu\|^2_{H^1(\Omega_E)} \Big)^{1/2} \Big(\sum_{E \in \mathcal{E}_{\mathrm{int}}} h_E \|[\![\vec{p}_h]\!]\|^2_{L^2(E)} \Big)^{1/2}\,.$$

Then, by (126), we see that

$$\sup_{u \in H^1_0(\Omega)} \frac{\langle \mathrm{tr}\, \mu, \vec{p}_h \cdot \vec{n} \rangle_h^2}{\|\mu\|^2_{H^1(\Omega)}} \lesssim \sum_{E \in \mathcal{E}_{\mathrm{int}}} h_E \|[\![\vec{p}_h]\!]\|^2_{L^2(E)}\,.$$

To arrive at the efficiency estimate (i.e., the lower bound), recall (125a) and observe that

$$\big\|[\![\vec{p}_h]\!]\big\|^2_{L^2(E)} \lesssim \int_E [\![\vec{p}_h]\!]^2 \widehat{b}_E = ([\![\vec{p}_h]\!] \widehat{b}_E, [\![\vec{p}_h]\!])_E\,.$$

Let $E \in \mathcal{E}_{\rm int}$ be arbitrary and assume that $[\![\vec{p}_h]\!]_E \neq 0$. Then, invoking (125b), notice that

$$h_E^{1/2} \big\| [\![\vec{p}_h]\!] \big\|_{L^2(E)} \lesssim \frac{([\![\vec{p}_h]\!]\widehat{b}_E, [\![\vec{p}_h]\!])_E}{|\mathscr{E}([\![\vec{p}_h]\!]\widehat{b}_E)|_{H^1(\Omega_E)}} \leq \sup_{\widehat{\mu}_E \in C_0^\infty(E)} \frac{(\widehat{\mu}_E, [\![\vec{p}_h]\!])_E}{|\mathscr{E}(\widehat{\mu}_E)|_{H^1(\Omega_E)}} \,. \tag{130}$$

Equation (130) shows that the scaled jump contribution for each edge, $h_E^{1/2} \big\| [\![\vec{p}_h]\!] \big\|_{L^2(E)}$, is bounded by a local version of the supremum in (121). It is now necessary to accumulate every such jump contribution, for each edge $E \in \mathcal{E}_{\rm int}$, and demonstrate that it is bounded by the entire supremum in (130). This is accomplished as follows:

$$\begin{aligned}
\sum_{E \in \mathcal{E}_{\rm int}} h_E \big\| [\![\vec{p}_h]\!] \big\|_{L^2(E)}^2 &\lesssim \sum_{E \in \mathcal{E}_{\rm int}} \sup_{\widehat{\mu}_E \in C_0^\infty(E)} \frac{(\widehat{\mu}_E, [\![\vec{p}_h]\!])_E^2}{|\mathscr{E}(\widehat{\mu}_E)|_{H^1(\Omega_E)}^2} \\
&= \sup_{\widehat{\mu} \in \prod_{E \in \mathcal{E}_{\rm int}} C_0^\infty(E)} \frac{\big(\sum_{E \in \mathcal{E}_{\rm int}} (\widehat{\mu}_E, [\![\vec{p}_h]\!])_E \big)^2}{\sum_{E \in \mathcal{E}_{\rm int}} |\mathscr{E}(\widehat{\mu}_E)|_{H^1(\Omega_E)}^2} && \text{by Corollary 2.3} \\
&= \sup_{\widehat{\mu} \in \prod_{E \in \mathcal{E}_{\rm int}} C_0^\infty(E)} \frac{\langle \widehat{\mu}, \vec{p}_h \cdot \vec{n} \rangle_h^2}{\sum_{E \in \mathcal{E}_{\rm int}} |\mathscr{E}(\widehat{\mu}_E)|_{H^1(\Omega_E)}^2} && \text{by (129)} \\
&\lesssim \sup_{\widehat{\mu} \in \prod_{E \in \mathcal{E}_{\rm int}} C_0^\infty(E)} \frac{\langle \widehat{\mu}, \vec{p}_h \cdot \vec{n} \rangle_h^2}{|\mathscr{E}(\widehat{\mu})|_{H^1(\Omega)}^2} && \text{by (127a)} \\
&\leq \sup_{u \in H_0^1(\Omega)} \frac{\langle \operatorname{tr} u, \vec{p}_h \cdot \vec{n} \rangle_h^2}{|u|_{H^1(\Omega)}^2} \lesssim \sup_{u \in H_0^1(\Omega)} \frac{\langle \operatorname{tr} u, \vec{p}_h \cdot \vec{n} \rangle_h^2}{\|u\|_{H^1(\Omega)}^2} \,. && \square
\end{aligned}$$

Proof of Lemma 7.13. Let us begin with the observation that $\big| [\![v_h]\!] \big|_{H^1(E)} \eqsim h_E^{-1} \big\| [\![v_h]\!] \big\|_{L^2(E)}$ for each edge $E \in \mathcal{E}$. This follows from the Poincarè inequality, an inverse inequality, and a scaling argument. Therefore, $h_E \big\| [\![v_h]\!] \big\|_{H^1(E)}^2 \eqsim h_E^{-1} \big\| [\![v_h]\!] \big\|_{L^2(E)}^2$ and it only remains to prove the first relation in (122).

Let $\vec{\sigma} \in H(\operatorname{div}, \Omega)$ and $\vec{\sigma}_h^1, \vec{\sigma}_h^2 \in \mathcal{RT}^0(\mathcal{T})$ be arbitrary. Let $\varphi \in H^1(\Omega)$ and $\vec{\psi} \in (H^1(\Omega))^d$, where $\vec{\sigma} = \operatorname{curl}(\varphi) + \vec{\psi}$ constitute a regular decomposition of $\vec{\sigma}$ and notice that for each $i = 1, 2$, $\operatorname{tr}_n(\vec{\sigma}_h^i)|_E = \text{const}$. Therefore, because $\int_E [\![v_h]\!]_E = 0$ for all $E \in \mathcal{E}$,

$$\langle \operatorname{tr}_n(\vec{\sigma}), v_h \rangle_h = \langle \operatorname{tr}_n(\vec{\sigma} - \vec{\sigma}_h^1 - \vec{\sigma}_h^2), v_h \rangle_h = \langle \operatorname{tr}_n(\operatorname{curl} \varphi - \vec{\sigma}_h^1), v_h \rangle_h + \langle \operatorname{tr}_n(\vec{\psi} - \vec{\sigma}_h^2), v_h \rangle_h \,.$$

Take $\vec{\sigma}_h^1 = \vec{\mathcal{I}} \operatorname{curl} \varphi$ and observe that

$$\langle \operatorname{tr}_n(\operatorname{curl} \varphi - \vec{\sigma}_h^1), v_h \rangle_h = \sum_{K \in \mathcal{T}} ((\operatorname{curl} \varphi - \vec{\mathcal{I}} \operatorname{curl} \varphi) \cdot \vec{n}_K, v_h)_{\partial K} = \sum_{K \in \mathcal{T}} (\operatorname{curl}(\varphi - \mathcal{I}\varphi) \cdot \vec{n}_K, v_h)_{\partial K} \,.$$

Moreover,

$$\begin{aligned}(\operatorname{curl}(\varphi-\mathcal{I}\varphi)\cdot\vec{n}_K, v_h)_{\partial K} &= (\operatorname{div}\operatorname{curl}(\varphi-\mathcal{I}\varphi), v_h)_K + (\operatorname{curl}(\varphi-\mathcal{I}\varphi), \operatorname{grad} v_h)_K \\ &= (\operatorname{curl}(\varphi-\mathcal{I}\varphi), \operatorname{grad} v_h)_K \\ &= (\varphi-\mathcal{I}\varphi, \vec{n}_K^{\perp}\cdot\operatorname{grad} v_h)_{\partial K}\,.\end{aligned}$$

Therefore, observe that

$$\langle \operatorname{tr}_n(\operatorname{curl}\varphi), v_h\rangle_h = \sum_{K\in\mathcal{T}} (\varphi-\mathcal{I}\varphi, \vec{n}_K^{\perp}\cdot\operatorname{grad} v_h)_{\partial K} = \sum_{E\in\mathcal{E}}\int_E (\varphi-\mathcal{I}\varphi)[\![\operatorname{grad} v_h]\!]^{\perp}\,.$$

Notice that on each edge E, $[\![\operatorname{grad} v_h]\!]^{\perp}$ is the tangential derivative of the jump in v_h so we may identify $\|[\![\operatorname{grad} v_h]\!]^{\perp}\|_{L^2(E)}$ with $|[\![v_h]\!]|_{H^1(E)}$. Then, isolating each term on the right-hand side above,

$$\int_E (\varphi-\mathcal{I}\varphi)[\![\operatorname{grad} v_h]\!]^{\perp} \le \|\varphi-\mathcal{I}\varphi\|_{L^2(E)}|[\![v_h]\!]|_{H^1(E)} \lesssim h_E^{1/2}\|\vec{\sigma}\|_{H(\operatorname{div},\Omega_E)}|[\![v_h]\!]|_{H^1(E)}\,.$$

Therefore,

$$\langle \operatorname{tr}_n(\operatorname{curl}\varphi), v_h\rangle_h \lesssim \Big(\sum_{E\in\mathcal{E}} \|\vec{\sigma}\|^2_{H(\operatorname{div},\Omega_E)}\Big)^{1/2}\Big(\sum_{E\in\mathcal{E}} h_E|[\![v_h]\!]|^2_{H^1(E)}\Big)^{1/2}. \tag{131}$$

Take $\vec{\sigma}_h^2 = \vec{\mathcal{I}}\vec{\psi}$ and observe that

$$\langle \operatorname{tr}_n(\vec{\psi}-\vec{\mathcal{I}}\vec{\psi}), v_h\rangle_h = \sum_{K\in\mathcal{T}} ((\vec{\psi}-\vec{\mathcal{I}}\vec{\psi})\cdot\vec{n}_K, v_h)_{\partial K} = \sum_{E\in\mathcal{E}}\int_E (\vec{\psi}-\vec{\mathcal{I}}\vec{\psi})\cdot[\![v_h]\!]\,.$$

Moreover, for each fixed edge E,

$$\int_E (\vec{\psi}-\vec{\mathcal{I}}\vec{\psi})\cdot[\![v_h]\!] \le \|(\vec{\psi}-\vec{\mathcal{I}}\vec{\psi})\cdot\vec{n}\|_{L^2(E)}\|[\![v_h]\!]\|_{L^2(E)} \lesssim h_E^{1/2}\|\vec{\sigma}\|_{H(\operatorname{div},\Omega_E)}\|[\![v_h]\!]\|_{L^2(E)}\,.$$

Therefore,

$$\langle \operatorname{tr}_n(\vec{\psi}), v_h\rangle_h \lesssim \Big(\sum_{E\in\mathcal{E}} \|\vec{\sigma}\|^2_{H(\operatorname{div},\Omega_E)}\Big)^{1/2}\Big(\sum_{E\in\mathcal{E}} h_E\|[\![v_h]\!]\|^2_{L^2(E)}\Big)^{1/2}. \tag{132}$$

Invoking both (131) and (132), we easily deduce

$$\langle \operatorname{tr}_n(\vec{\sigma}), v_h\rangle_h^2 \lesssim \|\vec{\sigma}\|^2_{H(\operatorname{div},\Omega)}\Big(\sum_{E\in\mathcal{E}} h_E\|[\![v_h]\!]\|^2_{H^1(E)}\Big),$$

which, since $\vec{\sigma}$ was chosen arbitrarily, demonstrates the required reliability (upper) bound.

The efficiency (lower) bound continues in two steps. First, observe that $\big\|[\![v_h]\!]\big\|_{L^2(E)}^2 = ([\![v_h]\!],[\![v_h]\!])_E$. Therefore, by (125c),

$$h_E^{1/2}\big\|[\![v_h]\!]\big\|_{L^2(E)} \lesssim \frac{([\![v_h]\!],[\![v_h]\!])_E}{\|\mathscr{E}_n([\![v_h]\!]\cdot\vec{n})\|_{H(\mathrm{div},\Omega_E)}} \leq \sup_{\vec{\sigma}_E\in(C^\infty(E))^d} \frac{\int_E \vec{\sigma}_E\cdot[\![v_h]\!]}{\|\mathscr{E}_n(\vec{\sigma}_E\cdot\vec{n})\|_{H(\mathrm{div},\Omega_E)}}\,. \tag{133}$$

Second, define $\phi_E = \mathscr{E}([\![\operatorname{grad} v_h]\!]^\perp b_E)$. Observe that

$$|[\![v_h]\!]|_{H^1(E)} = \|[\![\operatorname{grad} v_h]\!]^\perp\|_{L^2(E)} \lesssim (\phi_E,[\![\operatorname{grad} v_h]\!]^\perp)_E\,.$$

Moreover, fixing any element $K\subseteq\Omega_E$,

$$(\phi_E,[\![\operatorname{grad} v_h]\!]^\perp)_E = (\phi_E,[\![\operatorname{grad} v_h]\!]^\perp)_{\partial K} = \int_{\partial K}\operatorname{curl}\phi_E\cdot[\![v_h]\!] = \int_E \operatorname{curl}\phi_E\cdot[\![v_h]\!]\,.$$

Therefore, by (125b), $h_E^{1/2}|[\![v_h]\!]|_{H^1(E)}|\phi_E|_{H^1(\Omega_E)} \lesssim \int_E \operatorname{curl}\phi_E\cdot[\![v_h]\!]$. Finally, observe that $|\phi_E|_{H^1(\Omega_E)} = \|\operatorname{curl}\phi_E\|_{H(\mathrm{div},\Omega_E)}$ in 2D and $\|\mathscr{E}_n(\operatorname{curl}\phi_E\cdot\vec{n})\|_{H(\mathrm{div},\Omega_E)} \leq \|\operatorname{curl}\phi_E\|_{H(\mathrm{div},\Omega_E)}$. With these observations, we arrive at the following second local lower bound for each $E\in\mathcal{E}$:

$$h_E^{1/2}|[\![v_h]\!]|_{H^1(E)} \lesssim \frac{\int_E\operatorname{curl}\phi_E\cdot[\![v_h]\!]}{\|\operatorname{curl}\phi_E\|_{H(\mathrm{div},\Omega_E)}} \leq \sup_{\vec{\sigma}_E\in(C^\infty(E))^d}\frac{\int_E\vec{\sigma}_E\cdot[\![v_h]\!]}{\|\mathscr{E}_n(\vec{\sigma}_E\cdot\vec{n})\|_{H(\mathrm{div},\Omega_E)}}\,. \tag{134}$$

Just as in the proof of Lemma 7.12, (133) and (134) deliver a local lower bound on the H^1 norm of the jump in v_h which can be accumulated over all edges to deliver a global lower bound:

$$\begin{aligned}
\sum_{E\in\mathcal{E}} h_E\big\|[\![v_h]\!]\big\|_{H^1(E)}^2 &\lesssim \sum_{E\in\mathcal{E}}\sup_{\vec{\sigma}_E\in(C^\infty(E))^d}\frac{\Big(\int_E\vec{\sigma}_E\cdot[\![v_h]\!]\Big)^2}{\|\mathscr{E}_n(\vec{\sigma}_E\cdot\vec{n})\|_{H(\mathrm{div},\Omega_E)}^2}\\
&= \sup_{\vec{\sigma}\in\prod_{E\in\mathcal{E}}(C^\infty(E))^d}\frac{\big(\sum_{E\in\mathcal{E}}\int_E\vec{\sigma}_E\cdot[\![v_h]\!]\big)^2}{\sum_{E\in\mathcal{E}}\|\mathscr{E}_n(\vec{\sigma}_E\cdot\vec{n})\|_{H(\mathrm{div},\Omega_E)}^2} && \text{by Corollary 2.3}\\
&= \sup_{\vec{\sigma}\in\prod_{E\in\mathcal{E}}(C^\infty(E))^d}\frac{\langle\vec{\sigma}\cdot\vec{n},v_h\rangle_h^2}{\sum_{E\in\mathcal{E}}\|\mathscr{E}_n(\vec{\sigma}_E\cdot\vec{n})\|_{H(\mathrm{div},\Omega_E)}^2}\\
&\lesssim \sup_{\vec{\sigma}\in\prod_{E\in\mathcal{E}}(C^\infty(E))^d}\frac{\langle\vec{\sigma}\cdot\vec{n},v_h\rangle_h^2}{\|\mathscr{E}_n(\vec{\sigma}\cdot\vec{n})\|_{H(\mathrm{div},\Omega)}^2} && \text{by (127b)}\\
&\leq \sup_{\vec{\sigma}\in H(\mathrm{div},\Omega)}\frac{\langle\mathrm{tr}_n(\vec{\sigma}),v_h\rangle_h^2}{\|\vec{\sigma}\|_{H(\mathrm{div},\Omega)}^2}\,. && \square
\end{aligned}$$

7.5 Adaptive mesh refinement

By iteratively generating a loosely optimized approximation space, adaptive mesh refinement (AMR) strategies derived from rigorous *a posteriori* error analysis accelerate the convergence of finite element solutions. The general approach, however, strongly depends on how the error is estimated. For instance, all early strategies employed a global *energy error* estimate of the approximation error. By driving this type of error to zero at each refinement step, the total global error can be controlled and the overall accuracy of the solution can be efficiently enhanced. Goal-oriented adaptive mesh refinement (GMR) strategies, on the other hand, consider the error in a chosen quantity of interest (QOI) which, in our case, is a continuous linear functional output of the solution.

In the literature, there are a number of established goal-oriented adaptive mesh refinement (GMR) and error estimation strategies. Generally, the most common approaches are built on the dual-weighted residual (DWR) method [10, Section 5], which requires the solution of a dual problem on an enriched test space or additional post-processing of the computed dual solution. This method delivers a direct estimate of the QOI error—which may be above or below the true value—and relies on a functional relationship to the error, like (91), which can be localized to the element level and directly used in element marking.

Other competitive strategies instead deliver upper and/or lower bounds on the QOI error [97,110,118]. In these alternative strategies, element marking is usually driven by localized contributions of these bounding error estimates or sometimes by a separate estimate entirely. For instance, some such marking strategies, which rely on a crude QOI error upper bound similar to (110), can be proven to deliver optimal GMR convergence rates [9, 64, 85, 102]. Meanwhile, even though many of the other GMR strategies work extremely well in practice [3, 35, 83, 110, 122], constructing rigorous proofs of optimal convergence rates with them has remained largely unprofitable. The approaches to GMR taken in our experiments is similar to these recent strategies and also [2, 110] in that it employs two independent error estimators, η and η^*, for both the primal and dual (i.e. DPG and DPG*) methods, respectively.

In order to properly introduce our approaches, it is necessary to feature them in the context of other established AMR algorithms for finite element analysis. A brief summary of this is given below.

7.5.1 Refinement indicators and refinement strategies

In the context of finite element analysis, the most pervasive strategy for AMR can be described by the following loop [55, 58]:

$$\ldots \text{solve; estimate; mark; refine; } \ldots$$

Normally, at the first stage of the "estimate" step, a special dedicated estimate of the global error is computed. If this estimate is below a specific tolerance, the loop is broken. Otherwise, it continues until plenteous time or computational resources have been consumed. Usually, this global *a posteriori* error estimate, η_{QOI}, comes in the form of dedicated upper and, sometimes, lower bounds [97, 119, 123, 124]. We will forgo the global error estimate feature of a complete AMR strategy and leave the analysis of this ingredient for future study.

Recall Corollary 7.3, which gives us:

$$|G(u) - G(u_h)| \lesssim \|\mathcal{B}u_h - F\|_{\mathcal{V}'} \|\mathcal{B}'v_h - G\|_{\mathcal{U}'} \,. \tag{135}$$

In this manuscript, GMR will involve the following repeated sequence of actions: (1) compute the approximate solutions u_h and v_h; (2) construct computable *a posteriori* error estimators

$$\|\mathcal{B}u_h - F\|_{\mathcal{V}'} \eqsim \eta(u_h) \qquad \text{and} \qquad \|\mathcal{B}'v_h - G\|_{\mathcal{U}'} \eqsim \eta^*(v_h)\,;$$

(3) decompose the error estimators $\eta(u_h)$ and $\eta^*(v_h)$ into element-wise components, called *refinement indicators*,

$$\eta(u_h)^2 = \sum_{K\in\mathcal{T}} (\eta_K)^2 \qquad \text{and} \qquad \eta^*(v_h) = \sum_{K\in\mathcal{T}} (\eta_K^*)^2\,;$$

(4) mark suitable elements for refinement using a marking convention influenced by contributions of both sets of refinement indicators $\{\eta_K\}_{K\in\mathcal{T}}$ and $\{\eta_K^*\}_{K\in\mathcal{T}}$; (5) refine all marked elements with the intention of driving the upper bound (135) to zero with an exceptional rate.

The general AMR strategy above is summarized in Algorithm 1. In this manuscript, the only difference between GMR and solution-adaptive mesh refinement (SMR), which has usually driven AMR with DPG in the past, will be in the marking strategy, i.e., step (4).

Algorithm 1 Adaptive mesh refinement

Input: initial mesh $\mathcal{T}$, marking strategy, tolerance $\eta_{\text{tol.}}$.
loop
(1) Solve for u_h and v_h on $\mathcal{T}$.
(2) Compute global error estimate η_{QOI}.
if $\eta_{\text{QOI}} < \eta_{\text{tol.}}$ **then**
break
(3) Compute the two sets of refinement indicators $\{\eta_K\}_{K\in\mathcal{T}}$ and $\{\eta_K^*\}_{K\in\mathcal{T}}$.
(4) Mark elements for refinement $\mathcal{M}\subseteq\mathcal{T}$, as dictated by *the marking strategy.*
(5) Refine all marked elements $K\in\mathcal{M}$ and construct new mesh $\mathcal{T}$.
return acceptably accurate solution u_h and QOI $G(u_h)$.

Note that the product $\eta(u_h)\cdot\eta^*(v_h)$ could be used for the global error estimate η_{QOI}. However, in many scenarios, this can be such a poor overestimate of the true global error, it would be inefficient to formulate a stopping criterion based on it alone.

At step (5) of Algorithm 1, we assume that a minimal set of additional elements are also refined, beyond the set of marked elements $\mathcal{M}\subseteq\mathcal{T}$, to ensure that the new mesh has only one level of hanging nodes. Under this assumption, the freedom to construct refinement indicators $\{\eta_K\}_{K\in\mathcal{T}}$ and $\{\eta_K^*\}_{K\in\mathcal{T}}$ as well as the choice of marking strategy are the largest sources of flexibility in GMR.

We will consider only one class of refinement indicators for the DPG problem $\{\eta_K\}_{K\in\mathcal{T}}$, presented in Section 7.3. This class coincides with the same well established, standard energy norm refinement indicators that have been used for SMR in many previous DPG studies [27, 34, 50, 66, 90, 126, 137]. On the other hand, we will consider three different classes of refinement indicators $\{\eta_K^*\}_{K\in\mathcal{T}}$ for the DPG* problem. One of these immediately follows from the analysis in the previous section. The other two are derived only in [93].

7.5.2 Marking strategies

Marking strategies are the primary distinguishing factor in many AMR algorithms [9,55, 64,102]. In our experience, the so-called "greedy" strategy is usually very competitive with—and sometimes even outperforms—more complicated Dörfler-influenced marking strategies. In each of the experiments in Chapter 9, we have used one the following greedy marking strategies.

SMR marking strategy Naive (greedy) solution-oriented/energy-based strategy

---

Input: Refinement parameter $0 < \theta < 1$, set of refinement indicators $\{\eta_K\}_{K\in\mathcal{T}}$.
Calculate

$$\eta_{\max} = \max_{K\in\mathcal{T}} \eta_K \,.$$

Mark all elements $K \in \mathcal{T}$ such that

$$\theta\, \eta_{\max} \leq \eta_K \,.$$

GMR marking strategy #1 Naive (greedy) goal-oriented strategy

---

Input: Refinement parameter $0 < \theta < 1$, sets of refinement indicators $\{\eta_K\}_{K\in\mathcal{T}}$ and $\{\eta_K^*\}_{K\in\mathcal{T}}$.
Calculate

$$\eta_{\max} = \max_{K\in\mathcal{T}} \eta_K \eta_K^* \,.$$

Mark all elements $K \in \mathcal{T}$ such that

$$\theta^2\, \eta_{\max} \leq \eta_K \eta_K^* \,.$$

GMR marking strategy #2 Naive (greedy) goal-oriented strategy

---

Input: Refinement parameter $0 < \theta < 1$, sets of refinement indicators $\{\eta_K\}_{K\in\mathcal{T}}$ and $\{\eta_K^*\}_{K\in\mathcal{T}}$.
Calculate

$$\eta_{\max}^2 = \max_{K\in\mathcal{T}} \eta(u_h)^2 \cdot (\eta_K^*)^2 + (\eta_K)^2 \cdot \eta^*(v_h)^2 \,.$$

Mark all elements $K \in \mathcal{T}$ such that

$$\theta\, \eta_{\max}^2 \leq \eta(u_h)^2 \cdot (\eta_K^*)^2 + (\eta_K)^2 \cdot \eta^*(v_h)^2 \,.$$

The SMR strategy above is presently the predominant marking strategy for SMR in DPG studies. Note that this marking strategy does not require any DPG* refinement indicators,

so the DPG* problem need not be solved at all when using it. For GMR, Algorithm 1, paired with the GMR marking strategy #1, was used in the experiments in Section 9.2. Algorithm 1, paired instead with the GMR marking strategy #2, was used in Section 9.3.5.

