ADVANCED DISCRETIZATIONS AND MULTIGRID METHODS FOR LIQUID CRYSTAL CONFIGURATIONS

by

David B. Emerson

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy in Mathematics

TUFTS UNIVERSITY

May 2015

Advisors: James H. Adler and Scott P. MacLachlan

Abstract

Liquid crystals are substances that possess mesophases with properties intermediate between liquids and crystals. Here, we consider nematic liquid crystals, which consist of rod-like molecules whose average pointwise orientation is represented by a unit-length vector, $\mathbf{n}(x, y, z) = (n_1, n_2, n_3)^T$. In addition to their self-structuring properties, nematics are dielectrically active and birefringent. These traits continue to lead to many important applications and discoveries. Numerical simulations of liquid crystal configurations are used to suggest the presence of new physical phenomena, analyze experiments, and optimize devices.

This thesis develops a constrained energy-minimization finite-element method for the efficient computation of nematic liquid crystal equilibrium configurations based on a Lagrange multiplier formulation and the Frank-Oseen free-elastic energy model. First-order optimality conditions are derived and linearized via a Newton approach, yielding a linear system of equations. Due to the nonlinear unit-length constraint, novel well-posedness theory for the variational systems, as well as error analysis, is conducted. The approach is shown to constitute a convergent and well-posed approach, absent typical simplifying assumptions. Moreover, the energy-minimization method and well-posedness theory developed for the free-elastic case are extended to include the effects of applied electric fields and flexoelectricity.

In the computational algorithm, nested iteration is applied and proves highly effective at reducing computational costs. Additionally, an alternative technique is studied, where the unit-length constraint is imposed by a penalty method. The performance of the penalty and Lagrange multiplier methods is compared. Furthermore, tailored trust-region strategies are introduced to improve robustness and efficiency. While both approaches yield effective algorithms, the Lagrange multiplier method demonstrates superior accuracy per unit cost. In addition, we present two novel, optimally scaling, multigrid approaches for these systems based on Vanka- and Braess-Sarazin-type relaxation. Both approaches outperform direct methods

and represent highly efficient and scalable iterative solvers.

Finally, a three-dimensional problem considering the effects of geometrically patterned surfaces is presented, which gives rise to a nonlinear anisotropic reaction-diffusion equation. Well-posedness is shown for the intermediate linearization systems of the proposed Newton linearization. The configurations under consideration are part of ongoing physics research seeking new bistable configurations induced by geometric nano-patterning.

This thesis is dedicated to my beautiful and loving wife, Caitlin Bailey, and my parents, David and Kathleen Emerson, without whom this would not have been possible.

Acknowledgments

There are a great many people who have contributed, in ways large and small, to my graduate experience, achievements, and academic growth, the culmination of which is the work presented in this thesis. I am deeply grateful for the support and encouragement that I have received from so many throughout my journey. No one achieves anything of substance alone, and this work reflects not only my endeavors but the selfless efforts of numerous individuals.

I would like to start by thanking my advisors James Adler and Scott MacLachlan. I am profoundly appreciative of their relentless support in my academic development and encouragement of my research. They created a positive, enjoyable, and productive atmosphere in which to learn and work. They have devoted more hours of their free time to research discussions, impromptu problem solving, and general advisement than I can ever hope to repay. I have benefited enormously from their consistent mentorship, both as a mathematician and a person. I also thank Tim Atherton for serving on my thesis committee as well as his guidance and enthusiasm throughout this work. Without his commitment to collaborative exploration this research would not have been possible. Moreover, his contribution to my understanding of physics has been invaluable. I look forward to working on future applications of his design.

I extend my thanks and appreciation to Tom Manteuffel for serving on my thesis committee and his support in this research. Discussing research and receiving encouragement from an individual of such stature has been a wonderful experience. His contributions to my development and the work presented below are immense. In addition, I thank Xiaozhe Hu for serving on my committee and for his own guidance throughout this year. I have benefitted in no small way from his expertise as a mathematician.

I owe many thanks to Tom Benson for tackling a myriad of scientific computing predicaments with me. His computational expertise has been extremely appreciated.

I am indebted to him for allowing us to modify his multigrid code for use with our algorithms, his consistent willingness to help when I ran into road blocks, and his impressive ability to work with Trilinos. Additionally, I thank Ludmil Zikatanov and Johnny Guzmán for their comments, ideas, and expertise. I would also like to extend my gratitude to my officemates, especially Meghan, Melody, Stephanie, and Jiani, for their technical and moral support, as well as their friendship throughout my time at Tufts.

I would like to acknowledge and thank my wife, Caitlin. Without her unwavering support and confidence I would be lost. Her continuous optimism and spirit will always inspire me. She is my greatest blessing. Finally and foremost, I thank my parents for their encouragement, their understanding, and their love. They have always set an incredible example and embodied a commitment to hard work. I thank them for encouraging me to follow and fight for my dreams. Their confidence and fierce support is the reason I am who I am today.

Contents

List of Tables

List of Figures

Advanced Discretizations and Multigrid Methods for Liquid Crystal Configurations

Chapter 1

Introduction

Modern scientific research and applications require large-scale computational simulation of physical phenomena. Such simulations are used in many different ways, including predictive analysis, exploratory design, and theory validation. In order to carry out high-fidelity numerical simulations, accurate, efficient, and robust numerical methods are necessary. This thesis focuses on the computational simulation of liquid crystal configurations. Liquid crystals are used in diverse ways, most famously in display technologies, and the frontier of scientific applications and discoveries continues to expand. Emerging applications include liquid crystal-functionalized polymer fibers [74], nanoparticle organization [58, 126], photorefractive cells [29], and liquid crystal elastomers designed to produce effective actuator devices such as light driven motors [128] and artificial muscles [119].

Numerical simulation of liquid crystal equilibrium configurations are used to test and examine theory, suggest the presence of new physical phenomena [5], analyze experiments, and optimize device designs. Many current technologies and experiments, including bistable devices [30, 90], require simulation of anisotropic physical constants on two- and three-dimensional domains with complicated boundary conditions.

Many mathematical and computational models of liquid crystal continuum theory lead to complicated systems involving unit-length constrained vector fields. Currently, the complexity of such systems has restricted the existence of known analytical solutions to simplified geometries in one (1-D) or two dimensions (2-D), often under strong simplifying assumptions. When coupled with electric fields and other effects, far fewer analytical solutions exist, even in 1-D [117]. In addition, associated systems of partial differential equations, such as the equilibrium equations [46, 117], suffer from non-unique solutions, which must be distinguished via energy arguments.

Due to such difficulties, efficient, theoretically supported, numerical approaches to the modeling of nematic liquid crystals under free elastic and augmented electric effects are of great importance. A concise overview of existing research on the computational modeling of nematic liquid crystals is given in Section 3.1.

Herein, this thesis develops an energy-minimization finite-element approach for the computational simulation of static liquid crystal configurations based on the Frank-Oseen free-energy model. A theoretical framework supporting the accuracy and effectiveness of the approach is constructed, and tailored multigrid methods are explored for the arising linear systems. The overarching objective is a robust, efficient, and theoretically supported approach to the modeling of liquid crystal configurations.

In addition, we consider the numerical simulation of three-dimensional liquid crystal equilibrium problems incorporating substrates with certain two-dimensional geometric patterns. Surfaces with two-dimensional patterns have shown the potential to promote novel bistable arrangements, which are important in the design of display devices [3, 4]. We propose a variational system and finite-element method approach and analyze the well-posedness of the linearized systems in the context of both Dirichlet and Robin boundary conditions.

1.1 Thesis Outline

The structure of this thesis is as follows. Chapter 2 discusses the free-energy model to be used in simulating the nematic liquid crystal structures. This includes the free-elastic energy model, as well as extensions to include both external electric fields and internal electric fields due to flexoelectric effects. The chapter also includes a brief overview of the existing work on nematic equilibrium simulation.

In Chapter 3, an energy-minimization method using continuous Lagrange multipliers is developed for nematic liquid crystals under the influence of free-elastic effects. Theory proving the well-posedness of the linearization systems derived within the energy-minimization framework is established and error analysis is conducted to

demonstrate convergence. Finally, numerical results are presented using the minimization approach. Chapter 4 establishes an alternative energy-minimization framework where the necessary pointwise unit-length constraint is enforced via a penalty method. Well-posedness theory for the resulting linearization systems is again established. Trust-region methods designed for both the penalty and Lagrange multiplier approaches with finite-element discretizations are proposed in this chapter. The performance of the unit-length constraint techniques, trust-region methods, and nested iteration implementations is investigated in the numerical results section of the chapter.

Extensions of the energy-minimization approach with Lagrange multipliers are derived in Chapter 5 for external electric fields and flexoelectricity. The well-posedness theory constructed for the free-elastic model in Chapter 3 is expanded here to include both types of electric fields. In Chapter 6, two multigrid relaxation schemes specifically tailored to the block saddle-point linear systems arising in the discretization of the electrically coupled systems discussed in Chapter 5 are proposed and numerically vetted. These relaxation schemes lead to an optimally-scaling monolithic multigrid method well-suited for the electrically coupled discrete systems. Finally, numerical experiments applying the ensuing multigrid technique and demonstrating the energy-minimization method's performance for liquid crystal configurations in the presence of electric fields are performed.

Chapter 7 discusses the application of Newton linearization and finite-element methods for a nonlinear reaction-diffusion partial differential equation associated with a three-dimensional free-elastic liquid crystal configuration problem. Existence and uniqueness theory is proven for the emerging linearized systems in the presence of both Dirichlet and Robin boundary conditions. Numerical results computing configurations of physical interest are shown and ongoing experimental work is discussed.

Lastly, Chapter 8 gives some concluding remarks. The chapter also outlines some interesting problems and discusses a number of opportunities for future work. It addresses projects already under investigation, including extensions to liquid crystal

dynamics problems, as well as questions to be targeted in the future, such as parallel multigrid implementations and adaptive refinement techniques.

Chapter 2

Liquid Crystal Model

Liquid crystals, whose discovery is attributed to Reinitzer in 1888 [106], are substances that possess mesophases with properties intermediate between liquids and crystals. That is, liquid crystals are fluid yet exhibit long-range structured ordering. The mesophases exist at different temperatures or solvent concentrations. While the first observed liquid crystal structures were naturally occurring materials, a wide variety of synthesized chemical compounds have been produced [117]. There are many different liquid-crystal molecular structures including cholesterics, smectics, and their associated subclasses. However, in this thesis, we focus exclusively on nematic liquid crystal phases, which consist of rod-like molecules that self-assemble into an ordered structure, such that the molecules tend to align along a preferred orientation. The first nematic liquid crystals were synthesized in 1890 [55], while the first nematics existing stably at room temperature were not produced until 1969 by Kelker and Scheurle [72].

The preferred average direction for nematic liquid crystals at any point in a domain, Ω, is known as the director, denoted $\mathbf{n}(x, y, z) = (n_1, n_2, n_3)^T$; see Figure 2.1. The director is taken to be of unit length at every point and headless, that is $\mathbf{n}$ and $-\mathbf{n}$ are indistinguishable, reflecting the observed experimental symmetry of the phase. Thorough overviews of liquid crystal physics and properties are found in [22, 37, 117]. In the following sections, we discuss the various free-energy models to be used throughout this dissertation.

2.1 Free-Elastic Effects

At equilibrium, absent any external forces, fields, or boundary conditions, the free-elastic energy present in a liquid crystal sample is given by an integral functional, $\mathcal{F}$, which depends on the state variables of the system. A liquid crystal sample

Figure 2.1: Nematic liquid crystals with director orientation $\mathbf{n}$ [74].

tends toward configurations exhibiting minimal free energy. While a number of free-energy models exist cf. [31, 36], this thesis considers the Frank-Oseen free-elastic model [50, 117, 124]. The Frank-Oseen equations represent the free-elastic energy density, w_F, in a sample as

$$w_F = \frac{1}{2}K_1(\nabla\cdot\mathbf{n})^2+\frac{1}{2}K_2(\mathbf{n}\cdot\nabla\times\mathbf{n})^2+\frac{1}{2}K_3|\mathbf{n}\times\nabla\times\mathbf{n}|^2+\frac{1}{2}(K_2+K_4)\nabla\cdot[(\mathbf{n}\cdot\nabla)\mathbf{n}-(\nabla\cdot\mathbf{n})\mathbf{n}].$$

Throughout this thesis, the standard Euclidean inner product and norm are denoted $(\cdot,\cdot)$ and $|\cdot|$, respectively. The K_i, $i = 1,2,3,4$, are known as the Frank elastic constants [50], which vary depending on temperature and liquid crystal type [37,61]. By Ericksen's inequalities [47], $K_j \geq 0$ for $j = 1,2,3$. Each term represents an energy penalty for the presence of splay, twist, bend, and saddle-splay deformations, respectively.

As noted in [117],

$$\nabla\cdot[(\mathbf{n}\cdot\nabla)\mathbf{n} - (\nabla\cdot\mathbf{n})\mathbf{n}] = [\mathrm{tr}((\nabla\mathbf{n})^2) - (\nabla\cdot\mathbf{n})^2].$$

Using the 2-tensor definition

$$[\nabla\mathbf{v}]_{ij} = v_{i,j} = \frac{\partial v_i}{\partial x_j},$$

we write the trace term as

$$\mathrm{tr}((\nabla \mathbf{n})^2) = \nabla n_1 \cdot \frac{\partial \mathbf{n}}{\partial x} + \nabla n_2 \cdot \frac{\partial \mathbf{n}}{\partial y} + \nabla n_3 \cdot \frac{\partial \mathbf{n}}{\partial z}.$$

Therefore,

$$\nabla \cdot [(\mathbf{n} \cdot \nabla)\mathbf{n} - (\nabla \cdot \mathbf{n})\mathbf{n}] = \nabla n_1 \cdot \frac{\partial \mathbf{n}}{\partial x} + \nabla n_2 \cdot \frac{\partial \mathbf{n}}{\partial y} + \nabla n_3 \cdot \frac{\partial \mathbf{n}}{\partial z} - (\nabla \cdot \mathbf{n})^2. \tag{2.1}$$

Additionally, let

$$\mathbf{Z} = \kappa \mathbf{n} \otimes \mathbf{n} + (\mathbf{I} - \mathbf{n} \otimes \mathbf{n}) = \mathbf{I} - (1 - \kappa)\mathbf{n} \otimes \mathbf{n}, \tag{2.2}$$

where $\kappa = K_2/K_3$; in general, we consider the case that $K_2, K_3 > 0$. Denote the classical $L^2(\Omega)$ inner product and norm as $\langle \cdot, \cdot \rangle_0$ and $\| \cdot \|_0$, respectively. Employing (2.1), (2.2), and the fact that $\mathbf{n}$ has unit length, the total free energy for a domain, Ω, is

$$\begin{aligned}\int_\Omega w_F \, dV = {} & \frac{1}{2}(K_1 - K_2 - K_4)\|\nabla \cdot \mathbf{n}\|_0^2 + \frac{1}{2}K_3 \langle \mathbf{Z}\nabla \times \mathbf{n}, \nabla \times \mathbf{n}\rangle_0 \\ & + \frac{1}{2}(K_2 + K_4)\Big(\langle \nabla n_1, \frac{\partial \mathbf{n}}{\partial x}\rangle_0 + \langle \nabla n_2, \frac{\partial \mathbf{n}}{\partial y}\rangle_0 + \langle \nabla n_3, \frac{\partial \mathbf{n}}{\partial z}\rangle_0\Big).\end{aligned}$$

For simplicity, we define the following functional, scaled by 2, to be used in the minimization framework,

$$\begin{aligned}\mathcal{F}_1(\mathbf{n}) = {} & (K_1 - K_2 - K_4)\|\nabla \cdot \mathbf{n}\|_0^2 + K_3 \langle \mathbf{Z}\nabla \times \mathbf{n}, \nabla \times \mathbf{n}\rangle_0 \\ & + (K_2 + K_4)\Big(\langle \nabla n_1, \frac{\partial \mathbf{n}}{\partial x}\rangle_0 + \langle \nabla n_2, \frac{\partial \mathbf{n}}{\partial y}\rangle_0 + \langle \nabla n_3, \frac{\partial \mathbf{n}}{\partial z}\rangle_0\Big).\end{aligned} \tag{2.3}$$

For the special case of full Dirichlet boundary conditions, we consider a fixed director, $\mathbf{n}$, at each point on the boundary of Ω. Considering the integration carried out on the terms in (2.1),

$$\frac{1}{2}(K_2 + K_4) \int_\Omega \nabla \cdot [(\mathbf{n} \cdot \nabla)\mathbf{n} - (\nabla \cdot \mathbf{n})\mathbf{n}] \, dV$$

$$= \frac{1}{2}(K_2 + K_4) \int_{\partial\Omega} [(\mathbf{n} \cdot \nabla)\mathbf{n} - (\nabla \cdot \mathbf{n})\mathbf{n}] \cdot \nu \, dS, \tag{2.4}$$

by the divergence theorem, where ν is the outward facing unit normal. Further, since $\mathbf{n}$ is fixed along $\partial\Omega$, the free energy contributed by $\mathbf{n}$ on the boundary is constant regardless of the configuration of $\mathbf{n}$ on the interior of Ω. Thus, in the associated minimization to follow, the free energy contribution from this term is ignored. For this reason, (2.4) is often referred to as a null Lagrangian [124]. Note that the above identity is also applicable to a rectangular domain with mixed Dirichlet and periodic boundary conditions. Such conditions are considered extensively, in addition to Dirichlet boundary conditions, in the theory and numerical experiments to follow. This simplifies the free-energy functional in (2.3) to

$$\mathcal{F}_2(\mathbf{n}) = K_1 \|\nabla \cdot \mathbf{n}\|_0^2 + K_3 \langle \mathbf{Z}\nabla \times \mathbf{n}, \nabla \times \mathbf{n} \rangle_0. \tag{2.5}$$

We proceed with the functional in (2.3) in building a framework for minimization under general boundary conditions, including the possibility of Neumann or Robin conditions arising in the context of a free surface or weak anchoring [37, 117]. However, in the treatment of existence and uniqueness theory below, we assume the application of full Dirichlet or mixed Dirichlet and periodic boundary conditions and, thus, utilize the simplified form in (2.5).

2.2 Applied Electric Fields

In addition to their internal elastic properties, nematic liquid crystals are dielectrically active. Thus, their configurations are affected by electric fields. Moreover, since these materials are birefringent, with refractive indices that depend on the polarization of light, they can be used to control the propagation of light through a nematic structure.

Liquid crystal interactions with electric fields are strongly coupled as nematic polarization and electric displacement, in turn, affect the original electric field. This

coupling is captured by an auxiliary term added to the Frank-Oseen equations above, such that the total system free energy has the form

$$\int_\Omega \left(w_F - \frac{1}{2}\mathbf{D}\cdot\mathbf{E}\right) dV, \tag{2.6}$$

where $\mathbf{D}$ is the electric displacement vector induced by polarization and $\mathbf{E}$ is the local electric field [37]. This electric displacement vector is written

$$\mathbf{D} = \epsilon_0\epsilon_\perp\mathbf{E} + \epsilon_0\epsilon_a(\mathbf{n}\cdot\mathbf{E})\mathbf{n}.$$

Here, $\epsilon_0 > 0$ is the permittivity of free space. The dielectric anisotropy constant is $\epsilon_a = \epsilon_\parallel - \epsilon_\perp$, where the constant variables $\epsilon_\parallel > 0$ and $\epsilon_\perp > 0$ represent the parallel and perpendicular dielectric permittivity, respectively, specific to the liquid crystal. If $\epsilon_a > 0$, the director is attracted to parallel alignment with the electric field, and if $\epsilon_a < 0$, the director tends to align perpendicular to $\mathbf{E}$. Thus,

$$\mathbf{D}\cdot\mathbf{E} = \epsilon_0\epsilon_\perp\mathbf{E}\cdot\mathbf{E} + \epsilon_0\epsilon_a(\mathbf{n}\cdot\mathbf{E})^2.$$

Note that the magnitude and difference between $\epsilon_\parallel$ and $\epsilon_\perp$ influence the strength of polarization in the nematic as well as the coupling interaction with the electric field.

Equation (2.6) is expanded as

$$\int_\Omega w_F\, dV - \int_\Omega \frac{1}{2}\mathbf{D}\cdot\mathbf{E}\, dV = \int_\Omega w_F\, dV - \frac{1}{2}\epsilon_0\epsilon_\perp\langle\mathbf{E},\mathbf{E}\rangle_0 - \frac{1}{2}\epsilon_0\epsilon_a\langle\mathbf{n}\cdot\mathbf{E},\mathbf{n}\cdot\mathbf{E}\rangle_0. \tag{2.7}$$

The addition of the electric field not only increases the complexity of the functional, it introduces an inherent saddle-point structure into the equilibria for the liquid crystal samples. Energy minima are those that minimize the contribution of the free-elastic energy, while maximizing the negative contribution of the electric field terms. Moreover, the relevant Maxwell's equations for a static electric field, $\nabla\cdot\mathbf{D} = 0$ and $\nabla\times\mathbf{E} = \mathbf{0}$, known as Gauss' and Faraday's laws, respectively, must be satisfied.

In light of the necessary Maxwell equations and the fact that we are considering

static fields, we define a functional based on the system free energy in (2.7) using an electric potential function, ϕ, such that $\mathbf{E} = -\nabla\phi$. Applying similar scaling to that of (2.3), let

$$\begin{aligned}\mathcal{F}_3(\mathbf{n},\phi) = (K_1 - K_2 - K_4)\|\nabla\cdot\mathbf{n}\|_0^2 + K_3\langle\mathbf{Z}\nabla\times\mathbf{n},\nabla\times\mathbf{n}\rangle_0 \\ + (K_2+K_4)\big(\langle\nabla n_1,\frac{\partial\mathbf{n}}{\partial x}\rangle_0 + \langle\nabla n_2,\frac{\partial\mathbf{n}}{\partial y}\rangle_0 + \langle\nabla n_3,\frac{\partial\mathbf{n}}{\partial z}\rangle_0\big) \\ - \epsilon_0\epsilon_\perp\langle\nabla\phi,\nabla\phi\rangle_0 - \epsilon_0\epsilon_a\langle\mathbf{n}\cdot\nabla\phi,\mathbf{n}\cdot\nabla\phi\rangle_0. \qquad (2.8)\end{aligned}$$

Observe that using the potential function implies

$$\nabla\times\mathbf{E} = \nabla\times(-\nabla\phi) = \mathbf{0}.$$

So Faraday's law is automatically satisfied. It is noted below that, in the framework of energy-minimization, Gauss' law is also satisfied at an energy minimum.

2.3 Flexoelectric Phenomena

Flexoelectricity is a property demonstrated by certain dielectric materials, including liquid crystals. It is a spontaneous polarization of the liquid crystal induced by present curvature; it is caused by shape asymmetry of the constituent molecules of the liquid crystal material. The initial suggestion of this type of property in liquid crystals was introduced by Meyer [95]. The phenomenon is analogous to the accumulation of electric charge due to strain in solids, known as the piezoelectric effect [27]. Therefore, in some literature, flexoelectricity is referred to as piezoelectricity.

Flexoelectric phenomena can, for instance, be useful in the conversion of mechanical energy to electrical energy via large deformations of the boundary containing a liquid crystal sample [66]. It can also play a significant role in determining the equilibrium states of liquid crystal samples with patterned surface boundaries. For example, it is an important effect in the bistable configuration of the Zenithal Bistable Device (ZBD) [30].

The effect of flexoelectricity on the alignment of a liquid crystal bulk is modeled by an augmentation of the electric displacement vector $\mathbf{D}$, discussed above, and additional terms for the bulk free-energy functional. The electric displacement vector is modified [43] such that

$$\mathbf{D} = \epsilon_0\epsilon_\perp\mathbf{E} + \epsilon_0\epsilon_a(\mathbf{n}\cdot\mathbf{E})\mathbf{n} + \mathbf{P}_{\text{flexo}}.$$

Following the notation and sign convention of Rudquist [109], we write

$$\mathbf{P}_{\text{flexo}} = e_s\mathbf{n}(\nabla\cdot\mathbf{n}) + e_b(\mathbf{n}\times\nabla\times\mathbf{n}),$$

where e_s and e_b are material constants specific to a given liquid crystal. It is also common in physics literature to denote these constants as e_1 and e_3 under a separate sign convention [37, 43, 95].

As expressed in [43], the system free energy, when considering flexoelectricity, is written

$$\int_\Omega w_F\,dV - \frac{1}{2}\epsilon_0\epsilon_\perp\langle\mathbf{E},\mathbf{E}\rangle_0 - \frac{1}{2}\epsilon_0\epsilon_a\langle\mathbf{n}\cdot\mathbf{E},\mathbf{n}\cdot\mathbf{E}\rangle_0 - \langle\mathbf{P}_{\text{flexo}},\mathbf{E}\rangle_0.$$

Thus, the system free energy contributed by flexoelectric polarization is expanded as

$$-\int_\Omega e_s(\nabla\cdot\mathbf{n})(\mathbf{E}\cdot\mathbf{n}) + e_b(\mathbf{n}\times\nabla\times\mathbf{n})\cdot\mathbf{E}\,dV.$$

Substituting an electric potential function, $\mathbf{E} = -\nabla\phi$, and scaling by a factor of 2, the flexoelectric free-energy functional to be minimized is expressed,

$$\mathcal{F}_4(\mathbf{n},\phi) = \mathcal{F}_3(\mathbf{n},\phi) + 2e_s\langle\nabla\cdot\mathbf{n},\mathbf{n}\cdot\nabla\phi\rangle_0 + 2e_b\langle\mathbf{n}\times\nabla\times\mathbf{n},\nabla\phi\rangle_0. \tag{2.9}$$

Additionally, note that the Maxwell's equations, $\nabla\cdot\mathbf{D} = 0$ and $\nabla\times\mathbf{E} = \mathbf{0}$, must still be satisfied. As before, the use of the electric potential implies that Faraday's law is automatically satisfied.

In the presence of full Dirichlet or mixed Dirichlet and periodic boundary conditions on a rectangular domain, the simplification in (2.4) is applied to eliminate the $(K_2 + K_4)$ free-elastic terms from Functionals (2.8) and (2.9).

Chapter 3

Free-Elastic Energy Minimization

In this section, a general approach to computing the free-elastic equilibrium state for $\mathbf{n}$ is derived. Therefore, only free-elastic effects, governed by the functional in (2.3), are considered here. This equilibrium state corresponds to the configuration which minimizes the system free energy subject to the local constraint that $\mathbf{n}$ is of unit length throughout the sample volume, Ω. That is, the minimizer must satisfy $\mathbf{n} \cdot \mathbf{n} = 1$ pointwise throughout the domain. Pointwise unit-length constraints are present in other types of physical problems, such as ferromagnetics [73]. These constraints offer unique computational challenges as they are nonlinear, local, and differ significantly from the well-studied constraints treated in fluid theory.

3.1 Existing Approaches and Simplifications

A number of computational techniques for liquid crystal equilibrium [25, 57, 64, 103, 117] and dynamics problems [83, 84, 86, 127, 129] exist, including least-squares finite-element methods [5], discrete Lagrange multiplier approaches [54, 105], and penalty methods [57,64]. In addition, numerical experiments involving finite-element methods with Lagrange multipliers, applied to the equilibrium equations, have been successful in capturing certain liquid crystal characteristics [103].

Many of the methods described above utilize the so called one-constant approximation that $K_1 = K_2 = K_3$ and $K_4 = 0$ [25, 54, 57, 64, 83, 84, 86, 105, 117, 127, 129], in order to significantly simplify the free-elastic energy density to

$$\hat{w}_F = \frac{1}{2} K_1 |\nabla \mathbf{n}|^2, \qquad \text{where } |\nabla \mathbf{n}|^2 = \sum_{i,j=1}^{3} \left(\frac{\partial n_i}{\partial x_j} \right)^2 .$$

This expression for the free-energy density is more amenable to theoretical development and computational techniques. However, while this is an accurate approximation in certain scenarios, especially when the relationship of the Frank constants is unknown, there are many applications for which this approximation does not suitably capture liquid crystal behavior [3, 4, 6, 76]. Therefore, we endeavor to fully exclude this type of approximation.

