The Dissertation Committee for Sriram Nagaraj
certifies that this is the approved version of the following dissertation:

DPG Methods for Nonlinear Fiber Optics

Committee:

Leszek F. Demkowicz, Supervisor

Luis Caffarelli

Björn Engquist

Tan Bui-Thanh

Christopher Simmons

Ivo Babuška

DPG Methods for Nonlinear Fiber Optics

by

Sriram Nagaraj

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

May 2018

Dedicated to my family, *Guha Gruham*

Nadappathellam Nanmaike
Ancient Tamil Proverb

Acknowledgments

I thank my advisor, Dr. Leszek Demkowicz for his constant encouragement, insights and advice all through the years I have been at UT Austin. His advice, understanding and patience have been inspirational. I thank my committee members: Dr. Luis Caffarelli, Dr. Björn Engquist, Dr. Tan Bui-Than, Dr. Christopher Simmons and Dr. Ivo Babuška for graciously agreeing to serve on my dissertation committee. I thank the DPG group: Brendan, Federico, Socratis, Jaime, Stefan for important discussions and insights. In addition, I thank Brendan for our many discussions on veganism, fitness and politics and his infectious enthusiasm for research; Federico for our many afternoon coffee talks on a whole variety of topics: race, religion, politics, evolution, science, ethics etc.; Socratis for all the car rides back home, advice on research and life, Greek culture I have learned about, the infinite number of times he has helped me resolve bugs and his patient hearing of my many complaints about the USCIS; Jaime for helping me with the sum factorization; Stefan for our discussions on parallelism. I thank my collaborators: Dr. Jay Gopalakrishnan, Dr. Paulina Sepulveda and Dr. Jacob Grosek for their help and assistance. Thanks in particular to Jake: your insight into the physics of the Raman gain was vital to keep me on the right track. My thanks also to Stephanie and Lauren: your patience and help during these years at ICES have invaluable. Thanks to all the ICES staff for making my time at ICES a wonderful expe-

rience. Last, but certainly not the least, my family. Thanks Amma, Appa, Srikrishna and Nandhini for being there for me, with love, through thick and thin, and for believing in me even when I didn't believe in myself. I hope I have made you all proud. Nandhini, we had to wait for more than a year to be together again, I thank you for your love, support and patience.

DPG Methods for Nonlinear Fiber Optics

by

Sriram Nagaraj, Ph.D.
The University of Texas at Austin, 2018

Supervisor: Leszek F. Demkowicz

In recent years, the Discontinuous Petrov-Galerkin (DPG) method has been the subject of significant study. It comes with a collection of desirable properties, including uniform/mesh independent stability, localizable test norms via broken test spaces, and a canonical error indicator that is incorporated as part of the solution. In this work, the DPG method is applied to problems arising in fiber optics. Accurate modeling of wave propagation in nonlinear media is an important task in fiber optics applications. Nonlinear Maxwell equations in the context of optical fibers have been studied extensively in the past. Analysis of these intensity-dependent nonlinearities are based on several simplifying approximations which result in a nonlinear Schrödinger (NLS) type equation. The Schrödinger equation from a spacetime DPG perspective is discussed. In particular, a 2^{nd} order L^2 stable ultraweak formulation of the Schrödinger equation is constructed by introducing the notion of an auxiliary boundary operator. This theoretical device requires an

operator-specific conforming element to develop optimal convergence rates. Numerical studies show how, modulo (expected) roundoff issues, the theoretical convergence rates are delivered. Next, the use of the DPG method in modeling and simulating optical fiber laser amplifiers with nonlinear Raman gain is studied. In this application, the interaction of two time harmonic electromagnetic fields (the signal and pump fields) governed by two weakly coupled nonlinear Maxwell equations results in the amplification phenomenon. A novel Raman gain model for describing the phenomenon is proposed and an ultra weak DPG formulation is used for the discretization of the proposed model. The nonlinearity is handled by using simple iterations between the two systems. DPG implementation of a perfectly matched layer (PML) at the exit end of the fiber is essential in this model, as is the use of sum factorization for element computations. The presented results show that the signal field indeed gains power along the fiber, thereby justifying the use of the model. Auxiliary results presented in this dissertation include the construction of DPG Fortin operators for 2^{nd} order problems.

Table of Contents

Chapter 1

Introduction

Modeling and simulation of optical devices is a significant endeavor with a host of industrial and military applications. For instant, optical phenomena play a key role in laser-based biomedical devices, optical communication systems, laser guidance systems, optical metamaterials etc. Accurate simulations of optical devices used in these applications are critical in minimizing the potential damage that is possible with faulty design and/or implementations.

From the mathematical perspective, interesting optical phenomenon is often *nonlinear*. Indeed, optically active materials are used for exciting higher order optical nonlinearities that lead to both useful and harmful nonlinear effects such as stimulated Brillouin scattering, the Kerr effect, four-wave mixing self (and cross) modulation etc. It is thus important to amplify the useful effects while mitigating (or minimizing) the harmful effects.

Mathematical models of optical phenomenon are roughly of two kinds: time dependent models and time independent (i.e., time-harmonic) models. Both models start with the full vectorial set of Maxwell's equations, and derive distinct mathematical systems using different assumptions and approximations.

Time Dependent Models: These models often arise in modeling ultra-short pulse propagation in optical media, most commonly optical fibers. A common approach used in these models is to reduce the full vector Maxwell system to a nonlinear Schrödinger system in two variables (time and the spatial direction of propagation) using a slowly varying envelope approximation and dropping higher-order terms. The resulting Schrödinger system has often been solved using split-step Fourier methods

Time Harmonic Models: These models are used for wave propagation in long (several meter length) optical fibers where light travels at predominantly a single, fixed frequency. Using the cylindrical geometry of the optical fiber, one can decompose the three dimensional Maxwell system into a two-dimensional Helmholtz system in the transverse (radial) direction and a first- or second-order ordinary differential equation (ODE) in the direction of propagation. Approaches that use this technique are termed "beam propagation methods", and may use a combination of finite element and finite difference methods for solving the wave propagation problem.

The results in this dissertation address both these models from a *Discontinuous Petrov-Galerkin (DPG)* perspective. In recent years, the Discontinuous Petrov-Galerkin (DPG) method of Demkowicz and Gopalakrishnan ([23]) has afforded finite element method practitioners with a host of desirable properties: uniform/mesh independent stability, localizable test norms via broken test spaces and a canonical error indicator that is incorporated as

part of the solution. Moreover, the DPG philosophy is applicable, without any loss in its superior stability properties, to any well-posed weak formulation. In addition, the use of space-time elements within the DPG framework has been studied in the recent past ([32, 30]).

For linear problems, the DPG methodology has three equivalent points of view (the "three hats"), and each viewpoint sheds light on a specific aspect of the framework. In short, it is an optimally stable, minimum residual mixed method. As a minimum residual method, DPG achieves an orthogonal projection in the trial space, and it yields a Hermitian system *regardless* of the symmetry of the problem under question. The superior stability is obtained via inversion of the (approximate) *test space Riesz operator*. The use of *broken test spaces* allows for a conveniently parallel, element-wise local implementation. Finally, the mixed method viewpoint allows for efficient coding strategies of the resulting linear system of equations.

1.1 Goals of Dissertation: DPG and Applications in Optics

The main goals and contributions of this dissertation are in the application of the DPG method to problems arising in linear and nonlinear optics. This is motivated by the host of desirable properties of the DPG method and the promise of using the DPG method to solve challenging problems in fiber optics. In particular, both time dependent and time-harmonic approaches are considered in this dissertation.

1.1.1 Spacetime DPG for the Schrödinger Equation

The Schrödinger equation arises naturally in fiber optics applications as an accurate model of the amplitude variation of light propagating in a dispersive medium. However, using the DPG (or any Galerkin-like) methodology for numerically solving the Schrodinger equation has an immediate non-trivial difficulty: the time dependent Schrödinger equation has no natural L^2 stable first order space-time formulation. In addition, standard Sobolev space based trace theory is inapplicable. These two issues pose a signficant theoretical challenge for the development of stable numerical discretization of the Schrödinger system. In this dissertation, a 2nd order L^2 stable ultraweak formulation of the Schrödinger equation is developed by introducing the notion of an auxiliary boundary operator that hitherto was used, in less generality, for 1st order Friedrichs systems. This theoretical device requires an operator-specific conforming element to develop optimal convergence rates. The numerical studies in this dissertation show how, modulo (expected) roundoff issues, one can deliver the theoretical convergence rates.

1.1.2 A DPG Based Full Vector 3D Maxwell Raman Gain Model

Modeling optical fiber laser amplifiers is an important task with great military signficance. This thesis presents a unique full 3D Maxwell DPG simulation of a passive optical fiber amplifier that experiences stimulated Raman scattering. When calibrated as an amplifier, an optical fiber is used to convert optical power from a "pump" electromagnetic field to a "signal" electromag-

netic field. The signal field experiences a gain along the length of the fiber, and thereby increases in power, while absorbing power from the pump field, which in turn decays along the length of the fiber. In contrast with popular beam propagation models, this dissertation considers a truly *full vectorial approach*. This highly general approach, while computationally intensive, provides an ideal way to model a variety of nonlinear pheonomena as well as fiber orientations and pumping configurations, most of which are beyond the scope of traditional beam propagation approaches. The aim is to develop computational tools for the most general model with the fewest simplifying approximations, with the intent to eventually develop a high-fidelity, multi-physics fiber model that can handle much more complex problems with realistic fiber lengths. The primary interest is to establish the numerical approach and validate its feasibility by observing the qualitative characteristics of Raman gain. Towards that end, the superiority of ultraweak DPG formulation of the coupled Maxwell system is demonstrated numerically, and is implemented in the model using a DPG based perfectly matched layer (PML). The use of sum factorization for making the element computations tractable is critical to the success of this endeavor. It is successfully shown that a nonlinear iterative method is able to handle the nonlinear gain term. The work presented in this dissertation is, to the author's best knowledge, the *first full vector 3D Maxwell* simulation for optical fiber laser amplifiers with Raman gain in the context of Galerkin-based numerical discretization.

1.1.3 Other Contributions

Fortin operators arise naturally in the analysis of the DPG method. Indeed, the "ideal" DPG method is, in general, computationally intractable because of the need to invert an infinite dimensional test space Riesz operator, which in turn is an infinite dimensional optimization problem. Truncating the inversion (and thus optimization) procedure to a large, yet finite dimensional test space, yields a computationally feasible method, which is referred to as the "practical" DPG method. The change in stability properties while moving from the "ideal" to the "practical" scenario is quantified by appropriate Fortin operators. Discussed in this dissertation is the construction of DPG Fortin operators for 2nd order problems.

Generalizing DPG from the Hilbert space setting to a Banach space setting is an important first step for theoretical analysis of nonlinear schemes. Indeed, linearized versions of nonlinear PDEs are often lead to variational formulations defined on Banach spaces. Towards this end, an appendix to this dissertation has preliminary theoretical results in the analysis of minimizing residuals in L^p spaces instead of Hilbert spaces using an iteratively reweighted least squares approach.

1.2 Relavance of the Work

The work presented in this dissertation is of great relevance to the nonlinear optics community. Indeed, the spacetime DPG approach to the

Schrödinger equation is a novel theoretical contribution which extends the Friedrichs system approach to higher order operators. The Raman gain model and simulations are the first full vector 3D Maxwell simulation for optical fiber laser amplifiers with Raman gain in the context of Galerkin-based numerical discretization. The simulation of Raman gain with DPG required not only the implementation of a complex stretched PML, but also sum factorization of element computations for speeding up the element-wise integration of DPG field variables. The many positive features of the DPG methodology can be used fruitfully for stable, adaptive solutions to the nonlinear problems arising in the optics literature.

1.2.1 Outline of Dissertation

After this introductory chapter, a brief overview of the equations of nonlinear optics specific to the goals of the dissertation is presented in chapter 2. In particular, both the time dependent nonlinear Schrödinger and time harmonic Maxwell approaches are discussed. Chapter 3 will be a rapid introduction to the core ideas of the DPG framework and the construction of DPG Fortin operators along with numerical studies. Chapter 4 will focus on spacetime DPG for the Schrödinger equation, where both theoretical stability properties as well as a novel operator-specific finite element space construction are considered. Chapter 5 derives the Raman gain model in the DPG context while 6 is devoted to the numerical simulation studies of the proposed Raman model. Chapter 7 is a comprehensive conclusion to this dissertation. Adequate

review of relevant literature is provided on a chapter by chapter basis. The appendices provide a comparison of the primal and ultraweak formulations of Maxwell's equations, details of the DPG implementation of the perfectly matched layer used in the Raman model simulations, details of the sum factorization used in the 3D Maxwell simulations and miscellaneous results used in the construction of the Fortin operators. The final appendix contains the iteratively reweighted least squares approach to minimizing L^p residuals. A detailed list of references follows the appendices.

1.2.2 Acknowledgements

The work presented in this dissertation has been made possible by UT Austin's ICES program CSEM Fellowship, Teaching Assistant positions in UT Austin's ICES program, DOD Air Force (AFOSR) grants FA9550-12-1-0484 P00002, DOD Air Force (AFOSR) Air Force Research Lab FA9550-17-1-0090 and National Science Foundation (NSF) grant DMS-1418822. The support is much appreciated.

Chapter 2

An Overview of the Equations of Nonlinear Optics

In this chapter, we provide a self-contained overview of the equations of nonlinear optics that we shall deal with in this dissertation. We hasten to add that this chapter is meant for the purpose of collecting together the required details of the overarching physics that motivates the mathematical and numerical studies in this dissertation.

2.1 Background

We begin our discussion with the classical fomulation of Maxwell's equations in dielectrics with no magnetic polarization. Most of the discussion in this chapter has been motivated by and based on the presentation in the books [2, 1, 7, 75]. We begin with the time dependent formulation before specializing to the time harmonic case. Let Ω be a bounded Lipschitz domain in $\mathbb{R}^d$, where $d = 2, 3$. The classical form of Maxwell's equations is:

$$\nabla \times \mathbb{E} = -\mu_0 \frac{\partial \mathbb{H}}{\partial t},$$
$$\nabla \times \mathbb{H} = \epsilon_0 \frac{\partial \mathbb{E}}{\partial t} + \frac{\partial \mathbb{P}}{\partial t},$$
$$\nabla \cdot \mathbb{E} = \frac{\rho}{\epsilon_0},$$
$$\nabla \cdot \mathbb{H} = 0,$$

where, as usual, $\mathbb{E}, \mathbb{H}$ are the electric and magnetic fields, ϵ_0 is the free-space electric permittivity, μ_0 is the free-space permeability, ρ the electric charge density and $\mathbb{P}$ the electric polarization vector. Since we will be interested in the case of nonlinear optical fibers, the assumption that the magnetic polarization is absent is a valid one.

In the linear situation, the electric polarization vector $\mathbb{P}$ is proportional to the electric field $\mathbb{E}$:

$$\mathbb{P} = \chi_1 \epsilon_0 \mathbb{E}$$

where χ_1 is the electric susceptibility and ϵ_0 denotes the permittivity of free space. We can then speak of the relative permittivity ϵ_r and the permittivity (dielectric constant) $\epsilon = \epsilon_0 \epsilon_r$. Incorporation of the permittivity into the free space version of Maxwell's equations simply means the use of ϵ in place of ϵ_0.

Appropriate sources and boundary conditions for the Maxwell system depend both on the geometry of the domain in question as well as the kind of physics being modeled. For instance, free space propagation involves radiation boundary conditions, which requires that the electromagnetic fields decay as

the distance from the source goes to infinity. Absorbing boundary conditions such as impedance and/or Robin type boundary conditions arise in waveguide problems. Perfectly matched layers (PML) are often implemented to simulate infinite domain problems (see [6, 79, 16, 43, 60, 61, 81]).

When dealing with an optical fiber, wherein the fiber is made of silica material (glass), we have a *refractive index* tensor $\mathbf{n}$, and the so-called linear background polarization takes the form:

$$\mathbb{P}_{\text{background}} = (\mathbf{n}^2 - \mathbb{I})\epsilon_0 \mathbb{E},$$

where $\mathbb{I}$ is the identity tensor.

2.1.1 Nonlinear Optics

Wave propagation in optically active media is much more exotic than simple linear wave propagation. These nonlinearities give rise to both useful and harmful nonlinear effects such as Raman scattering, stimulated Brillouin scattering, the Kerr effect, four-wave mixing self (and cross) modulation etc. Simulating these nonlinear phenomena is an important step to understand how to make use of the positive effects while mitigating (or atleast minimizing) the harmful effects. The most common nonlinear model used in the literature is the nonlinear dependence of the polarization vector upon the electric field:

$$\mathbb{P} = \epsilon_0(\chi_1 \cdot \mathbb{E} + \chi_2 : \mathbb{E} \otimes \mathbb{E} + \chi_3 \vdots \mathbb{E} \otimes \mathbb{E} \otimes \mathbb{E}), \tag{2.1.1}$$

where χ_i, $i = 1, 2, 3$ are the first, second and third order susceptibility tensors. Note that the above is a commonly used third order approximation: one can include further higher order susceptibility terms in the expansion of the polarization vector ([2, 7]). This can be expressed alternatively as:

$$\mathbb{P} = \mathbb{P}_L + \mathbb{P}_{NL}.$$

where $\mathbb{P}_L$ takes into account the linear parts of the polarization while $\mathbb{P}_{NL}$ accounts for the nonlinear part of the polarization. The linear part models, for instance, the background polarization, while the nonlinear polarization models effects such as Raman and/or Brillouin scattering.

We can now eliminate the magnetic field and obtain the following wave equation for the electric field:

$$\nabla \times \nabla \times \mathbb{E} = -\frac{1}{c^2}\frac{\partial^2 \mathbb{E}}{\partial t^2} - \mu_0 \frac{\partial^2 \mathbb{P}}{\partial t^2}, \tag{2.1.2}$$

which accounts for the total (linear + nonlinear) contributions of the electric polarization $\mathbb{P}$. The above fundamental equation describes the behaviour of light within an optical fiber.

2.1.2 The Schrödinger Equation in Nonlinear Optics

We now develop the fundamental approximation technique used widely in fiber optics. Consider the time dependent Maxwell system. The principal model of Maxwell equations in fiber optics, (see [2]), adopts the slowly varying envelope approximation:

$$\mathbb{E}(\mathbf{r},t) = \frac{1}{2}\hat{\mathbf{x}}(E(\mathbf{r},t)e^{i\omega_0 t} + \text{c.c.}), \tag{2.1.3}$$

where $\hat{\mathbf{x}}$, is the polarization unit vector, $E(\mathbf{r},t)$ is a slowly varying function of time (relative to the central "carrier" frequency ω_0) and c.c. stands for the additional complex conjugate term. Neglecting the Raman effects, and after several additional assumptions and approximations (which may not be valid under all circumstances), we arrive at an approximate solution in cylindrical coordinates (r,θ,z):

$$\mathbb{E}(r,\theta,z) = \frac{1}{2}\hat{\mathbf{x}}(F(r,\theta)A(z,t)e^{i(\omega_0 t-\beta_0 z)} + \text{c.c.}). \tag{2.1.4}$$

The function $F(r,\theta)$ satisfies the eigenvalue problem:

$$\frac{1}{r}\frac{\partial}{\partial r}(\frac{\partial F}{\partial r}) + \frac{1}{r^2}\frac{\partial^2 F}{\partial \theta^2} + (\epsilon(\omega)k_0^2 - \beta^2) = 0, \tag{2.1.5}$$

where $k_0 = \frac{\omega}{c}$ is the wave number and $\beta = \beta(\omega)$ is the frequency dependent eigenvalue of the nonlinear problem viewed as a perturbation of the corresponding eigenvalue of the linear problem. We remark that β can also be viewed as a (frequency dependent) propagation constant.

Finally, the complexified amplitude function $A(z,t)$ can be shown to satisfy a nonlinear Schrödinger (NLS) type equation (we refer the reader to [2, 75] for more discussions and analysis):

$$\frac{\partial A}{\partial z} + \beta_1\frac{\partial A}{\partial t} + i\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} + \frac{\alpha}{2}A = i\gamma(\omega_0)|A|^2 A. \tag{2.1.6}$$

Here, $\beta_1 = \beta'(\omega_0), \beta_2 = \beta''(\omega_0)$, and $\alpha = \alpha(\omega_0)$ is the attenuation constant. The parameter $\gamma(\omega_0)$ is a computable function of $F(r, \theta)$. Note that the variable t is in fact a moving observation window, and not absolute time.

The NLS equation and its generalizations have been the basis for various theories in pulse propagation in nonlinear optics, providing a basis for simulations of lasers, four wave mixing, self/cross phase modulation, and many other topics [2, 1].

2.1.3 Time Harmonic Models

We can now consider a time harmonic version of the Maxwell system by using an ansatz of the form:

$$\mathbb{E}_0(x, y, z, t) = \mathbb{E}(x, y, z)e^{i\omega t},$$

$$\mathbb{H}_0(x, y, z, t) = \mathbb{H}(x, y, z)e^{i\omega t}.$$

This implies that the polarization too behaves as:

$$\mathbb{P}_0(x, y, z, t) = \mathbb{P}(x, y, z)e^{i\omega t}.$$

The time harmonic version of the Maxwell system of equations becomes:

$$\begin{aligned} \nabla \times \mathbb{E} &= -i\omega\, \mu_0 \mathbb{H}, \\ \nabla \times \mathbb{H} &= i\omega\, \epsilon_0 \mathbb{E} + i\omega \mathbb{P}, \\ \nabla \cdot \mathbb{E} &= \tfrac{\rho}{\epsilon_0}, \\ \nabla \cdot \mathbb{H} &= 0. \end{aligned} \tag{2.1.7}$$

The linear part $\mathbb{P}_L$ still depends upon the electric field through a linear relation but, in the frequency domain, the magnetic susceptibility χ_1 depends upon

frequency ω. Additionally, for anisotropic materials, susceptibility becomes tensor-valued.

The first step in solving equation 2.1.2 is to assume the nonlinear contributions $\mathbb{P}_{NL}$ to the polarization $\mathbb{P}$ is a small perturbation to the total polarization. Thus, we first solve equation 2.1.2 assuming $\mathbb{P}_{NL} = 0$, and then use standard perturbation theory to account for the nonlinear part. With $\mathbb{P}_{NL} = 0$, we obtain the following relation in the time harmonic case.

$$\nabla \times \nabla \times \mathbb{E} + \epsilon(\omega)\frac{\omega^2}{c^2}\mathbb{E} = 0.$$

The situation gets more complicated when nonlinear effects are included. After Taylor expansion of $\mathbb{P}$ at zero, quadratic and/or cubic terms are kept resulting in different nonlinear versions of the Maxwell equations.

Frequency Dependence of Refractive Index

Considering again, for a moment, the case of $\mathbb{P}_{NL} = 0$, we readily see the frequency dependence of the the dielectric constant $\epsilon(\omega) = 1 + \chi_1(\omega)$. Since in general $\chi_1(\omega)$ may be complex the dielectric constant $\epsilon(\omega)$ may also be complex. We can now write (see [2] for complete derivations) the following relation between $\epsilon(\omega)$ and the (linear) refractive index n_L of the material:

$$\epsilon(\omega) = (n_L(\omega) + ic\frac{\alpha_L(\omega)}{2\omega})^2,$$

where α_L is the frequency dependent (linear) absorption coefficient, $i = \sqrt{-1}$, and c is the speed of light in the material. We can then express the refractive index in terms of χ_1 as follows:

$$n_L(\omega) = 1 + \frac{1}{2}\text{Real}\{\chi_1(\omega)\},$$
$$\alpha_L(\omega) = \frac{\omega}{cn(\omega)}\text{Imag}\{\chi_1(\omega)\}.$$

The fundamental complication of the fully nonlinear equation is the additional dependence of $n_{NL}(\omega)$ on the *amplitude* of the wave:

$$n_{NL}(\omega) = n_L(\omega) + n_2|E|^2,$$

where $n_2 = \frac{3}{8n_L}Re\{\frac{\partial^4\chi_3}{\partial x^4}\}$. Likewise, a similar relation exists for the nonlinear absorption coefficient α_{NL}.

Physically, this means that in the nonlinear situation, different parts of the wave will "see" different refractive indices depending on the amplitude. The dependence of refractive index on the amplitude at any given spatial location leads to self-modulation of the wave: the beam within an optical fiber interacts with itself, because different parts of the wave move with different velocities and hence will self-intersect. In other words, parts of the wave with higher amplitude move *faster* in the material while conversely, parts of the wave with lower amplitude move *slower*. This self-modulation is an important feature of nonlinear wave propagation in optical fibers.

Linear Fiber Modes

The discussions in this subsections pertain to so-called "fiber modes". We refer the reader to [2, 7, 75] and references therein for more elaborate descriptions of the following content. Consider now light propagation in an circulary symmetric optical fiber. The geometry of the fiber can be modeled as a long, right circular cylindrical rod. It consists of two regions: a cylindrical core surrounded by a hollow cylindrical cladding. The radius of the core is denoted by r_{core} and the radius of the cladding by r_{cladding}. Both core and cladding are usually made of silica. The refractive index of the core (n_{core}) is larger than the refractive index of the cladding (n_{cladding}), which thereby allows for total internal reflection (i.e., by Snell's law) at the core-cladding interface which in turn is the means by which the entire optical fiber can behave as a waveguide. Reducing the full set of Maxwell's equations to a Helmholtz system allows for defining and computing modes of an optical fiber. With appropriate continuous boundary conditions, the optical fiber waveguide can support multiple modes. Exact closed form solution of fiber modes are generally derived by a separation of variables approach, and using the cylindrical symmetry of the fiber, one can derive Bessel function solutions to the resulting eigenvalue problem. We refer the reader to [2, 1] for details on fiber modes. In our work, we are interested in the general case, and we avoid explicit modal descriptions. We also remark that the techniques used to compute the fiber modes motivate the so-called beam propagation methods (BPMs see [68, 74, 63, 83, 4] and references therein). Though both semi-vectorial and full vectorial BPM

approaches have been implemented (see [49, 50, 71, 72, 36] and references therein), we are interested in a fiber model that is a full Maxwell boundary value problem rather than a Helmholtz reduction. We define a non-dimensional quantity called the numerical aperture (N.A.) as:

$$\text{N.A.} := \sqrt{n_{\text{core}}^2 - n_{\text{cladding}}^2}.$$

For a time-harmonic wave propagating at frequency ω with a wavelength λ, one can define the so-called "V-number" or normalized frequency as:

$$\text{V} := 2\pi \frac{r_{\text{core}}}{\lambda} \text{N.A.}.$$

One can show (see [2, 1]) that fibers tuned such that $\text{V} < 2.405$ support exactly a single mode (the "fundamental" or LP_{01} mode). Higher LP_{mn}, $m = 2, \ldots$, $n = 1, \ldots$ also exist, but are involved in inter-modal instabilities. Finally, it is well known ([2]) that the fundamental mode can be well-approximated by a Gaussian radial profile.

Chapter 3

The DPG Philosophy and its Three Hats

This chapter is devoted to a detailed review of the DPG methodology[1]. We also highlight the construction of DPG Fortin operators for second order problems in the latter part of this chapter.

Author contributions: The contents of this chapter are taken largely from the published multi-author article [67] which is co-authored by the author of this dissertation. The author of this dissertation contributed to the development of the theory and numerical results presented in [67], including the mathematical constructions/derivations as well as the writing of the manuscript.

3.1 Introduction

The DPG methodology, established by Demkowicz and Gopalakrishnan [25], enjoys a host of desirable properties that have been theoretically and numerically explored, and validated in the recent past. The theoretical

[1]The material in this chapter is taken largely from the published work [67], Copyright (c)2017 Elsevier. All rights reserved.

foundations of the subject were established in [23, 24, 12, 46, 13]. This active area of research has been successfully employed to problems in linear elasticity [52, 41], time harmonic wave propagation (including DPG versions of the PML) [69, 81], compressible and incompressible Navier-Stokes [32, 30, 15, 31], fluid flow [54], viscoelasticity [39] and space-time formulations [27, 48, 32, 29]. Moreover, versions of DPG for polygonal meshes have been introduced in [80]. Theoretical advances in goal-oriented adaptivity using DPG have been done in [56]. Practical implementation issues regarding conditioning of DPG systems are addressed in [55]. Coupling of different DPG formulations with each other is studied in [41] while coupling DPG with standard Galerkin methods is considered in [42].

Indeed, the ideal DPG method (with optimal broken test functions) has been shown to provide a uniform, mesh-independent stable discretization for *any* well-posed variational formulation [23, 13]. The computationally tractable, practical DPG method [46], upon discretization of the so-called trial-to-test operator, retains the guaranteed stability with a numerically estimable stability constant [67]. The DPG method uses element-wise defined test spaces with no global conformity ("broken" test spaces), which allow for parallelism. Since the method can be recast as minimum residual, and also a mixed method with a built-in error indicator (the residual), one can have automatic *hp* adaptivity starting from an arbitrarily coarse mesh, which has importance in problems involving singularities. Finally, the method always delivers a sparse Hermitian (symmetric) system making iterative conjugate

gradient based solvers ideal for large systems that cannot be handled by direct solvers [69, 47, 5, 70].

A generic variational formulation of a second order equation is usually posed in the following way. Given a continous bilinear form $b(\cdot,\cdot)$ defined on reflexive Banach spaces U, V, and a continuous linear functional $l(\cdot)$ on V, we seek a solution $u \in U$ of the problem:

$$b(u,v) = l(v), \forall v \in V. \tag{3.1.1}$$

It is clear that the bilinear form $b(\cdot,\cdot)$ generates two operators $B : U \to V'$ and $B' : V \to U'$ defined canonically as

$$\langle Bu, v\rangle_{V'\times V} = b(u,v), v \in V, \tag{3.1.2}$$

$$\langle B'v, u\rangle_{U'\times U} = b(u,v), u \in U. \tag{3.1.3}$$

Thus, our initial variational formulation (3.1) is fully equivalent to the operator equation $Bu = l$.

We assume that the continous inf-sup condition on $b(\cdot,\cdot)$ holds:

$$\gamma = \inf_{u\in U}\sup_{v\in V}\frac{|b(u,v)|}{\|u\|_U\|v\|_V} > 0, \tag{3.1.4}$$

where the constant γ is the inf-sup constant.

The inf-sup condition is simply Banach's closed range theorem in disguise: the inf-sup constant γ corresponds to the "boundedness below" constant of the closed range theorem [22].

Solving the variational problem proceeds by introducing finite dimensional trial and test subspaces U_h, V_h of U, V respectively and finding $u_h \in U_h$ that solves the discrete problem:

$$b(u_h, v_h) = l(v_h), \forall v_h \in V_h. \tag{3.1.5}$$

The question of well-posedness of the discrete problem is governed by Babuška's theorem ([3, 22]) which guarantees that if the *discrete* inf-sup condition holds, i.e., if

$$\gamma_h = \inf_{u_h \in U_h} \sup_{v_h \in V_h} \frac{|b(u_h, v_h)|}{\|u_h\|_U \|v_h\|_V} > 0, \tag{3.1.6}$$

then the discrete problem is well-posed.

3.1.1 Hat 1: Optimal Test Functions

Unfortunately, arbitrary choices of discrete trial and test spaces may lead to lack of discrete stability. The "optimal test function" hat of the DPG philosophy seeks to answer the following question: can one *choose* the test space in such a way to ensure stability? Said differently, can we construct the space V_h for a given U_h in such a way that we acheive the discrete inf-sup condition?

At this stage, the standard DPG approach specializes to the Hilbert setting with U, V being Hilbert spaces.

Now, for each $w \in U$, it is clear that Bw is a functional on V and hence, by the Riesz representation theorem, there is a unique $\hat{w} \in V$ such that

$$\langle Bw, v \rangle_{V' \times V} = (\hat{w}, v)_V. \tag{3.1.7}$$

By defining $T : U \to V$ to be the operator that assigns to each $w \in U$ the unique image under the Riesz map R_V of Bw, we generate the so-called "optimal test space". For obvious reasons, we refer to $T = R_V^{-1}B$ as the "trial-to-test" operator and T is characterized by $(Tu, v)_V = b(u, v)$.

At the discrete level, *given any* discrete trial space U_h, if we define

$$V_h^{\text{opt}} := T(U_h), \tag{3.1.8}$$

and we consider the Petrov-Galerkin problem with the optimal test space, we guarantee discrete inf-sup condition *by construction*.

3.1.2 Hat 2: Minimum Residual Formulation

The minimum residual hat of the DPG philosophy begins by considering the residual $Bu - l \in V'$ of the operator equation $Bu = l$. Clearly, the residual is zero if $u = u_h$. From this point of view, we consider the problem of finding $u_h \in U_h$ as that of minimizing a "residual" error. Indeed, solving $Bu = l$ in U_h means that we make an error of $Bu_h - l$ (the "residual")

It is thus natural to seek a solution $u_h \in U_h$ that minimizes the residual $Bu_h - l$. We thus consider

$$u_h = \text{argmin}_{w_h \in U_h} \frac{1}{2} \|Bw_h - l\|_{V'}^2, \tag{3.1.9}$$

and, using the isometric property of the Riesz map, it follows that

$$u_h = \text{argmin}_{w_h \in U_h} \frac{1}{2} \|R_V^{-1}(Bw_h - l)\|_V^2. \tag{3.1.10}$$

Taking a Gateâux derivative in the direction $\delta u_h \in U_h$, we see that the minimizer u_h satisfies

$$(R_V^{-1}(Bu_h - l), R_V^{-1}B\delta u_h)_V = 0, \tag{3.1.11}$$

and, noticing that $R_V^{-1}B\delta u_h = T\delta u_h$ where T is the trial-to-test operator, we conclude that

$$(R_V^{-1}(Bu_h - l), \delta v_h^{opt})_V = 0, \tag{3.1.12}$$

where $\delta v_h^{opt} = T(\delta u_h)$ are the optimal test functions corresponding to δu_h.

3.1.3 Hat 3: Mixed Formulation

Taking the formulation

$$(R_V^{-1}(Bu_h - l), R_V^{-1}B\delta u_h)_V = 0 \tag{3.1.13}$$

as our starting point, we define the "error representation function" $\psi = R_V^{-1}(Bu_h - l)$. Treating ψ which as an additional unknown, we arrive at a mixed formulation, first discussed by Dahmen et. al. [18]:

$$\begin{cases} u_h, \psi & u_h \in U_h,\ \psi \in V \\ (\psi, v)_V - b(u_h, v) = -l(v), & v \in V \\ b(w_h, \psi) = 0. & w_h \in U_h \end{cases} \tag{3.1.14}$$

Clearly, the mixed formulation is equivalent to the mixed formulation, with ψ being the Riesz inverse of the residual, i.e., $\psi = R_V^{-1}(Bu_h - l)$.

3.1.4 Additional Remarks on the DPG Methodology

It is clear that the DPG approach has three equivalent points of view. Some further remarks are in order. First, being a minimum residual method, we are guaranteed a symmetric (or Hermitian) stiffness matrix. As a Ritz method, we do not suffer from pre-asymptotic instability. Furthermore, if we equip the trial space U with the *energy norm* $\|u\|_E = \|Bu\|'_V$, the DPG method delivers the best approximation error (BAE). In addition, the FE approximation error $\|u - u_h\|_E = \|Bu - Bu_h\|_{V'} = \|R_V^{-1}(l - Bu_h)\|_V = \|\psi\|_V$, where ψ is the error representation function. Now, if viewed from the mixed method point of view, this shows that there is no need for a separate error indicator: the method comes with a cannonical error indicator. Thus, one can start adaptive refinements with this natural aposteriori error estimator.

However, this is the ideal situation, and we assume we can compute the optimal test functions exactly. A practical, computationally tractable approach involves approximating the Riesz map while computing the optimal test space. The ramifications of the practical DPG methodology are discussed in [46],[67].

3.1.5 Practical DPG Method

Recall that the distinguishing feature of the DPG methodology is the use of problem-dependent optimal test functions. As we saw in the introductory section, determination of optimal test functions involves computing the inverse of the Riesz map R_V defined on the entire test space V. However,

since V is infinite dimensional, inverting R_V is a computationally intractable problem. In practice, one replaces V with a large yet finite dimensional "enriched" test space $V_r \subset V$ and one determines the optimal test functions on V_r instead of V. The loss of stability by such a "practical" DPG scheme has been analyzed in the context of Fortin operators in [67, 46, 13].

A second computational issue is that of *localizability*. Given a well-chosen test space norm $\| \|_V$, it is possible to compute the total norm of a vector $v \in V$ as the summation of its contribution *elementwise*. When this is the case, we say that the test norm is *localizable*. Clearly, localizability is a highly desirable property, especially since this implies that the inversion of the Riesz map and determination of optimal test functions can be done on different elements independently, thereby making the entire computation parallelizable. In order to put the notion of localizability on firm mathematical ground, one must define the notion of *broken test spaces*.

3.1.6 Broken Test Spaces

A broken variational formulation arises naturally by taking an unbroken formulation (such as a standard Galerkin formulation), and introducing an elementwise defined (thus "broken") test space. The use of broken test spaces is required, in the DPG context, for ensuring that the Riesz map inversion can be performed elementwise with a localizable test norm (see [13]). Such use of a broken test space results in additional interface unknowns (the "trace" variables) at the element level, which must also be solved for. However, the entire

procedure becomes local, and one can thereby invert, at the element level, the enriched Riesz operator. Moreover, the test norm becomes localizable. Thus, the extra computational cost of additional unknowns is a price worth paying for ensured discrete stability. We refer the reader to [13] for an in-depth analysis of the concept and implications of using broken test spaces. We define the notion of an (abstract) broken variational formulation. A (continuous) broken variational formulation consists of a quadruple (U, V, b, l), where U, V are Hilbert spaces (called the trial and test spaces respectively), b is a continuous bilinear (or sesquilinear) form on $U \times V$ and l is a continuous linear (or conjugate-linear) form on V. The Hilbert space U is usually presented as a product of Hilbert spaces $U_0 \times \hat{U}$, while the bilinear form $b(\cdot, \cdot)$ decomposes as

$$b((u, \hat{u}), v) = b_0(u, v) + \hat{b}(\hat{u}, v)$$

with $b(\cdot, \cdot), \hat{b}(\cdot, \cdot)$ being continuous bilinear (or sesquilinear) forms on $U \times V$ and $\hat{U} \times V$ respectively. Here, U_0 corresponds to the space of "field" variables while $\hat{U}$ is the interface space of trace variables. Given such a quadruple (U, V, b, l) the variational problem we are interested in is the following. Find $(u, \hat{u}) \in U$ such that for all $v \in V$, we have:

$$b((u, \hat{u}), v) = l(v). \tag{3.1.15}$$

The broken weak formulations of most second order equations arising in physical applications can be cast in the above abstract setting [13].

A proper understanding of the well-posedness (i.e., existence, uniqueness and stability) of such variational formulations is important to determine

optimal discretization schemes. In order to determine when such an abstract broken formulation is well-posed, we make the following two assumptions:

Assumption 1 $b_0(\cdot,\cdot)$ satisfies the inf-sup condition, i.e., there exists a $\gamma > 0$ such that for all $(u, v) \in U_0 \times V$, we have:

$$\gamma \leq \inf_{u\neq 0} \sup_{v\neq 0} \frac{|b(u,v)|}{\|u\|_{U_0}\|v\|_V}$$

Assumption 2 Define

$$V_0 := \{v \in V : \hat{b}(\hat{u}, v) = 0 \ \forall \hat{u} \in \hat{U}\}.$$

With this V_0, we must ensure the triviality of the kernel Z_0, which is defined as

$$Z_0 = \{v \in V_0 : b_0(u, v) = 0 \ \forall u \in U_0\}.$$

Finally, we assume $\hat{b}(\cdot,\cdot)$ satisfies the inf-sup condition, i.e., there exists a $\hat{\gamma} > 0$ such that for all $(\hat{u}, v) \in \hat{U} \times V$, we have:

$$\hat{\gamma} \leq \inf_{\hat{u}\neq 0} \sup_{v\neq 0} \frac{|\hat{b}(\hat{u},v)|}{\|\hat{u}\|_{\hat{U}}\|v\|_V}$$

Theorem 3.1 of [13] ensures that with assumptions (1) and (2), we have a well-posed variational problem corresponding to the quadruple (U, V, b, l). In the sequel, we will, by abuse of notation, refer to the quadruple (U, V, b, l) itself as the broken variational formulation in place of the broken variational problem defined by the quadruple. Henceforth, we assume that assumptions (1) and (2) hold and we have identified a well-posed broken variational formulation.

All of our previous discussions of the DPG method hold in the broken case, and henceforth, we assume that the variational formulations in questions are broken formulations with appropriate field and trace spaces.

3.2 Ideal vs Practical

The notion of optimal test functions we have considered can be referred to as the "ideal" optimal test functions, since they guarantee discrete stability for any choice of trial space. While desirable, the computation of ideal optimal test functions involves an infinite dimensional optimization problem which is computationally prohibitive. In order to obtain a more realistic set of optimal test functions, one needs to approximate the trial-to-test operator T by another operator $T_r : U \to V_r$, where V_r is a large but finite dimensional subspace of V. The analysis of the "practical" DPG method was done in [46]. T_r satisfies

$$(T_r u, r)_V = b(u, v_r),\ v_r \in V_r, \tag{3.2.1}$$

in other words, we restrict the inversion of the Riesz operator to V_r to obtain $T_r = R_{V_r}^{-1} B$ [46][45]. The question of computing of T_r is now considered only over the finite dimensional subspace V_r and hence is a tractable quest. A natural question to ask at the moment is: how does the move from ideal to practical optimal test functions affect stability, and can this be quantified in some way? This chapter attempts to provide an answer to this question.

3.2.1 Fortin Operator

We now address the question of measuring the change in stability while moving from the ideal to practical notions of optimal test spaces. In order to do so, we introduce the idea of a Fortin operator [9][46]. A Fortin operator $\Pi : V \to V_r$ is a linear map that satisfies the conditions:

$$\begin{cases} \|\Pi v\|_V \leq C(r)\|v\|_V & 0 < C(r) < \infty,\ v \in V \\ b(w_h, \Pi v - v) = 0. & w_h \in U_h,\ v \in V \end{cases} \tag{3.2.2}$$

where $C(r)$, which we shall call the Fortin constant, is the operator norm of Π and depends, in particular, on the dimension of V_r. The second condition can be viewed as a b-orthogonality requirement on $\Pi v - v$. We then have the following relation [46] that shows how the stability is altered due to using the practical optimal test functions.

$$\|u - u_h\|_U \leq \frac{\|b\|}{\gamma_h} C(r) \inf_{w_h \in U_h} \|u - w_h\|. \tag{3.2.3}$$

It is therefore clear that we would like to have the constant $C(r)$ as close to unity as possible to ensure the least loss of stability.

3.2.2 Chapter Aims

The main aim of this chapter is to study the variation of the stability of the discontinuous Petrov-Galerkin method while changing from ideal to practical test functions using a suitable Fortin operator as a means to do so. As shown in [46, 14, 13, 67], the existence of a continuous Fortin operator ensures

(in fact, is equivalent to) the discrete stability of the variational problem. We shall restrict ourselves to the two dimensional case, and we shall provide the construction for the H^1 and $H(\mathrm{div})$ spaces using the Helmholtz and acoustic equations as motivation. Using discontinuous test functions and scaling arguments, we reduce our construction to be on a master triangular element and derive sufficient conditions to solve for the Fortin operator. We take a two-prong approach to the analysis of the Fortin operator. First, we derive an *upper bound* on the Fortin constant using the inf-sup constant γ_h associated with an *auxiliary* bilinear form. We are able to construct only an upper bound for the Fortin constant since a direct computation of the Fortin constant is not possible: evaluating the norm of the Fortin operator involves an infinite dimensional optimization problem. Based on this upper bound, we consider a numerical procedure that estimates γ_h and we thereby obtain an order of magnitude estimate on the Fortin constant. As a second line of analysis, we construct a sequence of *approximate* Fortin operators, each member of which is defined on an increasingly larger, yet finite dimensional subspace of the trial space. We then *exactly* compute the continuity constants of the *approximate* Fortin operators, which presumably converge to the *exact* Fortin constant of the *true* Fortin operator. We have not investigated a rigorous proof of the convergence of the approximate Fortin operators to the exact Fortin operator, however, the approximate Fortin constants provide a lower bound to the exact Fortin constant. In summary, our aim is to approximate the Fortin constant from above and below, thereby yielding a numerical range of how the overall

stability of the DPG method is affected by using practical test functions.

In [46], the authors provide the construction of a Fortin operator arising from the Poisson problem, and show the corresponding stability of the discrete method, although an explicit value of the Fortin constant is not provided. As we have indicated, it is very desirable to have an order of magnitude estimate on the Fortin constant to conclude how well the optimal test functions are resolved. Also, the value of the Fortin constant, and especially its dependence on the order p, indicates the possibility of hp-adaptivity. In particular, we will be interested in the p-(in)dependence of the Fortin constant.

3.2.3 Organization of Chapter

After this preliminary introduction, we detail our H^1 construction of the Fortin and approximate Fortin operators in Section 3.3, along with a numerical procedure of estimating their corresponding continuity constants. Section 3.4 details a similar construction and analysis for the $H(\text{div})$ case. The final section details our numerical results.

3.3 Construction of H^1 DPG Fortin operator

We shall construct and analyze our Fortin operator defined on a broken H^1 space. To a large extent, the operator construction is problem independent, and is applicable to a general class of second order linear problems. To be more concrete, we shall use the Helmholtz equation as a motivating example for the construction.

Notation and Mesh Assumptions Let $\Omega \subset \mathbb{R}^2$ be a polygonal domain. We shall consider the standard energy spaces:

$$H^1(\Omega) := \{u \in L^2(\Omega) : \nabla u \in (L^2(\Omega))^2\},$$
$$H(\mathrm{div}, \Omega) := \{u \in (L^2(\Omega))^2 : \nabla \cdot u \in L^2(\Omega)\},$$

where the operators ∇ and $\nabla\cdot$ are understood in the sense of distributions. Given a triangular mesh $\mathcal{T}_h$ covering Ω and an integer $p \geq 1$, we consider the standard Finite Element (FE) spaces:

$$W^p := \{u \in H^1(\Omega) : u|_K \in \mathcal{P}^p(K),\ K \in \mathcal{T}_h\},$$
$$RT^p := \{u \in H(\mathrm{div}, \Omega) : u|_K \in (\mathcal{P}^{p-1}(K))^2 \bigoplus x\widetilde{\mathcal{P}}^{p-1}(K)\ , K \in \mathcal{T}_h\},$$

where $\mathcal{P}^p(K)$ denotes the polynomials of (total) order less than or equal to p, and $\widetilde{\mathcal{P}}^{p-1}(K)$ denotes the subspace of homogeneous polynomials of order p.

The H^1 and $H(\mathrm{div})$ trace operators map W^p and RT^p onto the space of continuous polynomials $\mathcal{P}^p_c(\partial K)$ and the space of discontinuous polynomials $\mathcal{P}^{p-1}_d(\partial K)$,

$$\mathrm{tr} :\ u \in W^p \to u|_{\partial K} \in \mathcal{P}^p_c(\partial K),$$
$$\mathrm{tr} :\ u \in RT^p \to u|_{\partial K} \cdot n \in\ \mathcal{P}^{p-1}_d(\partial K).$$

Finally, by $H^1(\Omega_h)$ and $H(\mathrm{div}, \Omega_h)$, we mean the broken counterparts of the $H^1(\Omega)$ and $H(\mathrm{div}, \Omega)$ spaces that are discretized with the same elements of order $r > p$ but with no conformity requirements.

For our construction, we will make the following assumption on the FE mesh: a structured 2D mesh consisting of identical triangles where each element K can be obtained by scaling the master triangle $\hat{K}$ with element size h.