Chapter 8

Implementation

This chapter abstractly describes the implementation of DPG and DPG* methods. Much of the material here is based on work in [91] and [48]. A notable feature of DPG methods—that they can be formulated in a least-squares setting—is emphasized and expounded upon in this chapter. However, a much more detailed account of this special property can be found in [91].

8.1 Forming the saddle-point linear system

Recall (53). Namely,

$$\begin{cases} (v_h, \nu)_{\mathcal{V}} + b(w_h, \nu) = F(\nu) & \forall\, \nu \in \mathcal{V}_h, \\ b(\mu, v_h) \qquad\qquad\quad = G(\mu) & \forall\, \mu \in \mathcal{U}_h. \end{cases}$$

Upon the choice of bases $\{v_i\}$ and $\{w_j\}$ for the discrete spaces $\mathcal{V}_h$ and $\mathcal{U}_h$, this discrete system can be identified with the following system of linear equations:

$$\begin{bmatrix} \mathsf{G} & \mathsf{B} \\ \mathsf{B}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \mathsf{v} \\ \mathsf{w} \end{bmatrix} = \begin{bmatrix} \mathsf{f} \\ \mathsf{g} \end{bmatrix}. \tag{136}$$

Here, B is a rectangular matrix with coefficients determined by the bilinear form, $\mathsf{B}_{ij} = b(w_j, v_i)$, and, by notational convention, G is a Gram matrix governed by the chosen inner product, $\mathsf{G}_{ik} = (v_i, v_k)_{\mathcal{V}}$. Naturally, the vectors $\mathsf{f}_i = F(v_i)$, $\mathsf{g}_j = G(w_j)$ are identified with the two loads in (53) and the vectors v and w correspond to the coefficients of the chosen basis functions. In the broken space setting (52), the Gram matrix G can be block-diagonal. In that case, inverting G is computationally feasible and the Schur complement of (136) (cf. (16) and (18)) may be used to solve for the single vector w, independent of v, with a much smaller system:

$$\mathsf{B}^{\mathsf{T}}\mathsf{G}^{-1}\mathsf{B}\mathsf{w} = \mathsf{B}^{\mathsf{T}}\mathsf{G}^{-1}\mathsf{f} - \mathsf{g}\,. \tag{137}$$

Notice that the DPG stiffness matrix, $\mathsf{B}^{\mathsf{T}}\mathsf{G}^{-1}\mathsf{B}$, is always symmetric and positive-definite and that after solving for w via (137), v can always be recovered with only local cost; i.e., $\mathsf{v} = \mathsf{G}^{-1}(\mathsf{f} - \mathsf{B}\mathsf{w})$.

8.2 The overdetermined system

Let $M = \dim(\mathcal{V}_h)$ and $N = \dim(\mathcal{U}_h)$ and consider the equation $\mathsf{B}\mathsf{u} = \mathsf{f}$, with B, u, and f defined above. This linear system is obviously overdetermined due to the fact that $M > N$. In this case, it may still be uniquely solved in the following least-squares sense:

$$\min_{\mathsf{u}\in\mathbb{R}^N} \|\mathsf{B}\mathsf{u} - \mathsf{f}\|_{\mathsf{G}^{-1}}\,. \tag{138}$$

Here, the Gram matrix G^{-1} naturally holds the role of the weighting matrix in the least-squares problem.

In place of a general basis $\{v_i\}$, consider if we had used an orthonormal basis $\{\widetilde{v}_i\}$ in constructing the Gram matrix: $\widetilde{\mathsf{G}}_{ik} = (\widetilde{v}_i, \widetilde{v}_k)_{\mathcal{V}} = \delta_{ik}$. In this case, (138) simply becomes an ordinary linear least-squares problem:

$$\min_{\mathsf{u}\in\mathbb{R}^N} \|\widetilde{\mathsf{B}}\mathsf{u} - \widetilde{\mathsf{f}}\|_2\,, \tag{139}$$

where $\widetilde{\mathsf{B}}_{ij} = b(u_j, \widetilde{v}_i)$ and $\widetilde{\mathsf{f}}_i = F(\widetilde{v}_i)$.

Clearly, both (138) and (139) are equivalent to the DPG problem coming from (53). As discussed in Section 8.4, a DPG problem in the form of (139) has numerical advantages stemming from the fact the normal equation $\widetilde{\mathsf{B}}^{\mathsf{T}}\widetilde{\mathsf{B}}\mathsf{u} = \widetilde{\mathsf{B}}^{\mathsf{T}}\widetilde{\mathsf{f}}$ can be avoided when solving for u. As we immediately demonstrate, precomputing an orthonormal basis $\{\widetilde{v}_i\}$ is not necessary in order to form $\widetilde{\mathsf{B}}$ or $\widetilde{\mathsf{f}}$ and these matrices can be computed during the process of inverting G. Indeed, first notice that because G is positive definite it is congruent to the identity matrix. Therefore, one may arrive at (139) by defining $\widetilde{\mathsf{B}} = \mathsf{W}^{-1}\mathsf{B}$ and $\widetilde{\mathsf{f}} = \mathsf{W}^{-1}\mathsf{f}$, where W is any matrix solving the equation $\mathsf{W}\mathsf{W}^{\mathsf{T}} = \mathsf{G}$. Such a factorization is intimately related to finding a $\mathcal{V}$-orthogonal change-of-basis transformation. For instance, consider an arbitrary basis vector $v_i \in \{v_i\}$ represented in an orthonormal basis $\{\widetilde{v}_i\}$; i.e., $v_i = \sum_{k=1}^{M} \mathsf{W}_{ik}\widetilde{v}_k$. Using the latter

expression twice, any Gram matrix G can be represented as

$$\mathsf{G}_{ij} = (v_i, v_j)_{\mathcal{V}} = \sum_{k,\,l=1}^{M} \mathsf{W}_{ik}\mathsf{W}_{jl}(\widetilde{v}_k, \widetilde{v}_l)_{\mathcal{V}} = \sum_{k=1}^{M} \mathsf{W}_{ik}\mathsf{W}_{jk}\,.$$

That is, $\mathsf{G} = \mathsf{W}\mathsf{W}^{\mathsf{T}}$.

If $M > 1$, the equation $\mathsf{W}\mathsf{W}^{\mathsf{T}} = \mathsf{G}$ has an infinite number of solutions, but a convenient choice is the (unique) Cholesky factorization $\mathsf{G} = \mathsf{L}\mathsf{L}^{\mathsf{T}}$. In this case, we may rewrite (139) as

$$\min_{\mathsf{u}\in\mathbb{R}^N} \left\|\mathsf{L}^{-1}(\mathsf{B}\mathsf{u} - \mathsf{f})\right\|_2$$

and the matrices $\widetilde{\mathsf{B}} = \mathsf{L}^{-1}\mathsf{B}$ and $\widetilde{\mathsf{f}} = \mathsf{L}^{-1}\mathsf{f}$ can be computed efficiently by back-substitution. In the statistics and signal processing communities, the factoring and row-weighting procedure described above is known as "whitening" or "sphering." We refer the reader to [94] for a discussion on the benefits of the some other possibilities for the change-of-basis matrices W in that context.

8.3 The underdetermined system

Following the notation of the previous two sections, consider the underdetermined linear system $\mathsf{B}^{\mathsf{T}}\mathsf{v} = \mathsf{g}$. In order to select a unique solution of this equation, we may consider the following constrained minimum norm problem:

$$\min_{\mathsf{v}\in\mathbb{R}^M} \|\mathsf{v}\|_{\mathsf{G}} \quad \text{such that} \quad \mathsf{B}^{\mathsf{T}}\mathsf{v} = \mathsf{g}\,. \tag{140}$$

Similar to the relationship between DPG methods and the discrete minimization problem (138), this problem is naturally associated with DPG* methods. In the present work, we have only explored solving (140) by first forming (137) (with $\mathsf{f} = 0$), solving for the auxiliary variable w featured there, and post-processing the computed solution w to recover $\mathsf{v} = \mathsf{G}^{-1}\mathsf{B}\mathsf{w}$. However, for completeness, an approach analogous to one we took with the DPG system would transform (140) into

$$\min_{\widetilde{\mathsf{v}}\in\mathbb{R}^M} \|\widetilde{\mathsf{v}}\|_2 \quad \text{such that} \quad \widetilde{\mathsf{B}}^{\mathsf{T}}\widetilde{\mathsf{v}} = \mathsf{g}\,.$$

8.4 Solution algorithms for the overdetermined system

There are several prevalent strategies to solve generalized linear least-squares problems

$$\min_{\mathsf{u}\in\mathbb{R}^N} \|\mathsf{B}\mathsf{u}-\mathsf{f}\|_{\mathsf{G}^{-1}}\,. \tag{DLS}$$

Under infinite precision arithmetic, each approach is essentially equivalent. However, in terms of time-to-solution and round-off error sensitivity, they are significantly different. A survey the most important direct methods is given in this subsection.

8.4.1 The normal equation

In order to solve for u, we could begin by forming the normal equation:

$$\mathsf{A}\mathsf{u}=\mathsf{b}\,,\quad \text{where}\quad \mathsf{A}=\mathsf{B}^{\mathsf{T}}\mathsf{G}^{-1}\mathsf{B}\quad\text{and}\quad \mathsf{b}=\mathsf{B}^{\mathsf{T}}\mathsf{G}^{-1}\mathsf{f}\,. \tag{NE}$$

Beginning in this way is standard practice in the DPG community and is advantageous in that the stiffness matrix A and load vector b can be constructed locally and assembled in a sparse format using standard finite element assembly algorithms. Moreover, $\mathsf{A}=\widetilde{\mathsf{B}}^{\mathsf{T}}\widetilde{\mathsf{B}}$ is symmetric positive-definite and therefore has a structure amenable to efficient linear solvers not usually available for many challenging problems. On the other hand, the condition number of A will grow quadratically with the condition number of $\widetilde{\mathsf{B}}$ and, likewise, so will the upper bound on the round-off error of the normal equation solution. Importantly, however, this is not to say that when the normal equation is formed the growth of the condition number will be greater than most other standard finite element methods. Indeed, various DPG methods can be proved to induce a normal equation stiffness matrix with the ordinary condition number growth $\mathcal{O}(h^{-2})$ [82]. A similar result holds in the context of first-order system least-squares (FOSLS) methods [13, 14] which themselves induce linear systems closely related to the normal equation above (see [91]).

Although the properties above make a strong case for the normal equation approach, the scaling constant controlling the condition number of the stiffness matrix A can still be very large due to parameters in the problem being solved. If the condition number is large enough,

this can become an impediment to producing accurate solutions to very difficult problems and can induce large and sometimes overwhelming numerical errors. This is the primary reason that it is sometimes convenient to consider other approaches which deal explicitly with the matrices B, W, and f, and avoid the normal equation altogether.

8.4.2 Orthogonal decompositions

The most practical alternative to the normal equation when solving for u is to deal directly with the matrices $\widetilde{\mathsf{B}}$ and $\widetilde{\mathsf{f}}$ coming from the (sparsely weighted) linear least-squares problem

$$\text{(LS)} \qquad \min_{\mathsf{u}\in\mathbb{R}^N} \|\widetilde{\mathsf{B}}\mathsf{u}-\widetilde{\mathsf{f}}\|_2\,, \quad \text{where} \quad \widetilde{\mathsf{B}}=\mathsf{W}^{-1}\mathsf{B}\,, \quad \widetilde{\mathsf{f}}=\mathsf{W}^{-1}\mathsf{f}\,, \quad \text{and} \quad \mathsf{W}\mathsf{W}^{\mathsf{T}}=\mathsf{G}\,.$$

The most common of these approaches is the orthogonalization algorithm called QR-factorization (Householder, Givens, MGS) first introduced for least-squares problems in [78]. Other direct approaches are SVD, complete orthogonal decomposition, and Peters-Wilkinson decomposition as well as various hybrid methods [12]. Each of these approaches are usually less computationally efficient than solving via the normal equation, but are sometimes preferred because they are more numerically stable.

For the purposes we have in mind, the matrix $\widetilde{\mathsf{B}}$ will be large and sparse and therefore, because not all of the methods above are well suited for sparse matrices or amenable to parallel computing, we will focus only on the QR approach. As shown in various textbooks [12, 80, 136], the relative round-off error in the solution from a least-squares QR solve is controlled by $\mathrm{cond}(\widetilde{\mathsf{B}})+\rho(\widetilde{\mathsf{B}},\widetilde{\mathsf{f}})\cdot\mathrm{cond}(\widetilde{\mathsf{B}})^2$, where

$$\rho(\widetilde{\mathsf{B}},\widetilde{\mathsf{f}})=\frac{\|\widetilde{\mathsf{B}}\mathsf{u}-\widetilde{\mathsf{f}}\|_2}{\|\widetilde{\mathsf{B}}\|_2\|\mathsf{u}\|_2}\,, \tag{141}$$

and therefore depends upon the load vector.

Due to the compatibility condition $F\in\mathrm{Range}\,\mathcal{B}$, the residual (the numerator in (141)) is expected to tend to zero as the mesh is refined. The reader may even observe that $\rho(\widetilde{\mathsf{B}},\widetilde{\mathsf{f}})$ vanishes when $\widetilde{\mathsf{f}}\in\mathrm{Range}\,\widetilde{\mathsf{B}}$. Even in the general case, however, $\rho(\widetilde{\mathsf{B}},\widetilde{\mathsf{f}})\to 0$ as $h\to 0$. Of course,

the validity of this convergence as well as its rate will be determined by the interpolation spaces used in the discretization and, in many common scenarios, the *a priori* bound can be proven to decrease at a rate of at least $\mathcal{O}(h)$. Indeed, for many cases we have in mind, this rate is of the form $\mathcal{O}(h^p)$ where $p \geq 1$ is a parameter indicating the polynomial order used in the discretization.

In summary, the quadratic condition number term controlling the round-off error in a QR solve will be offset by a converging solution. For instance, recall that $\text{cond}(\widetilde{\mathsf{B}}) = \text{cond}(\mathsf{A})^{1/2}$. Therefore, for many DPG methods, the condition number growth in $\widetilde{\mathsf{B}}$ is only $\mathcal{O}(h^{-1})$. Moreover, if as described above the residual (141) is at least $\mathcal{O}(h)$, then $\rho(\widetilde{\mathsf{B}}, \widetilde{\mathsf{f}}) \cdot \text{cond}(\widetilde{\mathsf{B}})^2$ may be no worse than $\mathcal{O}(h^{-1})$. In such conventional scenarios, the numerical sensitivity of the least-squares solution is controlled simply by the inverse of the mesh size and will be far more accurate than any normal equation approach! That is, a QR-based algorithm for most common DPG methods will deliver an error bound of

$$\|\mathfrak{e}\|_2 \leq \epsilon_{mach.} \|\mathsf{u}\|_2 C h^{-1} \,,$$

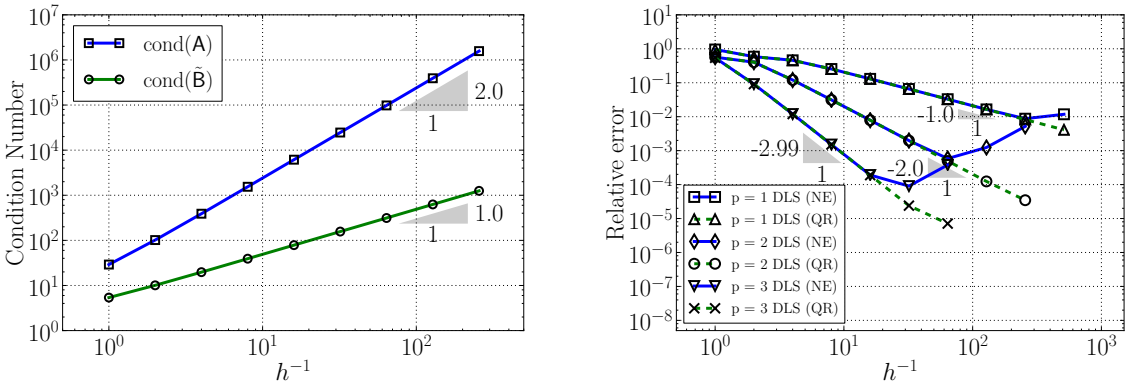
where $\mathfrak{e}$ is the round-off error in the computation of the least-squares solution, $\epsilon_{mach.}$ is machine precision, and C is a mesh-independent constant.

8.4.3 Generalized least-squares

For the most badly behaved problems, [12, 80] suggest beginning with the generalized least-squares problem

$$\text{(GLS)} \qquad \min_{\mathsf{u} \in \mathbb{R}^N,\, \mathsf{r} \in \mathbb{F}^M} \|\mathsf{r}\|_2 \quad \text{subject to} \quad \mathsf{B}\mathsf{u} + \mathsf{W}\mathsf{r} = \mathsf{f} \,, \quad \text{where} \quad \mathsf{W}\mathsf{W}^{\mathsf{T}} = \mathsf{G} \,.$$

For a direct method, the solution coefficients u are then suggested to be computed using a QR factorization approach which was first described in [116]. Although there are a couple of strict advantages of this idea, the size of the resulting saddle-point system is much larger than the systems in the previously discussed methods. Moreover, the QR-based solution algorithm given in [116] is seemingly impractical for any reasonably large sparse system because it involves storing and applying large and probably dense orthogonal matrices. In [117], an efficient algorithm is



(A) Condition numbers.

(B) Poisson's equation (single precision).

Figure 8.1: (A) : Condition number growth of the DPG stiffness matrices A and $\widetilde{\mathsf{B}}$ coming from the ultraweak formulation of Poisson's equation (78a). Here, $p = 2$ and $\mathrm{d}p = 1$. (B) : Differences in the discrete solution are clearly observed between the solution obtained when forming the normal equation (NE) and the solution obtained using the QR factorization of the overdetermined system directly. Here $p \in \{1, 2, 3\}$ and $\mathrm{d}p = 1$. See [91] for further details.

proposed for the case that G is block diagonal; however, we are unaware of any multi-frontal implementations so have not explored it in our experiments.

Notably, this saddle-point approach is analogous to the finite element methods described in [33] and [40]. Moreover, although this generalized least-squares starting point may be too expensive to be practical for direct linear solvers, it may have benefits for iterative solution algorithms [11, 79, 143]. One such very promising technique in the DPG context, inspired by PDE-constrained optimization, is developed in [23] in a similar setting where G is not factored.

Chapter 9

Numerical experiments

This chapter presents a number of numerical experiments which verify the *a priori* error estimation theory for DPG* methods in Chapter 6 and demonstrate the efficacy of the goal-oriented adaptive mesh refinement (GMR) strategies developed in Chapter 7 for both linear and nonlinear PDE models. Each of experiment here was performed with one of the following three finite element software packages which have been extensively used to implement DPG methods.

Camellia [127, 128]: A user-friendly C++ toolbox developed by Dr. Nathan V. Roberts which relies on Sandia's Trilinos library of packages [84].

***hp*2D [41]:** A sophisticated suite of Fortan routines with support for hierarchical and anisotropic adaptive h- and p-refinements on hybrid meshes consisting of quadrilateral and triangular elements [41]. In addition, *hp*2D interfaces with the complete de Rham sequence energy space-conforming ESEAS library of orientation-embedded shape functions [68, 88].

***hp*3D [53]:** Similar to *hp*2D, this is an extremely sophisticated suite of Fortan routines supporting hierarchical and anisotropic adaptive h- and p-refinements; in this case, on hybrid meshes consisting of hexahedral, tetrahedral, prismatic, and pyramidal elements [53]. It also interfaces with the ESEAS orientation-embedded shape function package.

9.1 A DPG* method for Poisson's equation

This section serves to verify the mathematical theory on *a priori* and *a posteriori* error estimation for DPG* methods developed in Chapters 6 and 7. Poisson's equation is chosen here for its simplicity. In our first set of experiments, we used Camellia for the *a priori* convergence rate verification on the model square domain $\Omega_\square = [0,1]^2$; see Section 9.1.2. In our second set of experiments, we used *hp*2D to implement a simple *hp*-adaptive algorithm for a singular

solution on the canonical L-shaped domain, $\Omega_{\boxplus} = (-1,1)^2 \setminus [0,1] \times [-1,0]$; see Section 9.1.3. This second set of experiments parallels a similar study with the analogous DPG method [45].

9.1.1 Set-up

Let $\Omega \in \{\Omega_{\square}, \Omega_{\boxplus}\}$ and let Γ_{D} and Γ_{N} be disjoint and relatively open subsets comprising $\partial\Omega$; $\Gamma_{\mathrm{D}} \cap \Gamma_{\mathrm{N}} = \emptyset$, $\overline{\Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{N}}} = \partial\Omega$. All of our experiments investigated some form of Poisson's equation:

$$\begin{cases} -\Delta v = f & \text{in } \Omega\,, \\ \quad v = v_0 & \text{on } \Gamma_{\mathrm{D}}\,, \\ \dfrac{\partial v}{\partial n} = p_n & \text{on } \Gamma_{\mathrm{N}}\,, \end{cases} \tag{142}$$

where the load $f \in L^2(\Omega)$ and the boundary data v_0 and p_n are appropriately smooth.

This section includes an *hp*-adaptive mesh refinement example. Therefore, it is appropriate to define what an *hp* mesh will be understood as. For each quadrilateral element $K \in \mathcal{T}$, we associate a unique (anisotropic) polynomial order $p_K, q_K \geq 1$, for each Cartesian direction, respectively. Each associated polynomial order can be naturally related to a (p_K, q_K)-order *conforming* finite element de Rham sequence. For instance, begin with the standard Nédélec spaces of the first type,

$$\mathcal{Q}^{p_K,q_K}(K) \xrightarrow{\text{curl}} \mathcal{Q}^{p_K,q_K-1} \times \mathcal{Q}^{p_K-1,q_K}(K) \xrightarrow{\text{div}} \mathcal{Q}^{p_K-1,q_K-1}(K)\,,$$

where $\mathcal{Q}^{p_K,q_K}(K)$ is the space of bivariate polynomials over K with degree at most p_K horizontally and q_K vertically. Now, consider the mesh-dependent sequence $W_{hp} \xrightarrow{\text{curl}} V_{hp} \xrightarrow{\text{div}} Y_{hp}$, where

$$\begin{aligned} W_{hp} &= \{w \in H_0^1(\Omega) : w|_K \in \mathcal{Q}^{p_K,q_K}(K)\ \forall K \in \mathcal{T}\}\,, \\ V_{hp} &= \{\vec{q} \in H(\mathrm{div},\Omega) : \vec{q}|_K \in \mathcal{Q}^{p_K,q_K-1}(K) \times \mathcal{Q}^{p_K-1,q_K}(K)\ \forall K \in \mathcal{T}\}\,, \\ Y_{hp} &= \{w \in L^2(\Omega) : w|_K \in \mathcal{Q}^{p_K-1,q_K-1}(K)\ \forall K \in \mathcal{T}\}\,. \end{aligned}$$

Recall the definitions of the trace operators tr and tr_n in (79). We define the (anisotropic) *hp* trial space to be $\mathcal{U}_h = Y_{hp} \times Y_{hp}^2 \times \mathrm{tr}(W_{hp}) \times \mathrm{tr}_n(V_{hp})$, where p_K and q_K are allowed to

vary freely throughout the mesh. With this definition, notice that the polynomial order of an hp interface function, when restricted to a single shared edge $E \in \mathcal{E}_h$, $E = \bigcap_{\overline{K} \cap E \neq \emptyset} \overline{K}$, will naturally be restricted by the lowest polynomial order of all elements $\overline{K} \cap E \neq \emptyset$ sharing that edge. Such a space obeys the so-called *minimum rule*.

To define the broken hp test functions $v \in \mathcal{V}_h$, begin with the following spaces:

$$\widetilde{W}_{hp,\mathrm{d}p} = \{v \in H^1(\mathcal{T}) : v|_K \in \mathcal{Q}^{p_K+\mathrm{d}p, q_K+\mathrm{d}p}(K) \ \forall K \in \mathcal{T}\},$$
$$\widetilde{V}_{hp,\mathrm{d}p} = \{\vec{q} \in H(\mathrm{div}, \mathcal{T}) : \vec{q}|_K \in \mathcal{Q}^{p_K+\mathrm{d}p, q_K+\mathrm{d}p-1}(K) \times \mathcal{Q}^{p_K+\mathrm{d}p-1, q_K+\mathrm{d}p}(K) \ \forall K \in \mathcal{T}\}.$$

In all of our numerical experiments here, we used $\mathcal{V}_h = \widetilde{W}_{hp,\mathrm{d}p} \times \widetilde{V}_{hp,\mathrm{d}p}$, where $\mathrm{d}p \in \{0, 1, 2\}$.