Certain approaches, such as those in [103], have numerically resolved physically expected liquid crystal behavior under simplifying assumptions but do so in the context of the equilibrium equations [46], which offer many theoretical and computational development challenges. Similarly, methods using a first-order system least-squares approach [20, 21] applied to the equilibrium equations have predicted new physical phenomena [5] but still require theory supporting ellipticity and have encountered difficulties in properly capturing applied electric field effects.

In the following, we develop a theoretically supported method that directly targets energy minimization in the continuum. The method and accompanying theory are applicable for a wide range of physical parameters. This allows for significantly improved modeling of physical phenomena not captured in many approaches. Moreover, as will be seen in subsequent chapters, the approach readily accommodates applied electric fields and flexoelectric effects.

3.2 Energy-Minimization Approach

Throughout the derivation of the energy-minimization framework, we use the general functional, $\mathcal{F}_1(\mathbf{n})$, in (2.3) and consider the spaces

$$H(\mathrm{div}, \Omega) = \{\mathbf{v} \in L^2(\Omega)^3 : \nabla \cdot \mathbf{v} \in L^2(\Omega)\},$$
$$H(\mathrm{curl}, \Omega) = \{\mathbf{v} \in L^2(\Omega)^3 : \nabla \times \mathbf{v} \in L^2(\Omega)^3\}.$$

Further, let

$$H_0(\mathrm{div}, \Omega) = \{\mathbf{v} \in H(\mathrm{div}, \Omega) : \nu \cdot \mathbf{v} = 0 \text{ on } \partial\Omega\},$$

$$H_0(\text{curl}, \Omega) = \{\mathbf{v} \in H(\text{curl}, \Omega) : \nu \times \mathbf{v} = \mathbf{0} \text{ on } \partial\Omega\},$$

where ν is the outward unit normal for $\partial\Omega$. Define

$$\mathcal{H}^{DC}(\Omega) = \{\mathbf{v} \in H(\text{div}, \Omega) \cap H(\text{curl}, \Omega) : B(\mathbf{v}) = \mathbf{g}\},$$

with norm $\|\mathbf{v}\|_{DC}^2 = \|\mathbf{v}\|_0^2 + \|\nabla \cdot \mathbf{v}\|_0^2 + \|\nabla \times \mathbf{v}\|_0^2$ and appropriate boundary conditions $B(\mathbf{v}) = \mathbf{g}$. Here, we assume that $\mathbf{g}$ satisfies appropriate compatibility conditions for operator B. For example, if B represents Dirichlet boundary conditions and Ω has a Lipschitz continuous boundary, it is assumed that $\mathbf{g} \in H^{\frac{1}{2}}(\partial\Omega)^3$ [56]. Further, let $\mathcal{H}_0^{DC}(\Omega) = \{\mathbf{v} \in H(\text{div}, \Omega) \cap H(\text{curl}, \Omega) : B(\mathbf{v}) = \mathbf{0}\}$. Note that if Ω is a Lipshitz domain and B imposes Dirichlet boundary conditions, then $\mathcal{H}_0^{DC}(\Omega) = H_0^1(\Omega)^3$ [56, Lemma 2.5]. Finally, denote the unit sphere as $\mathcal{S}^2$. The desired minimization becomes

$$\mathbf{n}_* = \underset{\mathbf{n} \in \mathcal{S}^2 \cap \mathcal{H}^{DC}(\Omega)}{\text{argmin}} \mathcal{F}_1(\mathbf{n}).$$

3.3 First-Order Optimality and Newton Linearization

Since $\mathbf{n}$ must be of unit length, it is natural to employ a Lagrange multiplier approach. This length requirement represents a pointwise equality constraint, such that $(\mathbf{n}, \mathbf{n}) - 1 = 0$. Following the notation in [88], we have $H(\mathbf{n}) = (\mathbf{n}, \mathbf{n}) - 1 = 0$ and, hence, the Lagrange multiplier $z^* \in L^2(\Omega)^*$, where $L^2(\Omega)^*$ is the dual space for $L^2(\Omega)$. The Lagrangian is written

$$\mathcal{L}(\mathbf{n}, z^*) = \mathcal{F}_1(\mathbf{n}) + \langle H(\mathbf{n}), z^* \rangle.$$

Since $L^2(\Omega)$ is a Hilbert space, by the Riesz representation theorem [108] there exists a $\lambda(\mathbf{x}) \in L^2(\Omega)$ such that the Lagrangian becomes

$$\mathcal{L}(\mathbf{n}, \lambda) = \mathcal{F}_1(\mathbf{n}) + \int_\Omega \lambda(\mathbf{x})((\mathbf{n}, \mathbf{n}) - 1)\, dV. \tag{3.1}$$

In order to minimize (2.3), we compute the Gâteaux derivatives of $\mathcal{L}$ with respect to $\mathbf{n}$ and λ in the directions $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$ and $\gamma \in L^2(\Omega)$, respectively. Hence, the necessary continuum first-order optimality conditions are

$$\mathcal{L}_{\mathbf{n}}[\mathbf{v}] = \frac{\partial}{\partial \mathbf{n}} \mathcal{L}(\mathbf{n}, \lambda)[\mathbf{v}] = 0, \qquad \forall \mathbf{v} \in \mathcal{H}_0^{DC}(\Omega), \tag{3.2}$$

$$\mathcal{L}_{\lambda}[\gamma] = \frac{\partial}{\partial \lambda} \mathcal{L}(\mathbf{n}, \lambda)[\gamma] = 0, \qquad \forall \gamma \in L^2(\Omega). \tag{3.3}$$

Computing these derivatives yields

$$\begin{aligned} \mathcal{L}_{\mathbf{n}}[\mathbf{v}] =& 2(K_1 - K_2 - K_4)\langle \nabla \cdot \mathbf{n}, \nabla \cdot \mathbf{v} \rangle_0 + 2K_3 \langle \mathbf{Z}(\mathbf{n}) \nabla \times \mathbf{n}, \nabla \times \mathbf{v} \rangle_0 \\ &+ 2(K_2 - K_3)\langle \mathbf{n} \cdot \nabla \times \mathbf{n}, \mathbf{v} \cdot \nabla \times \mathbf{n} \rangle_0 + 2(K_2 + K_4)\Big(\langle \nabla n_1, \frac{\partial \mathbf{v}}{\partial x} \rangle_0 \\ &+ \langle \nabla n_2, \frac{\partial \mathbf{v}}{\partial y} \rangle_0 + \langle \nabla n_3, \frac{\partial \mathbf{v}}{\partial z} \rangle_0 \Big) + 2 \int_{\Omega} \lambda(\mathbf{n}, \mathbf{v})\, dV, \end{aligned}$$

and

$$\mathcal{L}_{\lambda}[\gamma] = \int_{\Omega} \gamma((\mathbf{n}, \mathbf{n}) - 1)\, dV.$$

The variational system contains nonlinearities in both (3.2) and (3.3). Therefore, Newton iterations are employed by computing a generalized first-order Taylor series expansion, requiring computation of the Hessian [11, 39, 100].

Let $\mathbf{n}_k$ and λ_k be the current approximations for $\mathbf{n}$ and λ, respectively. Additionally, let $\delta\mathbf{n} = \mathbf{n}_{k+1} - \mathbf{n}_k$ and $\delta\lambda = \lambda_{k+1} - \lambda_k$ be updates to these approximations. Then, the Newton iterations are denoted

$$\begin{bmatrix} \mathcal{L}_{\mathbf{nn}} & \mathcal{L}_{\mathbf{n}\lambda} \\ \mathcal{L}_{\lambda\mathbf{n}} & \mathcal{L}_{\lambda\lambda} \end{bmatrix} \begin{bmatrix} \delta\mathbf{n} \\ \delta\lambda \end{bmatrix} = - \begin{bmatrix} \mathcal{L}_{\mathbf{n}} \\ \mathcal{L}_{\lambda} \end{bmatrix}, \tag{3.4}$$

where each of the system components are evaluated at $\mathbf{n}_k$ and λ_k. The matrix-vector multiplication indicates the direction that the derivatives in the Hessian are taken.

That is,

$$\mathcal{L}_{\mathbf{nn}}[\mathbf{v}]\cdot\delta\mathbf{n} = \frac{\partial}{\partial\mathbf{n}}\left(\mathcal{L}_{\mathbf{n}}(\mathbf{n}_k,\lambda_k)[\mathbf{v}]\right)[\delta\mathbf{n}], \quad \mathcal{L}_{\mathbf{n}\lambda}[\mathbf{v}]\cdot\delta\lambda = \frac{\partial}{\partial\lambda}\left(\mathcal{L}_{\mathbf{n}}(\mathbf{n}_k,\lambda_k)[\mathbf{v}]\right)[\delta\lambda],$$
$$\mathcal{L}_{\lambda\mathbf{n}}[\gamma]\cdot\delta\mathbf{n} = \frac{\partial}{\partial\mathbf{n}}\left(\mathcal{L}_{\lambda}(\mathbf{n}_k,\lambda_k)[\gamma]\right)[\delta\mathbf{n}], \quad \mathcal{L}_{\lambda\lambda}[\gamma]\cdot\delta\lambda = \frac{\partial}{\partial\lambda}\left(\mathcal{L}_{\lambda}(\mathbf{n}_k,\lambda_k)[\gamma]\right)[\delta\lambda],$$

where the partials denote Gâteaux derivatives in the respective variables.

Since $\mathcal{L}(\mathbf{n},\lambda)$ is linear in λ, $\mathcal{L}_{\lambda\lambda}[\gamma]\cdot\delta\lambda = 0$. Hence, the Hessian in (3.4) simplifies to a saddle-point structure,

$$\begin{bmatrix} \mathcal{L}_{\mathbf{nn}} & \mathcal{L}_{\mathbf{n}\lambda} \\ \mathcal{L}_{\lambda\mathbf{n}} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \delta\mathbf{n} \\ \delta\lambda \end{bmatrix} = -\begin{bmatrix} \mathcal{L}_{\mathbf{n}} \\ \mathcal{L}_{\lambda} \end{bmatrix}. \tag{3.5}$$

The discrete form of this Hessian leads to a saddle-point matrix, which poses unique difficulties for the efficient computation of the solution to the resulting linear system. Such structures commonly appear in constrained optimization and other settings; for a comprehensive overview of discrete saddle-point problems; see [10]. Here, we focus only on the linearization step rather than the underlying linear solvers. Multigrid solvers specifically designed for saddle-point systems are developed in Chapter 6. Computing the Gâteaux derivatives yields

$$\mathcal{L}_{\mathbf{n}\lambda}[\mathbf{v}]\cdot\delta\lambda = 2\int_\Omega \delta\lambda(\mathbf{n}_k,\mathbf{v})\,dV, \tag{3.6}$$

$$\mathcal{L}_{\lambda\mathbf{n}}[\gamma]\cdot\delta\mathbf{n} = 2\int_\Omega \gamma(\mathbf{n}_k,\delta\mathbf{n})\,dV, \tag{3.7}$$

$$\begin{aligned}\mathcal{L}_{\mathbf{nn}}[\mathbf{v}]\cdot\delta\mathbf{n} = {}& 2(K_1-K_2-K_4)\langle\nabla\cdot\delta\mathbf{n},\nabla\cdot\mathbf{v}\rangle_0 + 2K_3\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\delta\mathbf{n},\nabla\times\mathbf{v}\rangle_0 \\ & + 2(K_2-K_3)\Big(\langle\delta\mathbf{n}\cdot\nabla\times\mathbf{v},\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0 + \langle\mathbf{n}_k\cdot\nabla\times\mathbf{v},\delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k\rangle_0 \\ & + \langle\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\delta\mathbf{n}\rangle_0 + \langle\mathbf{n}_k\cdot\nabla\times\delta\mathbf{n},\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 \\ & + \langle\delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big) + 2(K_2+K_4)\Big(\langle\nabla\delta n_1,\frac{\partial\mathbf{v}}{\partial x}\rangle_0 \\ & + \langle\nabla\delta n_2,\frac{\partial\mathbf{v}}{\partial y}\rangle_0 + \langle\nabla\delta n_3,\frac{\partial\mathbf{v}}{\partial z}\rangle_0\Big) + 2\int_\Omega \lambda_k(\delta\mathbf{n},\mathbf{v})\,dV.\end{aligned} \tag{3.8}$$

Constructing (3.5) using (3.6)-(3.8) yields a linearized variational system, which is

fully expanded in Appendix A.1. For these iterations, we compute $\delta\mathbf{n}$ and $\delta\lambda$ satisfying this system for all $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$ and $\gamma \in L^2(\Omega)$ with the current approximations $\mathbf{n}_k$ and λ_k. The current approximations are then corrected with the solutions $\delta\mathbf{n}$ and $\delta\lambda$ to produce $\mathbf{n}_{k+1}$ and λ_{k+1}.

If we are considering a system with Dirichlet or mixed periodic and Dirichlet boundary conditions, as described above, we eliminate the $(K_2 + K_4)$ terms from (3.5), simplifying the linearization. This simplified system is also represented in Appendix A.1.

3.4 Uniform Symmetric Positive Definiteness of Z

In subsequent sections, theory establishing the existence and uniqueness of solutions to the Newton linearizations is developed. A key property exploited in these proofs is that $\mathbf{Z}$ is uniformly symmetric positive definite (USPD) under reasonable assumptions.

It is clear from the definition that $\mathbf{Z}$ is symmetric. Recall that by Ericksen's inequalities [47], $K_2, K_3 \geq 0$. Throughout this thesis, we consider the case where the inequality is strict; thus, $\kappa > 0$. We also assume that, in the Newton iterations, control has been maintained over the director length such that

$$\alpha \leq n_1^2 + n_2^2 + n_3^2 \leq \beta, \qquad \forall \mathbf{x} \in \Omega, \tag{3.9}$$

with constants $0 < \alpha \leq 1 \leq \beta$.

Lemma 3.4.1 *Assume that $\alpha \leq (\mathbf{n}, \mathbf{n}) \leq \beta$ for all $\mathbf{x} \in \Omega$. If $\kappa \geq 1$, then $\mathbf{Z}$ is USPD on Ω. For $0 < \kappa < 1$, if $\beta < \frac{1}{1-\kappa}$, then $\mathbf{Z}$ is USPD on Ω.*

Proof: Rewrite $\mathbf{Z}$ as

$$\mathbf{Z} = \mathbf{I} - \frac{\mathbf{n} \otimes \mathbf{n}}{\mathbf{n} \cdot \mathbf{n}} + (1 + (\kappa - 1)(\mathbf{n} \cdot \mathbf{n}))\frac{\mathbf{n} \otimes \mathbf{n}}{\mathbf{n} \cdot \mathbf{n}}.$$

For any $\mathbf{x} \in \Omega$, consider $\xi \in \mathbb{R}^3$. Decompose ξ as $\xi = a_1\mathbf{v} + a_2\mathbf{n}$, where $\mathbf{v}\cdot\mathbf{n} = 0$. Then,

$$\frac{\xi^T\mathbf{Z}(\mathbf{x})\xi}{\xi^T\xi} = \frac{a_1^2\mathbf{v}\cdot\mathbf{v} + (1+(\kappa-1)(\mathbf{n}\cdot\mathbf{n}))a_2^2(\mathbf{n}\cdot\mathbf{n})}{a_1^2\mathbf{v}\cdot\mathbf{v} + a_2^2(\mathbf{n}\cdot\mathbf{n})}.$$

Thus,

$$\min(1, 1+(\kappa-1)(\mathbf{n}\cdot\mathbf{n})) \leq \frac{\xi^T\mathbf{Z}(\mathbf{x})\xi}{\xi^T\xi} \leq \max(1, 1+(\kappa-1)(\mathbf{n}\cdot\mathbf{n})).$$

Case 1. $\kappa \geq 1$.

Note that

$$1 \leq 1+(\kappa-1)\alpha \leq (1+(\kappa-1)(\mathbf{n}\cdot\mathbf{n})) \leq 1+(\kappa-1)\beta, \qquad \forall \mathbf{x} \in \Omega.$$

Hence,

$$1 \leq \frac{\xi^T\mathbf{Z}(\mathbf{x})\xi}{\xi^T\xi} \leq 1+(\kappa-1)\beta, \qquad \forall \mathbf{x} \in \Omega, \xi \in \mathbb{R}^3.$$

Case 2. $0 < \kappa < 1$.

For this case,

$$1+(\kappa-1)\beta \leq (1+(\kappa-1)(\mathbf{n}\cdot\mathbf{n})) \leq 1+(\kappa-1)\alpha \leq 1, \qquad \forall \mathbf{x} \in \Omega.$$

Along with the assumption that $\beta < \frac{1}{1-\kappa}$, this implies that

$$0 < 1+(\kappa-1)\beta \leq \frac{\xi^T\mathbf{Z}(\mathbf{x})\xi}{\xi^T\xi} \leq 1, \qquad \forall \mathbf{x} \in \Omega, \xi \in \mathbb{R}^3.$$

□

Thus, $\mathbf{Z}$ is USPD for any $\kappa > 0$, as long as sufficient control is maintained on the length of $\mathbf{n}$. Let η and Λ denote the lower and upper bounds, respectively, in the proof above such that

$$0 < \eta \leq \frac{\xi^T\mathbf{Z}(\mathbf{x})\xi}{\xi^T\xi} \leq \Lambda, \qquad \forall \mathbf{x} \in \Omega, \xi \in \mathbb{R}^3.$$

These USPD bounds for $\mathbf{Z}$ play an important role in the proofs of existence and uniqueness of solutions to the linearizations undertaken below.

3.5 Existence and Uniqueness for the Linearizations

Here and in the following subsections, we routinely make use of the following set of assumptions.

Assumption 3.5.1 *Consider an open, bounded domain, Ω, with a Lipschitz-continuous boundary. Further, assume that there exist constants $0 < \alpha \leq 1 \leq \beta$, such that $\alpha \leq |\mathbf{n}_k|^2 \leq \beta$ and $\mathbf{Z}(\mathbf{n}_k(\mathbf{x}))$ remains USPD with lower and upper bounds on its Rayleigh quotient, η and Λ, respectively. Finally, assume that Dirichlet boundary conditions are applied. Therefore, both $\delta\mathbf{n}$ and $\mathbf{v}$ are in $H_0(div, \Omega) \cap H_0(curl, \Omega)$.*

In the continuum, the above Newton systems are written in a general form as

$$a(\delta\mathbf{n}, \mathbf{v}) + b(\mathbf{v}, \delta\lambda) = F(\mathbf{v}), \qquad \forall \mathbf{v} \in \mathcal{H}_0^{DC}(\Omega), \tag{3.10}$$

$$b(\delta\mathbf{n}, \gamma) = G(\gamma), \qquad \forall \gamma \in L^2(\Omega), \tag{3.11}$$

where $a(\cdot,\cdot)$ is a symmetric bilinear form, $b(\cdot,\cdot)$ is a bilinear form, and F and G are linear functionals. Note that in the presence of Dirichlet boundary conditions or mixed periodic and Dirichlet boundary conditions on a rectangular domain, the linearized system is reduced by the application of (2.4). For simplicity, throughout this section, we drop the notation of $\delta\mathbf{n}$, $\delta\lambda$. Thus,

$$\begin{aligned} a(\mathbf{u}, \mathbf{v}) =& K_1\langle \nabla\cdot\mathbf{u}, \nabla\cdot\mathbf{v}\rangle_0 + K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{u}, \nabla\times\mathbf{v}\rangle_0 \\ &+ (K_2 - K_3)\Big(\langle \mathbf{u}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0 + \langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \mathbf{u}\cdot\nabla\times\mathbf{n}_k\rangle_0 \\ &+ \langle \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{u}\rangle_0 + \langle \mathbf{n}_k\cdot\nabla\times\mathbf{u}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 \\ &+ \langle \mathbf{u}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big) + \int_\Omega \lambda_k(\mathbf{u}, \mathbf{v})\, dV, \end{aligned} \tag{3.12}$$

and

$$b(\mathbf{v},\gamma) = \int_\Omega \gamma(\mathbf{n}_k,\mathbf{v})\,dV.$$

Moreover,

$$F(\mathbf{v}) = -\Big(K_1\langle\nabla\cdot\mathbf{n}_k,\nabla\cdot\mathbf{v}\rangle_0 + K_3\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{n}_k,\nabla\times\mathbf{v}\rangle_0 + (K_2-K_3)\langle\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 + \int_\Omega \lambda_k(\mathbf{n}_k,\mathbf{v})\,dV\Big),$$

and

$$G(\gamma) = -\frac{1}{2}\int_\Omega \gamma((\mathbf{n}_k,\mathbf{n}_k)-1)\,dV.$$

In this section, we aim to show that the system in (3.10) and (3.11) is well-posed. Therefore, continuity, coercivity, and weak coercivity results are desired for the bilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$. Due to the complexity of the bilinear forms, deriving theoretical results in the continuum is challenging. However, the following lemmas hold.

Lemma 3.5.2 *Under Assumption 3.5.1 and the assumption that λ_k is pointwise non-negative, if $\kappa = 1$, there exists an $\alpha_0 > 0$ such that $\alpha_0\|\mathbf{v}\|_{DC}^2 \le a(\mathbf{v},\mathbf{v})$ for all $\mathbf{v}\in\mathcal{H}_0^{DC}(\Omega)$.*

Proof: The proof of this lemma is identical to that of Lemma 3.5.8 below. □

If additional regularity is asserted, such that $\delta\mathbf{n}$ and $\mathbf{v}$ are elements of $\mathcal{H}_0^{DC^1}(\Omega) = \{\mathbf{w}\in\mathcal{H}_0^{DC}(\Omega) : \nabla\times\mathbf{w}\in H^1(\Omega)^3\}$ with norm $\|\mathbf{w}\|_{DC^1}^2 = \|\mathbf{w}\|_0^2 + \|\nabla\cdot\mathbf{w}\|_0^2 + \|\nabla\times\mathbf{w}\|_1^2$, where $\|\cdot\|_1$ denotes the standard norm on $H^1(\Omega)$, then the next two lemmas hold for arbitrary κ.

Lemma 3.5.3 *Under Assumption 3.5.1, F and G are bounded linear functionals on $\mathcal{H}_0^{DC^1}(\Omega)$ and $L^2(\Omega)$, respectively.*

Proof: A simple application of the Cauchy-Schwarz inequality shows that $G(\gamma)$ is a bounded linear functional.

For $F(\mathbf{v})$, observe that

$$|F(\mathbf{v})| \leq K_1|\langle \nabla \cdot \mathbf{n}_k, \nabla \cdot \mathbf{v}\rangle_0| + K_3|\langle \mathbf{Z}(\mathbf{n}_k)\nabla \times \mathbf{n}_k, \nabla \times \mathbf{v}\rangle_0| \\ + |K_2 - K_3||\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0| + |\int_\Omega \lambda_k(\mathbf{n}_k, \mathbf{v})\, dV|, \tag{3.13}$$

by the triangle inequality. Applying Cauchy-Schwarz inequalities to (3.13), one obtains

$$\begin{aligned} |F(\mathbf{v})| &\leq K_1\|\nabla \cdot \mathbf{n}_k\|_0\|\nabla \cdot \mathbf{v}\|_0 + K_3\|\mathbf{Z}(\mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0\|\nabla \times \mathbf{v}\|_0 \\ &\quad + |K_2 - K_3||\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0| + \|\lambda_k \mathbf{n}_k\|_0\|\mathbf{v}\|_0 \\ &\leq K_1\|\nabla \cdot \mathbf{n}_k\|_0\|\mathbf{v}\|_{DC^1} + K_3\|\mathbf{Z}(\mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0\|\mathbf{v}\|_{DC^1} \\ &\quad + |K_2 - K_3||\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0| + \|\lambda_k \mathbf{n}_k\|_0\|\mathbf{v}\|_{DC^1}. \end{aligned} \tag{3.14}$$

In order to bound $|F(\mathbf{v})|$, consider the final three summands separately. Note that since $|\mathbf{Z}(\mathbf{n}_k)| \leq \Lambda$, where Λ is the relevant upper bound from Lemma 3.4.1, it is evident that

$$\|\mathbf{Z}(\mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0 \leq \Lambda\|\nabla \times \mathbf{n}_k\|_0, \tag{3.15}$$

and that

$$\|\lambda_k \mathbf{n}_k\|_0^2 \leq \beta \int_\Omega \lambda_k^2\, dV = C_L^2, \tag{3.16}$$

where β is the upper bound in (3.9). Finally, consider

$$|\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0| = |\langle (\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k, \mathbf{v}\rangle_0|.$$

Applying the Cauchy-Schwarz inequality,

$$|\langle (\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k, \mathbf{v}\rangle_0| \leq \|(\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0\|\mathbf{v}\|_0$$

$$\leq \|(\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0 \|\mathbf{v}\|_{DC^1}. \tag{3.17}$$

By Corollary 1.1 in [56], since $\nabla \times \mathbf{n}_k \in H^1(\Omega)^3$, $\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k \in L^2(\Omega)$. Note that

$$\begin{aligned}(\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k \cdot (\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k &= (\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)^2(\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k) \\ &\leq (|\mathbf{n}_k| \cdot |\nabla \times \mathbf{n}_k|)^2(\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k) \\ &\leq \beta \cdot (\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)^2.\end{aligned}$$

Employing this in (3.17) and letting $\|\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k\|_0 = C_N$,

$$\begin{aligned}\|(\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0 &\leq \sqrt{\beta}\Big(\int_\Omega (\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)^2 \, dV\Big)^{1/2} \\ &\leq \sqrt{\beta} C_N. \end{aligned} \tag{3.18}$$

Therefore, using (3.14)-(3.16), and (3.18),

$$\begin{aligned}|F(\mathbf{v})| \leq & K_1 \|\nabla \cdot \mathbf{n}_k\|_0 \|\mathbf{v}\|_{DC^1} + K_3 \Lambda \|\nabla \times \mathbf{n}_k\|_0 \|\mathbf{v}\|_{DC^1} \\ & + |K_2 - K_3| \sqrt{\beta} C_N \|\mathbf{v}\|_{DC^1} + C_L \|\mathbf{v}\|_{DC^1}. \end{aligned}$$ □

Lemma 3.5.4 *Under Assumption 3.5.1, $a(\mathbf{u}, \mathbf{v})$ and $b(\mathbf{v}, \gamma)$ are continuous for the norms $\|\cdot\|_{DC^1}$ and $\|\cdot\|_0$.*

Proof: First consider

$$\begin{aligned}|b(\mathbf{v}, \gamma)| &= |\int_\Omega \gamma(\mathbf{v}, \mathbf{n}_k) \, dV| \\ &\leq \|\gamma\|_0 \|\mathbf{v} \cdot \mathbf{n}_k\|_0 \\ &\leq \|\gamma\|_0 \sqrt{\beta} \|\mathbf{v}\|_0, \end{aligned}$$

by Hölder's inequality and (3.9). Therefore, $b(\mathbf{v}, \gamma)$ is a continuous bilinear form.