3.3.1 Construction

To motivate our construction, we first consider the Helmholtz equation with Dirichlet and Neumann boundary conditions:

$$\begin{cases} -\Delta u - \omega^2 u = f & \text{in } \Omega, \\ u = u_0 & \text{on } \Gamma_u, \\ \dfrac{\partial u}{\partial n} = t_0 & \text{on } \Gamma_t. \end{cases} \tag{3.3.1}$$

Here, Γ_u and Γ_t are two disjoint parts of the boundary (corresponding to the Dirichlet and Neumann boundaries respectively), $\omega > 0$ denotes the (angular) frequency and f, u_0, t_0 are given data. We assume $\overline{\Gamma_u} \cup \overline{\Gamma_t} = \partial\Omega$.

We proceed now with an elementary derivation of the primal DPG variational formulation. We multiply the Helmholtz equation with a test function v, integrate by parts over each element K, and sum over all elements to obtain:

$$\sum_K \{(\nabla u, \nabla v)_K - \omega^2 (u, v) + \langle \frac{\partial u}{\partial n_K}, v \rangle_{\partial K}\} = \sum_K (f, v)_K \,,$$

where $(\cdot,\cdot)_K$ and $\langle \cdot,\cdot \rangle_{\partial K}$ denote the $L^2(K)$ product, $H^{-\frac{1}{2}}(\partial K) \times H^{\frac{1}{2}}(\partial K)$ duality pairing on ∂K respectively, and n_K is the normal to K. In the standard

case, we assume that the test function v is globally conforming, $v \in H^1(K)$, which reduces $\sum_K \langle \frac{\partial u}{\partial n}, v \rangle_{\partial K}$ to $\langle \frac{\partial u}{\partial n}, v \rangle_{\partial \Omega}$ with n being the usual normal vector to $\partial\Omega$. We set $v = 0$ on Γ_u (i.e., do not test there) and replace $\frac{\partial u}{\partial n}$ with known boundary data t_0 on Γ_t, which is then moved to the right hand side.

In the DPG method, we test with discontinuous test functions $v \in H^1(\Omega_h)$, and make no additional assumptions about v on $\partial\Omega$. The normal derivative $\frac{\partial u}{\partial n}$ is identified as an extra unknown, the *flux* $\hat{t}$. More precisely, we take a field $t \in H(\mathrm{div}, \Omega)$, restrict it to element K, so that $t|_K \in H(\mathrm{div}, K)$, and consider its normal trace:

$$\langle \hat{t}, v \rangle_{\Gamma_h} := \sum_K \langle t|_{\partial K} \cdot n_K, v \rangle_{\partial K} .$$

The space of all such restrictions to the mesh skeleton $\Gamma_h = \bigcup_K \partial K$ is denoted by $H^{-\frac{1}{2}}(\Gamma_h)$ and is equipped with the quotient (minimum energy extension) norm:

$$\|\hat{t}\|_{H^{-\frac{1}{2}}(\Gamma_h)} := \inf_{t|_{\partial K} \cdot n_K = \hat{t}} \|t\|_{H(\mathrm{div},\Omega)} ,$$

where $t \in H(\mathrm{div}, \Omega)$. Notice that, by construction, the flux is single valued. The final variational formulation is now obtained as:

$$\begin{cases} u \in H^1(\Omega),\ u = u_0 \text{ on } \Gamma_u, \\ \hat{t} \in H^{-\frac{1}{2}}(\Gamma_h),\ \hat{t} = t_0 \text{ on } \Gamma_t, \\ (\nabla u, \nabla_h v) - \omega^2 (u, v) + \langle \hat{t}, v \rangle_{\Gamma_h} = (f, v) \text{ and } v \in H^1(\Omega_h). \end{cases} \tag{3.3.2}$$

The symbol h in $\nabla_h(\cdot)$ indicates that the gradient is computed elementwise.

Discretization

The two unknowns are discretized now with standard conforming elements:

$$u_h \in W^p \, , \hat{t}_h \in \mathrm{tr}_{\Gamma_h} RT^p.$$

If $b_K((u, \hat{t}), v)$ denotes the elementwise bilinear form corresponding to the variational formulation mentioned earlier, i.e.,

$$b_K((u, \hat{t}), v) = (\nabla u, \nabla_h v)_K - \omega^2 (u, v)_K + \langle \hat{t}, v \rangle_{\partial K} = (-\Delta u - \omega^2 u, v)_K + \langle \hat{t}, v \rangle_{\partial K},$$

then the ideal DPG method minimizes the residual

$$(u_h, \hat{t}_h) = \arg\min \Big(\sum_K \big(\sup_{v \in H^1(\Omega_h)} \frac{|b_K((u, \hat{t}), v)|}{\|v\|_{H^1(K)}} \big)^2 \Big)^{\frac{1}{2}},$$

with the $\arg\min$ taken over $u \in W^p \, , \hat{t} \in \mathrm{tr}_{\Gamma_h} RT^p$.

In the *practical* DPG method, the supremum of $v \in H^1(K)$ is replaced with a computable supremum of $v \in \mathcal{P}^r(K)$ with $r > p$.

Local nature of the construction

At this stage, we make a crucial remark. While dealing with a conforming mesh, we would need to deal with contributions from possibly the entire domain, whereas with broken test spaces, our computations are purely local. While the use of broken test spaces forces us to solve for an additional unknown (the term $\hat{t} = \frac{\partial u}{\partial n}$), we gain the ability to do purely local computations. In particular, this means that the use of broken test spaces allows us,

without loss of generality, to concentrate on the Fortin operator construction on a single element.

If we manage to construct an operator for a single element that satisfies elementwise version of conditions with an element independent constant C, the global conditions will easily follow. We will thus focus on constructing Π on a single triangular element K, obtained by a simple scaling of the master triangle $\hat{K}$.

Orthogonality relations

Given the elementwise bilinear form $b_K(\cdot,\cdot)$ over all elements K, one can obtain the bilinear form $b(\cdot,\cdot)$ on Ω:

$$b((u,\hat{t}),v) = \sum_K b_K((u,\hat{t}),v).$$

In constructing a Fortin operator, we are looking for a linear operator

$$\Pi : H^1(\Omega_h) \to \mathcal{P}^{p+\Delta p}(\Omega_h)$$

that satisfies two conditions:

$$\begin{aligned} &\text{(i)} \quad b((u,\hat{t}), v-\Pi v) = 0 \quad \forall u \in W_p,\ \hat{t} \in \mathrm{Tr}_{\Gamma_h} RT^p \\ &\text{(ii)} \qquad\qquad \|\Pi v\|_{H^1(\Omega_h)} \leq C\|v\|_{H^1(\Omega_h)} \end{aligned} \tag{3.3.3}$$

with a mesh independent constant C. Thus, in this case, $V_r = \mathcal{P}^{p+\Delta p}(\Omega_h) := \mathcal{P}^r(\Omega_h)$ and the index $r = p + \Delta p$.

Ideally, we would like the continuity constant C of Π to be independent of p. Now, while one may attempt to solve for Πv as stated, we will take a

slightly different approach by augmenting the Fortin definition conditions by a set of stronger conditions:

$$\begin{aligned} &\int_K (v - \Pi v)\, \phi = 0 \quad \forall \phi \in \mathcal{P}^p(K)\,, \\ &\int_e (v - \Pi v)\, \phi = 0 \quad \forall \phi \in \mathcal{P}^{p-1}(e),\ e = 1, 2, 3\,. \end{aligned} \tag{3.3.4}$$

Clearly, the stronger conditions are sufficient to conclude the orthogonality which we sought originally, and while the stronger conditions may result in a more pessimistic estimate for C, they lead to a computationally tractable problem. Notice that additional conditions are to a large extent *problem independent*. Indeed, the integration by parts argument that have led to them will work for a general class of second order operators as long as the material data are constant elementwise.

Ensuring uniqueness

Although we have identified a possible candidate for Πv, due to the underdetermined nature of the constraints, we may have more than one element in $\mathcal{P}^r$ that satisfies our requirements. In order to complete the definition of Π we request that in addition to the orthogonality conditions, the norm $\|\Pi v\|_{H^1(K)}$ to be minimal. Thus, we can view our construction of Π as a constrained minimization problem:

$$\begin{cases} \Pi v = \operatorname{argmin} \|v^*\|_{H^1(K)} & v^* \in \mathcal{P}^r,\ v \in V \\ (\phi, v^*)_{L^2(K)} = (\phi, v)_{L^2(K)}, & \phi \in \mathcal{P}^p(K) \\ \langle \phi_e, v^* \rangle_{L^2(\partial K)} = \langle \phi_e, v \rangle_{L^2(\partial K)}. & \phi_e \in \mathcal{P}_d^{p-1}(e),\ e = 1, 2, 3 \end{cases} \tag{3.3.5}$$

We can re-write this constrained minimization problem as a mixed (saddle-point) problem as follows:

$$\begin{cases} (v^*, \delta v^*)_{H^1(K)} + (\phi, \delta v^*)_{L^2(K)} + \langle \phi_e, \delta v^* \rangle_{L^2(\partial K)} = 0 \quad \delta v^* \in \mathcal{P}^r, \\ (\delta\phi, v^*)_{L^2(K)} = (\delta\phi, v)_{L^2(K)}, \ \delta\phi \in \mathcal{P}^p(K) \\ \langle \delta\phi_e, v^* \rangle_{L^2(\partial K)} = \langle \delta\phi_e, v \rangle_{L^2(\partial K)}, \end{cases}$$

where $\delta\phi_e \in \mathcal{P}^{p-1}(e)$, $e = 1, 2, 3$. Thus, our construction of the Fortin operator is the same as solving the above mixed problem.

Use of alternate H^1 norm

Given any $v \in H^1(\Omega_h)$, we can always split it into its average and the zero average parts,

$$v = \bar{v} + (v - \bar{v}), \quad \Pi v = \Pi\bar{v} + \Pi(v - \bar{v})\,. \tag{3.3.6}$$

We will denote by $H^1_{avg}(\Omega_h)$ the set of $H^1(\Omega_h)$ functions with zero average. In this context, we remark that the usual H^1 norm is equivalent to the following norm:

$$\|u\|^2_{H^1(\Omega_h)} = \sum_K \left(\|\nabla u\|^2_{L^2(K)} + \|\bar{u}\|^2_{L^2(K)} \right) \tag{3.3.7}$$

where

$$\bar{u} := \frac{1}{|K|} \int_K u \, dK$$

is the average value of function u in element K. This fact is proved in the appendix. The use of this norm in place of the standard H^1 norm will be

critical as we continue our analysis. Henceforth by H^1 norm we shall mean the above mentioned norm, as opposed to the usual H^1 norm.

This concludes the H^1 construction for the h-version of the DPG Fortin operator construction. Of course, we do not know at this point how large the continuity constant is and how it depends upon $r = p + \Delta p$. This is taken up next.

Numerical procedure

Recall that the Fortin operator orthogonality conditions resulted in an underdetermined system of equations for the image $w = \Pi v$. Therefore, in addition to the orthogonality conditions, we imposed an additional condition on the image $w = \Pi v$, namely, that it was of minimal norm in the appropriate space, in order to guarantee a unique solution to the constraint equations. We now derive a variational formulation for obtaining the optimal $w = \Pi v$ and thereby estimate the inf-sup constant γ_h. We use the definitions in [67]:

$$\begin{cases} w \in \mathcal{P}^r(K) \\ \widetilde{b}(w, (\phi, \phi_e)) = \widetilde{b}(v, (\phi, \phi_e)) \quad \phi \in \mathcal{P}^p(K),\ \phi_e \in \mathcal{P}^p(e) \end{cases}$$

where

$$\widetilde{b}(w, (\phi, \phi_e)) = \int_K w\phi + \sum_{e=1}^{3} \int_e w\phi_e \, .$$

This new bilinear form $\widetilde{b}(\cdot, \cdot)$ is an auxiliary bilinear form obtained from the orthogonality conditions imposed by the definition of Π. Since the auxiliary bilinear form continuity constant M, i.e.,

$$|\widetilde{b}(v, u)| \leq M \|v\|_V \|u\|_U \tag{3.3.8}$$

is $O(1)$, and the Fortin continuity constant $C = \frac{M}{\gamma_h}$, we have that overall $C = O(\frac{1}{\gamma_h})$.

Upon the use of the alternate H^1 norm based on the zero-average splitting of the H^1 space, minimization of the full H^1 norm is equivalent to the minimization of the H^1 semi-norm. Consider now the cost functional

$$J(w) = \frac{1}{2}\|\nabla w\|^2_{L^2} - \widetilde{b}(w, (\phi, \phi_e)). \tag{3.3.9}$$

As before, setting the derivative of the perturbed cost, $\frac{dJ(w+\epsilon\delta w)}{d\epsilon}|_{\epsilon=0} = 0$, we arrive at the variational formulation

$$(\nabla w, \nabla \delta w)_{L^2} = \widetilde{b}(\delta w, (\phi, \phi_e)) \text{ i.e.,} \tag{3.3.10}$$

$$\begin{cases} w \in \mathcal{P}^r(K) \\ \displaystyle\int_K \nabla w \nabla \delta w = \int_K \phi \delta w + \sum_e \int_e \phi_e \delta w, \quad \delta w \in \mathcal{P}^{p+\Delta p}(K)\,. \end{cases}$$

Introducing operator $T : (\phi, \phi_e) \to w$, we have,

$$\gamma = \inf_{(\phi,\phi_e)} \frac{\|T(\phi, \phi_e)\|_{H^1(K)}}{\|(\phi, \phi_e)\|}\,.$$

Thus γ is the smallest eigenvalue corresponding to the generalized eigenvalue problem:

$$(T(\phi, \phi_e), T(\delta\phi, \delta\phi_e)) = \lambda((\phi, \phi_e), (\delta\phi, \delta\phi_e)) \quad \forall \delta\phi, \delta\phi_e\,.$$

Steps involved in the computation The computation of γ involves the following generic steps:

- Select a basis e_i for w, and a basis g_j for (ϕ, ϕ_e) or (Ψ, η, ϕ) as the case may be

$$w = \sum_j w_i e_i, \quad (\phi, \phi_e) \text{ or } (\Psi, \eta, \phi) = \sum_j \phi_j g_j \,,$$

- Compute matrix representation of operator T in those bases,

$$w_i = T_{ij} g_j \,,$$

- Compute the Gram matrices corresponding to the two inner products,

$$H_{kl} := (e_k, e_l)_{H^1(K)}, \quad M_{kl} := (g_k, g_l) \,,$$

- Solve the generalized eigenvalue problem,

$$(T^T H T)\phi = \lambda M \phi \,.$$

As we saw earlier, the γ we are interested in can be seen as the (square root of) the smallest eigenvalue of the generalized eigenvalue problem in the above steps.

3.3.2 Approximate Fortin operators

We now take a different approach to the problem of computing the continuity constant of the Fortin operator. Recall that our difficulty is in exactly computing the constant due to the infinite dimensional nature of norm optimization. In this section, we construct a sequence of *approximate* Fortin operators, each member of which is defined on a large, yet finite dimensional

subspace $V^{r+\Delta r}$ containing the enriched test space V^r. In other words, we construct a sequence $\Pi_{\Delta r} : V^{r+\Delta r} \to V^r$ indexed by Δr in place of the *exact* Fortin operator we constructed earlier, $\Pi : V \to V^r$. However, since each approximate Fortin operator $\|\Pi_{\Delta r}\|$ is defined on a finite dimensional subspace $V^{r+\Delta r}$ of V, we can *exactly* compute the continuity constant of $\|\Pi_{\Delta r}\|$.

With increasing Δr, it is natural to expect $\|\Pi_{\Delta r}\| \to \|\Pi\|$. Indeed, as we have seen, computing Fortin constants leads to generalized eigenvalue problems. As shown in [51], asymptotic estimates of eigenvalues in the context of Galerkin approximations is a non-trivial problem. In our case, investigating the convergence of the appropriate Fortin operators to the exact Fortin operator is an even more challenging problem to study as a function of Δr. Yet, the sequence of approximate Fortin constants in non-decreasing, so we may view the approximate Fortin constants as being a lower bound on the exact Fortin constant. Finally, one expects a tighter lower estimate via the approximate Fortin constants, i.e., the approximate Fortin constants are expected to be "closer" to the exact Fortin constant than the upper bound estimate we derived earlier.

We do this analysis for two reasons. First, we will be able to get a realistic view of how sharp our previous estimates of the *exact* Fortin constants were. Since we were working with sufficient conditions, we know a priori that our previous estimates correspond to the worst case scenario. However, it may be the case that in practice, we always do better than the estimates' promise. The exact computation of $\|\Pi_{\Delta r}\|$ will shed light on how close we actually

are to the estimate. In other words, the approximate Fortin constants will give a heuristic lower bound of the exact Fortin constant. Second, in view of a numerical realization of the Fortin operator, one is naturally led to constructing a discrete approximation to the continuous Fortin operator for use in solving practical problems.

Construction of approximate Fortin operators

The construction of the approximate H^1 Fortin operators Π_Δ follows the exact case, with the only difference being that the domain of $\Pi_{\Delta r}$ is the finite dimensional space $V^{r+\Delta r}$ containing the enriched test space V^r instead of the infinite dimensional V. Note that the approximate Fortin operators $\Pi_{\Delta r}$ depend on the trial function approximation order p, the enrichment order Δp as well as the new (additional) Fortin enrichment order Δr. We can easily define the approximate Fortin operators using the constrained optimization definition of the exact Fortin operator we gave earlier.

Recall the exact H^1 Fortin operator is defined via the solution of the following constrained optimization problem:

$$\begin{cases} \Pi v = \text{argmin}\|v^*\|_{H^1(K)} & v^* \in \mathcal{P}^r,\ v \in V \\ (\phi, v^*)_{L^2(K)} = (\phi, v)_{L^2(K)}, & \phi \in \mathcal{P}^p(K) \\ \langle \phi_e, v^* \rangle_{L^2(\partial K)} = \langle \phi_e, v \rangle_{L^2(\partial K)}. & \phi_e \in \mathcal{P}^{p-1}(e),\ e = 1,2,3 \end{cases} \tag{3.3.11}$$

To define the approximate Fortin operator, we simply restrict our space of constraints to lie in the finite dimensional subspace $V^{r+\Delta r}$ instead of the

full space V:

$$\begin{cases} \Pi_{\Delta r} v = \text{argmin} \|v^*\|_{H^1(K)} & v^* \in \mathcal{P}^r,\ v \in \mathcal{P}^{r+\Delta r} \\ (\phi, v^*)_{L^2(K)} = (\phi, v)_{L^2(K)}, & \phi \in \mathcal{P}^p(K) \\ \langle \phi_e, v^* \rangle_{L^2(\partial K)} = \langle \phi_e, v \rangle_{L^2(\partial K)}. & \phi_e \in \mathcal{P}^{p-1}(e),\ e = 1,2,3 \end{cases} \tag{3.3.12}$$

We can then derive a mixed (saddle-point) formulation analagous to the exact case. One can then form a matrix representation of $\Pi_{\Delta r}$ and compute the *exact* value of the continuity constant of $\Pi_{\Delta r}$ as:

$$\begin{cases} \|\Pi_{\Delta r}\| = \max \dfrac{\|\Pi v\|_{H^1}}{\|v\|_{H^1}}, & v \in V^{r+\Delta r} \\ \text{or, } (\Pi v, \Pi \delta v)_{H^1} = \lambda^2 (v, \delta v)_{H^1}, & \delta v \in V^{r+\Delta r} \\ \text{or, } (P^* G P) v = \lambda^2 G v, \end{cases} \tag{3.3.13}$$

where P is the matrix representation of $\Pi_{\Delta r}$ and G is the Gram matrix corresponding to the alternate (zero-average split) H^1 inner product:

$$G = (\nabla v_i, \nabla v_j)_{L^2} \tag{3.3.14}$$

and $\text{span}\{v_i\} = V^{r+\Delta r}$. The maximum eigenvalue λ_m of the above generalized eigenvalue problem is the required norm of $\Pi_{\Delta r}$.

3.4 Construction of $H(\text{div}, \Omega)$ DPG Fortin operator

We now consider the construction of a DPG Fortin operator for the $H(\text{div}, \Omega)$ case. Our analysis will mirror that of the H^1 case, as will be seen

shortly.

3.4.1 Construction

Recall that the Helmholtz problem is obtained by eliminating velocity u from the linear acoustics equations:

$$\begin{cases} i\omega p + \text{div } u = 0, & \text{in } \Omega \\ i\omega u + \nabla p = 0, & \text{in } \Omega \\ p = p_0, \text{ on } \Gamma_p \\ u \cdot n = u_0, \text{ on } \Gamma_u \end{cases} \tag{3.4.1}$$

The equations are obtained by linearizing the isentropic Euler equations around a hydrostatic solution $u = 0$, $p = p_0 =$ constant. The first equation represents conservation of mass, and the second one conservation of linear momentum. The equations have been non-dimensionalized to obtain a unit sound speed.

The so-called *ultraweak variational formulation* for the system is obtained by multiplying the first equation with a test function q, the second equation with a test function v, integrating by parts over an element K, and then summing up the element contributions. Similar to the primal method, the boundary terms of u and p are identified as new unknowns. The final formulation is:

$$\begin{cases} p \in L^2(\Omega), u \in (L^2(\Omega))^2 \\ \hat{p} \in H^{\frac{1}{2}}(\Gamma_h),\ \hat{p} = p_0 \ \text{ on } \Gamma_p \\ \hat{t} \in H^{-\frac{1}{2}}(\Gamma_h),\ \hat{t} = u_0 \ \text{ on } \Gamma_u \\ i\omega(p,q) - (u, \nabla_h q) + \langle \hat{t}, q \rangle_{\Gamma_h} = 0, \quad q \in H^1(\Omega_h) \\ i\omega(u,v) - (p, \operatorname{div}_h v) + \langle \hat{p}, v \cdot n \rangle_{\Gamma_h} = 0, \quad v \in H(\operatorname{div}, \Omega_h), \end{cases} \tag{3.4.2}$$

where $\hat{p}$ is a trace of a global $p \in H^1(\Omega)$ to the mesh skeleton Γ_h, and

$$\langle \hat{p}, q \rangle_{\Gamma_h} = \sum_K \langle p, q \rangle_{\partial K}.$$

As in the case of $H^{-\frac{1}{2}}(\Gamma_h)$, $\hat{p}$ is measured in the quotient (minimum energy extention) norm:

$$\|\hat{p}\|_{H^{\frac{1}{2}}(\Gamma_h)} := \inf_{p|_{\partial K} = \hat{p}} \|p\|_{H^1(\Omega)},$$

where $p \in H^1(\Omega)$.

Discretization

Consistent with the exact sequence structure, the L^2 unknowns p, u are discretized with discontinuous polynomials of order p, traces $\hat{p}$ are discretized with the traces of W^p functions, i.e., continous polynomials of order p on the mesh skeleton Γ_h, and traces $\hat{t}$ are approximated with traces of RT^p on Γ_h, i.e. discontinuous polynomials of order $p-1$.

The practical DPG method is based on minimizing the residuals in the norms dual to the $H^1(\Omega_h)$ and $H(\operatorname{div}, \Omega_h)$ approximated by taking the

supremum with respect to discontinuous test spaces, i.e., $W^r(\Omega_h), RT^r(\Omega_h)$ with $r > p$.

As we did in the H^1 case, consider the element bilinear form $b_K(\cdot,\cdot)$:

$$b_K((p,u,\hat{p},\hat{t}),(q,v)) := (i\omega p + \text{div } u, q)_K + \langle \hat{t} - u\cdot n, q\rangle_{\partial K} + (i\omega u + \nabla p, v)_K + \langle \hat{p} - p, v\cdot n\rangle_{\partial K},$$

and the practical DPG method is:

$$(p_h,u_h,\hat{p}_h,\hat{t}_h) = \arg\min\Big(\sum_K \Big(\sup_{q\in\mathcal{P}^r, v\in RT^r} \frac{|b_K((p,u,\hat{p},\hat{t}),(q,v))|}{(\|q\|^2_{H^1(K)} + \|v\|^2_{H(\text{div},K)})^{\frac{1}{2}}}\Big)^2\Big)^{\frac{1}{2}},$$

It is well known that in order to obtain correct scaling arguments for the Fortin operator Π, the operator must satisfy the commuting exact sequence property:

$$\begin{array}{ccc} H(\text{div},K) & \xrightarrow{\text{div}} & L^2(K) \\ \downarrow \Pi & & \downarrow P \\ RT^r(K) & \xrightarrow{\text{div}} & \mathcal{P}^{r-1} \end{array}$$

i.e.,

$$\text{div } (\Pi v) = P(\text{div } v)$$

Here, in principle, $P(\cdot)$ is any well-defined continous operator but, in order to minimize the coninuity constant for $P(\cdot)$, it is natural to assume that $P(\cdot)$ is the L^2 projection operator.

Orthogonality relations

We now construct the operator on the master element $\hat{K}$ and subsequently develop scaling arguments for the generic element K. At the outset, it is clear we will enforce the orthogonality constraints on the element interior and boundary separately. As was the case with the H^1 construction, we will enforce stronger conditions which are sufficient for the orthogonality we seek. We see that in order for the orthogonality to hold,

$$\begin{aligned} \int_K \Psi \cdot (v - \Pi v) \, &= 0 \qquad \forall \Psi \in RT^p(K)\,, \\ \int_{\partial K} \phi \, [(v - \Pi v) \cdot n] \, &= 0 \quad \forall \phi \in \mathcal{P}_c^p(\partial K). \end{aligned} \tag{3.4.3}$$

are sufficient. However, for ease of analysis, we enforce the stronger conditions in our definition. Moreover, choosing $\Psi = \nabla \eta$ in the first condition yields

$$\begin{aligned} &\textstyle\int_K \Psi \cdot (v - \Pi v) \, = 0 \\ &= \textstyle\int_K \nabla \eta \cdot (v - \Pi v) \\ &= - \textstyle\int_K \eta \, \mathrm{div}(v - \Pi v) + \int_{\partial K} \eta [(v - \Pi v) \cdot n] = 0, \end{aligned} \tag{3.4.4}$$

which, using the fact that

$$\int_{\partial K} \phi \, [(v - \Pi v) \cdot n] \, = 0 \, \forall \, \phi \in \mathcal{P}_c^p(\partial K),$$

implies the final condition

$$\int_K \eta \, \mathrm{div} \, (v - \Pi v) = 0 \, \forall \, \eta \in \mathcal{P}^p(K).$$

However, we would like to ensure that the divergence of Πv yields the L^2 projection, i.e., we would like to have div $\Pi v = P(\mathrm{div}\ v)$ where $P(\cdot)$ is the L^2

projection. Having this condition would greatly simplify our estimates on the total $H(\text{div})$ norm of Πv. Towards this end, we note that we are essentially asking for orthogonality up to $\mathcal{P}^{r-1}(K)$, which, combined with the fact that we already have orthogonality up to $\mathcal{P}^p(K)$, means we need to enforce the last condition for η coming from the quotient space $\mathcal{P}^{r-1}(K)/\mathcal{P}^p(K)$. We thus see that our orthogonality requirement has given us three constraints:

$$\begin{aligned} &\textstyle\int_K \Psi \cdot (v - \Pi v) \, = 0 \qquad && \forall \Psi \in RT^p(K)\,, \\ &\textstyle\int_K \eta \,\text{div}(v - \Pi v) = 0 \qquad && \forall \eta \in \mathcal{P}^{r-1}(K)/\mathcal{P}^p(K)\,, \\ &\textstyle\int_{\partial K} \phi \,[(v - \Pi v) \cdot n] \, = 0 \quad && \forall \phi \in \mathcal{P}^p_c(\partial K). \end{aligned} \tag{3.4.5}$$

Norm minimization and mixed formulation.

As in the H^1 case, we have identified a possible candidate for Πv, but due to the underdetermined nature of the constraints, we need to identify a unique $\Pi v \in RT^r(K)$ that satisfies our requirements. We therefore request that in addition to the orthogonality conditions, the norm $\|\Pi v\|_{H(\text{div})}$ be minimal:

$$\begin{cases} \Pi v = \text{argmin} \|v^*\|_{H(\text{div})} & v^* \in RT^r,\, v \in V \\ (\Psi, v^*)_{L^2(K)} = (\Psi, v)_{L^2(K)}, & \Psi \in RT^p(K) \\ (\eta, \text{div } v^*)_{L^2(K)} = (\eta, \text{div } v)_{L^2(K)}, & \eta \in \mathcal{P}^{r-1}(K)/\mathcal{P}^p(K) \\ \langle \phi, v^* \rangle_{L^2(\partial K)} = \langle \phi, v \rangle_{L^2(\partial K)}. & \phi \in \mathcal{P}^p_c(\partial K). \end{cases} \tag{3.4.6}$$

We can re-write this constrained minimization problem as a mixed (saddle-point) problem as follows:

$$\begin{cases} (v^*, \delta v^*)_{H(\text{div})} + (\Psi, \delta v^*)_{L^2(K)} + \\ (\eta, \text{div } \delta v^*)_{L^2(K)} + \langle \phi, \delta v^* \rangle_{L^2(\partial K)} = 0 \quad \delta v^* \in \mathcal{P}^r, \\ (\delta\Psi, v^*)_{L^2(K)} = (\delta\Psi, v)_{L^2(K)}, \quad \delta\Psi \in RT^p(K) \\ (\delta\eta, \text{div } v^*)_{L^2(K)} = (\delta\eta, \text{div } v)_{L^2(K)}, \quad \delta\eta \in \mathcal{P}^{r-1}(K)/\mathcal{P}^p(K) \\ \langle \delta\phi, v^* \rangle_{L^2(\partial K)} = (\delta\phi, v)_{L^2(\partial K)}, \quad \delta\phi \in \mathcal{P}_c^p(\partial K). \end{cases} \tag{3.4.7}$$

The solution to the Fortin conditions coincides with the solution of the above mixed problem.

Numerical procedure

Recall that we required Πv to satisfy the orthogonality conditions, as well as having least $H(\text{div})$ norm, so that we are interested in

$$\Pi v = \arg \min_{w \in RT^r} \|w\|_{H(\text{div})}. \tag{3.4.8}$$

As in the H^1 case, we have the auxiliary bilinear form:

$$\widetilde{b}(w, (\Psi, \eta, \phi)) = \int_K \Psi \cdot w + \int_K \eta \, \text{div } w + \int_{\partial K} \phi \, [w \cdot n] \tag{3.4.9}$$

with $\Psi \in RT^p(K), \eta \in \mathcal{P}^{r-1}(K)/\mathcal{P}^p(K), \phi \in \mathcal{P}_c^p(\partial K)$ (see [67] for more details). Consider now the cost functional

$$J(w) = \frac{1}{2}\|w\|^2_{H(\text{div})} - \widetilde{b}(w, (\Psi, \eta, \phi)). \tag{3.4.10}$$

Setting the derivative of the perturbed cost, $\frac{dJ(w+\epsilon\delta w)}{d\epsilon}|_{\epsilon=0} = 0$, we arrive at the variational formulation

$$(w, \delta w)_{H(\text{div})} = \widetilde{b}(\delta w, (\Psi, \eta, \phi)) \text{ i.e.,} \tag{3.4.11}$$

$$\begin{cases} w \in RT^r(K) \\ \displaystyle\int_K (w, \delta w)_{L^2} = \int_K \Psi \cdot \delta w + \int_K \eta \mathrm{div} \delta w + \int_{\partial K} \phi \delta w \,. \end{cases}$$

As in the H^1 case, we introduce the operator $T : (\Psi, \eta, \phi) \to w$ and we have,

$$\gamma = \inf_{(\Psi,\eta,\phi)} \frac{\|T(\Psi, \eta, \phi)\|_{H(\mathrm{div},K)}}{\|(\Psi, \eta, \phi)\|} \,.$$

and γ is the smallest eigenvalue corresponding to the generalized eigenvalue problem:

$$(T(\Psi, \eta, \phi), T(\delta\Psi, \delta\eta, \delta\phi)) = \lambda((\Psi, \eta, \phi), (\delta\Psi, \delta\eta, \delta\phi)) \quad \forall \delta\Psi, \delta\eta, \delta\phi \,.$$

The steps involved in the numerical computation of γ are the same as in the H^1 case.

Approximate $H(\mathbf{div})$ Fortin operators

We come now to the approximate $H(\mathrm{div})$ Fortin operators. The definition of the approximate $H(\mathrm{div})$ Fortin operators follows the H^1 case:

$$\begin{cases} \Pi_{\Delta r} v = \mathrm{argmin} \|v^*\|_{H(\mathrm{div})} & v^* \in RT^r,\ v \in RT^{r+\Delta r} \\ (\Psi, v^*)_{L^2(K)} = (\Psi, v)_{L^2(K)}, & \Psi \in RT^p(K) \\ (\eta, \mathrm{div}\ v^*)_{L^2(K)} = (\eta, \mathrm{div}\ v)_{L^2(K)}, & \eta \in \mathcal{P}^{r-1}(K)/\mathcal{P}^p(K) \\ \langle \phi, v^* \rangle_{L^2(\partial K)} = \langle \phi, v \rangle_{L^2(\partial K)}. & \phi \in \mathcal{P}^p_c(\partial K), \end{cases} \tag{3.4.12}$$

and the equivalent saddle point problem reads:

$$\begin{cases} (v^*, \delta v^*)_{H(\text{div})} + (\Psi, \delta v^*)_{L^2(K)} + \\ (\eta, \text{div } \delta v^*)_{L^2(K)} + \langle \phi, \delta v^* \rangle_{L^2(\partial K)} = 0 \quad \delta v^* \in \mathcal{P}^r, \\ (\delta\Psi, v^*)_{L^2(K)} = (\delta\Psi, v)_{L^2(K)}, \quad \delta\Psi \in RT^p(K) \\ (\delta\eta, \text{div } v^*)_{L^2(K)} = (\delta\eta, \text{div } v)_{L^2(K)}, \quad \delta\eta \in \mathcal{P}^{r-1}(K)/\mathcal{P}^p(K) \\ \langle \delta\phi, v^* \rangle_{L^2(\partial K)} = (\delta\phi, v)_{L^2(\partial K)}, \quad \delta\phi \in \mathcal{P}_c^p(\partial K). \end{cases} \tag{3.4.13}$$

Again, we can compute $\|\Pi_{\Delta r}\|$ as:

$$\begin{cases} \|\Pi_{\Delta r}\| = \max \dfrac{\|\Pi v\|_{H(\text{div})}}{\|v\|_{H(\text{div})}}, & v \in RT^{r+\Delta r} \\ \text{or, } (\Pi v, \Pi \delta v)_{H(\text{div})} = \lambda^2 (v, \delta v)_{H(\text{div})}, & \delta v \in RT^{r+\Delta r} \\ \text{or, } (P^* G P) v = \lambda^2 G v, \end{cases} \tag{3.4.14}$$

where P is the matrix representation of $\Pi_{\Delta r}$ and G is the Gram matrix corresponding to the $H(\text{div})$ inner product:

$$G = (v_i, v_j)_{H(\text{div})} \tag{3.4.15}$$

and $\text{span}\{v_i\} = RT^{r+\Delta r}$. The maximum eigenvalue λ_m of the above generalized eigenvalue problem is the required norm of $\Pi_{\Delta r}$.

3.5 Numerical Results

We now discuss our computational results in detail, starting with the results of the upper bound estimates for the exact Fortin operator. Our computations use the shape functions described in [40].

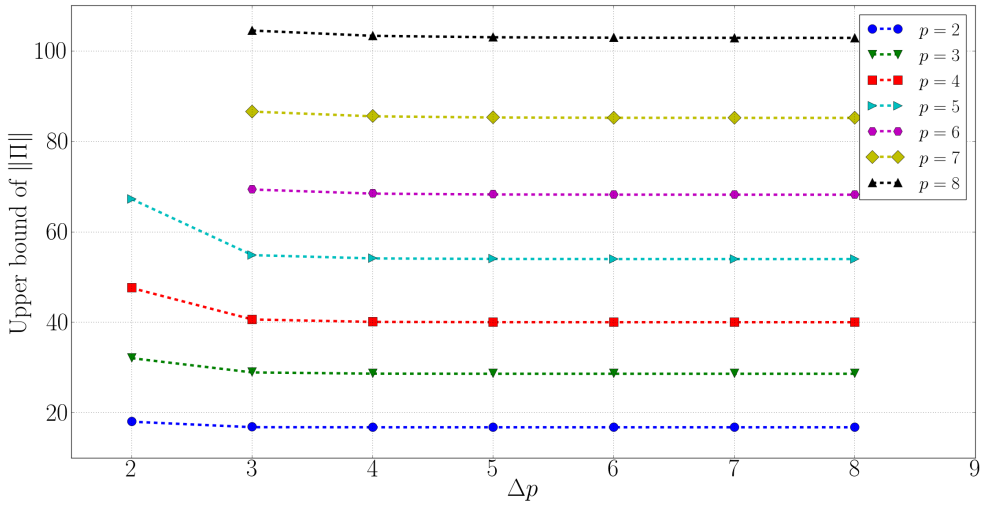

Figure 3.1: Upper bound of the H^1 DPG Fortin constant as a function of Δp for various p

H^1 Fortin operator upper bound Figure 3.1 shows the upper bound of the H^1 DPG Fortin constant as a function of p and Δp. We observe a few interesting results. First, with increasing p, we find that the upper bound of the Fortin constant also increases. This directly translates to the fact that with higher p, we loose stability, which of course is to be expected since increasing p makes resolving the optimal test functions more difficult. Second, we find that increasing Δp only marginally increases stability, that too, only for low p. This means that we do not gain significant stability by an indiscriminate increase in Δp. Also, for fixed p, the Fortin constant is a decreasing function of Δp. This is expected, since we search for the minimum eigenvalue of the generalized eigenvalue problem over a larger space with increasing Δp. Finally, we see that we need at least $\Delta p = 3$ for $p \geq 6$, for any reasonable stability, as

was indicated by the theory.

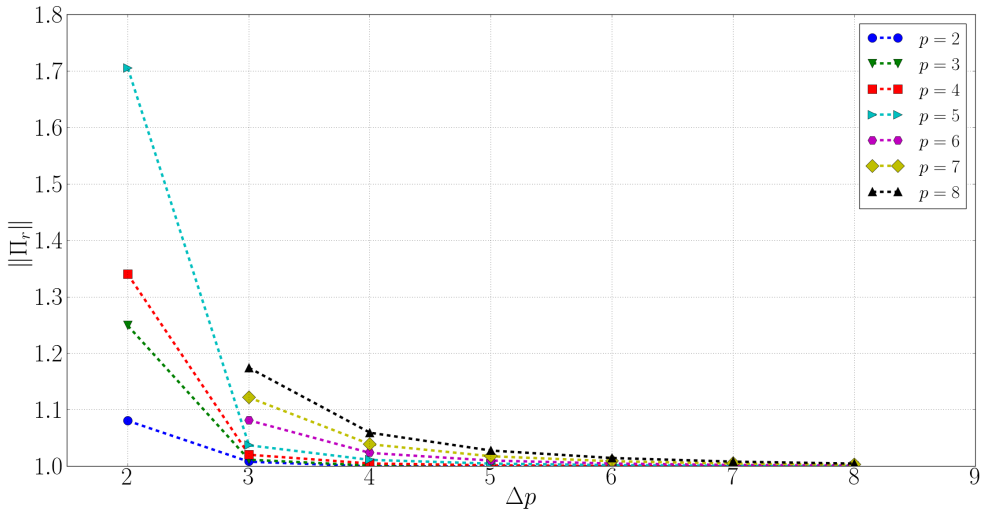

Figure 3.2: Exact Fortin constant of approximate H^1 DPG Fortin operators. Plots show values of $\|\Pi_{\Delta r}\|$ after convergence with sufficiently large Δr as a function of Δp for various p

Approximate H^1 Fortin constant Figure 3.2 shows the plots of the exact the Fortin constant of the *approximate* H^1 DPG Fortin operators. The plots display values of $\|\Pi_{\Delta r}\|$ after convergence with sufficiently large Δr as a function of Δp for various p. As we see from the plots, the Fortin constant is $O(1)$. Moreover, we get rapid convergence (for sufficiently large Δr) with increasing Δp. Again, a detailed analysis with respect to increasing Δr is beyond the scope of this chapter. Finally, as expected, the convergence is from above, i.e., the values decay with increasing Δp. These results are very optimistic, and indicate that we have, in practice, no recognizable loss of stability in the H^1 case with the use of practical test functions.

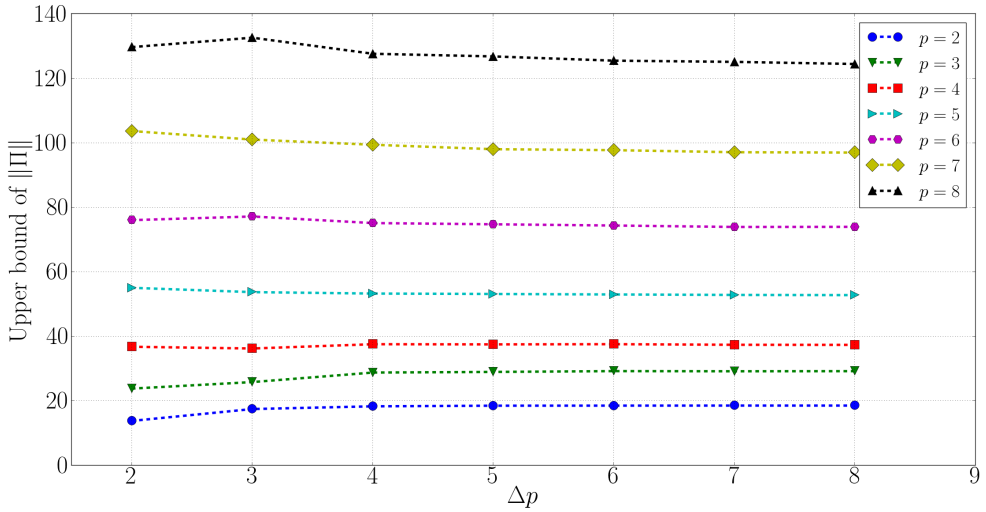

Figure 3.3: Upper bound of the $H(\text{div})$ DPG Fortin constant as a function of Δp for various p

$H(\text{div})$ Fortin operator upper bound Figure 3.3 shows the upper bound of the $H(\text{div})$ DPG Fortin constant as a function ofΔp for various p. Here, some of the results are, at first glance, a bit anomalous. At the very outset, we see that the actual values of the Fortin constant are much larger than in the H^1 case. This is due to the fact that we impose a greater number of constraints in the $H(\text{div})$ case than the H^1 case. As in the H^1 case, with increasing p, we find that the upper bound of the Fortin constant also increases. Also, in the $H(\text{div})$ case as well, increasing Δp only marginally increases stability, mainly for low p. However, for fixed p, we do not find monotonic decrease of the Fortin constant with increasing Δp, which may seem odd. However, this is clarified upon closer inspection of the constraints we have imposed. In order to obtain an L^2 projection on the divergence part of the $H(\text{div})$ norm, we

imposed a Δp dependent constraint, which means that increasing Δp changes the set of constraints, and therefore, the image of the Fortin operator. It is thus not reasonable to expect a monotonic decrease of the Fortin constant with increasing Δp in the $H(\mathrm{div})$ case. This is unavoidable with our construction, as we require the L^2 projection constraint on the divergence part of the Fortin operator in order to ensure that the scaling arguments can be used to reduce the construction to the master element.

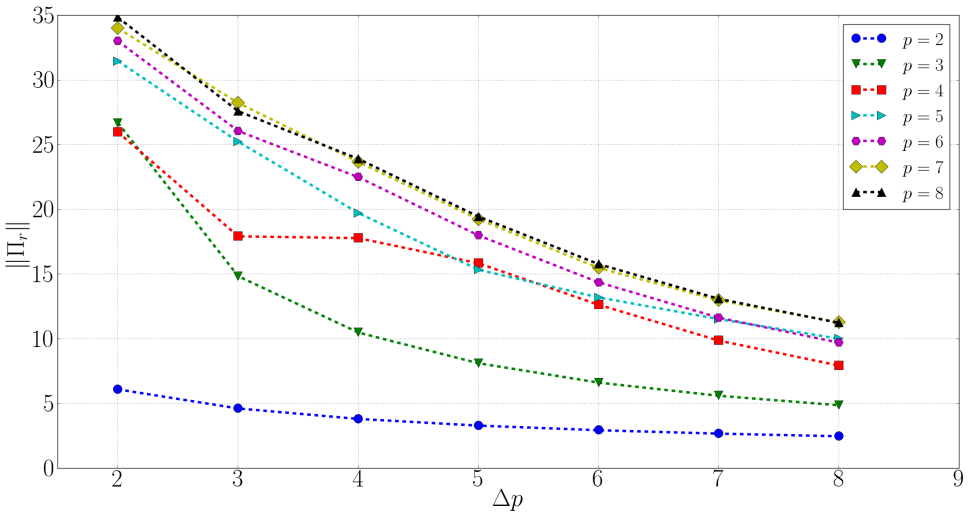

Figure 3.4: Exact Fortin constant of approximate $H(\mathrm{div})$ DPG Fortin operators. Plots show values of $\|\Pi_{\Delta r}\|$ after convergence with sufficiently large Δr as a function of Δp for various p

Approximate $H(\mathbf{div})$ Fortin constant Figure 3.4 shows the plots of the exact the Fortin constant of the *approximate* $H(\mathrm{div})$ DPG Fortin operators. Again, for ease of comparison with the earlier plots, we display values of $\|\Pi_{\Delta r}\|$ (after convergence with sufficiently large Δr) as a function of Δp for various

p. In this case, the Fortin constants are larger than the H^1 values, being an order of magnitude larger. However, we do still see a decay with increasing Δp. The decay is non-monotonic due to the same reasons we observed in the $H(\text{div})$ estimates earlier: we imposed a Δp dependent constraint, which means that increasing Δp changes the set of constraints, and therefore, the image of the Fortin operator. Finally, although the values are larger than the H^1 case, we still have a more optimistic result than the $H(\text{div})$ estimates, which were almost two orders of magnitude larger.

Chapter 4

The Linear Schrödinger Equation

In this chapter[1], we shall deal with variational formulations of the second order time dependent Schrödinger equation.

Author contributions: The contents of this chapter are taken largely from the published multi-author article [27], Copyright (c)2017 Society for Industrial and Applied Mathematics. Reprinted with permission. All rights reserved. The author of this dissertation is a co-author of the work [27] and contributed to the development of the theory and the numerical results (including coding the discretization described herein) presented in [27].

Recall that the nonlinear Schrödinger equation (NLS) appears in time-dependent models of pulse propagation in optical fibers. The derivation of the NLS in this context is well-known (see for instance [2, 75]). Indeed, the with certain simplifying assumptions, the full vector 3 dimensional Maxwell equations are reduced to a nonlinear Schrödinger (NLS) type equation (or NLSE) in the variable A (a complexified amplitude [2, 75]):

[1]The content of this chapter is taken from [27], Copyright (c)2017 Society for Industrial and Applied Mathematics. Reprinted with permission. All rights reserved.

$$i\frac{\partial A}{\partial x} - \frac{\beta}{2}\frac{\partial^2 A}{\partial t^2} + \gamma|A|^2 A = 0, \tag{4.0.1}$$

where the fiber length is along the x direction and t is an observation window (time) and β is a material constant. We refer the reader to [2, 75] for more details and derivations of the NLS from Maxwell equations. Note that in contrast with the standard NLSE, the NLS type equation (4) has time and space swapped: the equation is second order in time and first order in space. Nevertheless, the equation is known to be an accurate model in fiber optic communication. We refer the reader to [27] for complete details, derivations and analysis of the discussion in this chapter.