9.1.2 Pure Dirichlet boundary conditions on a square domain

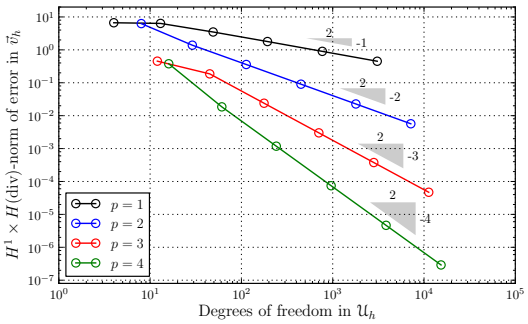
Let $\Omega = \Omega_\square$ and $\Gamma_\mathrm{D} = \partial\Omega$. In this first set of experiments, we considered two seemingly benign cases for the load in (142): (i) $f = 2\pi^2 \sin(\pi x)\sin(\pi y)$ and $v_0 = 0$; and (ii) $f = 0$ and $v_0 = 1$. In both cases, the exact solution v is infinitely smooth. Indeed, in case (i), $v = \sin(\pi x)\sin(\pi y)$ and, in case (ii), $v = 1$.

Recall from Theorem 6.7 that the best approximation error of a DPG* method involves the Lagrange multiplier $\vec{\lambda} = (\lambda, \vec{\zeta}, \widehat{\lambda}, \widehat{\zeta})$ as well as the DPG* solution variable $\vec{v} = (v, \vec{p})$. Assume that v is smooth. With the norm $\|(\nu, \vec{\tau})\|_\mathcal{V}^2 = \|\nu\|_{H^1(\mathcal{T})}^2 + \|\vec{\tau}\|_{H(\mathrm{div},\mathcal{T})}^2$, λ solves

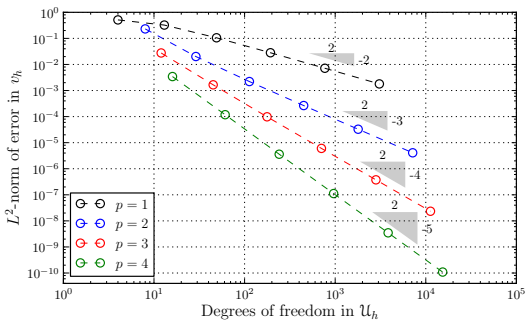
$$\begin{cases} -\Delta\lambda = g & \text{in } \Omega, \\ \lambda = 0 & \text{on } \partial\Omega, \end{cases} \tag{144}$$

where $g = v - 2\Delta v + \Delta^2 v$. Indeed, consider (97) and observe that $-\Delta\lambda = -\Delta f - \Delta e = \Delta(\Delta v) + (v + 2f) = v - 2\Delta v + \Delta^2 v$, where we have also used that $-\Delta v = f$ and $-\Delta e = v + 2f$. In case (i), $g = (1 + 4\pi^2 - 4\pi^4)\sin(\pi x)\sin(\pi y)$; meanwhile, in case (ii), $g = 1$. Notably, in both cases, $g \in C^\infty(\overline{\Omega})$.

In the first case, λ can easily be shown to be infinitely smooth. Therefore, by Corollary 6.10, the convergence rate of the DPG* method under uniform h-refinement is limited only by the underlying de Rham sequence polynomial order p. Indeed, Figure 9.1 (A) demonstrates

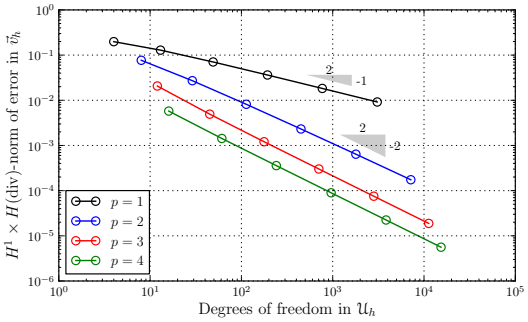


(A) Optimal rates.

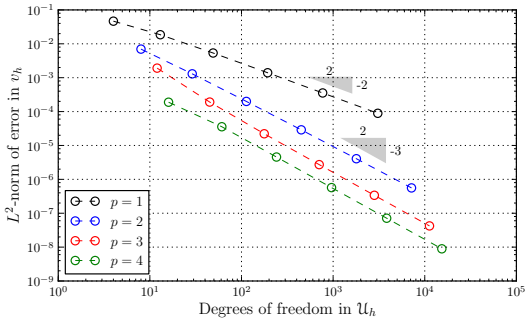


(B) Optimal rates in the L^2-norm.

Figure 9.1: Convergence under h-uniform mesh refinements with the manufactured solution $v(x, y) = \sin(\pi x) \sin(\pi y)$. (Here, $\mathrm{d}p = 1$.)



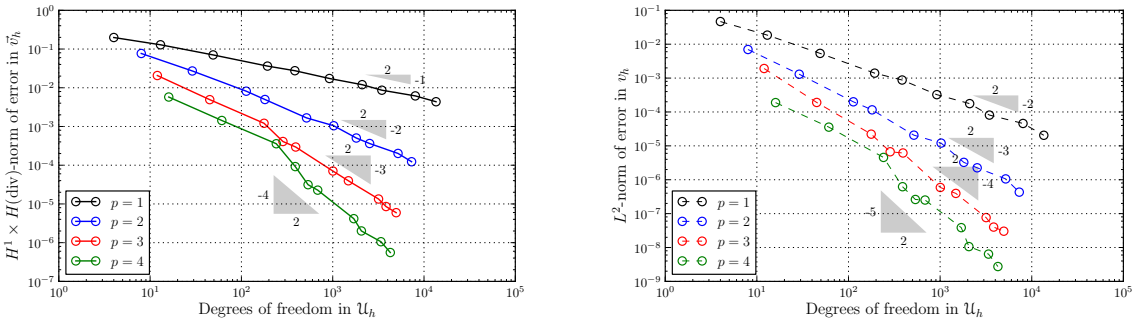
(A) Limited optimal rates.



(B) Limited optimal rates in the L^2-norm.

Figure 9.2: Convergence under h-uniform mesh refinements with the manufactured solution $v(x, y) = 1$. (Here, $\mathrm{d}p = 1$.)

the convergence of the corresponding discrete solution $\vec{v}_h = (v_h, \vec{p}_h)$ to the exact solution, $\vec{v} = (v, \operatorname{grad} v)$, measured in the full test norm above, starting with an single-element mesh with (isotropic) polynomial order $p_K = q_K = p \in \{1, 2, 3, 4\}$. Figure 9.1 (B) presents the convergence of only the solution variable v_h, measured in the $L^2(\Omega)$-norm. Although both figures correspond only to a test space enrichment of $\mathrm{d}p = 1$, similar results were observed for each choice $\mathrm{d}p \in \{0, 1, 2\}$.



(A) Recovered optimal rates.

(B) Recovered optimal rates in the L^2-norm.

Figure 9.3: Convergence under h-adaptive mesh refinements with the manufactured solution $v(x, y) = 1$. (Here, $\mathrm{d}p = 1$.)

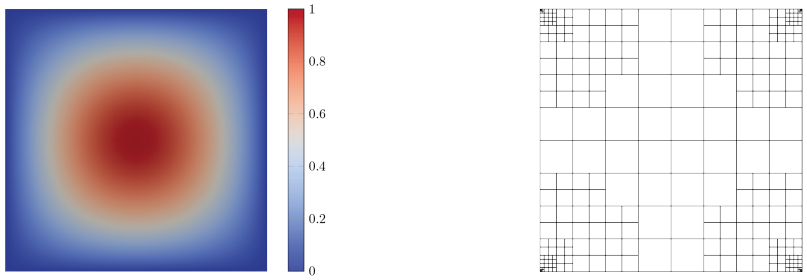
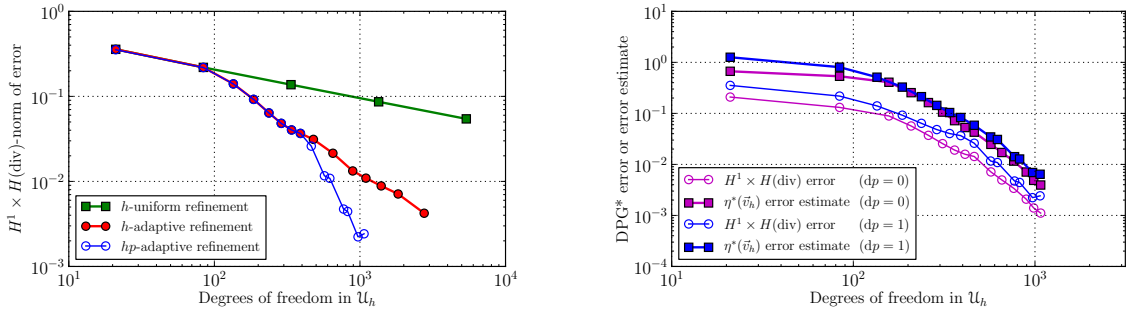


Figure 9.4: Left: The DPG* Lagrange multiplier solution component λ when $v = 1$. (Color scale represents the solution values.) Right: The corresponding quadtree mesh coming from the h-adaptive algorithm after ten refinements. (Here, $p = 4$ and $\mathrm{d}p = 1$.)

In the second case, because of the irregularity of the boundary, the Lagrange multiplier λ is *not* infinitely smooth.[3] Therefore, even though the majority of standard methods would be able to exactly reproduce the exact solution, the DPG* method experiences rate-limited convergence under uniform h-refinements. This is evidenced by Figure 9.2. However, as demonstrated in Figure 9.3, optimal convergence rates can still be easily recovered using the solution-oriented adaptive mesh refinement strategy described in Section 7.5. See Figure 9.4 for a visual depiction of the solution of the corresponding auxiliary problem (144) as well as a corresponding adaptively refined mesh.

[3]Standard results for elliptic regularity theory can be used to show that λ is, however, infinitely smooth in the interior of the domain [61].



(A) Effect of refinement strategies.

(B) Effect of dp.

Figure 9.5: (A) : Comparison of the convergence of various refinement strategies when d$p = 1$. (B) : Convergence of the error in the DPG* solution variable $\vec{v}_h$ and the error estimator $\eta^*(\vec{v}_h)$ for two values of dp with the hp-adaptive algorithm.

9.1.3 Mixed boundary conditions on an L-shaped domain

In this example, set $\Omega = \Omega_{\boxplus}$, $\Gamma_{\mathrm{D}} = [0, 1) \times \{0\} \cup \{0\} \times [0, -1)$, and $v_0 = 0$ in (142). Additionally, set p_n to be the normal derivative of the exact solution $v(r, \theta) = r^{2/3} \sin(\frac{2}{3}\theta)$. For this problem, it is well known that the solution $v \in H^{1+s}(\Omega)$ for all $s < 2/3$.

In each of our experiments here, we began with a single three-element mesh composed of congruent squares and uniform order $p_K = q_K = 2$ and d$p = 1$ in all three elements. Figure 9.5 (A) demonstrates the convergence of the solution error we witnessed under h-uniform, h-adaptive (as described above), and hp-adaptive refinements using a flagging strategy where all marked elements adjacent to the origin (i.e. the singular point) are h-refined and all other marked elements are p-refined. As shown in Figure 9.5 (B), the error estimator $\eta^*(\vec{v}_h)$ generally overestimated the solution error and the dependence upon dp was not seen to be particularly significant. Figures 9.6 and 9.7 depict both the computed solution and the hp mesh after fifteen refinement steps.

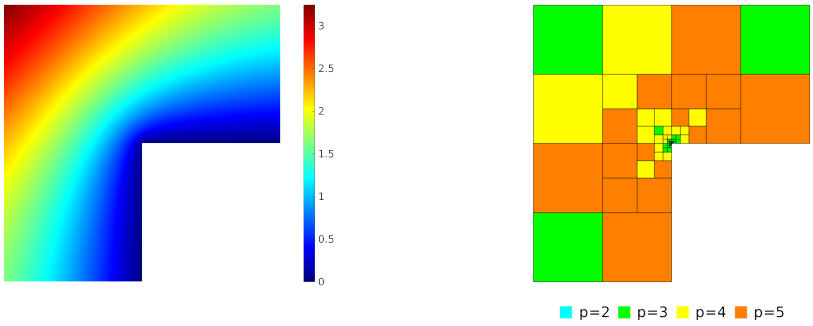


Figure 9.6: Left: The DPG* solution v. (Color scale represents solution values.) Right: The corresponding *hp* quadtree mesh delivered by the *hp*-adaptive algorithm after fifteen refinements. (Colors represent polynomial degree.)

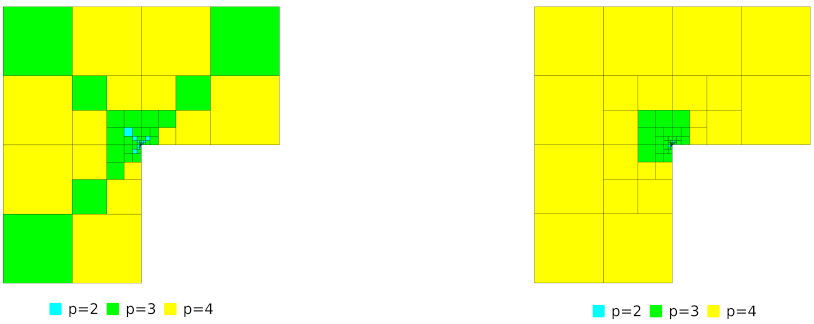


Figure 9.7: Left: The *hp* quadtree mesh delivered by a similar *hp*-adaptive algorithm using a maximum rule [45] after fifteen refinements. Right: The *hp* quadtree mesh delivered by another similar *hp*-adaptive algorithm using a different maximum rule after fifteen refinements. (Colors represent polynomial degree.)

9.2 A study on goal-oriented adaptive mesh refinement

This section serves to demonstrate the efficacy of goal-oriented adaptive mesh refinement (GMR) with DPG methods. Throughout this section, we focus only on Poisson's equation, given in Section 5.1.

Recall that Algorithm 1 requires two sets of error indicators. For the DPG error indicators η_K, we use localized components of the standard implicit DPG error estimator introduced in Section 7.3. For the DPG* error indicators η_K^*, however, we use localized components of three *different* estimators drawn from [93] but still relevant to the present analysis. The first of these, $\eta_K^{*,\mathrm{expl}}$, is simply the three-dimensional analogue of the explicit estimator defined in (124a). The second of these, $\eta_K^{*,\mathrm{impl}}$, comes from a new *implicit* error estimator defined in [93, Section 7.3]. With this second estimator, a special local problem must be solved, for each element $K \in \mathcal{T}$, on an enriched polynomial space. The final set of error indicators, deemed *ad hoc* error indicators, are denoted $\eta_K^{*,\mathrm{a.h.}}$. These error indicators can be arrived at from a clever identity that holds only for the optimal norm defined in Section 3.7. Indeed, when the optimal norm is used with an ultraweak formulation like (80) and the corresponding goal functional $G \in \mathcal{U}'$ only involves interior terms (e.g., $G(\vec{u}) = (u, g_1)_\Omega$) then the Lagrange multiplier $\vec{\lambda}$ can be expressed explicitly (e.g., $\vec{\lambda} = (g_1, \vec{0}, 0, 0)$). When using the graph norm, this property still holds up to a tunable error (cf. Proposition 6.9) and the Lagrange multiplier itself can be directly used to estimate the error. For further details, see [93, Section 7.4].

9.2.1 Set-up

To compare our GMR algorithm with the conventional solution-adaptive mesh refinement (SMR) algorithm (see Section 7.5.1), we use a manufactured solution with two regions of isolated steep and shallow gradients in the convex domain $\Omega = [0,4] \times [0,1] \times [0,1]$ (see Figure 9.9). The exact expression for this manufactured solution, $u = u^{\mathrm{man}} \in C_0^\infty(\Omega)$, is given by

$$u^{\mathrm{man}}(x,y,z) = f(x/4)f(y)f(z)\,, \quad \text{where} \quad f(x) = x(1-x)\big((x/4) + (1-4x)^2\big)\,. \tag{145}$$

The remaining components of $\vec{u} = (u, \vec{\sigma}, \widehat{u}, \widehat{\sigma})$ can easily be derived from this expression.

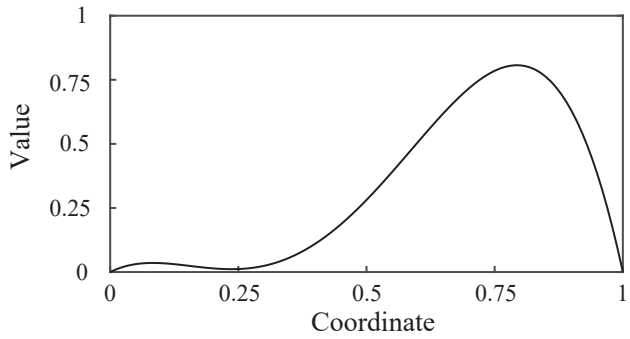


Figure 9.8: Graph of the function $f(x)$ in (145).

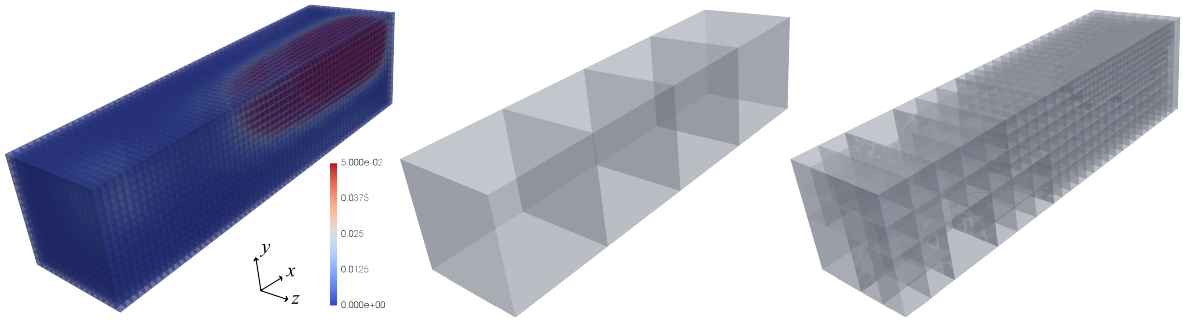


Figure 9.9: Left: converged solution. Center: initial mesh. Right: mesh after twelve SMRs.

Heuristically speaking, the expected behavior of an SMR strategy is to induce the majority of mesh refinements in the region with highest gradients in the solution. That is, refinements will occur where the length scale of the solution is the smallest and the solution is the most difficult to resolve. For the manufactured solution given in (145), the expected behavior is indeed exhibited in Figure 9.9.

Now, consider a goal functional $G \in \mathcal{U}'$ defined in terms of the solution in a region far away from the largest gradients. In this circumstance, it is conceivable that the best possible QOI estimate, for a fixed computational expenditure, would require a mesh with a refinement pattern very different than one coming from the standard series of SMRs. The results in this section will clearly verify this conjecture.

9.2.2 Experimental design

In this section, it is convenient to express $G \in \mathcal{U}$ as

$$G(\vec{u}) = \int_\Omega g_1 \cdot u + \int_\Omega \vec{g}_2 \cdot \vec{\sigma} + \int_{\Gamma_\mathrm{N}} \widehat{g}_3 \cdot \widehat{u} + \int_{\Gamma_\mathrm{D}} \widehat{g}_4 \cdot \widehat{\sigma} \qquad \forall\, \vec{u} = (u, \vec{\sigma}, \widehat{u}, \widehat{\sigma}) \in \mathcal{U}\,. \tag{146}$$

Always beginning with the same four-element mesh depicted in Figure 9.9 (center), we analyzed five different QOIs in the form above, including one modified to pertain to pointwise values of the solution (see Section 9.2.7). In our experiments, $g_1 \in L^2(\Omega)$, $\vec{g}_2 \in L^2(\Omega)^3$, $\widehat{g}_3 \in L^2(\Gamma_\mathrm{N})$, and $\widehat{g}_4 \in H^1_0(\Gamma_\mathrm{D})$ are each piecewise polynomial. Because the ad hoc indicator $\eta_K^{*,\mathrm{a.h.}}$ is not suitable for goal functionals with the two boundary terms in (146), when $G(\vec{u})$ involved nonzero $\widehat{g}_3$ or nonzero $\widehat{g}_4$ and $\eta_K^{*,\mathrm{a.h.}}$ was employed, we instead introduced a sequence of modified goal functionals (see Sections 9.2.5 and 9.2.6). In Section 9.2.7, an extension of this technique is presented for a pointwise-value QOI, $G_{\vec{x}_0}(\vec{u}) = u(\vec{x}_0)$.

All of our computations in this section were performed with *hp*3D. In our third experiment (Section 9.2.5), non-homogeneous Neumann boundary conditions were applied to a non-trivial subset of the boundary $\Gamma_\mathrm{N} \subseteq \partial\Omega$. In order to apply this essential boundary condition to the $\widehat{\sigma}$-variable, we used projection-based interpolation [42]. For each of the energy spaces above, this is a fully-supported feature of the *hp*3D software.

Parameters

Before presenting the results of our experiments, we now list the outstanding parameters in our algorithms and our choices for them in the experiments:

- Like in Section 9.1.1, discretizations of the trial space $\mathcal{U}_h$ and the test space $\mathcal{V}_h$ came from a de Rham sequence of polynomial order $p = 2$ and $p + \mathrm{d}p = 3$, respectively [88]. Note that the polynomial order of the manufactured solution (145) is too high for it to be fully recovered with this trial space discretization.
- In the implicit refinement indicator $\eta_K^{*,\mathrm{impl}}$, the local problems (see [93, Section 7.3]) were solved on individual elements K from the same mesh $\mathcal{T}$ as the global DPG and DPG* problems. Here, an enriched trial space $\mathcal{U}_H$ and further enriched test space $\mathcal{V}_H$ were

constructed as previously described for $\mathcal{U}_h$ and $\mathcal{V}_h$, but with basis functions taken from de Rham sequences of polynomial order $P = p + 1 = 3$ and $P + \mathrm{d}p = 4$, respectively.

- The graph norm (57b) was used, $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_{H(\mathcal{L}^*)}$.
- A refinement factor of $\theta = 0.5$ was set for both the SMR and GMR marking strategies (see Section 7.5.1).

9.2.3 Temperature in a subdomain

In this subsection, we consider the goal functional given in (146) where

$$g_1(x,y,z) = \begin{cases} 1\,, & x \leq 1\,, \\ 0\,, & \text{otherwise,} \end{cases} \qquad \vec{g}_2 = \vec{0}\,, \qquad \widehat{g}_3 = 0\,, \qquad \text{and} \qquad \widehat{g}_4 = 0\,. \tag{147}$$

Physically, this corresponds to a QOI which is the average value of the temperature u in the subdomain $0 \leq x \leq 1$. In these experiments, we used homogeneous Dirichlet boundary conditions, $u|_{\partial\Omega} = 0$.

Define the relative error in the QOI to be $|G(\vec{u} - \vec{u}_h)|/|G(\vec{u})|$. In Figure 9.10, we present the relative error vs. the degrees of freedom in each successive solution. It is immediately evident that each GMR step was far more efficient at reducing the relative error in the QOI than each SMR step. Moreover, taking into account the entire sequence of refinements, each GMR strategy performed very similarly, until nearly the final refinement.

In Figure 9.10 (right), we see the final refined mesh after twelve adaptive mesh refinements with the GMR marking strategy and the ad hoc refinement indicator $\eta_K^{*,\text{a.h.}}$. However, because there are two other classes of DPG* refinement indicators which preformed well for this problem and QOI, Figure 9.11 is provided to compare all three corresponding final solution and meshes. From the given perspective, it appears that the final meshes are extremely similar, but significantly different from the SMR mesh in Figure 9.9 (right). A strong visual similarity in the final GMR meshes was also exhibited in each of our studies. Therefore, from now on, we only provide one representative GMR mesh for illustration.

Finally, we provide Figure 9.12 for a visual depiction of the local temperature error in the region $0 \leq x \leq 1$. In some contexts, a goal functional of the form (147) is chosen to

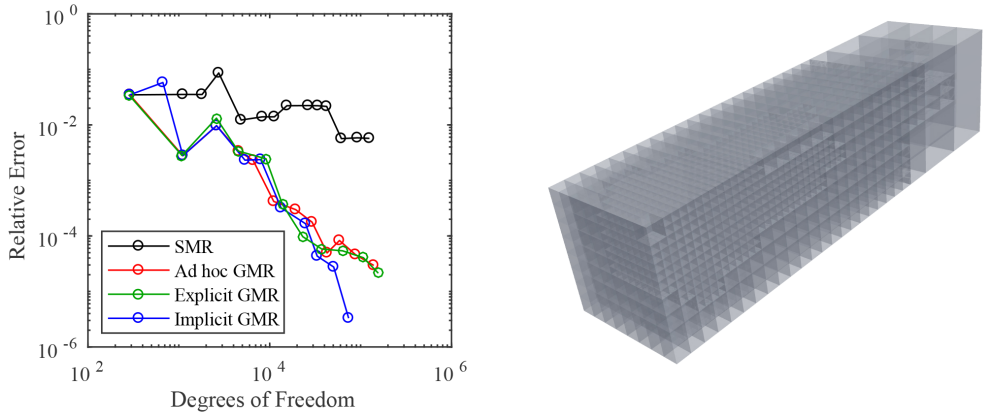


Figure 9.10: Left: The error in the QOI for the first example: average temperature u. Right: The final adaptively refined mesh using the ad hoc GMR approach.