For the continuity of $a(\mathbf{u}, \mathbf{v})$, observe that

$$|a(\mathbf{u}, \mathbf{v})| \leq K_1 |\langle \nabla \cdot \mathbf{u}, \nabla \cdot \mathbf{v} \rangle_0| + K_3 |\langle \mathbf{Z}(\mathbf{n}_k) \nabla \times \mathbf{u}, \nabla \times \mathbf{v} \rangle_0|$$

$$+|K_2-K_3|\Big(|\langle \mathbf{u}\cdot\nabla\times\mathbf{v},\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0|+|\langle \mathbf{n}_k\cdot\nabla\times\mathbf{v},\mathbf{u}\cdot\nabla\times\mathbf{n}_k\rangle_0|$$
$$+|\langle \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{u}\rangle_0|+|\langle \mathbf{n}_k\cdot\nabla\times\mathbf{u},\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0|$$
$$+|\langle \mathbf{u}\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0|\Big)+|\int_\Omega \lambda_k(\mathbf{u},\mathbf{v})\,dV|, \tag{3.19}$$

by the triangle inequality. For simplicity, consider the components of the sum above separately. Note that

$$|\langle\nabla\cdot\mathbf{u},\nabla\cdot\mathbf{v}\rangle_0|\le\|\nabla\cdot\mathbf{u}\|_0\|\nabla\cdot\mathbf{v}\|_0\le\|\mathbf{u}\|_{DC^1}\|\mathbf{v}\|_{DC^1}. \tag{3.20}$$

Considering $|\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{u},\nabla\times\mathbf{v}\rangle_0|$, using (3.15) implies that

$$\begin{aligned}|\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{u},\nabla\times\mathbf{v}\rangle_0|&\le\|\nabla\times\mathbf{v}\|_0\|\mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{u}\|_0\\&\le\Lambda\|\mathbf{v}\|_{DC^1}\|\nabla\times\mathbf{u}\|_0\\&\le\Lambda\|\mathbf{v}\|_{DC^1}\|\mathbf{u}\|_{DC^1}.\end{aligned} \tag{3.21}$$

From the imbedding in Lemma 2.5 of [56], if Ω is an open, bounded domain, with Lipschitz-continuous boundary, then for any $\mathbf{w}\in H_0(\mathrm{div},\Omega)\cap H_0(\mathrm{curl},\Omega)$ there exists a $C_{\mathrm{imb}}>0$ such that

$$\|\mathbf{w}\|_1^2\le C_{\mathrm{imb}}\|\mathbf{w}\|_{DC^1}^2.$$

Furthermore, $\mathbf{w}\in H_0^1(\Omega)^3$ by [56, Lemma 2.5]. Consider $|\langle\mathbf{u}\cdot\nabla\times\mathbf{v},\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0|$ from (3.19). By Corollary 1.1 in [56], the map $\mathbf{u}\cdot\nabla\times\mathbf{v}$ is a *continuous* bilinear map, $H^1(\Omega)^3\times H^1(\Omega)^3\to L^2(\Omega)$. Therefore, there exists a $C_{\mathrm{con}}>0$ such that

$$\|\mathbf{u}\cdot\nabla\times\mathbf{v}\|_0\le C_{\mathrm{con}}\|\mathbf{u}\|_1\|\nabla\times\mathbf{v}\|_1.$$

By the Cauchy-Schwarz inequality

$$|\langle\mathbf{u}\cdot\nabla\times\mathbf{v},\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0|\le\|\mathbf{u}\cdot\nabla\times\mathbf{v}\|_0\|\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\|_0. \tag{3.22}$$

Let $C' = C_{\text{con}} C_{\text{imb}}$ and note that

$$\|\mathbf{u} \cdot \nabla \times \mathbf{v}\|_0 \leq C_{\text{con}} \|\mathbf{u}\|_1 \|\nabla \times \mathbf{v}\|_1 \tag{3.23}$$

$$\leq C' \|\mathbf{u}\|_{DC^1} \|\mathbf{v}\|_{DC^1}, \tag{3.24}$$

where (3.23) is given by continuity and (3.24) is given by the imbedding. Hence,

$$\begin{aligned} |\langle \mathbf{u} \cdot \nabla \times \mathbf{v}, \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k \rangle_0| &\leq C' \|\mathbf{u}\|_{DC^1} \|\mathbf{v}\|_{DC^1} \|\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k\|_0 \\ &\leq C' \sqrt{\beta} \|\nabla \times \mathbf{n}_k\|_0 \|\mathbf{u}\|_{DC^1} \|\mathbf{v}\|_{DC^1}. \end{aligned} \tag{3.25}$$

The next summand from (3.19) is

$$|\langle \mathbf{n}_k \cdot \nabla \times \mathbf{v}, \mathbf{u} \cdot \nabla \times \mathbf{n}_k \rangle_0| \leq \|\mathbf{n}_k \cdot \nabla \times \mathbf{v}\|_0 \|\mathbf{u} \cdot \nabla \times \mathbf{n}_k\|_0.$$

Again bound

$$\|\mathbf{n}_k \cdot \nabla \times \mathbf{v}\|_0 \leq \sqrt{\beta} \|\mathbf{v}\|_{DC^1}.$$

Since $\mathbf{u} \in H_0^1(\Omega)^3$ and $\nabla \times \mathbf{n}_k \in H^1(\Omega)^3$,

$$\begin{aligned} \|\mathbf{u} \cdot \nabla \times \mathbf{n}_k\|_0 &\leq C_{\text{con}} \|\mathbf{u}\|_1 \|\nabla \times \mathbf{n}_k\|_1 \\ &\leq C' \|\mathbf{u}\|_{DC^1} \|\nabla \times \mathbf{n}_k\|_1. \end{aligned}$$

Therefore,

$$|\langle \mathbf{n}_k \cdot \nabla \times \mathbf{v}, \mathbf{u} \cdot \nabla \times \mathbf{n}_k \rangle_0| \leq \sqrt{\beta} C' \|\nabla \times \mathbf{n}_k\|_1 \|\mathbf{u}\|_{DC^1} \|\mathbf{v}\|_{DC^1}. \tag{3.26}$$

Now consider $|\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{u} \rangle_0|$ and note that this inner product is the same as that in (3.22) with the roles of $\mathbf{u}$ and $\mathbf{v}$ reversed. Since $\mathbf{u}$ and $\mathbf{v}$ are from the same space, the steps for deriving (3.25) are equally valid here. Thus,

$$|\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{u} \rangle_0| \leq C' \sqrt{\beta} \|\nabla \times \mathbf{n}_k\|_0 \|\mathbf{u}\|_{DC^1} \|\mathbf{v}\|_{DC^1}. \tag{3.27}$$

Similarly, the inequality for $|\langle \mathbf{n}_k \cdot \nabla \times \mathbf{u}, \mathbf{v} \cdot \nabla \times \mathbf{n}_k \rangle_0|$ is derived in an analogous manner to that of (3.26). Thus,

$$|\langle \mathbf{n}_k \cdot \nabla \times \mathbf{u}, \mathbf{v} \cdot \nabla \times \mathbf{n}_k \rangle_0| \leq \sqrt{\beta} C' \|\nabla \times \mathbf{n}_k\|_1 \|\mathbf{u}\|_{DC^1} \|\mathbf{v}\|_{DC^1}. \tag{3.28}$$

Next, examine

$$|\langle \mathbf{u} \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k \rangle_0| \leq \|\mathbf{u} \cdot \nabla \times \mathbf{n}_k\|_0 \|\mathbf{v} \cdot \nabla \times \mathbf{n}_k\|_0.$$

Since $\nabla \times \mathbf{n}_k \in H^1(\Omega)^3$,

$$\|\mathbf{u} \cdot \nabla \times \mathbf{n}_k\|_0 \leq C_{\text{con}} \|\mathbf{u}\|_1 \|\nabla \times \mathbf{n}_k\|_1 \leq C' \|\mathbf{u}\|_{DC^1} \|\nabla \times \mathbf{n}_k\|_1,$$
$$\|\mathbf{v} \cdot \nabla \times \mathbf{n}_k\|_0 \leq C_{\text{con}} \|\mathbf{v}\|_1 \|\nabla \times \mathbf{n}_k\|_1 \leq C' \|\mathbf{v}\|_{DC^1} \|\nabla \times \mathbf{n}_k\|_1.$$

Thus,

$$|\langle \mathbf{u} \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k \rangle_0| \leq (C')^2 \|\nabla \times \mathbf{n}_k\|_1^2 \|\mathbf{u}\|_{DC^1} \|\mathbf{v}\|_{DC^1}. \tag{3.29}$$

Finally,

$$\begin{aligned} \left| \int_\Omega \lambda_k(\mathbf{u}, \mathbf{v}) \, dV \right| &\leq \|\lambda_k\|_0 \|\mathbf{u} \cdot \mathbf{v}\|_0 \\ &\leq \|\lambda_k\|_0 C_{\text{con}} \|\mathbf{u}\|_1 \|\mathbf{v}\|_1 \\ &\leq \|\lambda_k\|_0 C' C_{\text{imb}} \|\mathbf{u}\|_{DC^1} \|\mathbf{v}\|_{DC^1}. \end{aligned} \tag{3.30}$$

Combining (3.20), (3.21), and (3.25)-(3.30),

$$\begin{aligned} a(\mathbf{u}, \mathbf{v}) \leq \Big(& K_1 + K_3 \Lambda + |K_2 - K_3| \big(2C' \sqrt{\beta} \|\nabla \times \mathbf{n}_k\|_0 + 2\sqrt{\beta} C' \|\nabla \times \mathbf{n}_k\|_1 \\ & + (C')^2 \|\nabla \times \mathbf{n}_k\|_1^2 \big) + \|\lambda_k\|_0 C' C_{\text{imb}} \Big) \|\mathbf{u}\|_{DC^1} \|\mathbf{v}\|_{DC^1}. \end{aligned}$$

□

In addition to these lemmas, Appendix B discusses a lemma proven in the course of studying the weak coercivity properties of the bilinear form, $b(\mathbf{v}, \gamma)$. While the framework in which this lemma was considered is not sufficient for demonstrating

weak coercivity, the lemma statement and proof are given in the appendix in order to inform further research addressing unit-length constrained problems.

The auxiliary regularity above poses a number of theoretical problems. For the well-posedness of the continuum system, coercivity and weak coercivity must be shown in the more intricate $\mathcal{H}^{DC^1}(\Omega)$-norm. Moreover, conforming finite elements for this space, such as Bogner-Fox-Schmit elements [13], are undesirably cumbersome and present notable difficulties in demonstrating stability for this linearization system. However, in the discrete setting, results guaranteeing the existence and uniqueness of solutions to the discrete Newton systems at each step are attained under less strict assumptions.

The existence and uniqueness theory to follow is explicitly developed in the presence of full Dirichlet boundary conditions. However, the theory is equally applicable for a rectangular domain with mixed Dirichlet and periodic boundary conditions. Such a domain is considered in the numerical experiments presented herein. In order for the theory discussed below to extend to the mixed Dirichlet and periodic boundary conditions case, it suffices to extend the results of Remark 2.7 in [56]. Here, for brevity, we present the extension under a slab-domain assumption, used in the numerical experiments below, such that $\mathbf{v}$ may have a nonzero z-component but $\frac{\partial \mathbf{v}}{\partial z} = \mathbf{0}$.

Lemma 3.5.5 *If Ω is a rectangular domain and $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$ with mixed periodic and Dirichlet boundary conditions, then there exists $C_p > 0$ such that*

$$\|\nabla \mathbf{v}\|_0^2 \leq C_p(\|\nabla \cdot \mathbf{v}\|_0^2 + \|\nabla \times \mathbf{v}\|_0^2).$$

Proof: Without loss of generality assume periodicity of $\mathbf{v}(x, y)$ in x with unit-length period. Let $\mathbf{v}_0$ denote the extension of $\mathbf{v}$ by zero in the y-direction outside of Ω. Denote the space for which $\mathbf{v}_0$ is defined as $\Omega^E = [0, 1] \times \mathbb{R}$. As in [56, Lemma 2.5], $\mathbf{v}_0 \in H(\mathrm{div}, \Omega^E) \cap H(\mathrm{curl}, \Omega^E)$ with

$$\|\nabla \cdot \mathbf{v}_0\|_{0,\, \Omega^E} = \|\nabla \cdot \mathbf{v}\|_0,$$

$$\|\nabla \times \mathbf{v}_0\|_{0,\,\Omega^E} = \|\nabla \times \mathbf{v}\|_0,$$

$$\|\mathbf{v}_0\|_{0,\,\Omega^E} = \|\mathbf{v}\|_0.$$

Since $\mathbf{v}_0(x, y)$ is periodic in x, we define the Fourier transform on Ω^E as

$$\mathscr{F}\mathbf{v}_0 = \int_0^1 \int_{-\infty}^{\infty} \mathbf{v}_0(x, y) e^{-2i\pi\sigma x} e^{-2i\pi\mu y}\, dy\, dx.$$

Note that $\mathscr{F}\mathbf{v}_0(\sigma, \mu)$ is a function of $\mathbb{Z} \times \mathbb{R}$ and

$$\|\mathscr{F}\mathbf{v}_0\|^2_{0,\,\mathbb{Z}\times\mathbb{R}} = \sum_{\sigma=-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \mathscr{F}\mathbf{v}_0 \cdot \overline{\mathscr{F}\mathbf{v}_0}\, d\mu \right),$$

where the overbar notation represents complex conjugation. By Plancherel's theorem, as in [96], if $f \in L^2(\Omega^E)$, then

$$\|\mathscr{F}(f)\|_{0,\,\mathbb{Z}\times\mathbb{R}} = \|f\|_{0,\,\Omega^E}.$$

Denote the components of $\mathbf{v}_0$ as $\mathbf{v}_0^{(j)}$, for $j = 1, 2, 3$. Computing derivatives, assuming $\frac{\partial \mathbf{v}}{\partial z} = \mathbf{0}$,

$$\mathscr{F}(\nabla \times \mathbf{v}_0) = 2i\pi(\mu\mathscr{F}\mathbf{v}_0^{(3)}, -\sigma\mathscr{F}\mathbf{v}_0^{(3)}, \sigma\mathscr{F}\mathbf{v}_0^{(2)} - \mu\mathscr{F}\mathbf{v}_0^{(1)}),$$
$$\mathscr{F}(\nabla \cdot \mathbf{v}_0) = 2i\pi(\sigma\mathscr{F}\mathbf{v}_0^{(1)} + \mu\mathscr{F}\mathbf{v}_0^{(2)}).$$

Furthermore,

$$\|\mathscr{F}\Big(\frac{\partial \mathbf{v}_0^{(j)}}{\partial y}\Big)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}} = \|2i\pi\mu\mathscr{F}\mathbf{v}_0^{(j)}\|^2_{0,\,\mathbb{Z}\times\mathbb{R}},$$
$$\|\mathscr{F}\Big(\frac{\partial \mathbf{v}_0^{(j)}}{\partial x}\Big)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}} = \|2i\pi\sigma\mathscr{F}\mathbf{v}_0^{(j)}\|^2_{0,\,\mathbb{Z}\times\mathbb{R}},$$

and

$$\|\mathscr{F}(\nabla \times \mathbf{v}_0)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}} + \|\mathscr{F}(\nabla \cdot \mathbf{v}_0)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}} =$$

$$\sum_{\sigma=-\infty}^{\infty}\int_{-\infty}^{\infty}\big(\mathscr{F}(\nabla\times\mathbf{v}_0)\cdot(\overline{\mathscr{F}(\nabla\times\mathbf{v}_0)})+(\mathscr{F}(\nabla\cdot\mathbf{v}_0))(\overline{\mathscr{F}(\nabla\cdot\mathbf{v}_0)})\big)\,d\mu,$$

with

$$\begin{aligned}\mathscr{F}(\nabla\times\mathbf{v}_0)\cdot\overline{\mathscr{F}(\nabla\times\mathbf{v}_0)} &= 4\pi^2\Big((\mu\mathscr{F}\mathbf{v}_0^{(3)})^2+(\sigma\mathscr{F}\mathbf{v}_0^{(3)})^2+(\sigma\mathscr{F}\mathbf{v}_0^{(2)})^2+(\mu\mathscr{F}\mathbf{v}_0^{(1)})^2\\ &\qquad -2\sigma\mu\mathscr{F}\mathbf{v}_0^{(1)}\mathscr{F}\mathbf{v}_0^{(2)}\Big),\\ (\mathscr{F}(\nabla\cdot\mathbf{v}_0))(\overline{\mathscr{F}(\nabla\cdot\mathbf{v}_0)}) &= 4\pi^2\Big((\mu\mathscr{F}\mathbf{v}_0^{(2)})^2+(\sigma\mathscr{F}\mathbf{v}_0^{(1)})^2+2\sigma\mu\mathscr{F}\mathbf{v}_0^{(1)}\mathscr{F}\mathbf{v}_0^{(2)}\Big).\end{aligned}$$

So,

$$\begin{aligned}&\mathscr{F}(\nabla\times\mathbf{v}_0)\cdot\overline{\mathscr{F}(\nabla\times\mathbf{v}_0)}+(\mathscr{F}(\nabla\cdot\mathbf{v}_0))(\overline{\mathscr{F}(\nabla\cdot\mathbf{v}_0)})=\\ &\qquad 4\pi^2\Big((\mu\mathscr{F}\mathbf{v}_0^{(2)})^2+(\sigma\mathscr{F}\mathbf{v}_0^{(1)})^2+(\mu\mathscr{F}\mathbf{v}_0^{(3)})^2+(\sigma\mathscr{F}\mathbf{v}_0^{(3)})^2\\ &\qquad +(\sigma\mathscr{F}\mathbf{v}_0^{(2)})^2+(\mu\mathscr{F}\mathbf{v}_0^{(1)})^2\Big).\end{aligned}$$

This implies that

$$\|\mathscr{F}(\nabla\times\mathbf{v}_0)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}}+\|\mathscr{F}(\nabla\cdot\mathbf{v}_0)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}}=\sum_{k=1}^{3}\|\mathscr{F}\big(\frac{\partial\mathbf{v}_0^{(k)}}{\partial y}\big)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}}+\sum_{j=1}^{3}\|\mathscr{F}\big(\frac{\partial\mathbf{v}_0^{(j)}}{\partial x}\big)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}}.$$

Hence,

$$\begin{aligned}\|\mathscr{F}\big(\frac{\partial\mathbf{v}_0^{(j)}}{\partial x}\big)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}} &\le \|\mathscr{F}(\nabla\times\mathbf{v}_0)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}}+\|\mathscr{F}(\nabla\cdot\mathbf{v}_0)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}}\\ &= \|\nabla\cdot\mathbf{v}_0\|^2_{0,\,\Omega^E}+\|\nabla\times\mathbf{v}_0\|_{0,\,\Omega^E}\\ &= \|\nabla\cdot\mathbf{v}\|_0^2+\|\nabla\times\mathbf{v}\|_0^2.\end{aligned}$$

Similarly,

$$\|\mathscr{F}\big(\frac{\partial\mathbf{v}_0^{(j)}}{\partial y}\big)\|^2_{0,\,\mathbb{Z}\times\mathbb{R}}\le\|\nabla\cdot\mathbf{v}\|_0^2+\|\nabla\times\mathbf{v}\|_0^2.$$

Therefore, there exists a $C_p>0$ such that

$$\|\nabla\mathbf{v}\|_0^2=\|\nabla\mathbf{v}\|^2_{0,\,\Omega^E}\le C_p(\|\nabla\cdot\mathbf{v}\|_0^2+\|\nabla\times\mathbf{v}\|_0^2),$$

and

$$\|\mathbf{v}\|_1^2 \leq (C_p + 1)(\|\mathbf{v}\|_0^2 + \|\nabla \cdot \mathbf{v}\|_0^2 + \|\nabla \times \mathbf{v}\|_0^2). \qquad \square$$

It is important to note that this proof generalizes to non-slab domain cases.

3.5.1 Discrete System Preliminaries

Performing the outlined Newton iterations for free-elastic effects necessitates solving the Newton systems, discussed above, to obtain update functions $\delta\mathbf{n}$ and $\delta\lambda$. Thus, finite elements are used to numerically approximate the updates. Finite-dimensional spaces $V_h \subset \mathcal{H}_0^{DC}(\Omega)$ and $\Pi_h \subset L^2(\Omega)$ are considered, yielding the discrete variational problem

$$a(\delta\mathbf{n}_h, \mathbf{v}_h) + b(\mathbf{v}_h, \delta\lambda_h) = F(\mathbf{v}_h), \qquad \forall \mathbf{v}_h \in V_h, \tag{3.31}$$

$$b(\delta\mathbf{n}_h, \gamma_h) = G(\gamma_h), \qquad \forall \gamma_h \in \Pi_h. \tag{3.32}$$

Throughout the rest of this section, the developed theory applies exclusively to discrete spaces. Therefore, except when necessary for clarity, we drop the subscript h along with the notation $\delta\mathbf{n}$ and $\delta\lambda$. For instance, we write $a(\mathbf{u}, \mathbf{v})$ to indicate the bilinear form in (3.31) operating on the discrete space $V_h \times V_h$.

Let $\{\mathcal{T}_h\}$, $0 < h \leq 1$, be a family of quadrilateral subdivisions of Ω, such that

$$\max\{\text{diam } T : T \in \mathcal{T}_h\} \leq h \text{ diam } \Omega. \tag{3.33}$$

Further, assume that $\{\mathcal{T}_h\}$ is quasi-uniform so that there exists a $\rho > 0$, such that

$$\min\{\text{diam } B_T : T \in \mathcal{T}_h\} \geq \rho\ h \text{ diam } \Omega, \tag{3.34}$$

for all $h \in (0, 1]$, where B_T is the largest ball contained in T, such that T is star-shaped with respect to B_T [16]. Denote the measure of any $T \in \mathcal{T}_h$ as $|T|$. Furthermore, let Q_p denote piecewise C^0 polynomials of degree $p \geq 1$ on $\mathcal{T}_h$ and P_0 denote

the space of piecewise constants on $\mathcal{T}_h$. Next, define a bubble space

$$V_h^b = \{\mathbf{v} \in C_c(\bar{\Omega})^3 : \mathbf{v}|_T = a_T b_T \mathbf{n}_k|_T, \forall T \in \mathcal{T}_h\},$$

where $C_c(\bar{\Omega})$ denotes the space of compactly supported continuous functions on the closure of Ω, b_T is the bi- or tri-quadratic bubble function [99], depending on dimension, that vanishes on $\partial T \in \mathcal{T}_h$, and a_T is a constant coefficient associated with b_T. The bubble functions are constructed [104], such that

$$\int_T b_T \, dV = 1, \qquad \forall T \in \mathcal{T}_h, \tag{3.35}$$

$$b_T > 0, \qquad \forall \mathbf{x} \in T. \tag{3.36}$$

Then, we consider the pair of spaces

$$\Pi_h = P_0, \tag{3.37}$$

$$V_h = \{\mathbf{v} \in Q_m \times Q_m \times Q_m \oplus V_h^b : \mathbf{v} = \mathbf{0} \text{ on } \partial\Omega\}. \tag{3.38}$$

In the following sections, to demonstrate the existence and uniqueness of solutions to the system given by (3.31) and (3.32), we show that $a(\mathbf{u}, \mathbf{v})$ is a coercive and continuous bilinear form and that $b(\mathbf{v}, \gamma)$ is a continuous and weakly coercive bilinear form [7, 12, 14, 16] for the above spaces, V_h and Π_h. We further assume that $\mathbf{n}_k \in Q_p \times Q_p \times Q_p$, for some $p \geq 1$, so that $V_h \subset Q_l \times Q_l \times Q_l$ for $l = \max(m, p+2)$.

3.5.2 Discrete Continuity

In this section, we show that the right-hand-sides of (3.31) and (3.32) are continuous linear functionals and that the bilinear forms $a(\mathbf{u}, \mathbf{v})$ and $b(\mathbf{v}, \gamma)$ are continuous for the assumptions discussed above.

Lemma 3.5.6 *Under Assumption 3.5.1, F and G are bounded linear functionals on V_h and Π_h, respectively.*

Proof: A simple application of the Cauchy-Schwarz inequality shows that $G(\gamma)$ is a bounded linear functional.

For $F(\mathbf{v})$, observe that

$$|F(\mathbf{v})| \leq K_1|\langle \nabla \cdot \mathbf{n}_k, \nabla \cdot \mathbf{v}\rangle_0| + K_3|\langle \mathbf{Z}(\mathbf{n}_k)\nabla \times \mathbf{n}_k, \nabla \times \mathbf{v}\rangle_0| + |K_2 - K_3||\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0| + \left|\int_\Omega \lambda_k(\mathbf{n}_k, \mathbf{v})\, dV\right|, \tag{3.39}$$

by the triangle inequality. Applying Cauchy-Schwarz inequalities to (3.39), one obtains

$$\begin{aligned} |F(\mathbf{v})| &\leq K_1\|\nabla \cdot \mathbf{n}_k\|_0\|\nabla \cdot \mathbf{v}\|_0 + K_3\|\mathbf{Z}(\mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0\|\nabla \times \mathbf{v}\|_0 \\ &\quad + |K_2 - K_3||\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0| + \|\lambda_k \mathbf{n}_k\|_0\|\mathbf{v}\|_0 \\ &\leq K_1\|\nabla \cdot \mathbf{n}_k\|_0\|\mathbf{v}\|_{DC} + K_3\|\mathbf{Z}(\mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0\|\mathbf{v}\|_{DC} \\ &\quad + |K_2 - K_3||\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0| + \|\lambda_k \mathbf{n}_k\|_0\|\mathbf{v}\|_{DC}. \end{aligned} \tag{3.40}$$

In order to bound $|F(\mathbf{v})|$, consider the final three summands separately. Note that since $|\mathbf{Z}(\mathbf{n}_k)| \leq \Lambda$, where Λ is the relevant upper bound from Lemma 3.4.1, it is evident that

$$\|\mathbf{Z}(\mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0 \leq \Lambda\|\nabla \times \mathbf{n}_k\|_0, \tag{3.41}$$

and that

$$\|\lambda_k \mathbf{n}_k\|_0^2 \leq \beta \int_\Omega \lambda_k^2\, dV = C_1^2, \tag{3.42}$$

where β is the upper bound in (3.9). Finally, consider

$$|\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0| = |\langle (\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k, \mathbf{v}\rangle_0|.$$

Applying the Cauchy-Schwarz inequality,

$$|\langle (\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k, \mathbf{v}\rangle_0| \leq \|(\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0\|\mathbf{v}\|_0$$

$$\leq \|(\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0 \|\mathbf{v}\|_{DC}. \tag{3.43}$$

Next, note that

$$\begin{aligned}(\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k \cdot (\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k &= (\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)^2 (\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k) \\ &\leq (|\mathbf{n}_k| \cdot |\nabla \times \mathbf{n}_k|)^2 (\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k) \\ &\leq \beta \cdot (\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)^2. \end{aligned} \tag{3.44}$$

Furthermore, $\nabla \times \mathbf{n}_k$ is a vector of piecewise polynomials. Therefore, $\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k \in L^2(\Omega)$. Employing (3.44) and letting $\|\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k\|_0 = C_2$,

$$\begin{aligned}\|(\mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)\nabla \times \mathbf{n}_k\|_0 &\leq \sqrt{\beta}\Big(\int_\Omega (\nabla \times \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k)^2 \, dV\Big)^{1/2} \\ &\leq \sqrt{\beta} C_2. \end{aligned} \tag{3.45}$$

Therefore, using (3.40)-(3.43), and (3.45),

$$\begin{aligned}|F(\mathbf{v})| \leq & K_1 \|\nabla \cdot \mathbf{n}_k\|_0 \|\mathbf{v}\|_{DC} + K_3 \Lambda \|\nabla \times \mathbf{n}_k\|_0 \|\mathbf{v}\|_{DC} \\ & + |K_2 - K_3| \sqrt{\beta} C_2 \|\mathbf{v}\|_{DC} + C_1 \|\mathbf{v}\|_{DC},\end{aligned}$$

implying $F(\mathbf{v})$ is a bounded linear functional on V_h. □

Lemma 3.5.7 *Under Assumption 3.5.1, $a(\mathbf{u}, \mathbf{v})$ and $b(\mathbf{v}, \gamma)$ are continuous.*

Proof: First consider

$$\begin{aligned}|b(\mathbf{v}, \gamma)| &= \left| \int_\Omega \gamma(\mathbf{v}, \mathbf{n}_k) \, dV \right| \\ &\leq \|\gamma\|_0 \|\mathbf{v} \cdot \mathbf{n}_k\|_0 \\ &\leq \|\gamma\|_0 \sqrt{\beta} \|\mathbf{v}\|_0, \end{aligned}$$

by Hölder's inequality and (3.9). Therefore, $b(\mathbf{v}, \gamma)$ is a continuous bilinear form.