4.1 Introduction

Motivated by optics applications, we will, for the remainder of the section, study the standard linear Schrödinger equation (LSE) in multiple space dimensions from the DPG perspective. We will therefore consider a bounded Lipschitz domain $\Omega_0 \subset \mathbb{R}^n$ and time $t \in [0, T]$ with $T < \infty$. We set our spacetime domain $\Omega = \Omega_0 \times [0, T]$. Unless otherwise mentioned, we reserve the symbol i for $\sqrt{-1}$.

$$i\frac{\partial u}{\partial t} - \Delta u = f, \qquad x \in \Omega_0,\ 0 < t < T, \tag{4.1.1a}$$

$$u(x, t) = 0, \qquad x \in \partial\Omega_0,\ 0 < t < T, \tag{4.1.1b}$$

$$u(x, 0) = 0, \qquad x \in \Omega_0. \tag{4.1.1c}$$

Our discussion in this chapter will focus on the linear case. As we shall see, even the linear problem provides us with significant challenges that must first be addressed satisfactorily before we may venture into the nonlinear setting.

4.1.1 Previous Work on LSE and NLSE

Schrödinger type equations, both linear and nonlinear, have been the object of significant study (see the monograph [78] and references therein). The usual functional setting for the SE has been couched in the language of semigroup theory. Moreover, estimates are usually given for the unbounded domain case, as this most naturally models the physics behind the equation. The analytic tools involve *Strichartz estimates* along with the Banach fixed point theorem for proving well-posedness of NLSE type equations. Our interest, however, is restricted to the *bounded domain* case for obvious reasons: given our motivation in optical fiber communication, unbounded domains are not practical. Moreover, keeping computations in mind, we are naturally led to bounded (temporal and spatial) domains.

4.1.2 Inapplicability of First Order Formulations

Most physically relevant second order problems are usually obtained by reducing a system of first order equations by eliminating one of two variables. The two first order equations are often a conservation law and constitutive relation respectively. For instance, equations of linear elasticity [53] can be

written as:

$$\begin{cases} \sigma - C : \epsilon(u) = 0 \text{ in } \Omega, \\ \qquad -\text{div } \sigma = f \text{ in } \Omega, \end{cases} \tag{4.1.2}$$

where Ω is a bounded Lipschitz domain, the unknown variable u is displacement, C is the stiffness tensor, σ is Cauchy stress and $\epsilon(u)$ is the engineering strain.

In addition, stable (i.e. *inf-sup* stable) first order variational reformulations open up the possibility of using well-known *exact sequence* based discretizations. These exact-sequence conforming discretizations have well established interpolatory estimates and convergence guarantees ([21], [40] and references therein).

The general philosophy in applying the DPG method to linear problems defined using standard energy spaces $(H^1, H(\text{curl}), H(\text{div}), L^2)$ consists of the following steps:

- First prove the continuous *inf-sup* condition for a given formulation, i.e., the continuous problem is well-posed.
- Consider the corresponding *broken* formulation and apply the results of [13] that guarantee the *inf-sup* condition for the *broken* formulation.
- Apply the DPG method which automatically guarantees discrete stability (discrete *inf-sup*) by construction of optimal broken test spaces.

- Account for the approximation of optimal test functions by introducing appropriate Fortin operators [67], [46].

Thus far, the DPG method was applied to problems where well-posedness of a single formulation was equivalent to the well-posedness of *all* possible variational formulations. This meant that, simultaneously, the DPG method became applicable to various variational formulations and the pros and cons of each formulation could be studied extensively [53, 22]. Further, since different formulations imply convergence in different norms, one could try to select an "optimal" formulation for a given problem. Finally, since the problems considered thus far were defined using standard energy spaces, approximability was never an issue due to the applicability of existing optimal interpolatory error estimates.

However, in the case of the LSE, the equation is *naturally* a second order equation which is not obtained from elimination from first order equations. Thus, a naive first order reformulation is not physically meaningful.

Thus, in the case of LSE, we have encountered for the first time a problem where the (strong) second order formulation is *inf-sup* stable, but no reasonable first order formulation is stable in the L^2 setting. In addition, standard discrete energy spaces do not apply and we must consider FE discretization using operator-specific conforming elements.

Relation to Variational Formulations of Parabolic Problems:

Functional settings for variational formulations of time-dependent parabolic problems have been studied in works such as [73], [20] etc. The associated function spaces are described in the language of semigroup theory. In particular, the class of parabolic problems considered are of the form $\frac{du}{dt} + Au$ with A being a coercive operator. The coercive part of the parabolic problems considered allow for stable first-order reformulations of second order problems. Such an analysis is not available to the LSE due to the non-coercive nature of the steady state (time-independent) equation: the time-independent LSE is of the form iAu, which is not coercive.

H^1 Instability of LSE

Consider a "reasonable" first order reformulation of the LSE:

$$i\frac{\partial u}{\partial t} - \text{div } \tau = f, \qquad x \in \Omega_0,\ 0 < t < T, \tag{4.1.3a}$$

$$\nabla u - \tau = g, \qquad x \in \Omega_0,\ 0 < t < T, \tag{4.1.3b}$$

$$u(x,t) = 0, \qquad x \in \partial\Omega_0,\ 0 < t < T, \tag{4.1.3c}$$

$$u(x,0) = 0, \qquad x \in \Omega_0. \tag{4.1.3d}$$

We can re-write the operator above compactly as:

$$T\begin{pmatrix} u \\ \tau \end{pmatrix} = \begin{pmatrix} i\frac{\partial u}{\partial t} - \text{div } \tau \\ \nabla u - \tau \end{pmatrix}. \tag{4.1.4}$$

Clearly, for a well-posed formulation for $T(\cdot)$, we must, for a generic $\begin{pmatrix} f \\ g \end{pmatrix} \in L^2(\Omega) \times (L^2(\Omega))^n$ be able to conclude the $L^2(\Omega)$ control of u, τ and, therefore, ∇u. Thus, we must be able find constants $C_1, \ldots, C_6$ so that:

$$\begin{aligned} \|u\|_{L^2(\Omega)} &\leq C_1 \|f\|_{L^2(\Omega)} + C_2 \|g\|_{L^2(\Omega)} \\ \|\nabla u\|_{L^2(\Omega)} &\leq C_3 \|f\|_{L^2(\Omega)} + C_4 \|g\|_{L^2(\Omega)} \\ \|\tau\|_{L^2(\Omega)} &\leq C_5 \|f\|_{L^2(\Omega)} + C_6 \|g\|_{L^2(\Omega)}. \end{aligned} \tag{4.1.5}$$

However, as we shall presently see, this is impossible. We shall show this using an elementary separation of variables argument. This separation of variables technique in the PDE literature is also known as the *Galerkin method* (see for example [35]), although we hasten to add that this has no relation to the Galerkin scheme of discretization used for numerical methods of partial differential equations (PDEs).

Briefly, the method is to consider series expansions of the variables in appropriate function spaces (for instant, eigenfunction expansions) and reduce the PDE to a system of ODE's. The existence/uniqueness theory of ODE's can then be used to conclude the existence and uniqueness of the original PDE.

Let us define $Lu := -\Delta u$ and $Au := iu_t + Lu$. Given the Dirichlet boundary conditions, we may conclude the existence of an $L^2(\Omega_0)$ orthonormal basis of eigenfunctions of the Laplace operator L. Thus, we have an eigenbasis $e_k(x), k = 1, \ldots,$ and eigenvalues ω_k^2 that satisfy, for each k,

$$Le_k(x) = \omega_k^2 e_k(x). \tag{4.1.6}$$

Note that the ω_k^2 are unbounded in k, i.e., $0 < \omega_1^2 \leq \omega_2^2 \leq \ldots \omega_k^2 \to \infty$.

By elliptic regularity theory applied to L, we conclude that $e_k(\cdot) \in H^1(\Omega_0)$. Using this eigenbasis, we can write

$$u(t,x) = \sum_{k=1}^{\infty} u_k(t) e_k(x),$$

where the equality is in the $L^2(\Omega)$ sense. From here, we conclude

$$Au(t,x) = (i\dot{u}_k(t) + \omega_k^2 u_k(t)) e_k(x).$$

Here, the dot $(\dot{u}_k)$ indicates the time derivative.

Now, $Au = f$ then implies the following system of ODEs for the coefficients $u_k(t)$:

$$i\dot{u}_k(t) + \omega_k^2 u_k(t) = f_k(t),$$

where $f_k(t) = (f, e_k(x))_{L^2(\Omega)}$.

From the initial condition $u = 0$ at $t = 0$, we have the following solution for the $u_k(t)$:

$$u_k(t) = -i \int_0^t e^{i\omega_k^2 (t-s)} f_k(s) ds,$$

and so

$$u(t,x) = \sum_{k=1}^{\infty} (-i \int_0^t e^{i\omega_k^2 (t-s)} f_k(s) ds) e_k(x).$$

We can clearly conclude that $\|u\|_{L^2(\Omega)} \leq C_1 \|f\|_{L^2(\Omega)}$ for some constant $C_1 > 0$. However, given only that $f \in L^2(\Omega)$, we cannot conclude that there

is some $C_2 > 0$ such that $\|\nabla u\|_{L^2(\Omega)} \leq C_2 \|f\|_{L^2(\Omega)}$. Indeed, the choice of $f(t,x) \in L^2(\Omega)$ such that:

$$f_k(t) = \frac{1}{k} e^{i\omega_k^2 t},$$

shows that the corresponding solution of $Au = f$ has unbounded gradient, i.e., $\|\nabla u\|_{L^2(\Omega)} = \infty$ (see [27] for full details of this analysis).

The existence of such solutions immediately rules out the L^2 *inf-sup* stability of first order formulations such as (4.1.3).

Relation with Gelfand Triples:

A possible fix to the H^1 instability would be to allow only regular (say at least $H^1(\Omega)$) right hand sides f, g. However, if we were to require more regular loads, the corresponding variational formulation would necessarily require a *less* regular test space. While this would result in a stable first order formulation, the functional setting would take us out of the ambit of *Gelfand triples [44]*, where $L^2(\Omega)$ is identified with its dual and serves as the pivot space. This complication will be avoided and we maintain $L^2(\Omega)$ as the pivot space. We therefore are in search of stable variational formulations within the Gelfand triple framework.

As we shall see in the following section, the *strong* operator equation $Au = f$ gives rise to a second order variational formulation (the *strong* formulation) which is *inf-sup* stable. In addition, once we prove the *inf-sup* stability

of the strong formulation, we can prove the *inf-sup* stability of the so-called "ultraweak" (UW) variational formulation. Thus, one still can use the DPG method, but among *all* possible variational formulations, only two satisfy the *inf-sup* condition.

This situation is similar to the time-dependent linear acoustics equation or time-dependent linear Maxwell's equation. Here too one has *inf-sup* stability in the L^2 setting for two formulations: the strong (or least squares) and the UW formulations. The main conclusion we can draw from this observation is that these equations (and the LSE) do not allow for selective spacetime relaxation: one must either relax *both* space *and* time, or *neither* space *nor* time.

4.1.3 Relation With Previous Work

Various authors have studied the concept of generalized boundary operators using notions very similar to our operational definition. Most notably, the work of Friedrichs [38] was an early precursor in the area with the development of so-called *Friedrichs* systems as a means of studying elliptic and hyperbolic problems within a single unified framework. In [34],[33], the authors present a thorough analysis of Friedrichs' systems and recast the theory in an elegant functional setting. A DPG version of the Friedrichs theory was analyzed in [10]. Our work differs substantially from the Friedrichs system approach. First, Friedrichs systems assume the operator A satisfies $\|(A + A^*)\phi\|_{L^2(\Omega)} \leq C\|\phi\|_{L^2(\Omega)}$ for all ϕ in the domain of the operator. We

clearly do not have such a bound for the Schrödinger operator. Moreover, Friedrichs systems typically apply to first-order or systems of first-order equations. In our case, we are forced to deal directly with the second order operator. While the work in [84] comes close to ours, it still is limited to the first order setting. Our work thus is a generalization of these previous approaches.

4.2 DPG Variational Formulations

As an outline to this chapter, we identify appropriate functional spaces and prove *inf-sup* stability of the strong and ultraweak (UW) variational formulations in this functional setting for the LSE. In order to do this, we define the notion of a generalized "auxiliary" duality map that is used in place of the trace map, due to the non-avalibility of a trace energy space specific to the linear Schrödinger operator. Our results generalize the results of [13] and [84] to the case of a general (higher order) differential operator. In addition, we develop optimal interpolatory error estimates (in one space dimension) for a custom made FE space that conforms to our duality map. We also provide numerical evidence that corroborates our theoretical estimates. We refer the interested reader to [27] for full details and proofs of theorems in this chapter, some of which are omitted here for sake of brevity.

4.2.1 The Strong and UW Variational Formulations

Let $\Omega_0 \subset \mathbb{R}^n$ $(n \geq 1)$ be an open bounded Lipschitz domain. The spatial variable x lies in Ω_0 and the temporal variable $t \in (0, T)$ with $T < \infty$.

We let $\Omega = \Omega_0 \times (0,T)$ and define these parts of $\partial\Omega$:

$$\Gamma = \partial\Omega_0 \times [0,T] \cup \Omega_0 \times \{0\}, \qquad \Gamma^* = \partial\Omega_0 \times [0,T] \cup \Omega_0 \times \{T\}$$

The Schrödinger initial boundary value problem (IBVP) is:

$$i\frac{\partial u}{\partial t} - \Delta u = f, \qquad x \in \Omega_0,\ 0 < t < T, \tag{4.2.1a}$$

$$u(x,t) = 0, \qquad x \in \partial\Omega_0,\ 0 < t < T, \tag{4.2.1b}$$

$$u(x,0) = 0, \qquad x \in \Omega_0. \tag{4.2.1c}$$

Here f is any function in $L^2(\Omega)$.

We set A to be the Schrödinger operator:

$$Au := i\frac{\partial u}{\partial t} - \Delta u.$$

Note that A is formally self-adjoint: $A = A^*$. We first define the space W as: $W = W^* = \{u \in L^2(\Omega) : \ i\partial_t u - \Delta_x u \in L^2(\Omega)\}$. Next, we can then define the boundary operator $D = D^* : W \to W'$ $\langle Dw, \widetilde{w}\rangle_W = (Aw, \widetilde{w})_\Omega - (w, A\widetilde{w})_\Omega$ for all $w, \widetilde{w} \in W$.

Finally, we define the domain of A as

$$\mathrm{dom}(A) := \{u \in W : \langle Dv, u\rangle_W = 0,\ \forall v \in \mathcal{V}^*\}, \tag{4.2.2}$$

where $\mathcal{V}^* = \{\phi \in \mathcal{D}(\bar{\Omega}) : \ \phi|_{\Gamma^*} = 0\}$.

In a very similar fashion, we can define $\mathrm{dom}(A^*)$ through $\mathcal{V}$, where $\mathcal{V}$ is defined (similar to $\mathcal{V}^*$) as the space of smooth distributions vanishing on

Γ. We set $V = \mathrm{dom}(A)$ with the tacit understanding that V is given the W-topology, and we likewise set $V^* = \mathrm{dom}(A^*)$ with the tacit understanding that V^* is given the W^*-topology. We make the following density assumption which is proved in [27] for the 1D space case:

Assumption 1. Assume $\mathcal{V}^*$ is dense in V^* and $\mathcal{V}$ is dense in V.

This assumption is proved in 1 space dimension in [27]. We now have our main theorems:

Theorem 4.2.1. *Suppose Assumption 1 holds. Then the linear Schrödinger operator $A : V \to L^2(\Omega)$ is a continuous bijection. Therefore, the strong formulation is* inf-sup *stable.*

Proof. See [27]. ■

Now we consider the "ultraweak" formulation. This is a mesh-dependent formulation. The reader is referred to [27] for the definitions of the spaces W_h, Q and linear/bilinear forms etc. in the next theorem statement.

Problem 4.2.2 (Ultraweak formulation). *Given $F \in W_h'$, find $u \in L^2(\Omega)$ and $q \in Q$ such that*

$$b((u,q),v) = F(v), \qquad \forall v \in W_h.$$

Theorem 4.2.3. *Suppose Assumption 1 holds. Then Problem 4.2.2 is well posed, i.e., there is a $C > 0$ such that given any $F \in W_h'$, there is a unique solution $(u,q) \in L^2(\Omega) \times Q$ to Problem 4.2.2 and it satisfies*

$$\|u\|_\Omega^2 + \|q\|_Q^2 \leq C \|F\|_{W_h'}^2.$$

Proof. See [27]. ■

4.3 Error Estimates for the ideal DPG method

We now proceed to analyze the convergence of the ideal DPG method for Problem 4.2.2. Again, we refer the reader to [27] for the details in this section. The ideal DPG method finds u_h and q_h in finite dimensional subspaces $U_h \subset L^2(\Omega)$ and $Q_h \subset Q$ respectively, satisfying

$$b((u_h, q_h), v) = F(v), \qquad \text{for all } v \in T(U_h \times Q_h). \tag{4.3.1}$$

Here $T : L^2(\Omega) \times Q \to W_h$ is defined by $(T(z, r), v) = b((z, r), v)$ for all $v \in W_h$ and any $(z, r) \in L^2(\Omega) \times Q$. The main feature of the ideal DPG method is that the wellposedness of Problem 4.2.2 implies quasioptimality of the method's error [24]. The wellposedness of Problem 4.2.2 follows from Theorem 4.2.3. Hence to obtain convergence rates for specific subspaces, we need only develop interpolation error estimates. Since the interpolation properties of the L^2-conforming U_h are standard, we need only discuss those of Q_h. To study this, we will create a spacetime finite element space $V_h \subset V$, then identify Q_h as $D_h(V_h)$, and finally establish interpolation estimates for Q_h using those for V_h. Note that V_h will be used only in the proof (and not in the computations).

To transparently present the ideas, we shall limit ourselves to the very simple case of a uniform mesh Ω_h of spacetime square elements of side length h. Let $\mathcal{E}_h$ denote the set of edges of Ω_h. On any $E \in \mathcal{E}_h$, let $P_p(E)$ denote the space of polynomials on the edge of degree at most p. On any $K \in \Omega_h$, let

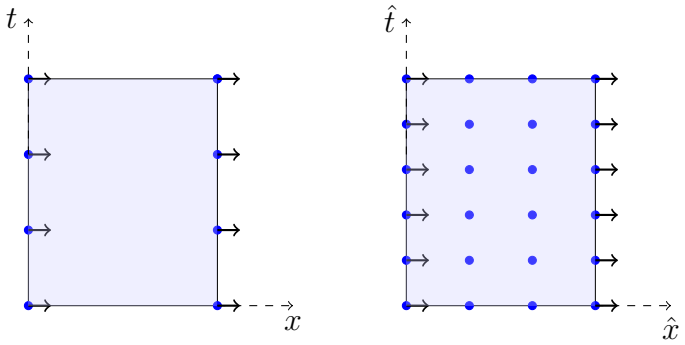

Figure 4.1: Degrees of freedom in the $p = 3$ (left) and $p = 5$ (right) cases.

$Q_p(K)$ denote the space of polynomials of degree at most p in x and at most p in t. To begin the finite element construction, we consider the reference element $\hat{K} = (0,1) \times (0,1)$ and the element space $Q_p(\hat{K})$, endowed with the following degrees of freedom: For any $w \in H^3(K)$, and for each $i \in \{0, 1, \ldots, p-2\}$ and $j \in \{0, 1, 2, \ldots, p\}$, write $x_i = i/(p-2)$ and $t_j = j/p$ and set

$$\sigma_{ij}(w) = w(x_i, t_j), \qquad \sigma_j^0(w) = \partial_x w(0, t_j), \qquad \sigma_j^1(w) = \partial_x w(1, t_j).$$

Together, these form a set Σ with $(p-1)(p+1) + 2(p+1)$ linear functionals. The triple $(\hat{K}, Q_k(\hat{K}), \Sigma)$ is a unisolvent finite element, in the sense of [17], as we show next.

Lemma 4.3.1. *Suppose $p \geq 3$. Then any polynomial $w \in Q_p(\hat{K})$ is uniquely defined by the values of its degrees of freedom σ in Σ.*

Proof. From [27]. Suppose $w \in Q_p(\hat{K})$ and $\sigma(w) = 0$ for all $\sigma \in \Sigma$. Then $w_j(x) = w(x, t_j)$ is a polynomial of degree p in one variable (x). The Hermite

and Lagrange degrees of freedom on $t = t_j$ imply $w_j = 0$. Now, fixing x, observe that the polynomial $w(x,t)$ is of degree at most p in the variable t and has $p+1$ zeros. Hence $w \equiv 0$ and the proof is complete since $\dim Q_p(\hat{K})$ equals the number of degrees of freedom. ■

Next, consider the global finite element space $W_h^p(\Omega) = \{w \in L^2(\Omega) : \partial_t w$ and $\partial_{xx} w$ are in $L^2(\Omega)$ and $w|_K \in Q_p(K)$ for all $K \in \Omega_h\}$. Each element $K \in \Omega_h$ is obtained by mapping the reference element $\hat{K}$ by $T_K : \hat{K} \to K$, $T_K(\hat{x}, \hat{t}) = (h\hat{x} + x_K, h\hat{t} + t_K)$, where (x_K, t_K) is the lower left corner vertex of K, and the element space $Q_p(K)$ is the pull back of the reference element space $Q_p(\hat{K})$ under this map. The space $W_h^p(\Omega)$ can be controlled by a global set of degrees of freedom obtained by mapping the reference element degrees of freedom and, as usual, coalescing those that coincide at the mesh element interfaces.

On the reference element $\hat{K}$, the degrees of freedom define an interpolation operator

$$\hat{\Pi} w = \sum_{\sigma \in \Sigma} \sigma(w)\, \varphi_\sigma$$

where, as usual, $\{\varphi_\eta \in Q_p(\hat{K}) : \eta \in \Sigma\}$ is the set of shape functions obtained as the dual basis of Σ. By the Sobolev inequality in two dimensions, $\hat{\Pi} : H^3(\hat{K}) \to Q_p(\hat{K})$ is continuous. Similarly, the global degrees of freedom define an interpolation operator $\Pi : H^3(\Omega) \to W_h^p(\Omega)$ satisfying

$$(\Pi w) \circ T_K = \hat{\Pi}(w \circ T_K). \tag{4.3.2}$$

Lemma 4.3.2. *If $w \in H^{p+1}(\Omega)$, then for all $p \geq 3$,*

$$\|w - \Pi w\|_\Omega \leq Ch^{p+1}|w|_{H^{p+1}(\Omega)}$$
$$\|\partial_t(w - \Pi w)\|_\Omega \leq Ch^{p}|w|_{H^{p+1}(\Omega)}$$
$$\|\partial_{xx}(w - \Pi w)\|_\Omega \leq Ch^{p-1}|w|_{H^{p+1}(\Omega)}.$$

Proof. From [27]. Changing variables $(x,t) = T_K(\hat{x},\hat{t})$ as $(\hat{x},\hat{t})$ runs over $\hat{K}$, integrating, and using (4.3.2),

$$\|w - \Pi w\|_K = h\|\hat{w} - \hat{\Pi}\hat{w}\|_{\hat{K}} \tag{4.3.3a}$$
$$\|\partial_t(w - \Pi w)\|_K = \|\partial_{\hat{t}}(\hat{w} - \hat{\Pi}\hat{w})\|_{\hat{K}} \tag{4.3.3b}$$
$$\|\partial_{xx}(w - \Pi w)\|_K = h^{-1}\|\partial_{\hat{x}\hat{x}}(\hat{w} - \hat{\Pi}\hat{w})\|_{\hat{K}}. \tag{4.3.3c}$$

On the reference element, since $H^{p+1}(\hat{K}) \hookrightarrow H^3(\hat{K})$, the interpolation operator $\hat{\Pi} : H^{p+1}(\hat{K}) \to Q_p(\hat{K})$ is continuous. Moreover $\hat{\Pi}\hat{w} = \hat{w}$ for all $\hat{w} \in Q_p(\hat{K})$. Hence, the Bramble-Hilbert Lemma yields a $\hat{C} > 0$ such that $\|\hat{w} - \hat{\Pi}\hat{w}\|_{H^3(\hat{K})} \leq \hat{C}|\hat{w}|_{H^{p+1}(\hat{K})}$ for all $\hat{w} \in H^{p+1}(\hat{K})$. Since $|\hat{w}|_{H^{p+1}(\hat{K})} \leq Ch^p|w|_{H^{p+1}(K)}$, combining with (4.3.3) and summing over all the elements in Ω_h, we obtain the result. ■

Now we are ready to present the main result of this section. Set $V_h = W_h^p(\Omega) \cap V$ and

$$Q_h = D_h(V_h), \quad U_h = \{u \in L^2(\Omega) : u|_K \in Q_{p-1}(K) \text{ for all } K \in \Omega_h\}. \tag{4.3.4}$$

Theorem 4.3.3. *Let $p \geq 3$. Suppose $u \in V \cap H^{p+1}(\Omega)$ and $q = D_h u$ solve Problem 4.2.2 and suppose $U_h \times Q_h$ is set by* (4.3.4)*. Then, there exists a*

constant C independent of h such that the discrete solution $u_h \in U_h$ and $q_h \in Q_h$ solving (4.3.1) *satisfies*

$$\|u - u_h\|_\Omega + \|q - q_h\|_Q \leq Ch^r |u|_{H^{r+2}(\Omega)} \tag{4.3.5}$$

for $2 \leq r \leq p-1$.

Proof. From [27]. By [24, Theorem 2.2] the ideal DPG method is quasioptimal:

$$\begin{aligned}\|(u,q) - (u_h,q_h)\|^2_{U\times Q} \leq\ & C \inf_{(z_h,r_h)\in U_h\times Q_h} \|(u,q) - (z_h,r_h)\|^2_{U\times Q} \\ &= C \inf_{(z_h,r_h)\in U_h\times Q_h} \left(\|u - z_h\|^2_\Omega + \|q - r_h\|^2_Q\right).\end{aligned}$$

Because of the standard approximation estimate $\inf_{z_h\in U_h} \|u-z_h\|_\Omega \leq Ch^r|u|_{H^r(\Omega)}$ for $0 \leq r \leq p-1$, it suffices to focus on $\|q - r_h\|_Q$. Since $q = D_h u$, by the definition of Q-norm (see [27]), and the fact that any r_h in Q_h equals $D_h v_h$ for some $v_h \in V_h$, we have

$$\inf_{r_h\in Q_h} \|q - r_h\|_Q \leq \inf_{v_h\in V_h} \|u - v_h\|_W \leq \|u - \Pi u\|_W.$$

Applying Lemma 4.3.2, the result follows. ■

We conclude this section by examining a property of Q_h that is useful for computations. Let $\mathcal{E}_h^{\shortmid}$ and $\mathcal{E}_h^{-}$ denote the set of vertical and horizontal (closed) mesh edges, respectively, and $\mathcal{E}_h^+ = \mathcal{E}_h^{\shortmid} \cup \mathcal{E}_h^{-}$. Let $E_h^{\shortmid}$ and E_h^+ denote the closed set formed by the union of all edges in $\mathcal{E}_h^{\shortmid}$ and $\mathcal{E}_h^+$, respectively. Let $Q_h^{\shortmid} = \{r \in L^2(E_h^{\shortmid}) : r|_F \in P_p(F) \text{ for all } F \in \mathcal{E}_h^{\shortmid}\}$ and $Q_h^+ = \{r \in L^2(E_h^+) : r$ is continuous on E_h^+ and $r|_F \in P_p(F)$ for all $F \in \mathcal{E}_h^+$ and $r|_\Gamma = 0\}$. For any

$v_h \in V_h$, since v_h is a polynomial on each element, we may integrate by parts element by element to get

$$\begin{aligned}\langle D_h v_h, \psi\rangle_h &= (A_h v_h, \psi)_h - (v_h, A_h\psi)_h \\ &= \sum_{K\in\Omega_h} \int_{\partial K} i n_t v_h \bar{\psi} + \int_{\partial K} v_h n_x (\partial_x \bar{\psi}) - \int_{\partial K} n_x (\partial_x v_h)\bar{\psi},\end{aligned}$$

for all $\psi \in \Delta(\bar{\Omega})$. Thus $q = D_h v_h$ satisfies

$$\langle q, \psi\rangle_h = \sum_{K\in\Omega_h} \int_{\partial K} q^+ (i n_t \bar{\psi}) + \int_{\partial K} q^+ n_x (\partial_x \bar{\psi}) - \int_{\partial K} q^{\shortmid} (n_x \bar{\psi}),$$

where $q^+ = v_h|_{E_h^+}$ and $q^{\shortmid} = \partial_x v_h|_{E_h^{\shortmid}}$. In computations, one may therefore identify Q_h with the interfacial polynomial space $Q_h^+ \times Q_h^{\shortmid}$ whose components are of degree at most p.

4.4 Numerical Results

We now present our numerical results. We consider the standard LSE in one space dimension with Dirichlet conditions but with a coefficient β for the u_{xx} term:

$$i u_t - \beta u_{xx} = f. \tag{4.4.1}$$

Figure 4.2 shows the rates of convergence for two different manufactured solutions. We plot the error (L^2 error) in the field variable $u(t, x)$, i.e., we plot $\|u - u_h\|_{L^2(\Omega)}$ versus the number of degrees of freedom N_h. First, on the right of figure 4.2 is the convergence plot corresponding to the complex Gaussian solution:

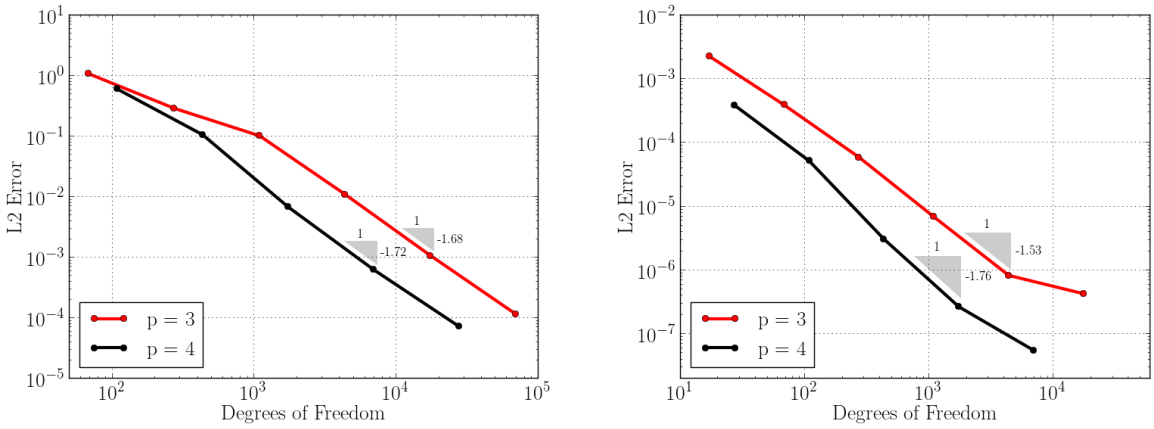

Figure 4.2: Convergence plots for the DPG method applied to the ultraweak (UW) formulation of the one dimensional LSE.

$$u(x,t) = \frac{MT_0}{\sqrt{T_0^2 - i\beta t}} e^{-\frac{x^2}{2(T_0^2 - i\beta t)}}, \tag{4.4.2}$$

where M, T_0, and β are fiber-dependent constants (see [75]). The origin is shifted to avoid possible singularities. Our simulations used non-dimensionalized units of $M = T_0 = 1.5$ and $\beta = 2.5$. On the left is the covergence plot corresponding to the manufactured solution with a standard Gaussian beam which is rotated by 45°.

Given that the minimum polynomial order that we require for the conforming element described earlier is $p = 3$, our simulations use polynomial order $p = 3, 4$.

We observe that due to the discretization with second order derivatives, we should expect the conditioning of the DPG system to come into the spotlight at some stage. Indeed, the DPG system with $p = 3$ or $p = 4$ has, after

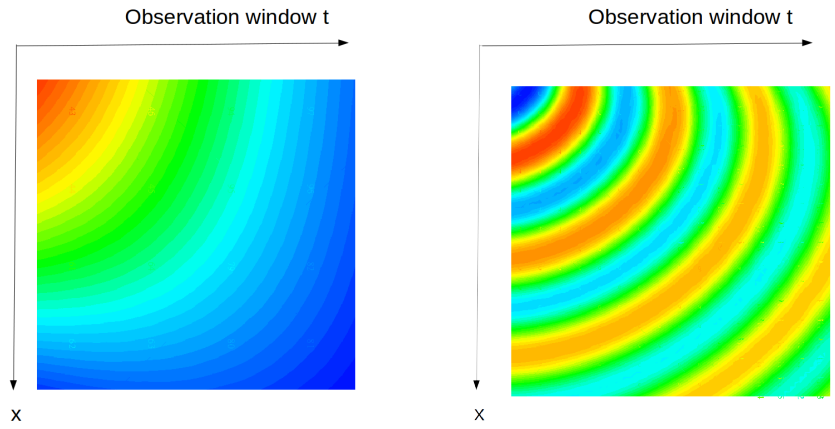

Figure 4.3: Plots of manufactured solutions: complex Gaussian (left), and Gaussian beam solution with $\omega = 20$ (right)

4-5 uniform refinements, a condition number in the vicinity of $O(10^{10})$. Therefore, the roundoff effect becomes apparent after we achieve an error threshold around 10^{-6} or 10^{-7}. This is reflected in the "flattening out" of the rates in the convergence plot corresponding to the complex Gaussian solution. Note that since we work in dimension $d = 2$, the rate r is theoretically expected to be $\frac{p-1}{d}$. We see rates of $\frac{p}{d}$ for the $p = 3$ case and $\frac{p-\frac{1}{2}}{d}$ for the $p = 4$ case

In the localized Gaussian beam case, we start with a higher error $O(1)-O(10)$ due to our use of a resonably high wave number. Thereafter, we see rates of roughly $\frac{p}{d}$ for the $p = 3$ case and $\frac{p-\frac{1}{2}}{d}$ for the $p = 4$ case, as was the case with the Gaussian exact solution.

We also report here adaptivity of our numerical method. We restricted

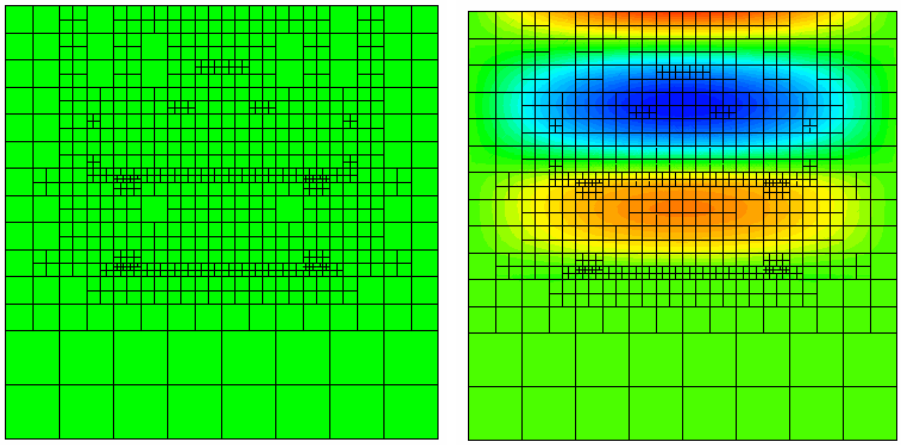

Figure 4.4: h Adaptive mesh and solution

our attention to h-adaptivity with $p = 3$. In figure 4.4, we show the resulting mesh after 10 adaptive h refinements of the inhomogeneous equation 4.4.1 with a piecewise constant load of $10i$ in the rectangular region $[.25, .75] \times [.4, .6]$ and zero elsewhere in the unit square. The adaptivity was driven by the DPG residual. In the log-log plot of figure 4.5 we see the residual decaying.

Details of Numerical Simulations: We first comment on the details of the numerical discretization used in our simulations. The theory we developed treats the boundary variable $D_h u = q$ as an indepedent unknown in the ultra-weak formulation. However, from a discretization point of view, the *action* of the discrete boundary operator $D_h u$ (see [27]) on the boundary can be viewed as a combination of two *independent* boundary actions of variables $\hat{u}$ and $\hat{u}_x$ where $\hat{u}$ is globally continuous on the element boundary while $\hat{u}_x$ is continuous on the edges parallel to the t-axis. The theory dictates that both $\hat{u}$ and $\hat{u}_x$ are

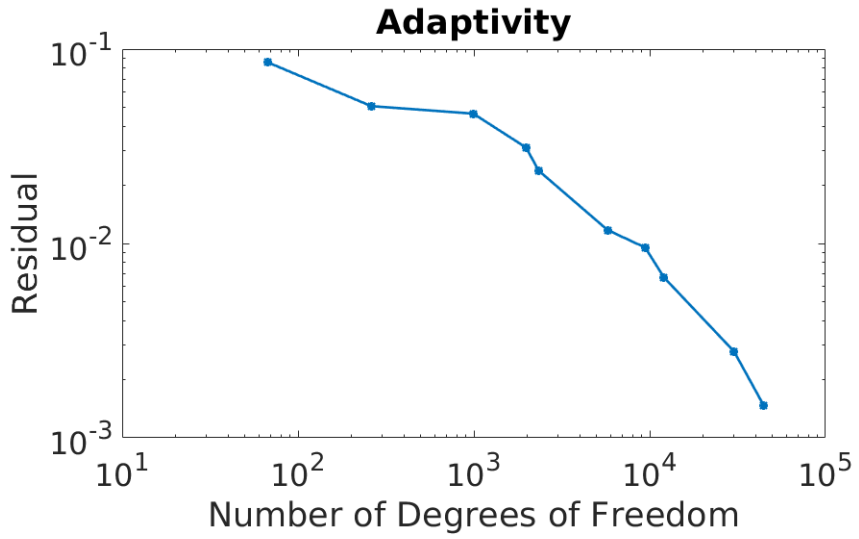

Figure 4.5: Decay of residual

to be of the same polynomial order.

However, our implementation has been done in the standard Petrov-Galerkin code supporting the exact sequence elements of the first type [40]. Consequently, $\hat{u}$ is discretized with (continuous) traces of H^1 conforming elements of order p but $\hat{u}_x$ is discretized with (discontinuous) traces of $H(\mathrm{div})$ conforming elements of order $p-1$, i.e., one order less than required by the presented interpolation theory[2]. In turn, the L^2 variable u is discretized with elements of order $p-1$.

Second, we note that we use the "practical" DPG method which involves inversion of the approximate test Riesz map ([67], [46]). In order to

[2]Still, we do not see any lower rates of convergence which may be one more indication of the suboptimality of our analysis

invert the approximate Riesz map, our simulations used additional enrichment orders $\Delta p = 1, \Delta p = 2$. No significant differences are seen with increasing Δp, and the presented numerical results have been obtained with $\Delta p = 1$.

Chapter 5

Raman Gain Model

The main aim of this chapter[1] and the next is to present a full Maxwell, three dimensional (3D) Discontinuous Petrov-Galerkin (DPG) simulation of a fiber amplifier, using Raman gain [82, 57, 77, 59] in a typical passive, step-index, core-pumped optical fiber amplifier as the test case for initial validation purposes.

Author contributions: The contents of this chapter are taken largely from the multi-author article "A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers" S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, J. Mora, in preparation. The article has not yet been submitted for journal publication. The author of this dissertation contributed to model development, and code/numerical implementation of the model and analysis of the results.

[1]The content of this chapter is taken from the manuscript "A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers" S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, J. Mora, in preparation. Information approved for public release on 08 May 2018 by AFRL OPSEC/PA OPS-18-19547.

5.1 Introduction

In this chapter, we present several novel advances, both in the modeling as well as in the methodology used to study Raman amplification in a full vectorial model. First, our propagation model makes minimal assumptions on the electromagnetic fields in question, unlike the scalar beam propagation method (BPM, see [68, 74, 63, 83, 4] and references therein), which assumes a polarization maintaining propagation of the electromagnetic fields in an optical fiber, whereas our treatment is truly vectorial. Though both semi-vectorial and full vectorial BPM approaches have already been implemented (see [49, 50, 71, 72, 36] and references therein), we are introducing a fiber model that is a full boundary value problem rather than an initial value problem. In addition, we employ 3D isoparametric curvilinear elements to model the curved fiber (core and inner cladding) geometry, which can also later be used for studying microstructure fibers or hollow-core gas-filled fiber lasers. Indeed, most scalar fiber modeling techniques assume, starting with the initial condition, that only one of the three electric (and corresponding magnetic) field components dominate in magnitude during propagation, and thus treats the non-dominant components as zero. This is due to the assumption that the source light is robustly linearly polarized, and is thus only launched into one of the three electric field components, usually also neglecting the corresponding magnetic field component. Also, by assuming that the fiber is polarization maintaining (either by design or by active control), we can expect negligible field coupling as the light propagates through the fiber. While this assumption reduces the

complexity of the model from a vectorial curl-curl Maxwell system to a scalar Helmholtz system, it may be the case that such assumptions may not hold to the degree required for the model to be accurate, especially in the presence of injected light that is not perfectly linearly polarized, or when there are high intensities, manufacturing defects, fiber bending, thermal effects, and/or the presence of embedded microstructures. In other words, the weakly coupled polarization states assumption may not be true in general, which would result in non-trivial coupling between the electromagnetic field components as the light propagates down the fiber.

Second, we propose a novel full vectorial time-harmonic 3D model for Raman gain. We show how Raman gain, which typically is viewed as a non-linear third-order susceptibility component of the electric polarization, can be derived by assuming that it originates from a mostly imaginary perturbation to the refractive index, just as active gain is usually derived. The proposed Raman model is particularly significant, since this fits well with, if not instrumental for, our full Maxwell simulation efforts, even though this effort centers on the validation of the numerical approach and not on a demonstration of polarization coupling.

Our model incorporates the fact that the ultraweak (UW) DPG formulation, used for solving the electromagnetic equations, provide us with both electric and magnetic fields. Thus, we are able to compute the time-averaged Poynting vector (irradiance) using the DPG trace variables. In this context, we also note the fact that we utilize a frequency domain perfectly matched

layer (PML, see [6, 79, 16, 43, 60, 61, 81]), which is also implemented using the ultraweak DPG formulation. As we shall explain, the use of a PML is critically important, and one cannot adequately observe the gain phenomenon with simpler impedance boundary conditions. Moreover, for element computations, we employ *sum factorization* to integrate the local DPG matrices, which significantly accelerates the otherwise temporally expensive element integration [62, 58].

Thus, to our knowledge, the contents of this chapter are the first attempts at a general, full vectorial simulation of 3D Maxwell equations with a nonlinear gain term, equipped with a PML, in the context of higher-order Galerkin-based simulations. Also, the discussion in this chapter introduces an innovative formulation of Raman gain [82, 57, 77, 59] amenable to a vectorial simulation, which presumes that only the measured bulk Raman gain coefficient is available to the computer modeling team, as is almost always the case. Our simulations are, at this stage, not scaled to perform on supercomputing infrastructures, as would be needed for modeling fiber amplifiers of realistic sizes. However, the novelty of applying such a generalized approach to solving a vectorial, nonlinear fiber amplifier model with advanced 3D DPG technologies that provide the necessary accuracy and the unprecedented computational efficiency (for this type of methodology) is the major contribution of results of this chapter. Subsequent investigations into code optimization, scaling and parallelism will allow for mega-scale simulations of not only Raman gain but other phenomenon endemic to high-power fiber amplifiers such as stimulated

Brillouin scattering (SBS), the transverse mode instability (TMI), thermal lensing, fiber bending, etc. [2, 1, 7, 65, 66]. Indeed, the gain results presented here serve as an important step in (and are motivated by) the need to study thermal instabilities that arise due to heating of optical fibers when used in high power regimes. Also, this formulation of the governing equations, along with the resulting simulation approach, can serve as a basis for studying new amplifier configurations and/or for optimizing microstructure designs in future efforts.

We emphasize that our aim is to obtain *qualitative* results that indicate the feasibility of the methodology applied to the full Maxwell model, and, as such, we use an artificially large Raman gain coefficient so as to be able to see the gain effects in a short enough fiber length that all calculations can be accomplished on a single laptop or workstation. The remainder of this chapter is organized as follows. In Section 5.2, we provide details of the physics underlying our model, introduce the novel, full vectorial electric polarization term that accounts for Raman gain, and delineate the system of equations that we will be solving. Section 6.1 briefly outlines how the DPG methodology is applied to general broken variational formulations. The discussions in Sections 5.2 and 6.1 are unified in Section 6.2, which provides details of the variational formulations, the nonlinear iterative scheme, and the time-harmonic Poynting theorem that are used in our model. We discuss in detail our results in Section 6.3. Three appendices to this dissertation provide details of the PML, the sum factorization implementation, and the theoretical underpinnings of the DPG

approach via a comparison of the numerical differences between the primal and ultraweak formulation of Maxwell's equations.

5.2 3D Maxwell Raman Gain Model

5.2.1 Fiber Model

This model considers a continuous wave (cw), double clad, non-dispersive, circularly symmetric, weakly guided, step-index fiber amplifier, where the core and cladding regions are isotropic and homogeneous (see Fig. 5.1). The outer layer (second cladding) of the fiber is a polymer coating that covers the inner cladding of fused silica. The refractive index of the core (n_{core}) and cladding (n_{cladding}) satisfy $n_{\text{core}} - n_{\text{cladding}} \ll 1$. Since it is assumed that all of the light (pump and signal laser fields) in this fiber is guided in the core region by total internal reflection, the subsequent model will ignore the polymer jacket, given that it has almost no effect on the core guided light. Because both the pump and signal fields are seeded into the core region at the beginning of the fiber ($z = 0$), it is assumed that the light only propagates in the forward direction, which is a typical approximation for a co-pumped passive fiber amplifier. Such a configuration also suggests that both the pump and the signal are already highly coherent, which means that this amplifier acts only as a frequency converter, instead of also as a brightness enhancer.

For the purposes of this Raman gain analysis, the electromagnetic fields are treated as time-harmonic. This is justified by the fact that real core-pumped Raman amplifiers are indeed usually seeded by lasers that produce

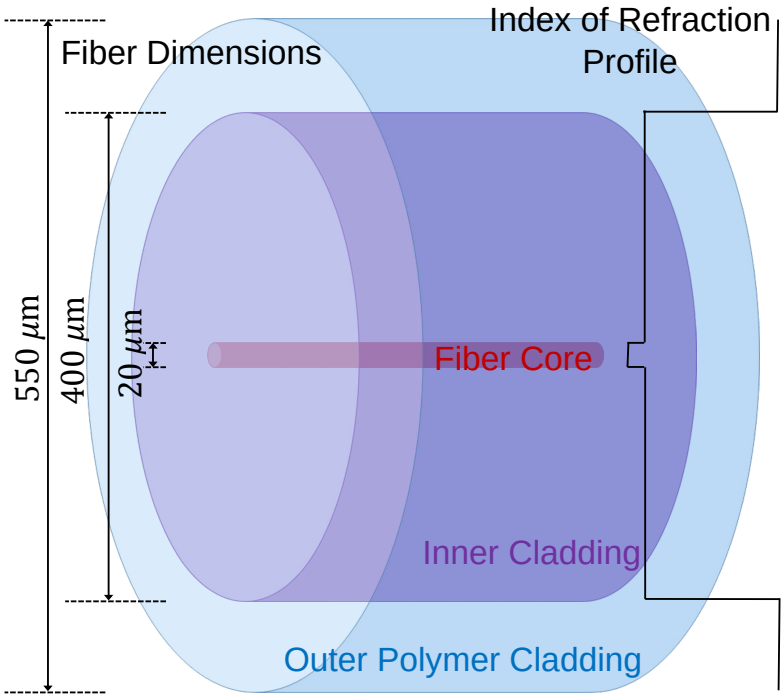

Figure 5.1: A typical circularly symmetric, double-clad, step index fiber amplifier, with a core region made of silica glass, a cladding region also made of silica glass, but with a slightly lower index of refraction than the core region, and a polymer coating, the outer cladding, with a substantially lower index of refraction than the inner cladding region. Such fibers are usually ∼5-100s meters long, even though it is depicted here as only being a few hundred microns long.

near monochromatic light, and because any other sources of time dependent behaviour, most notably thermal effects, in passive fibers only occur at significantly slower varying time scales than the optical frequencies of the light present in the fiber. Thus, the following time-harmonic ansatz is assumed for all involved electromagnetic fields:

$$\mathbb{E}_0(x,y,z,t) = \mathbb{E}(x,y,z)e^{i\omega t} + \text{ c.c. and}$$

$$\mathbb{H}_0(x,y,z,t) = \mathbb{H}(x,y,z)e^{i\omega t} + \text{ c.c.,}$$

where ω is the frequency of propagation, $i = \sqrt{-1}$ and c.c. indicates complex conjugate of the previous term.