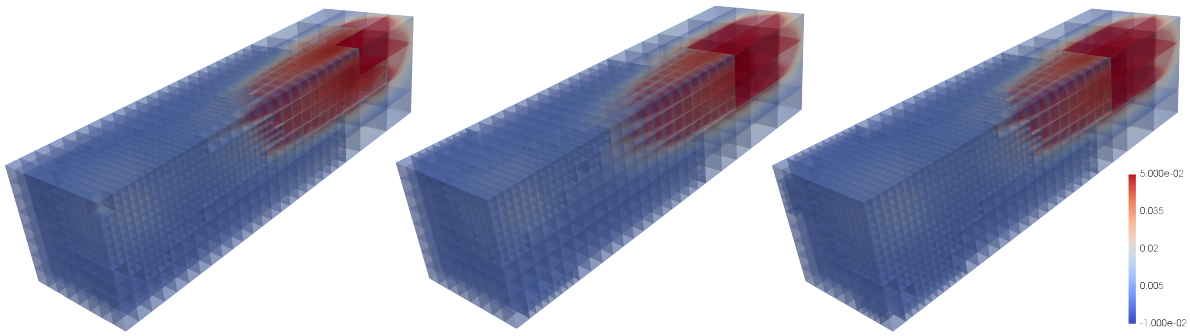


Figure 9.11: The solution after twelve refinement steps using the: (left) ad hoc GMR; (center) explicit GMR; (right) implicit GMR.

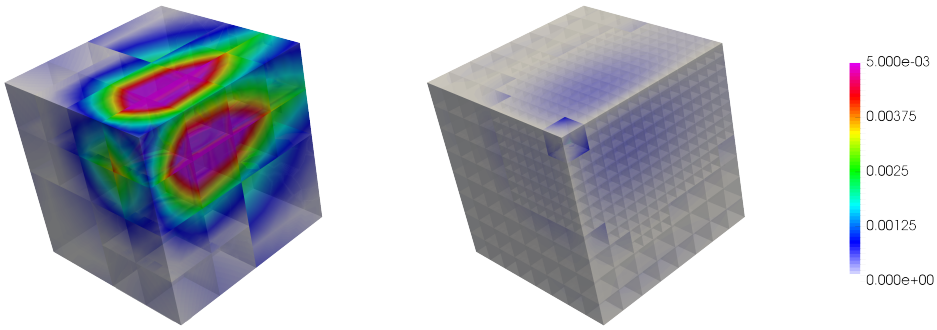


Figure 9.12: The error in the solution component u in $0 \leq x \leq 1$ at the final mesh using: (left) conventional DPG SMR; (right) ad hoc DPG GMR.

drive adaptivity with the intention of significantly reducing the error in a particular solution variable—in this case, it is the temperature u—in a *region of interest*. Although this can also be done more accurately by considering a nonlinear goal functional [110], simply using GMR with a closely related linear QOI often provides a sufficient improvement. With this understanding, Figure 9.12 clearly demonstrates that the total error in the temperature variable u in the region of interest is far lower as a result of the GMRs as opposed to the conventional SMR for a similar number of degrees of freedom. In Figure 9.12, we visualize the error from the ad hoc approach. The results from the other two GMR approaches were nearly indistinguishable upon visualization.

9.2.4 Flux in a subdomain

In this subsection, we consider the goal functional given in (146) where

$$g_1 = 0\,, \qquad \vec{g}_2(x,y,z) = \begin{cases} (1,0,0)^{\mathsf{T}}\,, & x \leq 1\,, \\ \vec{0}\,, & \text{otherwise,} \end{cases} \qquad \widehat{g}_3 = 0\,, \qquad \text{and} \qquad \widehat{g}_4 = 0\,. \tag{148}$$

Physically, this corresponds to a QOI which is the average value of the x-component of the flux, σ_x, in the subdomain $0 \leq x \leq 1$. In these experiments, homogeneous Dirichlet boundary conditions, $u|_{\partial\Omega} = 0$, were used.

Define the relative error in the QOI to be $|G(\vec{u} - \vec{u}_h)|/|G(\vec{u})|$. As with the previous experiment, we present the relative error in this QOI with each of the AMR strategies. Again,

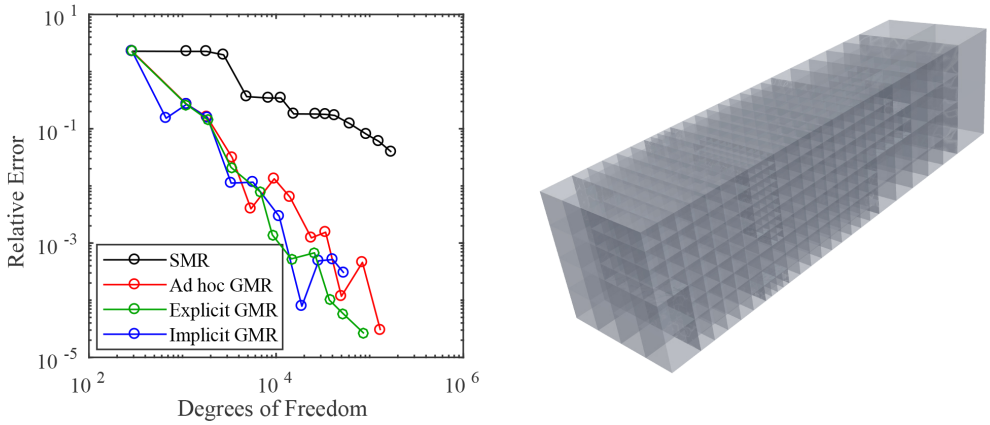


Figure 9.13: Left: The error in the QOI for the second example: average flux σ_x. Right: The final adaptively refined mesh using the explicit GMR approach.

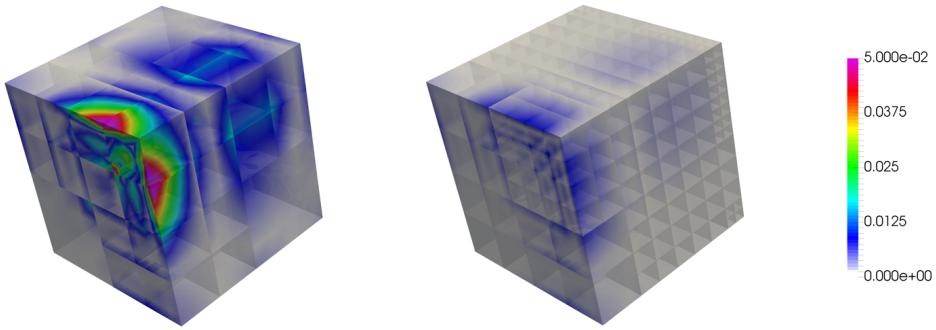


Figure 9.14: The final error in the solution component σ_x in the subdomain $0 \leq x \leq 1$ using: (left) SMR; (right) explicit GMR.

by inspecting Figure 9.13, it is clear that each of the GMR strategies are far more efficient than conventional DPG SMR.

For a visual comparison of the error in σ_x in the region of interest $0 \leq x \leq 1$, we provide Figure 9.14. Even though they both come from meshes inducing a similar number of degrees of freedom, notice that the local error from the GMR strategy is at least two orders of magnitude smaller than the local error from the SMR strategy.

9.2.5 Temperature on the boundary

In this subsection, consider the goal functional G given in (146) where

$$g_1 = 0\,, \qquad \vec{g}_2 = \vec{0}\,, \qquad \widehat{g}_3(x,y,z) = \begin{cases} 1\,, & x = 0\,, \\ 0\,, & \text{otherwise,} \end{cases} \qquad \text{and} \qquad \widehat{g}_4 = 0\,. \tag{149}$$

Physically, this corresponds to a QOI which is the average value of the temperature u on the subboundary $x = 0$. Here, homogeneous Dirichlet and homogeneous Neumann boundary conditions were used on disjoint regions of the boundary, $u|_{\Gamma_{\mathrm{D}}} = 0$ and $\frac{\partial u}{\partial n}|_{\Gamma_{\mathrm{N}}} = 0$, where $\Gamma_{\mathrm{D}} = \{(x,y,z) \in \partial\Omega \mid x > 0\}$ and $\overline{\Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{N}}} = \partial\Omega$.

An alternative functional for ad hoc refinement indicators

Given that the ad hoc refinement indicators $\eta_K^{*,\mathrm{a.h.}}$ were not derived for goal functionals involving nonzero $\widehat{g}_3$ or $\widehat{g}_4$, we actually used a different goal functional (150) for this indicator.

Let $\Gamma = \{(x,y,z) \in \partial\Omega \mid x = 0\}$ and define the set of all elements neighboring Γ as $\mathcal{T}_\Gamma = \{K \in \mathcal{T} \mid \overline{K} \cap \Gamma \neq \varnothing\}$. Lastly, define the subdomain occupied by this set of elements as $\Omega_\Gamma = \bigcup_{K \in \mathcal{T}_\Gamma} \overline{K}$. Now, instead of (149), consider the modified goal functional

$$g_1(\vec{x}) = \begin{cases} h_{K,x}^{-1}\,, & \vec{x} \in \Omega_\Gamma\,, \\ 0\,, & \text{otherwise,} \end{cases} \qquad \vec{g}_2 = \vec{0}\,, \qquad \widehat{g}_3 = 0\,, \qquad \text{and} \qquad \widehat{g}_4 = 0\,, \tag{150}$$

where, $h_{K,x}$ is the length, in the x-dimension, of the element $K \in \mathcal{T}_\Gamma$ enclosing the point $\vec{x} \in K$. Notice that because g_1 operates on the u-component of the solution in (146), as the mesh becomes finer near Γ, this functional will also limit to a characterization of the average temperature on the boundary. The primary novelty of (150) is that the definition is mesh-dependent. This is demonstrated in Figure 9.15, where Ω_Γ is highlighted in red on different adaptively refined meshes.

Results

Notice that $u|_{\partial\Omega} = 0$ from (145). Therefore, from the definition of G given in (149), $G(\vec{u}) = \int_{\Gamma_{\mathrm{N}}} u = \int_\Gamma u = 0$. Define the relative error in the QOI to be $\left|\int_\Gamma \widehat{u}_h\right| / \left|\int_\Gamma \widehat{u}_{h_{\mathrm{init}}}\right|$, where

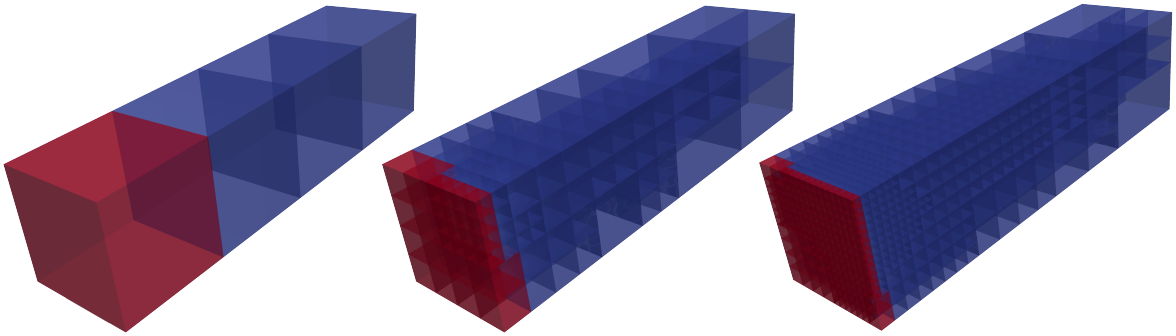


Figure 9.15: The region of interest Ω_Γ, marked in red, for: (left) the initial mesh; (center) the mesh after six GMR steps with the ad hoc approach; (right) the mesh after twelve ad hoc GMR steps.

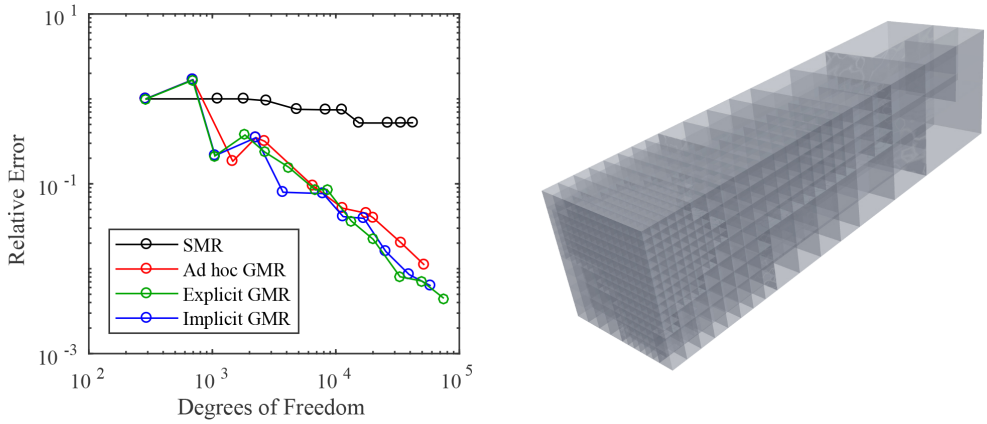


Figure 9.16: Left: The error in the QOI for the third example: average temperature $\widehat{u}$. Right: The final adaptively refined mesh using the ad hoc GMR approach.

$\widehat{u}_{h_{\text{init}}}$ is the third component of approximate solution $\vec{u}_h$ computed on the initial mesh, which all approaches having in common (see Figure 9.9 (left)). From only a cursory inspection of Figure 9.16, it is again evident that each of the GMR strategies were far more efficient than conventional SMR.

In Figure 9.17, the visual comparison given of the error in the temperature variable on the region of interest, $x = 0$, demonstrates a substantial improvement over the conventional SMR strategy, even with the ad hoc GMR approach employing the modified goal functional (150).

Figure 9.17: The final error in the solution $\widehat{u}$ on $x = 0$ from: (left) conventional SMR; (right) the ad hoc GMR approach.

9.2.6 Flux on the boundary

Consider the goal functional given in (146) where

$$g_1 = 0\,, \qquad \vec{g}_2 = \vec{0}\,, \qquad \widehat{g}_3 = 0\,, \qquad \text{and} \qquad \widehat{g}_4(x,y,z) = \begin{cases} 1\,, & x = 0\,, \\ 0\,, & \text{otherwise.} \end{cases} \tag{151}$$

Physically, this corresponds to the QOI being the average value of the flux $\vec{\sigma} \cdot \vec{n}$ through the subboundary $x = 0$. In our experiments, we used homogeneous Dirichlet boundary conditions everywhere, $u|_{\partial\Omega} = 0$.

Energy space considerations

In this experiment, $\Gamma_{\mathrm{D}} = \partial\Omega$ and so $\left(H_{\mathrm{N}}^{-1/2}(\partial\Omega)\right)' = H^{1/2}(\partial\Omega)$ and $H_0^1(\Gamma_D) = H^1(\partial\Omega)$. Notice that $\widehat{g}_4 \in L^2(\partial\Omega)$ but $\widehat{g}_4 \notin H^1(\partial\Omega)$. Therefore, the assumptions of (146) are not met. In fact, there is no prerequisite reason why G, as defined by (151), should even be a bounded linear functional on $\mathcal{U}$. Indeed, because it has a nontrivial jump discontinuity, $\widehat{g}_4 \notin H^{1/2}(\partial\Omega) \supseteq H^1(\partial\Omega)$ and so $G \notin \mathcal{U}'$.

Unfortunately, violating the energy setting is not simply a mathematical concern. Indeed, we found spuriously concentrated refinements near the discontinuity in $\widehat{g}_4$, when using (151). Therefore, instead of (151), we chose to mollify the physically ideal (but discontinuous) $\widehat{g}_4$ so that it obeys the proper energy setting. For our explicit and implicit GMR experiments, specifically, we used the ramp function depicted in Figure 9.18, in place of the function $\widehat{g}_4$ defined in (151).

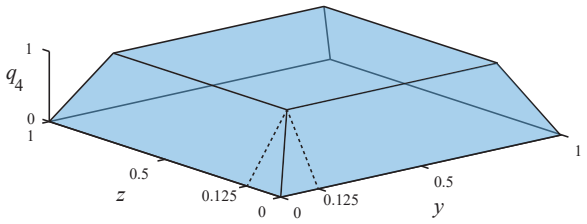


Figure 9.18: Illustration of the function $\widehat{g}_4 \in H^1(\partial\Omega)$ (on the face $x = 0$) used for explicit and implicit GMR.

Another alternative functional for ad hoc refinement indicators

Analogous to the previous experiment, here we considered an alternative to the goal functional (151) for the ad hoc DPG* refinement indicator. Namely, we used

$$(152) \qquad g_1 = 0 \qquad \vec{g}_2(\vec{x}) = \begin{cases} (h_{K,x}^{-1}, 0, 0)^{\mathsf{T}}, & \vec{x} \in \Omega_\Gamma, \\ \vec{0}, & \text{otherwise,} \end{cases}, \qquad \widehat{g}_3 = 0, \qquad \text{and} \qquad \widehat{g}_4 = 0,$$

with the same definitions for Ω_Γ and $h_{K,x}$ as in (150). Fortunately, with this definition, $\vec{g}_2 \in L^2(\Omega)$ and so $G \in \mathcal{U}'$ for all meshes and the energy space issues for (151) are again avoided.

Results

Define the relative error in the QOI to be $\left|\int_\Gamma(\vec{\sigma}\cdot\vec{n} - \widehat{\sigma}_h)\right| / \left|\int_\Gamma \vec{\sigma}\cdot\vec{n}\right|$. An inspection of Figure 9.19 clearly illustrates that the GMR strategies were far more efficient than the conventional SMR strategy for this QOI. In fact, with the conventional strategy, the error in this QOI did not even decrease until the eighth mesh refinement was performed!

As for the visual comparison of the error in the flux on the region of interest, Figure 9.20 again demonstrates a significant improvement over conventional DPG SMR.

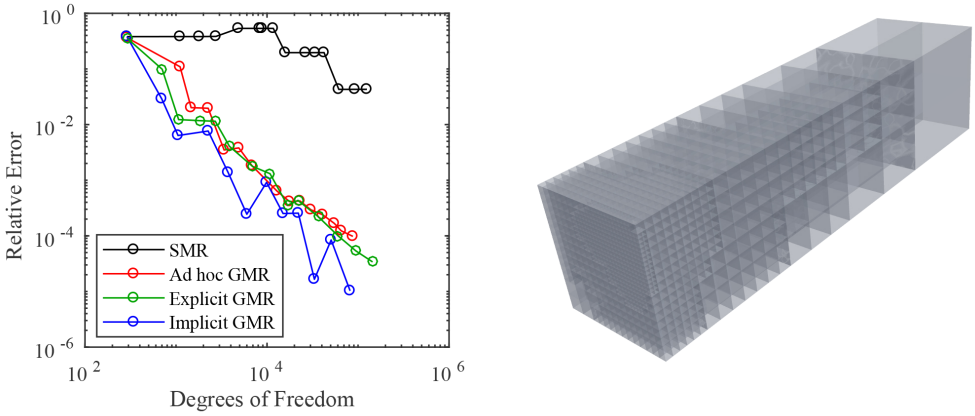


Figure 9.19: Left: The error in the QOI for the fourth example: average normal flux $\widehat{\sigma}$. Right: The final adaptively refined mesh using the implicit GMR approach.

Figure 9.20: The final error in the solution $\widehat{\sigma}$ at $x = 0$ using: (left) SMR; (right) implicit GMR.

9.2.7 Temperature at a point

In this final experiment, we considered the goal functional $G_{\vec{x}_0}(\vec{u}) = u(\vec{x}_0)$, where $\vec{x}_0 \in \Omega$ is a specified point in the domain. Markedly, this QOI does not fall into the theory of this dissertation because $G_{\vec{x}_0} \notin \mathcal{U}'$. To overcome this issue, we employ a mesh-dependent goal functional like (150) and (152). Define the set of all elements containing the point $\vec{x}_0$ in their closure as $\mathcal{T}_{\vec{x}_0} = \{K \in \mathcal{T} \mid \vec{x}_0 \in \overline{K}\}$ and define the subdomain occupied by this set of elements $\Omega_{\vec{x}_0} = \bigcup_{K \in \mathcal{T}_{\vec{x}_0}} \overline{K}$. Now, redefine the goal functional as

$$g_1(\vec{x}) = \begin{cases} \mathrm{vol}(\Omega_{\vec{x}_0})^{-1}\,, & \vec{x} \in \Omega_{\vec{x}_0}\,, \\ 0\,, & \text{otherwise,} \end{cases} \qquad \vec{g}_2 = \vec{0}\,, \qquad \widehat{g}_3 = 0\,, \qquad \text{and} \qquad \widehat{g}_4 = 0\,.$$

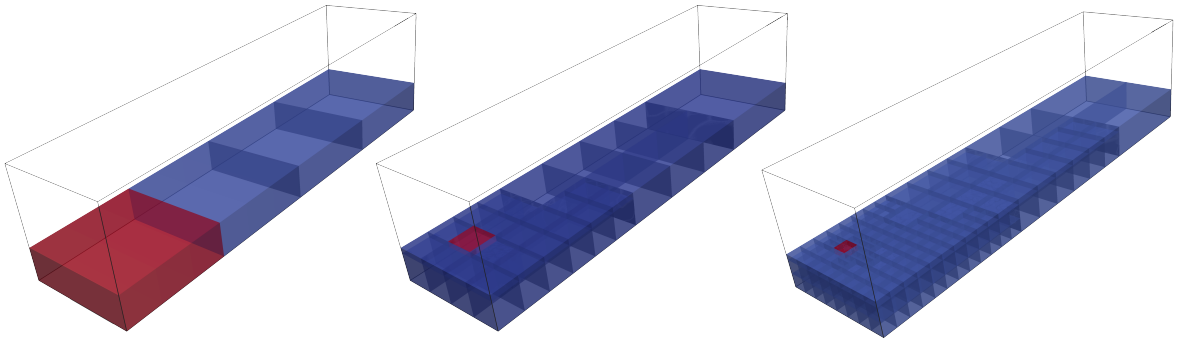


Figure 9.21: The region of interest $\Omega_{\vec{x}_0}$ enclosing the point $\vec{x}_0 = (0.3, 0.3, 0.3)^\intercal$ for: (left) the initial mesh; (center) the mesh after six refinements with the implicit approach; (right) the mesh after twelve refinements. For the sake of illustration, only $y \leq 0.3$ is shown.

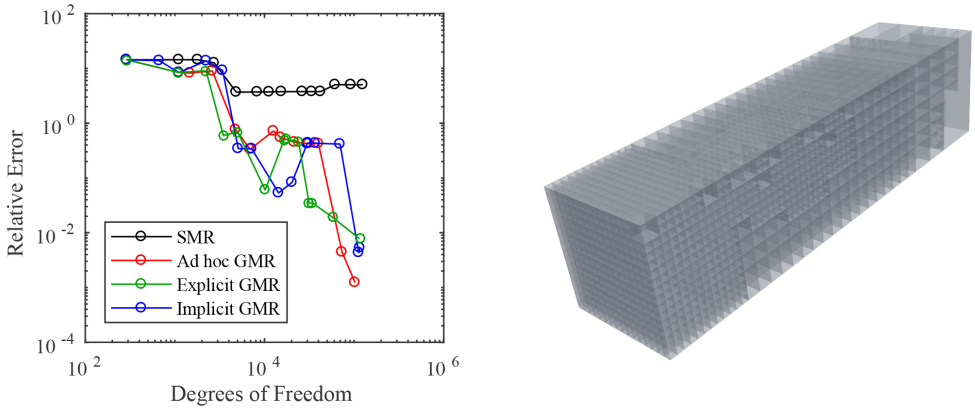


Figure 9.22: Left: The error in the QOI for the fifth example: temperature u at a fixed point. Right: the final adaptively refined mesh using the explicit GMR approach.

All results presented in this section are from our experiments with the mesh-dependent definition above and $\vec{x}_0 = (0.3, 0.3, 0.3)^\intercal$. The evolution of this functional is visually depicted in Figure 9.21, where the region of interest $\Omega_{\vec{x}_0}$ is highlighted in red on a selection of meshes.

Inspect Figure 9.22. Here, the relative error is defined to be $|u(\vec{x}_0) - u_h(\vec{x}_0)|/|u(\vec{x}_0)|$. As in every previous experiment, each GMR approach vastly outperformed conventional SMR. However, in this experiment, the convergence behavior of each GMR approach was quite erratic.