For the continuity of $a(\mathbf{u},\mathbf{v})$, observe that

$$\begin{aligned}|a(\mathbf{u},\mathbf{v})| \leq & K_1|\langle \nabla\cdot\mathbf{u}, \nabla\cdot\mathbf{v}\rangle_0| + K_3|\langle \mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{u}, \nabla\times\mathbf{v}\rangle_0| \\ & + |K_2-K_3|\Big(|\langle \mathbf{u}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0| + |\langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \mathbf{u}\cdot\nabla\times\mathbf{n}_k\rangle_0| \\ & + |\langle \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{u}\rangle_0| + |\langle \mathbf{n}_k\cdot\nabla\times\mathbf{u}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0| \\ & + |\langle \mathbf{u}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0|\Big) + \left|\int_\Omega \lambda_k(\mathbf{u},\mathbf{v})\,dV\right|, \end{aligned} \tag{3.46}$$

by the triangle inequality. For simplicity, consider the components of the sum above. Note that

$$|\langle \nabla\cdot\mathbf{u}, \nabla\cdot\mathbf{v}\rangle_0| \leq \|\nabla\cdot\mathbf{u}\|_0\|\nabla\cdot\mathbf{v}\|_0 \leq \|\mathbf{u}\|_{DC}\|\mathbf{v}\|_{DC}. \tag{3.47}$$

Considering $|\langle \mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{u}, \nabla\times\mathbf{v}\rangle_0|$, using (3.41) implies that

$$\begin{aligned}|\langle \mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{u}, \nabla\times\mathbf{v}\rangle_0| &\leq \|\nabla\times\mathbf{v}\|_0\|\mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{u}\|_0 \\ &\leq \Lambda\|\mathbf{v}\|_{DC}\|\nabla\times\mathbf{u}\|_0 \\ &\leq \Lambda\|\mathbf{v}\|_{DC}\|\mathbf{u}\|_{DC}. \end{aligned} \tag{3.48}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}|\langle \mathbf{u}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0| &= |\langle (\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k)\mathbf{u}, \nabla\times\mathbf{v}\rangle_0| \\ &\leq \|(\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k)\mathbf{u}\|_0\|\nabla\times\mathbf{v}\|_0. \end{aligned} \tag{3.49}$$

Note that

$$(\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k)^2 \leq |\mathbf{n}_k|^2|\nabla\times\mathbf{n}_k|^2 \leq \beta|\nabla\times\mathbf{n}_k|^2.$$

Furthermore, since $\nabla\times\mathbf{n}_k$ is a vector of piecewise polynomials, $|\nabla\times\mathbf{n}_k|^2$ is bounded. Letting $C_{\sup} = \sup_{\mathbf{x}\in\Omega}|\nabla\times\mathbf{n}_k|^2$,

$$\begin{aligned}\|(\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k)\mathbf{u}\|_0 &= \left(\int_\Omega (\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k)^2(\mathbf{u}\cdot\mathbf{u})\,dV\right)^{1/2} \\ &\leq \sqrt{\beta}\left(\int_\Omega |\nabla\times\mathbf{n}_k|^2(\mathbf{u}\cdot\mathbf{u})\,dV\right)^{1/2} \end{aligned}$$

$$\leq \sqrt{\beta C_{\sup}}\|\mathbf{u}\|_0.$$

Hence,

$$|\langle \mathbf{u}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0| \leq \sqrt{\beta C_{\sup}}\|\mathbf{u}\|_{DC}\|\mathbf{v}\|_{DC}. \tag{3.50}$$

The next summand from (3.46) is

$$|\langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \mathbf{u}\cdot\nabla\times\mathbf{n}_k\rangle_0| \leq \|\mathbf{n}_k\cdot\nabla\times\mathbf{v}\|_0\|\mathbf{u}\cdot\nabla\times\mathbf{n}_k\|_0,$$

with

$$\|\mathbf{n}_k\cdot\nabla\times\mathbf{v}\|_0 \leq \sqrt{\beta}\|\mathbf{v}\|_{DC}.$$

Furthermore,

$$\|\mathbf{u}\cdot\nabla\times\mathbf{n}_k\|_0 \leq \sqrt{C_{\sup}}\|\mathbf{u}\|_0.$$

Therefore,

$$|\langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \mathbf{u}\cdot\nabla\times\mathbf{n}_k\rangle_0| \leq \sqrt{\beta C_{\sup}}\|\mathbf{v}\|_{DC}\|\mathbf{u}\|_{DC}. \tag{3.51}$$

Now consider $|\langle \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{u}\rangle_0|$ and note that this inner product is the same as that in (3.49) with the roles of $\mathbf{u}$ and $\mathbf{v}$ reversed. Since $\mathbf{u}$ and $\mathbf{v}$ are from the same space, the steps for deriving (3.50) are equally valid. Thus,

$$|\langle \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{u}\rangle_0| \leq \sqrt{\beta C_{\sup}}\|\mathbf{u}\|_{DC}\|\mathbf{v}\|_{DC}. \tag{3.52}$$

Similarly, the inequality for $|\langle \mathbf{n}_k\cdot\nabla\times\mathbf{u}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0|$ is derived in an analogous manner to that of (3.51). Thus,

$$|\langle \mathbf{n}_k\cdot\nabla\times\mathbf{u}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0| \leq \sqrt{\beta C_{\sup}}\|\mathbf{v}\|_{DC}\|\mathbf{u}\|_{DC}. \tag{3.53}$$

Next, examine

$$|\langle \mathbf{u}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0| \leq \|\mathbf{u}\cdot\nabla\times\mathbf{n}_k\|_0\|\mathbf{v}\cdot\nabla\times\mathbf{n}_k\|_0.$$

Since $\nabla \times \mathbf{n}_k$ is a vector of piecewise polynomials,

$$\|\mathbf{u} \cdot \nabla \times \mathbf{n}_k\|_0 \leq \sqrt{C_{\sup}}\|\mathbf{u}\|_0,$$
$$\|\mathbf{v} \cdot \nabla \times \mathbf{n}_k\|_0 \leq \sqrt{C_{\sup}}\|\mathbf{v}\|_0.$$

Thus,

$$|\langle \mathbf{u} \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k \rangle_0| \leq C_{\sup}\|\mathbf{u}\|_{DC}\|\mathbf{v}\|_{DC}. \tag{3.54}$$

Finally, since λ_k is piecewise constant, λ_k^2 is bounded. Letting $C_\lambda = \sup_{\mathbf{x}\in\Omega} \lambda_k^2$,

$$\begin{aligned}\left|\int_\Omega \lambda_k(\mathbf{u},\mathbf{v})\, dV\right| &\leq \|\lambda_k \mathbf{u}\|_0\|\mathbf{v}\|_0 \\ &\leq \sqrt{C_\lambda}\|\mathbf{u}\|_0\|\mathbf{v}\|_{DC} \\ &\leq \sqrt{C_\lambda}\|\mathbf{u}\|_{DC}\|\mathbf{v}\|_{DC}. \end{aligned} \tag{3.55}$$

Combining (3.47), (3.48), and (3.50)-(3.55),

$$a(\mathbf{u},\mathbf{v}) \leq \Big(K_1 + K_3\Lambda + |K_2 - K_3|\big(4\sqrt{\beta C_{\sup}} + C_{\sup}\big) + \sqrt{C_\lambda}\Big)\|\mathbf{u}\|_{DC}\|\mathbf{v}\|_{DC}. \qquad \square$$

3.5.3 Discrete Coercivity

In this section, two proofs of the coercivity of $a(\mathbf{u},\mathbf{v})$ are given. The first is for the case when $\kappa = K_2/K_3 = 1$. The second addresses coercivity when κ lies in a neighborhood of unity. For both proofs, we use the additional assumption that the approximation is close enough to the solution such that the Lagrange multiplier, λ_k, is pointwise non-negative. This assumption is reasonable since at the solution, $\mathbf{n}_*$, λ_* may be chosen arbitrarily.

Lemma 3.5.8 *Under Assumption 3.5.1 and the assumption that λ_k is pointwise non-negative, if $\kappa = 1$, there exists an $\alpha_0 > 0$ such that $\alpha_0\|\mathbf{v}\|_{DC}^2 \leq a(\mathbf{v},\mathbf{v})$ for all $\mathbf{v} \in V_h$.*

Proof: Note that since $\kappa = 1$, $(K_2 - K_3) = 0$, and

$$a(\mathbf{v},\mathbf{v}) = K_1\langle \nabla\cdot\mathbf{v}, \nabla\cdot\mathbf{v}\rangle_0 + K_3\langle \nabla\times\mathbf{v}, \nabla\times\mathbf{v}\rangle_0 + \int_\Omega \lambda_k(\mathbf{v},\mathbf{v})\,dV.$$

Thus, it remains to show that there exists $\alpha_0 > 0$ such that

$$\alpha_0\|\mathbf{v}\|_{DC}^2 \leq K_1\langle \nabla\cdot\mathbf{v}, \nabla\cdot\mathbf{v}\rangle_0 + K_3\langle \nabla\times\mathbf{v}, \nabla\times\mathbf{v}\rangle_0 + \int_\Omega \lambda_k(\mathbf{v},\mathbf{v})\,dV.$$

From Remark 2.7 in [56], there exists $C_3 > 0$ such that

$$\|\nabla\mathbf{v}\|_0^2 \leq C_3^2\big(\|\nabla\cdot\mathbf{v}\|_0^2 + \|\nabla\times\mathbf{v}\|_0^2\big).$$

Moreover, recall that $\|\mathbf{v}\|_0^2 \leq C_4\|\nabla\mathbf{v}\|_0^2$ by the classical Poincaré-Friedrichs' inequality. Hence, for $C = C_4C_3^2 > 0$,

$$\|\mathbf{v}\|_0^2 \leq C\big(\|\nabla\cdot\mathbf{v}\|_0^2 + \|\nabla\times\mathbf{v}\|_0^2\big). \tag{3.56}$$

Since $\|\mathbf{v}\|_{DC}^2 = \|\mathbf{v}\|_0^2 + \|\nabla\cdot\mathbf{v}\|_0^2 + \|\nabla\times\mathbf{v}\|_0^2$, then

$$\|\mathbf{v}\|_{DC}^2 \leq (C+1)\big(\|\nabla\cdot\mathbf{v}\|_0^2 + \|\nabla\times\mathbf{v}\|_0^2\big).$$

Letting $K = \min(K_1, K_3) > 0$ and $\alpha_0 = K/(C+1)$, it follows that

$$\alpha_0\|\mathbf{v}\|_{DC}^2 \leq K\big(\|\nabla\cdot\mathbf{v}\|_0^2 + \|\nabla\times\mathbf{v}\|_0^2\big) \leq K_1\|\nabla\cdot\mathbf{v}\|_0^2 + K_3\|\nabla\times\mathbf{v}\|_0^2. \tag{3.57}$$

Finally, it was assumed that λ_k is pointwise non-negative, implying

$$\int_\Omega \lambda_k(\mathbf{v},\mathbf{v})\,dV \geq 0.$$

Therefore, (3.57) implies that

$$\alpha_0\|\mathbf{v}\|_{DC}^2 \leq K_1\langle \nabla\cdot\mathbf{v}, \nabla\cdot\mathbf{v}\rangle_0 + K_3\langle \nabla\times\mathbf{v}, \nabla\times\mathbf{v}\rangle_0 + \int_\Omega \lambda_k(\mathbf{v},\mathbf{v})\,dV. \qquad \square$$

The assumption that $\kappa = 1$ is a common modeling approach. In fact, this supposition represents a weaker constraint than is seen in the many models that utilize the one-constant approximation, cf. [25, 105, 117, 127]. However, it is possible to loosen the restriction that $\kappa = 1$ and still maintain the coercivity of $a(\mathbf{u}, \mathbf{v})$ with a small data type assumption on κ. That is, we assume that κ varies within a certain, possibly small, range of unity. Small data assumptions are common, for instance, in the study of solutions to the Navier-Stokes' equations [52, 77, 93], where bounds are imposed on certain norms of the initial data in order to demonstrate existence and uniqueness of solutions.

Lemma 3.5.9 (Small Data) *Under Assumption 3.5.1 and the assumption that λ_k is pointwise non-negative, there exists $\epsilon_1, \epsilon_2 > 0$, dependent on $\beta = \max |\mathbf{n}_k|^2$, such that if $\kappa \in (1 - \epsilon_2, 1 + \epsilon_1)$, then $a(\mathbf{u}, \mathbf{v})$ is coercive.*

Proof: Since $\mathbf{Z}(\mathbf{n}_k)$ is USPD by assumption,

$$\eta K_3 \langle \nabla \times \mathbf{v}, \nabla \times \mathbf{v} \rangle_0 \leq K_3 \langle \mathbf{Z}(\mathbf{n}_k) \nabla \times \mathbf{v}, \nabla \times \mathbf{v} \rangle_0,$$

where η is the relevant lower bound from Lemma 3.4.1. Defining $K' = \min(K_1, \eta K_3) > 0$ and $\alpha_1 = K'/(C + 1)$, where $C = C_4 C_3^2$ is the constant defined in (3.56), then,

$$\alpha_1 \|\mathbf{v}\|_{DC}^2 \leq K_1 \langle \nabla \cdot \mathbf{v}, \nabla \cdot \mathbf{v} \rangle_0 + \eta K_3 \langle \nabla \times \mathbf{v}, \nabla \times \mathbf{v} \rangle_0.$$

Thus, using the assumption that λ_k is pointwise non-negative,

$$\alpha_1 \|\mathbf{v}\|_{DC}^2 \leq K_1 \langle \nabla \cdot \mathbf{v}, \nabla \cdot \mathbf{v} \rangle_0 + K_3 \langle \mathbf{Z}(\mathbf{n}_k) \nabla \times \mathbf{v}, \nabla \times \mathbf{v} \rangle_0 + \int_\Omega \lambda_k(\mathbf{v}, \mathbf{v}) \, dV. \quad (3.58)$$

It should be noted that the constant η may depend on κ. Thus, the following three cases are considered.

Case 1. $\kappa = 1 + \epsilon_1$, for $\epsilon_1 > 0$.

If this case holds, then $\eta = 1$. Hence, α_1, defined for (3.58), is independent of κ.

Since $K_2 - K_3 = K_3(\kappa - 1)$, the discrete bilinear form of (3.12) becomes

$$\begin{aligned} a(\mathbf{v},\mathbf{v}) =& K_1\langle \nabla\cdot\mathbf{v}, \nabla\cdot\mathbf{v}\rangle_0 + K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{v}, \nabla\times\mathbf{v}\rangle_0 \\ &+ \epsilon_1 K_3\Big(2\langle \mathbf{v}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0 + 2\langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 \\ &+ \langle \mathbf{v}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big) + \int_\Omega \lambda_k(\mathbf{v},\mathbf{v})\, dV. \end{aligned} \tag{3.59}$$

Observe that from (3.58),

$$\begin{aligned} \alpha_1\|\mathbf{v}\|_{DC}^2 \leq& K_1\langle \nabla\cdot\mathbf{v}, \nabla\cdot\mathbf{v}\rangle_0 + K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{v}, \nabla\times\mathbf{v}\rangle_0 + \int_\Omega \lambda_k(\mathbf{v},\mathbf{v})\, dV \\ &+ \epsilon_1 K_3\langle \mathbf{v}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0. \end{aligned} \tag{3.60}$$

Consider the magnitude of the terms in (3.59) not bounded from below in (3.60), denoted as $\mathcal{G}(\mathbf{v},\mathbf{v})$,

$$\begin{aligned} |\mathcal{G}(\mathbf{v},\mathbf{v})| &= |2\epsilon_1 K_3\big(\langle \mathbf{v}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0 + \langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\big)| \\ &\leq 2\epsilon_1 K_3\big(|\langle \mathbf{v}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0| + \|\mathbf{n}_k\cdot\nabla\times\mathbf{v}\|_0\|\mathbf{v}\cdot\nabla\times\mathbf{n}_k\|_0\big). \end{aligned}$$

Using bounds derived in the proof of Lemma 3.5.7,

$$|\mathcal{G}(\mathbf{v},\mathbf{v})| \leq 4\epsilon_1 K_3\sqrt{\beta C_{\sup}}\|\mathbf{v}\|_{DC}^2.$$

Denoting $\alpha_3 = 4K_3\sqrt{\beta C_{\sup}}$, then

$$|\mathcal{G}(\mathbf{v},\mathbf{v})| \leq \epsilon_1\alpha_3\|\mathbf{v}\|_{DC}^2.$$

Utilizing (3.60),

$$a(\mathbf{v},\mathbf{v}) \geq \alpha_1\|\mathbf{v}\|_{DC}^2 - \epsilon_1\alpha_3\|\mathbf{v}\|_{DC}^2 = (\alpha_1 - \epsilon_1\alpha_3)\|\mathbf{v}\|_{DC}^2.$$

It is, thus, sufficient to have $\epsilon_1 < \alpha_1/\alpha_3$, guaranteeing that $(\alpha_1 - \epsilon_1\alpha_3) > 0$.

Case 2. $\kappa = 1 - \epsilon_2 > 0$, for $\epsilon_2 > 0$, and $K_1 < K_3$.

Since $\kappa < 1$, $\eta = 1+(\kappa-1)\beta = (1-\epsilon_2\beta)$. For $K_1 < K_3$, there exists an ϵ_2 small enough, such that $K_1 < (1-\epsilon_2\beta)K_3$. This implies that, for small enough ϵ_2,

$$\alpha_1 = \frac{\min(K_1, (1-\epsilon_2\beta)K_3)}{(C+1)} = \frac{K_1}{(C+1)}.$$

Therefore, α_1 is again independent of κ. Since $K_2 - K_3 = K_3(\kappa - 1)$, the discrete bilinear form of (3.12) becomes

$$\begin{aligned} a(\mathbf{v},\mathbf{v}) =& K_1\langle \nabla\cdot\mathbf{v}, \nabla\cdot\mathbf{v}\rangle_0 + K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{v}, \nabla\times\mathbf{v}\rangle_0 \\ &- \epsilon_2 K_3\Big(2\langle \mathbf{v}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0 + 2\langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 \\ &+ \langle \mathbf{v}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big) + \int_\Omega \lambda_k(\mathbf{v},\mathbf{v})\, dV. \end{aligned} \tag{3.61}$$

The terms of (3.61), not already bounded from below in (3.58), are bounded as

$$\begin{aligned} |\mathcal{G}(\mathbf{v},\mathbf{v})| =& |\epsilon_2 K_3\Big(2\langle \mathbf{v}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0 \\ &+ 2\langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 + \langle \mathbf{v}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big)| \\ \le& \epsilon_2 K_3\Big(2|\langle \mathbf{v}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0| \\ &+ 2\|\mathbf{n}_k\cdot\nabla\times\mathbf{v}\|_0\|\mathbf{v}\cdot\nabla\times\mathbf{n}_k\|_0 + \|\mathbf{v}\cdot\nabla\times\mathbf{n}_k\|_0\|\mathbf{v}\cdot\nabla\times\mathbf{n}_k\|_0\Big). \end{aligned}$$

Again using the bounds derived in the proof of Lemma 3.5.7,

$$|\mathcal{G}(\mathbf{v},\mathbf{v})| \le \epsilon_2 K_3\Big(4\sqrt{\beta C_{\text{sup}}} + C_{\text{sup}}\Big)\|\mathbf{v}\|^2_{DC}.$$

Denoting $\alpha_4 = K_3\Big(4\sqrt{\beta C_{\text{sup}}} + C_{\text{sup}}\Big)$, then,

$$|\mathcal{G}(\mathbf{v},\mathbf{v})| \le \epsilon_2\alpha_4\|\mathbf{v}\|^2_{DC}.$$

Using (3.58) implies,

$$a(\mathbf{v},\mathbf{v}) \ge \alpha_1\|\mathbf{v}\|^2_{DC} - \epsilon_2\alpha_4\|\mathbf{v}\|^2_{DC} = (\alpha_1 - \epsilon_2\alpha_4)\|\mathbf{v}\|^2_{DC}.$$

Thus, possibly requiring ϵ_2 to be even smaller, $\epsilon_2 < \alpha_1/\alpha_4$, so that $(\alpha_1 - \epsilon_2\alpha_4) > 0$.

In the case that $\kappa < 1$, the additional restriction that $\beta < \frac{1}{1-\kappa}$ for $\mathbf{Z}$ to be USPD is necessary, which implies that $\epsilon_2\beta < 1$ is required. Therefore, for any fixed choice of β, ϵ_2 must also be taken small enough to satisfy this condition. Hence,

$$\epsilon_2 < \min\left(\frac{\alpha_1}{\alpha_4}, \frac{K_3 - K_1}{\beta K_3}, \frac{1}{\beta}\right).$$

Case 3. $\kappa = 1 - \epsilon_2 > 0$, for $\epsilon_2 > 0$, and $K_3 \leq K_1$.

Here, again, $\eta = (1 - \epsilon_2\beta)$. For this case, it is clear that $(1 - \epsilon_2\beta)K_3 < K_1$. Thus,

$$\alpha_1 = \frac{(1 - \epsilon_2\beta)K_3}{(C + 1)}.$$

Using the same α_4 as in the previous case and similar arguments,

$$a(\mathbf{v}, \mathbf{v}) \geq \alpha_1\|\mathbf{v}\|_{DC}^2 - \epsilon_2\alpha_4\|\mathbf{v}\|_{DC}^2 = (\alpha_1 - \epsilon_2\alpha_4)\|\mathbf{v}\|_{DC}^2.$$

Hence, in order for $(\alpha_1 - \epsilon_2\alpha_4) > 0$ to hold, it is necessary that

$$\epsilon_2 < \frac{K_3}{K_3\beta + \alpha_4(C + 1)}.$$

Finally, ϵ_2 must still be chosen sufficiently small with respect to β such that $\epsilon_2\beta < 1$, as in Case 2. Therefore,

$$\epsilon_2 < \min\left(\frac{K_3}{K_3\beta + \alpha_4(C + 1)}, \frac{1}{\beta}\right).$$

Thus, if ϵ_1, $\epsilon_2 > 0$ satisfy the applicable conditions in the cases above, then at each Newton iteration, $a(\mathbf{u}, \mathbf{v})$ is coercive for $\kappa \in (1 - \epsilon_2, 1 + \epsilon_1)$. □

Remark 3.5.10 The size of the interval about $\kappa = 1$ depends, in part, on C_{sup}. At each iteration, C_{sup} is a well-defined constant for the variational problem given in (3.31)-(3.32). However, no uniformity is guaranteed without stronger assumptions;

notably, if the iterates, $\mathbf{n}_k$, approach a function with singular curl, then C_{sup} may grow either as the iteration proceeds or with grid refinement. To achieve uniformity, an assumption that the true solution, $\mathbf{n}_*$, is suitably smooth is needed. By the Sobolev Imbedding Theorem [1], if Ω is a bounded, Lipschitz domain in $\mathbb{R}^2$ or $\mathbb{R}^3$ and $\mathbf{n}_* \in H^3(\Omega)^3$, then $\nabla \times \mathbf{n}_*$ is continuous and bounded. However, further assumptions would be necessary to ensure that the Newton iterations remain in a neighborhood of $\mathbf{n}_*$ where their curls could be uniformly bounded. $\diamond$

3.5.4 Discrete Weak Coercivity

For this section, we consider the weak coercivity of $b(\cdot,\cdot)$, under Assumption 3.5.1, with the restriction that Ω is a polyhedral domain. That is, we show that there exists a $\zeta > 0$ such that

$$\zeta \|\gamma\|_0 \leq \sup_{\mathbf{v} \in V_h} \frac{|b(\mathbf{v},\gamma)|}{\|\mathbf{v}\|_{DC}}, \qquad \forall \gamma \in \Pi_h. \tag{3.62}$$

Before proving the weak coercivity result for V_h and Π_h, we prove two critical lemmas. Let $N = 2, 3$ denote the dimension of Ω.

Lemma 3.5.11 *For the bubble functions, b_T, satisfying* (3.35) *and* (3.36) *on a rectangle T, $\sup_{\mathbf{x} \in T} b_T = C_d/|T|$, where $C_d = (\frac{3}{2})^N$.*

Proof: For $N = 2$, without loss of generality, assume that T is a rectangle at the origin given by $[0,a] \times [0,b]$. Let $\bar{b}_T = xy(a-x)(b-y)$ on T and zero elsewhere. Note that $\bar{b}_T$ is the bubble function on T that has not been normalized such that (3.35) holds. Integrating over T yields

$$\int_T \bar{b}_T \, dV = \frac{|T|^3}{36}. \tag{3.63}$$

Computing the maximum value of $\bar{b}_T$ shows that $\sup_{\mathbf{x} \in T} \bar{b}_T = \dfrac{|T|^2}{16}$. Normalizing $\bar{b}_T$, using (3.63), to define b_T implies that

$$\sup_{\mathbf{x} \in T} b_T = \frac{|T|^2/16}{|T|^3/36} = \frac{9}{4|T|}.$$

The case for $N = 3$ is derived analogously for T, the rectangular box $[0,a] \times [0,b] \times [0,c]$, and $\bar{b}_T = xyz(a-x)(b-y)(c-z)$. The corresponding b_T satisfies

$$\sup_{\mathbf{x} \in T} b_T = \frac{|T|^2/64}{|T|^3/216} = \frac{27}{8|T|}. \qquad \square$$

Following the notation in [16], consider two finite elements $(T, \mathcal{P}, \mathcal{N})$ and $(\hat{T}, \hat{\mathcal{P}}, \hat{\mathcal{N}})$, where T and $\hat{T}$ are element domains, $\mathcal{P}$ and $\hat{\mathcal{P}}$ are the respective sets of basis functions, and $\mathcal{N}$ and $\hat{\mathcal{N}}$ are the associated dual bases. We say that $(\hat{T}, \hat{\mathcal{P}}, \hat{\mathcal{N}})$ is affine equivalent to $(T, \mathcal{P}, \mathcal{N})$ if there exists an affine mapping, $G : T \to \hat{T}$, such that for $\mathbf{x} \in T$

$$G\mathbf{x} = \mathbf{x}_0 + M\mathbf{x},$$

with non-singular matrix M, satisfying

- $G(T) = \hat{T}$,
- $G^* \hat{\mathcal{P}} = \mathcal{P}$, and
- $G_* \mathcal{N} = \hat{\mathcal{N}}$.

Here, the pullback G^* is defined by $G^*(\hat{f}) := \hat{f} \circ G$, and the push-forward G_* is defined by $(G_* N)(\hat{f}) := N(G^*(\hat{f}))$.

Lemma 3.5.12 *Consider a rectangular reference element $(T, \mathcal{P}, \mathcal{N})$, where $\mathcal{P}$ is the basis of shape functions for T associated with $V_h \times \Pi_h$, defined above. If, for all $\hat{T} \in \mathcal{T}_h$, $(\hat{T}, \hat{\mathcal{P}}, \hat{\mathcal{N}})$ is affine equivalent to $(T, \mathcal{P}, \mathcal{N})$, then $\sup_{\hat{\mathbf{x}} \in \hat{T}} b_{\hat{T}} = C_d/|\hat{T}|$, where $b_{\hat{T}}$ is the normalized bubble function satisfying* (3.35) *and* (3.36) *on $\hat{T}$.*

Proof: Note that the non-normalized bubble function on $\hat{T}$, $\bar{b}_{\hat{T}}$, is given by

$$\bar{b}_{\hat{T}} = b_T \circ G^{-1},$$

where b_T is the normalized bubble function on T. Therefore, the maximum value for $\bar{b}_{\hat{T}}$ corresponds to the maximum value for b_T, which, as shown in Lemma 3.5.11,

is $C_d/|T|$. Observe that

$$\begin{aligned}\int_{\hat{T}} \bar{b}_{\hat{T}}\, dV &= \int_T b_T |\det M|\, dV \\ &= |\det M|,\end{aligned}$$

where $\det M$ denotes the determinant of the matrix M. Thus, $b_{\hat{T}}$ is given by dividing $\bar{b}_{\hat{T}}$ by $|\det M|$. Therefore,

$$\begin{aligned}\sup_{\hat{\mathbf{x}}\in\hat{T}} b_{\hat{T}} &= \frac{1}{|\det M|} \sup_{\mathbf{x}\in T} b_T \\ &= \frac{C_d}{|\det M||T|} \\ &= \frac{C_d}{|\hat{T}|}.\end{aligned}$$

□

In the following, we make use of the second set of assumptions below when necessary.