In our application, we have *two* sets of time harmonic Maxwell equations: one corresponding to the signal field $(\mathbb{E}_s, \mathbb{H}_s)$ at frequency ω_s and the pump field $(\mathbb{E}_p, \mathbb{H}_p)$ at frequency ω_p. We shall use the index $l = s, p$ to distinguish between the signal and pump fields. The fact that the pump (p) and signal (s) fields are monochromatic and well-separated from one another, allows for solving two separate sets of Maxwell equations, which are coupled together through the Raman gain:

$$\begin{aligned} \nabla \times \mathbb{E}_l &= -i\omega_l \, \mu_0 \mathbb{H}_l, \\ \nabla \times \mathbb{H}_l &= i\omega_l \, \varepsilon_0 \mathbb{E}_l + i\omega_l \mathbb{P}_l, \\ \nabla \cdot \mathbb{E}_l &= \tfrac{\rho}{\varepsilon_0}, \\ \nabla \cdot \mathbb{H}_l &= 0, \end{aligned} \tag{5.2.1}$$

where $l = \mathrm{p}, \mathrm{s}$ is the index for the two frequencies of light, and $\mathbb{E}_l$ and $\mathbb{H}_l$ are the time-harmonic electric and magnetic fields respectively of the signal ($l = s$ and pump $l = p$). Thus, the above equation is a compressed version of *two* sets of Maxwell's equations. The free-space electric permittivity and magnetic permeability are denoted by ε_0 and μ_0, respectively. The electric charge density ρ is zero for silica fibers, and $\mathbb{P}_l$ represents the electric polarization term.

5.2.2 Polarization Model

Since silica fibers have negligible magnetic susceptibilities, all of the interactions between the electromagnetic fields and the medium can be formu-

lated mathematically through the electric polarization term ($\mathbb{P}_l$). The electric polarization can be expanded in terms of the electric field and susceptibility tensors $\boldsymbol{\chi}^{(i)}, i = 1, 2, \ldots$ as follows...

$$\mathbb{P} = \epsilon_0 \left(\underbrace{\boldsymbol{\chi}^{(1)} \cdot \mathbb{E}}_{\substack{\text{background refractive index (real)} \\ \text{active laser gain (imaginary)}}} + \boldsymbol{\chi}^{(2)} : \mathbb{E} \otimes \mathbb{E} + \underbrace{\boldsymbol{\chi}^{(3)} \vdots \mathbb{E} \otimes \mathbb{E} \otimes \mathbb{E}}_{\text{Raman gain} \propto |\mathbb{E}|^2 \mathbb{E}} + \ldots \right) \quad [2].$$

An adequate model for this demonstration of a typical co-pumped passive fiber amplifier that experiences significant Raman gain must include the background index of refraction of the fiber, which will be denoted as $\mathbb{P}_l^{\text{background}}$ and is expressed through the real part of the first-order susceptibility. Also, the model must include the contribution of the Raman gain to the electric polarization, which will be denoted as $\mathbb{P}_l^{\text{Raman}}$ and is considered to be a component of the third-order susceptibility tensor. Active laser gain ($\mathbb{P}_l^{\text{active gain}}$) in a fiber amplifier is often seen as mostly imaginary perturbation to the refractive index, and is thus expressed as part of the first-order susceptibility term. This perturbation to the refractive index can be expressed as

$$\mathbf{n}_l^2 + 2\delta n^{\text{gain}} \frac{\mathbf{n}_l}{|\mathbf{n}_l|} \approx \frac{\boldsymbol{\varepsilon}^l}{\varepsilon_0},$$

where $\boldsymbol{\varepsilon}^l$ is the dielectric tensor of the medium and $\delta n^{\text{gain}} = \delta n^{\text{gain}}(\omega_l)$ is a complex perturbation to the refractive index that causes a gain in the optical field. As will be shown presently, Raman gain can also be derived from the perspective that it is a mostly imaginary perturbation to the refractive index. A more complete model might include other effects such as linear loss ($\mathbb{P}_l^{\text{loss}}$),

thermal effects ($\mathbb{P}_l^{\text{thermal}}$) and/or other optical nonlinearities ($\mathbb{P}_l^{\text{opt. nonlin.}}$) such as stimulated Brillouin scattering (SBS), the Kerr nonlinearity, and/or four-wave mixing.

For the purposes of this chapter, the electric polarization model takes the form of

$$\mathbb{P}_l(\mathbb{E}_l) = \mathbb{P}_l^{\text{background}}(\mathbb{E}_l) + \mathbb{P}_l^{\text{Raman}}(\mathbb{E}_l),$$

where

$$\mathbb{P}_l^{\text{background}}(\mathbb{E}_l) \approx \varepsilon_0\big(\mathbf{n}_l^2 - \mathbb{I}\big)\mathbb{E}_l, \tag{5.2.2}$$

given that $\mathbb{I}$ is the identity tensor and $\mathbf{n}_l$ is the real-valued index of refraction tensor that accounts for the differences between the refractive indices of the fiber core region, the inner cladding region, and the polymer jacket region of the fiber [2, 7].

Raman scattering is an inelastic optical nonlinearity that occurs as incident light (the pump), at a sufficiently high-intensity, vibrates the molecules of the medium, resulting in optical phonons and scattered photons (the Stokes field, see [82]), usually of a lower frequency than the incident photons. This process can start from noise, but in this model the Raman scattering is stimulated by having a seeded signal field offset in frequency from the pump field so as to achieve peak Raman gain and coinciding perfectly with the Stokes field frequency. Though stimulated Raman scattering (SRS) is an optical nonlinearity, its contribution to the electric polarization can be derived in the same way that active gain is derived (see [82]), but with a different gain function. This is

somewhat surprising given that active gain is usually seen as a predominantly imaginary perturbation to the index of refraction, which is a first-order electric susceptibility term, while Raman scattering is considered to be a third-order electric susceptibility term. This perturbation to the refractive index can be expressed as

$$\mathbf{n}_l^2 + 2\delta n^{\text{gain}} \frac{\mathbf{n}_l}{|\mathbf{n}_l|} \approx \frac{\boldsymbol{\varepsilon}^l}{\varepsilon_0},$$

where $\boldsymbol{\varepsilon}^l$ is the dielectric tensor of the medium and $\delta n^{\text{gain}} = \delta n^{\text{gain}}(\omega_l)$ is a complex perturbation to the refractive index that causes a gain in the optical field.

5.2.3 Derivation of the Raman model

We provide here a generic derivation of the Raman gain polarization term in terms of the signal and pump fields. We will show how the final expression for the Raman polarization can be related to the third order susceptibility. To our best knowledge, this is a novel derivation of a general full vector 3 dimensional Maxwell equation based Raman gain term. Indeed, most other derivations are specific to beam propagation methods, and may not easily generalize to the full Maxwell case.

In order to derive the Raman gain contribution to the electric polarization, first consider how one might derive the contribution of active laser gain to the electric polarization. This approach is outlined in [82] using a scalar electric field; however, the process can be extended to a vectorial field. Even in high gain amplifiers, the gain is still a perturbation to the refractive index,

and thus one should not expect that the gain would significantly contribute to the divergence of the electric field: $\boldsymbol{\nabla} \cdot \mathbb{P}_l^{\text{gain}} \approx 0$. The gain contribution to the first-order susceptibility can be denoted as $\boldsymbol{\chi}_{\text{g}}^{(1)} = \boldsymbol{\chi}_{\text{g}}^{(1)}(x, y, z, t)$ and can be decomposed into its real and imaginary components: $\boldsymbol{\chi}_{\text{g}}^{(1)}(x, y, z, t) = \boldsymbol{\chi}_{\text{g}}^{\text{Re}}(x, y, z, t) + i\boldsymbol{\chi}_{\text{g}}^{\text{Im}}(x, y, z, t)$, where $\boldsymbol{\chi}_{\text{g}}^{\text{Re/Im}}(x, y, z, t) \in \mathbb{R}$. It is reasonable to assume that the electric field grows according to a given gain function: $g_l = g_l(x, y, z, t)$, with units of m^{-1}. The electric field vector with gain is expressed as

$$\mathbb{E}_0^l(x, y, z, t) \approx \frac{1}{2}\mathbb{E}_l(x, y, z, t)e^{\frac{\langle g_l \rangle z}{2} + i(\omega_l t - \boldsymbol{\beta}^l \cdot \mathbf{r})} + \text{ c.c.},$$

where each component of propagation constant vector $\boldsymbol{\beta}^l$ is a positive real value, $\mathbf{r} = [x\ y\ z]^{\text{T}}$, the electric field envelop $\mathbb{E}_l$ is also slowly varying in time, $\omega_l > 0$, and

$$\langle g_l \rangle(z, t) = \frac{1}{\text{A}_{D_{\text{gain}}}} \iint_{D_{\text{gain}}} g_l(x, y, z, t)\, dxdy \text{ (by the 2D Mean Value Theorem)},$$

where $\text{A}_{D_{\text{gain}}}$ represents the transverse area of the domain of the gain. The expression for the electric field assumes that both the electric field amplitude ($\mathbb{E}_l$) and the gain function (g_l) are slowly varying compared to longitudinal oscillations at a frequency of β_z^l and to temporal oscillations at a frequency of ω_l. It is reasonably assumed that if the gain function obeys slowly varying envelope approximations than so does the gain contribution to the first-order susceptibility ($\boldsymbol{\chi}_{\text{g}}^{(1)}$). Therefore,

$$\begin{aligned} \left|\partial_{zz} f\right| \ll \beta\left|\partial_z f\right| \ll \beta^2\left|f\right| \\ \left|\partial_{tt} f\right| \ll \omega\left|\partial_t f\right| \ll \omega^2\left|f\right| \end{aligned} \text{ where } f \in \{\boldsymbol{\chi}_{\text{g}}^{\text{Re/Im}}, g, \mathbb{E}\}. \tag{5.2.3}$$

Furthermore, $\mathbb{E}_l$ is assumed to be slowly varying in x and y compared to the oscillations at the frequencies β_x^l and β_y^l in the x- and y-directions respectively.

It will be assumed that the vectorial Helmholtz equation robustly holds when applied to the slowly varying electric field amplitude:

$$\left[\boldsymbol{\Delta} + \frac{\mathbf{n}_l^2 \omega_l^2}{c^2}\right] \left(\mathbb{E}_0^l e^{-\frac{\langle g_l \rangle z}{2}}\right) = 0,$$

basically indicating that the light propagates in the fiber even if there is no gain; i.e., the fiber is a waveguide. Moreover, it has been assumed that the light propagates only in the z-direction in order to simplify the mathematics, which means that Poynting vector $\mathbb{S}$, which parallel to the propagation constant vector $\boldsymbol{\beta} = \mathbf{n}_{\text{eff}}\omega/c$, is assumed to have a dominant z-component (and negligible x- and y-components). The gain has an isotropic effect, and thus its perturbative contribution to the refractive index occurs such that each direction of the refractive index is altered the same as any other direction even if the refractive index is birefringent (anisotropic). Mathematically, this is captured by $\boldsymbol{\chi}_{\text{g}}^{\text{Re/Im}} = \chi_{\text{g}}^{\text{Re/Im}} \mathbf{n}_l / |\mathbf{n}_l|$. Finally, the main idea of this derivation is to express the gain contribution to the electric polarization as a function of the first-order susceptibility due to gain:

$$\mathbb{P}_l^{\text{gain}}(x, y, z, t) = \varepsilon_0 \left(\boldsymbol{\chi}_{\text{g}}^{(1)}(x, y, z, t) \cdot \mathbb{E}_0^l(x, y, z, t)\right) = \varepsilon_0 \left(\boldsymbol{\chi}_{\text{g}}^{\text{Re}} + i\boldsymbol{\chi}_{\text{g}}^{\text{Im}}\right) \cdot \mathbb{E}_0^l. \tag{5.2.4}$$

Starting with the electric field wave equation in a dielectric medium with the gain contribution to the electric polarization term, and then applying the slowly varying envelope approximations and the vectorial Helmholtz

equation:

$$\frac{\mathbf{n}_l}{|\mathbf{n}_l|}\left[\boldsymbol{\Delta}\mathbb{E}_0^l - \frac{\mathbf{n}_l^2}{c^2}\frac{\partial^2\mathbb{E}_0^l}{\partial t^2}\right] \approx \mu_0\frac{\partial^2\mathbb{P}_l^{\text{gain}}}{\partial t^2} = \varepsilon_0\mu_0\frac{\partial^2}{\partial t^2}\left[\left(\boldsymbol{\chi}_{\text{g}}^{\text{Re}} + i\boldsymbol{\chi}_{\text{g}}^{\text{Im}}\right)\cdot\mathbb{E}_0^l\right],$$

one derives:

$$\underbrace{\left(\frac{g_l}{2}\right)^2 + \left(\frac{\omega_l}{c}\right)^2\chi_{\text{g}}^{\text{Re}}}_{\text{real-valued}} \approx \underbrace{i\left[\left(\frac{\omega_l}{c}\right)^2\chi_{\text{g}}^{\text{Im}} + g_l\beta_z^l\right]}_{\text{purely imaginary}}.$$

Now note that a real-valued function can only equal an imaginary-valued function when both functions are identically zero. Therefore, setting each side of the relation equal to zero produces the following relations:

$$\chi_{\text{g}}^{\text{Re}}(x,y,z,t) \approx -\left(\frac{g_l(x,y,z,t)c}{2\omega_l}\right)^2 \leftarrow \frac{g_l(x,y,z,t)c}{\omega_l} \ll 1 < n_l^2 \qquad (5.2.5)$$

$$\chi_{\text{g}}^{\text{Im}}(x,y,z,t) \approx g_l(x,y,z,t)\beta_z^l\left(\frac{c}{\omega_l}\right)^2 \approx \frac{n_{\text{eff}}^l c}{\omega_l}g_l(x,y,z,t). \qquad (5.2.6)$$

Continuing the derivation, one finds that

$$\delta n_l^{\text{gain}}(x,y,z,t) \approx \frac{icg_l(x,y,z,t)}{2\omega_l} \approx -\frac{\sigma_l(x,y,z,t)}{\omega_l\varepsilon_0},$$

where σ_l is the dielectric conductivity, which is another way of viewing gain in a fiber amplifier. This approximation also indicates that any contribution to the real component of the refractive index (5.2.5) by the presence of gain in the fiber is negligible in comparison to the imaginary component contribution (5.2.6). In fact, for light in the 1-2 μm wavelength range, the perturbation to real component of the index of refraction due to gain is about 7 orders of magnitude smaller than it is for the imaginary component. Therefore, one can approximate $\chi_{\text{g}}^{\text{Re}}(\mathbf{r},t)$ to be zero, and re-express the contribution of gain to

the electric polarization (5.2.4) using the derived relation for the imaginary component of the susceptibility due to gain (5.2.6) to get:

$$\mathbb{P}_l^{\text{gain}}(x,y,z,t) \approx \frac{i\varepsilon_0 c \mathbf{n}_l}{\omega_l} g_l(x,y,z,t)\mathbb{E}_l(x,y,z,t). \tag{5.2.7}$$

Unfortunately, experimentalists consistently measure the bulk Raman gain coefficient as the primary means of determining how susceptible a fiber may be to experiencing the onset of Raman scattering. This means that simulations cannot produce a model better than the limitations imposed by this constant, and the methodology used to determine its value. It is important to understand that experimentalists ascertain the Raman gain coefficient measurement from a coupled set of ODEs for power (P_l) of the pump ($l = \text{p}$) and Raman Stokes ($l = \text{S}$) fields along the length of the fiber, which can be derived from multiple simplifying assumptions applied to Maxwell's equations [57, 77, 59]. Written concisely, without including extra terms for starting the Raman scattering from noise, these coupled set of ODEs take the form of

$$\begin{aligned} \mathrm{d}P_{\text{p}}z(z) &= \frac{\Upsilon_{\text{R}}^{\text{p}} g_{\text{R}}}{\text{A}_{\text{eff}}} P_{\text{p}}(z)P_{\text{S}}(z) \text{ and } \mathrm{d}P_{\text{S}}z(z) = \frac{\Upsilon_{\text{R}}^{\text{S}} g_{\text{R}}}{\text{A}_{\text{eff}}} P_{\text{p}}(z)P_{\text{S}}(z) \text{ with} \\ \text{A}_{\text{eff}} &= \frac{\iint_{\text{A}_{\text{clad}}} (\varphi^{\text{p}})^2 \, dxdy \iint_{\text{A}_{\text{clad}}} (\varphi^{\text{S}})^2 \, dxdy}{\iint_{\text{A}_{\text{clad}}} (\varphi^{\text{p}})^2 (\varphi^{\text{S}})^2 \, dxdy}, \end{aligned} \tag{5.2.8}$$

where $g_{\text{R}} \in \mathbb{R}^+$ is the measured bulk Raman gain coefficient, and $\varphi^l = \varphi^l(x,y)$ is a single transverse mode of the fiber; presumably the fundamental mode. The dimensionless parameter Υ_{R}^l, with $l \in \{\text{p}, \text{S}\}$, allows for photon flux conservation when

$$\Upsilon_{\text{R}}^l = \begin{cases} \frac{-\omega_{\text{p}}}{\omega_{\text{S}}}, & \text{when } l = \text{p}, \\ 1, & \text{when } l = \text{S}. \end{cases}$$

These ODEs assume that the light, at either frequency, only resides in one transverse mode, which may be a limiting factor when considering large mode area (LMA) fibers. Also, note that the power at a particular point along the fiber is already independent of the transverse direction, and that the effective area calculation further washes out any transverse dependencies. Finally, recall that, in this simulation, the pump frequency is higher than the Stokes (or signal) frequency: $\omega_{\mathrm{p}} > \omega_{\mathrm{S}}$, or equivalently, $\lambda_{\mathrm{S}} > \lambda_{\mathrm{p}}$, so that the term $|\frac{-\omega_{\mathrm{p}}}{\omega_{\mathrm{S}}}| > 1$, which results in energy transfer from the pump field into the Stokes (signal) field.

This can be shown by assuming that the light is only propagating in the $+z$-direction in a guided fiber amplifier:

$$
\begin{aligned}
\sum_{l}\left[\frac{1}{\hbar\omega_l}\frac{\partial I_l}{\partial z}\right] = \frac{1}{\hbar\omega_{\mathrm{p}}}\frac{\partial I_{\mathrm{p}}}{\partial z} + \frac{1}{\hbar\omega_{\mathrm{S}}}\frac{\partial I_{\mathrm{S}}}{\partial z} = 0 \;\leftarrow\; \left[\frac{\text{photons}}{\mathrm{m}^3\cdot\mathrm{sec}}\right] \\
\text{But, } \frac{\partial I_{\mathrm{p}}}{\partial z} \approx \Upsilon_{\mathrm{R}}^{\mathrm{p}} g_{\mathrm{R}} I_{\mathrm{S}} I_{\mathrm{p}} \text{ and } \frac{\partial I_{\mathrm{S}}}{\partial z} \approx \Upsilon_{\mathrm{R}}^{\mathrm{S}} g_{\mathrm{R}} I_{\mathrm{p}} I_{\mathrm{S}} \text{ so that} \\
\frac{\Upsilon_{\mathrm{R}}^{\mathrm{p}} g_{\mathrm{R}} I_{\mathrm{p}} I_{\mathrm{S}}}{\hbar\omega_{\mathrm{p}}} + \frac{\Upsilon_{\mathrm{R}}^{\mathrm{S}} g_{\mathrm{R}} I_{\mathrm{p}} I_{\mathrm{S}}}{\hbar\omega_{\mathrm{S}}} \approx 0 \\
\frac{g_{\mathrm{R}} I_{\mathrm{p}} I_{\mathrm{S}}}{\hbar}\left[\frac{\Upsilon_{\mathrm{R}}^{\mathrm{p}}}{\omega_{\mathrm{p}}} + \frac{\Upsilon_{\mathrm{R}}^{\mathrm{S}}}{\omega_{\mathrm{S}}}\right] \approx 0 \\
\frac{\Upsilon_{\mathrm{R}}^{\mathrm{p}}}{\omega_{\mathrm{p}}} + \frac{\Upsilon_{\mathrm{R}}^{\mathrm{S}}}{\omega_{\mathrm{S}}} \approx 0 \\
\Upsilon_{\mathrm{R}}^{\mathrm{p}} \approx -\Upsilon_{\mathrm{R}}^{\mathrm{S}}\frac{\omega_{\mathrm{p}}}{\omega_{\mathrm{S}}} \\
\Upsilon_{\mathrm{R}}^{\mathrm{p}} \approx -\frac{\omega_{\mathrm{p}}}{\omega_{\mathrm{S}}} \;\leftarrow\; \text{by choosing } \Upsilon_{\mathrm{R}}^{\mathrm{S}} = 1 \\
\therefore\; \Upsilon_{\mathrm{R}}^{l} = \begin{cases} -\frac{\omega_{\mathrm{p}}}{\omega_{\mathrm{s}}}, & l = \mathrm{p} \\ 1, & l = \mathrm{S}. \end{cases}
\end{aligned}
\tag{5.2.9}
$$

Thus, even though the measurement and use of bulk Raman gain coefficient in computer models is a limiting factor, at least the total photon flux can still be conserved by choosing

$$\Upsilon_{\mathrm{R}}^{l} = \begin{cases} \frac{-\omega_{\mathrm{p}}}{\omega_{\mathrm{S}}}, & \text{when } l = \mathrm{p} \\ 1, & \text{when } l = \mathrm{S} \end{cases}.$$

A slight generalization to these power evolution ODEs (5.2.8), is the set of transverse dependent PDEs for the evolution of the irradiance ($I_l := |\mathrm{Re}(\mathbb{E}_l \times \mathbb{H}_l^*)|$) along the fiber:

$$\frac{\partial I_{\mathrm{p}}}{\partial z} = \Upsilon_{\mathrm{R}}^{\mathrm{p}} g_{\mathrm{R}} I_{\mathrm{p}} I_{\mathrm{S}} \text{ and } \frac{\partial I_{\mathrm{S}}}{\partial z} = \Upsilon_{\mathrm{R}}^{\mathrm{S}} g_{\mathrm{R}} I_{\mathrm{p}} I_{\mathrm{S}}. \tag{5.2.10}$$

These PDEs help illuminate the gain function (g_l) for Raman scattering.

Accepting that the bulk Raman gain coefficient is the primary means of determining Raman gain, it is prudent to introduce this constant directly into the derivation of gain; specifically by including g_{R} into the gain function (g_l). The gain function for Raman scattering can be extracted from the coupled irradiance PDEs (5.2.10) by choosing $g_l(x, y, z, t) = \Upsilon_{\mathrm{R}}^{l} g_{\mathrm{R}} I_k(x, y, z, t)$, where $k \neq l \in \{\mathrm{p}, \mathrm{s}\}$ and $\mathrm{S} \equiv \mathrm{s}$ for this simulation. Using this form of the Raman gain function in the expression for the contribution of gain to the electric polarization (5.2.7), which meets the necessary criteria of having units of m^{-1} and obeying the slowly varying envelop approximations (5.2.3), yields a novel and practical formulation of the Raman gain contribution to the electric polarization:

$$\mathbb{P}_l^{\mathrm{Raman}}(\mathbb{E}_l) \approx \frac{i\varepsilon_0 \mathbf{n}_l c}{\omega_l} (\Upsilon_{\mathrm{R}}^{l} g_{\mathrm{R}} I_k) \mathbb{E}_l. \tag{5.2.11}$$

Recalling that the intensity can be related to the square of the electric field, and thus so is the irradiance, it is clear that this expression sets $\mathbb{P}_{\text{Raman}} \propto |\mathbb{E}|^2\mathbb{E}$ as would be expected for a third-order susceptibility component of the electric polarization since $\mathbb{P}^l_{\text{Raman}} \propto I_k\mathbb{E}_l$ and $I_k \propto |\mathbb{E}_k|^2$. These expressions show that the Raman gain is the source of the nonlinearity for this model of a coupled system of Maxwell equations.

5.2.4 Non-dimensionalization of Governing Equations

The above mentioned equations are dimensional. The non-dimensional version of the equations are derived in order to distinguish the physics from the system of units, especially since fibers have very disparate geometric scales. Indeed, the physical dimensions of a typical high-power fiber amplifier are 1 mm in diameter (at most) and about 5-100 m in length, but possibly even longer for some Raman amplifiers.

Let l_0 be a generic spatial scaling such that $x = l_0\hat{x}$, where $\hat{x}$ is the non-dimensional spatial variable. For the rest of the chapter, the hat symbol $(\hat{\cdot})$ will indicate a non-dimensional parameter or variable. One can now derive that

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial \hat{x}}\frac{\partial \hat{x}}{\partial x} = \frac{1}{l_0}\frac{\partial}{\partial \hat{x}}.$$

Likewise, consider the dimensionless versions of the electromagnetic fields, and frequency parameter, which can be expressed as

$$\mathbb{E}_l = E_0\hat{\mathbb{E}}_l,\ \mathbb{H}_l = H_0\hat{\mathbb{H}}_l,\ \text{and}\ \omega_l = \omega_0\hat{\omega}_l.$$

In this simulation, the frequencies of the pump and signal fields are both near to $\omega_0 = 10^{15}$ rads/sec, which will be considered as the chosen value for this parameter. The other parameters that have been introduced for the non-dimensionalization of the governing equations will be chosen as follows:

$$l_0 = \frac{c}{\omega_0},\ H_0 = \frac{1}{c}\sqrt{\frac{\omega_0\kappa_a}{\mu_0 g_{\mathrm{R}}}},\ \text{and}\ E_0 = \sqrt{\frac{\mu_0\omega_0\kappa_a}{g_{\mathrm{R}}}}.$$

With these choices, and with the identity $c^2 = (\varepsilon_0\mu_0)^{-1}$, one can determine that

$$\frac{\omega_0\mu_0 H_0 l_0}{E_0} = 1,\ \frac{\omega_0\varepsilon_0 E_0 l_0}{H_0} = 1,\ \ \text{and}\ \frac{c g_{\mathrm{R}} E_0 H_0}{\omega_0} = \kappa_a.$$

In order to augment the Raman gain phenomenon within a short fiber (several tens of wavelengths), we have introduced the artificial scaling parameter κ_a. This non-physical parameter scales the intensity values (through boundary conditions) throughout the fiber, and thereby allows us to simulate very short fiber lengths while allowing for the gain of signal power from the pump field. Now the first-order Maxwell system (5.2.1), with the expressions for the background refractive index (5.2.2) and Raman gain (5.2.11) contributions to the electric polarization, can be non-dimensionalized:

$$\begin{aligned}
&\hat{\nabla}\times\hat{\mathbb{E}}_l = -i\hat{\omega}_l\hat{\mathbb{H}}_l\\
&\hat{\nabla}\times\hat{\mathbb{H}}_l = i\hat{\omega}_l\hat{\mathbb{E}}_l + i(\mathbf{n}_l^2 - \mathbb{I})\hat{\omega}_l\hat{\mathbb{E}}_l + i\hat{\omega}_l\frac{i\mathbf{n}_l\kappa_a\Upsilon_{\mathrm{R}}^l}{\hat{\omega}_l}\big|\mathrm{Real}(\hat{\mathbb{E}}_k\times\hat{\mathbb{H}}_k^*)\big|\hat{\mathbb{E}}_l,
\end{aligned} \tag{5.2.12}$$

which result in:

$$\begin{aligned}
&\hat{\nabla}\times\hat{\mathbb{E}}_l = -i\hat{\omega}_l\hat{\mathbb{H}}_l\\
&\hat{\nabla}\times\hat{\mathbb{H}}_l = i\mathbf{n}_l^2\hat{\omega}_l\hat{\mathbb{E}}_l - \mathbf{n}_l\kappa_a\Upsilon_{\mathrm{R}}^l\big|\mathrm{Real}(\hat{\mathbb{E}}_k\times\hat{\mathbb{H}}_k^*)\big|\hat{\mathbb{E}}_l
\end{aligned} \tag{5.2.13}$$

This represents two nonlinear Maxwell systems, one with $l = \text{s}$ and $k = \text{p}$ and the other with $l = \text{p}$ and $k = \text{s}$, that are coupled together through the Raman gain term.

Boundary conditions

In this model, the light propagates along the z-axis in the fiber core, only in the forward ($+z$) direction. Since the model is formulated as a boundary value problem, it is paramount that the boundary conditions, especially on the output facet of the fiber, correctly capture the physics of the amplifier. The fiber is excited at the input end (corresponding to $z = 0$) with two light sources launched into the fiber core region. Recall that we have introduced the artifical scaling parameter κ_a. The non-dimensionalization relations show that by choosing $0 < \kappa_a < 1$, we artifically increase the field intensities, thereby injecting an increased amount of power within the short fiber at $z = 0$ in order to see sufficient gain in a short distance along the fiber. A zero boundary condition for the electromagnetic fields is set at the outer edge of the inner cladding, which is far enough away from the core region so as to not significantly affect the guided light. Indeed, for guided light, the fields decay exponentially within the cladding, and at the radial boundary, $\sqrt{x^2 + y^2} = r = r_{\text{cladding}}$, the fields are, within numerical precision, zero. Finally, at the exit end of the fiber ($z = L$, where L is the length of the fiber), appropriate out-flowing radiation boundary conditions are set. In order to facilitate this, a PML is introduced at the end of the fiber. The need for this is better understood by observing that

the gain polarization can be thought of as producing an electric conductivity within the material.

Implication of Conductivity on Boundary Condition: How the Raman gain can be viewed in terms of non-zero conductivity σ will now be addressed. We will drop the "hats" while referring to the non-dimensional equations derived earlier. Consider the term $i\omega_l \mathbb{P}_l$. The background part of the polarization behaves linearly:

$$i\omega_l \mathbb{P}_l^{\text{background,}}(\mathbb{E}_l) = i\omega_l(\mathbf{n}_l^2 - \mathbb{I})\mathbb{E}_l.$$

However, the Raman term yields:

$$i\omega_l \mathbb{P}_l^{\text{Raman}}(\mathbb{E}_l) = i\omega_l \frac{i\mathbf{n}_l}{\omega_l} \kappa_a \, \Upsilon_R^l |\text{Real}(\mathbb{E}_k \times \mathbb{H}_k^*)| \mathbb{E}_l = -(\mathbf{n}_l \kappa_a \Upsilon_R^l) |\text{Real}(\mathbb{E}_k \times \mathbb{H}_k^*)| \mathbb{E}_l.$$

The term $(\mathbf{n}_l \kappa_a \Upsilon_R^l)|\text{Real}(\mathbb{E}_k \times \mathbb{H}_k^*)|$ is purely real and hence can be interpreted as a material conductivity, which acts as a nonlinear coupling between the signal and pump fields. The entire amplification properties hinge on this term. Indeed, this nonlinear term is responsible for the power transfer from the pump field into the signal field, since $\Upsilon_R^s = 1$ while $\Upsilon_R^p = -\frac{\omega_p}{\omega_s} < -1$, which implies loss from the pump into the signal. Although the numerical value of the gain is increased significantly, it is a weak nonlinearity, since it does not induce any self-coupling in the signal and pump fields individually.

One implication for DPG implementation is apparent: a simple impedance like boundary condition at the terminal end of the fiber will not suffice. Indeed, impedance boundary conditions for waveguides work on the principle of

a single propagating mode in a lossless linear medium with an exactly known impedance constant, say γ. One then relates the $\mathbb{E}, \mathbb{H}$ fields on a boundary (with normal $\vec{n}$) as:

$$\mathbb{E} + \gamma \vec{n} \times \mathbb{H} = 0.$$

However, since this is a nonlinear problem, where the conductivity changes along the length of the fiber, an exact impedance-like relation between the $\mathbb{E}, \mathbb{H}$ on the terminal boundary is inapplicable, and would correspond to incorrect boundary behaviour. Thus, one must develop a perfectly matched layer (PML) at the exit end of the fiber, which would not hamper the behaviour of the fields within the domain. Towards this end, PMLs have been widely used in finite element implementations. Most notably, [8] and the recent work [81] uses DPG methods to implement ultraweak formulations for various wave propagation phenomenon. In this case, a stretched coordinate PML for the ultraweak formulation is used, with stretching along the z-axis, since outgoing waves need to be attenuated in only the z-direction. It is suggested that the reader refer to [81] and Appendix B for implementation details. Figures 5.2 and 5.3 indicate the use of ultraweak DPG PML for a rectangular waveguide with a pronounced exponentially growing wave. In particular, the imaginary and real parts of an exponentially growing wave terminated with a DPG PML that begins at roughly the middle of the rectangular waveguide are shown in figures 5.2 and 5.3. The entire setup is solved using the ultraweak DPG formulation of the Maxwell system. Notice how the wave attenuates completely after entering the PML region. A similar approach is used for the

fiber geometry.

Given the nature of this simulation, and the computational challenges it entails, the aim is to demonstrate *qualitative* results of the Raman gain action. This goal, for now, requires the use of a sufficiently short fiber so that all calculations can be completed on a regular laptop or workstation in a reasonable amount of time. This is done with the understanding that future efforts will parallelize this model and implement it on a supercomputing platform, where more realistically sized fibers can be studied. Therefore, this simulation sets the fiber length to be less than 0.1 mm ($\sim$50-100 wavelengths), and artificially increases the field intensities (and thereby powers) in (polarization maintaining) silica fibers, by many orders of magnitude in order to absorb significant amounts of the pump field in this short distance, allowing one to observe the Raman process.

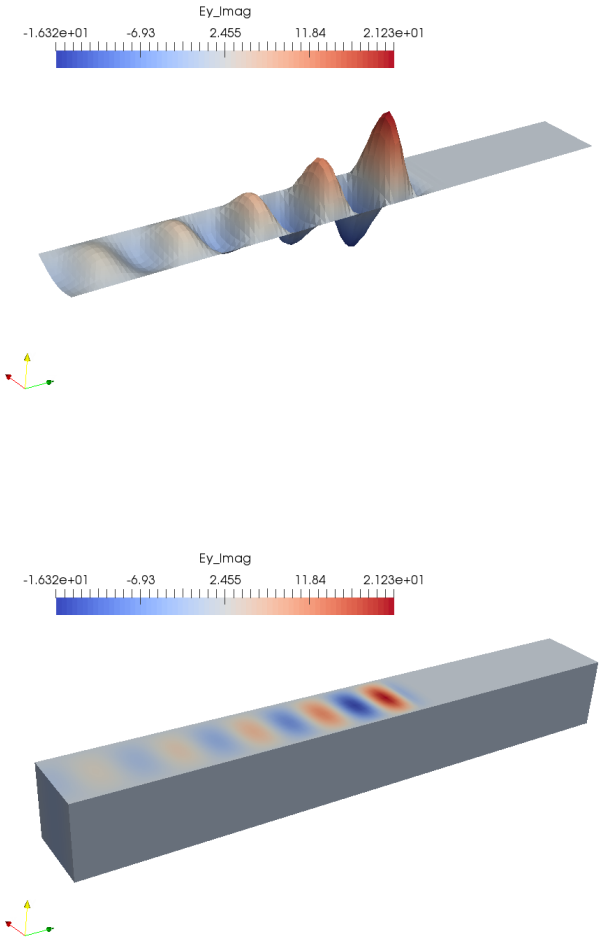

Figure 5.2: Ultraweak DPG PML for growing waves: Imaginary parts of an exponentially growing wave with domain terminated by a DPG PML

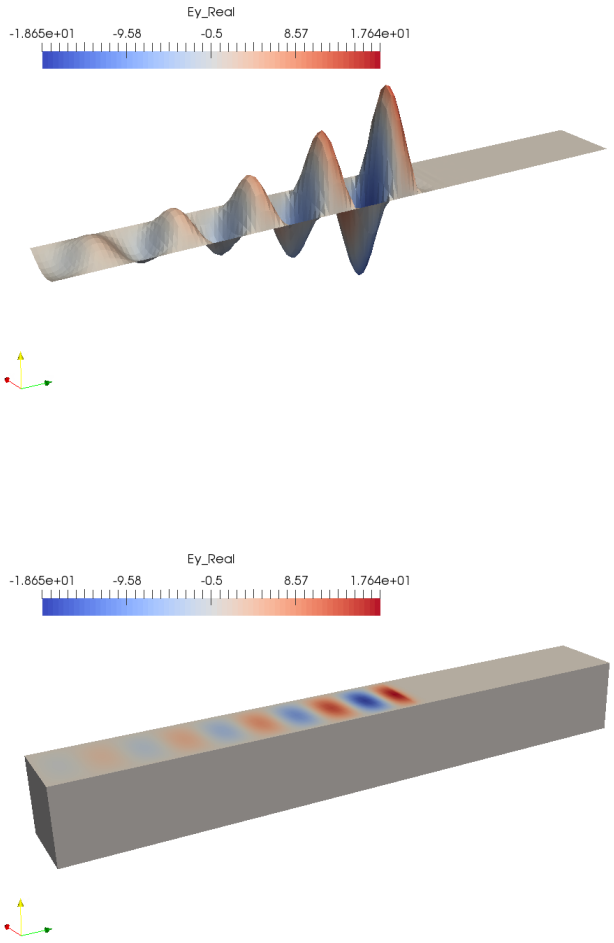

Figure 5.3: Ultraweak DPG PML for growing waves: Real parts of an exponentially growing wave with domain terminated by a DPG PML

Chapter 6

DPG for Raman Gain

This chapter[1] deals with the details of the DPG implementation of the Raman model for fiber laser amplifiers described in the previous chapter.

Author contributions: The contents of this chapter are taken largely from the multi-author article "A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers" S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, J. Mora, in preparation. The article has not yet been submitted for journal publication. The author of this dissertation contributed to model development, and code/numerical implementation of the model and analysis of the results.

We first define the required energy spaces that arise in the discretization of the time-harmonic Maxwell system, and the so-called ultraweak variational formulation, which shall be used for this Raman gain model, is defined. The following section will provide details of the numerical simulations including

[1]The content of this chapter is taken from the manuscript "A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers" S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, J. Mora, in preparation. Information approved for public release on 08 May 2018 by AFRL OPSEC/PA OPS-18-19547.

model parameters, nonlinear iterations for the coupled signal-pump system and optical power calculation. The final section provides the results of our numerical simulations.

6.1 DPG Technology

As we have seen, the DPG technology is a multi-faceted approach to the stable discretization of well-posed variational formulations. In essence, DPG methods (used with optimal test functions) come with several impressive properties: uniform, mesh independent stability, localizable test norms via broken test spaces and a built-in canonical error indicator.

6.1.1 Energy Spaces for Maxwell Equations

Returning to the DPG discretization of time-harmonic Maxwell equations, there are, like other equations of physics [22], four conceivable variational formulations of the Maxwell equations [13]. As described in appendix D, the ultraweak formulation is the formulation of choice in this Maxwell fiber amplifier problem because of its superior properties which are important in wave propagation applications. The remainder of this subsection defines the energy spaces required for the Maxwell system and defines the ultraweak formulation.

Consider a bounded, simply connected Lipschitz domain $\Omega \subset \mathbb{R}^3$ with boundary $\partial\Omega$ and unit normal vector n. The existence of a mesh Ω_h of finitely many open elements K, each with unit normal n_K, such that $\overline{\Omega} \subset \bigcup_{K\in\Omega_h} \overline{K}$

is assumed. Next, define:

$$\begin{array}{ll} L^2(\Omega) & := \{f : \Omega \to \mathbb{R} : \int_\Omega |f|^2 < \infty\}, \\ \mathbb{L}^2(\Omega) & := L^2(\Omega) \times L^2(\Omega) \times L^2(\Omega), \\ H(\text{curl}, \Omega) & := \{\mathbb{E} \in \mathbb{L}^2(\Omega) : \nabla \times \mathbb{E} \in \mathbb{L}^2(\Omega)\}, \\ H_0(\text{curl}, \Omega) & := \{\mathbb{E} \in H(\text{curl}, \Omega) : n \times \mathbb{E}|_{\partial\Omega} = 0\}. \end{array} \tag{6.1.1}$$

The broken counterpart of $H(\text{curl}, \Omega)$ is defined as:

$$H(\text{curl}, \Omega_h) := \{\mathbb{E} \in \mathbb{L}^2(\Omega) : \mathbb{E}|_K \in H(\text{curl}, K), K \in \Omega_h\} = \prod_{K \in \Omega_h} H(\text{curl}, K). \tag{6.1.2}$$

Notice that the broken counterpart of $\mathbb{L}^2(\Omega)$ is itself. The element-wise summed $\mathbb{L}^2(\Omega)$ inner product of the two arguments is denoted by $(\cdot, \cdot)_h$, and the element-wise summed duality pairing of appropriate dual spaces is represented by $\langle \cdot, \cdot \rangle_h$. The symbol $\| \cdot \|$ shall mean the $\mathbb{L}^2(\Omega)$ norm. As was shown in [13], the definition of trace operators are required in order to elegantly define the DPG interface spaces. First, define element trace operators:

$$\begin{array}{ll} t_{K,\top}(\mathbb{E}) & := (n_K \times \mathbb{E}) \times n_K|_{\partial K} \\ t_{K,\perp}(\mathbb{E}) & := (n_K \times \mathbb{E})|_{\partial K} \end{array} \tag{6.1.3}$$

Notice that these trace operators have range $H^{-1/2}(\text{curl}, \partial K)$ and $H^{-1/2}(\text{div}, \partial K)$ respectively, i.e.,

$$\begin{array}{ll} t_{K,\top} & : H(\text{curl}, K) \to H^{-1/2}(\text{curl}, \partial K), \\ t_{K,\perp} & : H(\text{curl}, K) \to H^{-1/2}(\text{div}, \partial K). \end{array} \tag{6.1.4}$$

Finally, the trace operators on the full broken $H(\text{curl}, \Omega_h)$ space are defined via the element-wise application of the element trace operators:

$$\begin{array}{ll} T_\top & : H(\text{curl}, \Omega_h) \to \prod_{K \in \Omega_h} H^{-1/2}(\text{curl}, \partial K), \\ T_\perp & : H(\text{curl}, \Omega_h) \to \prod_{K \in \Omega_h} H^{-1/2}(\text{div}, \partial K). \end{array} \tag{6.1.5}$$

The operators $T_\top, T_\perp$ are linear by construction. Finally, the spaces of interface variables (or interface spaces) can be defined as the images under the trace maps of the conforming $H(\text{curl}, \Omega)$ space:

$$\begin{array}{ll} H^{-1/2}(\text{div}, \partial\Omega_h) & := T_\top(H(\text{curl}, \Omega)) \\ H^{-1/2}(\text{curl}, \partial\Omega_h) & := T_\perp(H(\text{curl}, \Omega)). \end{array} \tag{6.1.6}$$

As shown in [13], the trace (quotient) norms on the two interface spaces are dual to each other.

Ultraweak Variational Formulation

The Maxwell operator is defined as

$$A \begin{pmatrix} \mathbb{E} \\ \mathbb{H} \end{pmatrix} := \begin{pmatrix} -(i\omega\epsilon + \sigma) & \nabla\times \\ -\nabla\times & -i\omega\mu \end{pmatrix} \begin{pmatrix} \mathbb{E} \\ \mathbb{H} \end{pmatrix}$$

with adjoint A^*:

$$A^* \begin{pmatrix} \mathbb{E} \\ \mathbb{H} \end{pmatrix} := \begin{pmatrix} (i\omega\epsilon - \sigma) & -\nabla\times \\ \nabla\times & i\omega\mu \end{pmatrix} \begin{pmatrix} \mathbb{E} \\ \mathbb{H} \end{pmatrix}$$

If $\sigma = 0$, then $A^* = -A$. The ultraweak formulation (see also appendix D) corresponds to the case where

$$X_0 = \mathbb{L}^2(\Omega) \times \mathbb{L}^2(\Omega), \hat{X} = H^{-1/2}(\text{curl}, \partial\Omega_h) \times H^{-1/2}(\text{curl}, \partial\Omega_h),$$

$$Y_0 = H(\text{curl}, \Omega) \times H_0(\text{curl}, \Omega), Y = H(\text{curl}, \Omega_h) \times H(\text{curl}, \Omega_h).$$

Denote by $u = (\mathbb{E}, \mathbb{H}) \in X_0$, $\hat{u} = (\hat{\mathbb{E}}, \hat{\mathbb{H}}) \in \hat{X}$ and $v = (\mathbb{R}, \mathbb{S}) \in Y$. The bilinear forms corresponding to the ultraweak formulation are:

$$b_0(u, v) = (u, A^* v)_h,$$

$$\hat{b}(\hat{u}, v) = \langle n \times \hat{\mathbb{E}}, \mathbb{R} \rangle_h + \langle n \times \hat{\mathbb{H}}, \mathbb{S} \rangle_h.$$

The ultraweak formulation comes equipped with the (scaled) adjoint graph norm:

$$\|v\|_Y^2 := \alpha \|v\|^2 + \|A^* v\|^2.$$

Modification of the true adjoint graph norm (consisting of only $\|A^* v\|^2$ term) by adding the above $\alpha \|v\|^2$ scaling term is required to make the norm localizable [26, 13]. We use $\alpha = 1$ in the non-dimensional setting. Next, the simulation of the Raman gain problem using the ultraweak DPG discretization will be considered.

6.2 Setup of Simulations

Having established the superiority of the ultraweak formulation in appendix D, this formulation will be used for the remainder of this work. For notational convenience, the "hats" that denoted the non-dimensional quantities derived in Section 5.2 shall be omitted.

Recall that for the Raman gain problem, we are interested in solving for

$$\begin{aligned} \nabla \times \mathbb{E}_l &= -i\omega_l \, \mathbb{H}_l, \\ \nabla \times \mathbb{H}_l &= i\omega_l \, \mathbb{E}_l + i\omega_l \mathbb{P}_l, \end{aligned} \tag{6.2.1}$$

where, $\mathbb{E}_l, \mathbb{H}_l$ are the electric and magnetic fields corresponding to the signal and pump frequencies ω_l with $l = s, p$.