9.3 A DPG method for viscoelastic fluid flow

This section considers the verification of the DPG method proposed in Section 5.2 on the so-called confined cylinder problem, for which many results are available in the literature [1, 37, 39, 63, 86, 95, 98, 114, 134]. This is a two-dimensional problem where the viscoelastic fluid passes through a narrow channel with a centrally placed cylinder impeding its flow. In this case, the ratio of the cylinder radius to the channel width to the length of the domain is taken to be precisely 1 : 2 : 15, as depicted in Figure 9.23. All computations in this section were performed with Camellia [127, 128].

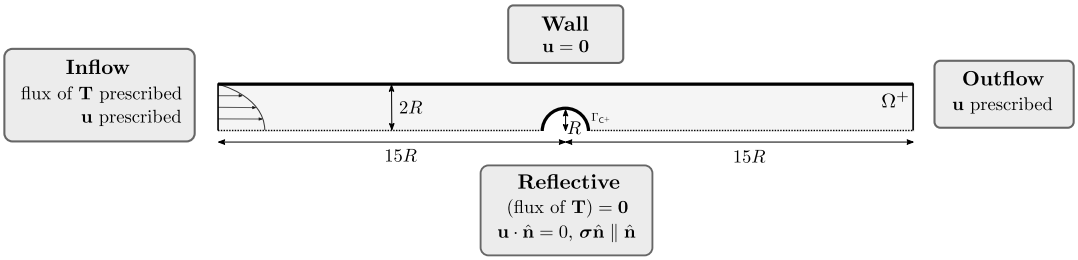


Figure 9.23: Diagram of flow domain with boundary names and the strong boundary conditions. Here, the prescribed flux of $\mathbf{T}$ is given in (154).

In this problem, the length of the channel is understood to be sufficiently large so that any physically reasonable outflow condition will have little effect upon the following quantity of interest, the drag coefficient:

$$C_{\mathrm{D}} = \frac{1}{\eta \bar{u}} \int_{\Gamma_{\mathsf{C}}} (\boldsymbol{\sigma} \boldsymbol{n}) \cdot \mathbf{e}_1 \,. \tag{153}$$

Here, $\bar{u}$ is the average inflow velocity and Γ_{C} is the entire cylinder boundary. The Cauchy stress, $\boldsymbol{\sigma}$, is defined in (81). In Figure 9.23, notice that we reduced the computational expense by considering only half of the flow domain with reflectively symmetric solutions. We thus compute on the upper-half of Ω, which we denote Ω^+.

This section proceeds by first further describing the boundary conditions applied to this problem. We then compare our results to the literature for a various values of the Weissenberg number Wi various values of the Reynolds number Re and various values of the mobility factor α. Of these comparisons, the first dedicated subsection examines the Oldroyd-B model without

Re	0, 0.01, 0.1, 1
Wi	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0
α	0, 0.001, 0.01, 0.1
η_{S}	0.59
η_{P}	0.41
p	2
$\mathrm{d}p$	2
θ	0.2
solver	MUMPS 5.0.1 [4, 99]

Table 9.1: Parameters used in this study.

inertial effects; i.e., Re = 0. This is the case for which the most extensive results appear in the literature and our comparisons are the most thorough. It is also in this subsection that we analyze our adaptive mesh refinement strategy and investigate the influence of this strategy on our estimates of C_{D}. In Sections 9.3.3 and 9.3.4 we use our techniques to examine the effects of non-zero Reynolds numbers followed by non-zero mobility factors. Comparisons with the literature are given in both cases. Lastly, in Section 9.3.5, we present very recent work from [92] using goal-oriented adaptive mesh refinement for this problem. Table 9.1 may serve as a reference for the parameters used in our experiments. In the experiments, we followed the adaptive mesh refinement strategies presented in Section 7.5 and consistently used the refinement threshold parameter $\theta = 0.2$.

9.3.1 Set-up

In this study, the Reynolds and Weissenberg numbers are defined as

$$\mathrm{Re} = \frac{\rho\,\bar{u}R}{\eta} \qquad \text{and} \qquad \mathrm{Wi} = \lambda\frac{\bar{u}}{R}\,,$$

respectively. As is standard in similar incompressible flow problems, we prescribed the inflow velocity to be that of the steady-state Poiseuille solution for flow in a channel, $\mathbf{u}(x,y) = \begin{pmatrix} u_1(y) \\ 0 \end{pmatrix}$, where $u_1(y) = \frac{3\bar{u}}{2}\big(1 - \frac{y^2}{4R^2}\big)$. Conveniently, this can be complemented with a simple solution for the extra-stress tensor, $\mathbf{T}(x,y) = \begin{pmatrix} T_{11}(y) & T_{12}(y) \\ T_{12}(y) & T_{22}(y) \end{pmatrix}$, where

$$T_{11}(y) = \frac{9\bar{u}^2\lambda\eta_{\mathrm{P}}y^2}{8R^4}\,, \qquad T_{12}(y) = \frac{-3\bar{u}\eta_{\mathrm{P}}y}{4R^2}\,, \qquad T_{22}(y) = 0\,.$$

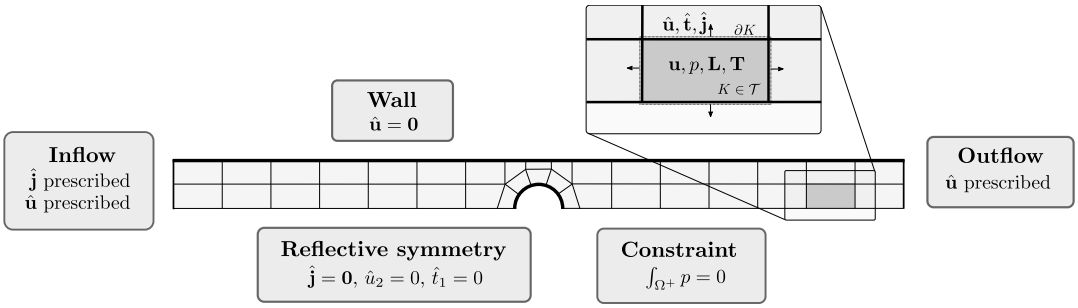


Figure 9.24: Initial mesh and boundary conditions. Note that $\widehat{\mathbf{u}} = \begin{pmatrix} \widehat{u}_1 \\ \widehat{u}_2 \end{pmatrix}$, $\widehat{\mathbf{t}} = \begin{pmatrix} \widehat{t}_1 \\ \widehat{t}_2 \end{pmatrix}$, and $\widehat{\mathbf{j}} = \begin{pmatrix} \widehat{j}_{11} & \widehat{j}_{12} \\ \widehat{j}_{12} & \widehat{j}_{22} \end{pmatrix}$.

In some related studies, the inflow value of $\mathbf{T}$ is prescribed (see, e.g., [95]). This is arguably unnatural since the flux term $\mathbf{T}\,\mathbf{u} \cdot \boldsymbol{n}$ actually appears on the inflow boundary through integration by parts. Prescribing the flux of $\mathbf{T}$ from the exact solution above, we set

$$\widehat{\mathbf{j}} = \mathbf{T}\,\mathbf{u} \cdot \boldsymbol{n} = -\mathbf{T} u_1 \tag{154}$$

at the inflow boundary. Notice that the dot product $\mathbf{u} \cdot \boldsymbol{n}$ may vanish at points on the boundary where $\mathbf{T}$ does not.

We chose the boundary conditions at the reflective boundary from physical intuition and motivation from the inflow boundary prescription. First of all, due to the symmetry of the solution, we anticipate zero *flux* of the extra stress tensor across the boundary. The logic for this is simple since any flux vector should exist in equal magnitude, but *reflected direction*, at the opposing point across the boundary. Following from this principle, the normal component of the flux must vanish at points where the flux is equal to its reflection, and so at all points on the reflective boundary, $\widehat{\mathbf{j}} = \mathbf{0}$. Likewise, the velocity of the fluid normal to the boundary, the *mass flux*, must also vanish, $\mathbf{u} \cdot \boldsymbol{n} = 0$. Indeed, we could conclude that the flux of $\mathbf{T}$ vanishes at the reflective boundary, plainly from this relationship on the fluid velocity, $\widehat{\mathbf{j}} = \mathbf{T}\,\mathbf{u} \cdot \boldsymbol{n} = \mathbf{0}$. The final boundary condition at the reflective boundary—the stress vector, $\widehat{\mathbf{t}} = \boldsymbol{\sigma}\boldsymbol{n}$, being parallel to the boundary normal can be argued similarly.

At the outflow boundary, many different choices of boundary conditions are possible for this problem. We have chosen to present results from experiments with a fully prescribed

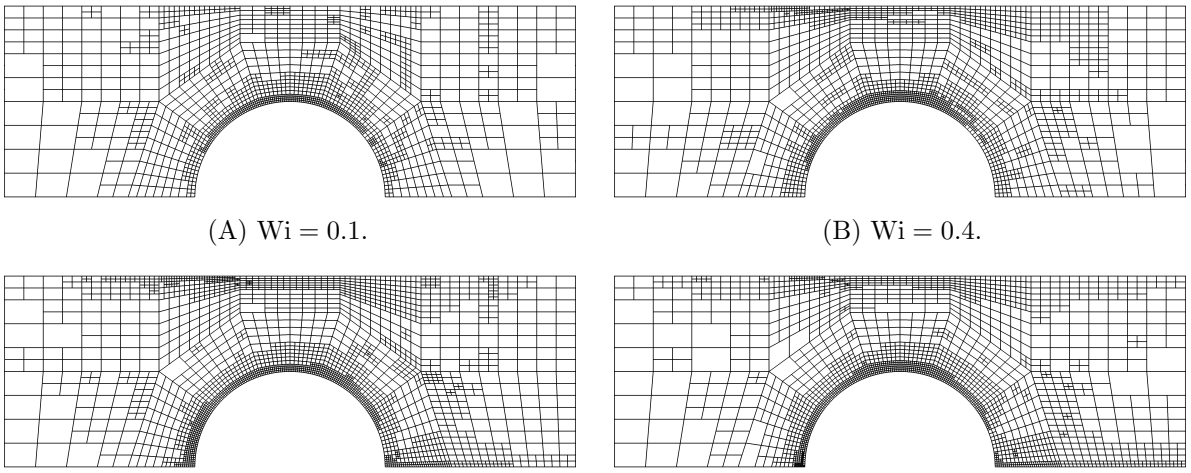
outflow velocity field. Other choices such as zero outflow traction or zero tangential velocity and zero normal stress were also tried; however, neither had any discernible influence upon the drag coefficient estimates in our experiments.

Finally, boundary conditions at the walls of the channel and obstructing cylinder were simply taken to be of the standard no-slip type, $\mathbf{u} = \mathbf{0}$. Here, the flux of the extra stress tensor was not prescribed. Of course, prescribing the velocity field at all boundaries of the computational domain necessitated introducing a uniqueness constraint on the pressure of the system. In our experiments, we chose to enforce a zero-average pressure, $\int_{\Omega^+} p = 0$.

Given that the DPG method generally requires boundary conditions to be prescribed through the interface variables, we have reinterpreted them in Figure 9.24. Here, we also depict the initial mesh used for each simulation. We always began our simulations with the same 36 elements and a zero initial guess for the solution. After each mesh refinement, the previous converged solution was projected onto the new mesh and used as a new initial guess. In doing this, we did not require any parameter continuation to achieve our results (cf. [95]).

9.3.2 Creeping flow with the Oldroyd-B model

Here, we present the computed values of the drag coefficient for the Oldroyd-B model with Stokes flow coupling. For select values of the Weissenberg number, Table 9.2 displays the precise evolution of the drag coefficient estimates under the growing mesh. Each computation ended when the MUMPS direct solver failed on our system. These failures could have been caused by insufficient computing resources for the problem size (an eventual barrier for all discretizations), ill conditioning, or simply ill-posedness and instability. Below, we attempt to determine, based on the data, the likely reason for each failure at each Wi. We also highlight the fact that the number of elements added to the mesh after each successive refinement could vary greatly with the Weissenberg number. With this in mind, even after just a handful of refinements, the mesh structure also varied greatly—as is demonstrated in Figure 9.25—with refinements in the wake of the cylinder becoming more pronounced with growing Wi. Figure 9.26 depicts the profiles of $\mathbf{T}$ around the cylinder when $\mathrm{Wi} = 0.4$.



(A) Wi = 0.1. (B) Wi = 0.4.

(C) Wi = 0.7. (D) Wi = 1.

Figure 9.25: Close-up of meshes from the energy refinement strategy after five refinements for select values of the Weissenberg number.

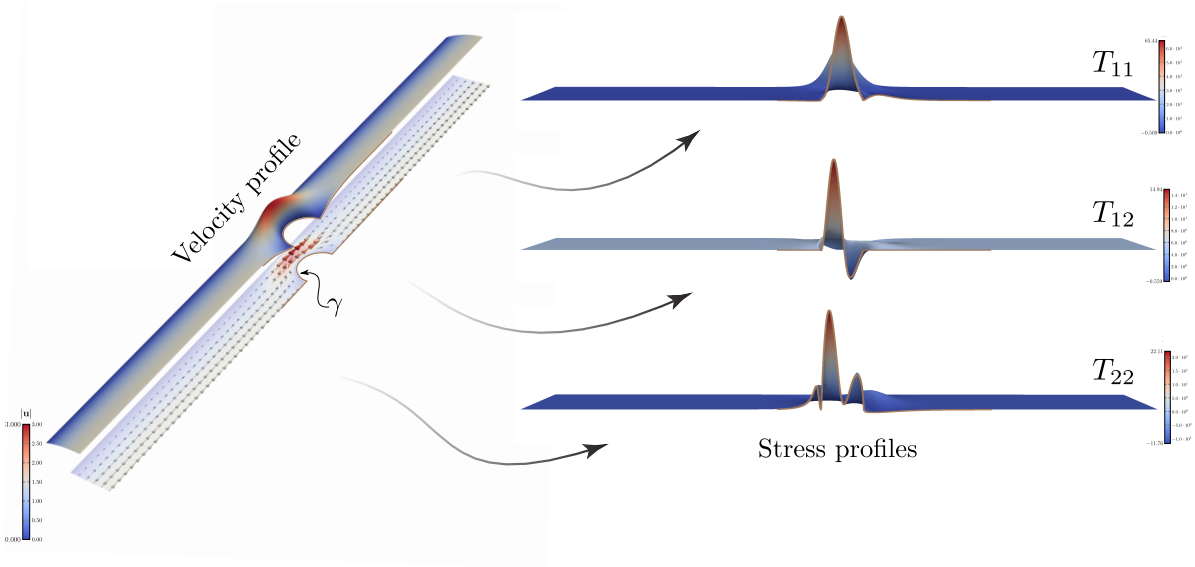


Figure 9.26: Extra-stress tensor components for Weissenberg number Wi = 0.4 in the Oldroyd-B fluid model. Computation from the tenth refinement with the energy strategy (see Table 9.2). Displayed in the velocity profile is a surface plot of the velocity magnitude underlaid by a vector field plot showing the magnitude and direction of the velocity. The brown curve defines $\boldsymbol{\gamma}$, along which the stress components are sampled in the coming figures.

Ref. #	Wi = 0.1					Wi = 0.4				
	DoF	Drag coefficient (C_{D})		Error estimate		DoF	Drag coefficient (C_{D})		Error estimate	
		Field	Flux	Drag	Energy		Field	Flux	Drag	Energy
0	5123	55.6760	56.1664	2.1454	0.3734	5123	33.9969	34.4629	1.7955	0.4296
1	19604	94.8348	95.0995	1.2605	0.2214	19604	91.1569	91.9478	2.0618	0.2358
2	24242	122.1375	122.1917	0.5008	0.09841	30893	110.3167	110.6781	0.8045	0.09273
3	42488	129.4744	129.5010	0.1501	0.03008	60839	116.6999	116.8479	0.3542	0.03740
4	94265	130.2408	130.2491	0.04778	0.009435	108599	119.5058	119.5410	0.1114	0.01670
5	229358	130.3381	130.3420	0.01543	0.003150	296321	120.4220	120.4297	0.03248	0.005749
6	571787	130.3566	130.3584	0.006452	0.001031	648935	120.5616	120.5642	0.01102	0.002306
7	1122647	130.3608	130.3616	0.003087	0.0004299	1265147	120.5818	120.5837	0.007196	0.001052
8	2477081	130.3620	130.3624	0.001540	0.0001561	1923035	120.5879	120.5884	0.003450	0.0006169
9	4067018	130.3624	130.3626	0.001950	0.00008547	3000239	120.5897	120.5902	0.002383	0.0003567
10						3920696	120.5904	120.5906	0.001718	0.0002493

Ref. #	Wi = 0.7					Wi = 1.0				
	DoF	Drag coefficient (C_{D})		Error estimate		DoF	Drag coefficient (C_{D})		Error estimate	
		Field	Flux	Drag	Energy		Field	Flux	Drag	Energy
0	5123	31.5416	32.1447	1.9049	0.4374	5123	33.9760	34.6702	2.0224	0.4322
1	19604	85.4511	85.9790	1.3912	0.2057	19604	79.9033	80.3355	1.1545	0.2037
2	34076	105.0442	105.1755	0.4563	0.08371	36110	103.2292	103.2436	0.4292	0.09038
3	78146	112.5455	112.6402	0.2461	0.03462	80138	115.7134	115.7109	0.2457	0.04052
4	187970	115.0789	115.1318	0.1421	0.01771	204866	119.0960	119.1157	0.1127	0.02097
5	373802	116.6757	116.6810	0.04498	0.008825	353648	118.8900	118.8965	0.06567	0.02599
6	670955	116.9821	116.9929	0.03149	0.005488	362882	118.8117	118.8179	0.06386	0.01382
7	1137575	117.2007	117.2027	0.01587	0.003770	759218	118.3835	118.3891	0.04660	0.05420
8	1273412	117.1716	117.1739	0.01537	0.003152	834446	118.5343	118.5362	0.03567	0.009143
9	1683764	117.2265	117.2285	0.01433	0.002573	1169099	118.0705	118.0741	0.02763	0.007946
10	1785419	117.2284	117.2303	0.01437	0.002455	1427945	117.9592	117.9617	0.02655	0.005679
11	1860731	117.2319	117.2340	0.01438	0.002380	1643489	117.9875	117.9877	0.02284	0.004940
12	1904042	117.2345	117.2365	0.01433	0.002307	1955600	118.0815	118.0818	0.02638	0.004513
13	1955027	117.2379	117.2396	0.01367	0.002224	2406731	118.0868	118.0870	0.02369	0.005986
14	2007593	117.2395	117.2413	0.01369	0.002148					
15	2069720	117.2412	117.2430	0.01371	0.002066					
16	2134904	117.2435	117.2455	0.01517	0.001987					
17	2217206	117.2446	117.2466	0.01523	0.001775					
18	2929895	117.2890	117.2899	0.005264	0.001295					
19	3048377	117.2919	117.2928	0.005258	0.001240					
20	3130976	117.2925	117.2934	0.005254	0.001205					
21	3201977	117.2937	117.2946	0.005266	0.001166					
22	3262238	117.2942	117.2952	0.005278	0.001140					
23	3320939	117.2945	117.2955	0.005341	0.001115					
24	3379862	117.2951	117.2961	0.005332	0.001092					
25	3444287	117.2958	117.2968	0.005264	0.001070					
26	3539396	117.2971	117.2981	0.005215	0.001038					
27	3630017	117.2977	117.2987	0.005201	0.001010					
28	3717284	117.2984	117.2994	0.005165	0.0009821					
29	3806375	117.2990	117.3000	0.005107	0.0009549					
30	3893738	117.3001	117.3011	0.004956	0.0008700					

Table 9.2: Computed drag coefficient values are given above for Weissenberg numbers $\mathrm{Wi} = 0.1, 0.4, 0.7,$ and 1.0. We give the drag coefficient as computed from the flux variable $\widehat{\mathbf{t}}$, as well as the drag coefficient as computed from the field variables. Reduction in the energy error as well as in the L^2 drag error is usually observed as each mesh is refined.

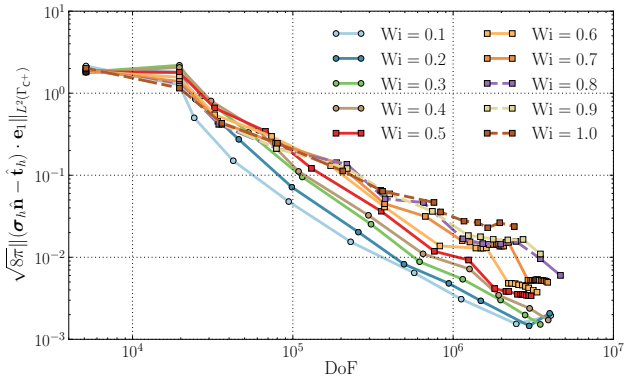


Figure 9.27: Convergence in drag error estimate versus degrees of freedom for the energy strategy. Markers indicate values at each refined mesh.

Notice that two different estimates of the drag coefficient are presented in Table 9.2. The explanation for this is that by computing both the traction, $\widehat{\mathbf{t}}$, as well as the field variable stress components, p, $\mathbf{L}$, and $\mathbf{T}$, we can make two *different* estimates of the drag coefficient:

(Flux) $$C_{\mathrm{D}}(\widehat{\mathbf{t}}_h) = \frac{2}{\mu \bar{u}} \int_{\Gamma_{\mathsf{C}+}} \widehat{\mathbf{t}}_h \cdot \mathbf{e}_1 \, ,$$

and

(Field) $$C_{\mathrm{D}}(\boldsymbol{\sigma}_h \boldsymbol{n}) = \frac{2}{\mu \bar{u}} \int_{\Gamma_{\mathsf{C}+}} (\boldsymbol{\sigma}_h \boldsymbol{n}) \cdot \mathbf{e}_1 \, ,$$

where $\boldsymbol{\sigma}_h = -p_h \mathbf{I} + \eta_{\mathrm{P}}(\mathbf{L}_h + \mathbf{L}_h^{\mathsf{T}}) + \mathbf{T}_h$. The first we call the *flux* estimate; and the second we call the *field* estimate.

Having two different estimates of the drag coefficient motivates a new *extrinsic* estimate of solution error. Define the drag error estimate

(155) $$\mathfrak{E}_{C_{\mathrm{D}}} = |\Gamma_{\mathsf{C}}|^{1/2} \| (\widehat{\mathbf{t}}_h - \boldsymbol{\sigma}_h \boldsymbol{n}) \cdot \mathbf{e}_1 \|_{L^2(\Gamma_{\mathsf{C}})} \, .$$

We anticipate that for smooth enough solutions, this error will converge to zero. Figure 9.27 presents the behavior of this value as the mesh was refined for various values of Wi. Here, we see a relatively steady decrease in the drag error with refinement for $0.1 \leq \mathrm{Wi} \leq 0.4$, a lower rate and growing number of mesh refinements in the range $0.5 \leq \mathrm{Wi} \leq 0.7$, and progressively poorer results in the range $0.8 \leq \mathrm{Wi} \leq 1$.

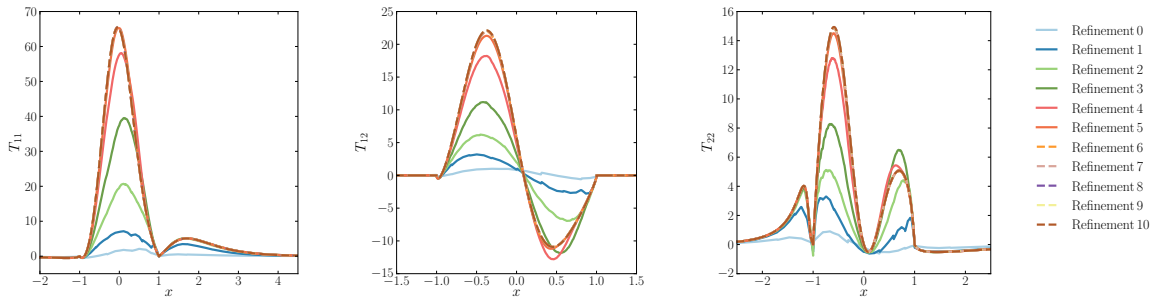


Figure 9.28: Convergence with mesh refinements in components of $\mathbf{T}$ along the curve $\boldsymbol{\gamma}$ (see Figure 9.26) for $\mathrm{Wi} = 0.4$.

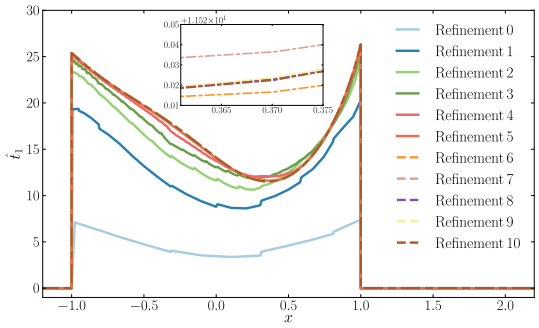


Figure 9.29: Convergence with mesh refinements in the first component of $\widehat{\mathbf{t}}$ along $\boldsymbol{\gamma}$ for $\mathrm{Wi} = 0.4$.