Assumption 3.5.13 *Let $\{\mathcal{T}_h\}$ be a family of quadrilateral subdivisions of a polyhedral domain Ω satisfying* (3.33) *and* (3.34)*. Moreover, assume that for each $T \in \mathcal{T}_h$, the element $(T, \mathcal{P}_T, \mathcal{N}_T)$ is affine equivalent to a rectangular reference element for all h.*

Prior to considering the following lemma, recall that α and β are the bounds on the length of $\mathbf{n}$ in (3.9), ρ is the quasi-uniform mesh parameter defined in (3.34), and C_d is the constant derived in Lemma 3.5.11 depending on N, the dimension of Ω.

Lemma 3.5.14 *Under Assumptions 3.5.1 and 3.5.13, V_h and Π_h constitute a pair satisfying* (3.62) *with constant $\zeta = h\left[\frac{2\alpha\rho^N}{9C_f C_* \sqrt{\beta C_d}}\right]$, for C_f and C_* defined below.*

Proof: Since $V_h \subset Q_l \times Q_l \times Q_l$, by [16, Theorem 4.5.11] there exists $C_* > 0$ depending only on ρ such that

$$\|\mathbf{v}\|_1 \leq C_* h^{-1} \|\mathbf{v}\|_0.$$

Furthermore, using the fact that $\|\mathbf{v}\|_{DC} \leq C_f \|\mathbf{v}\|_1$,

$$\sup_{\mathbf{v}\in V_h} \frac{|b(\mathbf{v},\gamma)|}{\|\mathbf{v}\|_{DC}} \geq \sup_{\mathbf{v}\in V_h} \frac{|b(\mathbf{v},\gamma)|}{C_f\|\mathbf{v}\|_1} \geq \sup_{\mathbf{v}\in V_h} \frac{|b(\mathbf{v},\gamma)|}{C_f C_* h^{-1}\|\mathbf{v}\|_0}. \tag{3.64}$$

Therefore, (3.62) is reduced to finding $\zeta > 0$ such that

$$\zeta\|\gamma\|_0 \leq \sup_{\mathbf{v}\in V_h} \frac{|b(\mathbf{v},\gamma)|}{C_f C_* h^{-1}\|\mathbf{v}\|_0}, \qquad \forall \gamma \in \Pi_h.$$

Now consider constructing $\mathbf{v}_0$ on each $T \in \mathcal{T}_h$ by letting $a_T = \gamma|_T$, where this denotes the restriction of γ to the element T, and defining

$$\mathbf{v}_0|_T = a_T b_T \mathbf{n}_k|_T.$$

Observe that, as defined, $\mathbf{v}_0 \in V_h$. Let $C_m = \max_{T\in\mathcal{T}_h} |T|$. Then,

$$\begin{aligned} b(\mathbf{v}_0,\gamma) = \sum_{T\in\mathcal{T}_h} \int_T \gamma(\mathbf{v}_0,\mathbf{n}_k) &\geq \alpha \sum_{T\in\mathcal{T}_h} \gamma^2 \int_T b_T \, dV \\ &= \alpha \sum_{T\in\mathcal{T}_h} \gamma^2 \geq \frac{\alpha}{C_m}\|\gamma\|_0^2. \end{aligned} \tag{3.65}$$

It is also the case that

$$\|\mathbf{v}_0\|_0^2 = \sum_{T\in\mathcal{T}_h} \int_T a_T^2 b_T^2 (\mathbf{n}_k,\mathbf{n}_k)\, dV \leq \beta \sum_{T\in\mathcal{T}_h} \gamma^2 \int_T b_T^2 \, dV.$$

Since the bubble functions are fixed, let

$$C_b = \max_{T\in\mathcal{T}_h} \int_T b_T^2 \, dV, \qquad C_T = \min_{T\in\mathcal{T}_h} |T|.$$

Thus,

$$\|\mathbf{v}_0\|_0^2 \leq \beta C_b \sum_{T\in\mathcal{T}_h} \gamma^2 \leq \frac{\beta C_b}{C_T}\|\gamma\|_0^2. \tag{3.66}$$

Therefore, combining (3.65) and (3.66),

$$\begin{aligned}\sup_{\mathbf{v}\in V_h}\frac{\int_\Omega \gamma(\mathbf{v},\mathbf{n}_k)\,dV}{\|\mathbf{v}\|_0} &\geq \frac{\int_\Omega \gamma(\mathbf{v}_0,\mathbf{n}_k)\,dV}{\|\mathbf{v}_0\|_0}\\ &\geq \frac{\frac{\alpha}{C_m}\|\gamma\|_0^2}{\sqrt{\frac{\beta C_b}{C_T}}\|\gamma\|_0} = \frac{\alpha\sqrt{C_T}}{C_m\sqrt{\beta C_b}}\|\gamma\|_0. \end{aligned} \tag{3.67}$$

Note that the final constant in (3.67) is mesh dependent. Let $N = 2, 3$ denote the dimension of Ω. Observe that

$$C_b \leq \max_{T\in\mathcal{T}_h}\sup_{\mathbf{x}\in T} b_T \int_T b_T\,dV = \max_{T\in\mathcal{T}_h}\sup_{\mathbf{x}\in T} b_T.$$

From Lemma 3.5.12, for arbitrary $T \in \mathcal{T}_h$,

$$\sup_{\mathbf{x}\in T} b_T = C_d/|T|,$$

where C_d depends only on the dimension of Ω. Therefore,

$$\max_{T\in\mathcal{T}_h}\sup_{\mathbf{x}\in T} b_T = \frac{C_d}{C_T}.$$

Hence,

$$\frac{\sqrt{C_T}}{C_m\sqrt{C_b}} \geq \frac{C_T}{C_m\sqrt{C_d}}. \tag{3.68}$$

Define the constants

$$\begin{aligned} C_{2,1} &= \frac{\pi}{4}, & C_{2,2} &= \pi, & &\text{for } N = 2,\\ C_{3,1} &= \frac{\pi}{6}, & C_{3,2} &= \frac{3\pi}{4}, & &\text{for } N = 3. \end{aligned}$$

Using Properties (3.33) and (3.34) with the constants above, it is straightforward to show that

$$\begin{aligned} C_T &\geq C_{N,1}(\min\{\text{diam } B_T : T \in \mathcal{T}_h\})^N \geq C_{N,1}\rho^N(h\text{diam }\Omega)^N,\\ C_m &\leq C_{N,2}(\max\{\text{diam } T : T \in \mathcal{T}_h\})^N \leq C_{N,2}(h\text{diam }\Omega)^N. \end{aligned}$$

Therefore,

$$\frac{C_T}{C_m} \geq \frac{C_{N,1}\rho^N}{C_{N,2}}. \tag{3.69}$$

Utilizing (3.68) and (3.69)

$$\frac{\alpha\sqrt{C_T}}{C_m\sqrt{\beta C_b}}\|\gamma\|_0 \geq \frac{\alpha C_{N,1}\rho^N}{C_{N,2}\sqrt{\beta C_d}}\|\gamma\|_0 \geq \frac{2\alpha\rho^N}{9\sqrt{\beta C_d}}\|\gamma\|_0,$$

where C_d depends only on the dimension of Ω. Hence, (3.62) is satisfied with constant $\zeta = h\left[\frac{2\alpha\rho^N}{9C_f C_*\sqrt{\beta C_d}}\right]$. Thus, V_h and Π_h represent a pair of spaces on which $b(\cdot,\cdot)$ is weakly coercive. □

For $\mathbf{n}_k \in Q_p$, with $V_h \subset Q_m \times Q_m \times Q_m \oplus V_h^b$, as in (3.38), and $l = \max(m, p+2)$, the above lemma yields an immediate corollary.

Corollary 3.5.15 *Under Assumptions 3.5.1 and 3.5.13,* $\mathbf{n}_k \in Q_p$ *implies that* $b(\cdot,\cdot)$ *is weakly coercive for the pair* Q_l–P_0. *The special case that* $\mathbf{n}_k \in P_0$ *implies that* $b(\cdot,\cdot)$ *is weakly coercive on the pair* $Q_{\max(m,2)}$–P_0.

Proof: Note that if $\mathbf{n}_k \in Q_p$, the bubble space defined above satisfies $V_h^b \subset Q_{p+2} \times Q_{p+2} \times Q_{p+2}$, since $b_T \in Q_2$. This implies that $V_h \subset Q_l \times Q_l \times Q_l$. Therefore, since $b(\cdot,\cdot)$ is weakly coercive for the pair V_h–P_0, weak coercivity must also hold for the pair Q_l–P_0. If $\mathbf{n}_k \in P_0$, then $V_h^b \subset Q_2 \times Q_2 \times Q_2$. Hence, $V_h \subset Q_{\max(m,2)} \times Q_{\max(m,2)} \times Q_{\max(m,2)}$. The lemma above is equally valid for $\mathbf{n}_k \in P_0$. Therefore, $b(\cdot,\cdot)$ is weakly coercive on the pair $Q_{\max(m,2)}$–P_0 for the given $\mathbf{n}_k$. □

In light of the lemmas discussed above, verification of weak coercivity allows for the formulation and proof of the following theorem.

Theorem 3.5.16 *Under Assumptions 3.5.1 and 3.5.13, existence of discrete solutions* $(\delta\mathbf{n}_h, \delta\lambda_h)$ *for each Newton linearization are guaranteed for the pair* V_h–Π_h. *In the case that* $\kappa = 1$ *or that* κ *satisfies the small data conditions of Lemma 3.5.9, such solutions are unique.*

Proof: Following a mixed formulation approach based on [12, 14, 16], Lemmas 3.5.6 and 3.5.7 guarantee the existence of a solution to the system given by (3.31) and (3.32). In the event that $\kappa = 1$ or that κ satisfies the small data assumptions, Lemma 3.5.8 or 3.5.9 coupled with Lemma 3.5.14 implies that the solution is also unique.□

3.6 Error Analysis

In the previous section, the derived weak coercivity constant depends on the mesh parameter h. Therefore, as h approaches zero so too does the weak coercivity constant for the pair V_h and Π_h. However, the convergence of the scheme for the enriched Lagrangian finite-element spaces composing V_h is only slightly compromised. In this section, we derive approximation error bounds for the discrete update solution. Throughout this section, it is assumed that Assumptions 3.5.1 and 3.5.13 apply. Let $(\mathbf{u}, q)$ represent a solution to the continuum variational system given by (3.10) and (3.11) and $(\mathbf{u}_h, q_h)$ be the unique solution to the discrete system in (3.31) and (3.32). As above, denote the dimension of Ω by $N = 2, 3$.

Theorem 3.6.1 *Let Π_h and V_h be defined as in* (3.37) *and* (3.38) *with $m = 2$. Under Assumptions 3.5.1 and 3.5.13, for $\mathbf{u} \in H^3(\Omega)^3$ and $q \in H^1(\Omega)$ there exists $C_a > 0$ such that*

$$\|\mathbf{u} - \mathbf{u}_h\|_{DC} \leq C_a h\big(\|\mathbf{u}\|_3 + \|q\|_1\big). \tag{3.70}$$

Proof: Let α_0 denote the coercivity constant from either Lemma 3.5.8 or 3.5.9. Furthermore, let ζ denote the h-dependent weak coercivity constant derived in Lemma 3.5.14. By Theorem 5.2.2 in [12],

$$\|\mathbf{u} - \mathbf{u}_h\|_{DC} \leq \frac{4C_A C_B}{\alpha_0 \zeta} E_u + \frac{C_B}{\alpha_0} E_q, \tag{3.71}$$

where C_A and C_B are the continuity constants associated with $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$, respectively, and

$$E_u = \inf_{v_h \in V_h} \|\mathbf{u} - \mathbf{v}_h\|_{DC}, \qquad E_q = \inf_{\gamma_h \in \Pi_h} \|q - \gamma_h\|_0.$$

Note that

$$\inf_{v_h \in V_h} \|\mathbf{u} - \mathbf{v}_h\|_{DC} \leq C_f \inf_{v_h \in V_h} \|\mathbf{u} - \mathbf{v}_h\|_1,$$

where C_f is the constant used in (3.64). Let $\mathcal{I}^h f$ denote the global interpolant of f over the appropriate finite-element space. Since $\{\mathcal{T}_h\}$ is quasi-uniform, it is, in particular, non-degenerate. Therefore, applying [16, Theorem 4.4.24] to the discrete space V_h, there exists a $C_5 > 0$, such that

$$\left(\sum_{T \in \mathcal{T}_h} \|\mathbf{v} - \mathcal{I}^h \mathbf{v}\|^2_{H^1(T)} \right)^{1/2} = \|\mathbf{v} - \mathcal{I}^h \mathbf{v}\|_1 \leq C_5 h^2 \|\mathbf{v}\|_3, \qquad \forall \mathbf{v} \in H^3(\Omega).$$

This implies that if $\mathbf{u} \in H^3(\Omega)^3$, then

$$\inf_{\mathbf{v}_h \in V_h} \|\mathbf{u} - \mathbf{v}_h\|_{DC} \leq C_f C_5 h^2 \|\mathbf{u}\|_3. \tag{3.72}$$

For Π_h, Theorem 3.1.6 in [23] implies that there exists a $C_6 > 0$ such that

$$\|\gamma - \mathcal{I}^h \gamma\|_0 \leq C_6 h \|\gamma\|_1, \qquad \forall \gamma \in H^1(\Omega).$$

Hence, if $q \in H^1(\Omega)$,

$$\inf_{\gamma_h \in \Pi_h} \|q - \gamma_h\|_0 \leq C_6 h \|q\|_1. \tag{3.73}$$

Combining (3.72) and (3.73) with (3.71) produces the error estimate

$$\begin{aligned} \|\mathbf{u} - \mathbf{u}_h\|_{DC} &\leq \frac{4 C_A C_B}{\alpha_0 \zeta} C_f C_5 h^2 \|\mathbf{u}\|_3 + \frac{C_B}{\alpha_0} C_6 h \|q\|_1 \\ &= \frac{18 C_A C_B C_f^2 C_* \sqrt{\beta C_d} C_5}{\alpha \rho^N \alpha_0} h \|\mathbf{u}\|_3 + \frac{C_B C_6}{\alpha_0} h \|q\|_1. \end{aligned}$$

Taking $C_a = \max\left(\frac{18 C_A C_B C_f^2 C_* \sqrt{\beta C_d} C_5}{\alpha \rho^N \alpha_0}, \frac{C_B C_6}{\alpha_0} \right)$, (3.70) is obtained. □

Thus, the approximation is convergent for V_h–Π_h but with an order of sub-optimality, due to the weak coercivity constant's dependence on the mesh parameter. Use of a discrete $H^{-1}(\Omega)$ norm for the space Π_h is being considered as a means of eliminating this mesh dependence. However, preliminary numerical results have

suggested that such an approach may not be viable.

Remark 3.6.2 The constant C_A is dependent on C_{sup} and, therefore, complexities across iterations similar to those discussed in Remark 3.5.10 may arise. The error analysis presented above deals with a fixed $\mathbf{n}_k$ across grids and a variational problem with true Newton correction $\mathbf{u}$. It demonstrates that, for the fixed variational problem, the solution on successively finer grids converges to $\mathbf{u}$ with order h. Assumptions similar to those in Remark 3.5.10 would be needed to ensure uniformity across iterations or grid refinements. ♢

3.7 Numerical Results

In this section, we present numerical results for the energy-minimization finite-element method discussed above. The algorithm to perform the minimization discussed in previous sections has three stages; see Algorithm 1. The outermost phase is nested iteration (NI) [94, 116], which begins on a specified coarsest grid level. Newton iterations are performed on each grid, updating the current approximation after each step. The stopping criterion for the Newton iterations at each level is based on a specified tolerance for the current approximation's conformance to the first-order optimality conditions in the standard Euclidean l_2-norm. For the numerical experiments in this section, this tolerance is held fixed at 10^{-4}. The resulting approximation is then interpolated to a finer grid. The current implementation performs uniform grid refinement after each set of Newton iterations.

The Newton iteration systems are constructed by applying finite-element discretizations on each grid. The resulting, relatively sparse, matrix has the anticipated saddle-point block structure,

$$\begin{bmatrix} A & B \\ B^T & \mathbf{0} \end{bmatrix}.$$

The matrix is inverted using LU decomposition in order to solve for the discrete updates $\delta\mathbf{n}_h$ and $\delta\lambda_h$. Finally, a damped Newton correction is performed. That is,

the new iterates are given by

$$\begin{bmatrix} \mathbf{n}_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{n}_k \\ \lambda_k \end{bmatrix} + \omega \begin{bmatrix} \delta\mathbf{n}_h \\ \delta\lambda_h \end{bmatrix}, \tag{3.74}$$

where $\omega \leq 1$. This is to ensure relatively strict adherence to the constraint manifold, which is necessary for the well-posedness discussed above. For the algorithm applied in this section, ω is chosen to begin at 0.2 on the coarsest grid and increases by 0.2, to a maximum of 1, after each grid refinement, so that as the approximation converges, larger Newton steps are taken. For complicated boundary conditions, such damped Newton steps are important in preventing method divergence. More sophisticated methods for choosing the Newton step size are discussed in the next chapter. The grid management and discretizations are implemented using the deal.II finite-element library, which is an aggressively optimized and parallelized open-source library widely used in scientific computing [8, 9].

Algorithm 1: Newton's method minimization algorithm with NI

0. Initialize $(\mathbf{n}_0, \lambda_0)$ on coarse grid.

while *Refinement limit not reached* **do**

  while *First-order optimality conformance threshold not satisfied* **do**

    1. Set up discrete linear system (3.5) on current grid, H.
    2. Solve for $\delta\mathbf{n}_H$ and $\delta\lambda_H$.
    3. Compute $\mathbf{n}_{k+1}$ and λ_{k+1} as in (3.74).

  end

  4. Uniformly refine the grid.
  5. Interpolate $\mathbf{n}_H \to \mathbf{n}_h$ and $\lambda_H \to \lambda_h$.

end

3.7.1 Practical Choice of Finite Elements

The bubble enrichment discussed above is non-standard in its incorporation of $\mathbf{n}_k$ in the construction of the bubbles. Therefore, during numerical implementation, it

was desirable to find an experimentally stable, standard, finite-element pair closely related to the spaces discussed above. It was observed that Q_1–Q_1 finite-element discretizations resulted in singular matrices. This implies that Q_1–Q_1 is not a pair for which $b(\cdot,\cdot)$ is weakly coercive. Such a phenomenon is not unique. For example, instabilities arise for equal order elements in Galerkin approaches to both the Stokes' equations [41] and the Navier-Stokes' equations [49].

On the other hand, in the numerical experiments to be discussed below, mixed finite-element approaches, such as Q_2–P_0 discretizations, experimentally appear to admit weak coercivity without the need for rising order finite-element implementations or bubble enrichments. In addition, Corollary 3.5.15 implies that for a piecewise constant initial iterate, the update element space Q_2–P_0 implies weak coercivity for the first Newton iteration. With this assurance, coupled with the empirical weak coercivity evidence for Q_2–P_0, we employ Q_2–P_0 spaces to approximate $\delta\mathbf{n}_h$ and $\delta\lambda_h$ for all iterations in the experiments of this section. In the event that instabilities occur for the Q_2–P_0 discretization of a particular problem, the bubble enriched finite-element pair V_h–Π_h, defined in (3.37) and (3.38), may be implemented and is particularly attractive because the rising order of the bubble functions, $b_T\mathbf{n}_k|_T$, on each element does not increase the number of unknowns at each Newton iteration.

3.7.2 Free Elastic Numerical Results

The general test problem in this section considers a classical domain with two parallel substrates placed at distance $d = 1$ apart. The substrates run parallel to the xz-plane and perpendicular to the y-axis. It is assumed that this domain represents a uniform slab in the xy-plane. That is, $\mathbf{n}$ may have a nonzero z component but $\frac{\partial \mathbf{n}}{\partial z} = \mathbf{0}$. Hence, we consider the 2-D domain $\Omega = \{(x, y) \mid 0 \le x, y \le 1\}$. The problem assumes periodic boundary conditions at the edges $x = 0$ and $x = 1$. Dirichlet boundary conditions are enforced on the y-boundaries. As discussed above, the simplification outlined in (2.4) is relevant for this domain and boundary conditions. In this chapter, we introduce work units (WUs) to quantify the efficiency of NI. For our numerical results, WUs are measured in terms of assembling and solving a single

linearization on the finest grid of a NI hierarchy and are computed by summing the number of non-zeroes in each matrix across all grids and dividing by the number of non-zeroes in the (fixed) sparsity pattern at the finest level. Assuming the presence of solvers that scale linearly with the number of non-zeroes in the matrix, WUs offer a metric for cost comparisons against runs without NI.

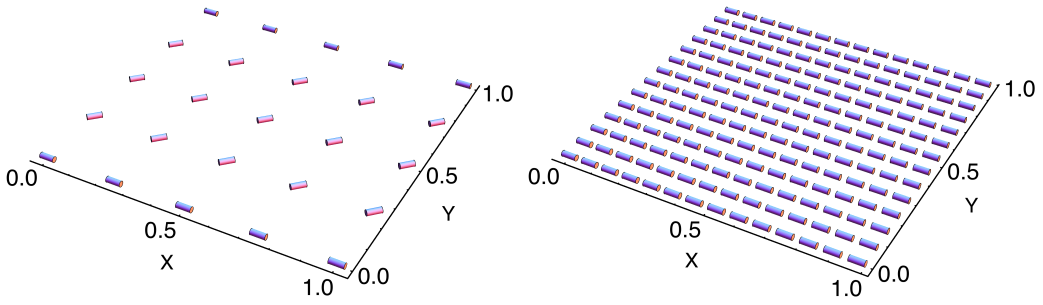


Figure 3.1: (a) Initial guess on 4×4 mesh with initial free energy of 5.467 and (b) resolved solution on 512×512 mesh (restricted for visualization) with final free energy of 0 for a uniformly aligned boundary.

The first numerical experiment is run on one of the simplest configurations of this type. Along each of the substrates the liquid crystal rods are uniformly aligned parallel to the x-axis. The relevant Frank constants are $K_1 = K_2 = K_3 = 1$. It should be noted that for the parameters defined throughout this thesis, the unit of length is taken to be microns. The problem is solved on a 4×4 coarse grid with seven successive uniform refinements resulting in a 512×512 fine grid. The initial guess and computed, converged solution are displayed in Figure 3.1.

The final minimized functional energy is $\mathcal{F}_1 = 0$, compared to the initial guess energy of 5.467. In Table 3.1, the number of Newton iterations per grid is detailed as well as the conformance of the solution to the first-order optimality conditions after the first and final Newton steps, respectively, on each grid. The total work required in these iterations is approximately 1.33 WUs. In contrast, without nested iteration, the algorithm requires 27 damped Newton steps on the 512×512 finest grid alone, to satisfy the tolerance limit. The application of damped Newton steps becomes even more important when beginning on finer grids with a rough initial guess, as divergence can be more prevalent. Table 3.1 also reveals the performance

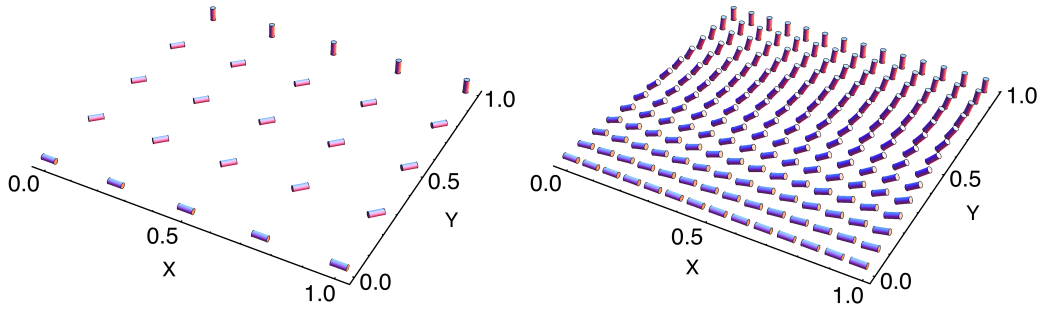
of the algorithm with respect to the pointwise constraint, presenting the progressively tighter minimum and maximum director deviations from unit length at the quadrature nodes. The computed equilibrium solution behaves as expected with the rods uniformly aligning parallel to the x-axis.

Grid Dim.	Newton Iter.	Init. Res.	Final Res.	Deviation in $\|\mathbf{n}\|^2$	Final Energy
4×4	22	4.34e-00	5.69e-05	8.00e-07, 7.19e-06	8.298e-10
8×8	1	3.16e-05	1.26e-05	1.62e-07, 2.92e-06	1.328e-10
16×16	1	7.10e-06	1.42e-06	1.63e-08, 5.89e-07	5.311e-12
32×32	1	8.32e-07	2.28e-13	0, 7.23e-13	2.239e-24
64×64	1	1.31e-13	5.57e-14	-4.00e-16, 1.00e-15	0
128×128	1	1.22e-13	1.01e-13	-4.00e-16, 0	0
256×256	1	2.18e-13	1.91e-13	-4.00e-16, 0	0
512×512	1	4.46e-13	3.97e-13	-4.00e-16, 0	0

Table 3.1: Grid and solution progression for uniform free-elastic boundary conditions with initial and final residuals for the first-order optimality conditions, minimum and maximum director deviations from unit length at the quadrature nodes, and final functional energy on each grid.

The second test, run for the free-elastic slab problem, incorporates twist boundary conditions and unequal Frank constants. On the lower slab, along $y = 0$, the nematic rods are aligned parallel to the x-axis. For the upper slab, the rods are uniformly aligned parallel to the z-axis. The relevant constants for this run are $K_1 = 1$, $K_2 = 1.2$, and $K_3 = 1$. This implies that $\kappa = K_2/K_3 > 1$. The solves are again performed on a 4×4 coarse grid, uniformly ascending to a 512×512 fine grid. The expected configuration for such boundary conditions is a twisted equilibrium solution along the y-axis. Indeed, the numerically resolved solution in Figure 3.2, displayed alongside the initial guess, demonstrates such a twist. The final minimized functional energy is $\mathcal{F}_1 = 1.480$, compared to the initial guess energy of 12.534. Table 3.2 enumerates the algorithm run attributes.

As in Table 3.1 above, a sizable majority of the Newton iteration computations are isolated to the coarsest grids, with the finest grids requiring only one Newton iteration to reach the residual tolerance limit. Therefore, most of the computational cost is also isolated to the cheaper coarse grids rather than the finer levels. Here, the total work required is approximately 1.36 WUs. Without nested iteration, 28 damped Newton steps are required on the finest grid to compute the solution.



(a) (b)

Figure 3.2: (a) Initial guess on 4×4 mesh with initial free energy of 12.534 and (b) resolved solution on 512×512 mesh (restricted for visualization) with final free energy of 1.480 for a twist boundary.

Grid Dim.	Newton Iter.	Init. Res.	Final Res.	Deviation in $\lvert\mathbf{n}\rvert^2$	Final Energy
4×4	23	6.71e-00	5.15e-05	-5.98e-05, 4.40e-05	1.481
8×8	7	1.80e-02	2.95e-05	-3.82e-06, 1.79e-06	1.480
16×16	4	4.51e-03	7.22e-06	-2.39e-07, 1.10e-07	1.480
32×32	2	1.13e-03	2.16e-14	-1.47e-08, 6.88e-09	1.480
64×64	2	2.82e-04	4.10e-14	-9.21e-10, 4.30e-10	1.480
128×128	1	7.05e-05	1.35e-12	-5.75e-11, 2.69e-11	1.480
256×256	1	1.76e-05	1.63e-13	-3.60e-12, 1.68e-12	1.480
512×512	1	4.41e-06	3.09e-13	-2.25e-13, 1.05e-13	1.480

Table 3.2: Grid and solution progression for the free-elastic problem and twist boundary conditions with initial and final residuals for the first-order optimality conditions, minimum and maximum director deviations from unit length at the quadrature nodes, and final functional energy on each grid.

In the final numerical run, letting $r = 0.25$ and $s = 0.95$, the boundary conditions considered are

$$n_1 = 0, \tag{3.75}$$

$$n_2 = \cos\left(r(\pi + 2\tan^{-1}(X_m) - 2\tan^{-1}(X_p))\right), \tag{3.76}$$

$$n_3 = \sin\left(r(\pi + 2\tan^{-1}(X_m) - 2\tan^{-1}(X_p))\right), \tag{3.77}$$

where $X_m = \frac{-s\sin(2\pi(x+r))}{-s\cos(2\pi(x+r))-1}$ and $X_p = \frac{-s\sin(2\pi(x+r))}{-s\cos(2\pi(x+r))+1}$. Such boundary conditions are meant to simulate nano-patterned surfaces important in current research [5, 6]. Even in the absence of electric fields, such patterned surfaces result in complicated director configurations throughout the interior of Ω.