Moreover, the polarization vector decomposes as:

$$\mathbb{P}_l(\mathbb{E}_l) = \mathbb{P}_{\text{background},\, l}(\mathbb{E}_l) + \mathbb{P}_{\text{Raman},\, l}(\mathbb{E}_l), \tag{6.2.2}$$

where:

$$\begin{aligned} \mathbb{P}_{\text{background},\, l}(\mathbb{E}_l) &= (\mathbf{n}^2 - I)\mathbb{E}_l, \\ \mathbb{P}_{\text{Raman},\, l}(\mathbb{E}_l) &= \tfrac{i\mathbf{n}}{\omega_l}(\Upsilon_l^R\, g_l^R)\mathbb{E}_l. \end{aligned} \tag{6.2.3}$$

Here,

$$\Upsilon_l^R = \begin{cases} \frac{-\omega_p}{\omega_s}, & \text{if } l = p \\ 1, & l = s, \end{cases}$$

and

$$g_l^R(\mathbb{E}_l) = \begin{cases} |\text{Real}(\mathbb{E}_p \times \mathbb{H}_p^*)| & \text{if } l = s, \\ |\text{Real}(\mathbb{E}_s \times \mathbb{H}_s^*)| & \text{if } l = p. \end{cases}$$

6.2.1 Model Implementation

For the Raman fiber amplifier simulations, shape functions developed in [40] are used, which support 3D elements of all shapes (hexahedron, prism, tetrahedron and pyramid). The coding for this problem was done in the $hp3D$ infrastructure detailed in the book [28]. As is noted in [55], the DPG method can be implemented in any standard finite element code supporting the exact sequence energy spaces. An all-hexahedron mesh is used in order to take advantage of the fast quadrature developed in [62]. The space of polynomials of order p are denoted as $\mathcal{P}^p$, with $\mathcal{Q}^{(p,q,r)} := \mathcal{P}^p \otimes \mathcal{P}^q \otimes \mathcal{P}^r$ and

$$\begin{aligned} \mathcal{W}_p &:= \mathcal{Q}^{(p,q,r)}, \\ \mathcal{Q}_p &:= \mathcal{Q}^{(p-1,q,r)} \times \mathcal{Q}^{(p,q-1,r)} \times \mathcal{Q}^{(p,q,r-1)}, \\ \mathcal{V}_p &:= \mathcal{Q}^{(p,q-1,r-1)} \times \mathcal{Q}^{(p-1,q,r-1)} \times \mathcal{Q}^{(p-1,q-1,r)}, \\ \mathcal{Y}_p &:= \mathcal{Q}^{(p-1,q-1,r-1)}. \end{aligned} \tag{6.2.4}$$

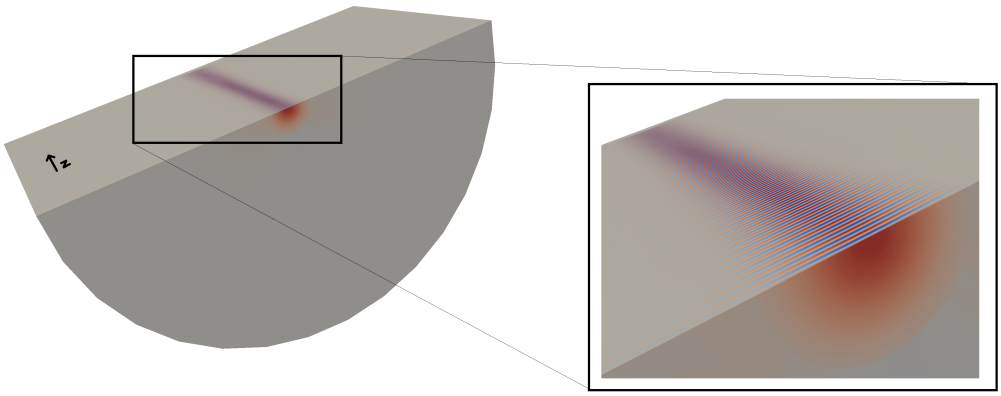

Figure 6.1: Cross sectional view of an example of our simulation described in this chapter with ≈ 80 wavelengths and all hexahedron curvilinear geometry. The core is discretized with 5 hexahedral elements while the cladding has 4 hexahedral elements in the initial mesh. The zoomed part shows the a close-up of the core region of the fiber. Here, the core radius is roughly one tenth the radius of the cladding.

The Maxwell system utilizes the Nedelec hexahedron of first type characterized by the exact sequence [28]:

$$\mathbb{R} \xrightarrow{\text{id}} \mathcal{W}_p \xrightarrow{\nabla} \mathcal{Q}_p \xrightarrow{\nabla\times} \mathcal{V}_p \xrightarrow{\nabla\cdot} \mathcal{Y}_p \longrightarrow 0 \ .$$

In the process of coding the Raman problem within $hp3D$, separate data structures for both signal and pump variables are supported, but the memory is allocated for each solve separately. This is possible due to the weak coupling between the two sets of fields through the Raman gain. Thus, while solving for the signal fields, memory is allocated only for the signal, and likewise while solving for the pump. The solvers used in this work come from the MUMPS (MUltifrontal Massively Parallel sparse direct Solver, see at http://mumps.enseeiht.fr/) library and the Intel MKL Pardiso solver.

6.2.2 Model Parameters

This test problem of a core-pumped, step-index Raman amplifier sets the non-dimensional core and inner cladding radii to $\hat{r}_{\text{core}} = 0.25\sqrt{2}$ and $\hat{r}_{\text{cladding}} = 2.5\sqrt{2}$. The cladding refractive index is set to $n_{\text{cladding}} = 1.45$. Using a numerical aperture of NA ≈ 0.0659, and knowing that $\text{NA} = \sqrt{n_{\text{core}}^2 - n_{\text{cladding}}^2}$, the core refractive index can be calculated to be $n_{\text{core}} \approx 1.4515$. This means that the normalized frequency (or V-number) of the fiber is

$$\text{V} = \frac{2\pi r_{\text{core}}}{\lambda}\text{NA} \approx 2.198,$$

given a signal wavelength in air of $\lambda = \lambda_{\text{s}} = 1.116\ \mu$m. Note that the V-number can also be expressed in terms of non-dimensional quantities as:

$$\text{V} = \frac{\omega}{c} r_{\text{core}}\text{NA} = \frac{l_0\omega_0}{c}\hat{\omega}\hat{r}_{\text{core}}\text{NA} = \hat{\omega}\hat{r}_{\text{core}}\text{NA},$$

where $\hat{\omega}$ is the non-dimensional frequency and $\hat{r}_{\text{core}}$ is the non-dimensional core radius. In our simulations, we use $\hat{\omega} = 30\pi$ and $\hat{r}_{\text{core}} = 0.25\sqrt{2} \approx 0.3536$, so that V ≈ 2.198. Because $V < 2.405$, the fiber is robustly single-mode. The pump wavelength in air is $\lambda_{\text{p}} = 1.064\ \mu$m.

6.2.3 Iterative Solve For the Nonlinearity

How the nonlinear problem is solved is addressed here. It is sufficient to resort to a simple iteration scheme, where the signal and pump system is solved, and then the gain is updated, and the entire system is solved again in an iterative fashion as shown in the following algorithm: Here, $u_{l,n} = (\mathbb{E}_{l,n}, \mathbb{H}_{l,n})$

Algorithm 1 Simple Iterations

procedure

$u_{l,0} = 0$, $l = s, p$

$\Delta = \Delta_0 = 1, n = 0$

do while ($\Delta >$ tol):

Solve for $u_{s,n+1}$.

Update $g_R^{p,n+1}$.

Solve for $u_{p,n+1}$.

Update $g_R^{s,n+1}$.

$\Delta = \Delta_{n+1} = \frac{\|u_{s,n+1} - u_{s,n}\|}{\|u_{s,n}\|}$

enddo

is defined to be the electromagnetic field solutions of signal/pump ($l = s, p$) at iteration n and $g_R^{l,n} = \mathbf{n}_l \Upsilon_{\mathrm{R}}^l I_k$ the corresponding gain. This process is repeated until convergence. It is worth pointing out that at each nonlinear step, a *new* (scaled) adjoint graph (test) norm is computed, which carries within it the gain contributions from the previous step:

$$\|v_n\|_{Y_n}^2 := \|v_n\|^2 + \|A_n^* v_n\|^2, \tag{6.2.5}$$

where

$$A_{n+1} \begin{pmatrix} \mathbb{E} \\ \mathbb{H} \end{pmatrix} := \begin{pmatrix} -(i\omega + \mathbb{P}_n) & \nabla\times \\ -\nabla\times & -i\omega \end{pmatrix} \begin{pmatrix} \mathbb{E} \\ \mathbb{H} \end{pmatrix},$$

and $\mathbb{P}_n$ is the electric polarization from the previous step. Thus, this methodology assures that the optimality properties of the ultraweak formulation are carried over at each iteration. In other words, at each step n, the current system of linear problems is guaranteed to be optimal. Note that by updating the test norm between each iteration, the test space is also effectively redefined

between iterations. In other words, at step n, the test space Y_n is defined by the norm 6.2.3, and the embedding $Y_n \hookrightarrow L^2(\Omega)$ is tacitly assumed for all n.

6.2.4 Optical Power Calculation

The overall quantity of interest is the cross-sectional power through the fiber at any given z-value along the length of the fiber, but especially at the end of the fiber ($z = L$). Indeed, the existence of gain can be seen through the fact that energy is transferred from pump wavelength to the signal wavelength. Towards this end, one should note that the time-averaged power is computed using the (complex) Poynting vector. The (mean-squared) complex Poynting vector is defined as:

$$\mathbb{S} := \mathbb{E} \times \mathbb{H}^*,$$

where the real part $\mathbb{S}_r = \text{Real}\{\mathbb{S}\}$ is the quantity of interest. Let $z = z_0$ be a position along the fiber and $\vec{n}$ be the corresponding normal vector to the cross-sectional face of the fiber at $z = z_0$. Given that most of the power in the fiber flows in the forward direction, the net power flowing in the direction determined by $\vec{n}$ through a cross-section $z = z_0$ of the fiber is computed as:

$$\mathcal{P} := \left| \int_{z=z_0} \vec{n} \cdot \mathbb{S}_r \, dS \right|.$$

In order to make rigorous mathematical sense of this term in the context of energy spaces, notice that for any domain V, we have:

$$\int_V (\nabla \times \mathbb{E}, \mathbb{H}^*) = \int_V (\mathbb{E}, \nabla \times \mathbb{H}^*) + \int_{\partial V} \langle \vec{n} \times \mathbb{E}, \mathbb{H}^* \rangle,$$

however, $(\vec{n} \times \mathbb{E}) \cdot \mathbb{H}^* = \vec{n} \cdot \mathbb{E} \times \mathbb{H}^*$, so that:

$$\mathcal{P} \leq \|\mathbb{E}\|_{\mathrm{H(curl)}} \|\mathbb{H}\|_{\mathrm{H(curl)}}.$$

Thus,

$$\vec{n} \cdot \mathbb{S}_r = \mathrm{Real}\{\vec{n} \cdot \mathbb{E} \times \mathbb{H}^*\} = \mathrm{Real}\{(\vec{n} \times \mathbb{E}) \cdot \mathbb{H}^*\},$$

and the last term $(\vec{n} \times \mathbb{E}) \cdot \mathbb{H}^*$ can be viewed (on the surface $z = z_0$) as a duality pairing between $H^{-1/2}(\mathrm{div}, \partial\Omega_h) \times H^{-1/2}(\mathrm{curl}, \partial\Omega_h)$. Since the UW formulation of DPG has trace variables coming from the trace spaces $H^{-1/2}(\mathrm{div}, \partial\Omega_h) \times H^{-1/2}(\mathrm{curl}, \partial\Omega_h)$, we are able to compute the power without resorting to any post-processing. Thus, the equation for $\mathcal{P}$, viewed in light of the duality pairing, has a rigorous definition.

6.3 Results

The simulation results are obtained on two workstations with 256 GB memory and 24-28 cores. Initial numerical experiments on a rectangular waveguide indicated that implementing DPG with a polynomial order $p = 5$, and with 4 elements per wavelength (anisotropically), was able to resolve the propagating wave. Those parameters were set likewise in the fiber amplifier, and we used $p = 5$ elements for the fiber amplifier, though we needed a few more elements per wavelength for the longer fibers used in the simulations.

6.3.1 Code Verification

The first verification of the model consists of performing uniform h-convergence studies on the cylindrical core geometry comprised of curvilinear hexahedral elements. The initial mesh consists of five curvilinear hexahedra in the fiber core region. This test uses a manufactured solution of $\mathbb{E} = \sin(\omega x)\,\sin(\omega y)\,\sin(\omega z)\hat{\mathbf{e}}_x$, allowing one to find the analytical expression for the load term that is needed to produce this solution. Figure 6.2 depicts the expected convergence rates for both the signal (excited with $\omega = 1.001$) and pump (excited with $\omega = 1.05$) relative error for polynomial order $p = 1, \ldots, 5$, which theory predicts to be $-\frac{p}{3}$.

6.3.2 Linear Problem

The next verification test studies the linear case, which corresponds to setting $\mathbb{P}_l^{\text{Raman}}(\mathbb{E}_l) = 0$; in other words, this considers a simple, lossless fiber waveguide problem. In this case, only one frequency of light is needed, the signal field, denoted by $\mathbb{E}_{\text{s}} = \mathbb{E}$. Given that the fiber is single-mode, one ought to expect to observe the propagation of only the fundamental mode (called the LP_{01} mode in scalar models) in the x-component of the $\mathbb{E}$ field and in the y-component of the $\mathbb{H}$ field. Moreover, even when the light is only launched into the E_x component, after some distance into the fiber, one ought to expect that all of the components of the electromagnetic field acquire a non-zero value. This occurs because the light does not have to propagate perfectly in the z-direction, but instead is guided by total internal reflection

in the core, allowing it to spread out to a small maximum angle off of the z-axis, which is controlled by the numerical aperture of the fiber. Only in a full vectorial model, with both electric and magnetic field components, could this phenomenon be observed.

The output images of Figs. 6.3-6.14 show that a Gaussian-shaped fundamental mode does propagate through the fiber, and that all of the field components are non-zero; though E_x and H_y have the largest magnitudes, as expected, since the light is launched only into E_x at $z = 0$ and there are no other polarization coupling factors that would cause energy transfer between the electromagnetic field components. Along with the output plots of the real and imaginary parts of each field component, the real and imaginary parts of the cross-sectional view of the fiber parallel to the z-axis are also displayed.

Another important check associated with this test is to ensure that the light does not lose power (energy) along the length of the fiber. By computing the power at various positions along the fiber (no less than one wavelength apart), the two plots of Fig. 6.15 are created. The start of the PML is indicated by a vertical line. These plots show that the power is conserved as the light propagates, and, as would be expected, the solve time for the numerical model increases linearly with fiber length since. Indeed, since the mesh refinements are performed anisotropically, only in the z-direction, the number of elements grows linearly, as does the cost of element computations. Second, the multifrontal solver has linear complexity, and the overall time grows linearly. We note that the times reported are average times for the linear solve over

many runs. Specifically, fiber lengths of $L \approx 8, 16, 32, 64, 80$ wavelengths are used in the test.

Since these are fibers of ultra-short lengths, elementary ray optics arguments set an upper bound in terms of the number of wavelengths required for the launched signal energy to settle into the physically correct solution of the waveguide. This can be roughly estimated to be $\frac{r_{\text{core}}}{\tan(\text{NA})} \approx \frac{r_{\text{core}}}{(\text{NA})}$, which is $\sim$100-250 wavelengths for typical fibers. However, this example seems to exhibit its physically relevant solution within 10 wavelengths.

6.3.3 Gain Problem

The final validation of the model includes the Raman gain action along the fiber. This requires that there is both a pump wavelength and a signal wavelength, which are separated in frequency space from one another by -13.2 THz, corresponding to the peak Raman gain in fused silica glass. As discussed previously, the nonlinearity of the gain is handled by simple iterations. Plot of Fig. 6.16 illustrates the convergence of these iterations for different values of the artificial scaling κ_a. For comparison purposes, note that this test must track both frequencies of light; thus doubling the dataset size of the dependent variables. Again runs are completed for fibers of lengths: $L \approx 8, 16, 32, 64, 80$ wavelengths. Note also that in plots (both linear and nonlinear case) with title/legend indicating number of wavelengths, we mean an approximate number of wavelengths.

6.3.4 Co-pumped and Counter Pumped Configurations

As with all optical nonlinearities, Raman gain per unit length increases with the intensity (irradiance) of the optical fields present in the fiber (see the coupled irradiance PDEs (5.2.10) or recall the gain function for Raman scattering: $g_l = \Upsilon^l_{\mathrm{R}} g_{\mathrm{R}} I_k$). Choosing a core-pumped amplifier, rather than a cladding-pumped amplifier configuration, ensures that the pump optical field is of a higher intensity than if it was spread out through the inner cladding and core of the fiber. Such a core-pumped amplifier can be configured in at least two different ways: co-pumped and counter pumped configurations. Figure 6.19 illustrates these two configurations. In the co-pumped configuration, both the signal and pump fields are injected at the same entrance end of the fiber, whereas in the counter pumped configuration, they are injected at opposing ends of the fiber. Our model can be easily adapted to handle both these configurations. In the co-pumped case, we have boundary conditions for both signal and pump at the entrance end of the fiber and PMLs for both at the exit end, whereas in the counter pumped case, we have boundary data for the signal field at the entrance end and for the pump field at the exit end. Likewise, we implement two different PML configurations: one at the exit end for the signal and another for the pump at the entrance end. We note that the counter pumped configuration is beyond the scope of most traditional scalar BPM models. The first plot of Fig. 6.17 depicts the pump field transferring energy to the signal field along the fiber length, as is expected from a Raman amplifier. This is for a co-pumped configuration, where both signal and pump

are injected at the same fiber end ($z = 0$). The plot of Fig. 6.18 shows a counter-pumped configuration where the signal light is injected at $z = 0$, while the pump is introduced at $z = L$, the end of the fiber. Note that such a configuration entails the use of separate a PML for the signal and pump fields at opposite ends of the fiber. We add that such a configuration cannot be so easily modeled by a scalar BPM approach. The plots of Fig. 6.20- 6.43 are cross-sectional views of the fiber with cross sections normal to the z and y axis.

The qualitative results obtained through this novel 3D vectorial DPG fiber amplifier model provide the validations needed to conclude that the methodology and implementation are sufficient for studying simple fiber amplifier configurations, and provide confidence that future efforts may prove successful in studying more complicated fiber designs under more realistic high-power operation conditions, as is the ultimate goal of this project.