Additionally, having two available estimates, we attempted to see if one was generally more accurate than the other. Our inspection was both qualitative as well as quantitative. First, by plotting the profiles of the extra-stress components along the curve $\boldsymbol{\gamma}$ presented in Figure 9.26, we inspected the convergence with mesh refinement of $\mathbf{T}$ compared to $\widehat{\mathbf{t}}$ for several values of the Weissenberg number. As seen by comparing Figures 9.28 and 9.29, the profile of t_1 was generally less variable than any single component of $\mathbf{T}$. This suggests that $C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$ could be more accurate simply because the accumulation of relative error in the field variables (used to form $\boldsymbol{\sigma}_h$) is avoided. Another justification can be found by a simple examination of the theoretical energy spaces but we will not explore this here. It was, of course, the empirical evidence that was the strongest suggestion that $C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$ is the most accurate estimate. By observing the convergence behavior of both estimates and comparing them with the drag coefficient values reported in the literature, we found that the flux estimate was always slightly more accurate. In

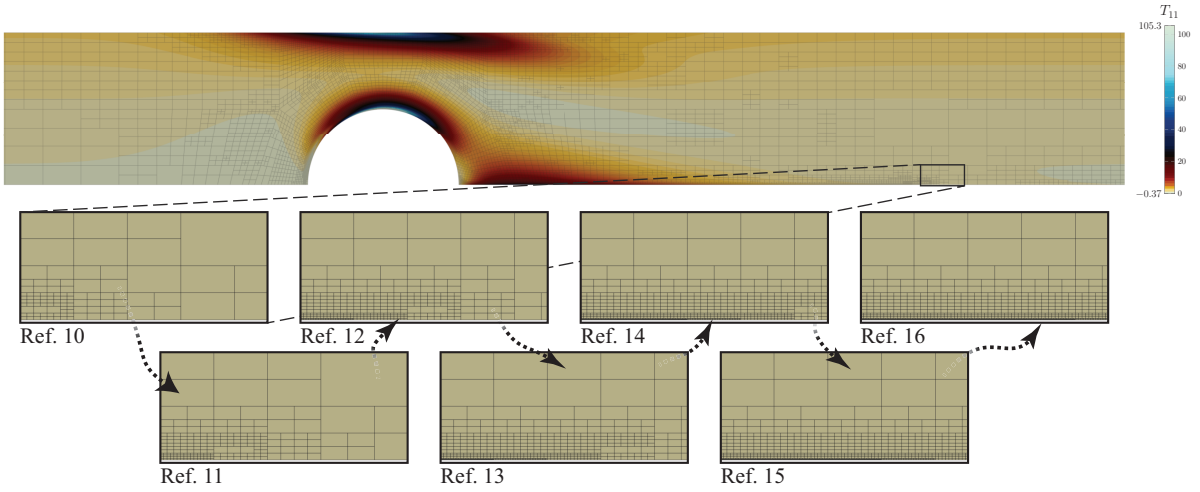


Figure 9.30: Close-up of a sequence of meshes from the Wi = 0.7 computation with the energy strategy. Notice the slow and sequential development of the mesh along the reflective boundary.

fact, although we are not sure of the reason, the field estimate was always slightly smaller than the flux estimate and, moreover, both approximations appeared to always converge to a steady value *from below*. Therefore, from now on, we only present the computed values of $C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$ in our results; however, we will continue to display the error estimate (155).

The accuracy of the drag coefficient for Weissenberg numbers in the range $0.5 \leq \mathrm{Wi} \leq 0.7$ was less per degree of freedom than for the range $0 \leq \mathrm{Wi} \leq 0.4$. Some explanation for this can be given simply in the refinement pattern our chosen strategy delivered. Figure 9.30 demonstrates the growth of the mesh with the energy strategy for $\mathrm{Wi} = 0.7$. Because the majority of refinements depicted there are performed downstream, it is obvious that the energy error in this region in the wake of the cylinder eventually controls the mesh growth. For this reason, after a certain number of refinements, the mesh was rarely again refined near the cylinder and it is of little surprise that eventually the drag coefficient estimate was no longer strongly improved by each subsequent adaptive mesh refinement. It is well known that T_{11} develops a strong internal layer in the wake of the confined cylinder as the Weissenberg number grows. This is depicted in Figure 9.31 and explains the refinement pattern we saw. Figure 9.32 demonstrates the evolution of this component of $\mathbf{T}$ along $\boldsymbol{\gamma}$ as the mesh was refined when $\mathrm{Wi} = 0.7$. For this Weissenberg

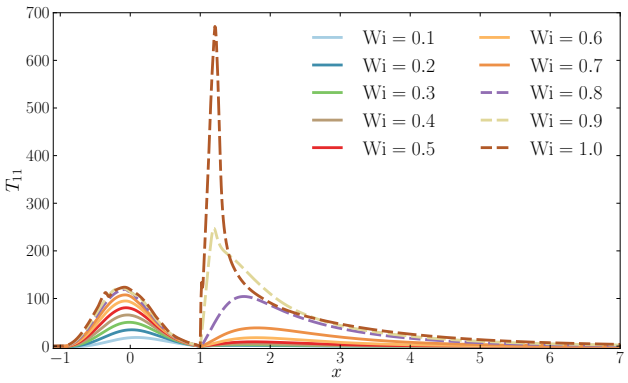


Figure 9.31: Profile curves for the T_{11} component of the extra stress tensor along $\boldsymbol{\gamma}$ among all Weissenberg numbers. Values taken from the energy strategy solutions at their final meshes (see Table 9.3).

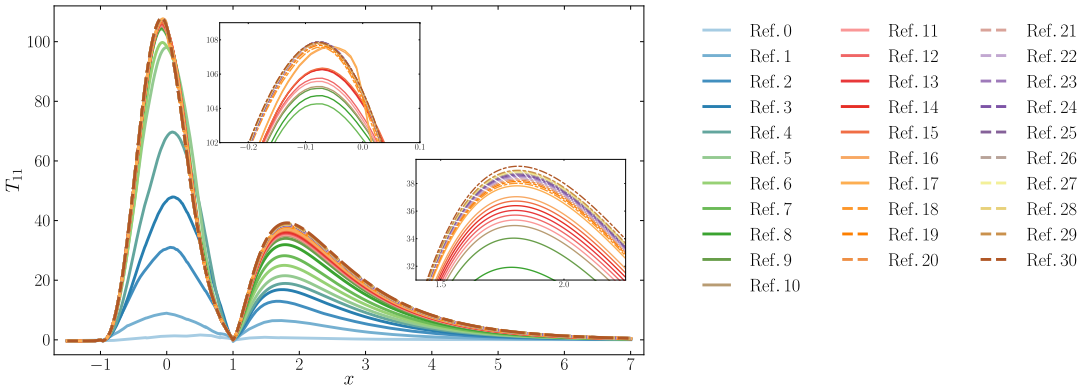


Figure 9.32: Convergence with mesh refinements in T_{11} along $\boldsymbol{\gamma}$ for $\mathrm{Wi} = 0.7$.

number, this internal layer is known to be very difficult to reliably capture and our results, in accord with much of the literature, do not show convergence of the profile of T_{11} in the wake of the cylinder.

As the drag coefficient is measured from solution values on the cylinder, we tested two ad-hoc refinement strategies which would necessarily refine the mesh more often near the cylinder and so hopefully increase the accuracy of our estimates. In Strategy #1, we began with the same energy strategy exhibited above, with the same threshold parameter $\theta = 0.2$, except that at every step we also refined every element with an edge lying on the cylinder boundary (whether or not it was originally scheduled for refinement). In Strategy #2, we similarly began

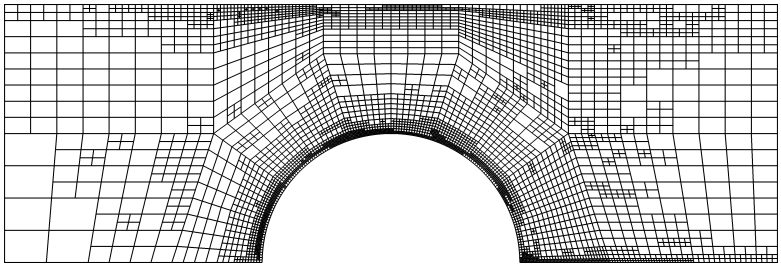
Wi	Energy strategy		Ad-hoc strategy #1		Ad-hoc strategy #2		Drag coefficient				
	DoF(#Refs)	$C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$	DoF(#Refs)	$C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$	DoF(#Refs)	$C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$	[95]	[37]	[86]	[114]	[63]
0.1	4067018(9)	130.3626	2630114(8)	130.3625	2649878(7)	130.3618	130.3626	130.364	130.363		130.36
0.2	4001165(10)	126.6251	3358379(9)	126.6210	2504387(7)	126.6241	126.6252	126.626	126.626		126.62
0.3	3498518(10)	123.1909	2214719(8)	123.1904	2573072(7)	123.1897	123.1912	123.192	123.193		123.19
0.4	3920696(10)	120.5906	2172425(8)	120.5889	2777531(7)	120.5885	120.5912	120.593	120.592		120.59
0.5	3065843(17)	118.8229	2093630(8)	118.8150	2673770(7)	118.8132	118.8260	118.826	118.836	118.827	118.83
0.6	3338165(19)	117.7687	1858451(8)	117.7370	2798306(7)	117.7581	117.7752	117.776	117.775	117.775	117.77
0.7	3893738(30)	117.3011	1649231(8*)	117.1923	2606123(7*)	117.2951	117.3157	117.316	117.315	117.291	117.32
0.8	4672934(12*)	117.2973	1888487(8*)	117.2091	2597321(7*)	117.3057	117.3454	117.368	117.373	117.237	117.36
0.9	3503723(12*)	117.5502	1847429(8*)	117.5248	2607365(7*)	117.6907	117.7678	117.812	117.787	117.503	117.79
1.0	2365391(14*)	118.0873	930626(7*)	118.7843	2506139(7*)	118.5970		118.550	118.501	118.030	118.49

Table 9.3: Comparison with results in literature with Stokes flow coupling. The superscript-* indicates that second order convergence was not reached in the final mesh.

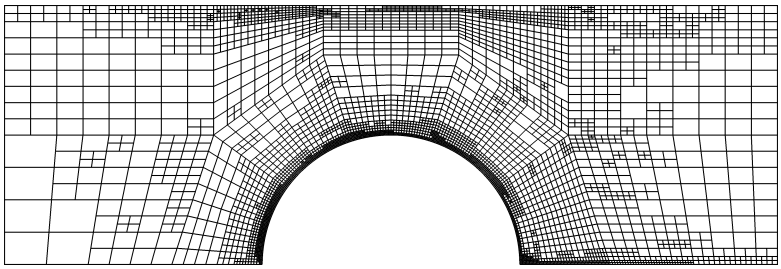
with the energy error strategy except that at every step we enforced the refinement of each element with and edge lying within a distance of 0.1 from the cylinder boundary. Close-ups of the sixth refined meshes for each of the three strategies when $\mathrm{Wi} = 0.7$ are given in Figure 9.33.

Unfortunately, the ad-hoc strategies that we tested introduced new issues of their own. In the first strategy, the relative scales of element sizes in the later meshes produced conditioning issues that led to earlier failures in our solver. In the second strategy, the size of the narrow band about which we were enforcing mesh refinements was just large enough that all of our computations failed upon attempting the eighth refined mesh. In this second scenario, a slightly thinner band would likely have returned a more desirable final mesh; one with few enough degrees of freedom that our solver would not have crashed and we would have gotten a more accurate drag coefficient for our final data point. Determining the optimal band length was eventually abandoned as it was not in line with our research interests. In conclusion: the energy strategy is not optimal for developing an accurate estimate of the drag coefficient. A goal-oriented approach would have been more desirable in this context and such an approach is described in Section 9.3.5.

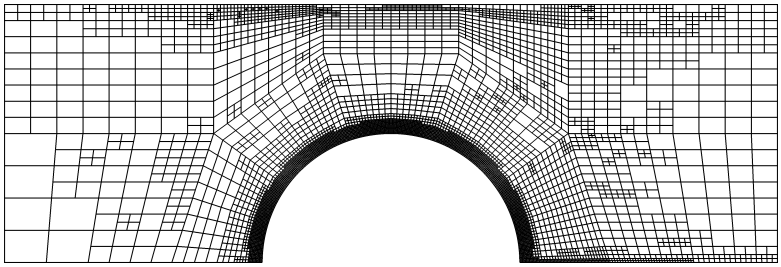
Table 9.3 compares the computed values of the drag coefficient with each of the three different adaptive strategies. Here, we see the best agreement with the literature in the energy strategy and therefore—while still recognizing its flaws—we decided to use it exclusively in our other studies.



(A) Energy strategy.



(B) Ad-hoc strategy #1.



(C) Ad-hoc strategy #2.

Figure 9.33: Meshes near the cylinder boundary Γ_{C^+} from the three different non-goal oriented refinement strategies after six adaptive mesh refinements. $\mathrm{Wi} = 0.7$. At this point, the only large differences between the meshes are near the cylinder boundary.

A note must be made on our computations for Weissenberg numbers between 0.8 and 1. In this parameter interval, our nonlinear iterations failed to establish the expected quadratic rate of convergence in the Newton iterations as the mesh was being developed. Indeed, usually for coarse meshes, quadratic convergence of our Newton iterations was easily attained for all studied values of Wi. Only as the meshes were adaptively refined did the expected rate of converge falter for large Weissenberg numbers. This was true regardless of the refinement strategy we considered. Some reasons for this could possibly include degeneration of solution regularity but—considering similar results in the literature—most likely indicates that the problem we

Wi	Re = 0.01				Re = 0.1				Re = 1			
	DoF(#Refs)	Drag coefficient		Error	DoF(#Refs)	Drag coefficient		Error	DoF(#Refs)	Drag coefficient		Error
		$C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$	[37]	$\mathfrak{E}_{C_{\mathrm{D}}}$		$C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$	[37]	$\mathfrak{E}_{C_{\mathrm{D}}}$		$C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$	[37]	$\mathfrak{E}_{C_{\mathrm{D}}}$
0.1	4065047(9)	130.3628	130.364	0.0019	4208813(9)	130.3667	130.368	0.0019	4146473(9)	130.6075	130.609	0.0020
0.2	3993227(10)	126.6259	126.627	0.0020	4635707(11)	126.6352	126.636	0.0023	2849903(9)	126.9366	126.938	0.0015
0.3	3489548(10)	123.1926	123.194	0.0015	3691595(10)	123.2095	123.211	0.0014	3760697(10)	123.5957	123.597	0.0014
0.4	3933806(10)	120.5933	120.595	0.0019	4666562(10)	120.6197	120.622	0.0016	4485530(10)	121.1038	121.106	0.0017
0.5	3097631(17)	118.8269	118.831	0.0031	4472378(14)	118.8654	118.868	0.0024	3573788(12)	119.4567	119.460	0.0027
0.6	3718352(13)	117.7752	117.781	0.0030	3535721(13)	117.8234	117.831	0.0032	4364192(12)	118.5380	118.542	0.0024
0.7	3900887(30)	117.3079	117.323	0.0049	3924098(30)	117.3717	117.387	0.0050	3982037(13)	118.2221	118.233	0.0042
0.8	3503915(11*)	117.2765	117.379	0.0095	3587420(14*)	117.3533	117.459	0.0113	3306764(11)	118.3971	118.455	0.0092
0.9	3560189(12*)	117.5592	117.827	0.0110	3165626(22*)	117.6836	117.925	0.0126	4106792(14)	118.8741	119.096	0.0134
1.0	2708810(15*)	118.1303	118.563	0.0281	2436611(13*)	118.2237	118.697	0.0245	3063377(13)	120.1455	120.057	0.0260

Table 9.4: Comparison with results in literature for the Oldroyd-B model with Navier-Stokes coupling. The superscript-* indicates that second order convergence was not reached in the final mesh. Notice that quadratic convergence was obtained for all Weissenberg numbers only in the case of the **largest** Reynolds number considered.

are solving is no longer well-posed. Indeed, some researchers have indicated that the solution becomes transient in this interval [37]. Others have indicated that the problem may be entirely ill-posed in this range due to the model allowing infinite extension of the viscoelastic fluid under finite elongation rates [86,95]. Ultimately, we consider each result for the interval of Weissenberg numbers $0.8 \leq \mathrm{Wi} \leq 1$—as well as for all our later results where quadratic convergence was not exhibited—dubious.

The loss of quadratic convergence indicates an issue, not in the resolution or stability of the problem, but either in our discretization or in the underlying well-posedness itself. In order to gain further algorithmic insight, we have continued our analysis for the inertial Oldroyd-B model and the Giesekus model.

9.3.3 Effects of inertia in the Oldroyd-B model

In this subsection we investigate the effects of the advective term in the kinematic equations upon the Oldroyd-B model. Here, the drag coefficient was computed using the same energy strategy described above for Reynolds numbers $\mathrm{Re} = 0.01$, 0.1, and 1. The results are collected together in Table 9.4. For a fixed Weissenberg number, the drag coefficient *grows* with the Reynolds number due to increasing velocity gradients.

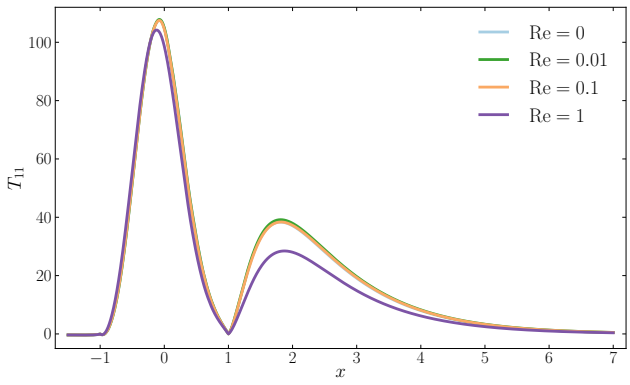


Figure 9.34: Profile of the T_{11} component of the extra stress tensor for $\mathrm{Wi} = 0.7$ along $\boldsymbol{\gamma}$ with various Reynolds numbers.

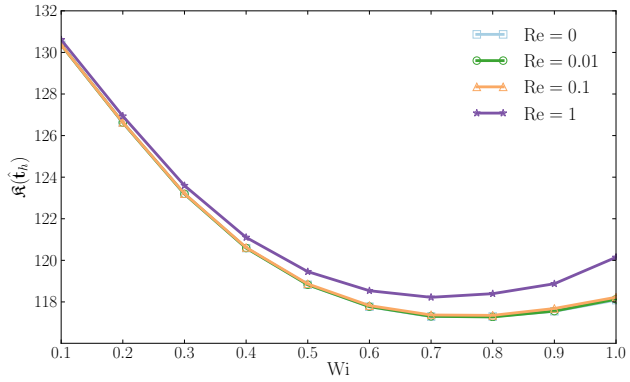


Figure 9.35: Dependence of the drag coefficient upon Reynolds number for $\mathrm{Wi} = 0.7$.

Perhaps surprisingly, we see that our results match most closely with those in the literature as the Reynolds number is increased. The reason for this is suggested in Figure 9.34 where the profile of T_{11} is plotted for each Reynolds number. Notably, as the Reynolds number grows, the values of T_{11} decrease in the wake of the cylinder much faster than on the cylinder surface. This suggests that a smaller proportion of the elements will be marked for refinement downstream as Re grows, resulting in more accurate drag coefficient estimates. An inspection of the various meshes (not shown) verified this hypothesis.

Figure 9.35 illustrates how the drag coefficient changes as the Reynolds number is increased. We see that the behavior is nearly identical until $\mathrm{Re} = 1$.

Wi	$\alpha = 0.001$				$\alpha = 0.01$				$\alpha = 0.1$			
	DoF(#Refs)	Drag coefficient		Error	DoF(#Refs)	Drag coefficient		Error	DoF(#Refs)	Drag coefficient		Error
		$C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$	[37]	$\mathfrak{E}_{C_{\mathrm{D}}}$		$C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$	[37]	$\mathfrak{E}_{C_{\mathrm{D}}}$		$C_{\mathrm{D}}(\widehat{\mathbf{t}}_h)$	[37]	$\mathfrak{E}_{C_{\mathrm{D}}}$
0.1	4016738(9)	130.2894	130.291	0.0019	3872162(9)	129.6696	129.671	0.0018	4307282(12)	125.5871	125.587	0.0058
0.2	3991931(10)	126.3946	126.396	0.0020	4606553(10)	124.6686	124.670	0.0021	4319063(12)	117.1127	117.113	0.0038
0.3	3966476(9)	122.7765	122.778	0.0014	3338810(10)	120.0840	120.085	0.0015	4638866(13)	111.0985	111.098	0.0037
0.4	4023257(10)	119.9797	119.981	0.0016	3999086(10)	116.5157	116.513	0.0015	4291865(13)	106.8551	106.855	0.0019
0.5	3134828(18)	118.0022	118.005	0.0031	3612272(10)	113.8652	113.861	0.0017	4196786(12)	103.7331	103.733	0.0019
0.6	2720891(12)	116.7135	116.719	0.0033	4535513(17)	111.9025	111.906	0.0017	4171346(12)	101.3416	101.341	0.0020
0.7	4415051(12)	115.9751	115.982	0.0035	4452977(13)	110.4167	110.422	0.0020	4155383(12)	99.4481	99.448	0.0023
0.8	3446018(13)	115.6382	115.679	0.0065	4188863(14)	109.2506	109.258	0.0025	4298489(12)	97.9093	97.909	0.0022
0.9	4529801(13)	115.6060	115.664	0.0064	3595223(13)	108.2981	108.307	0.0033	4312274(14)	96.6317	96.631	0.0022
1.0	3499793(14)	115.7115	115.868	0.0108	4147868(15)	107.4928	107.508	0.0032	4555955(12)	95.5525	95.552	0.0022

Table 9.5: Comparison with results in literature for the Giesekus model with Stokes flow coupling.

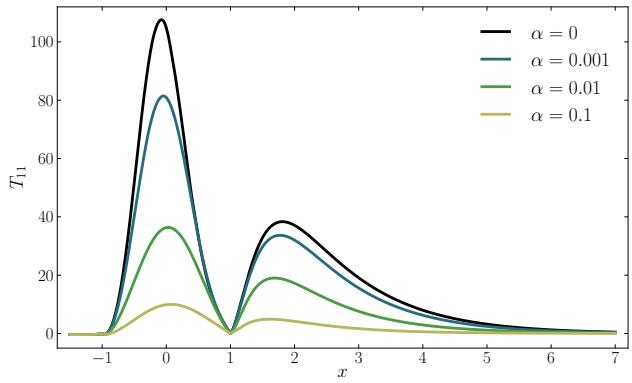


Figure 9.36: Profile of the T_{11} component of the extra stress tensor for $\mathrm{Wi} = 0.7$ along $\boldsymbol{\gamma}$ with various values of α.

9.3.4 Creeping flow with the Giesekus model

In this subsection we investigate the effects of the mobility factor in the Giesekus model with Stokes flow coupling. Here, the drag coefficient was computed using the same energy strategy described in Section 9.3.2 for $\alpha = 0.01$, 0.1, and 1. The results are collected together in Table 9.5. For a fixed Weissenberg number, the drag coefficient *decreases* as the mobility factor is increased. The decrease in the drag coefficient is due to the shear-thinning properties of the Giesekus fluid model.

Figure 9.36 depicts the profile of T_{11} for each value of α and fixed $\mathrm{Wi} = 0.7$. Notice that the scale of this variable is strongly dependent upon the order of magnitude of the mobility factor.

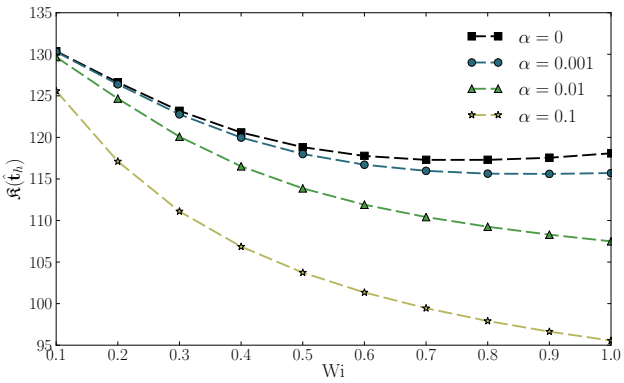


Figure 9.37: Dependence of the drag coefficient upon α for $\mathrm{Wi} = 0.7$.

Figure 9.37 demonstrates how the drag coefficient pattern changed with Wi as the mobility factor was increased. Notably, the relationship became monotonic over the parameter range considered once $\alpha \geq 0.01$.