A similar grid progression to the cases above is applied. The Frank elastic constants for the experiment are $K_1 = 1$, $K_2 = 0.62903$, and $K_3 = 1.32258$. This results in $\kappa < 1$. The final solution, as well as the initial guess, are displayed in Figure 3.3. Table 3.3, again, details the relevant computation data. The computed equilibrium configuration demonstrates the expected alignment and symmetries given the patterned surfaces.

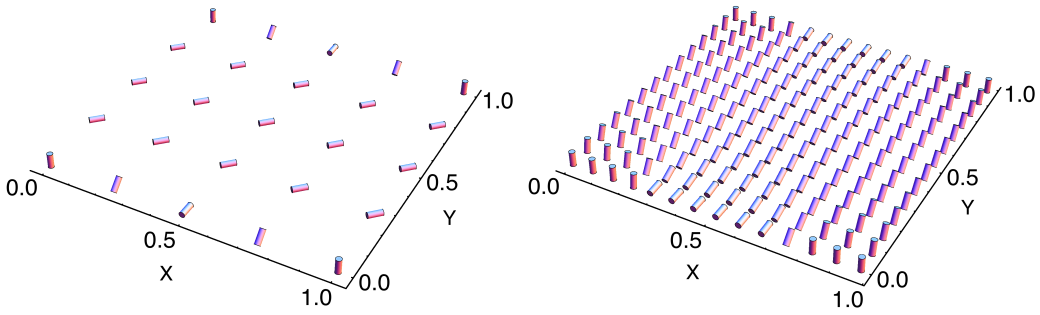


Figure 3.3: (a) Initial guess on 4×4 mesh with initial free energy of 13.242 and (b) resolved solution on 512×512 mesh (restricted for visualization) with final free energy of 3.890 for a nano-patterned boundary.

The minimized functional energy is $\mathcal{F}_1 = 3.890$, compared to the initial guess free energy of 13.242. The work required is approximately 2.75 WUs. On the other hand, without nested iterations, 28 damped Newton steps are required on the finest grid. Therefore, in all cases discussed, nested iteration is successful in significantly reducing the computational work necessary to compute an equilibrium solution.

Grid Dim.	Newton Iter.	Init. Res.	Final Res.	Deviation in $\lvert\mathbf{n}\rvert^2$	Final Energy
4×4	24	7.04e-00	3.67e-05	-9.09e-02, 4.67e-02	2.521
8×8	12	1.20e-00	2.01e-05	-8.20e-02, 4.58e-02	3.194
16×16	7	1.06e-00	1.34e-05	-6.69e-02, 3.96e-02	3.674
32×32	3	8.22e-01	3.41e-12	-4.31e-02, 2.78e-02	3.885
64×64	3	5.04e-01	4.56e-14	-1.73e-02, 1.26e-02	3.900
128×128	3	2.24e-01	9.12e-14	-3.51e-03, 2.81e-03	3.890
256×256	2	6.94e-02	1.49e-10	-4.63e-04, 3.63e-04	3.890
512×512	2	1.78e-02	7.39e-13	-6.92e-05, 5.89e-05	3.890

Table 3.3: Grid and solution progression for patterned boundary conditions with initial and final residuals for the first-order optimality conditions, minimum and maximum director deviations from unit length at the quadrature nodes, and final functional energy on each grid.

Chapter 4

A Penalty Method and Trust Regions

In the previous chapter, a theoretically supported energy-minimization finite-element approach using Newton linearization and a Lagrange multiplier for the pointwise constraint was developed. The approach effectively enforces the unit-length constraint while converging to energy-minimizing configurations. However, alternative approaches to efficiently impose unit-length conformance exist, such as the renormalized Newton method presented in [53]. Penalty methods have also been applied to liquid crystal equilibrium problems [5, 57, 64] and are utilized extensively to simplify the Ericksen-Leslie equations [44, 78] in nematohydrodynamics simulations [83, 85, 86]. In addition, penalty methods are used for unit-length constraints in certain ferromagnetic problems [73].

In this chapter, we aim to compare the performance of techniques enforcing the unit-length constraint via Lagrange multipliers or with penalty methods employing augmentations to the free-elastic energy functional. Due to their broad use, the accuracy and efficiency of penalty methods, relative to the Lagrange multiplier approach outlined above, are of great interest in the context of our energy-minimization algorithm. The constraint enforcement approaches are discussed and well-posedness for the intermediate Newton linearizations that arise in the penalty method formulation is established. In the numerical experiments to follow, we compare the cost and precision of the constraint techniques in order to determine the most effective approach.

In addition, several tailored trust-region methods are investigated in order to consider improvements to the damped Newton stepping method discussed in Section 3.7. These trust-region approaches include one- and two-dimensional subspace minimization techniques [18, 19, 100]. In the context of optimization, trust-region methods are quite successful at increasing convergence robustness while generally

improving precision and time to solution. A modified penalty method, which normalizes the director after each step, is also introduced in this section. The resulting algorithms are tested on three benchmark free-elastic problems. Throughout this chapter, we utilize the null Lagrangian simplification discussed in Equation (2.4).

4.1 Penalty Method Energy Minimization

In order to define the penalty approach, the free-energy functional in (2.5) is augmented with a positive-definite term to form the functional

$$\mathcal{P}(\mathbf{n}) = K_1\langle \nabla\cdot\mathbf{n}, \nabla\cdot\mathbf{n}\rangle_0 + K_3\langle \mathbf{Z}\nabla\times\mathbf{n}, \nabla\times\mathbf{n}\rangle_0 + \zeta\langle \mathbf{n}\cdot\mathbf{n}-1, \mathbf{n}\cdot\mathbf{n}-1\rangle_0, \tag{4.1}$$

where $\zeta > 0$ represents a constant weight, energetically penalizing deviations of the solution from the unit-length constraint. Thus, in the limit of large ζ values, unconstrained minimization of (4.1) is equivalent to the constrained minimization of (2.5). In order to minimize $\mathcal{P}(\mathbf{n})$, we compute the Gâteaux derivative of $\mathcal{P}(\mathbf{n})$ with respect to $\mathbf{n}$ in the direction $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$. Hence, the first-order optimality condition is

$$\mathcal{P}_{\mathbf{n}}[\mathbf{v}] = \frac{\partial}{\partial \mathbf{n}}\mathcal{P}(\mathbf{n})[\mathbf{v}] = 0, \qquad \forall \mathbf{v} \in \mathcal{H}_0^{DC}(\Omega).$$

Computation of this derivative produces the variational problem

$$\begin{aligned}\mathcal{P}_{\mathbf{n}}[\mathbf{v}] = {} & 2K_1\langle \nabla\cdot\mathbf{n}, \nabla\cdot\mathbf{v}\rangle_0 + 2K_3\langle \mathbf{Z}\nabla\times\mathbf{n}, \nabla\times\mathbf{v}\rangle_0 \\ & + 2(K_2-K_3)\langle \mathbf{n}\cdot\nabla\times\mathbf{n}, \mathbf{v}\cdot\nabla\times\mathbf{n}\rangle_0 + 4\zeta\langle \mathbf{v}\cdot\mathbf{n}, \mathbf{n}\cdot\mathbf{n}-1\rangle_0 = 0,\end{aligned}$$

for all $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$.

As with the Lagrangian formulation, the variational problem above contains nonlinearities. Therefore, Newton iterations are again applied, requiring computation of the second-order Gâteaux derivative with respect to $\mathbf{n}$. Let $\mathbf{n}_k$ be the current approximation for $\mathbf{n}$ and $\delta\mathbf{n} = \mathbf{n}_{k+1} - \mathbf{n}_k$ be the update that we seek to compute.

Then, the Newton linearizations are written

$$\frac{\partial}{\partial \mathbf{n}} \left(\mathcal{P}_{\mathbf{n}}(\mathbf{n}_k)[\mathbf{v}]\right)[\delta \mathbf{n}] = -\mathcal{P}_{\mathbf{n}}(\mathbf{n}_k)[\mathbf{v}], \qquad \forall \mathbf{v} \in \mathcal{H}_0^{DC}(\Omega), \tag{4.2}$$

where

$$\begin{aligned}\frac{\partial}{\partial \mathbf{n}} \left(\mathcal{P}_{\mathbf{n}}(\mathbf{n}_k)[\mathbf{v}]\right)[\delta \mathbf{n}] = {} & 2K_1\langle \nabla \cdot \delta \mathbf{n}, \nabla \cdot \mathbf{v}\rangle_0 + 2K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla \times \delta \mathbf{n}, \nabla \times \mathbf{v}\rangle_0 \\ & + 2(K_2 - K_3)\Big(\langle \delta \mathbf{n} \cdot \nabla \times \mathbf{v}, \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k\rangle_0 \\ & + \langle \mathbf{n}_k \cdot \nabla \times \mathbf{v}, \delta \mathbf{n} \cdot \nabla \times \mathbf{n}_k\rangle_0 + \langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \delta \mathbf{n}\rangle_0 \\ & + \langle \mathbf{n}_k \cdot \nabla \times \delta \mathbf{n}, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0 + \langle \delta \mathbf{n} \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0\Big) \\ & + 4\zeta\Big(\langle \mathbf{n}_k \cdot \mathbf{n}_k - 1, \mathbf{v} \cdot \delta \mathbf{n}\rangle_0 + 2\langle \delta \mathbf{n} \cdot \mathbf{n}_k, \mathbf{v} \cdot \mathbf{n}_k\rangle_0\Big). \end{aligned} \tag{4.3}$$

Completing (4.2) with the above second-order derivative computation yields a linearized variational system. For each iteration, we compute $\delta \mathbf{n}$ satisfying (4.2) for all $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$ with the current approximation $\mathbf{n}_k$.

4.2 Well-Posedness of the Penalty Linearizations

Here, we adapt the well-posedness results established for the Lagrange multiplier method in Section 3.5 to the discrete form of the penalty method linearizations in Equation (4.2).

Let $a(\delta \mathbf{n}, \mathbf{v})$ denote the bilinear form defined in (4.3) for fixed $\mathbf{n}_k$ and $F(\mathbf{v})$ be the linear functional on the right-hand-side of the linearization in (4.2). Using finite elements to approximate the desired update, $\delta \mathbf{n}$, and considering a discrete space $V_h \subset \mathcal{H}_0^{DC}(\Omega)$ yields the discrete linearized system,

$$a(\delta \mathbf{n}_h, \mathbf{v}_h) = F(\mathbf{v}_h), \qquad \forall \mathbf{v}_h \in V_h. \tag{4.4}$$

Throughout the rest of this section, the developed theory applies exclusively to discrete spaces. Therefore, except when necessary for clarity, we drop the subscript h along with the notation, $\delta \mathbf{n}$. For instance, we write $a(\mathbf{u}, \mathbf{v})$ to indicate the bilinear

form in (4.4) operating on the discrete space $V_h \times V_h$. While the theory below, explicitly concerns full Dirichlet boundary conditions, the theory is equally applicable to mixed Dirichlet and periodic boundary conditions on a rectangular domain.

In order to establish the well-posedness of (4.4), we show that the functional, $F(\mathbf{v})$, is continuous and that the bilinear form, $a(\mathbf{u},\mathbf{v})$, is continuous and coercive. Decomposing the bilinear form, $a(\mathbf{u},\mathbf{v})$, and the linear form, $F(\mathbf{v})$, into terms that contain the penalty term and those that do not, $\hat{a}(\mathbf{u},\mathbf{v})$ and $\hat{F}(\mathbf{v})$,

$$\begin{aligned} a(\mathbf{u},\mathbf{v}) &= \hat{a}(\mathbf{u},\mathbf{v}) + 2\zeta\big(\langle \mathbf{n}_k\cdot\mathbf{n}_k - 1, \mathbf{v}\cdot\mathbf{u}\rangle_0 + 2\langle \mathbf{u}\cdot\mathbf{n}_k, \mathbf{v}\cdot\mathbf{n}_k\rangle_0\big), \\ F(\mathbf{v}) &= \hat{F}(\mathbf{v}) + 2\zeta\langle \mathbf{v}\cdot\mathbf{n}_k, \mathbf{n}_k\cdot\mathbf{n}_k - 1\rangle_0. \end{aligned}$$

We then extend the results in Section 3.5.

Lemma 4.2.1 *Under Assumption 3.5.1, F is a bounded linear functional on V_h.*

Proof: From the bounds derived in Lemma 3.5.6 and an application of the Cauchy-Schwarz inequality,

$$\begin{aligned} |F(\mathbf{v})| &\le |\hat{F}(\mathbf{v})| + 2\zeta\|\mathbf{n}_k\cdot\mathbf{n}_k - 1\|_0\|\mathbf{n}_k\cdot\mathbf{v}\|_0 \\ &\le C_F\|\mathbf{v}\|_{DC} + 2\zeta\|\mathbf{n}_k\cdot\mathbf{n}_k - 1\|_0\|\mathbf{n}_k\cdot\mathbf{v}\|_0, \end{aligned}$$

where C_F is the constant independent of mesh size, defined by Lemma 3.5.6. Note that by assumption $\alpha \le \mathbf{n}_k\cdot\mathbf{n}_k \le \beta$, where $0 < \alpha \le 1 \le \beta$. Then, letting $C_\mu = \max(1-\alpha, \beta-1)$,

$$\|\mathbf{n}_k\cdot\mathbf{n}_k - 1\|_0^2 = \int_\Omega (\mathbf{n}_k\cdot\mathbf{n}_k - 1)^2\, dV \le C_\mu^2 \int_\Omega dV = C_\mu^2|\Omega|.$$

Hence, $\|\mathbf{n}_k\cdot\mathbf{n}_k - 1\|_0 \le C_\mu|\Omega|^{\frac{1}{2}}$. In addition,

$$\|\mathbf{n}_k\cdot\mathbf{v}\|_0 \le \sqrt{\beta}\|\mathbf{v}\|_0 \le \sqrt{\beta}\|\mathbf{v}\|_{DC}.$$

Thus,

$$|F(\mathbf{v})| \leq C_F \|\mathbf{v}\|_{DC} + 2\zeta C_\mu |\Omega|^{\frac{1}{2}} \sqrt{\beta} \|\mathbf{v}\|_{DC}. \qquad \square$$

Lemma 4.2.2 *Under Assumption 3.5.1, $a(\mathbf{u}, \mathbf{v})$ is continuous.*

Proof: Applying the triangle inequality,

$$\begin{aligned} |a(\mathbf{u}, \mathbf{v})| &\leq |\hat{a}(\mathbf{u}, \mathbf{v})| + 2\zeta\big(|\langle \mathbf{n}_k \cdot \mathbf{n}_k - 1, \mathbf{u} \cdot \mathbf{v} \rangle_0| + 2|\langle \mathbf{u} \cdot \mathbf{n}_k, \mathbf{v} \cdot \mathbf{n}_k \rangle_0|\big) \\ &\leq C_A \|\mathbf{u}\|_{DC} \|\mathbf{v}\|_{DC} + 2\zeta\big(|\langle \mathbf{n}_k \cdot \mathbf{n}_k - 1, \mathbf{u} \cdot \mathbf{v} \rangle_0| + 2|\langle \mathbf{u} \cdot \mathbf{n}_k, \mathbf{v} \cdot \mathbf{n}_k \rangle_0|\big), \end{aligned}$$

where C_A is the continuity constant defined in Lemma 3.5.7. Note that,

$$|\langle \mathbf{n}_k \cdot \mathbf{n}_k - 1, \mathbf{u} \cdot \mathbf{v} \rangle_0| = |\langle (\mathbf{n}_k \cdot \mathbf{n}_k - 1)\mathbf{u}, \mathbf{v} \rangle_0| \leq \|(\mathbf{n}_k \cdot \mathbf{n}_k - 1)\mathbf{u}\|_0 \|\mathbf{v}\|_0.$$

Furthermore,

$$\|(\mathbf{n}_k \cdot \mathbf{n}_k - 1)\mathbf{u}\|_0^2 = \int_\Omega (\mathbf{n}_k \cdot \mathbf{n}_k - 1)^2 (\mathbf{u} \cdot \mathbf{u})\, dV \leq C_\mu^2 \|\mathbf{u}\|_0^2.$$

This implies that

$$\|(\mathbf{n}_k \cdot \mathbf{n}_k - 1)\mathbf{u}\|_0 \leq C_\mu \|\mathbf{u}\|_0,$$

and

$$|\langle \mathbf{n}_k \cdot \mathbf{n}_k - 1, \mathbf{u} \cdot \mathbf{v} \rangle_0| \leq C_\mu \|\mathbf{u}\|_{DC} \|\mathbf{v}\|_{DC}.$$

Noting that

$$|\langle \mathbf{u} \cdot \mathbf{n}_k, \mathbf{v} \cdot \mathbf{n}_k \rangle_0| \leq \beta \|\mathbf{u}\|_{DC} \|\mathbf{v}\|_{DC},$$

we bound

$$a(\mathbf{u}, \mathbf{v}) \leq \Big(C_A + 2\zeta\big(C_\mu + 2\beta\big)\Big) \|\mathbf{u}\|_{DC} \|\mathbf{v}\|_{DC}. \qquad \square$$

Following the theory established in Section 3.5.3, two coercivity lemmas for $a(\mathbf{u}, \mathbf{v})$ are proved. The first proof addresses the case when $\kappa = 1$. The second considers

coercivity when κ lies in a neighborhood of unity, $\kappa \in (1-\epsilon_2, 1+\epsilon_1)$. Let $\alpha_0 > 0$ be the coercivity constant from Lemma 3.5.8.

Lemma 4.2.3 *Under Assumption 3.5.1, if $\kappa = 1$ and $2\zeta|\alpha - 1| < \alpha_0$, there exists a $\beta_0 > 0$ such that $\beta_0 \|\mathbf{v}\|_{DC}^2 \le a(\mathbf{v},\mathbf{v})$ for all $\mathbf{v} \in V_h$.*

Proof:

$$a(\mathbf{v},\mathbf{v}) = \hat{a}(\mathbf{v},\mathbf{v}) + 2\zeta\langle \mathbf{n}_k\cdot\mathbf{n}_k - 1, \mathbf{v}\cdot\mathbf{v}\rangle_0 + 4\zeta\langle \mathbf{v}\cdot\mathbf{n}_k, \mathbf{v}\cdot\mathbf{n}_k\rangle_0$$

Using the coercivity of $\hat{a}(\mathbf{v},\mathbf{v})$ from Lemma 3.5.8 and the fact that $\langle \mathbf{v}\cdot\mathbf{n}_k, \mathbf{v}\cdot\mathbf{n}_k\rangle_0 \ge 0$,

$$\alpha_0\|\mathbf{v}\|_{DC}^2 \le \hat{a}(\mathbf{v},\mathbf{v}) + 4\zeta\langle \mathbf{v}\cdot\mathbf{n}_k, \mathbf{v}\cdot\mathbf{n}_k\rangle_0.$$

Observe that

$$\langle \mathbf{n}_k\cdot\mathbf{n}_k - 1, \mathbf{v}\cdot\mathbf{v}\rangle_0 = \int_\Omega (\mathbf{n}_k\cdot\mathbf{n}_k - 1)(\mathbf{v}\cdot\mathbf{v})\, dV.$$

If $\alpha \le \mathbf{n}_k\cdot\mathbf{n}_k \le \beta$ for all $\mathbf{x}\in\Omega$ with $0 < \alpha \le 1 \le \beta$, then $(\alpha - 1) \le 0$, and

$$\langle \mathbf{n}_k\cdot\mathbf{n}_k - 1, \mathbf{v}\cdot\mathbf{v}\rangle_0 \ge (\alpha-1)\int_\Omega \mathbf{v}\cdot\mathbf{v}\, dV \ge (\alpha-1)\|\mathbf{v}\|_{DC}^2. \tag{4.5}$$

Letting $\beta_0 = \alpha_0 - 2\zeta|\alpha - 1|$,

$$\beta_0\|\mathbf{v}\|_{DC}^2 \le \hat{a}(\mathbf{v},\mathbf{v}) + 4\zeta\langle \mathbf{v}\cdot\mathbf{n}_k, \mathbf{v}\cdot\mathbf{n}_k\rangle_0 + 2\zeta\langle \mathbf{n}_k\cdot\mathbf{n}_k - 1, \mathbf{v}\cdot\mathbf{v}\rangle_0.$$

Hence, if $2\zeta|\alpha-1| < \alpha_0$, then $\beta_0 > 0$. □

Therefore, $a(\mathbf{u},\mathbf{v})$ is coercive for $\kappa = 1$ if ζ is not so large in comparison to the pointwise lower bound on the director length as to overwhelm α_0.

As in Section 3.5.3, the assumption that $\kappa = 1$ can be loosened to include some anisotropy and retain coercivity of $a(\mathbf{u},\mathbf{v})$. Let $C > 0$ such that $\|\mathbf{v}\|_0^2 \le C\big(\|\nabla\cdot\mathbf{v}\|_0^2 + \|\nabla\times\mathbf{v}\|_0^2\big)$ (see Lemma 3.5.8). Further, let $\alpha_1 > 0$ be defined as in the proof of Lemma 3.5.9, where $K' = \min(K_1, \eta K_3)$ and $\alpha_1 = \frac{K'}{(C+1)}$. The following extends the results of Lemma 3.5.9 to the penalty method.

Lemma 4.2.4 (Small Data) *Under Assumption 3.5.1, if*

$$\beta_1 = \frac{\min(K_1, K_3)}{C+1} - 2\zeta|\alpha - 1| > 0,$$

there exists $\epsilon_1, \epsilon_2 > 0$*, dependent on* $\beta = \max|\mathbf{n}_k|^2$*, such that for* $\kappa \in (1-\epsilon_2, 1+\epsilon_1)$*,* $a(\mathbf{u}, \mathbf{v})$ *is coercive.*

Proof: Let

$$\begin{aligned}\tilde{a}(\mathbf{v}, \mathbf{v}) = K_1\langle \nabla \cdot \mathbf{v}, \nabla \cdot \mathbf{v}\rangle_0 + K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla \times \mathbf{v}, \nabla \times \mathbf{v}\rangle_0 + 4\zeta\langle \mathbf{v}\cdot\mathbf{n}_k, \mathbf{v}\cdot\mathbf{n}_k\rangle_0 \\ + 2\zeta\langle \mathbf{n}_k\cdot\mathbf{n}_k - 1, \mathbf{v}\cdot\mathbf{v}\rangle_0.\end{aligned}$$

From the proof of Lemma 3.5.9, the fact that $\zeta > 0$, and (4.5),

$$(\alpha_1 - 2\zeta|\alpha - 1|)\|\mathbf{v}\|_{DC}^2 \le \tilde{a}(\mathbf{v}, \mathbf{v}). \tag{4.6}$$

The USPD lower bound for $\mathbf{Z}(\mathbf{n}_k)$, η, may depend on κ; see Lemma 3.4.1. Thus, the proof is split into three cases.

Case 1. $\kappa = 1 + \epsilon_1$, for $\epsilon_1 > 0$.

If this case holds, then $\eta = 1$. Hence, α_1, defined for (4.6), is independent of κ. Since $K_2 - K_3 = K_3(\kappa - 1)$, the discrete bilinear form of (4.3) becomes

$$\begin{aligned}a(\mathbf{v}, \mathbf{v}) =& \tilde{a}(\mathbf{v}, \mathbf{v}) + \epsilon_1 K_3\Big(2\langle \mathbf{v}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0 + 2\langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 \\ &+ \langle \mathbf{v}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big).\end{aligned} \tag{4.7}$$

Observe that from (4.6),

$$(\alpha_1 - 2\zeta|\alpha - 1|) \le \tilde{a}(\mathbf{v}, \mathbf{v}) + \epsilon_1 K_3\langle \mathbf{v}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0. \tag{4.8}$$

Consider the magnitude of the terms in (4.7) not bounded from below in (4.8),

denoted as $\mathcal{G}(\mathbf{v},\mathbf{v})$. As in the proof of Lemma 3.5.9,

$$|\mathcal{G}(\mathbf{v},\mathbf{v})| \leq \epsilon_1\alpha_3\|\mathbf{v}\|^2_{DC}, \tag{4.9}$$

where α_3 is a constant defined in Lemma 3.5.9. Utilizing (4.8) and (4.9),

$$a(\mathbf{v},\mathbf{v}) \geq \alpha_1\|\mathbf{v}\|^2_{DC} - 2\zeta|\alpha-1|\|\mathbf{v}\|^2_{DC} - \epsilon_1\alpha_3\|\mathbf{v}\|^2_{DC} = (\beta_1 - \epsilon_1\alpha_3)\|\mathbf{v}\|^2_{DC}.$$

It is, thus, sufficient to have $\epsilon_1 < \beta_1/\alpha_3$, guaranteeing that $(\beta_1 - \epsilon_1\alpha_3) > 0$.

Case 2. $\kappa = 1-\epsilon_2 > 0$, for $\epsilon_2 > 0$, and $K_1 < K_3$.
Since $\kappa < 1$, $\eta = 1+(\kappa-1)\beta = (1-\epsilon_2\beta)$. For $K_1 < K_3$, there exists an ϵ_2 small enough, such that $K_1 < (1-\epsilon_2\beta)K_3$. This implies that, for small enough ϵ_2,

$$\alpha_1 = \frac{\min(K_1,(1-\epsilon_2\beta)K_3)}{(C+1)} = \frac{K_1}{(C+1)}.$$

Therefore, α_1 is again independent of κ. Since $K_2 - K_3 = K_3(\kappa-1)$, the discrete bilinear form of (4.3) becomes

$$\begin{aligned} a(\mathbf{v},\mathbf{v}) =& \tilde{a}(\mathbf{v},\mathbf{v}) - \epsilon_2 K_3\Big(2\langle \mathbf{v}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0 + 2\langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 \\ &+ \langle \mathbf{v}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big). \end{aligned} \tag{4.10}$$

The terms of (4.10), not already bounded from below in (4.6), are bounded, utilizing Lemma 3.5.9, as

$$|\mathcal{G}(\mathbf{v},\mathbf{v})| \leq \epsilon_2\alpha_4\|\mathbf{v}\|^2_{DC}, \tag{4.11}$$

where α_4 is a constant defined in Lemma 3.5.9. Using (4.6) and (4.11) implies that

$$a(\mathbf{v},\mathbf{v}) \geq \alpha_1\|\mathbf{v}\|^2_{DC} - 2\zeta|\alpha-1|\|\mathbf{v}\|^2_{DC} - \epsilon_2\alpha_4\|\mathbf{v}\|^2_{DC} \geq (\beta_1 - \epsilon_2\alpha_4)\|\mathbf{v}\|^2_{DC}.$$

Thus, possibly requiring ϵ_2 to be smaller, choose $\epsilon_2 < \beta_1/\alpha_4$, so that $(\beta_1 - \epsilon_2\alpha_4) > 0$.

In the case that $\kappa < 1$, the additional restriction that $\beta < \frac{1}{1-\kappa}$ for $\mathbf{Z}$ to be USPD is necessary, which implies that $\epsilon_2\beta < 1$ is required. Therefore, for any fixed β, ϵ_2 must also be taken small enough to satisfy this condition. Hence,

$$\epsilon_2 < \min\left(\frac{\beta_1}{\alpha_4}, \frac{K_3 - K_1}{\beta K_3}, \frac{1}{\beta}\right).$$

Case 3. $\kappa = 1 - \epsilon_2 > 0$, for $\epsilon_2 > 0$, and $K_3 \leq K_1$.