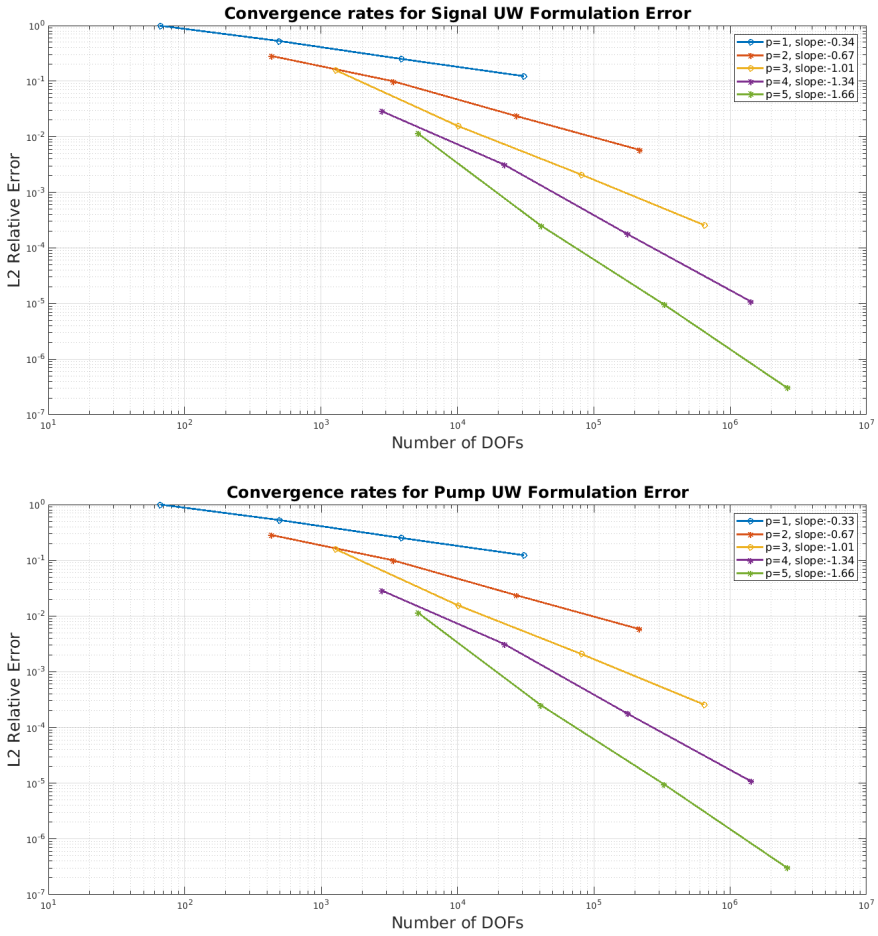

Figure 6.2: Uniform h-convergence rates for manufactured solution

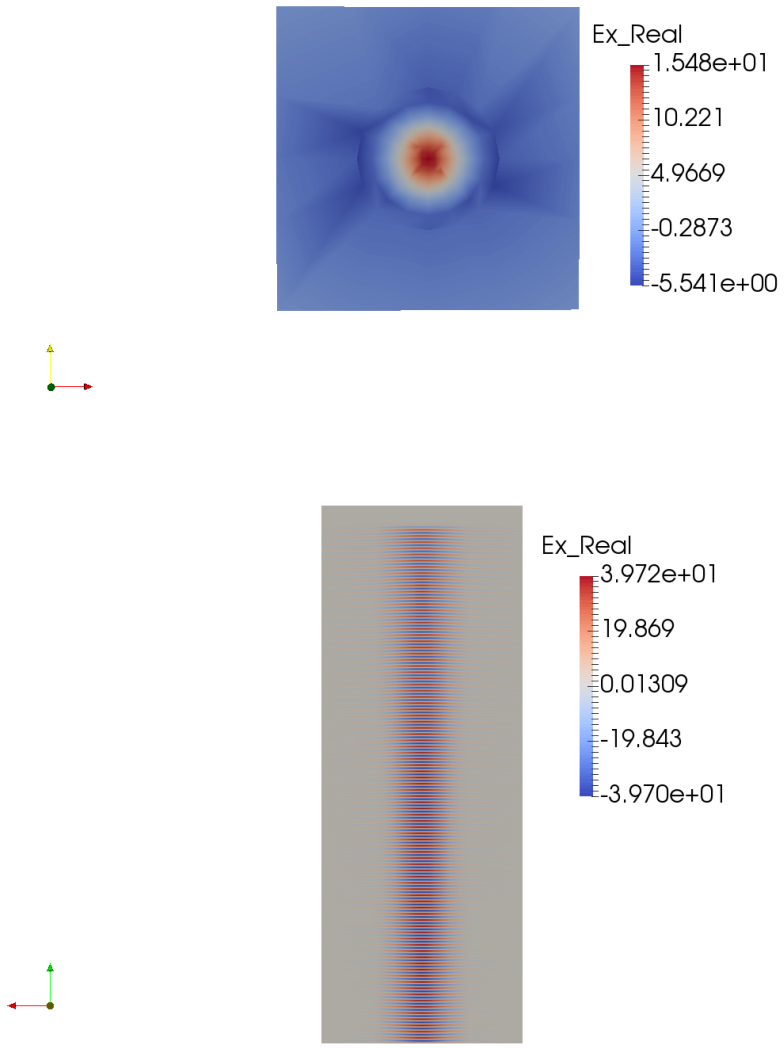

Figure 6.3: Linear Problem: Real Part of E_x

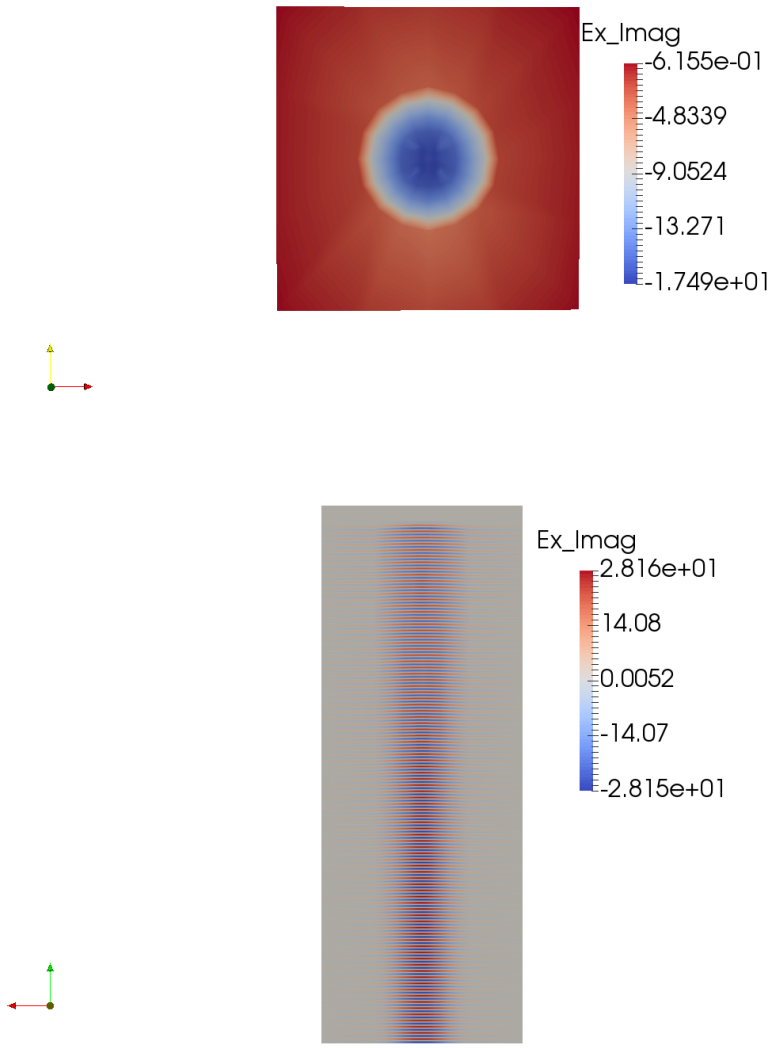

Figure 6.4: Linear Problem: Imaginary Part of E_x

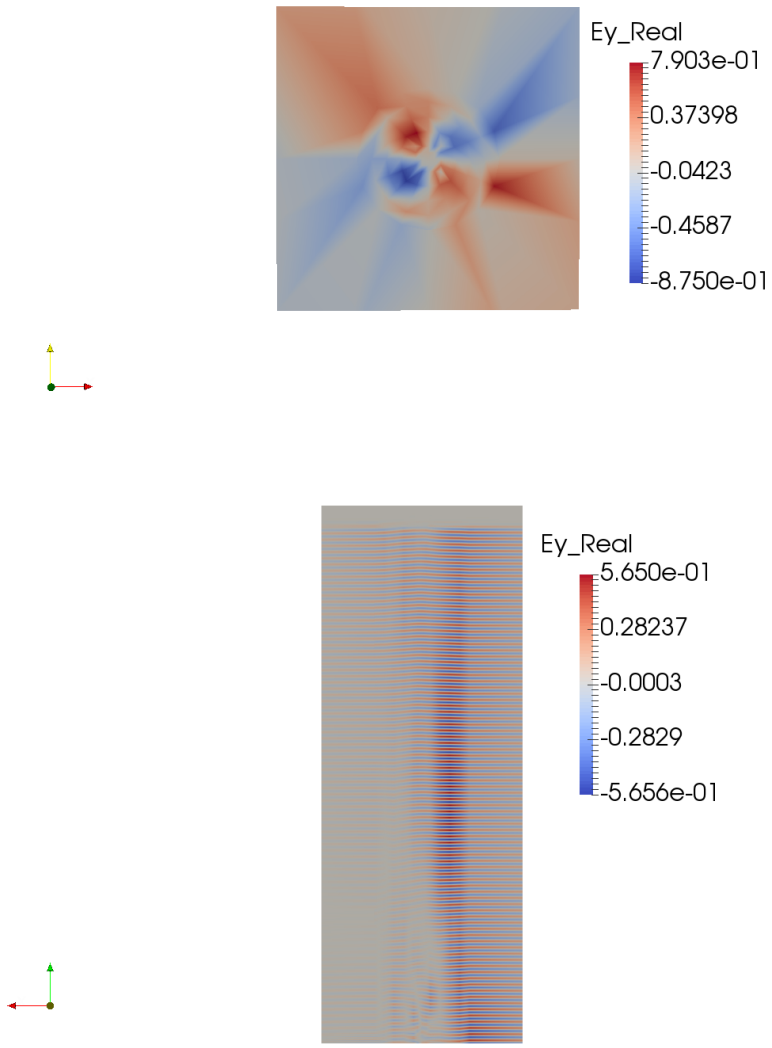

Figure 6.5: Linear Problem: Real Part of E_y

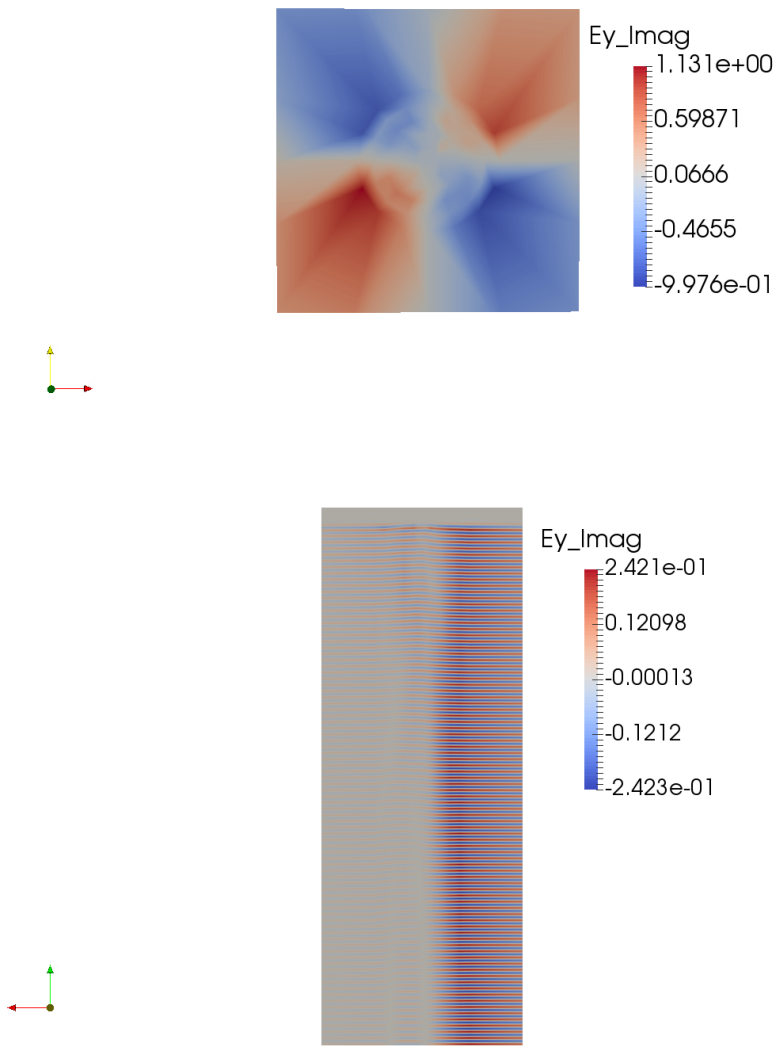

Figure 6.6: Linear Problem: Imaginary Part of E_y

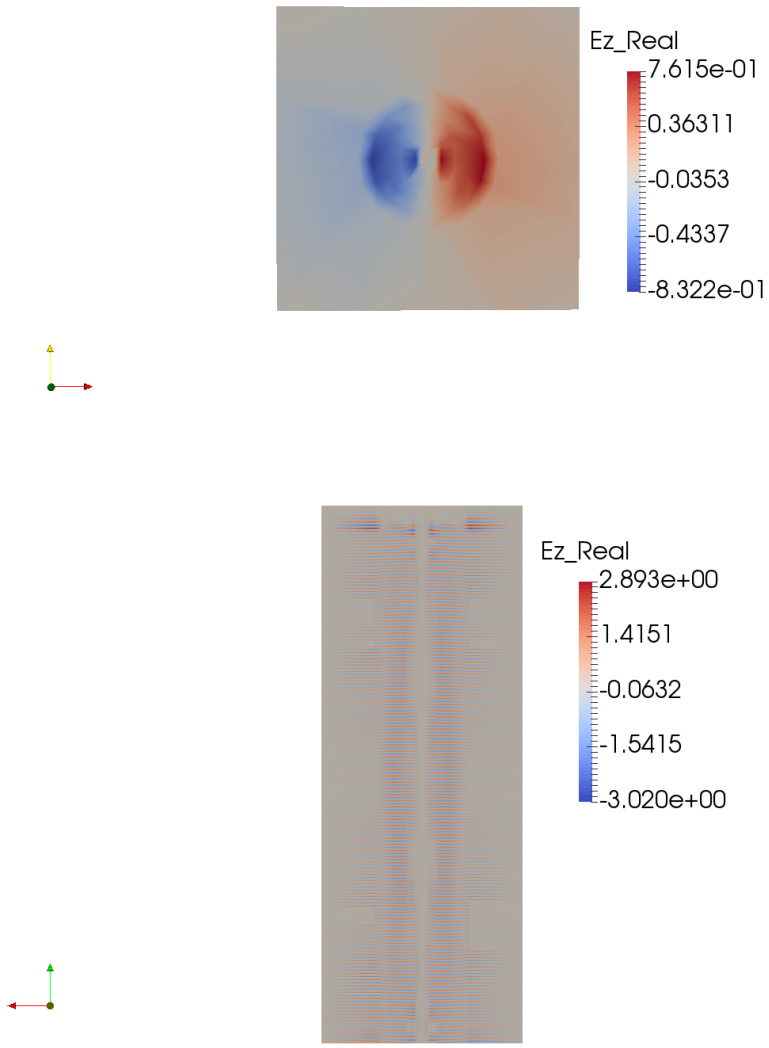

Figure 6.7: Linear Problem: Real Part of E_z

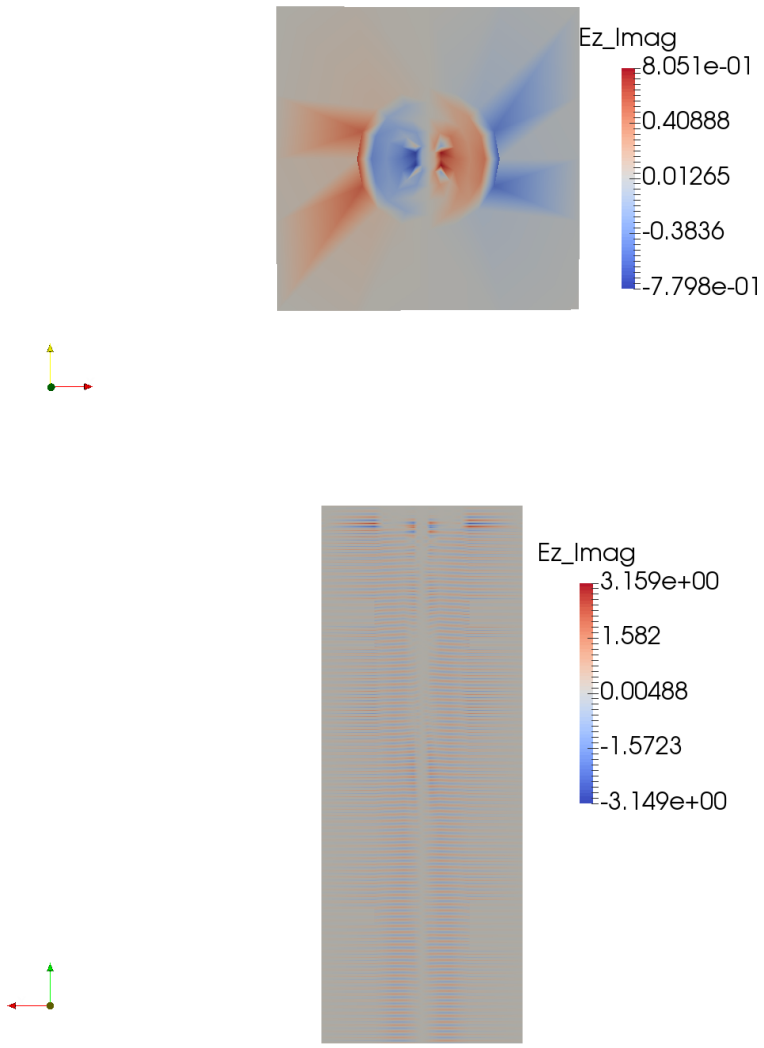

Figure 6.8: Linear Problem: Imaginary Part of E_z

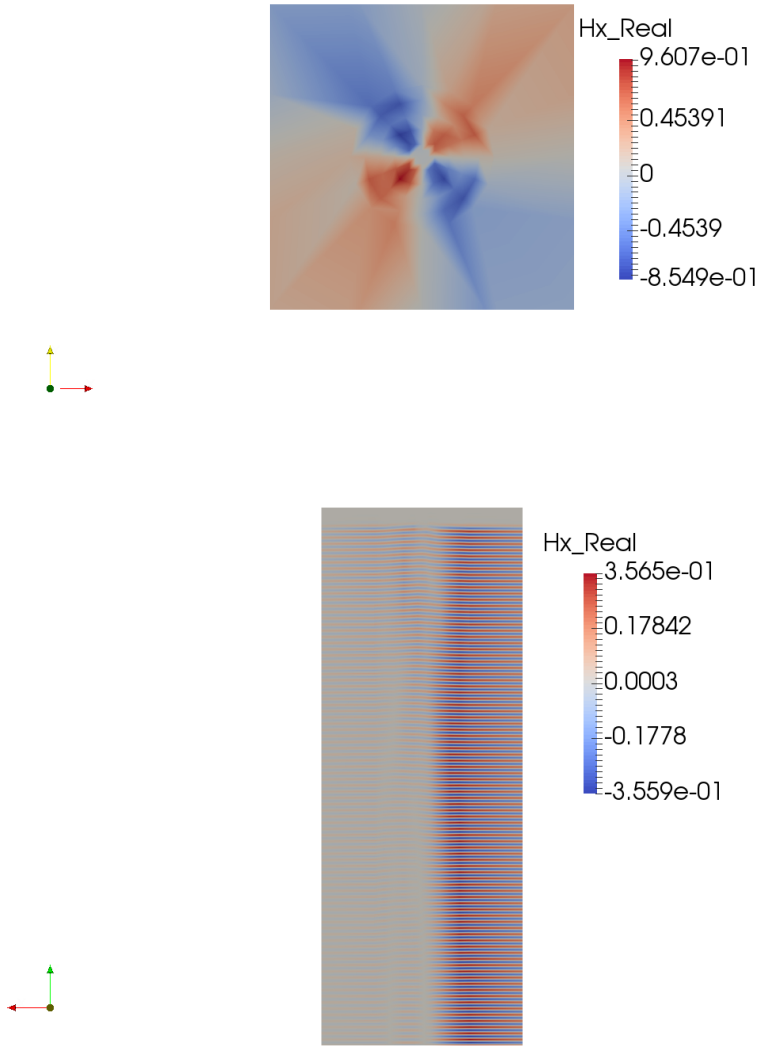

Figure 6.9: Linear Problem: Real Part of H_x

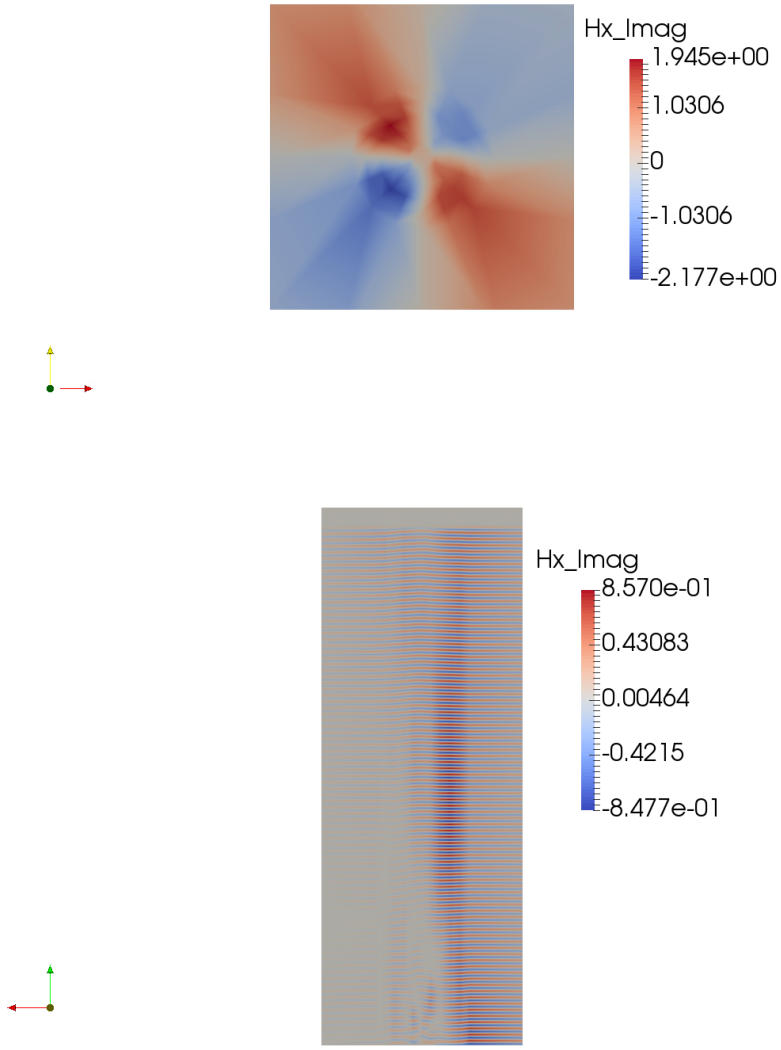

Figure 6.10: Linear Problem: Imaginary Part of H_x

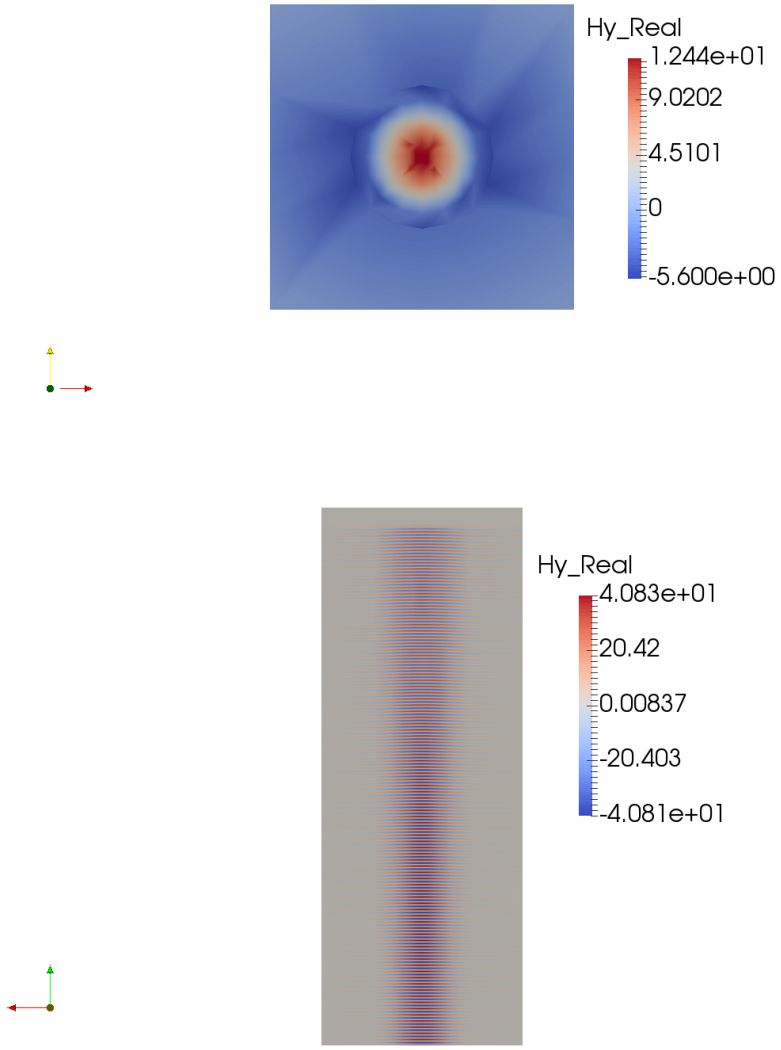

Figure 6.11: Linear Problem: Real Part of H_y

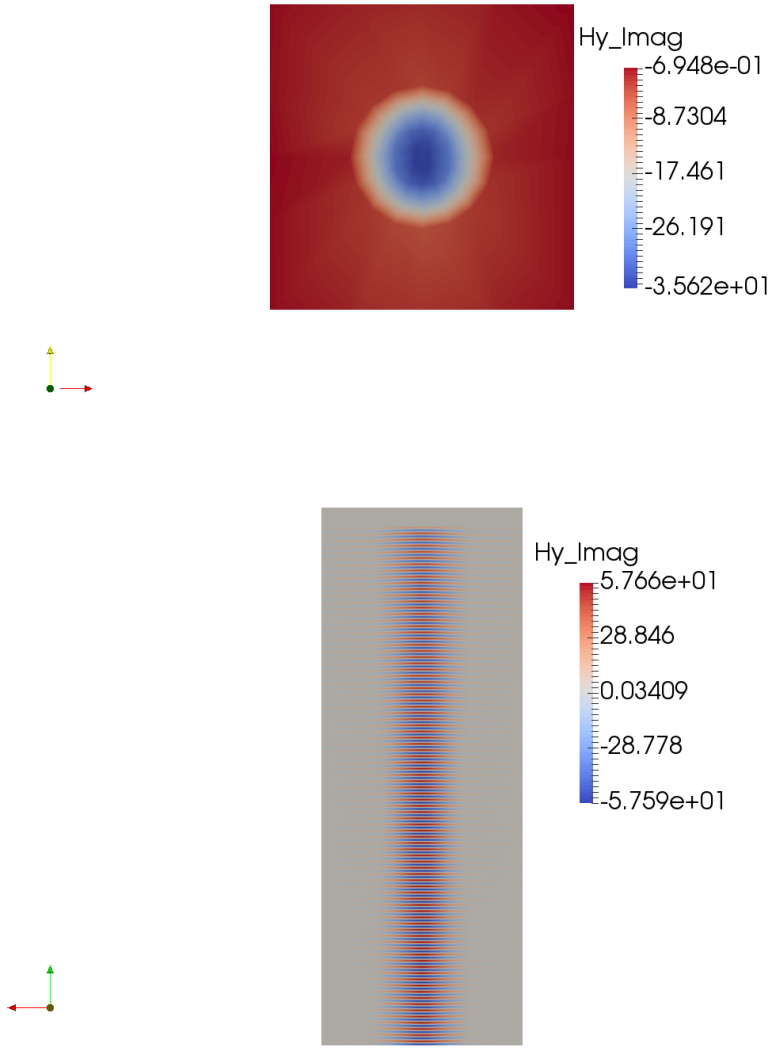

Figure 6.12: Linear Problem: Real and Imaginary Parts of H_y

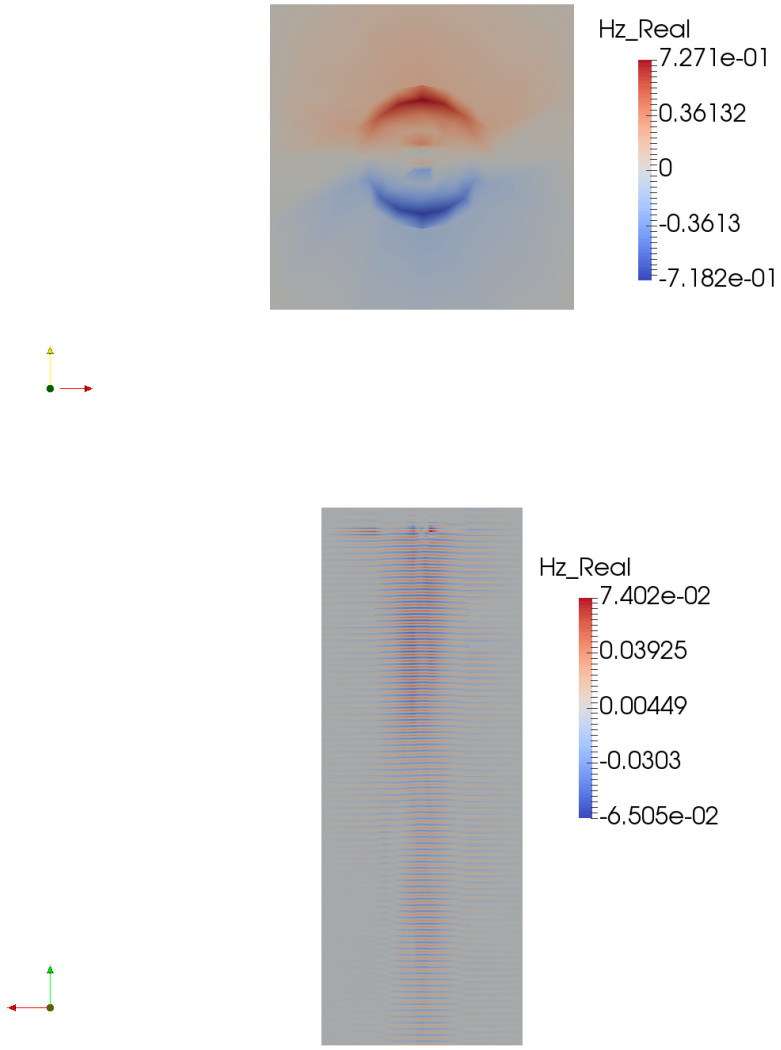

Figure 6.13: Linear Problem: Real Part of H_z

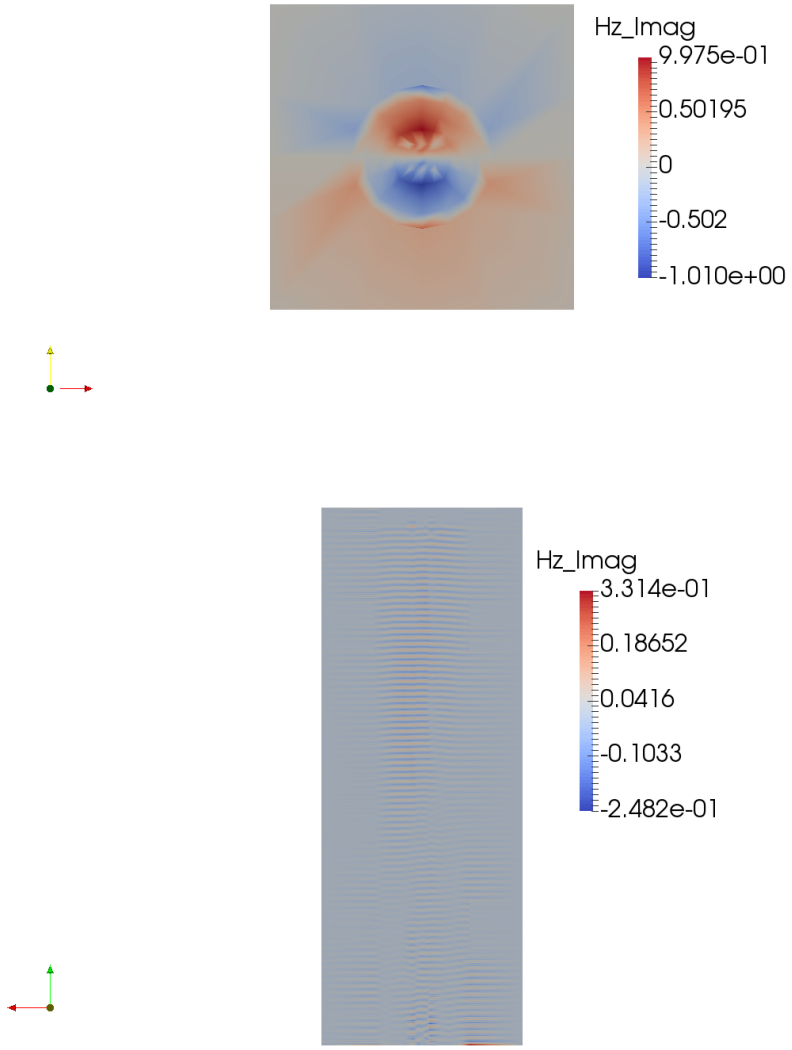

Figure 6.14: Linear Problem: Imaginary Parts of H_z

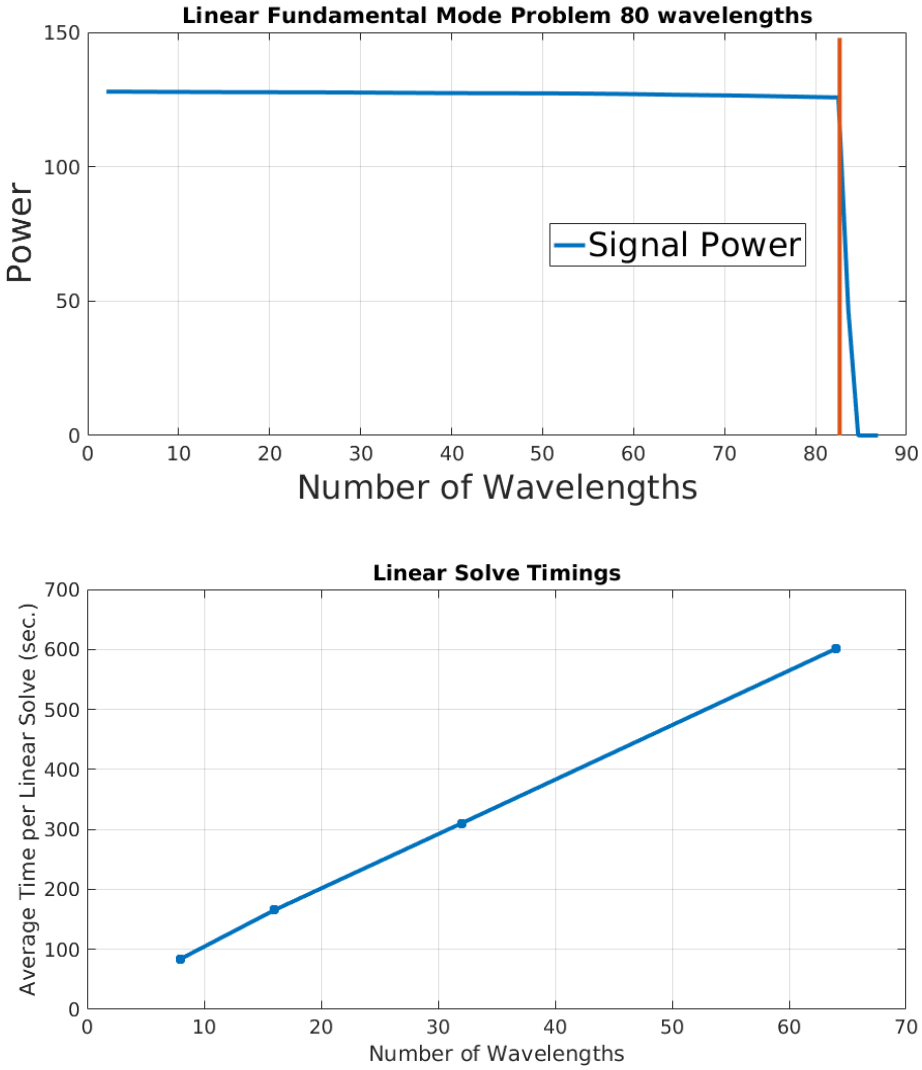

Figure 6.15: Conservation of power (top) and computational solve times for the linear model (bottom).

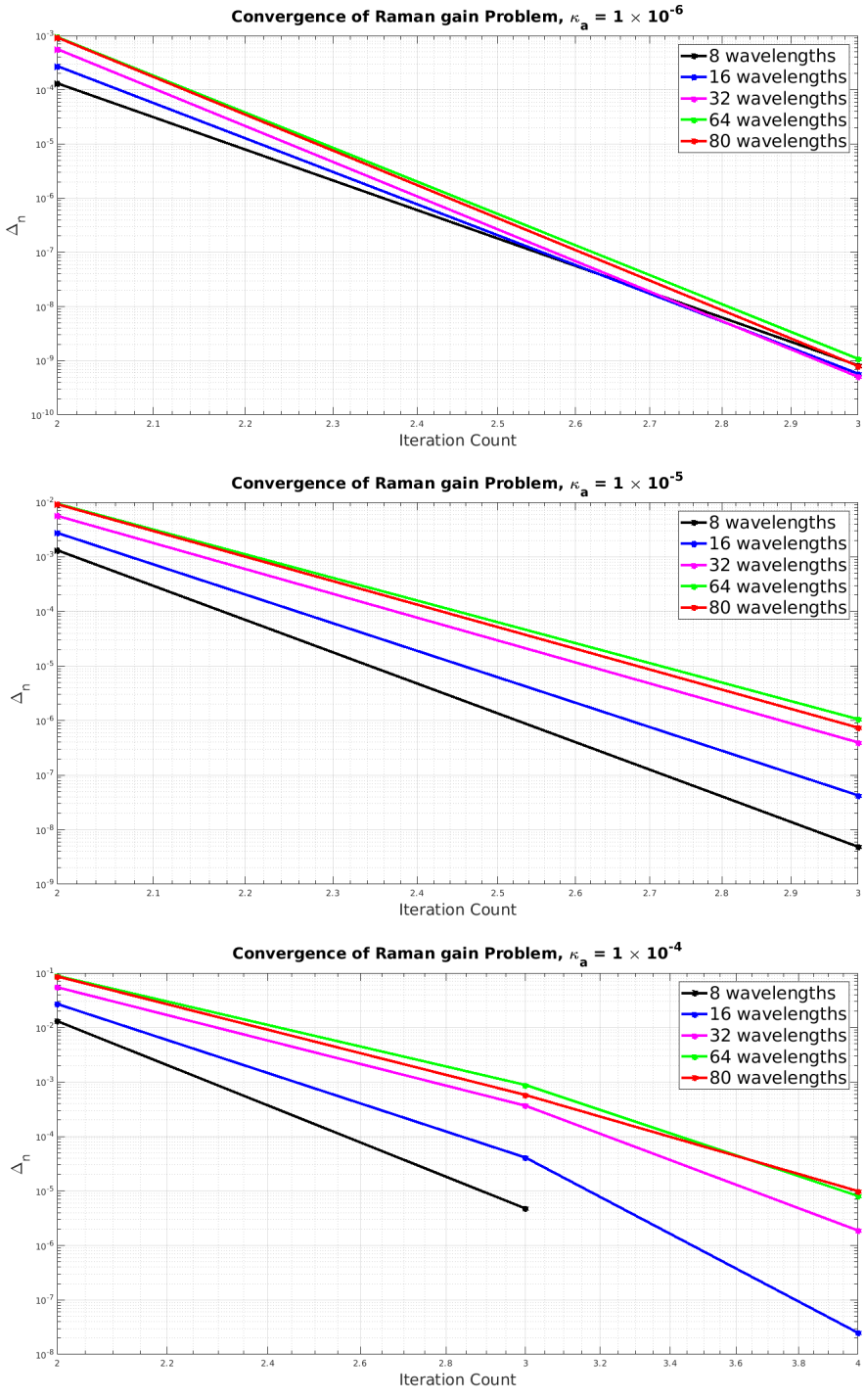

Figure 6.16: Nonlinear convergence

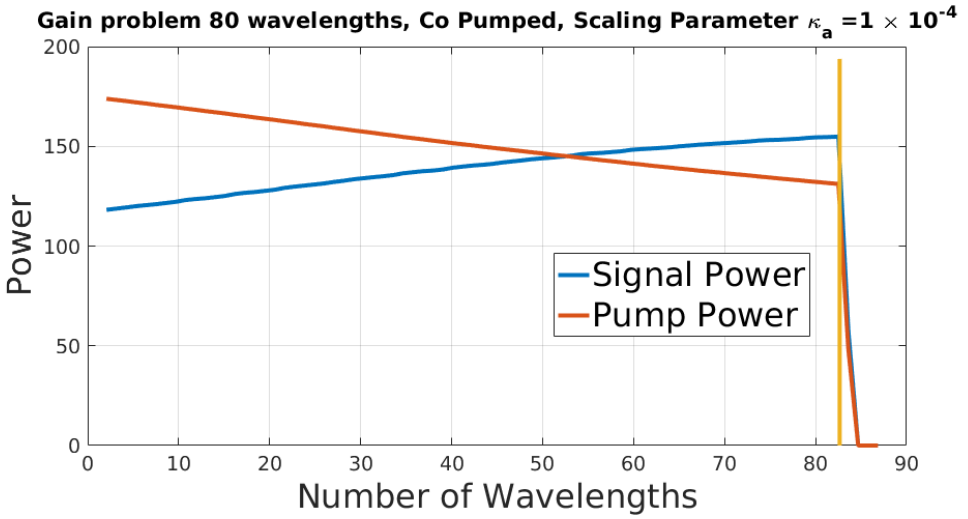

Figure 6.17: Gain for fiber of length 80 wavelengths

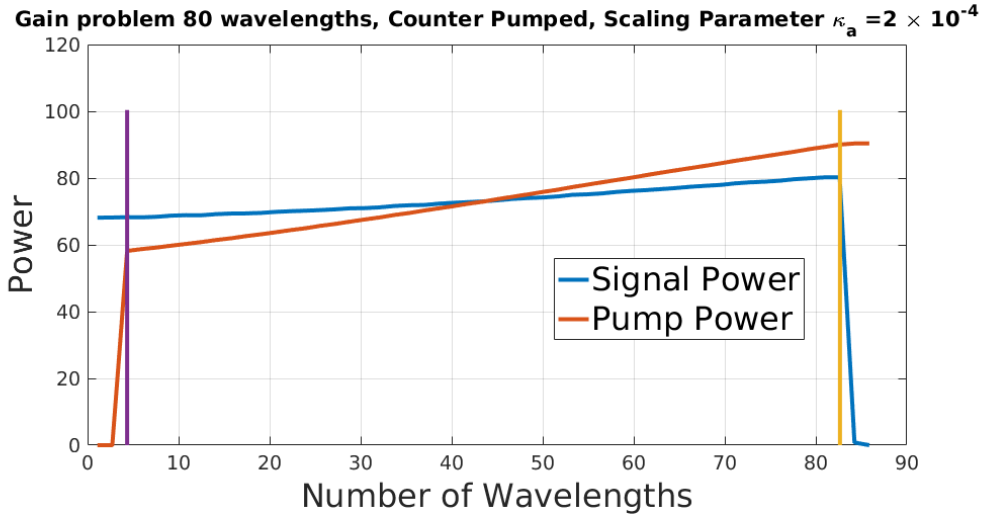

Figure 6.18: Gain for fiber of length 80 wavelengths

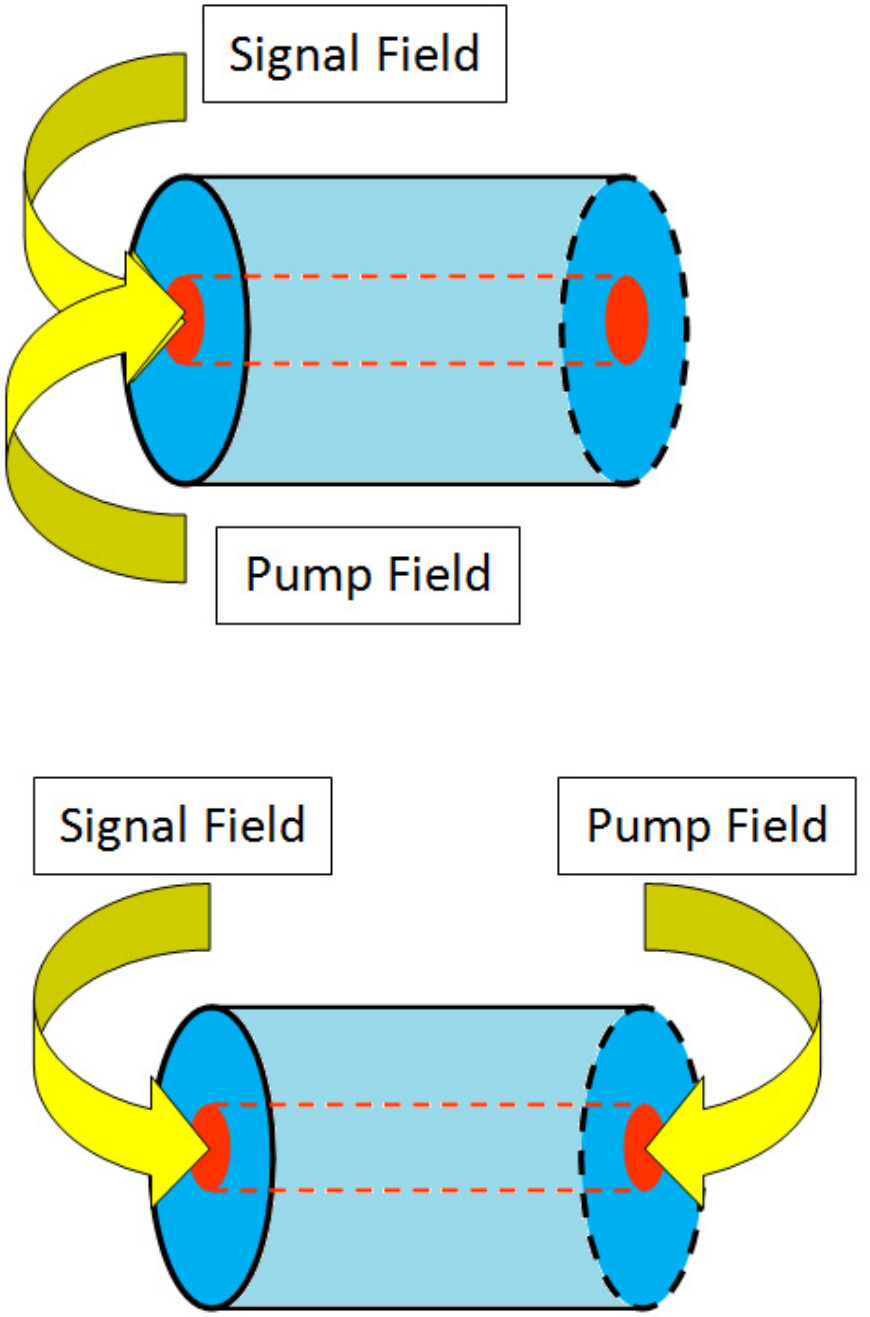

Figure 6.19: Co- pumped (top) and counter pumped (bottom) configuration schematic