9.3.5 Goal-oriented adaptive mesh refinement

In this subsection, initial results are presented for goal-oriented adaptive mesh refinement (GMR) with the DPG method for Oldroyd-B model above. In an attempt to accelerate the convergence of the drag coefficient C_{D}, given in (153), we adapted the explicit GMR strategy analyzed for Poisson's equation in Section 9.2 to the purposes here. In this case, at each refinement step, some number of Gauss–Newton iterations were performed to generate a primal (DPG) solution on the given mesh. Then, using the final linearization about the computed DPG solution (see (86)), the dual (DPG*) problem was formed.

Having the primal and dual solutions, the error indicators η_K and η_K^* could be computed. Indeed, first, the error indicators η_K are obviously the same as already computed for the previous solution-oriented strategies. Second, the η_K^* error indicators can be defined from the formula given in (124b), which holds for all ultraweak bilinear forms in the form of (120). For further details, see [92].

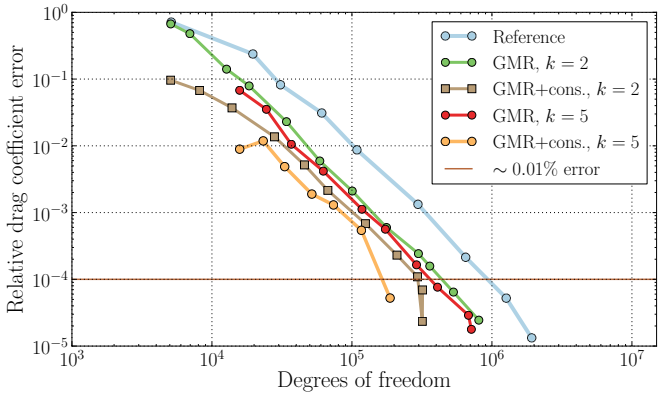


Figure 9.38: The error in the drag coefficient with goal-oriented adaptive mesh refinement (GMR) compared to the original method and strategy (light blue). "+cons." indicates volume conservation is also used. Polynomial degree is indicated by k and $\mathrm{Wi} = 0.4$.

In our GMR experiments, we also considered the introduction of constraints inducing local volume conservation and the effect of high order discretizations [57]. Combining all of these features, we see in Figure 9.38 that almost a significant accuracy improvement can be achieved over the standard DPG approach. Some ad hoc anisotropic and *hp* goal-oriented adaptive strategies were also briefly studied, but none of these were seen to deliver a clear advantage.

Chapter 10

Conclusion

10.1 Discussion

The theory developed in this dissertation strengthens the theoretical foundation upon which DPG methods rest. Previous to this work, DPG had already been identified as a Petrov–Galerkin method with so-called *optimal test functions*, as a minimum residual method, and as a mixed method, the main emphasis here. An important addition to this list of interpretations is featured in Chapter 8; the practical (i.e., fully discrete) method always yields an ordinary least-squares problem. There, it is emphasized that a DPG method induces an overdetermined system of linear equations which may be solved directly, in a least-squares sense, with algorithms which avoid ever forming the corresponding normal equation. Notably, this leads to new solution algorithms for DPG methods with far lower round-off error sensitivity when compared to the traditional approach. The least-squares point of view can be seen as the fourth and final interpretation of this class of methods.

DPG* methods, the duals to DPG methods, are introduced in this dissertation. This new class of methods has a similar set of interpretations as DPG methods, with each interpretation emphasizing DPG* methods as means for solving underdetermined finite element discretizations. Unlike DPG methods, none of these interpretations is shown here to lead to DPG* solution algorithms with less sensitivity to round off error; however, this question is also never actively pursued. A more significant detriment to DPG* methods is identified in Chapter 4. Namely, the accuracy of a DPG* solution is controlled, in part, by the regularity of an auxiliary Lagrange multiplier variable. As clearly illustrated in Section 9.1, this can induce sub-optimal convergence rates under uniform h-refinements. Fortunately, it is also demonstrated that this issue can be overcome via adaptive mesh refinement.

The development of goal-oriented adaptive mesh refinement strategies for DPG methods is an important contribution in this dissertation. The strategies described in Section 7.5 each involve the solution of a primal (DPG) and a dual (DPG*) problem. In the case of nonlinear DPG methods, the primal problem is solved successively, until the solution has sufficiently converged. Afterward, the induced dual problem is solved only once. In this DPG* problem, the load comes from an underlying quantity of interest. Leading up to each mesh refinement, the primal and dual solutions are used to form local error indicators which can be used to select which elements in the present mesh to refine next. Several examples of goal-oriented adaptive mesh refinement with DPG methods are illustrated in Section 9.2 and Section 9.3.5. Here, it is clearly evident that these goal-oriented strategies have a significant influence on the error in the corresponding quantities of interest, especially when compared to the standard DPG solution-oriented strategy.

Duality plays a crucial role in most aspects of this dissertation. In Chapter 2, analysis of dual embeddings of the simple operator equation $Bu = \ell$ leads to the general mixed variational formulation given in (13). In Chapter 3, dual minimization principles leads to the semi-discrete "ideal" methods given in (47). This also leads to Proposition 3.2, which expresses the interplay between the energy norm $|||\cdot|||_{\mathcal{U}}$ and the corresponding optimal norm $|||\cdot|||_{\mathcal{V}}$. Then, building off the semi-discrete methods mentioned above, duality leads to the general fully discrete problem (53) as well as to the discovery of DPG* methods. Duality is also a crucial component of the analysis involved in Chapter 6. In particular, Galerkin orthogonality in the dual problems plays a central role in the Aubin–Nitsche-type arguments which the proofs of Theorems 6.5 and 6.8 involve. Many of the deductions above are also essential in the *a posteriori* error estimate for linear quantities of interest featured in Corollary 7.3. This estimate heavily influences the several goal-oriented adaptive mesh refinement strategies described in Chapter 7 and is used in many of the numerical experiments reported on in Chapter 9.

In addition to the contributions described above, several new results specific to DPG and DPG* methods appear in this dissertation. For instance, the entire *a priori* error estimation theory for DPG* methods presented in Sections 6.4 and 6.5 is completely novel. Similarly, the *a*

posteriori error analysis for DPG* methods in Section 7.4 and the improved reliability estimates for DPG methods given in Theorems 7.7 and 7.8 may also be regarded as notable contributions.

10.2 Final remarks

This dissertation develops a paradigm for the analysis of DPG methods and introduces DPG* methods as their natural dual counterpart. Ultimately, much of the theory presented here exists in a framework which extends far beyond just these particular classes of finite element methods. Indeed, stemming from the idea of embedding the underlying PDE into a saddle-point problem, this dissertation provides a full perspective on the construction and analysis of finite element methods with trial and test spaces of unequal dimension. Some of the other principal insights gleaned by this research touch on classical topics like *a priori* and *a posteriori* error analysis, adaptive mesh refinement, and solution algorithms for least-squares problems. Both linear and nonlinear physical models are examined in this dissertation and the results indicate a promising basis for further study.

Addendum

Other work

This chapter contains short summaries of several additional research projects which the author was involved in throughout this PhD study. Most of this work is only tangentially related to the overall goal of this dissertation, so only brief summaries of each project are provided. The interested reader is encouraged to consult the cited texts for fuller accounts of the work mentioned here.

Orientation embedded high order shape functions for the exact sequence elements of all shapes [68, 88]

The first completed project of the author's work establishes a unifying methodology for the construction of hierarchical shape functions for hybrid *hp* finite element meshes. Originally conceived in [75] and involving several ideas from [133, 141], this work is collected together in a 105-page monograph [68] and the corresponding code is published online, in a package called ESEAS, at [88]. This shape function library, ESEAS, has been used extensively by the ICES Electromagnetics and Acoustics Group. Specifically, it has been relied upon for the numerical experiments performed in [28, 49, 65–67, 87, 89, 91, 93, 103, 105–107, 120, 131, 137].

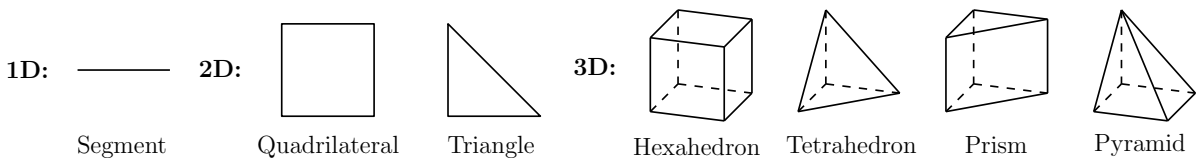


Figure A.1: The standard 3D elements.

The most notable scientific contributions of this work are:

- The first explicit construction of exact sequence-spanning, arbitrary polynomial order, hierarchical shape functions for the pyramid element. This construction closely follows the definitions of the pyramid spaces given in [108].
- The shape functions are conforming, hierarchical and compatible with other neighboring elements across shared boundaries so they may be used in hybrid meshes.
- The methodology of the construction is consistent throughout all element types (segment, quadrilateral, triangle, hexahedron, tetrahedron, triangular prism and pyramid). This readily allows corresponding software to be written using relatively very few subroutines.
- Construction is accompanied by open-source software [88].

The DPG methodology applied to different variational formulations of linear elasticity [89]

The second project the author was involved in their study had the following broadly defined purpose: to analyze contemporary nonlinear hyperelasticity models for arterial mechanics and, ultimately, incorporate the discretization of the corresponding PDEs into a comprehensive nonlinear DPG framework. Originally, the research was subsumed within this elasticity program. This project began with studying the equations of *linear* elasticity with the new DPG methodology developed in [28] in four different variational formulations. A follow-up project, discussed in the next section, grew out of this work. Eventually, the research focus was directed toward the duality theory which makes up the central body of this dissertation and the focus toward the analysis of DPG methods for hyperelasticity was ultimately abandoned.

The work considered here demonstrates the flexibility of the DPG methodology by solving the equations of linear elasticity with the DPG framework in four different variational formulations: strong, ultraweak, primal, and mixed. Relying upon the theory developed in [28], each of these variational formulations is shown to be well-posed after the corresponding test space is broken. Moreover, an important lemma from [28] is also generalized. The article also provides a proof of the mutual ill- or well-posedness of each (broken or unbroken) formulation

when using traditional energy spaces on the whole domain. Some numerical results from [89] are reproduced in Figures A.2 and A.3. Notably, this work, completed in collaboration with Dr. Federico Fuentes, formed a foundation for much of his dissertation work on linear thermo-viscoelasticity [65, 66]

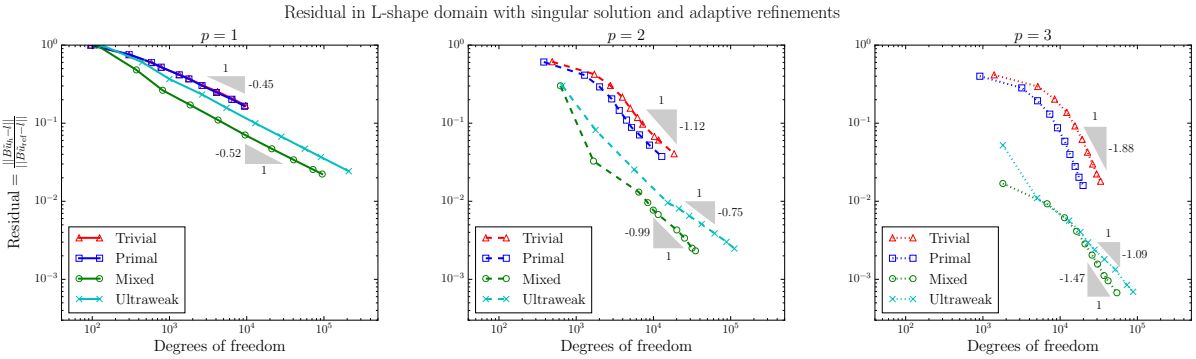


Figure A.2: Residual as a function of the degrees of freedom after adaptive anisotropic hexahedral h-refinements in the L-shaped domain with a singular solution.

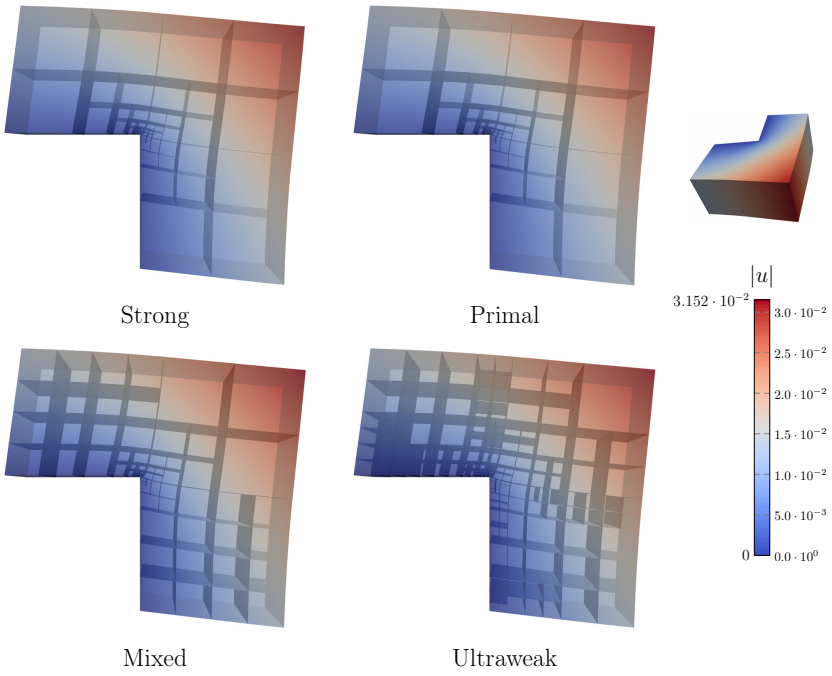


Figure A.3: The adaptive meshes for each method after five successive refinements. The domains are colored by the displacement magnitude, $|\boldsymbol{u}|$, and warped by a factor of 10.

The most notable scientific contributions of this work are:

- Proof of the mutual well-posedness of several variational formulations of the linear elasticity boundary value problem.
- Proof of Lemma 3.1.
- Three-dimensional anisotropic h-adaptive computations.

Coupled variational formulations of linear elasticity and the DPG methodology [67]

This work expands upon [89] in order to verify that not only is the DPG methodology suitable for a wide variety of different variational formulations of a problem throughout a single computational domain, but that it is also suitable for coupled variational formulations which are chosen differently inside a collection of disjoint subdomains. In this paper, we developed a straight forward method for coupling variational formulations together using the natural interface variables present in DPG methods. We then showed that these new coupled formulations are indeed well-posed and can be exploited to solve challenging problems more accurately and efficiently in a variety of physical scenarios where stability is an issue or a particular mode of convergence is desired in a specific subdomain.

The efficacy of the general coupling strategy is demonstrated in two numerical examples in the paper. The second example is the most interesting as it considers a physically-motivated sheathed hose problem with very large material contrast, as indicated in Figure A.4, given to us as a challenge by Prof. Patrick Le Tallec. The results of our experiments with uniform and non-uniform pressure loading are depicted in Figures A.5 and A.6, respectively.

The most notable scientific contributions of this work are:

- Proof of the well-posedness of the broken and unbroken coupled variational formulations.
- First demonstration of the DPG methodology for coupled broken variational formulations.
- Analysis of challenging high material contrast problem of the sheathed hose.

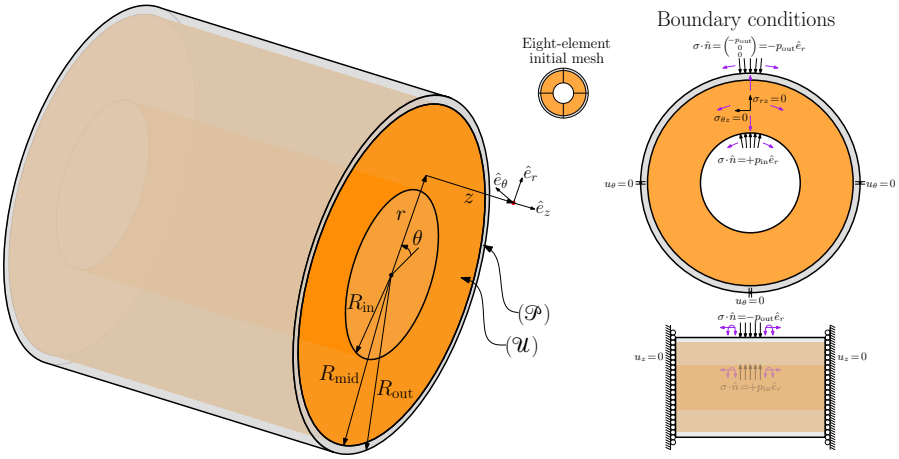


Figure A.4: Diagram of the sheathed hose problem with the configuration of variational formulations per subdomain and a schematic of the boundary conditions used.

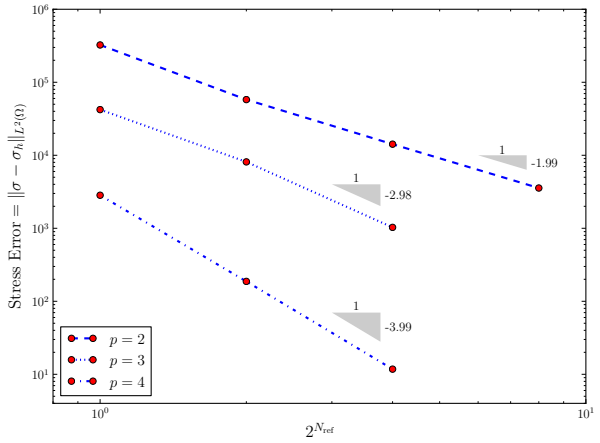


Figure A.5: Stress error (in Pa) as a function of the number of uniform refinements, N_{ref}. The value of $p = 1$ was not shown because the approximating isoparametric geometry was too inaccurate for the initial meshes.

On perfectly matched layers for discontinuous Petrov–Galerkin methods [138]

This final additional project arose from conversations with Dr. Ali Vaziri Astaneh while he was working as a postdoctoral researcher at ICES. Vaziri Astaneh's PhD research centered on the design and discretization of perfectly matched layers (PMLs) in classical Galerkin finite element methods [139]. Together, we expanded on the DPG theory surrounding PMLs, first documented in Dr. Jamie Bramwell's PhD dissertation [17]. In [138], we demonstrated

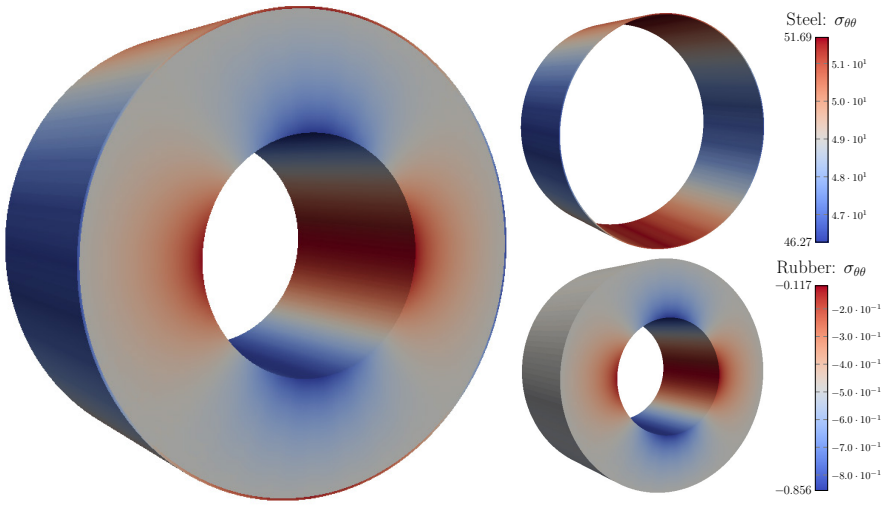


Figure A.6: Stress component $\sigma_{\theta\theta}$ (in MPa) from computed solution with $p = 2$ and nonuniform internal pressure loading after three uniform refinements of the eight-element initial mesh. Note the discontinuity across the material interface.

that ultraweak variational formulations permit two dual constructions of PMLs which, for a unique coordinate stretching function, do not generally coincide. This was a surprising discovery, because it is at odds with the analogous scenario found in classical methods. Ultimately, we went on to derive and analyze DPG PML formulations for time-harmonic acoustics, electromagnetics, and elastodynamics problems posed on unbounded domains in both two- and three-dimensions. Selected results from our experiments are presented in Figures A.7–A.9.

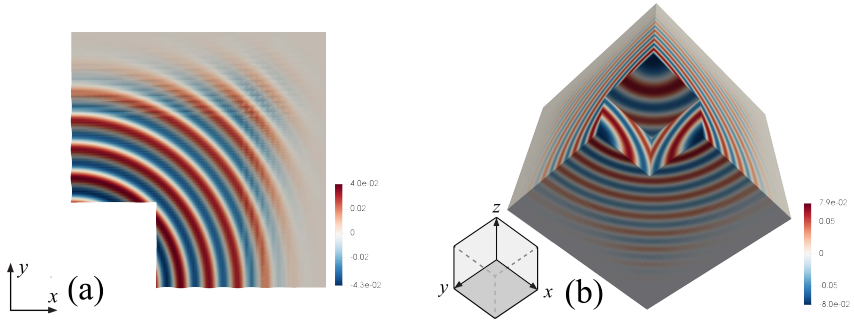


Figure A.7: Time-harmonic acoustic wave scattering in the discrete pressure variable p in: (a) 2D; (b) 3D. Here, only the real part of the solution is visualized.

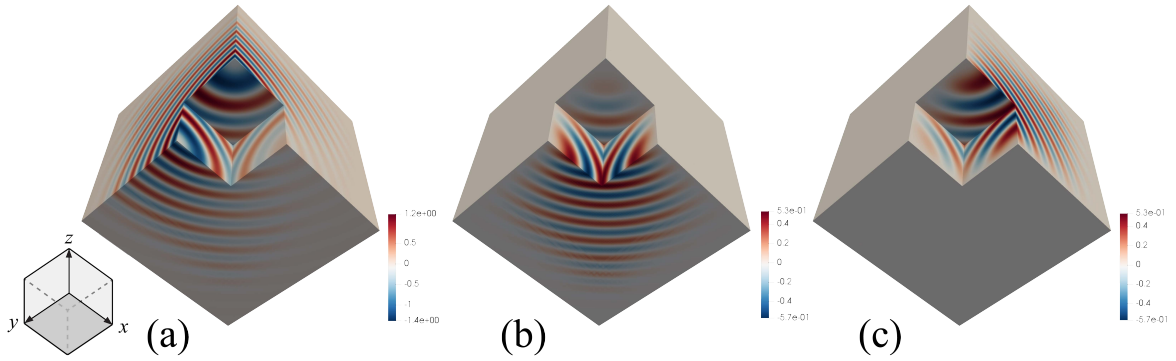


Figure A.8: Time-harmonic electromagnetic wave scattering of the discrete electric field $\mathbf{E}$ in three dimensions: (a) the x-component of the electric field, E_x; (b) the y-component of the electric field, E_y; (c) the z-component of the electric field, E_z. Again, only the real part of the solution is visualized.

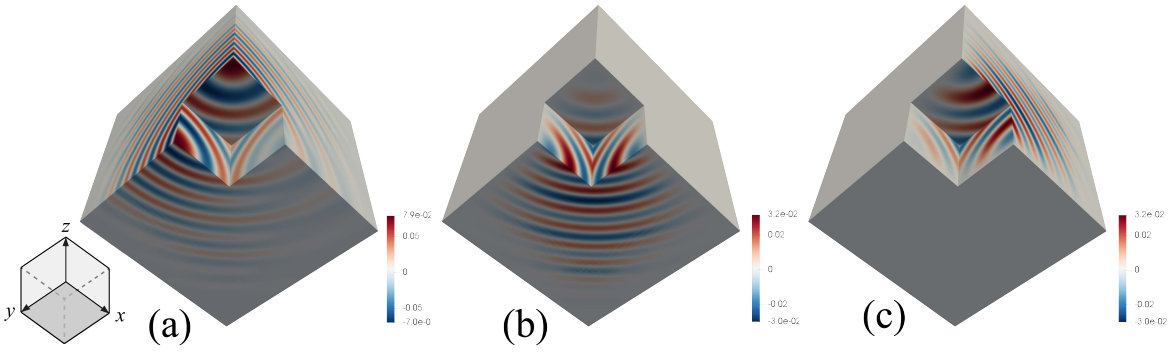


Figure A.9: Time-harmonic elastodynamic wave scattering of the discrete displacement $\boldsymbol{u}$ in three dimensions: (a) the x-component of the displacement, u_x; (b) the y-component of the displacement, u_y; (b) the z-component of the displacement, u_z. Again, only the real part of the solution is visualized.

Bibliography

[1] A. Afonso, P. Oliveira, F. Pinho, and M. Alves, *The log-conformation tensor approach in the finite-volume method framework*, J. Non-Newton. Fluid Mech., 157 (2009), pp. 55–65.

[2] M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, vol. 37, John Wiley & Sons, 2000.