Here, again, $\eta = (1 - \epsilon_2\beta)$. For this case, it is clear that $(1 - \epsilon_2\beta)K_3 < K_1$. Thus,

$$\alpha_1 = \frac{(1 - \epsilon_2\beta)K_3}{(C+1)}.$$

Using the same α_4 as in the previous case and similar arguments,

$$\begin{aligned} a(\mathbf{v}, \mathbf{v}) &\geq \alpha_1 \|\mathbf{v}\|_{DC}^2 - 2\zeta|\alpha - 1|\|\mathbf{v}\|_{DC}^2 - \epsilon_2\alpha_4\|\mathbf{v}\|_{DC}^2 \\ &= \left(\frac{K_3}{C+1} - 2\zeta|\alpha - 1| - \frac{\epsilon_2\beta K_3}{C+1} - \epsilon_2\alpha_4\right)\|\mathbf{v}\|_{DC}^2 \\ &\geq \left(\beta_1 - \frac{\epsilon_2\beta K_3}{C+1} - \epsilon_2\alpha_4\right)\|\mathbf{v}\|_{DC}^2. \end{aligned}$$

Hence, in order for $\left(\beta_1 - \frac{\epsilon_2\beta K_3}{C+1} - \epsilon_2\alpha_4\right) > 0$ to hold, it is necessary that

$$\epsilon_2 < \frac{\beta_1(C+1)}{K_3\beta + \alpha_4(C+1)}.$$

Finally, ϵ_2 must still be chosen sufficiently small with respect to β such that $\epsilon_2\beta < 1$, as in Case 2. Therefore,

$$\epsilon_2 < \min\left(\frac{\beta_1(C+1)}{K_3\beta + \alpha_4(C+1)}, \frac{1}{\beta}\right).$$

□

Although the bounds on ϵ_1 and ϵ_2 are complicated, the only constant therein that depends on ζ is β_1. The remainder of the constants are independent of ζ.

The above lemmas allow for the formulation of the following summary theorem.

Theorem 4.2.5 *Under Assumption 3.5.1, if the conditions of Lemma 4.2.3 or Lemma 4.2.4 are satisfied, the discrete variational problem in* (4.4) *is well-posed.*

Proof: Lemmas 4.2.1 and 4.2.2 imply that $F(\mathbf{v})$ and $a(\mathbf{u}, \mathbf{v})$ are continuous, respectively. Lemmas 4.2.3 or 4.2.4 imply that $a(\mathbf{u}, \mathbf{v})$ is coercive. Therefore, by the Lax-Milgram Theorem [14], (4.4) is a well-posed discrete variational problem. □

Therefore, the discretization of the linearizations arising in the penalty method are always well-posed under the assumption of small anisotropy in the system coefficients and sufficient conformance to the unit-length constraint. However, the penalty parameter must be chosen appropriately to achieve accurate representation of the unit-length constraint. If ζ is too small, constraint conformance becomes poor and the functional minimum does not accurately represent the constrained minimum. Alternatively, if ζ is too large, the solvability of the intermediate variational systems degrades in two possible ways. The neighborhood admitting coercivity around $\kappa = 1$ shrinks and the system becomes increasingly ill-conditioned due to a decreasing coercivity constant, or coercivity is lost entirely and the possibility of non-invertible matrices appears. On the other hand, the proof of such well-posedness does not require establishing an inf-sup condition that necessitates subtle choices of finite-element spaces, as used for the Lagrange multiplier approach.

4.3 Robust Newton Step Methods

The Newton method applied to the Lagrange multiplier formulation discussed in Chapter 3 employs naïve damped Newton stepping. That is, for a computed Newton update direction, $\delta\mathbf{n}$, a constant damping factor, $0 < \omega \leq 1$, is applied such that the new iterate is given as $\mathbf{n}_{k+1} = \mathbf{n}_k + \omega\delta\mathbf{n}$. Such an approach aims to improve convergence robustness when dealing with an inaccurate initial guess on coarse grids. However, this procedure may miss opportunities to take larger steps in "good" descent directions that effectively reduce the free energy.

Trust-region techniques are specifically designed to improve the robustness and efficiency of iterative procedures such as Newton's method. Updates are confined

to a neighborhood, known as a trust region, where the accuracy of the linearized first-order optimality conditions is "trusted". These neighborhoods are expanded or contracted based on a measure of the model fidelity for a computed update. Significant research has produced both theoretical support and practical applications of such techniques [100]. This section discusses the use of constrained and unconstrained trust-region methods for the Lagrangian and penalty approaches. For a general overview of trust-region methods, see [100].

4.3.1 Trust-Region Approaches for the Penalty Formulation

Using the penalty functional in (4.1), the desired energy minimization is unconstrained. For this subsection, we denote the discretized forms of $\frac{\partial}{\partial \mathbf{n}} \left(\mathcal{P}_{\mathbf{n}}(\mathbf{n}_k)[\mathbf{v}]\right)[\delta\mathbf{n}]$ and $\mathcal{P}_{\mathbf{n}}(\mathbf{n}_k)[\mathbf{v}]$ as U_k and $\mathbf{f}_k$, respectively. The quadratic model of the penalty functional, for a given $\mathbf{n}_k$, is written

$$M_k(\delta\mathbf{n}) = \mathcal{P}(\mathbf{n}_k) + \mathbf{f}_k^T \delta\mathbf{n} + \frac{1}{2}\delta\mathbf{n}^T U_k \delta\mathbf{n}. \tag{4.12}$$

As a consequence of the well-posedness theory developed in Section 4.2, the matrix U_k is positive definite for each iteration. Therefore, we follow the methodology in [19, 113], computing steps by solving a trust-region minimization problem.

We seek an efficiently computable correction, $\delta\mathbf{n}$, that approximately minimizes the model in (4.12). In the following, we introduce two approaches to computing a step length and direction for this problem. The performance of these techniques is vetted in the numerical experiments below.

Damped Newton stepping is equivalent to taking a small step in the descent direction, $-U_k^{-1}\mathbf{f}_k$. As seen in the previous chapter, this is an effective means of finding energy minimizing solutions for both the penalty and Lagrangian methods. Therefore, in the first approach, a simple step selection technique is used in which the step is chosen satisfying the constrained minimization problem

$$\delta\mathbf{n}(\Delta_k) = \operatorname{argmin}\{\mathcal{P}(\mathbf{n}_k) + \mathbf{f}_k^T \delta\mathbf{n} + \frac{1}{2}\delta\mathbf{n}^T U_k \delta\mathbf{n} : |\delta\mathbf{n}| \le \Delta_k,\ \delta\mathbf{n} = \mu U_k^{-1}\mathbf{f}_k\}, \tag{4.13}$$

where Δ_k indicates the trust-region radius for iterate $\mathbf{n}_k$. Candidate solutions of (4.13) are easily computed to be $-U_k^{-1}\mathbf{f}_k$, the full Newton step, which may or may not be inside the trust region, and $\pm\frac{\Delta_k}{|U_k^{-1}\mathbf{f}_k|}U_k^{-1}\mathbf{f}_k$, representing steps to the trust region boundary.

An important aspect of trust-region methods is the adjustment of the trust-region radius and application of a computed step. This typically involves a measure of a computed step's merit. For a computed step, $\delta\mathbf{n}$, we compute the ratio,

$$\rho_k = \frac{\mathcal{P}(\mathbf{n}_k) - \mathcal{P}(\mathbf{n}_k + \delta\mathbf{n})}{M_k(\mathbf{0}) - M_k(\delta\mathbf{n})},$$

of the actual to the predicted reduction in $\mathcal{P}$ due to the computed step. The closer ρ_k is to 1, the more accurately the quadratic model behavior matches that of the true functional.

If the ratio, ρ_k, is deemed acceptable, the step is applied and the trust region expands, remains static, or shrinks depending on the specific value of ρ_k. If ρ_k is too small, the step is rejected, the trust-region radius is shrunk, and the process repeated. To quantify, let $0 < \eta_3 < \eta_1 < \eta_2$ be positive constants, along with $0 < C_1 < 1 < C_3$. Further, let $\bar{\Delta}$ be a maximum limit on the trust-region size. Using these parameters, the specific decision trees for accepting the step, and subsequently adjusting the trust region, are given in Procedures 2 and 3, respectively.

Procedure 2: Solution update.

if $\rho_k > \eta_3$ **then**
| Accept step: $\mathbf{n}_{k+1} = \mathbf{n}_k + \delta\mathbf{n}$.
else
| Reject step: $\mathbf{n}_{k+1} = \mathbf{n}_k$.
end

Procedure 3: TR size adjustment.

if $\rho_k < \eta_1$ **then**
| Shrink region: $\Delta_{k+1} = C_1\Delta_k$.
else if $\rho_k > \eta_2$ *and* $|\delta\mathbf{n}| = \Delta_k$ **then**
| Expand region:
| $\Delta_{k+1} = \min(C_3\Delta_k, \bar{\Delta})$.
else
| Keep region constant:
| $\Delta_{k+1} = \Delta_k$.
end

For our algorithm, if the components of the ratio, ρ_k, are very small and the computed step lies on the interior of the trust region, representing a full step towards satisfying the first-order optimality conditions, we choose to apply the step regardless of ρ_k and the trust region remains static. In this way, the trust-region minimization approach is used until we trust in the application of full Newton steps to obtain the first-order optimality conditions. A set of typical values for the trust-region parameters discussed above are listed in Table 4.1 and used in the numerical methods below.

A number of well-founded techniques improving trust-region step selection exist, including dogleg and two-dimensional (2D) subspace methods [19, 48, 100, 115]. Because the 2D-subspace method subsumes both the simple step selection approach above and dogleg methods, it is chosen as the alternative step selection computation here. Steps are computed by solving

$$\delta\mathbf{n}(\Delta_k) = \operatorname{argmin}\{\mathcal{P}(\mathbf{n}_k) + \mathbf{f}_k^T\delta\mathbf{n} + \frac{1}{2}\delta\mathbf{n}^T U_k\delta\mathbf{n} : |\delta\mathbf{n}| \le \Delta_k,\ \delta\mathbf{n} = \mu_1\mathbf{f}_k + \mu_2 U_k^{-1}\mathbf{f}_k\}. \tag{4.14}$$

Again, the candidate solutions for (4.14) are efficiently computable, amounting to solving for the zeroes of a fourth-order polynomial. Say that $\delta\mathbf{n} = \mu_1\mathbf{f}_k + \mu_2 U_k^{-1}\mathbf{f}_k$. Let

$$a_1 = |\mathbf{f}_k|^2, \qquad a_2 = (\mathbf{f}_k, U_k\mathbf{f}_k),$$
$$a_3 = (\mathbf{f}_k, U_k^{-1}\mathbf{f}_k), \qquad a_4 = |U_k^{-1}\mathbf{f}_k|^2.$$

Computing candidate solutions for the minimization problem in (4.14) yields

$$\mu_1 = \frac{2a_3^2\gamma - 2a_1a_4\gamma}{-a_1^2 - 2a_1a_3\gamma + a_3a_2 - 4a_3^2\gamma^2 + 2a_4\gamma(a_2 + 2a_1\gamma)},$$
$$\mu_2 = \frac{-a_2a_3 + a_1^2}{-a_1^2 - 2a_1a_3\gamma + a_3a_2 - 4a_3^2\gamma^2 + 2a_4\gamma(a_2 + 2a_1\gamma)},$$

where γ is a Lagrange multiplier for the constrained minimization and is a real root

of the fourth order polynomial

$$
\begin{aligned}
&16\Delta_k^2\gamma^4(a_3^4 + a_1^2a_4^2 - 2a_1a_3^2a_4) + 16\Delta_k^2\gamma^3(a_1a_3^3 + a_1a_2a_4^2 - a_2a_3^2a_4 - a_1^2a_3a_4) \\
&\quad + 4\gamma^2\Big(\Delta_k^2(a_2^2a_4^2 - 2a_2a_3^3 + 3a_1^2a_3^2 - 2a_1^3a_4) - a_1(a_3^2 - a_1a_4)^2\Big) \\
&\quad + 4\gamma\Big(\Delta_k^2(a_1^3a_3 - a_1a_2a_3^2 - a_1^2a_2a_4 + a_2^2a_3a_4) - a_3(a_3^2 - a_1a_4)(a_1^2 - a_3a_2)\Big) \\
&\quad + \Big(\Delta_k^2(a_1^4 + a_2^2a_3^2 - 2a_1^2a_2a_3)\Big) - a_4(a_1^2 - a_3a_2)^2 = 0.
\end{aligned}
$$

4.3.1.1 A Renormalization Penalty Method

In addition to the standard penalty method discussed above, a modification is also considered in the numerical experiments below. Once the approximation to the solution has been updated with a computed and accepted step, the new approximation is renormalized at the finite-element nodes. That is, the updated approximation is projected onto the unit sphere at each finite-element node. This procedure is similar to that presented in [53] for a Lagrange multiplier formulation. There, the approach is derived within a nullspace method framework using the one-constant approximation. Here, renormalization is applied to the penalty method, with and without trust regions and nested iteration, for anisotropic Frank constants.

This renormalization aims at improving unit-length conformance for solutions computed by the penalty method. The expectation is that this will lead to enhanced constraint conformance at lower penalty weights. However, unless the renormalization scaling is relatively uniform across nodes, the Newton direction may be significantly altered. Throughout this chapter, this modification is referred to as the "renormalization" penalty method.

4.3.2 Trust-Regions for the Lagrange Multiplier Approach

Applications of trust-region techniques to optimization problems with nonlinear constraints have also been developed. However, certain challenges arise in the theory and practical use of such methods [92]. Here, we consider existing trust-region approaches in the context of finite-element methods. For this subsection, let W_k

be the matrix associated with a finite-element discretization of the second-order derivative of (2.5) (i.e., the functional without the Lagrange multiplier term), given by $\frac{\partial}{\partial \mathbf{n}} \left(\mathcal{F}_{\mathbf{n}}(\mathbf{n}_k)[\mathbf{v}] \right) [\delta \mathbf{n}]$. For the trust-region approach, write the constraint

$$c(\mathbf{n}) = \langle \mathbf{n} \cdot \mathbf{n} - 1, \mathbf{n} \cdot \mathbf{n} - 1 \rangle_0 = 0. \tag{4.15}$$

The Gâteaux derivative of (4.15) is

$$\frac{\partial}{\partial \mathbf{n}} c(\mathbf{n})[\mathbf{v}] = 4 \langle \mathbf{n} \cdot \mathbf{n} - 1, \mathbf{n} \cdot \mathbf{v} \rangle_0. \tag{4.16}$$

Finally, let $\mathbf{c}_k$ be the column vector representing the finite-element discretized form of (4.16) at iterate $\mathbf{n}_k$.

One of the significant advantages of finite-element discretizations is the inherent sparsity of the resulting matrices. Trust-region algorithms in the Byrd-Omojokun family [18, 102, 118] require computation of the generally non-sparse matrix N_k, whose columns form an orthonormal basis for the orthogonal complement of $\mathbf{c}_k$, as well as the formation and inversion of the matrix $N_k^T W_k N_k$. In general, the matrix $N_k^T W_k N_k$ is quite large and dense, as W_k has dimension $m \times m$ and N_k is $m \times (m-1)$, where m is the number of discretization degrees of freedom for $\mathbf{n}$. Storage and computation with these dense matrices proves to be prohibitive, even on relatively small grids. Therefore, any advantages garnered by the use of these trust regions is outweighed by loss of the finite-element sparsity. Similarly, trust-region methods based on the fundamental work in [123] suffer from sparsity fill-in issues for large matrices in the context of finite-element methods.

To preserve sparsity properties, while still maintaining some advantages of a trust-region approach, we implement a simple trust-region method specifically fitted to the Lagrange multiplier formulation of the minimization problem. For the Lagrange multiplier approach in Section 3.3, we compute a Newton update direction, $\delta\chi = [\delta \mathbf{n} \ \delta\lambda]^T$. This update is meant to bring $\mathbf{n}_k$ and λ_k closer to satisfying the first-order optimality conditions. Let $\mathcal{L}_0(\mathbf{n}_k, \lambda_k)$ represent the finite-element

discretized form of the right-hand-side of Equation (3.4) for $\mathbf{n}_k$ and λ_k. Define the proportions w_k and w_{lim}, such that $0 < w_{\text{lim}} \leq w_k \leq 1$, where w_{lim} is a lower bound for w_k. For a given step, $w_k\delta\chi$, the expected change in $|\mathcal{L}_0(\mathbf{n}_k, \lambda_k)|$ is equal to $w_k|\mathcal{L}_0(\mathbf{n}_k, \lambda_k)|$. Therefore, we define the ratio

$$\rho_k = \frac{|\mathcal{L}_0(\mathbf{n}_k, \lambda_k)| - |\mathcal{L}_0(\mathbf{n}_k + w_k\delta\mathbf{n}, \lambda_k + w_k\delta\lambda)|}{w_k|\mathcal{L}_0(\mathbf{n}_k, \lambda_k)|}.$$

This ratio compares the change in $\mathcal{L}_0(\mathbf{n}, \lambda)$ predicted by the linearized model to the true change in $\mathcal{L}_0(\mathbf{n}, \lambda)$ for a computed step.

Procedure 4: Solution update.

if $\rho_k > \eta_2$ *or* $w_k = w_{lim}$ **then**
  Accept step:
  $[\mathbf{n}_{k+1}\ \lambda_{k+1}]^T = [\mathbf{n}_k\ \lambda_k]^T + w_k\delta\chi.$
else
  Reject step:
  $[\mathbf{n}_{k+1}\ \lambda_{k+1}]^T = [\mathbf{n}_k\ \lambda_k]^T.$
end

Procedure 5: TR size adjustment.

if $\rho_k < \eta_2$ **then**
  Shrink region:
  $w_{k+1} = \max(w_{\text{lim}}, w_k - w_{\text{dec}}).$
else if $\eta_2 < \rho_k < \eta_1$ **then**
  Keep region constant:
  $w_{k+1} = w_k.$
else
  Expand region:
  $w_{k+1} = \min(w_k + w_{\text{inc}}, 1).$
end

Let $0 < \eta_2 < \eta_1$ and $w_{\text{inc}}, w_{\text{dec}} \in (0, 1]$. Since w_k is a scaling factor, rather than a radius length, step selection and trust-region adjustment differ slightly from the procedures discussed above and are given in Procedures 4 and 5, respectively.

4.4 Numerical Results

In this section, we compare the performance of the methods outlined above for three benchmark equilibrium problems. The general algorithm utilized by each method is similar to that outlined in Algorithm 1; see Algorithm 6. The outer stage again uses nested iteration (NI). The iterative solution updates are computed via one of the methods described in the previous sections. In general, the iteration stopping

criterion, on a given level, is based on a set tolerance for the approximation's conformance to the first-order optimality conditions in the standard Euclidean l_2-norm. For the renormalization penalty method, the Newton iteration tolerance is based on the reduction of the ratio of the energy from the previous step to the current step rather than conformance to the first-order optimality conditions. In the numerical experiments carried out below, both tolerances are held at 10^{-4}. As with the numerical experiments above, the nested grid hierarchy is formed by uniform refinements of the initial coarse grid. However, adaptive refinement could also be performed.

The components of the variational problems in Equations (3.4) and (4.2) are discretized with finite elements on each grid. Both formulations use $Q_2 \times Q_2 \times Q_2$ elements for $\mathbf{n}$, while the Lagrange multiplier approach uses P_0 elements for λ, as in Section 3.7. In this section, the arising matrices are again inverted using the UMFPACK LU decomposition [32–35]. The algorithm's discretizations and grid management are performed with the deal.II library.

Algorithm 6: General minimization algorithm with NI

0. Initialize solution approximation on coarse grid.

while *Refinement limit not reached* **do**

  while *Nonlinear iteration tolerance not satisfied* **do**

    1. Assemble discrete components of System (3.4) or (4.2) on current grid, H.
    2. Compute correction to current approximation.
    3. Update current approximation.

  end

  4. Uniformly refine the grid to size h.
  5. Interpolate solution $\mathbf{u}_H \to \mathbf{u}_h$.

end

Each of the problems below is posed on a unit-square domain in the xy-plane, such that $\Omega = \{(x, y) \mid 0 \leq x, y \leq 1\}$. As in Section 3.7.2, we assume a slab domain such that $\mathbf{n}$ may have nonzero z-component but $\frac{\partial \mathbf{n}}{\partial z} = \mathbf{0}$. Dirichlet boundary conditions are applied at the y-edges and periodic boundary conditions are assumed at

the boundaries $x = 0$ and $x = 1$. The experiments to follow consider an 8×8 coarse mesh ascending in six uniform refinements to a 512×512 mesh.

For the numerical experiments, each of the trust-region methods discussed above is applied. For the penalty trust-region methods, the initial trust-region radius is set to Δ_{init}. At each refinement level, the trust-region radius is reset to Δ_{init} plus an incremental increase, Δ_{inc}, with a maximum of $\bar{\Delta}$. The Lagrangian trust-region approach sets the initial value of w_k to w_{init}. After each refinement, w_k is reset to w_{init} plus w_{lev}, up to a maximum of 1. These increments are due to the increasing accuracy of the iterates at each grid level. These constants are outlined in Tables 4.1 and 4.2

$\eta_1 = 0.25$	$\eta_2 = 0.75$	$\eta_3 = 0.125$	$C_1 = 0.5$
$C_3 = 1.3$	$\Delta_{\text{inc}} = 0.3$	$\bar{\Delta} = 100$	$\Delta_{\text{init}} = 0.3$

Table 4.1: Trust-region parameters for the penalty formulation.

$\eta_1 = 0.5$	$\eta_2 = 0.25$	$w_{\text{inc}} = 0.1$	$w_{\text{dec}} = 0.1$
$w_{\text{lev}} = 0.1$	$w_{\text{min}} = 0.1$	$w_{\text{init}} = 0.2$	–

Table 4.2: Trust-region parameters for the Lagrangian formulation.

The non-trust-region, damped Newton stepping approach is also performed for both formulations as a comparison benchmark with an initial $\omega = 0.2$, increasing by 0.2 at each refinement to a maximum of 1. The performance of each of these methods is then compared. Note that in the results to follow, all reported free energies are computed using only the free elastic quantities without any augmentations, such as the penalty terms.

4.4.1 Twist Equilibrium Configuration

The first set of boundary conditions induce a classical twist equilibrium configuration similar to the one seen in the second numerical experiment of Section 3.7. For this experiment, and the tilt-twist experiment in the next subsection, let the general form of the solution be

$$\mathbf{n} = \big(\cos(\theta(y))\cos(\phi(y)), \cos(\theta(y))\sin(\phi(y)), \sin(\theta(y))\big). \tag{4.17}$$

Note that the known analytical solutions have a one-dimensional structure, but the numerical experiments below are full two-dimensional simulations. For the twist configuration, let $\theta_0 = \frac{\pi}{8}$. At the boundaries $\theta(0) = -\theta_0$, $\theta(1) = \theta_0$, and $\phi(0) = \phi(1) = 0$. The Frank constants for this problem are $K_1 = 1.0$, $K_2 = 1.2$, and $K_3 = 1.0$. The analytical equilibrium solution for these boundary conditions and Frank constants is derived, under a rotated coordinate system, in [117]. The solution is given by

$$\mathbf{n} = (\cos(\theta_0(2y-1)), 0, \sin(\theta_0(2y-1))),$$

with true free-elastic energy $2K_2\theta_0^2$. This corresponds to an expected free energy of 0.37011. The existence of an analytical solution for this problem allows for the computation of an L^2-error for each computed approximation.

Type	Free Energy	L^2-error	Min. Dev.	Max Dev.	Cost	TR Cost
Lagrangian	0.370110	2.076e-11	−1.43e-14	7.00e-15	1.350	1.340
Pen. $\zeta = 10^1$	0.358832	1.589e-02	−3.96e-02	−3.59e-05	1.371	1.354
Pen. $\zeta = 10^2$	0.368481	1.993e-03	−4.32e-03	−1.16e-05	1.376	1.355
Pen. $\zeta = 10^3$	0.369931	2.107e-04	−4.32e-04	−3.68e-06	1.440	1.418
Pen. $\zeta = 10^4$	0.370092	2.143e-05	−4.32e-05	−1.14e-06	1.448	1.420
Pen. $\zeta = 10^5$	0.370108	2.154e-06	−4.32e-06	−3.32e-07	1.447	1.426
Pen. $\zeta = 10^6$	0.370110	2.157e-07	−4.32e-07	−7.27e-08	–	1.436
Pen. $\zeta = 10^7$	0.370110	2.158e-08	−5.05e-08	−9.98e-09	–	1.465
Pen. $\zeta = 10^8$	0.370110	2.158e-09	−5.18e-09	−1.06e-09	–	1.516
Pen. $\zeta = 10^9$	0.370110	2.168e-10	−5.19e-10	−1.06e-10	–	1.639

Table 4.3: Statistics for the twist equilibrium solution with the different formulations and penalty weights. Included is the system free energy, the computed L^2-error on the finest grid, and the minimum and maximum deviations from unit director length at the quadrature nodes. Approximations of the cost in WUs for the corresponding method with no trust regions and simple trust regions are included. Dashes in the columns indicate divergence.

Table 4.3 compares the performance of the Lagrange multiplier method to the penalty method without renormalization. The runs were performed with nested iteration and the approximate work, measured in WUs, is given for the corresponding method with no trust regions and the simple trust region approaches, respectively. Observe that the non-trust-region, damped Newton stepping discussed above diverged for penalty parameters of $\zeta = 10^6$ and greater. However, smaller damping parameters may yield convergence. Both penalty-method trust-region approaches

converged without modification.

The table demonstrates the superior performance of the Lagrange multiplier method for this problem across all statistics with lower error, cost, and tighter conformance to the constraint. The penalty method does not match the free energy obtained by the Lagrangian formulation until reaching a penalty weight of 10^6 and, without trust regions, encounters divergence issues for these large penalty weights. While trust regions do not significantly reduce overall computations costs, Table 4.3 suggests that they significantly improve robustness.

	No Trust Region		Simple Trust Region		2D Trust Region	
Type	L^2-error	Cost	L^2-error	Cost	L^2-error	Cost
Pen. $\zeta = 10^1$	1.457e-02	1.338	1.457e-02	1.334	1.457e-02	1.334
Pen. $\zeta = 10^2$	8.932e-05	1.338	8.931e-05	1.334	8.931e-05	1.334
Pen. $\zeta = 10^3$	3.358e-06	1.339	3.357e-06	1.334	3.357e-06	1.335
Pen. $\zeta = 10^4$	1.523e-07	1.340	1.116e-07	1.336	1.116e-07	1.336
Pen. $\zeta = 10^5$	6.260e-08	8.113	3.595e-09	1.364	3.592e-09	1.340
Pen. $\zeta = 10^6$	6.356e-06	81.120	1.688e-02	73.052	1.098e-07	2.731

Table 4.4: A comparison of renormalization penalty methods, with and without trust-region approaches, for the twist solution. For each algorithm, the computed L^2-error on the finest grid and an approximation of the cost in WUs is included.

Type	Free Energy	L^2-error	Min. Dev.	Max Dev.	2D TR Cost
Pen. $\zeta = 10^1$	0.370168	1.457e-02	−4.58e-11	4.58e-11	1.334
Pen. $\zeta = 10^2$	0.370111	8.931e-05	−1.68e-11	1.68e-11	1.334
Pen. $\zeta = 10^3$	0.370110	3.357e-06	−5.18e-12	5.16e-12	1.335
Pen. $\zeta = 10^4$	0.370110	1.116e-07	−1.45e-12	1.43e-12	1.336
Pen. $\zeta = 10^5$	0.370110	3.592e-09	−3.16e-13	2.98e-13	1.340
Pen. $\zeta = 10^6$	0.370110	1.098e-07	−4.04e-14	2.20e-14	2.731

Table 4.5: Statistics for the twist equilibrium solution with different penalty weights. Here, the penalty method with renormalization and 2D-subspace minimization is considered. Included is the system free energy, the computed L^2-error on the finest grid, the minimum and maximum deviations from unit director length at the quadrature nodes, and an approximation of the cost in WUs for the corresponding method.