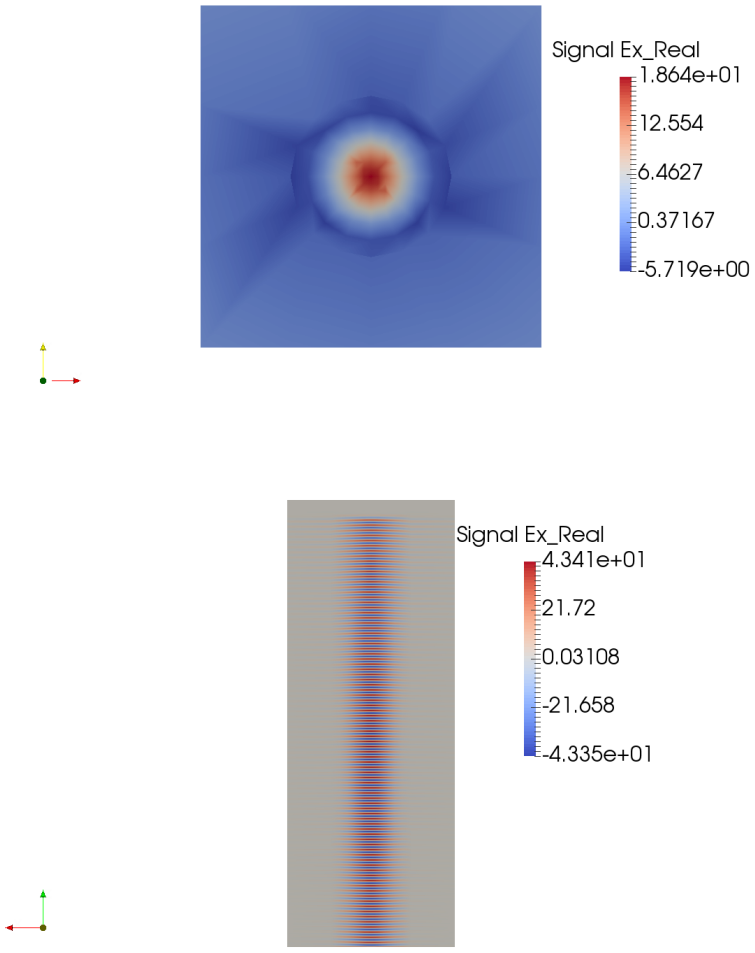

Figure 6.20: Nonlinear Problem, Signal Field: Real Part of E_x

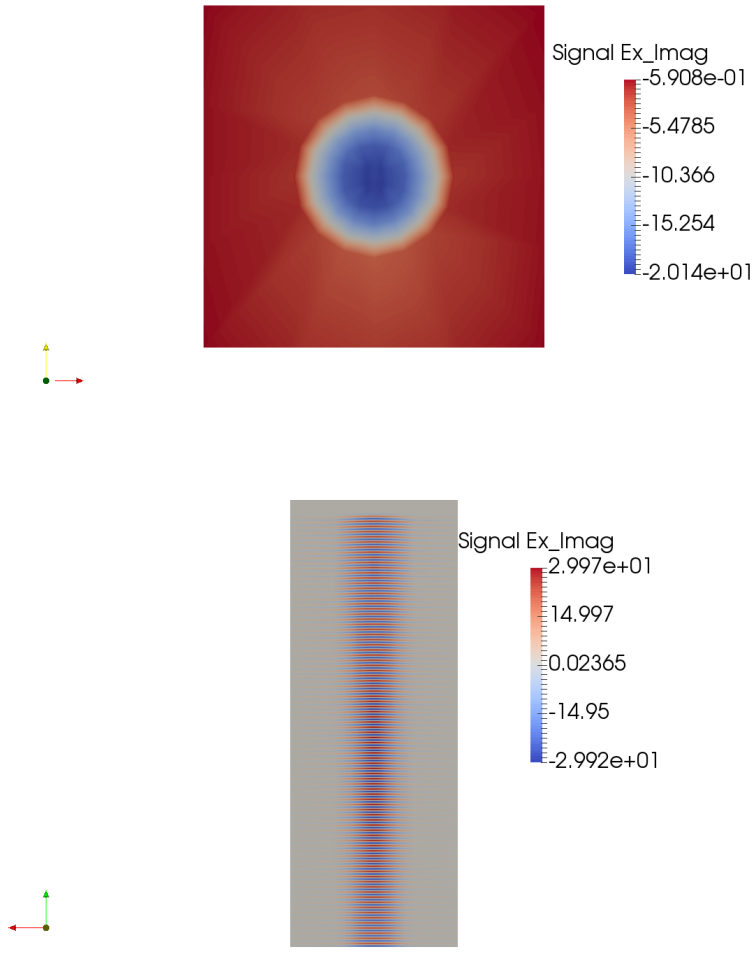

Figure 6.21: Nonlinear Problem, Signal Field: Imaginary Part of E_x

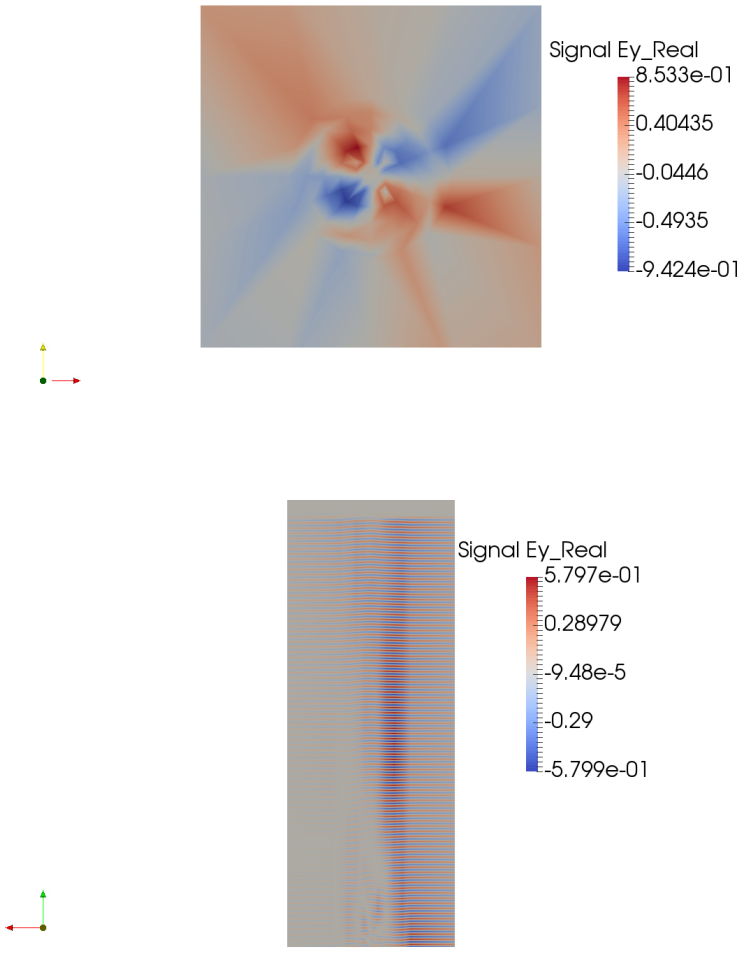

Figure 6.22: Nonlinear Problem, Signal Field: Real Part of E_y

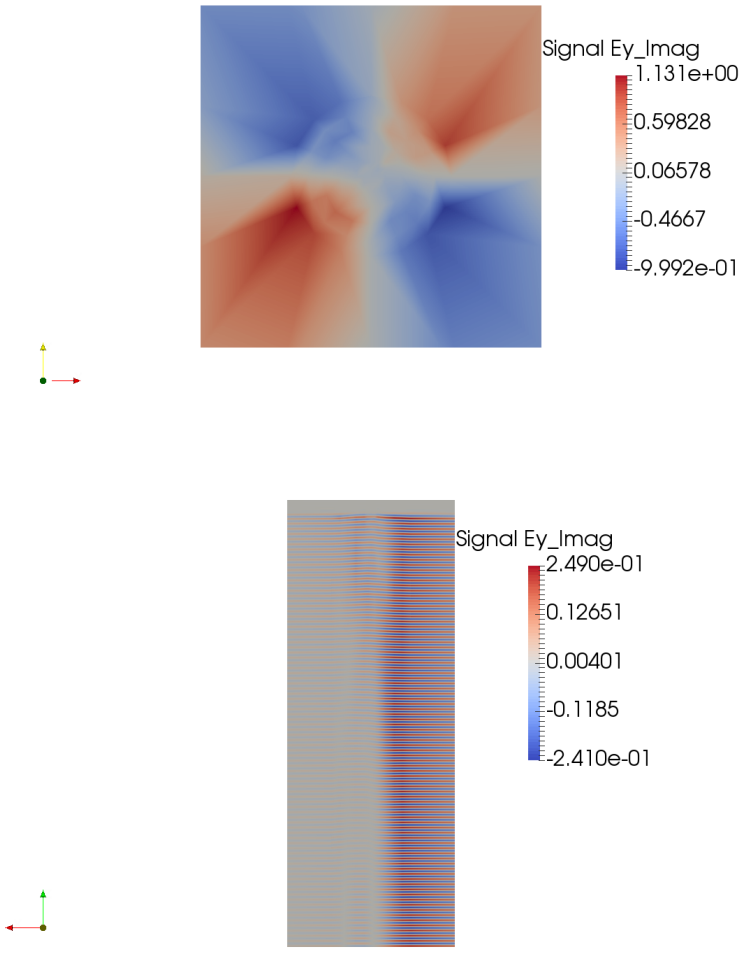

Figure 6.23: Nonlinear Problem, Signal Field: Imaginary Part of E_y

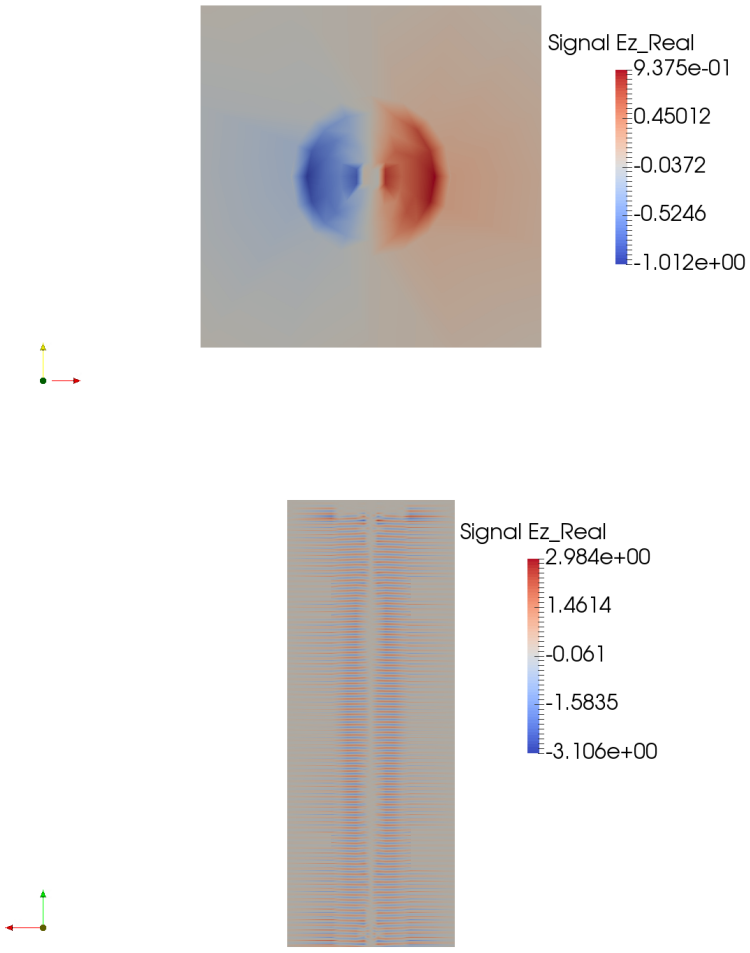

Figure 6.24: Nonlinear Problem, Signal Field: Real Part of E_z

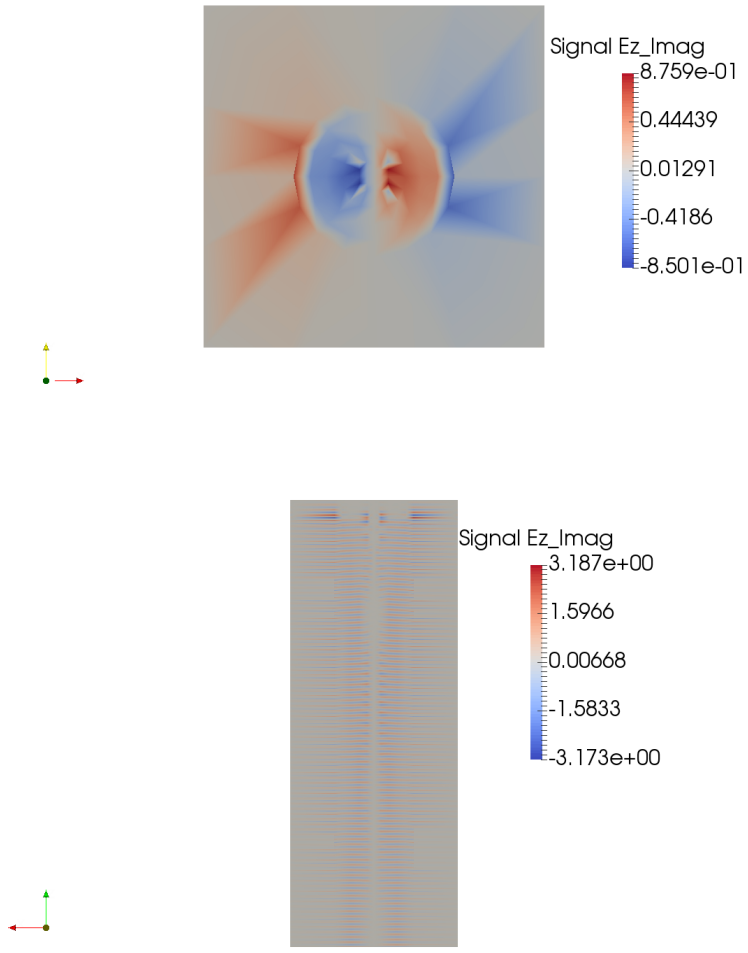

Figure 6.25: Nonlinear Problem, Signal Field: Imaginary Part of E_z

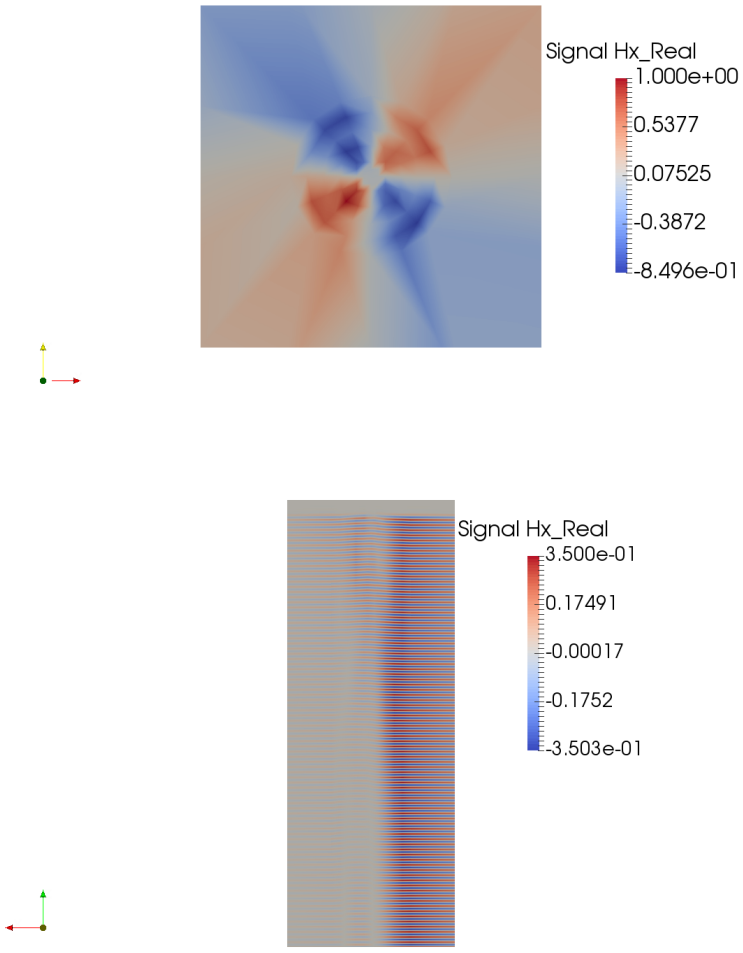

Figure 6.26: Nonlinear Problem, Signal Field: Real Part of H_x

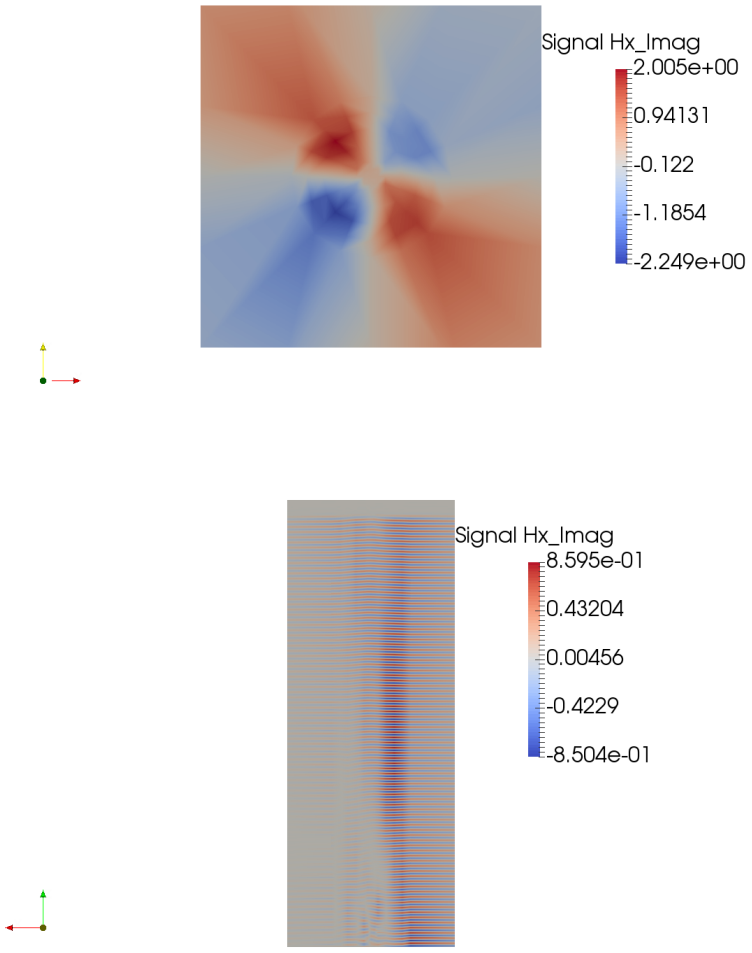

Figure 6.27: Nonlinear Problem, Signal Field: Imaginary Part of H_x

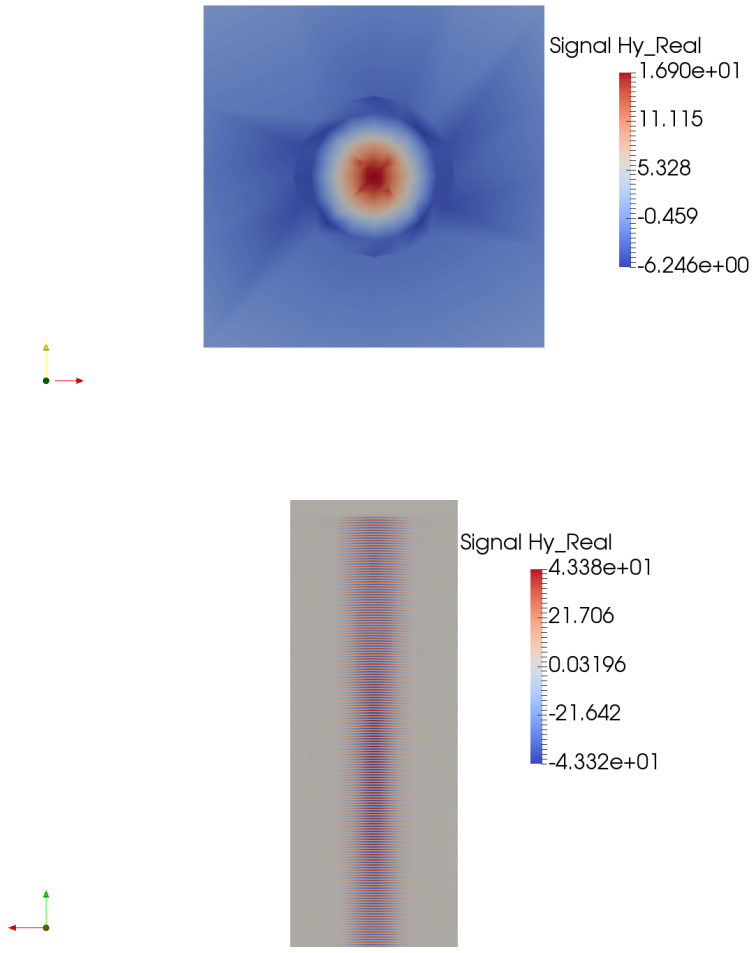

Figure 6.28: Nonlinear Problem, Signal Field: Real Part of H_y

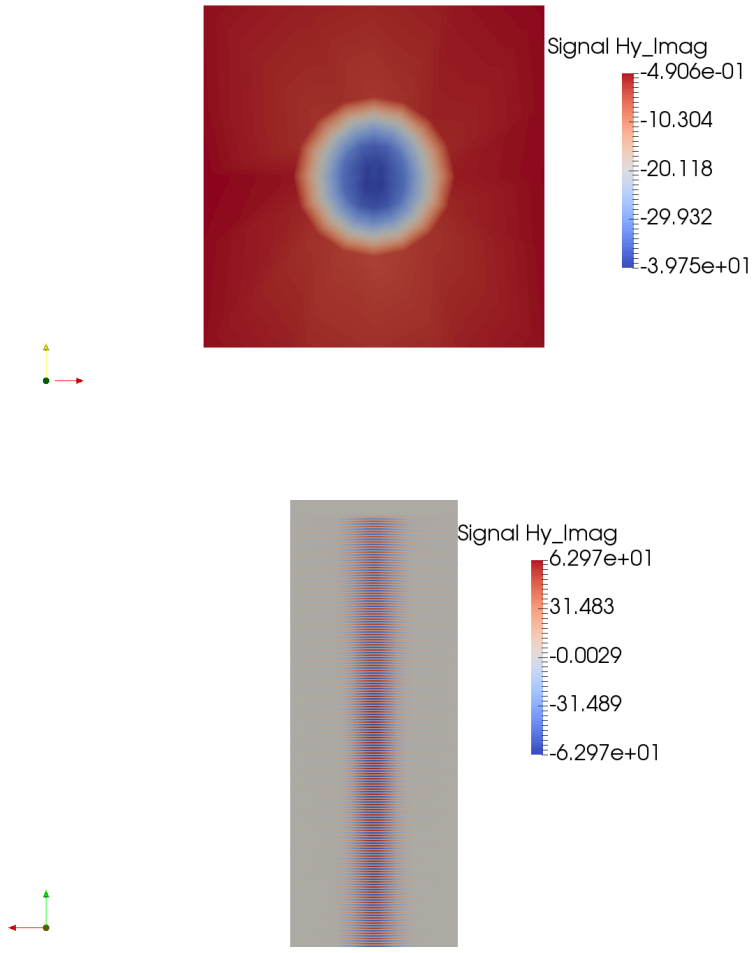

Figure 6.29: Nonlinear Problem, Signal Field: Imaginary Part of H_y

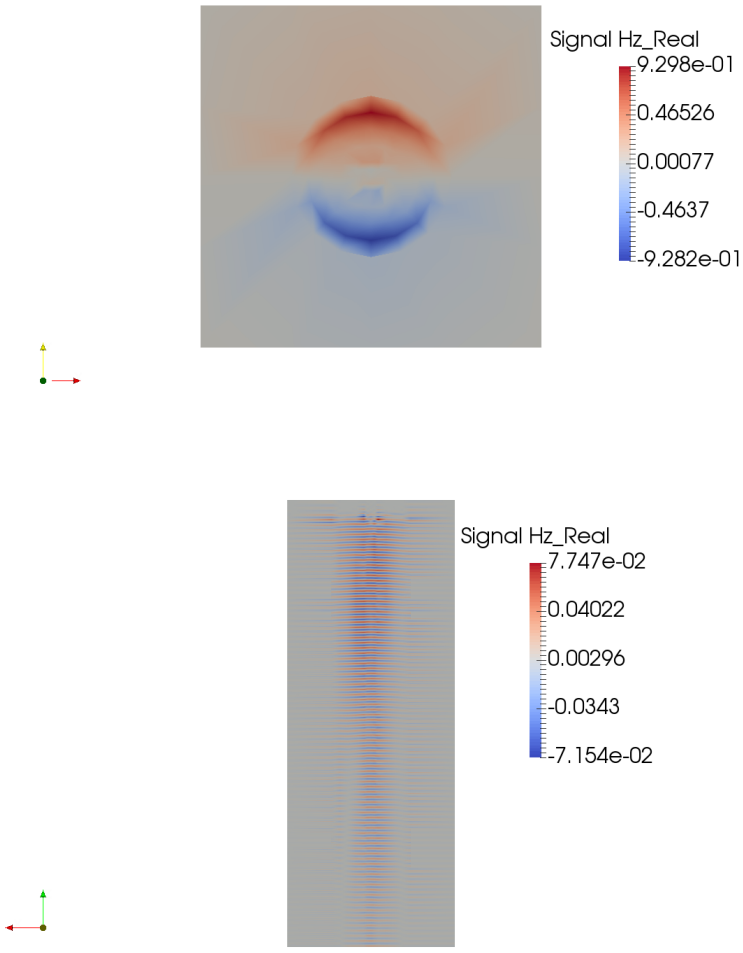

Figure 6.30: Nonlinear Problem, Signal Field: Real Part of H_z

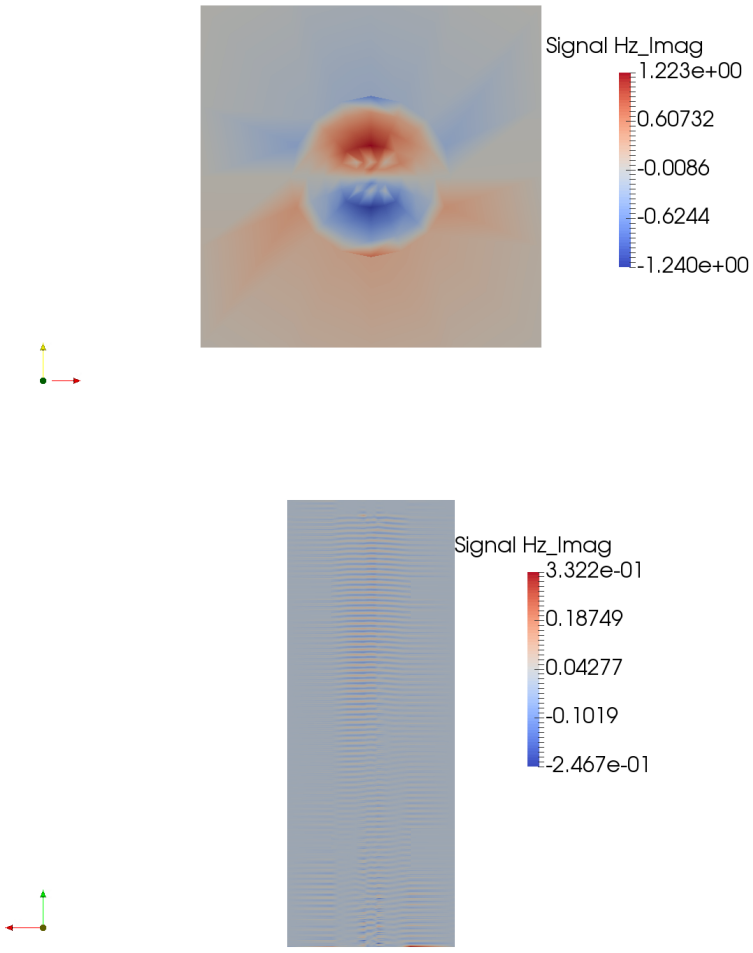

Figure 6.31: Nonlinear Problem, Signal Field: Imaginary Part of H_z

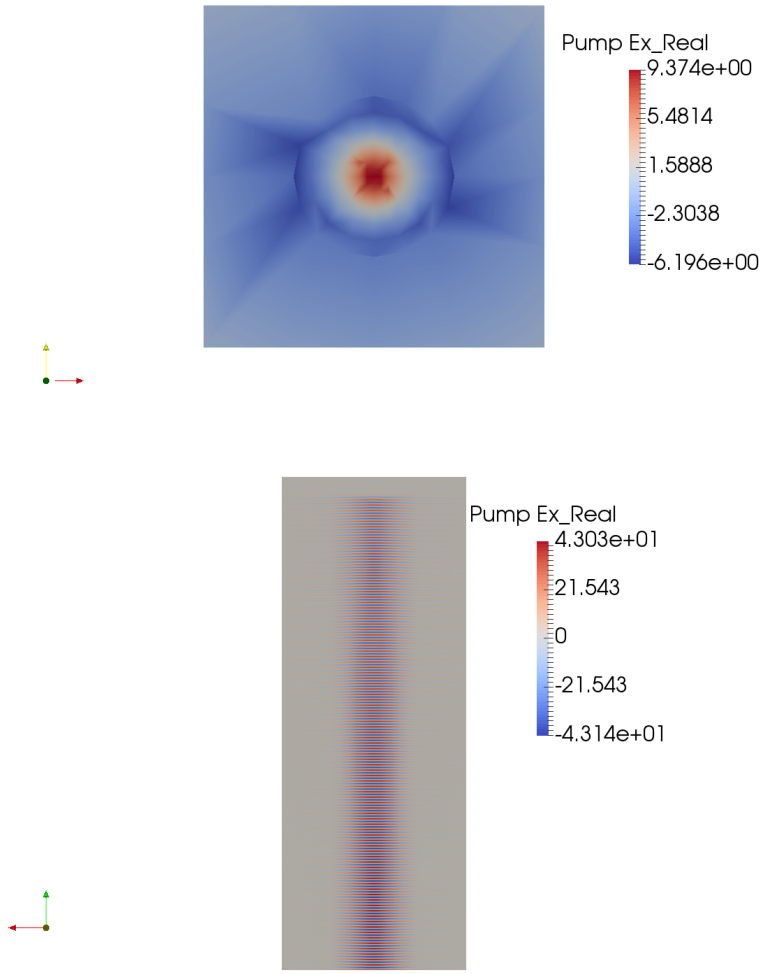

Figure 6.32: Nonlinear Problem, Pump Field: Real Part of E_x

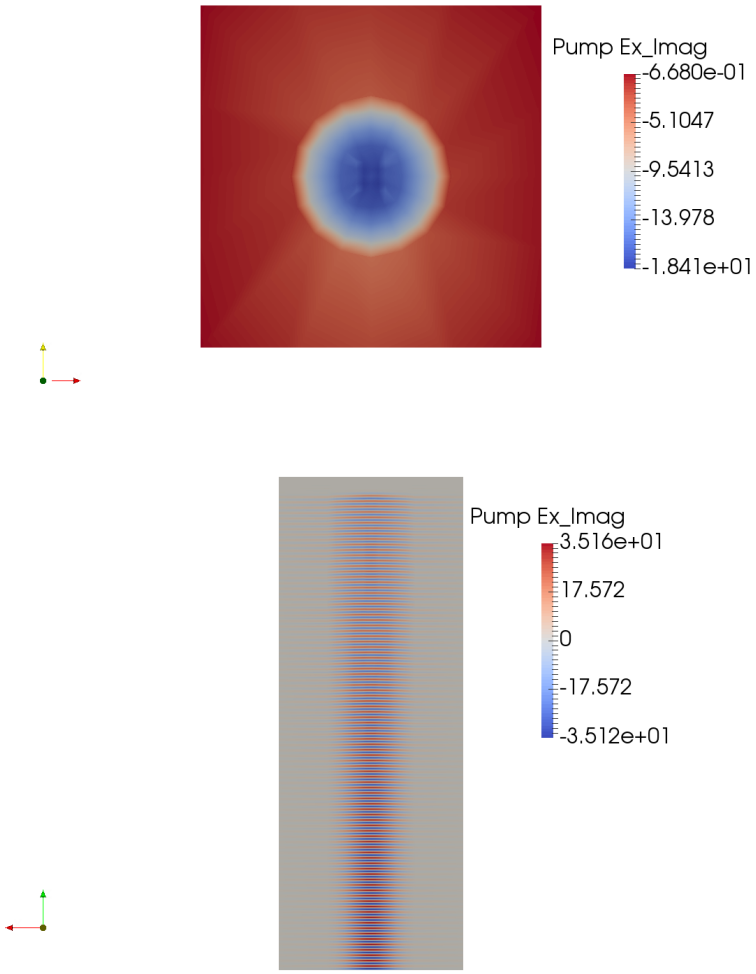

Figure 6.33: Nonlinear Problem, Pump Field: Imaginary Part of E_x

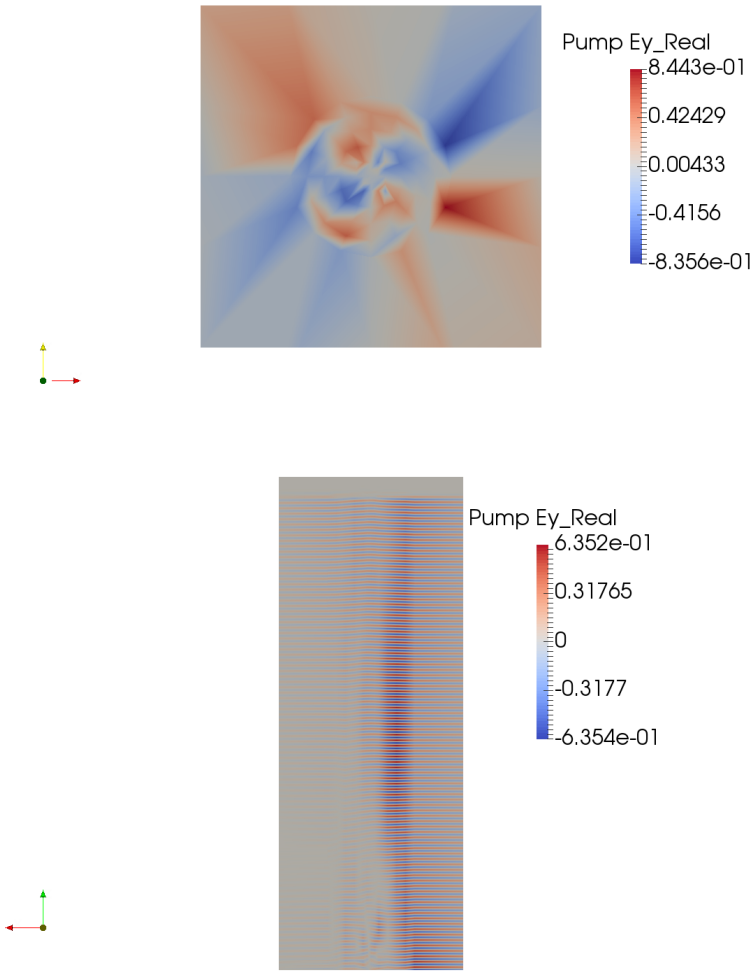

Figure 6.34: Nonlinear Problem, Pump Field: Real Part of E_y

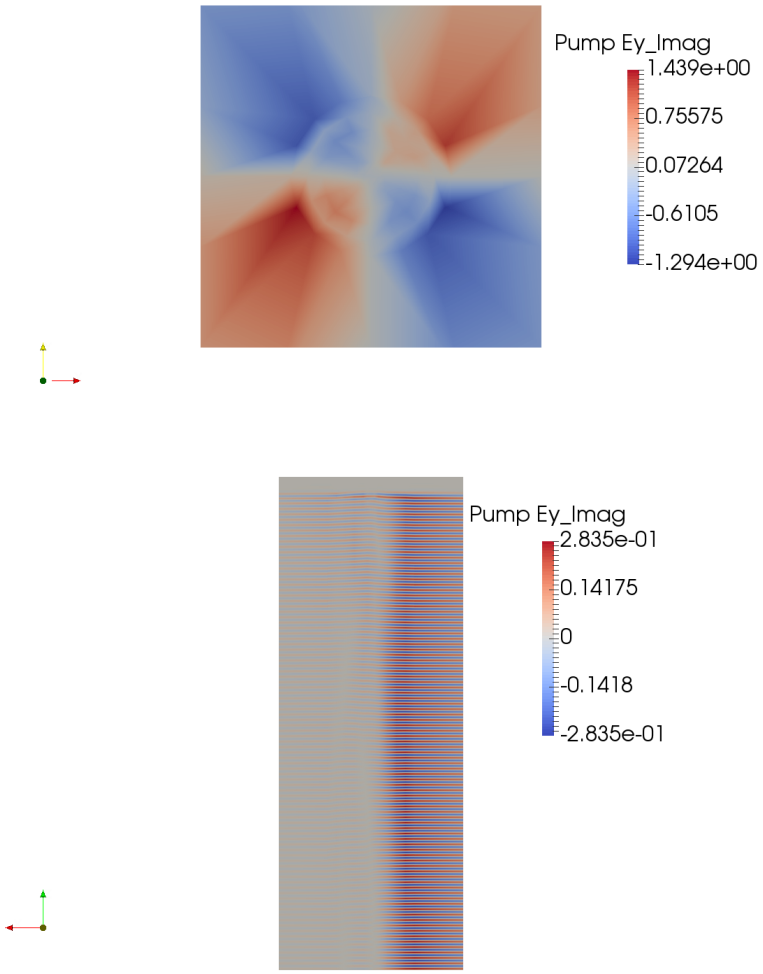

Figure 6.35: Nonlinear Problem, Pump Field: Imaginary Part of E_y

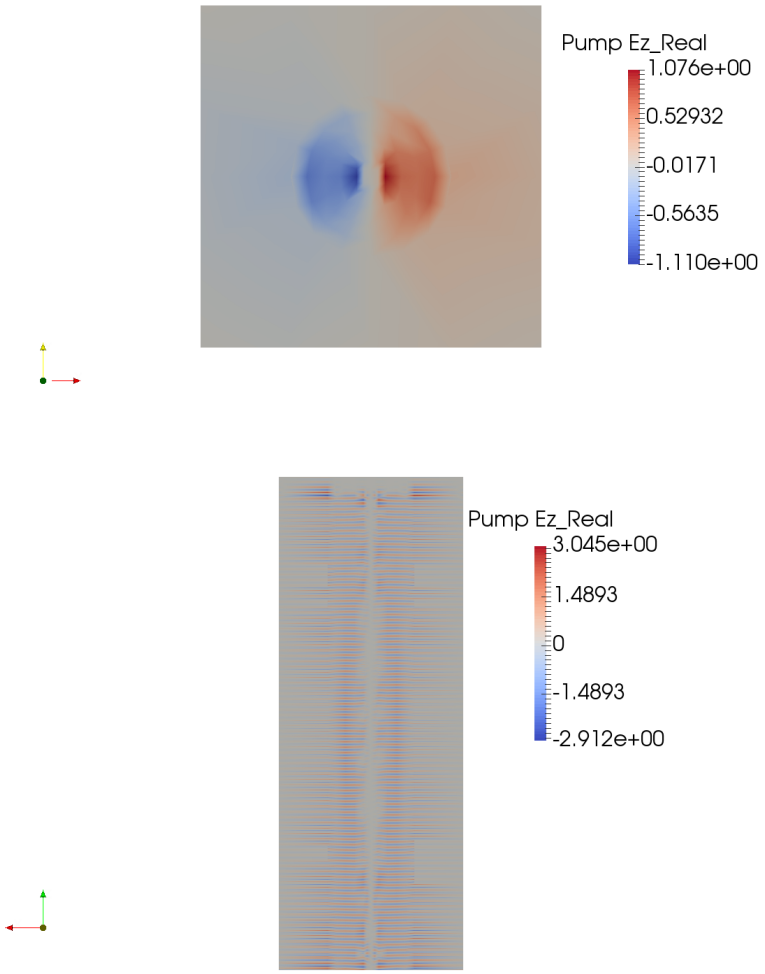

Figure 6.36: Nonlinear Problem, Pump Field: Real Part of E_z

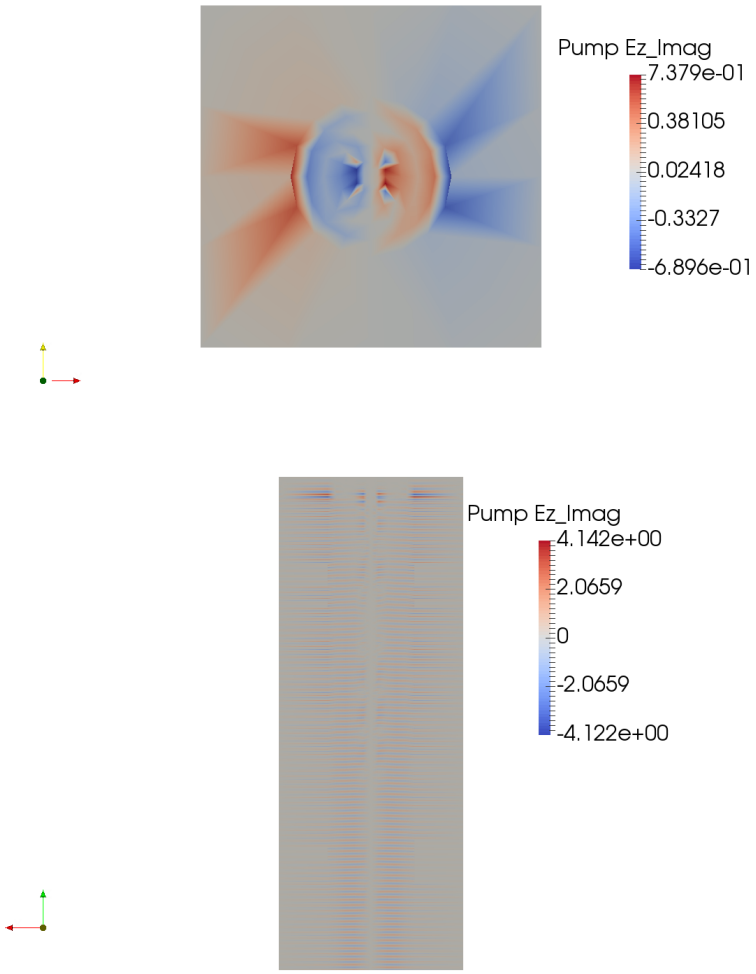

Figure 6.37: Nonlinear Problem, Pump Field: Imaginary Part of E_z

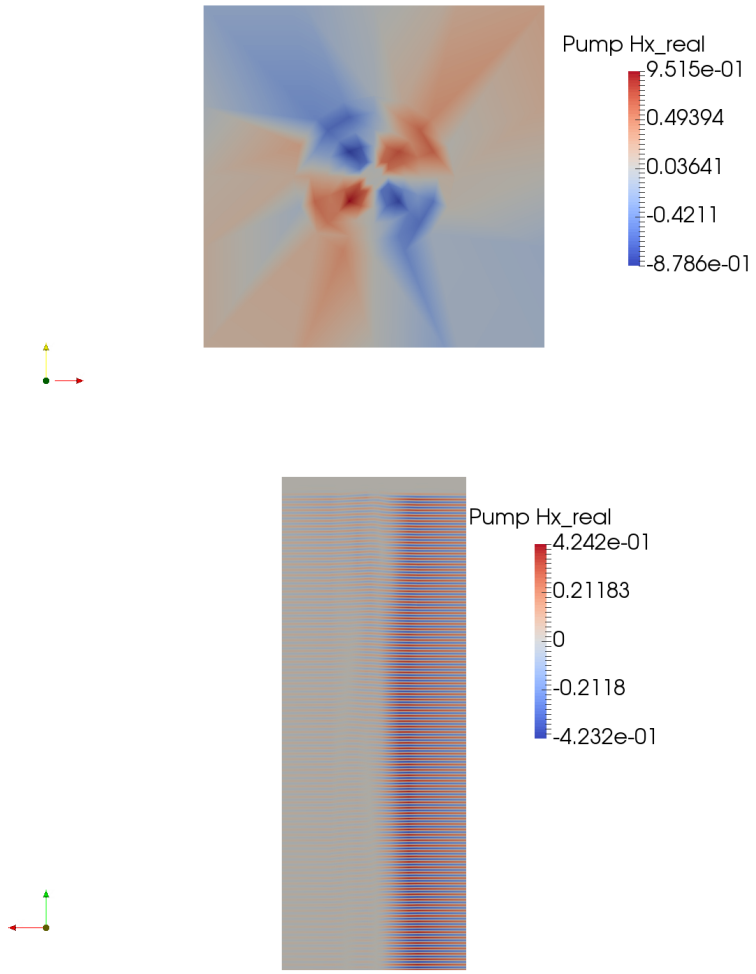

Figure 6.38: Nonlinear Problem, Pump Field: Real Part of H_x

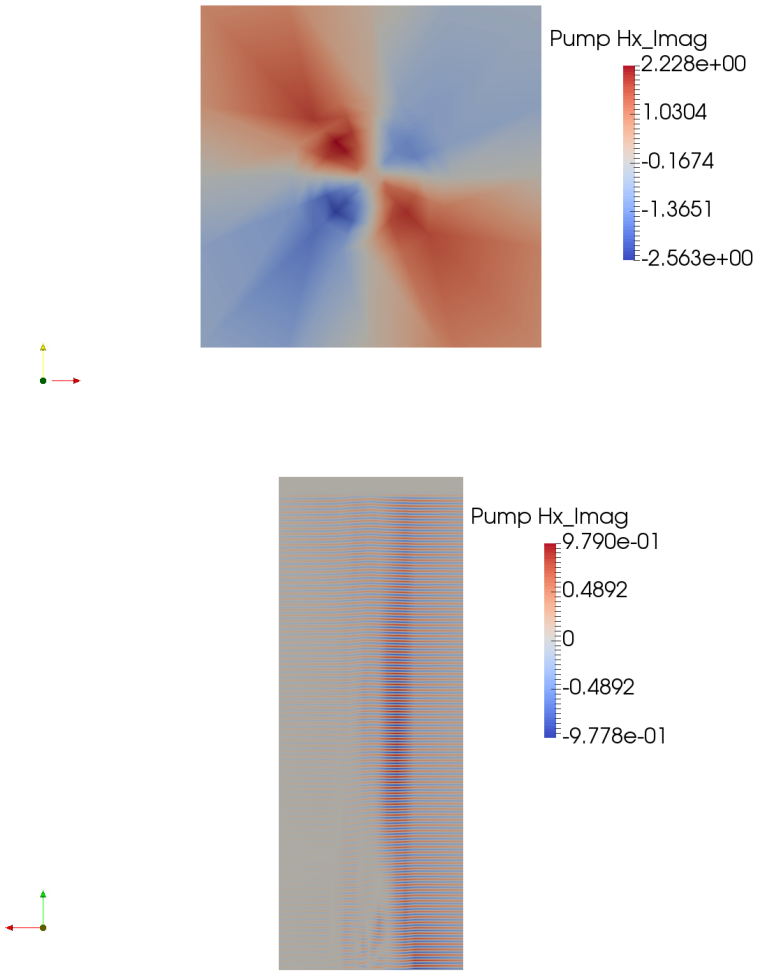

Figure 6.39: Nonlinear Problem, Pump Field: Imaginary Part of H_x

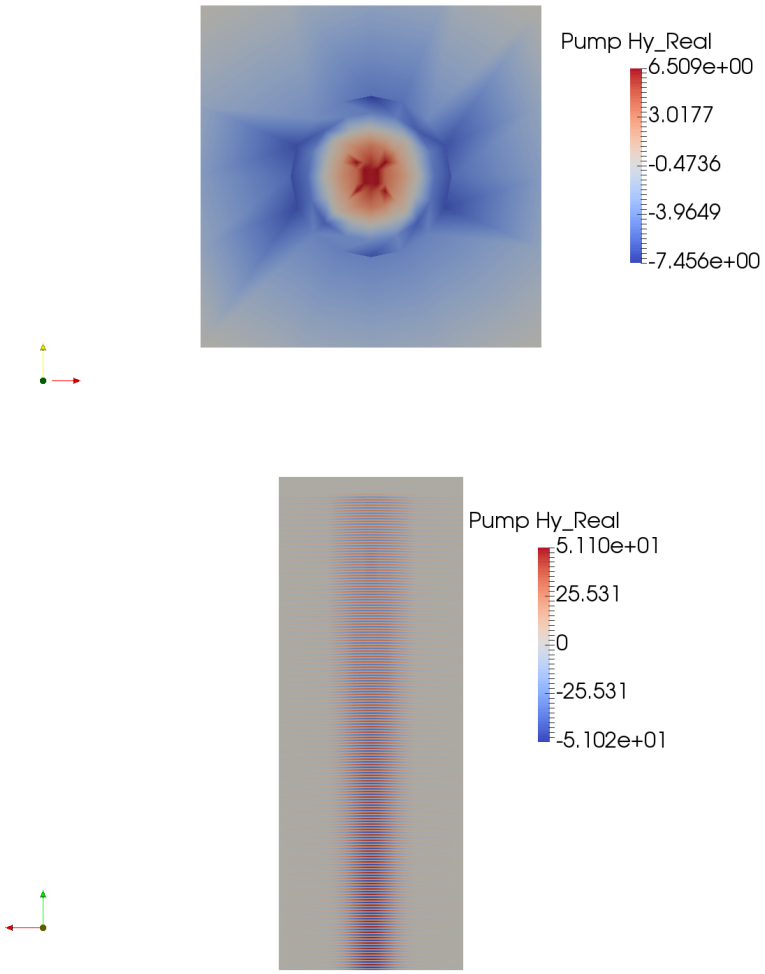

Figure 6.40: Nonlinear Problem, Pump Field: Real Part of H_y

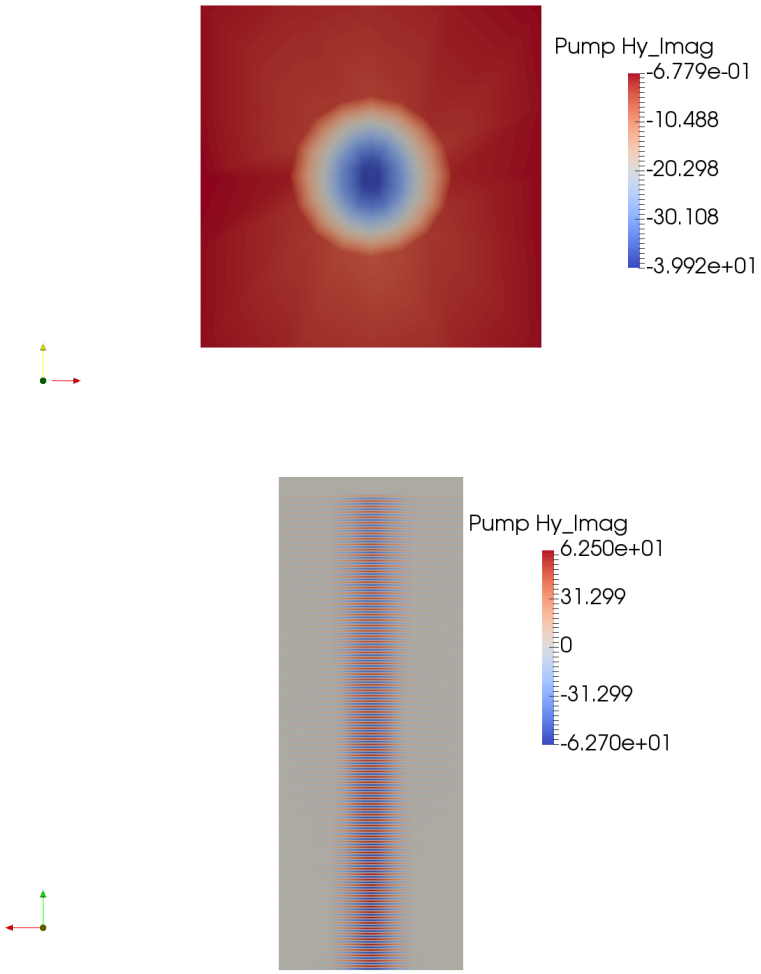

Figure 6.41: Nonlinear Problem, Pump Field: Imaginary Part of H_y

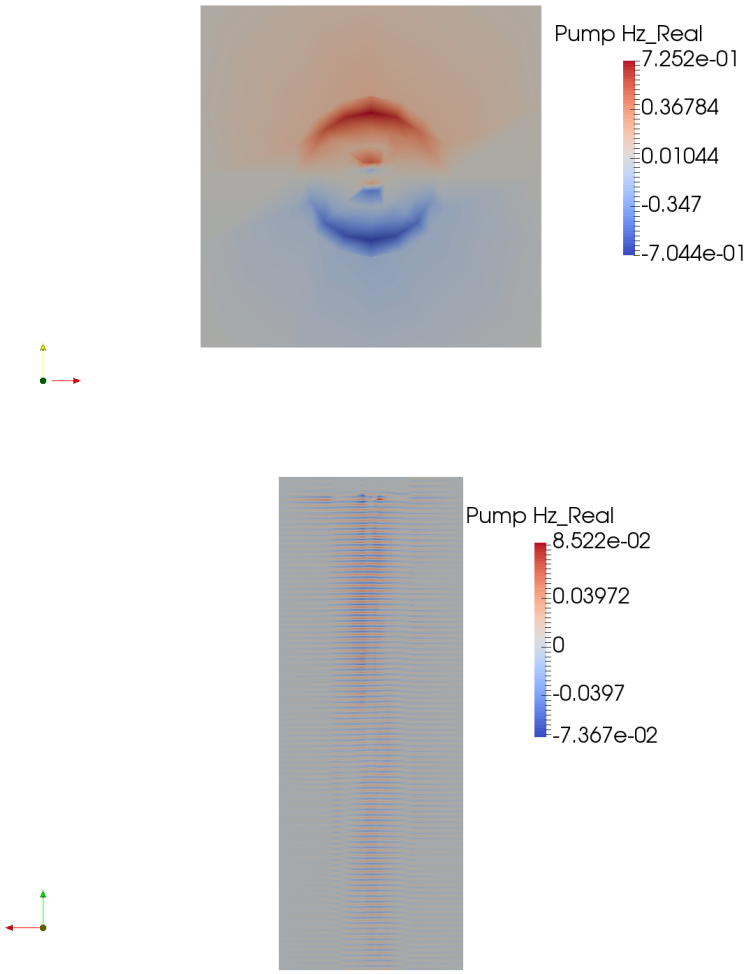

Figure 6.42: Nonlinear Problem, Pump Field: Real Part of H_z

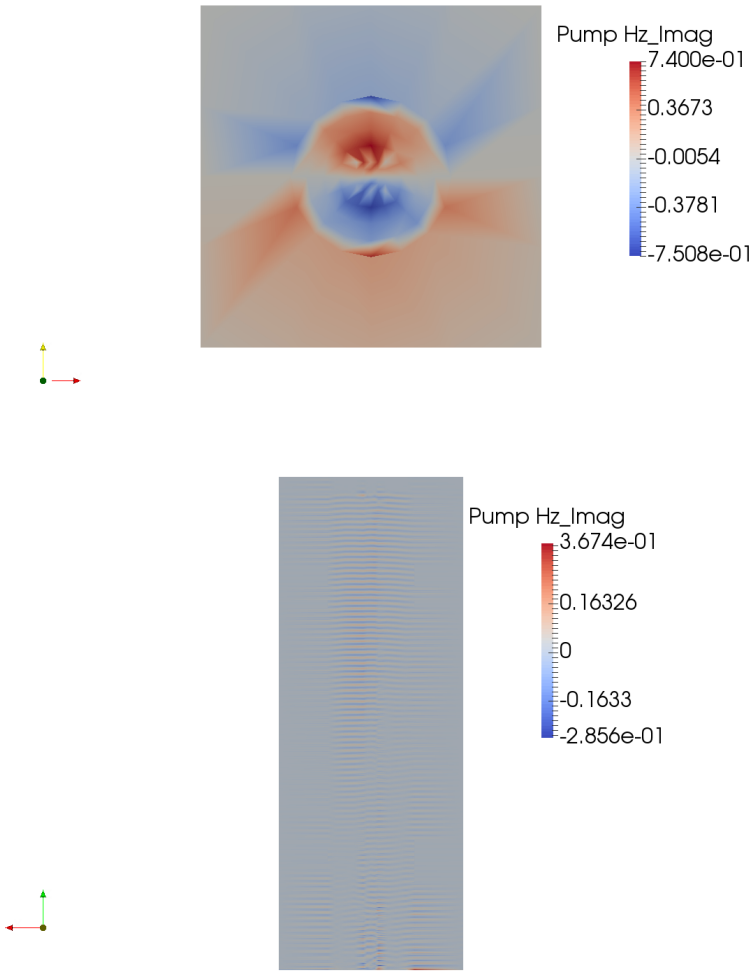

Figure 6.43: Nonlinear Problem, Pump Field: Imaginary Part of H_z

Chapter 7

Summary and Future Directions

The main aim of this work was to develop a DPG framework for applications in linear and nonlinear problems arising in optics, particularly fiber optics. Chapter 3 described the general features of the DPG methodology and the optimality properties of the "ideal" DPG method. The stability of a "practical" choice of optimal test functions in the DPG context was analyzed by constructing a DPG Fortin operator which was used to quantitatively measure the change in stability while approximately inverting the Riesz map in computing the optimal test space. This analysis was done for H^1 and $H(\mathrm{div})$ spaces defined on triangular elements. Chapter 4 was devoted to analyzing the linear time dependent Schrödinger equation (LSE) from a variational standpoint. The greatest obstacle was the non-existence of an L^2 stable first order reformulation of the LSE and the possibility of selective relaxation (i.e., integration by parts) was not available: the LSE must be dealt with as a second order equation with non-standard energy spaces and only two variational formulations: the strong and ultraweak (UW) formulations. The notion of an auxiliary "boundary" operator that generalized the notion of trace was developed and *inf-sup* stability of the strong and UW formulations was proved.

Theoretical convergence rates for the 1-space dimension case using interpolation error estimates based on a specialized A-conforming element. Numerical evidence to corroborate our theoretical convergence theory was provided.

Chapter 5 and 6 presented a unique full 3D Maxwell DPG simulation of a passive optical fiber amplifier that experiences stimulated Raman scattering. The aim was to develop computational tools for the most general model with the fewest simplifying approximations, with the intent to eventually develop a high-fidelity, multi-physics fiber model that can handle much more complex problems with realistic fiber lengths. However, in this chapter, the primary interest was establishing the numerical approach and validating its feasibility by observing the qualitative characteristics of Raman gain. Towards that end, the superiority of ultraweak DPG formulation of the coupled Maxwell system was demonstrated numerically, and was implemented in the model using a perfectly matched layer (PML). It was successfully shown that a nonlinear iterative method was able to handle the nonlinear gain. The use of sum factorization for making the element computations tractable was critical to the success of this endeavor. The model validation included a convergence test, energy conservation in a linear waveguide, and qualitative gain results from a typical amplification problem. Also included was the case of a counter-pumped configuration, which is beyond the scope of most traditional scalar BPM models. This also verifies the new full vectorial electric polarization term for Raman scattering, which emphasizes the fact that the bulk Raman gain coefficient is the primary measured value available to the computer modeling

team, as a practical approach to simulating Raman gain in fibers.

The appendices provide a comparison of the primal and ultraweak formulations of Maxwell's equations, details of the DPG implementation of the perfectly matched layer used in the Raman model simulations, details of the sum factorization used in the 3D Maxwell simulations and miscellaneous results used in the construction of the Fortin operators. The final appendix contains the iteratively reweighted least squares approach to minimizing L^p residuals.

7.0.1 Future Directions

Moving forward, there are key issues, both on the theory side and applications side, that can be explored. Explicit construction of general Fortin operators for problems with $H(\text{curl})$ energy spaces, for instance, is an immediate open problem. A general framework for studying nonlinear problems discretized using the DPG method is another open problem, although attempts have been made in [64, 11].

With regard to extensions of the applications of DPG to nonlinear optics, several points are in order. In order to extend the use of full vector 3 dimensional DPG Maxwell models to large-scale simulations, a significant increase in computational resources will be needed. In particular, investments must be made to develop a distributed memory architecture, which entails looking at possible MPI implementations. Second, a novel nested dissection solver, or some variant thereof, would take into advantage the possibility of

static condensation of the field variables in the ultraweak formulation resulting in further optimization of the code. Third, as the complexity of the model as well as size of the problem increases, more sophisticated nonlinear techniques may be required. From a modeling perspective, these results indicate that additional physical phenomena such as stimulated Brillouin scattering (SBS), transverse mode instability (TMI) as well as more sophisticated fiber designs and configurations, such as gain tailoring, microstructure fibers, bi-directional pumping, etc., can be easily accommodated within the current model. These additional modeling endeavors may require coupling DPG with other formulations [42], or implementing a coupling among various DPG formulations [41] in the Maxwell case.

Appendix A

Another Charaterization of the Optimal Test Space

As we have seen, the optimal test space (for a given trial space) consists of the vectors that achieve the inf-sup condition. Another way to view how the space of optimal test functions guarantees stability is as follows[1].

Author contributions: The contents of this appendix are taken largely from the published multi-author article [67] which is co-authored by the author of this dissertation. The author of this dissertation contributed to the development of the theory and numerical results presented in [67], including the mathematical constructions/derivations as well as the writing of the manuscript.

The standard Petrov-Galerkin method *fixes* both U_h and V_h and hence, in order to ensure the inf-sup condition, we must have $B(U_h) = R_V(V_h)$. However, depending on the exact form of B and the subspaces U_h, V_h, this may not always be true. Indeed, in general, $B(U_h)$ is just an arbitrary finite

[1]The material in this appendix is taken largely from the published work [67],

dimensional subspace of V', and there is no reason for us to believe that the operator B restricted to U_h must map to $R_V(V_h)$. This discrepency between $B(U_h)$ and $R_V(V_h)$ is fundamental cause for lack of stability at the discrete level.

Now, we present yet another characterization of the optimal test space. We start with a lemma:

Lemma A1 Given a Hilbert space V and a closed subspace $A \subset V$, we have $R_V(A) = (A^\perp)^\circ$, where $(S)^\circ$ denotes the annihilator of a set $S \subset V$, i.e., $(S)^\circ = \{f \in V^{'} | f(s) = 0 \forall s \in S\}$.

Proof: Let $f \in R_V(A)$, and $a^\perp \in A^\perp$. Then, $f(a^\perp = (R_V^{-1}(f), a^\perp)_V = 0$, since $R_V^{-1}(f) \in A$ and $a^\perp \in A^\perp$, so $R_V(A) \subset (A^\perp)^\circ$.

Next, let $f \in (A^\perp)^\circ$. Now, $f(a^\perp) = (R_V^{-1}(f), a^\perp) = 0$ for $a^\perp \in A^\perp$. Since $V = A \oplus A^\perp$, we have that $R_V^{-1}(f) \in A$, or, $f \in R_V(A)$. Thus, $R_V(A) = (A^\perp)^\circ \square$.

Let us denote $R_V(A)$ for a closed subspace of V as A'. We thus have the following situation. For an arbitrary choice of V_h, we may not have $V_h^{'} = B(U_h)$, and so may not have the discrete inf-sup condition. However, if we pick V_h as the optimal test space, i.e., $R_V^{-1} B(U_h) = V_h^{opt}$, then, by definition, $B(U_h) = (V_h^{opt})^{'}$, ensuring the discrete inf-sup condition.

A.1 Proof of H^1 norm equivalence:

Choice of norm for the broken test space We shall use the following norm (for the analysis):

$$\|u\|^2_{H^1(\Omega_h)} = \sum_K \Big(\|\nabla u\|^2_{L^2(K)} + \|\bar{u}\|^2_{L^2(K)} \Big) \tag{A.1.1}$$

where

$$\bar{u} := \frac{1}{|K|} \int_K u \, dK$$

is the average value of function u in element K. The norm is equivalent with the standard broken H^1-norm with mesh independent equivalence constants. Indeed,

$$\begin{aligned} \|\bar{u}\|^2_{L^2(K)} &= \int_K |\bar{u}|^2 = |k| \, |\bar{u}|^2 = |K|^{-1} \, | \int_K u|^2 \\ &\leq |K|^{-1} \int_K |u|^2 \, |K| \qquad \text{(Schwartz inequality for } \int_K u \cdot 1) \\ &= \|u\|^2_{L^2(K)} \, . \end{aligned}$$

Likewise, by Pythagoras theorem,

$$\|u\|^2_{L^2(K)} = \|u - \bar{u} + \bar{u}\|^2_{L^2(K)} = \|u - \bar{u}\|^2_{L^2(K)} + \|\bar{u}\|^2_{L^2(K)} \, .$$

Function $u - \bar{u}$ has a zero average, $u - \bar{u} \in H^1_{avg}(K)$, and so does the corresponding pullback $\hat{u} - \overline{\hat{u}}$. Recalling the Poincaré inequality for the master element,

$$\|\hat{u}\|^2_{L^2(\hat{K})} \leq \frac{\sqrt{2}}{\pi} \|\hat{\nabla} \hat{u}\|^2_{L^2(\hat{K})} \quad \forall \hat{u} \in H^1_{avg}(\hat{K}) \, , \tag{A.1.2}$$

and applying the scaling argument, we get,

$$\|u\|^2_{L^2(K)} \leq \frac{\sqrt{2}}{\pi} \, h^2 \, \|\nabla u\|^2_{L^2(K)} \quad \forall u \in H^1_{avg}(K) \, . \tag{A.1.3}$$

In conclusion,

$$\|u\|^2_{L^2(K)} \leq \frac{\sqrt{2}}{\pi} h^2 \|\nabla u\|^2_{L^2(K)} + \|\bar{u}\|^2_{L^2(K)} .$$

The first equivalence constant is one and, for small h, the second equivalence constant is very close to one, too.

Appendix B

PML Details

In this appendix[1], we provide details of the PML implementation.

Author contributions: The contents of this appendix are taken largely from the multi-author article "A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers" S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, J. Mora, in preparation. The article has not yet been submitted for journal publication. The author of this dissertation contributed to model development, and code/numerical implementation of the model and analysis of the results.

Recall that the use of a PML was required due to the non-zero Raman gain term, which acts as a nonlinear conductivity. Since the model pursued in this paper is a full 3D boundary value problem (BVP) model, we must specify appropriate boundary conditions at all boundaries of the domain. We have the source located at the input ($z = 0$) end of the fiber, and PEC boundary

[1]The content of this appendix is taken from the manuscript "A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers" S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, J. Mora, in preparation. Information approved for public release on 08 May 2018 by AFRL OPSEC/PA OPS-18-19547.

conditions set to 0 at the external radial boundary ($r = r_{\text{cladding}}$). Given that we are modeling a fiber an arbitrary length, we must have boundary conditions of absorbing type at the (computational) exit end of the fiber ($z = z_L$). The presence of the Raman gain makes the system nonlinear, and naive absorbing impedance (or Robin) boundary conditions would induce spurious solutions. These issues can be overcome with a perfectly matched layer (PML) at the exit end of the fiber that allows the signal field to gain power and the pump field to lose power within the computational domain, while effectively setting both to 0 outside the computational domain.

Towards this end, the implementation pursued in this paper is a DPG version of a stretched coordinate PML [81], with the requirement that the stretching is done only along the z-direction, due to this being the direction of propagation.

B.1 Complex stretching in z-direction

Let $\phi : \mathbb{R}^3 \to \mathbb{C}^3$ be a smooth invertible map with Jacobian $\mathbb{J}_{ij} = \frac{\partial \phi_i}{\partial x_j}$, where x_i are the real coordinates,

$$\phi(x_1, x_2, x_3) = (\phi_1(x_1, x_2, x_3), \phi_2(x_1, x_2, x_3), \phi_3(x_1, x_2, x_3))$$

and $i, j = 1, \ldots, 3$. We let $J = |\mathbb{J}|$ denote the determinant of the Jacobian, $\mathbb{J}^{-1}$ denote its inverse and $\mathbb{J}^{-T}$ denote its inverse transpose. The map ϕ will be our stretching map: ϕ acts as identity within the computational domain, while outside, it is designed to kill outgoing waves. In our case, we have that

$\phi_1, \phi_2 = 1$, since we need to stretch only the z-axis. Thus,

$$\mathbb{J} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{\partial \phi_3}{\partial x_3}. \end{pmatrix}$$

The choice of the stretching function ϕ_3 is particularly important. Given the growth of the signal field, the growth of ϕ_3 must be commensurate so that the signal field is killed effectively. The pump is assumed to be decaying anyway, so that it will be killed by the same PML nonetheless.

The Maxwell operator including the complex stretching takes the form:

$$\widetilde{A} \begin{pmatrix} \mathbb{E} \\ \mathbb{H} \end{pmatrix} := \begin{pmatrix} -(i\omega\epsilon + \sigma) J \mathbb{J}^{-1} \mathbb{J}^{-T} & \nabla \times \\ -\nabla \times & -i\omega\mu J \mathbb{J}^{-1} \mathbb{J}^{-T} \end{pmatrix} \begin{pmatrix} \mathbb{E} \\ \mathbb{H}, \end{pmatrix}$$

and we use the broken ultraweak formulation corresponding to $\widetilde{A}$. For exponential growth, i.e., a wave of the form $e^{(a-i\omega)x_3}$, with $a > 0$, the choice of ϕ_3 must be such that $a - \phi_3(x_1, x_2, x_3) < 0$, in order to ensure exponential decay. Figures 5.2 5.3, depicting the manufactured solution in the fiber waveguide is an example of a truly exponential growth, and thus $\phi_3(x_1, x_2, x_3)$ was of the form $\frac{a}{\omega}(x_3 - L)^n e^{(\frac{x_3}{L})}$ for x_3 values inside the PML region, where L is the length of the computational domain. For the simulations with Raman gain, however, such dramatic exponential growth was not observed for the fiber lengths considered, and thus $\phi_3(x_1, x_2, x_3)$ was of the form $\phi_3(x_1, x_2, x_3) = \frac{25}{\omega} \frac{(x_3 - L)^3}{\beta}$, where β is a fraction of the total length of the fiber used for the PML. For instance, $\beta = 0.2$ for the longest fiber we used.

Appendix C

Sum Factorization Details

In this appendix[1], the need and efficacy of the *sum factorization* technique used in these the 3D computations is briefly reported.

Author contributions: The contents of this appendix are taken largely from the multi-author article "A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers" S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, J. Mora, in preparation. The article has not yet been submitted for journal publication. The author of this dissertation contributed to model development, and code/numerical implementation of the model and analysis of the results.

The exact implementation details and algorithmic break-down of this numerical integration is provided in detail in [62]. The sum factorization idea is an efficient way to dramatically reduce the time involved in the element computations (integration of element stiffness and Gram matrices) by appealing to

[1]The content of this appendix is taken from the manuscript "A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers" S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, J. Mora, in preparation. Information approved for public release on 08 May 2018 by AFRL OPSEC/PA OPS-18-19547.

the tensor structure of the element shapes and associated shape functions. For instance, a hexahedral element can be viewed as the tensor product of three 1D segments, and the corresponding Gaussian quadrature points of the hexahedral element can be viewed as a collection of Gaussian quadrature points of each 1D segment.

On account of the nonlinearity in this fiber amplifier problem (the Raman gain term), direct use of substructuring/templating approaches to speed-up element computations are not possible. In other words, the geometry of the fiber, which remains invariant in the longitudinal direction, cannot be used directly for recycling the stiffness or Gram matrices since they change with each nonlinear solve iteration. Other ideas such as rank-1 updates between the nonlinear iterations would also not be acceptable, since one would then be losing the true adjoint graph norm in the ultraweak formulation. Thus, the sum factorization method offers the best approach for improving the computational efficiency of the DPG implementation. Also, note that the computational improvement gained from using the sum factorization approach generally increases as the polynomial order (p) increases.

Sum factorization, first developed in [58] and later extended in [62], can be implemented for isoparametric elements, thus achieving the computational benefits even in problems with curvilinear geometries, which is critical for the fiber amplifier application. The ultraweak formulation and having discontinuities across elements in the L^2 field variables, permits the use of the sum factorization scheme for the stiffness matrices as well as the Gram matrices.

In contrast, the primal formulation would have made implementing the sum factorization methodology for the stiffness matrices significantly more complicated by the need to account for the orientations of the conforming $H(\text{curl})$ trial variables used in such a formulation of Maxwell's equations.

Since the hexahedral element is fully tensor, benefits are maximized by exclusively using this element type throughout the mesh domain. While the sum factorization approach can be extended to prismatic elements, it would not only be more difficult to implementation, but also one cannot guarantee the same speed-up since the prismatic element has a tensor product structure only in one direction, not all three. The use of sum factorization technique, with polynomial order $p = 5$ and hexahedral elements, in this fiber model has resulted in a computational speed-up of 80 times. It is worth noting that the many of simulations reported in this problem for longer fiber lengths, specifically over 32 wavelengths, would have been prohibitive because of the time required for element computations without this sum factorization method.

Appendix D

Comparing Primal and Ultraweak Maxwell Formulations

Author contributions: The contents of this chapter are taken largely from the multi-author article "A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers" S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, J. Mora, in preparation. The article has not yet been submitted for journal publication. The author of this dissertation contributed to model development, and code/numerical implementation of the model and analysis of the results.

This appendix[1] consists of two parts. First, we provide a brief, yet thorough, definitions of the primal and ultraweak variational formulations of the time-harmonic forms of Maxwell's equations. Second, we provide numerical evidence comparing the primal and ultraweak formulations of the Maxwell system, validating that the ultraweak formulation has superior performance.

[1]The content of this appendix is taken from the manuscript "A 3D DPG Maxwell approach to nonlinear Raman gain in fiber laser amplifiers" S. Nagaraj, J. Grosek, S. Petrides, L. Demkowicz, J. Mora, in preparation. Information approved for public release on 08 May 2018 by AFRL OPSEC/PA OPS-18-19547.

D.1 Primal vs. Ultraweak Formulations

We devote this section to the comparison of two (distinct) formulations of the time harmonic Maxwell system: the primal and ultraweak formulations. We assume an ansatz of the electromagnetic fields of the form $\mathbb{E}_0(x,y,z,t) = \mathbb{E}(x,y,z)e^{i\omega t}$ and $\mathbb{H}_0(x,y,z,t) = \mathbb{H}(x,y,z)e^{i\omega t}$ where $\omega > 0$ is the non-dimensionalized propagating frequency and t is time. As usual, ϵ, μ will represent non-dimensionalized electric permittivity and magnetic permeability. Moreover, we let $\sigma \in \mathbb{R}$ be a (possibly non-zero) conductivity.