[3] B. Aksoylu, S. D. Bond, E. C. Cyr, and M. Holst, *Goal-oriented adaptivity and multilevel preconditioning for the Poisson–Boltzmann equation*, J. Sci. Comput., 52 (2012), pp. 202–225.

[4] P. R. Amestoy, I. S. Duff, J.-Y. L'Excellent, and J. Koster, *A fully asynchronous multifrontal solver using distributed dynamic scheduling*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 15–41.

[5] D. N. Arnold, R. S. Falk, and R. Winther, *Finite element exterior calculus, homological techniques, and applications*, Acta Numer., 15 (2006), pp. 1–155.

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

[7] I. Babuška and W. C. Rheinboldt, *A-posteriori error estimates for the finite element method*, Int. J. Numer. Meth. Eng., 12 (1978), pp. 1597–1615.

[8] I. Babuška, R. B. Kellogg, and J. Pitkäranta, *Direct and inverse error estimates for finite elements with mesh refinements*, Numer. Math., 33 (1979), pp. 447–471.

[9] R. Becker, E. Estecahandy, and D. Trujillo, *Weighted marking for goal-oriented adaptive finite element methods*, SIAM J. Numer. Anal., 49 (2011), pp. 2451–2469.

[10] R. Becker and R. Rannacher, *An optimal control approach to a posteriori error estimation in finite element methods*, Acta Numer., 10 (2001), pp. 1–102.

[11] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numer., 14 (2005), pp. 1–137.

[12] Å. Björck, *Numerical methods for least squares problems*, SIAM, Philadephia, 1996.

[13] P. Bochev, *Least-squares finite element methods for first-order elliptic systems*, Int. J. Numer. Anal. Model, 1 (2004), pp. 49–64.

[14] P. B. Bochev and M. D. Gunzburger, *Least-squares finite element methods*, vol. 166, Springer Science & Business Media, 2009.

[15] D. Boffi, M. Fortin, and F. Brezzi, *Mixed finite element methods and applications*, Springer series in computational mathematics, Springer, Berlin, Heidelberg, 2013.

[16] T. Bouma, J. Gopalakrishnan, and A. Harb, *Convergence rates of the DPG method with reduced test space degree*, Comput. Math. Appl., 68 (2014), pp. 1550–1561.

[17] J. Bramwell, *A discontinuous Petrov–Galerkin method for seismic tomography problems*, PhD dissertation, The University of Texas at Austin, Austin, Texas, U.S.A., 2013.

[18] J. Bramwell, L. Demkowicz, J. Gopalakrishnan, and Q. Weifeng, *A locking-free hp DPG method for linear elasticity with symmetric stresses*, Numer. Math., 122 (2012), pp. 671–707.

[19] F. Brezzi, *On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers*, Rev. fr. autom. inform. rech. opér., Anal. numér., 8 (1974), pp. 129–151.

[20] A. N. Brooks and T. J. Hughes, *Streamline upwind/Petrov–Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations*, Comput. Methods in Appl. Mech. Eng., 32 (1982), pp. 199–259.

[21] T. Bui-Thanh, L. Demkowicz, and O. Ghattas, *Constructively well-posed approximation methods with unity inf–sup and continuity constants for partial differential equations*, Math. Comput., 82 (2013), pp. 1923–1952.

[22] ——, *A unified discontinuous Petrov–Galerkin method and its analysis for Friedrichs' systems*, SIAM J. Numer. Anal., 51 (2013), pp. 1933–1958.

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

[24] Z. Cai, T. A. Manteuffel, S. F. McCormick, and J. Ruge, *First-order system $\mathcal{LL}^*$ (FOSLL*): Scalar elliptic partial differential equations*, SIAM J. Numer. Anal., 39 (2001), pp. 1418–1445.

[25] P. Cantin and N. Heuer, *A DPG framework for strongly monotone operators*, <hal-01690281>, (2018).

[26] C. Carstensen, P. Bringmann, F. Hellwig, and P. Wriggers, *Nonlinear discontinuous Petrov–Galerkin methods*, ArXiv e-prints, arXiv:1710.00529 [math.NA], (2017).

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

[28] ——, *Breaking spaces and forms for the DPG method and applications including Maxwell equations*, Comput. Math. Appl., 72 (2016), pp. 494–522.

[29] C. Carstensen and F. Hellwig, *Low-order discontinuous Petrov–Galerkin finite element methods for linear elasticity*, SIAM J. Numer. Anal., 54 (2016), pp. 3388–3410.

[30] J. Chan, *A DPG method for convection-diffusion problems*, PhD dissertation, The University of Texas at Austin, Austin, Texas, U.S.A., 2013.

[31] J. Chan, L. Demkowicz, and R. Moser, *A DPG method for steady viscous compressible flow*, Comput. Fluids, 98 (2014), pp. 69–90.

[32] J. Chan, L. Demkowicz, R. Moser, and N. Roberts, *A new discontinuous Petrov–Galerkin method with optimal test functions. Part V: Solution of 1D Burgers and Navier–Stokes equations*, ICES Report 10-25, The University of Texas at Austin, 2010.

[33] J. Chan, J. A. Evans, and W. Qiu, *A dual Petrov–Galerkin finite element method for the convection–diffusion equation*, Comput. Math. Appl., 68 (2014), pp. 1513–1529.

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

[35] J. H. Chaudhry, E. C. Cyr, K. Liu, T. A. Manteuffel, L. N. Olson, and L. Tang, *Enhancing least-squares finite element methods through a quantity-of-interest*, SIAM J. Numer. Anal., 52 (2014), pp. 3085–3105.

[36] P. G. Ciarlet, *Linear and Nonlinear Functional Analysis with Applications*, SIAM, Philadelphia, 2013.

[37] S. Claus and T. Phillips, *Viscoelastic flow around a confined cylinder using spectral/hp element methods*, J. Non-Newton. Fluid Mech., 200 (2013), pp. 131–146. Special Issue: Advances in Numerical Methods for Non-Newtonian Flows.

[38] P. Clément, *Approximation by finite element functions using local regularization*, Rev. fr. autom. inform. rech. opér., Anal. numér., 9 (1975), pp. 77–84.

[39] O. M. Coronado, D. Arora, M. Behr, and M. Pasquali, *A simple method for simulating general viscoelastic fluid flows with an alternate log-conformation formulation*, J. Non-Newton. Fluid Mech., 147 (2007), pp. 189–199.

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

[41] L. Demkowicz, *Computing with hp Finite Elements. I. One and Two Dimensional Elliptic and Maxwell Problems*, Chapman & Hall/CRC Press, New York, October 2006.

[42] ——, *Polynomial exact sequences and projection-based interpolation with application to Maxwell equations*, in Mixed Finite Elements, Compatibility Conditions, and Applications, D. Boffi and L. Gastaldi, eds., vol. 1939 of Lecture Notes in Mathematics, Springer, Berlin, 2008, pp. 101–158.

[43] ——, *Various variational formulations and closed range theorem*, ICES Report 15-03, The University of Texas at Austin, 2015.

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

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

[46] ——, *A class of discontinuous Petrov–Galerkin methods. II. Optimal test functions*, Numer. Methods Partial Differ. Equ., 27 (2011), pp. 70–105.

[47] ——, *A primal DPG method without a first-order reformulation*, Comput. Math. Appl., 66 (2013), pp. 1058–1064.

[48] L. Demkowicz, J. Gopalakrishnan, and B. Keith, *The DPG-star method*, In preparation, (2018).

[49] L. Demkowicz, J. Gopalakrishnan, S. Nagaraj, and P. Sepulveda, *A spacetime DPG method for the Schrödinger equation*, SIAM J. Numer. Anal., 55 (2017), pp. 1740–1759.

[50] L. Demkowicz, J. Gopalakrishnan, and A. H. Niemi, *A class of discontinuous Petrov–Galerkin methods. Part III: Adaptivity*, Appl. Numer. Math., 62 (2012), pp. 396–427.

[51] L. Demkowicz, J. Gopalakrishnan, and J. Schöberl, *Polynomial extension operators. Part III*, Math. Comput., 81 (2012), pp. 1289–1326.

[52] L. Demkowicz and N. Heuer, *Robust DPG method for convection-dominated diffusion problems*, SIAM J. Numer. Anal., 51 (2013), pp. 2514–2537.

[53] L. Demkowicz, J. Kurtz, D. Pardo, M. Paszyński, W. Rachowicz, and A. Zdunek, *Computing with hp Finite Elements. II. Frontiers: Three Dimensional Elliptic and Maxwell Problems with Applications*, Chapman & Hall/CRC, New York, October 2007.

[54] A. Demlow and A. N. Hirani, *A posteriori error estimates for finite element exterior calculus: The de Rham complex*, Found. Comput. Math., 14 (2014), pp. 1337–1371.

[55] W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., 33 (1996), pp. 1106–1124.

[56] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, vol. 28 of Classics in Applied Mathematics, SIAM, 1999.

[57] T. Ellis, L. Demkowicz, and J. Chan, *Locally conservative discontinuous Petrov–Galerkin finite elements for fluid problems*, Comput. Math. Appl., 68 (2014), pp. 1530–1549.

[58] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Introduction to adaptive methods for differential equations*, Acta Numer., 4 (1995), pp. 105–158.

[59] A. Ern, J.-L. Guermond, and G. Caplain, *An intrinsic criterion for the bijectivity of Hilbert operators related to Friedrichs' systems*, Comm. Partial Differential Equations, 32 (2007), pp. 317–341.

[60] J. Ernesti, *Space-Time Methods for Acoustic Waves with Applications to Full Waveform Inversion*, PhD dissertation, Karlsruhe, Germany, 2017.

[61] L. C. Evans, *Partial differential equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, second ed., 2010.

[62] R. Falk and R. Winther, *Local bounded cochain projections*, Math. Comput., 83 (2014), pp. 2631–2656.

[63] Y. Fan, R. Tanner, and N. Phan-Thien, *Galerkin/least-square finite-element methods for steady viscoelastic flows*, J. Non-Newton. Fluid Mech., 84 (1999), pp. 233–256.

[64] M. Feischl, D. Praetorius, and K. G. Van der Zee, *An abstract analysis of optimal goal-oriented adaptivity*, SIAM J. Numer. Anal., 54 (2016), pp. 1423–1448.

[65] F. Fuentes, *Various applications of discontinuous Petrov–Galerkin (DPG) finite element methods*, PhD dissertation, The University of Texas at Austin, Austin, Texas, U.S.A., 2018.

[66] F. Fuentes, L. Demkowicz, and A. Wilder, *Using a DPG method to validate DMA experimental calibration of viscoelastic materials*, Comput. Methods Appl. Mech. Engrg., 325 (2017), pp. 748–765.

[67] F. Fuentes, B. Keith, L. Demkowicz, and P. Le Tallec, *Coupled variational formulations of linear elasticity and the DPG methodology*, J. Comput. Phys., 348 (2017), pp. 715–731.

[68] F. Fuentes, B. Keith, L. Demkowicz, and S. Nagaraj, *Orientation embedded high order shape functions for the exact sequence elements of all shapes*, Comput. Math. Appl., 70 (2015), pp. 353–458.

[69] T. Führer, *Superconvergent DPG methods for second order elliptic problems*, ArXiv e-prints, arXiv:1712.07719 [math.NA], (2017).

[70] ——, *Superconvergence in a DPG method for an ultra-weak formulation*, Comput. Math. Appl., 75 (2018), pp. 1705–1718.

[71] T. Führer and N. Heuer, *Robust coupling of DPG and BEM for a singularly perturbed transmission problem*, Comput. Math. Appl., 74 (2017), pp. 1940–1954.

[72] ——, *Fully discrete DPG methods for the Kirchhoff–Love plate bending model*, ArXiv e-prints, arXiv:1805.08864 [math.NA], (2018).

[73] T. Führer, N. Heuer, and A. H. Niemi, *An ultraweak formulation of the Kirchhoff–Love plate bending model and DPG approximation*, ArXiv e-prints, arXiv:1805.07835 [math.NA], (2018).

[74] T. Führer, N. Heuer, and E. P. Stephan, *On the DPG method for Signorini problems*, IMA J. Numer. Anal., in press (2017).

[75] P. Gatto and L. Demkowicz, *Construction of H^1-conforming hierarchical shape functions for elements of all shapes and transfinite interpolation*, Finite Elem. Anal. Des., 46 (2010), pp. 474–486.

[76] H. Giesekus, *A simple constitutive equation for polymer fluids based on the concept of deformation-dependent tensorial mobility*, J. Non-Newton. Fluid Mech., 11 (1982), pp. 69–109.

[77] M. B. Giles and E. Süli, *Adjoint methods for PDEs: a posteriori error analysis and postprocessing by duality*, Acta Numer., 11 (2002), pp. 145–236.

[78] G. Golub, *Numerical methods for solving linear least squares problems*, Numer. Math, 7 (1965), pp. 206–216.

[79] G. H. Golub, C. Greif, and J. M. Varah, *An algebraic analysis of a block diagonal preconditioner for saddle point systems*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 779–792.

[80] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 4, The Johns Hopkins University Press, 2013.

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

[82] J. Gopalakrishnan and W. Qiu, *An analysis of the practical DPG method*, Math. Comput., 83 (2014), pp. 537–552.

[83] B. N. Granzow, M. S. Shephard, and A. A. Oberai, *Output-based error estimation and mesh adaptation for variational multiscale methods*, Comput. Methods Appl. Mech. Engrg., 322 (2017), pp. 441–459.

[84] M. A. Heroux, R. A. Bartlett, V. E. Howle, R. J. Hoekstra, J. J. Hu, T. G. Kolda, R. B. Lehoucq, K. R. Long, R. P. Pawlowski, E. T. Phipps, A. G. Salinger, H. K. Thornquist, R. S. Tuminaro, J. M. Willenbring, A. Williams, and K. S. Stanley, *An overview of the Trilinos project*, ACM Trans. Math. Softw., 31 (2005), pp. 397–423.

[85] M. Holst and S. Pollock, *Convergence of goal-oriented adaptive finite element methods for nonsymmetric problems*, Numer. Methods Partial Differ. Equ., 32 (2016), pp. 479–509.

[86] M. A. Hulsen, R. Fattal, and R. Kupferman, *Flow of viscoelastic fluids past a cylinder at high weissenberg number: Stabilized simulations using matrix logarithms*, J. Non-Newton. Fluid Mech., 127 (2005), pp. 27–39.

[87] B. Keith, L. Demkowicz, and J. Gopalakrishnan, *DPG* method*, ICES Report 17-25, The University of Texas at Austin, 2017.

[88] B. Keith, F. Fuentes, and L. Demkowicz, *The Exact Sequence for Elements of All Shapes (ESEAS) software package*. https://github.com/libESEAS/ESEAS, 2015.

[89] B. Keith, F. Fuentes, and L. Demkowicz, *The DPG methodology applied to different variational formulations of linear elasticity*, Comput. Methods Appl. Mech. Engrg., 309 (2016), pp. 579–609.

[90] B. Keith, P. Knechtges, N. V. Roberts, S. Elgeti, M. Behr, and L. Demkowicz, *An ultraweak DPG method for viscoelastic fluids*, J. Non-Newton. Fluid Mech., 247 (2017), pp. 107–122.

[91] B. Keith, S. Petrides, F. Fuentes, and L. Demkowicz, *Discrete least-squares finite element methods*, Comput. Methods Appl. Mech. Engrg., 327 (2017), pp. 226–255.

[92] B. Keith and N. V. Roberts, *A new DPG method for viscoelastic fluids with goal-oriented adaptivity*, In preparation, (2018).

[93] B. Keith, A. Vaziri Astaneh, and L. Demkowicz, *Goal-oriented adaptive mesh refinement for non-symmetric functional settings*, ArXiv e-prints, arXiv:1711.01996 [math.NA], (2017).

[94] A. Kessy, A. Lewin, and K. Strimmer, *Optimal whitening and decorrelation*, Am. Stat., (2017).

[95] P. Knechtges, M. Behr, and S. Elgeti, *Fully-implicit log-conformation formulation of constitutive laws*, J. Non-Newton. Fluid Mech., 214 (2014), pp. 78–87.

[96] P. Ladevèze, *Comparaison de modeles de milieux continus*, PhD dissertation, Université P. et M. Curie, Paris, France, 1975.

[97] P. Ladevèze, F. Pled, and L. Chamoin, *New bounding techniques for goal-oriented error estimation applied to linear problems*, Int. J. Numer. Meth. Eng., 93 (2013), pp. 1345–1380.

[98] A. W. Liu, D. E. Bornside, R. C. Armstrong, and R. A. Brown, *Viscoelastic flow of polymer solutions around a periodic, linear array of cylinders: comparisons of predictions for microstructure and flow fields*, J. Non-Newton. Fluid Mech., 77 (1998), pp. 153–190.

[99] J. W. H. Liu, *The multifrontal method for sparse matrix solution: Theory and practice*, SIAM Review, 34 (1992), pp. 82–109.

[100] C. Macosko, *Rheology: principles, measurements, and applications*, Advances in interfacial engineering series, Wiley-VCH, New York, 1994.

[101] J. E. Marsden and T. J. Hughes, *Mathematical foundations of elasticity*, Dover Publications, New York, 1994.

[102] M. S. Mommer and R. Stevenson, *A goal-oriented adaptive finite element method with convergence rates*, SIAM J. Numer. Anal., 47 (2009), pp. 861–886.

[103] J. Mora and L. Demkowicz, *Fast integration of DPG matrices based on tensorization*, ArXiv e-prints, arXiv:1711.00984 [math.NA], (2017).

[104] 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, arXiv:1511.04400, (2015).

[105] S. Nagaraj, *DPG Methods For Nonlinear Fiber Optics*, PhD dissertation, The University of Texas at Austin, Austin, Texas, U.S.A., 2018.

[106] S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, and J. Mora, *A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers*, ArXiv e-prints, arXiv:1805.12240 [math.NA], (2018).

[107] S. Nagaraj, S. Petrides, and L. Demkowicz, *Construction of DPG Fortin operators for second order problems*, Comput. Math. Appl., 74 (2017), pp. 1964–1980.

[108] N. Nigam and J. Phillips, *High-order conforming finite elements on pyramids*, IMA J. Numer. Anal., 32 (2012), pp. 448–483.

[109] J. T. Oden and S. Prudhomme, *New approaches to error estimation and adaptivity for the Stokes and Oseen equations*, Int. J. Numer. Methods Fluids, 31 (1999), pp. 3–15.

[110] ——, *Goal-oriented error estimation and adaptivity for the finite element method*, Comput. Math. Appl., 41 (2001), pp. 735–756.

[111] ——, *Estimation of modeling error in computational mechanics*, J. Comput. Phys., 182 (2002), pp. 496–515.

[112] J. T. Oden and J. N. Reddy, *Variational Methods in Theoretical Mechanics*, Universitext, Springer-Verlag, Berlin, 2nd ed., 1983.

[113] J. Oldroyd, *On the formulation of rheological equations of state*, in P. Roy. Soc. A, vol. 200, The Royal Society, 1950, pp. 523–541.

[114] R. G. Owens, C. Chauvière, and T. N. Philips, *A locally-upwinded spectral technique (LUST) for viscoelastic flows*, J. Non-Newton. Fluid Mech., 108 (2002), pp. 49–71. Numerical Methods Workshop S.I.

[115] R. G. Owens and T. N. Phillips, *Computational rheology*, Imperial College Press, London, 2002.

[116] C. C. Paige, *Computer solution and perturbation analysis of generalized linear least squares problems*, Math. Comput., 33 (1979), pp. 171–183.

[117] ——, *Fast numerically stable computations for generalized linear least squares problems*, SIAM J. Numer. Anal., 16 (1979), pp. 165–171.

[118] M. Paraschivoiu, J. Peraire, and A. T. Patera, *A posteriori finite element bounds for linear-functional outputs of elliptic partial differential equations*, Comput. Methods Appl. Mech. Engrg., 150 (1997), pp. 289–312.

[119] A. T. Patera and J. Peraire, *A general Lagrangian formulation for the computation of a posteriori finite element bounds*, in Error Estimation and Adaptive Discretization Methods in Computational Fluid Dynamics, Springer, 2003, pp. 159–206.

[120] S. Petrides and L. F. Demkowicz, *An adaptive DPG method for high frequency time-harmonic wave propagation problems*, Comput. Math. Appl., 74 (2017), pp. 1999–2017.

[121] W. Prager and J. L. Synge, *Approximations in elasticity based on the concept of function space*, Quart. Appl. Math., 5 (1947), pp. 241–269.

[122] S. Prudhomme and J. T. Oden, *On goal-oriented error estimation for elliptic problems: application to the control of pointwise errors*, Comput. Methods Appl. Mech. Eng., 176 (1999), pp. 313–331.

[123] ——, *Computable error estimators and adaptive techniques for fluid flow problems*, in Error estimation and adaptive discretization methods in computational fluid dynamics, Springer, 2003, pp. 207–268.

[124] S. Prudhomme, J. T. Oden, T. Westermann, J. Bass, and M. E. Botkin, *Practical methods for a posteriori error estimation in engineering applications*, Int. J. Numer. Meth. Eng., 56 (2003), pp. 1193–1224.

[125] S. Repin, S. Sauter, and A. Smolianski, *Two-sided a posteriori error estimates for mixed formulations of elliptic problems*, SIAM J. Numer. Anal., 45 (2007), pp. 928–945.

[126] N. V. Roberts, *A discontinuous Petrov–Galerkin methodology for incompressible flow problems*, PhD dissertation, The University of Texas at Austin, Austin, Texas, U.S.A., 2013.

[127] N. V. Roberts, *Camellia: A software framework for discontinuous Petrov–Galerkin methods*, Comput. Math. Appl., 68 (2014), pp. 1581–1604.

[128] N. V. Roberts, *Camellia v1.0 manual: Part I*, Tech. Rep. ANL/ALCF-16/3, Argonne National Laboratory, Argonne, Illinois, 2016.

[129] N. V. Roberts, T. Bui-Thanh, and L. Demkowicz, *The DPG method for the Stokes problem*, Comput. Math. Appl., 67 (2014), pp. 966–995.

[130] N. V. Roberts, L. Demkowicz, and R. Moser, *A discontinuous Petrov–Galerkin methodology for adaptive solutions to the incompressible Navier–Stokes equations*, J. Comput. Phys., 301 (2015), pp. 456–483.

[131] J. Salazar, J. Mora, and L. Demkowicz, *Alternative enriched test spaces in the DPG method for singular perturbation problems*, ICES Report 18-15, 2018.

[132] J. Schöberl, *A posteriori error estimates for Maxwell equations*, Math. Comput., 77 (2008), pp. 633–649.

[133] M. Shephard, S. Dey, and J. Flaherty, *A straightforward structure to construct shape functions for variable p-order meshes*, Comput. Methods Appl. Mech. Engrg., 147 (1997), pp. 209–233.

[134] J. Sun, M. Smith, R. Armstrong, and R. Brown, *Finite element method for viscoelastic flows based on the discrete adaptive viscoelastic stress splitting and the discontinuous Galerkin method: DAVSS-G/DG*, J. Non-Newton. Fluid Mech., 86 (1999), pp. 281–307.

[135] D. B. Szyld, *The many proofs of an identity on the norm of oblique projections*, Numer. Algorithms, 42 (2006), pp. 309–323.

[136] L. N. Trefethen and D. Bau III, *Numerical linear algebra*, SIAM, Philadelphia, 1997.

[137] A. Vaziri Astaneh, F. Fuentes, J. Mora, and L. Demkowicz, *High-order polygonal discontinuous Petrov—Galerkin (PolyDPG) methods using ultraweak formulations*, Comput. Methods Appl. Mech. Engrg., 332 (2018), pp. 686–711.

[138] A. Vaziri Astaneh, B. Keith, and L. Demkowicz, *On perfectly matched layers for discontinuous Petrov–Galerkin methods*, ArXiv e-prints, arXiv:1804.04496 [math.NA], (2018).

[139] A. Vaziriastaneh, *On the Forward and Inverse Computational Wave Propagation Problems.*, PhD dissertation, North Carolina State University, Raleigh, North Carolina, USA, 2016.

[140] R. Verfürth, *A review of a posteriori error estimation and adaptive mesh-refinement techniques*, John Wiley & Sons Inc, New York, 1996.

[141] C. H. Whiting, K. E. Jansen, and S. Dey, *Hierarchical basis for stabilized finite element methods for compressible flows*, Comput. Methods Appl. Mech. Engrg., 192 (2003), pp. 5167–5185.

[142] C. Wieners, *The skeleton reduction for finite element substructuring methods*, in Numerical Mathematics and Advanced Applications ENUMATH 2015, Springer, 2016, pp. 133–141.

[143] J. Y. Yuan, *Numerical methods for generalized least squares problems*, J. Comput. Appl. Math., 66 (1996), pp. 571–584.

[144] J. Zitelli, I. Muga, L. Demkowicz, J. Gopalakrishnan, D. Pardo, and V. M. Calo, *A class of discontinuous Petrov–Galerkin methods. Part IV: The optimal test norm and time-harmonic wave propagation in 1D*, J. Comput. Phys., 230 (2011), pp. 2406–2432.