D.1.1 Primal Formulation

The primal formulation corresponds to the case where

$$X_0 = Y_0 = H_0(\text{curl}, \Omega), \hat{X} = H^{-1/2}(\text{curl}, \partial\Omega_h), Y = H(\text{curl}, \Omega_h).$$

The bilinear forms are:

$$b_0(\mathbb{E}, \mathbb{F}) = (\nabla \times \mathbb{E}, \nabla \times \mathbb{F})_h + ((i\omega\mu\sigma - \omega^2\mu\epsilon)\mathbb{E}, \mathbb{F})_h,$$

$$\hat{b}(\hat{\mathbb{E}}, \mathbb{F}) = \langle n \times \hat{\mathbb{E}}, \mathbb{F} \rangle_h.$$

The primal formulation is thus a broken version of the standard Bubnov-Galerkin formulation. The test space is given the standard (or mathematician's) norm, i.e.,

$$\|v\|_Y^2 := \|\mathbb{F}\|^2 + \|\nabla \times \mathbb{F}\|^2.$$

D.1.2 Ultraweak Formulation

Recall the definition of the Maxwell operator

$$A\begin{pmatrix}\mathbb{E}\\ \mathbb{H}\end{pmatrix} := \begin{pmatrix}-(i\omega\epsilon+\sigma) & \nabla\times\\ -\nabla\times & -i\omega\mu\end{pmatrix}\begin{pmatrix}\mathbb{E}\\ \mathbb{H}\end{pmatrix}$$

with adjoint A^*:

$$A^*\begin{pmatrix}\mathbb{E}\\ \mathbb{H}\end{pmatrix} := \begin{pmatrix}(i\omega\epsilon-\sigma) & -\nabla\times\\ \nabla\times & i\omega\mu\end{pmatrix}\begin{pmatrix}\mathbb{E}\\ \mathbb{H}\end{pmatrix}$$

If $\sigma = 0$, we have $A^* = -A$. The ultraweak formulation corresponds to the case where

$$X_0 = \mathbb{L}^2(\Omega)\times\mathbb{L}^2(\Omega), \hat{X} = H^{-1/2}(\mathrm{curl}, \partial\Omega_h)\times H^{-1/2}(\mathrm{curl}, \partial\Omega_h),$$

$$Y_0 = H(\mathrm{curl}, \Omega)\times H_0(\mathrm{curl}, \Omega), Y = H(\mathrm{curl}, \Omega_h)\times H(\mathrm{curl}, \Omega_h).$$

Denote by $u = (\mathbb{E}, \mathbb{H}) \in X_0$, $\hat{u} = (\hat{\mathbb{E}}, \hat{\mathbb{H}}) \in \hat{X}$ and $v = (\mathbb{R}, \mathbb{S}) \in Y$. The bilinear forms corresponding to the ultraweak formulation are:

$$b_0(u, v) = (u, A^*v)_h,$$

$$\hat{b}(\hat{u}, v) = \langle n\times\hat{\mathbb{E}}, \mathbb{R}\rangle_h + \langle n\times\hat{\mathbb{H}}, \mathbb{S}\rangle_h,$$

equipped with the scaled adjoint graph norm:

$$\|v\|_Y^2 := \alpha\|v\|^2 + \|A^*v\|^2.$$

It is well-known that for scalar wave propagation problems, the ultraweak formulation with the scaled adjoint graph has superior pre-asymptotic behaviour and we are guaranteed a robust estimate of the approximation error [69, 26]. While this was known to be the case for the Maxwell system as well, we provide now numerical evidence to support this claim.

D.1.3 Energy Norm Projection and Pollution Studies

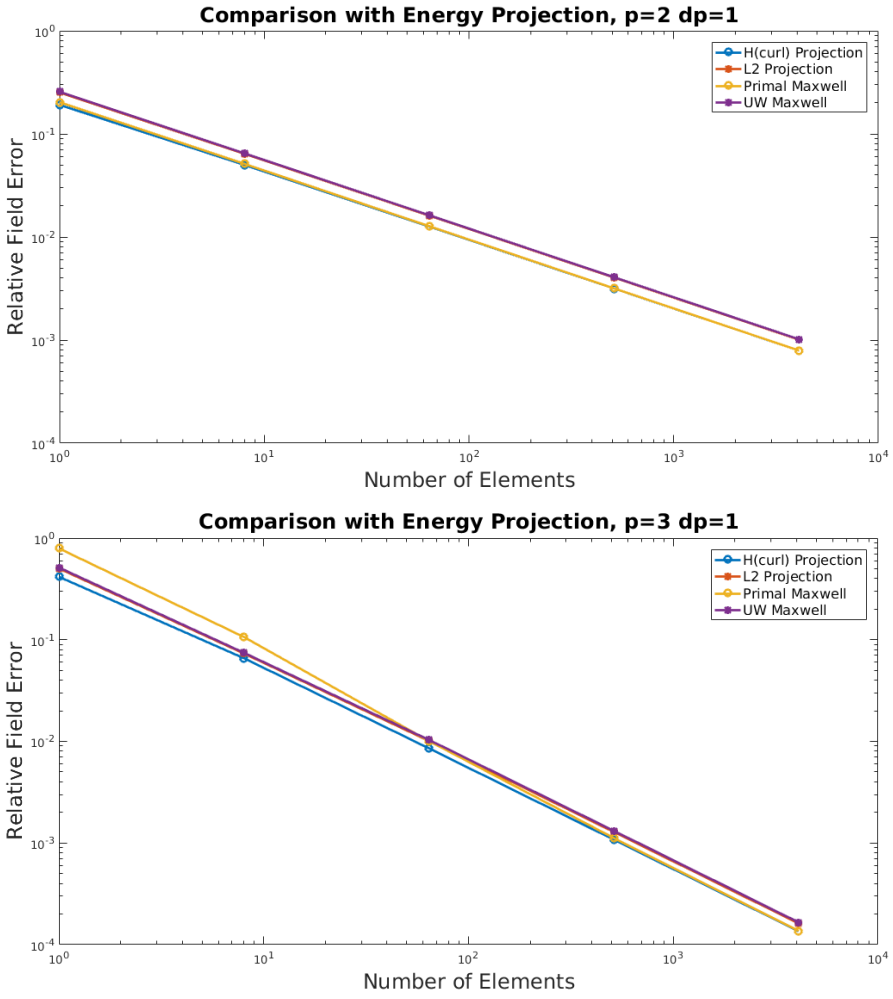

Figure D.1: Energy Norm Projections

In order to facilitate numerical comparisons of the primal and ultra-weak formulations, we consider two regimes of operations. The pre- and post-asymptotic regimes. In the pre-asymptotic regime, the propagating wave is not resolved (i.e., there are not enough degrees of freedom to capture the propa-

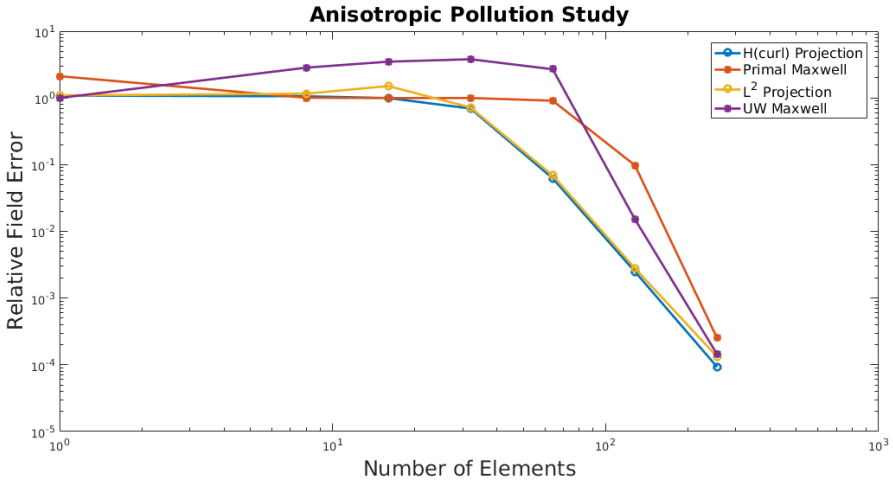

Figure D.2: Pollution Study

gation) while in the post-asymptotic regime, the wave is fully resolved. In all these cases, we assume $\sigma = 0$. Our theory indicates that the ultraweak formulation should have superior behaviour when compared with the primal formulation in both scenarios due to its use of the scaled adjoint graph norm. In the case of acoustics (Helmholtz) equation, a wavenumber explicit mathematical analysis of this behaviour is possible [26], while no such theory currently exists for the Maxwell system. The use of the scaled adjoint graph norm implies that the (ideal) unbroken ultraweak should, upon mesh refinement, deliver the L^2 projection in a robust fashion, while the primal has no such guarantees of convergence to the corresponding $H(\text{curl})$ projection. Figure D.1,D.2 show some comparisons between the primal and ultraweak formulations. The pollution study addresses the cases of pre-asymptotic behaviour. In this study, we consider a rectangular waveguide $\Omega = [0,1] \times [0,1] \times [0,16]$ and impose impedance

boundary conditions on the face $z = 16$ for both formulations. The waveguide was excited by the fundamental transverse electric (TE) mode with a non-dimensionalized frequency $\omega = \sqrt{5}\pi$, which corresponds to 16 wavelengths in the z-direction. We now anisotropically refine the waveguide only in the z-direction, and study how each formulation behaves vis-a-vis the corresponding energy norm projections with polynomial order $p = 5$. We note that with $p = 5$, we required roughly 4 elements per wavelength to resolve the wave. The choice of $p = 5$ was not arbitrary, yet, lower polynomial order will require significantly more refinements to achieve the same error levels. As expected, we see that the ultraweak formulation has superior pre-asymptotic behaviour. Indeed, the error of the ultraweak formulation coincides with the $\mathbb{L}^2$ projection error earlier, while the primal formulation is farther away from the $H(\text{curl})$ projection on the same mesh. In the energy norm projection study, we study the post-asymptotic behaviour. In this case $\Omega = [0, 1]^3$ and we use polynomial orders $p = 2, 3$ and perform uniform mesh refinements. Again, we see that the ultraweak formulation "catches up" to the $\mathbb{L}^2$ projection earlier (in terms of number of refinements) than the primal with the $H(\text{curl}$ projection. This means that the ultraweak formulation (with the scaled adjoint graph norm) delivers a solution closer to the $\mathbb{L}^2(\Omega)$ projection in both the pre- and post-asymptotic regime. Thus, Fig. D.1,D.2 provides numerical evidence that the ultraweak formulation is the better choice for the Maxwell system problem.

Appendix E

The L^p IRLS Algorithm

In this appendix, we consider the minimization of $\|Au - f\|_{L^p}$ for a continuous operator $A : U \to L^p$ with U a reflexive Banach space and $f \in L^p$ an L^p bounded load. We now describe our L^p IRLS algorithm. The original IRLS algorithm has a long history and has been used in signal processing applications for a number of years, most recently in sparse signal processing [19, 37]. Our approach to the Banach space version is based mainly on [76]. In particular, our main construction, proof methods and results are based on the l^p (finite dimensional) version presented in [76]. Indeed, this appendix is an attempt to generalize the construction in [76] to the L^p case.

In all that follows, p will denote a fixed real number, $p \in (2, \infty)$ and q its Hölder conjugate:

$$\frac{1}{p} + \frac{1}{q} = 1,$$

and we define

$$r = \frac{p}{p-2}$$

Let $\Omega \subset \mathbb{R}^d$ be a finite-measure Lipschitz domain in d dimensions, $d = 2, 3$. Also, $\mathbb{R}^+$ shall mean the interval $(0, \infty)$ and $\alpha \in \mathbb{R}^+$ will be a fixed constant.

We deal with the following spaces:

$$L^p(\Omega) := \{u : \Omega \to \mathbb{R} \,|\, \int_\Omega |u|^p < \infty\}.$$

Henceforth, since there will be no confusion, we shall suppress mentioning Ω explicitly in the above spaces, and all integrals will be over the domain Ω. By I_Ω we shall mean the indicator function on Ω. With $p > 2$ and Ω bounded, we have $L^p \subset L^2$ and hence, there is a constant D such that $\|f\|_{L^2} \leq D\|f\|_{L^p}$ for all $f \in L^p$. Consider a continuous linear operator $A : U \to L^p$ where U is assumed to be a reflexive Banach space. We make the following assumption:

Assumption 1: We assume A is L^2 bounded from below, i.e., there is a $\gamma > 0$ so that

$$\gamma\|u\|_U \leq \|Au\|_{L^2}$$

for all $u \in U$.

Note that this implies that A is L^p bounded from below as well. By the Banach closed range theorem, we are guaranteed a unique, bounded solution to $Au = f$ for any $f \in L^p$. In other words, there is a unique, bounded u^* such that:

$$u^* = \text{argmin}_{v \in U}\|Av - f\|_{L^p}.$$

Now, our aim is to replace the minimization over L^p by a weighted minimization over L^2. To this end, we observe:

$$\int |Av - f|^p = \int |Av - f|^{p-2}|Av - f|^2.$$

If we regard $w = |Av - f|^{p-2}$ as a weight function, we have, formally,

$$\|Av - f\|_{L^p}^p = \|Av - f\|_{L^2_w}^2,$$

where L^2_w is the weighted Hilbert space:

$$L^2_w = \{u : \Omega \to \mathbb{R} \,|\, \int w|u|^2 < \infty\}.$$

The idea of IRLS is to construct a sequence of weights $w_{n+1} = |Au_n - f|$ and solve the updated least squares problem for u_{n+1}, namely,

$$u_{n+1} = \text{argmin}_{v \in U} \|Av - f\|_{L^2_{w_{n+1}}}.$$

The hope is that $u_n \to u$, with u being the L^p minimizer. In order to establish this convergence, one sets up a functional depending on the current weight, solution and an auxiliary continuation parameter.

In finite dimensions, where strong and weak convergence coincide, the results of [76] indicate the conditions under which this hope can be realized. As we shall see in the infinite dimensional case, we will be forced to deal with weak convergence.

We now define the functional that shall be the main work horse of our method. Note the similarity with the one presented in [76]. We define:

$$\mathcal{J}(u, w, \epsilon^2) : U \times L^r \times \mathbb{R}^+ \to \mathbb{R}$$

as:

$$\mathcal{J}(u, w, \epsilon^2) = \int w(|Au - f|^2 + \epsilon^2 I_\Omega) - \int \frac{\alpha}{r} w^r.$$

Since $Au - f \in L^p$, we have $|Au - f|^2 \in L^{\frac{p}{2}}$, and, with $r = \frac{p}{p-2}$ the term $\int w|Au - f|^2$ is to be understood as the duality pairing between L^r and $L^{\frac{p}{2}}$.

We refer to the sequence of u_n as the sequence of solutions, the w_n as the

Algorithm 2 L^pIRLS

1: Set $\epsilon_1^2 = 1$, $w_1 = I_\Omega$. For $n = 1, 2, \ldots$:
2: Solve for $u_{n+1} = \text{argmin}_{u \in U} \mathcal{J}(u, w_n, \epsilon_n^2)$
3: Set $\epsilon_{n+1}^2 = \min(\epsilon_n^2, k_{n+1}\|Au_{n+1} - f\|^2_{L^2_{w_n}})$
4: Solve for $w_{n+1} = \text{argmin}_{w \in L^r} \mathcal{J}(u_{n+1}, w, \epsilon_{n+1}^2)$

sequence of weights and the ϵ_n as the sequence of relaxation parameters. In the sequel, we will need to compare the weighted L^2 norms and the L^2 norms of the residual. We make the following assumption:

Assumption 2: In algorithm 2 , we assume that there exist constants B, C such that for each $n \geq 1$ the w_n weighted L^2 norm is equivalent to the L^2 norm:

$$C\|v\|_{L^2} \leq \|v\|_{L^2_{w_n}} \leq B\|v\|_{L^2},$$

for all $v \in L^p$.

We have the following lemmas:

Lemma E.0.1. *In algorithm 2 ,*

$$w_{n+1} = (\frac{1}{\alpha}(|Au_{n+1} - f|^2 + \epsilon_{n+1}^2 I_\Omega))^{\frac{p-2}{2}}.$$

In particular, w_{n+1} is a valid non-negative weight for all n.

Proof We compute the first variation of $\mathcal{J}$ with fixed u_n in the direction δw:

$$D_w\mathcal{J}(\delta w) = \frac{d}{ds}\mathcal{J}(u_{n+1}, w+\delta w, \epsilon_{n+1}^2)|_{s=0} = \int \delta w(|Au_{n+1}-f|^2+\epsilon_{n+1}^2 I_\Omega)-\alpha \int w^{r-1}\delta w.$$

Thus, setting the first variation to 0, we obtain, for all δw:

$$\int \delta w(|Au_{n+1} - f|^2 + \epsilon_{n+1}^2 I_\Omega - \alpha w^{r-1}) = 0.$$

Notice that $r - 1 = \frac{2}{p-2}$. By the Hahn-Banach theorem, we conclude:

$$w_{n+1} = (\frac{1}{\alpha}(|Au_{n+1} - f|^2 + \epsilon_{n+1}^2 I_\Omega))^{\frac{p-2}{2}}.$$

■

Lemma E.0.2. *In algorithm 2 above, u_{n+1} is the weighted least squares solution of $Au = f$ with weight w_n, i.e.,*

$$(Au_{n+1}, A\,\delta u)_{L^2_{w_n}} = (f, A\delta u)_{L^2_{w_n}},$$

for all $\delta u \in U$

Proof We compute the first variation of $\mathcal{J}$ with fixed w_n in the direction δu:

$$D_u\mathcal{J}(\delta u) = \frac{d}{ds}\mathcal{J}(u + s\delta u, w_n, \epsilon_n^2)|_{s=0} = \int 2w_n(Au - f)A\delta u,$$

for all δu. Upon equating to 0, we obtain the variational equation for u_{n+1}:

$$\int w_n A u_{n+1}\, A\delta u = \int w_n f A\delta u.$$

In other words:

$$(Au_{n+1}, A\,\delta u)_{L^2_{w_n}} = (f, A\delta u)_{L^2_{w_n}},$$

for all $\delta u \in U$ ■

Lemma E.0.3. *We have the following expression for the functional* $\mathcal{J}(u_{n+1}, w_{n+1}, \epsilon_{n+1}^2)$*:*

$$\mathcal{J}(u_{n+1}, w_{n+1}, \epsilon_{n+1}^2) = \frac{2}{p\alpha^{\frac{p-2}{2}}} \int (|Au_{n+1} - f|^2 + \epsilon_{n+1}^2 I_\Omega)^{\frac{p}{2}}$$

Proof We have

$$\mathcal{J}(u_{n+1}, w_{n+1}, \epsilon_{n+1}^2) = \int w_{n+1}(|Au_{n+1} - f|^2 + \epsilon_{n+1}^2 I_\Omega) - \frac{\alpha}{r} \int w_{n+1}^r.$$

By lemma 1,

$$w_{n+1} = (\frac{1}{\alpha}(|Au_{n+1} - f|^2 + \epsilon_{n+1}^2 I_\Omega))^{\frac{p-2}{2}}.$$

Thus,

$$\begin{aligned} \mathcal{J}(u_{n+1}, w_{n+1}, \epsilon_{n+1}^2) = & \frac{1}{\alpha^{\frac{p-2}{2}}} \int (|Au_{n+1} - f|^2 + \epsilon_{n+1}^2 I_\Omega)^{\frac{p-2}{2}+1} \\ & - \frac{\alpha}{r} \int (\frac{1}{\alpha}(|Au_{n+1} - f|^2 + \epsilon_{n+1}^2 I_\Omega))^{(\frac{p-2}{2})r}. \end{aligned} \tag{E.0.1}$$

Notice that $(\frac{p-2}{2})r = \frac{p}{2}$. Thus,

$$\begin{aligned} \mathcal{J}(u_{n+1}, w_{n+1}, \epsilon_{n+1}^2) &= \frac{1}{\alpha^{\frac{p-2}{2}}} \int (|Au_{n+1} - f|^2 + \epsilon_{n+1}^2 I_\Omega)^{\frac{p}{2}} (1 - \frac{p-2}{p}) \\ &= \frac{2}{p\alpha^{\frac{p-2}{2}}} \int (|Au_{n+1} - f|^2 + \epsilon_{n+1}^2 I_\Omega)^{\frac{p}{2}} \end{aligned} \tag{E.0.2}$$

■

Lemma E.0.4. *The functional* $\mathcal{J}(u_n, w_n, \epsilon_n^2)$ *is monotonic in the following sense:*

$$\mathcal{J}(u_{n+1}, w_{n+1}, \epsilon_{n+1}^2) \leq \mathcal{J}(u_n, w_n, \epsilon_n^2)$$

Proof By the definition of w_{n+1} as the minimizer of $\mathcal{J}(u_{n+1}, \cdot, \epsilon_{n+1}^2)$, we have

$$\mathcal{J}(u_{n+1}, w_{n+1}, \epsilon_{n+1}^2) \leq \mathcal{J}(u_{n+1}, w_n, \epsilon_{n+1}^2).$$

By $\epsilon_{n+1}^2 \leq \epsilon_n^2$ we have

$$\mathcal{J}(u_{n+1}, w_n, \epsilon_{n+1}^2) \leq \mathcal{J}(u_{n+1}, w_n, \epsilon_n^2).$$

Finally, by the definition of u_{n+1} as the minimizer of $\mathcal{J}(\cdot, w_n, \epsilon_n^2)$, we conclude

$$\mathcal{J}(u_{n+1}, w_n, \epsilon_n^2) \leq \mathcal{J}(u_n, w_n, \epsilon_n^2).$$

■

In particular, we see that the sequence $\{\mathcal{J}(u_n, w_n, \epsilon_n^2)\}$ is bounded and $\mathcal{J}(u_n, w_n, \epsilon_n^2) \leq \mathcal{J}(u_1, w_1, \epsilon_1^2)$ for all $n = 1, 2, \ldots$.

Lemma E.0.5. *The sequence of iterates u_n is bounded, i.e., there exists an $M > 0$ such that for all $n = 1, 2, \ldots$, the vectors u_n lie in $B_U(0, M)$, the ball of radius M in U.*

Proof Let u^* be the unique solution of $Au = f$. Recall for all $u \in U$ we have $\gamma \|u\|_U \leq \|Au\|_{L^p}$. Now,

$$\begin{aligned} \|u_n\|_U \leq \|u_n - u^*\|_U + \|u^*\|_U &\leq \frac{1}{\gamma}(\|Au_n - Au^*\|_{L^p}) + \|u^*\|_U \\ &\leq \frac{1}{\gamma}(\|Au_n - f\|_{L^p} + \|Au^* - f\|_{L^p}) + \|u^*\|_U \\ &= \frac{1}{\gamma}\|Au_n - f\|_{L^p} + \|u^*\|_U. \end{aligned} \tag{E.0.3}$$

Notice now that

$$\begin{aligned}\|Au_n - f\|_{L^p}^p = \int |Au_n - f|^p &= \int (|Au_n - f|^2)^{\frac{p}{2}} \\ &\leq \int (|Au_n - f|^2 + \epsilon_n^2 I_\Omega)^{\frac{p}{2}} \\ &= \frac{p\alpha^{\frac{p-2}{2}}}{2} \mathcal{J}(u_n, w_n, \epsilon_n^2)(\text{ by lemma 1}) \\ &\leq \frac{p\alpha^{\frac{p-2}{2}}}{2} \mathcal{J}(u_1, w_1, \epsilon_1^2),\end{aligned} \tag{E.0.4}$$

thus, $\|Au_n - f\|_{L^p} \leq (\frac{p\alpha^{\frac{p-2}{2}}}{2} \mathcal{J}(u_1, w_1, \epsilon_1^2))^{\frac{1}{p}}$. Thus,

$$\|u_n\|_U \leq M := \frac{1}{\gamma}(\frac{p\alpha^{\frac{p-2}{2}}}{2} \mathcal{J}(u_1, w_1, \epsilon_1^2))^{\frac{1}{p}} + \|u^*\|_U. \tag{E.0.5}$$

Thus $u_n \in B_U(0, M)$ for all $n = 1, 2, \ldots$. ■

Lemma E.0.6. *The functional $\mathcal{J}(u, w_n, \epsilon_n^2)$ is uniformly convex in u for fixed w_n, ϵ_n^2 with constant $2C^2\gamma^2$.*

Proof Strong convexity of $\mathcal{J}(\cdot, w_n, \epsilon_n^2)$ is equivalent to the condition that the second variation of $\mathcal{J}(\cdot, w_n, \epsilon_n^2)$ is positive definite. We have already computed the first variation of $\mathcal{J}(\cdot, w_n, \epsilon_n^2)$ earlier:

$$D_u \mathcal{J}(\delta u) = (2Au - f, A\delta u)_{L^2_{w_n}}.$$

The second variation is:

$$D_u^2 \mathcal{J}(\delta u, \delta u) = 2(A\delta u, A\delta u)_{L^2_{w_n}} \geq 2C^2 \|A\delta u\|_{L^2}^2 > 2C^2\gamma^2 \|\delta u\|_U^2$$

for all nonzero $\delta u \in U$.

Note that since $\mathcal{J}(u, w_n, \epsilon_n^2)$ is uniformly convex in u, we can conclude that

$$\mathcal{J}(u, w_n, \epsilon_n^2) - \mathcal{J}(v, w_n, \epsilon_n^2) \geq 2C^2\gamma^2\|u - v\|_U^2,$$

for all $u, v \in U$. ■

Lemma E.0.7. *The terms of the sequence $\{u_n\}$ generated by algorithm 2 satisfy*

$$\|u_{n+1} - u_n\|_U \to 0.$$

Proof By the monotonocity of $\mathcal{J}(u_n, w_n, \epsilon_n^2)$, we have

$$\begin{aligned} |\mathcal{J}(u_n, w_n, \epsilon_n^2) - \mathcal{J}(u_{n+1}, w_{n+1}, \epsilon_{n+1}^2)| &\geq |\mathcal{J}(u_n, w_n, \epsilon_n^2) - \mathcal{J}(u_{n+1}, w_n, \epsilon_n^2)| \\ &\geq 2C^2\gamma^2\|u_{n+1} - u_n\|_U^2. \end{aligned} \tag{E.0.6}$$

Since

$$|\mathcal{J}(u_n, w_n, \epsilon_n^2) - \mathcal{J}(u_{n+1}, w_{n+1}, \epsilon_{n+1})| \to 0,$$

we conclude $\|u_{n+1} - u_n\|_U \to 0$. ■

Definition E.0.8. For $\epsilon > 0$, the ϵ- perturbed L^p residual $f_\epsilon(u) : L^p \to \mathbb{R}$ is defined as:

$$f_\epsilon(u) = \frac{2}{p\alpha^{\frac{p-2}{2}}} \int (|Au - f|^2 + \epsilon^2 I_\Omega)^{\frac{p}{2}}$$

Definition E.0.9. The (ϵ, v)- weight $w(\epsilon, v) \in L^r$ is defined as:

$$w(\epsilon, v) = (|Av - f|^2 + \epsilon^2 I_\Omega)^{\frac{p-2}{2}}.$$

Lemma E.0.10. *If*

$$\int |Av - f|^2 w(\epsilon, v) \leq \|A\widetilde{v} - f\|^2_{L^2_{w(\epsilon,v)}}$$

for all $\widetilde{v} \in U$*, then* $v \in \ argmin_{z \in U} f_\epsilon(z)$

Proof We have

$$\int |Av - f|^2 w(\epsilon, v) \leq \|A\widetilde{v} - f\|^2_{L^2_{w(\epsilon,v)}},$$

so that

$$\int (|Av-f|^2+\epsilon^2 I_\Omega) w(\epsilon, v) = \int (|Av-f|^2+\epsilon^2 I_\Omega)^{\frac{p}{2}} \leq \int (|A\widetilde{v}-f|^2+\epsilon^2 I_\Omega) w(\epsilon, v).$$

Notice that $|A\widetilde{v} - f|^2 + \epsilon^2 I_\Omega \in L^{\frac{p}{2}}$ and $w(\epsilon, v) \in L^r$ with $r^* + (\frac{p}{2})^* = 1$, so by Hölder's inequality, we have:

$$\int (|Av - f|^2 + \epsilon^2 I_\Omega) w(\epsilon, v) \leq \|(A\widetilde{v} - f)^2 + \epsilon^2 I_\Omega\|_{L^{\frac{p}{2}}} \|w(\epsilon, v)\|_{L^r}.$$

Now since $r = \frac{p}{p-2}$, we have:

$$\|w(\epsilon, v)\|_{L^r} = (\int (|Av - f|^2 + \epsilon^2 I_\Omega)^{\frac{p-2}{2} \frac{p}{p-2}})^{\frac{p-2}{p}} = (\int (|Av - f|^2 + \epsilon^2 I_\Omega)^{\frac{p}{2}})^{\frac{p-2}{p}}.$$

Thus, we have:

$$(\int (|Av - f|^2 + \epsilon^2 I_\Omega)^{\frac{p}{2}})^{1-\frac{p-2}{p}} \leq (\int (|A\widetilde{v} - f|^2 + \epsilon^2 I_\Omega)^{\frac{p}{2}})^{\frac{2}{p}}.$$

Raising both sides to the $\frac{p}{2}$ power and multiplying by the factor $\frac{2}{p\alpha^{\frac{p-2}{2}}}$, we see that

$$f_\epsilon(v) \leq f_\epsilon(\widetilde{v})$$

for all $\widetilde{v} \in L^p$. ■

As of now, we are able to conclude only that the difference of consecutive terms of the sequence $\{u_n\}$ generated by algorithm 2 get closer and closer. In general, this is, in the infinite dimensional case, not sufficient to guarantee convergence. We therefore make one additional assumption on the operator A:

Assumption 3: Let $\mathcal{S} := \{u_n\}$, the set consisting of the iterates of algorithm 2 . We assume that for every weakly convergent subsequence u_{n_k} of u_n, with $u_{n_k} \rightharpoonup \overline{u}$ there exists a further subsequence $u_{n_{k_s}}$ of u_{n_k} such that $Au_{n_{k_s}}$ converges strongly in L^p to $A\overline{u}$.

With this assumption, we state the following theorem:

Theorem E.0.11. *Let $f \in L^p$ and $A : U \to L^p$ be a continuous operator satisfying the assumptions 1-3 made earlier. Let u_n, w_n, ϵ_n be the sequence of iterates, weights and relaxation parameters generated by algorithm 2 . Let u^* be the unique minimizer in U of $\|Au - f\|_{L^p}$. Then, if $\epsilon_n \to 0$, then there is a subsequence of u_n which converges to u^*.*

Proof Consider first the case $\epsilon_n \to 0$. If $\epsilon_N = 0$ for some $N > 0$, with $\epsilon_{N-1} \neq 0$, then, by definition of ϵ_N, we have $\|Au_N - f\|_{L^p} = 0$. However, since u^* is the unique minimizer in U of $\|Au - f\|_{L^p}$, it follows that $u_N = u^*$.

Next, consider the case $\epsilon_n > 0 \, \forall n$, with $\epsilon_n \to 0$. Since u_n, is a bounded sequence in the reflexive Banach space U, there is a weakly convergent sequence

$u_{n_k} \rightharpoonup \overline{u}$ of u_n. By assumption 3, there is a further subsequence $u_{n_{k_s}}$ of u_{n_k} such that $Au_{n_{k_s}} \to A\overline{u}$ in L^p. Now, $\|A\overline{u} - f\|_{L^p} \leq \|A\overline{u} - Au_{n_{k_s}}\|_{L^p} + \|Au_{n_{k_s}} - f\|_{L^p}$. The term $\|A\overline{u} - Au_{n_{k_s}}\|_{L^p} \to 0$ by assumption 3, while the second term also approaches 0 by the assumption that $\epsilon_n \to 0$. Indeed, if $\epsilon_n \to 0$, then $\epsilon_{n_{k_s}} \to 0$, and hence, by the definition of $\epsilon_{n_{k_s}}$, we must have $\|Au_{n_{k_s}} - f\|_{L^p} \to 0$. Thus, $\|A\overline{u} - f\|_{L^p} = 0$, and by uniqueness of the minimizer, we have $\overline{u} = u^*$. ■

Bibliography

[1] G. P. Agrawal. *Applications of Nonlinear Fiber Optics, Second Edition.* Academic Press, Waltham, Massachusetts, USA, 2008.

[2] G. P. Agrawal. *Nonlinear Fiber Optics, Fifth Edition.* Academic Press, Waltham, Massachusetts, USA, 2012.

[3] I. Babuska. Error-bounds for finite element method. *Numerische Mathematik*, 16:322–333, 1971.

[4] A. Bahl, A. Teleki, P. K. Jakobsen, E. M. Wright, and M. Kolesik. Reflectionless beam propagation on a piecewise linear complex domain. *Journal of Lightwave Technology*, 32(22):3670–3676, 2014.

[5] A. T. Barker, V. Dobrev, J. Gopalakrishnan, and T. Kolev. A scalable preconditioner for a primal discontinuous Petrov-Galerkin method. *SIAM Journal on Scientific Computing*, 40:A1187–A1203, 2016.

[6] J. P. Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200, 1994.

[7] R. W. Boyd. *Nonlinear Optics, Third Edition.* Academic Press, Waltham, Massachusetts, USA, 2008.

[8] J. Bramwell. *A discontinuous Petrov-Galerkin method for seismic tomography problems*. PhD thesis, the University of Texas at Austin, 5 2013.

[9] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods (Springer Series in Computational Mathematics)*. Springer 1st ed., Providence, Rhode Island, USA, 1991.

[10] T. Bui-Thanh, L. F. Demkowicz, and O. Ghattas. A unified discontinuous Petrov-Galerkin method and its analysis for Friedrichs' systems. *SIAM Journal on Numerical Analysis*, 51:1933–1958, 2013.

[11] C. Carstensen, P. Bringmann, F. Hellwig, and P. Wriggers. Nonlinear discontinuous Petrov-Galerkin methods. *Numerische Mathematik*, 2018.

[12] C. Carstensen, L. F. Demkowicz, and J. Gopalakrishnan. A posteriori error control for DPG methods. *SIAM Journal on Numerical Analysis*, 52:1335–1353, 2014.

[13] C. Carstensen, L. F. Demkowicz, and J. Gopalakrishnan. Breaking spaces and forms for the DPG method and applications including Maxwell equations. *Computers and Mathematics With Applications*, 72, 2016.

[14] C. Carstensen and F. Hellwig. Low-order DPG-FEMs for linear elasticity. *SIAM Journal on Numerical Analysis*, 54:3388–3410, 2015.

[15] J. Chan, L. F. Demkowicz, and R. Moser. A DPG method for steady viscous compressible flow. *Computers and Fluids*, 98:69–90, 2014.

[16] W. C. Chew and W. H. Weedon. A 3d perfectly matched medium from modified Maxwell's equations with stretched coordinates. *Microwave and Optical Technology Letters.*, 7:599–604, 1994.

[17] Ph. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. Society for Industrial and Applied Mathematics (SIAM), 2 ed., 2002.

[18] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for 1st order transport equations. *SIAM Journal on Numerical Analysis*, 50:2420–2445, 2012.

[19] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gunturk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63:1–38, 2009.

[20] R. Dautray and J. L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology: Volume 5 Evolution Problems I*. Springer, 2013.

[21] L. F. Demkowicz. Polynomial exact sequences and projectioin-based interpolation with application to maxwell equations. *Mixed Finite Elements, Compatibility Conditions and Applications, vol. 1939 Lecture Notes in Mathematics*, 1939:101–158, 2006.

[22] L. F. Demkowicz. Various variational formulations and closed range theorem. *ICES Report, The Institute for Computational Engineering and Sciences, The University of Texas at Austin*, 15-03, 2015.

[23] L. F. Demkowicz and J. Gopalakrishnan. Analysis of the DPG method for the Poisson equation. *SIAM Journal on Numerical Analysis*, 49:1788–1809, 2011.

[24] L. F. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. part II: Optimal test functions. *Numerical Methods for Partial Differential Equations*, 27:70–105, 2011.

[25] L. F. Demkowicz and J. Gopalakrishnan. An overview of the discontinuous petrov galerkin method. *Feng X., Karakashian O., Xing Y. (eds) Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations.*, The IMA Volumes in Mathematics and its Applications, vol 157:149–180, 2013.

[26] L. F. Demkowicz, J. Gopalakrishnan, I. Muga, and J. Zitelli. Wavenumber explicit analysis for a DPG method for the multidimensional Helmholtz equation. *Computer Methods in Applied Mechanics and Engineering*, 213/216:126–138, March 2012.

[27] L. F. Demkowicz, J. Gopalakrishnan, S. Nagaraj, and P. Sepulveda. A spacetime DPG method for the Schrodinger equation. *SIAM Journal on Numerical Analysis*, 55:1740–1759, 2017.

[28] L. F. Demkowicz, J. Kurtz, D. Pardo, M. Paszynski, W. Rachowicz, and A. Zdunek. *Computing with hp-Adaptive Finite Elements, volume 2.* Chapman and Hall, CRC, Boca Raton FL, 1st edition, 2008.

[29] W. Dorfler, S. Findeisen, and C. Wieners. Space-time discontinuous Galerkin discretizations for linear first-order hyperbolic evolution. *Computational Methods in Applied Mathematics*, 16:409–428, 2016.

[30] T. E. Ellis, J.L. Chan, and L. F. Demkowicz. Robust DPG methods for transient convection-diffusion. *In: Barrenechea G., Brezzi F., Cangiani A., Georgoulis E. (eds) Building Bridges: Connections and Challenges in Modern Approaches to Numerical Partial Differential Equations. Lecture Notes in Computational Science and Engineering*, Lecture Notes in Computational Science and Engineering book series (LNCSE, volume 114), Springer, Cham, 2016.

[31] T. E. Ellis, L. F. Demkowicz, and J. Chan. Locally conservative discontinuous PetrovGalerkin finite elements for fluid problems. *Computers and Mathematics with Applications*, 68:1530–1549, 2014.

[32] T. E. Ellis, L. F. Demkowicz, J. L. Chan, and R. D. Moser. Space-time DPG: Designing a method for massively parallel CFD. *ICES Report, The Institute for Computational Engineering and Sciences, The University of Texas at Austin*, 14-32, 2014.

[33] A. Ern and J.-L. Guermond. Discontinuous Galerkin methods for Friedrichs' systems. I. general theory. *SIAM Journal on Numerical Analysis*, 44:753–778, 2006.

[34] A. Ern, J.-L. Guermond, and G. Caplain. An intrinsic criterion for the

bijectivity of Hilbert operators related to Friedrichs' systems. *Communications in Partial Differential Equations*, 32:317–341, 2007.

[35] L. C. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics vol. 19, American Mathematical Society, Providence, Rhode Island, USA, 1998.

[36] K. Fan, W. Cai, and X. Ji. A full vectorial generalized discontinuous Galerkin beam propagation method (GDG-BPM) for nonsmooth electromagnetic fields in waveguides. *Journal of Computational Physics*, 227:7178–7191, 2008.

[37] M. Fornasier, H. Rauhut, and R. Ward. Low-rank matrix recovery via iteratively reweighted least squares minimization. *SIAM Journal on Optimization*, 21:1614–1640, 2011.

[38] K. Friedrichs. Symmetric positive linear differential equations. *Communications on Pure and Applied Mathematics*, 11:333–418, 1958.

[39] F. Fuentes, L. F. Demkowicz, and A. Wilder. Using a DPG method to validate DMA experimental calibration of viscoelastic materials. *Computer Methods in Applied Mechanics and Engineering*, 325:748–765, 2017.

[40] F. Fuentes, B. Keith, L. F. Demkowicz, and S. Nagaraj. Orientation embedded high order shape functions for the exact sequence elements of all shapes. *Computers and Mathematics with Applications*, 70:353–458, 2015.

[41] F. Fuentes, B. Keith, L. F. Demkowicz, and P. Le Tallec. Coupled variational formulations of linear elasticity and the DPG methodology. *Journal of Computational Physics*, pages 715–731, 2017.

[42] T. Fuhrer, N. Heuer, M. Karkulik, and R. Rodriguez. Combining the DPG method with finite elements. *Computational Methods in Applied Mathematics*, 2017.

[43] S. D. Gedney. An anisotropic perfectly matched layer absorbing media for the truncation of FDTD latices. *IEEE Transactions on Antennas and Propagation*, 44:1630–1639, 1996.

[44] I. M. Gelfand and N. J. Vilenkin. *Generalized Functions vol. 4: Some Applications of Harmonic Analysis*. Academic Press, New York, 1964.

[45] J. Gopalakrishnan. Five lectures on DPG methods. *ArXiv e-prints*, arXiv:1306.0557 [math.NA], 2014.

[46] J. Gopalakrishnan and W. Qiu. An analysis of the practical DPG method. *Mathematics of Computation*, 83:537–552, 2014.

[47] J. Gopalakrishnan and J. Schoberl. Degree and wavenumber [in]dependence of a Schwarz preconditioner for the DPG method. *In: Kirby R., Berzins M., Hesthaven J. (eds) Spectral and High Order Methods for Partial Differential Equations ICOSAHOM 2014*, Lecture Notes in Computational Science and Engineering, vol 106:257–265, 2015.

[48] J. Gopalakrishnan and P. Sepulveda. A spacetime DPG method for acoustic waves. *ArXiv e-prints*, arXiv:1709.08268 [math.NA], 2017.

[49] Weiping Huang, Chenglin Xu, S-T Chu, and Sujeet K Chaudhuri. The finite-difference vector beam propagation method: analysis and assessment. *Journal of Lightwave Technology*, 10(3):295–305, 1992.

[50] WP Huang and CL Xu. Simulation of three-dimensional optical waveguides by a full-vector beam propagation method. *IEEE journal of quantum electronics*, 29(10):2639–2649, 1993.

[51] I. Babuska and J.E. Osborn. Finite element-Galerkin approximation of the eigenvalues and eigenvectors of selfadjoint problems. *Mathematics of Computation*, 52:275–297, 1989.

[52] B. Keith, F. Fuentes, and L. F. Demkowicz. The DPG methodology applied to different variational formulations of linear elasticity. *Computer Methods in Applied Mechanics and Engineering*, pages 579–609, 2016.

[53] B. Keith, F. Fuentes, and L. F. Demkowicz. The DPG methodology applied to different variational formulations of linear elasticity. *Computer Methods in Applied Mechanics and Engineering*, 309:579–609, 2016.

[54] B. Keith, P. Knechtges, N. V. Roberts, S. Elgeti, M. Behr, and L. F. Demkowicz. An ultraweak DPG method for viscoelastic fluids. *Journal of Non-Newtonian Fluid Mechanics*, 247:107–122, 2017.

[55] B. Keith, S. Petrides, F. Fuentes, and L. F. Demkowicz. Discrete least-squares finite element methods. *Computer Methods in Applied Mechanics and Engineering*, pages 226–255, 2017.

[56] B. Keith, A. Vaziri Astaneh, and L. F. Demkowicz. Goal-oriented adaptive mesh refinement for non-symmetric functional settings. *ArXiv e-prints*, arXiv:1711.01996 [math.NA], 2017.

[57] H. Kidorf, K. Rottwitt, M. Nissov, M. Ma, and E. Rabarijaona. Pump interactions in a 100-nm bandwidth raman amplifier. *IEEE Photonics Technology Letters*, 11(5):530–532, 1999.

[58] J. Kurtz. *Fully automatic hp-adaptivity for acoustic and electromagnetic scattering in three dimensions*. PhD thesis, the University of Texas at Austin, 5 2007.

[59] H. Masuda and K. Kitamura. Highly sensitive raman gain coefficient measurement by detecting spontaneous raman scattering power for distributed raman amplification systems. *IEICE Communications Express*, 1:1–6, 2017.

[60] P. J. Matuszyk and L. F. Demkowicz. Parametric finite elements, exact sequences and perfectly matched layers. *Computational Mechanics*, 51:35–45, 2013.

[61] C. Michler, L. F. Demkowicz, J. Kurtz, and D. Pardo. Improving the

performance of perfectly matched layers by means of hp-adaptivity. *Numerical Methods for Partial Differential Equations*, 23:832–858, 2007.

[62] J. Mora and L. F. Demkowicz. Fast integration of DPG matrices based on tensorization. *ArXiv e-prints*, arXiv:1711.00984 [math.NA], 2017.

[63] R Andrew Motes, Sami A Shakir, and Richard W Berdine. An efficient scalar, non-paraxial beam propagation method. *Journal of Lightwave Technology*, 30(1):4–8, 2012.

[64] 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 [math.NA], 2015.

[65] S. Naderi, I. Dajani, J. Grosek, and T. Madden. Theoretical and numerical treatment of modal instability in high-power core and cladding-pumped Raman fiber amplifiers. *Optics Express*, 24:16550–16565, 2016.

[66] S. Naderi, I. Dajani, T. Madden, and C. Robin. Investigations of modal instabilities in fiber amplifiers through detailed numerical simulations. *Optics Express*, 21:16111–16129, 2013.

[67] S. Nagaraj, S. Petrides, and L. F. Demkowicz. Construction of DPG Fortin operators for second order problems. *Computers and Mathematics with Applications*, 74:1964–1980, 2017.

[68] K. Okamoto. *Fundamentals of Optical Waveguides 2nd Edition*. Academic Press, Waltham, Massachusetts, USA, 2005.

[69] S. Petrides and L. F. Demkowicz. An adaptive DPG method for high frequency time-harmonic wave propagation problems. *Computers and Mathematics with Applications*, 74:1999–2017, 2017.

[70] N. V. Roberts and J. Chan. A geometric multigrid preconditioning strategy for DPG system matrices. *Computers and Mathematics with Applications*, 74:2018–2043, 2017.

[71] K. Saitoh and M. Koshiba. Full-vectorial finite element beam propagation method with perfectly matched layers for anisotropic optical waveguides. *Journal of Lightwave Technology*, 19:405–413, 2001.

[72] K. Saitoh and M. Koshiba. Full-vectorial imaginary-distance beam propagation method based on a finite element scheme: application to photonic crystal fibers. *IEEE Journal of Quantum Electronics*, 38:927–933, 2002.

[73] C. Schwab and R. Stevenson. Space-time adaptive wavelet methods for parabolic evolution problems. *Mathematics of Computation*, 78:1293–1318, 2009.

[74] Sami A Shakir, Raymond Andrew Motes, and Richard W Berdine. Efficient scalar beam propagation method. *IEEE Journal of Quantum Electronics*, 47(4):486–491, 2011.

[75] J. K. Shaw. *Mathematical Principles of Optical Fiber Communication*. CBMS-NSF Regional Conference Series in Applied Mathematics (76). SIAM: Society for Industrial and Applied Mathematics, 2004.

[76] J. Sigl. Nonlinear residual minimization by iteratively reweighted least squares. *Computational Optimization and Applications*, 64:755–792, 2016.

[77] R. Stegeman, L. Jankovic, H. Kim, C. Rivero, G. Stegeman, K. Richardson, P. Delfyett, Y. Guo, A. Schulte, and T. Cardinal. Tellurite glasses with peak absolute raman gain coefficients up to 30 times that of fused silica. *Optics Letters*, 28(13):1126–1128, 2003.

[78] T. Tao. *Nonlinear Dispersive Equations*. CBMS Regional Conference Series in Mathematics vol. 106, Published for the Conference Board of the Mathematical Sciences, Washington, DC by the American Mathematical Society, 2006.

[79] F. L. Teixeira and W. C. Chew. General closed-form PML constitutive tensors to match arbitrary bianisotropic and dispersive linear media. *IEEE Microwave and Guided Wave Letters*, 8:223–225, 1998.

[80] A. Vaziri Astaneh, F. Fuentes, J. Mora, and L. F. Demkowicz. High-order polygonal discontinuous Petrov-Galerkin (PolyDPG) methods using ultraweak formulations. *Computer Methods in Applied Mechanics and Engineering*, 332:686–711, 2017.

[81] A. Vaziri Astaneh, B. Keith, and L. F. Demkowicz. On perfectly matched layers for discontinuous Petrov-Galerkin methods. *ArXiv e-prints*, arXiv:1804.04496 [math.NA], 2017.

[82] J. Verdeyen. *Laser Electronics*. Prentice Hall, New Jersey, 3rd edition, 1995.

[83] Benjamin G Ward. Modeling of transient modal instability in fiber amplifiers. *Optics express*, 21(10):12053–12067, 2013.

[84] Christian Wieners. The skeleton reduction for finite element substructuring methods. *In: Karasozen B., Manguoglu M., Tezer-Sezgin M., Goktepe S., Ugur O. (eds) Numerical Mathematics and Advanced Applications ENUMATH 2015.*, Lecture Notes in Computational Science and Engineering, vol 112. Springer, Cham, 2016.

Vita

Sriram Nagaraj was born in Mumbai, India. After completing high school at Padma Seshadri Bala Bhavan (K K Nagar) School in 2005, he entered the University of Texas at Dallas in Richardson, Texas. He received the degree of Bachelor of Science in Electrical Engineering (BSEE) from UT Dallas in May 2009. He then moved to Rice University in Houston, Texas and received the degree of Master of Science from Rice University in May 2011. He entered the Graduate School at the University of Texas at Austin in the fall semester of 2013.

Email Address: sriram@ices.utexas.edu

This manuscript was typed by the author.