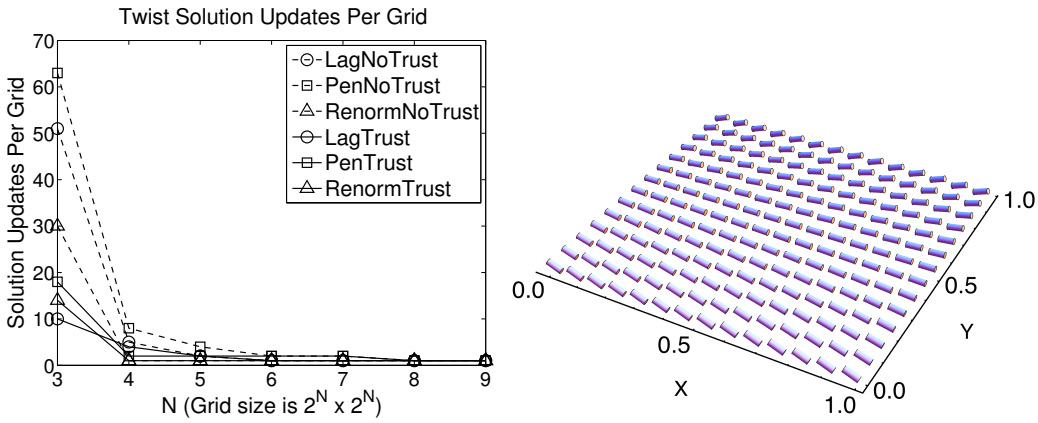
The results in Tables 4.4 and 4.5 show the performance of the renormalization penalty method with and without trust regions. Table 4.5 provides additional statistics for the 2D-subspace minimization trust-region approach discussed in Table 4.4. For the twist equilibrium solution, the renormalization penalty method obtains better error values for smaller penalty weights than the unmodified penalty method. In Table 4.5, using the 2D-subspace minimization trust-region approach, we obtain

an error of 3.592e-09 with a penalty weight of only $\zeta = 10^5$. Moreover, the minimum and maximum deviation of the director at the quadrature nodes is closer to that of the Lagrangian method. However, the performance improvements rely more heavily on the penalty parameter. While an error measure closer to the Lagrange multiplier formulation is achieved for $\zeta = 10^5$, performance degrades at $\zeta = 10^6$, with notable jumps in costs for all methods recorded in Table 4.4. The increases in error are due to the algorithm beginning to emphasize the unit-length constraint over proper director orientation. Correctly selecting the penalty weight represents a fundamental difficulty for this method.

Figure 4.1a displays the number of iterations required to reach the specified iteration tolerance within a nested iteration scheme alongside the final solution computed by the Lagrange multiplier formulation in Figure 4.1b. Counts for both the Lagrange multiplier approach and penalty formulation, with and without renormalization, for a penalty parameter $\zeta = 10^3$ are shown. In general, the trust-region methods significantly reduce iteration counts on the coarse grids. However, on the finer grids, this reduction is not sustained due to the efficiency of nested iteration. Because the improved iteration counts are confined to the coarsest grids, overall cost reduction is generally small. For example, the approximate cost for the Lagrange multiplier method was reduced very slightly from 1.350 WUs to 1.340 WUs, resulting in only a one second drop in overall time to solution.

Table 4.6 summarizes both the efficiency of nested iteration and highlights the strengths of certain applications of trust-region methods. For all of the constraint enforcement formulations, nested iteration offers very clear cost improvements. Coupling nested iteration with the Lagrange multiplier method for this problem is quite powerful, yielding the fastest overall run time and highest accuracy. Trust regions have a clear impact on time to solution in the absence of nested iteration but offer modest time to solution improvements when coupled with NI.

If the penalty method is used, pairing nested iteration with trust regions increases robustness and cost consistency. For example, the use of trust regions for



(a) (b)

Figure 4.1: (a) Number of iterations required to reach iteration tolerance for each method with NI. The penalty weight for the penalty formulation was $\zeta = 1000$. Only the 2D-subspace minimization trust-region approach is displayed, as the behavior of simple trust regions is similar. (b) The final computed solution for the Lagrangian formulation on a 512×512 mesh (restricted for visualization).

Lagrangian			
Method		Solve Cost	Run Time
No NI	No TR	61	17,975s
NI	No TR	1.350	550s
No NI	TR	10	3,071s
NI	TR	1.340	548s
Renormalization Penalty: $\zeta = 10^5$			
Method		Solve Cost	Run Time
No NI	No TR	38	11,838s
NI	No TR	8.113	2,272s
No NI	TR	29	9,172s
No NI	TR 2D	32	10,147s
NI	TR	1.364	585s
NI	TR 2D	1.340	584s

Unmodified Penalty: $\zeta = 10^5$			
Method		Solve Cost	Run Time
No NI	No TR	142	41,013s
NI	No TR	1.474	593s
No NI	TR	63	18,425s
No NI	TR 2D	64	19,287s
NI	TR	1.426	569s
NI	TR 2D	1.424	574s
Unmodified Penalty: $\zeta = 10^9$			
Method		Solve Cost	Run Time
No NI	No TR	–	–
NI	No TR	–	–
No NI	TR	1,016	294,349s
No NI	TR 2D	1,736	511,874s
NI	TR	1.639	641s
NI	TR 2D	1.958	764s

Table 4.6: Twist statistics comparison for NI and trust region combinations. The solve cost column displays an approximation of the work in WUs for the corresponding method. The overall time to solution is also presented. Dashes in the columns indicate divergence.

the unmodified penalty method with $\zeta = 10^9$ overcomes prominent divergence issues. In addition to the improved error performance in Table 4.4, for $\zeta = 10^5$, the renormalization penalty method is generally faster than the unmodified penalty approach with the same penalty weight. The slightly slower overall run times, in spite of the small WU costs, when NI is paired with trust regions, in comparison with the

unmodified penalty method, are due to the work involved in normalizing the director after each iteration. As discussed above, a shortcoming of the renormalization penalty method is sensitivity to parameter choice.

4.4.2 Tilt-Twist Equilibrium Configuration

For this problem, $\mathbf{n}$ retains the form in (4.17) and the same boundary conditions are applied with $\theta_0 = \frac{\pi}{4}$ and Frank constants of $K_1 = 1.0$, $K_2 = 3.0$, and $K_3 = 1.2$. Twist solutions incorporating a nonplanar tilt deviating from parallel alignment with the xz-plane are investigated in [81, 82]. It is shown that nonplanar twist solutions become energetically optimal at a computable threshold. This threshold is satisfied for the chosen parameters. The analytical, energy-minimizing, tilt-twist solution is defined implicitly for a rotated coordinate system in [81, 117].

For the coordinate system and boundary conditions here, let ϕ_m represent the maximum nonplanar tilt in the domain and define the functions

$$f(\phi) = K_1 \cos^2\phi + K_3 \sin^2\phi,$$
$$g(\phi) = (K_2 \cos^2\phi + K_3 \sin^2\phi)\cos^2\phi.$$

The following set of equations determine the analytical solution values of $\phi(y)$ and $\theta(y)$ implicitly for a given value $0 \le y \le \frac{1}{2}$. The corresponding values for $\frac{1}{2} < y \le 1$ are computed by symmetry as $\theta(y) = \theta(1-y)$ and $\phi(y) = \phi(1-y)$. First the value of ϕ_m is determined by solving the equation

$$\theta_0 = \sqrt{g(\phi_m)} \int_0^{\phi_m} \left(\frac{f(u)}{g(u)\big(g(u) - g(\phi_m)\big)} \right)^{1/2} du.$$

Using the computed value of ϕ_m, an intermediate value, b, is computed as

$$b = 2\sqrt{g(\phi_m)} \int_0^{\phi_m} \left(\frac{g(u)f(u)}{g(u) - g(\phi_m)} \right)^{1/2} du.$$

Using the values b and ϕ_m, the value of $\phi(y)$, for a given value of y, is determined

by solving the implicit equation

$$by = \sqrt{g(\phi_m)} \int_0^{\phi} \left(\frac{g(u)f(u)}{g(u) - g(\phi_m)} \right)^{1/2} du, \qquad 0 \le y \le \frac{1}{2},$$

for ϕ. Finally, the value for $\theta(y)$ is computed, using the calculated ϕ, as

$$\theta = -\theta_0 + \sqrt{g(\phi_m)} \int_0^{\phi} \left(\frac{f(u)}{g(u)\big(g(u) - g(\phi_m)\big)} \right)^{1/2} du, \qquad 0 \le y \le \frac{1}{2}.$$

The free energy associated with the nonplanar twist solution is minimizing if

$$b^2 < 4K_2\theta_0^2 g(\phi_m).$$

This inequality is satisfied for the chosen parameters and the associated analytical, free-elastic energy is 3.59294.

For the tilt-twist equilibrium solution, the damped Newton stepping approach converged for all of the penalty weights considered. Table 4.7 details the statistics for the unmodified penalty method compared with the Lagrange multiplier method. Again, the Lagrange multiplier method outperforms the penalty method in each category. The free energy of the Lagrange multiplier method is not obtained by the penalty method until ζ reaches 10^8.

Type	Free Energy	L^2-error	Min. Dev.	Max Dev.	Cost	TR Cost
Lagrangian	3.59294	4.717e-07	−7.89e-10	7.88e-10	1.463	1.447
Pen. $\zeta = 10^1$	2.15620	4.403e-01	−4.78e-01	−2.62e-04	1.458	1.333
Pen. $\zeta = 10^2$	3.38037	4.597e-02	−5.01e-02	−1.17e-04	1.732	1.665
Pen. $\zeta = 10^3$	3.56953	4.565e-03	−4.97e-03	−3.88e-05	1.732	1.665
Pen. $\zeta = 10^4$	3.59052	4.606e-04	−5.00e-04	−1.21e-05	2.735	2.665
Pen. $\zeta = 10^5$	3.59269	4.590e-05	−5.00e-05	−3.56e-06	2.743	2.667
Pen. $\zeta = 10^6$	3.59291	4.253e-06	−5.01e-06	−8.05e-07	2.782	2.678
Pen. $\zeta = 10^7$	3.59293	2.735e-07	−5.83e-07	−1.14e-07	2.809	2.723
Pen. $\zeta = 10^8$	3.59294	4.340e-07	−6.00e-08	−1.22e-08	2.885	2.747
Pen. $\zeta = 10^9$	3.59294	4.676e-07	−6.01e-09	−1.24e-09	3.218	2.879

Table 4.7: Statistics for the tilt-twist equilibrium solution with the different formulations and penalty weights. Included is the system free energy, the computed L^2-error on the finest grid, and the minimum and maximum deviations from unit director length at the quadrature nodes. Approximations of the cost in WUs for the corresponding method with no trust regions and simple trust regions are included.

It should be noted that the behavior of the error for the Lagrangian method, as well as the penalty method for weights greater than 10^7, is affected by the complexity of the equations, described above, implicitly defining the true solution. To compute the error, the equations describing the analytical solution are solved approximately at the appropriate quadrature points using Mathematica. Solving these equations involves successive root finding for complicated integral equations where the unknowns are limits of integration. Hence, approximation error creates an artificial limit for the computed solution error at accuracies smaller than 10^{-7}.

Considering Tables 4.8 and 4.9, the renormalization penalty method does not perform as well as in the previous problem. Note that Table 4.9 details additional statistics for the 2D-subspace minimization trust-region approach discussed in Table 4.8. Compared to the unmodified penalty method, computational costs remain steadier, with the exception of the run without trust regions and a penalty weight of 10^6, and adherence to the unit-length constraint is improved. However, the method fails to reach an equivalent accuracy before performance degrades. As with the simpler twist problem, performance of the renormalization method is sensitive to an appropriate choice of penalty weight.

	No Trust Region		Simple Trust Region		2D Trust Region	
Type	L^2-error	Cost	L^2-error	Cost	L^2-error	Cost
Pen. $\zeta = 10^1$	4.354e-01	1.384	4.319e-01	1.337	4.424e-01	1.377
Pen. $\zeta = 10^2$	3.691e-02	1.335	3.635e-02	1.333	3.578e-02	1.333
Pen. $\zeta = 10^3$	4.708e-03	1.335	4.533e-03	1.332	4.493e-03	1.332
Pen. $\zeta = 10^4$	1.085e-03	1.335	8.662e-04	1.332	8.536e-04	1.333
Pen. $\zeta = 10^5$	1.650e-03	1.336	9.487e-04	1.333	7.012e-04	1.333
Pen. $\zeta = 10^6$	9.414e-01	87.362	5.375e-04	1.341	7.344e-04	1.341

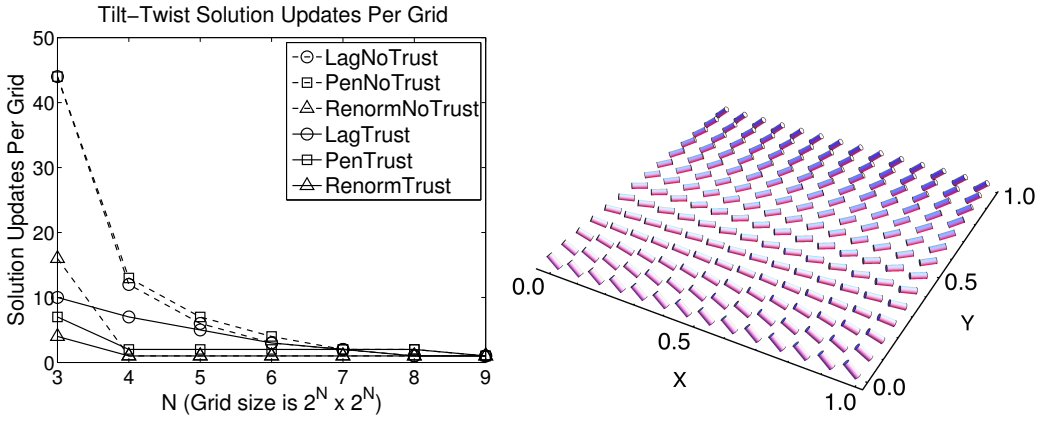
Table 4.8: A comparison of renormalization penalty methods, with and without trust-region approaches, for the tilt-twist solution. For each algorithm, the computed L^2-error on the finest grid and an approximation of the cost in WUs is included.

The method with renormalization does find the true free energy at a lower penalty weight than the approach without renormalization. At a penalty weight of $\zeta = 10^4$, the penalty method without renormalization has a slightly lower error measure, but has not fully matched the true energy. While the unmodified method

Type	Free Energy	L^2-error	Min. Dev.	Max Dev.	2D TR Cost
Pen. $\zeta = 10^1$	3.92827	4.424e-01	−4.62e-09	4.62e-09	1.377
Pen. $\zeta = 10^2$	3.59611	3.578e-02	−1.15e-09	1.14e-09	1.333
Pen. $\zeta = 10^3$	3.59298	4.493e-03	−7.77e-10	7.76e-10	1.332
Pen. $\zeta = 10^4$	3.59294	8.536e-04	−7.87e-10	7.85e-10	1.333
Pen. $\zeta = 10^5$	3.59294	7.012e-04	−7.91e-10	7.90e-10	1.333
Pen. $\zeta = 10^6$	3.59294	7.344e-04	−7.87e-10	7.86e-10	1.341

Table 4.9: Statistics for the tilt-twist equilibrium solution with varying penalty weights. Here, the penalty method with renormalization and 2D-subspace minimization is shown. Included is the system free energy, the computed L^2-error on the finest grid, the minimum and maximum deviations from unit director length at the quadrature nodes, and an approximation of the cost in WUs for the corresponding method.

more accurately resolves the orientation of the director in comparison with the renormalization method, it slightly shrinks the director length to attain the moderately smaller free energy.



(a) (b)

Figure 4.2: (a) Number of iterations required to reach iteration tolerance for each method with NI. The penalty weight for the penalty formulation was $\zeta = 1000$. Only the 2D-subspace minimization trust-region approach is displayed, as the behavior of simple trust regions is similar. (b) The final computed solution for the Lagrangian formulation on a 512×512 mesh (restricted for visualization).

Figure 4.2a presents similar behavior to Figure 4.1a, in that trust regions productively reduce the number of iterations on the coarsest grids but have less effect on iteration counts on the finest grids. This is again due to the efficiency of nested iteration. Figure 4.2b displays the solution computed by the Lagrange multiplier approach. Note that the solution correctly exhibits a non-planar tilt in addition to the simple planar twist. For the unmodified penalty method with NI and a penalty weight of $\zeta = 10^9$, trust regions only reduce the computational cost from 3.218 WUs

to 2.879 WUs. This results in only an 8.9% decrease in overall time to solution.

As shown in Table 4.10, improvements from the trust regions for the Lagrange multiplier method are minor, decreasing computational costs from 1.463 WUs to 1.447 WUs and reducing overall time to solution by only 0.82%. Here, the renormalization penalty method is faster than the unmodified approach and, in some cases, even slightly outpaces the Lagrange multiplier formulation. However, as shown in Tables 4.7 and 4.9 the associated error convergence is not comparable. In Table 4.10, results for $\zeta = 10^9$ are not reported due to untenably large run times without nested iteration.

Lagrangian			
Method		Solve Cost	Run Time
No NI	No TR	33	9,853s
NI	No TR	1.463	584s
No NI	TR	9	2,812s
NI	TR	1.447	579s
Renormalization Penalty: $\zeta = 10^5$			
Method		Solve Cost	Run Time
No NI	No TR	16	5,119s
NI	No TR	1.336	586s
No NI	TR	18	5,658s
No NI	TR 2D	18	5,817s
NI	TR	1.333	575s
NI	TR 2D	1.333	591s

Unmodified Penalty $\zeta = 10^5$			
Method		Solve Cost	Run Time
No NI	No TR	39	11,606s
NI	No TR	2.743	939s
No NI	TR	22	6,598s
No NI	TR 2D	22	6,680s
NI	TR	2.667	920s
NI	TR 2D	2.667	949s

Table 4.10: Tilt-twist statistics comparison for NI and trust region combinations. The solve cost column displays an approximation of the work in WUs for the corresponding method. The overall time to solution is also presented.

4.4.3 Nano-Patterned Boundary Conditions

In this numerical experiment, we use Frank constants $K_1 = 1.0$, $K_2 = 0.62903$, and $K_3 = 1.32258$. The applied boundary conditions are the same as those defined in Equations (3.75)-(3.77). These boundary conditions are pictured in Figure 4.3b. The result is a sharp transition from vertical to planar-aligned rods followed by a rapid transition back to vertical alignment. Such boundary conditions produce configuration distortions throughout the interior of the domain. Due to the boundary condition complexity, no analytical solution currently exists.

The more complicated nature of the nano-patterned boundary conditions is reflected in the data of Table 4.11. The overall approximate costs for the methods with and without trust regions are larger than previous examples and the unit-length constraint is more difficult to capture. Nonetheless, the Lagrange multiplier method provides an accurate and cost effective approach. The penalty method without trust regions diverges for penalty weights greater than $\zeta = 10^4$. At higher penalty weights, even the trust-region approach suffers jumps in computational costs. At $\zeta = 10^9$, the system becomes over constrained and accuracy begins to degrade. Hence, results for this weight are not included.

Type	Free Energy	Min. Dev.	Max Dev.	Cost	TR Cost
Lagrangian	3.89001	−6.92e-05	5.89e-05	2.864	2.779
Pen. $\zeta = 10^1$	3.83657	−8.84e-02	1.96e-03	2.864	2.748
Pen. $\zeta = 10^2$	3.86896	−4.01e-02	4.40e-03	2.864	2.749
Pen. $\zeta = 10^3$	3.88331	−1.80e-02	7.32e-03	2.868	2.749
Pen. $\zeta = 10^4$	3.88819	−6.58e-03	5.81e-03	2.886	2.757
Pen. $\zeta = 10^5$	3.88965	−1.60e-03	2.01e-03	–	2.805
Pen. $\zeta = 10^6$	3.88996	−2.90e-04	4.55e-04	–	3.736
Pen. $\zeta = 10^7$	3.89001	−7.92e-05	1.01e-04	–	4.797
Pen. $\zeta = 10^8$	3.89001	−6.76e-05	5.83e-05	–	22.328

Table 4.11: Statistics for the nano-patterned equilibrium solution with the different formulations and penalty weights. Included is the system free energy and the minimum and maximum deviations from unit director length at the quadrature nodes. Approximations of the cost in WUs for the corresponding method with no trust regions and simple trust regions are included. Dashes in the columns indicate divergence.

As with the tilt-twist equilibrium solution, the renormalization penalty method approaches the Lagrangian formulation's free energy and unit-length bounds earlier than the unmodified penalty method. It also yields a lower computational cost for most penalty weights. However, as was seen in the tilt-twist data, matching the energy earlier than the unmodified penalty approach does not directly indicate higher accuracy in resolving the correct orientation of the director. Moreover, in Table 4.12, divergence issues are apparent for the renormalization method at high penalty weights.

When considered with Table 4.11, Table 4.12 reinforces the conclusion that trust regions positively influence the robustness of penalty method approaches. While the simple trust-region approach works most effectively for the non-renormalization

penalty method, the 2D-subspace minimization approach is more favorable for the renormalization penalty formulation. Note that Table 4.13 details additional statistics for the 2D-subspace minimization trust-region approach discussed in Table 4.12.

	No Trust Region		Simple Trust Region		2D Trust Region	
Type	Free Energy	Cost	Free Energy	Cost	Free Energy	Cost
Pen. $\zeta = 10^1$	3.89319	1.680	3.89351	1.518	3.89308	1.686
Pen. $\zeta = 10^2$	3.89049	1.681	3.89051	1.666	3.89049	1.666
Pen. $\zeta = 10^3$	3.89006	1.682	3.89006	1.666	3.89006	1.666
Pen. $\zeta = 10^4$	3.89002	1.683	3.89002	1.669	3.89002	1.669
Pen. $\zeta = 10^5$	–	–	3.89001	2.133	3.89001	2.433
Pen. $\zeta = 10^6$	–	–	–	–	3.89001	5.418

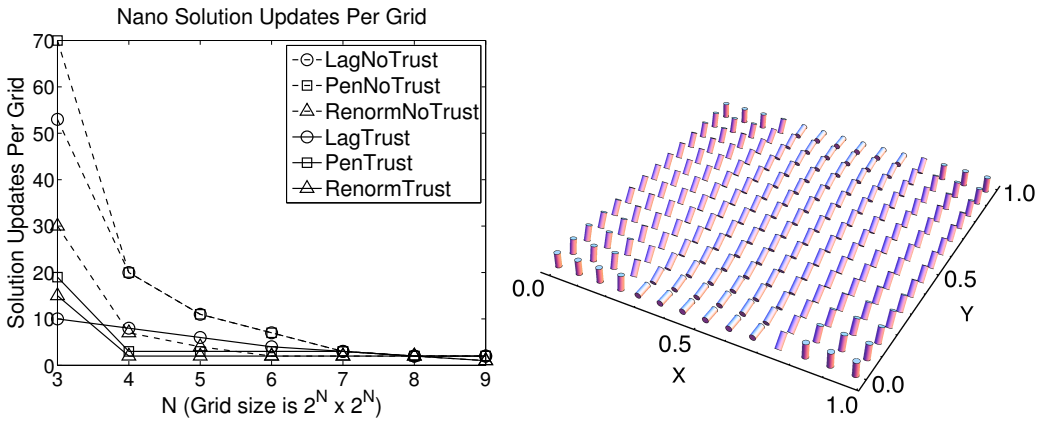
Table 4.12: A comparison of renormalization penalty methods, with and without trust-region approaches, for the nano-pattern solution. For each algorithm, the computed free energy on the finest grid and an approximation of the cost in WUs is included. Dashes in the columns indicate divergence.

Type	Free Energy	Min. Dev.	Max Dev.	2D TR Cost
Pen. $\zeta = 10^1$	3.89308	−7.06e-05	6.02e-05	1.686
Pen. $\zeta = 10^2$	3.89049	−7.07e-05	6.02e-05	1.666
Pen. $\zeta = 10^3$	3.89006	−7.09e-05	6.01e-05	1.666
Pen. $\zeta = 10^4$	3.89002	−7.09e-05	6.01e-05	1.669
Pen. $\zeta = 10^5$	3.89001	−7.08e-05	6.00e-05	2.433
Pen. $\zeta = 10^6$	3.89001	−7.07e-05	5.98e-05	5.418

Table 4.13: Statistics for the nano-patterned equilibrium solution with the different formulations and penalty weights. Here, the penalty method with renormalization and 2D-subspace minimization is used. Included is the system free energy, the minimum and maximum deviations from unit director length at the quadrature nodes, and an approximation of the cost in WUs for the corresponding method.

Figure 4.3a displays the iteration counts as a function of grid size for both the Lagrange multiplier formulation and the penalty method, with and without renormalization, at a penalty weight of $\zeta = 10^3$. The solution computed by the Lagrange multiplier method is shown in Figure 4.3b. Similar to the previous problems, trust regions reduce iteration counts on the coarsest grids with reduced efficacy at the finer levels, due to NI. The cost savings from trust regions within a nested iteration scheme are slightly higher for this problem but persist as small improvements overall.

Table 4.14 reiterates the efficacy of nested iteration for efficient computation and trust regions for robustness. For this problem, Table 4.14 additionally shows that



(a) (b)

Figure 4.3: (a) Number of iterations required to reach iteration tolerance for each method with NI. The penalty weight for the penalty formulation was $\zeta = 1000$. Only the 2D-subspace minimization trust-region approach is displayed, as the behavior of simple trust regions is similar. (b) The final computed solution for the Lagrangian formulation on a 512×512 mesh (restricted for visualization).

the renormalization penalty method with nested iteration and trust regions has a somewhat shorter overall run time than that of the Lagrange multiplier approach. Moreover, the renormalization approach matches the free energy and unit-length conformance of the Lagrangian formulation, see Tables 4.11 and 4.13. However, while the overall run time and approximate cost of the approach is slightly larger, the accuracy of the Lagrange multiplier formulation is expected to be much higher. For the Lagrange multiplier approach, the l_2-norm of the first-order optimality conditions is 7.386e-13, whereas the same measure for the renormalization penalty method is 1.603e-02. As with the corresponding tilt-twist table, Table 4.14 does not report runs with $\zeta = 10^9$.

In all of the experiments above, the accuracy per unit cost of the Lagrange multiplier method convincingly outperforms that of either of the penalty methods. Moreover, the experimental results imply that nested iteration should be used when considering any of the methods, as it proves to be exceedingly effective at reducing computational costs for all problems and approaches. While trust regions offer very slight improvements in computation time, they readily improve robustness of the penalty method. Due to their limited cost, it would be advantageous to include them for either method. The simple trust-region approach works best for the unmodified

Lagrangian			
Method		Solve Cost	Run Time
No NI	No TR	63	18,861s
NI	No TR	2.864	983s
No NI	TR	10	3,113s
NI	TR	2.779	960s
Renormalization Penalty: $\zeta = 10^5$			
Method		Solve Cost	Run Time
No NI	No TR	35	10,918s
NI	No TR	–	–
No NI	TR	32	9,893s
No NI	TR 2D	34	10,976s
NI	TR	2.133	789s
NI	TR 2D	2.433	901s

Unmodified Penalty $\zeta = 10^5$			
Method		Solve Cost	Run Time
No NI	No TR	169	49,654s
NI	No TR	–	–
No NI	TR	73	21,415s
No NI	TR 2D	75	22,366s
NI	TR	2.805	958s
NI	TR 2D	3.530	1,202s

Table 4.14: Nano-pattern statistics comparison for NI and trust region combinations. The solve cost column displays an approximation of the work in WUs for the corresponding method. The overall time to solution is also presented. Dashes in the columns indicate divergence.

penalty method with stopping tolerances based on the first-order optimality conditions, whereas the 2D-subspace minimization trust regions are most effective for the renormalization penalty method with an energy reduction based stopping tolerance. Though larger penalty weights are generally necessary, the unmodified penalty method offers more consistent error reduction and performance with respect to an increasing weight.

Since the Lagrange multiplier method coupled with nested iteration is the most accurate and efficient approach for enforcing the unit-length constraint, it is exclusively considered in the chapters to follow. The chapter to follow extends the Lagrange multiplier method and theory to include electric effects due to external and internal electric fields.

Chapter 5

Electric Effects

In Chapter 3, a general approach for computing the equilibrium state for $\mathbf{n}$ under free-elastic effects is derived. We apply this methodology to the augmented elastic-electric free energy. As in the free elastic setting, the director equilibrium state corresponds to the configuration which minimizes the system free energy subject to the local constraint that $\mathbf{n}$ is of unit length throughout the sample volume, Ω. As discussed above, liquid crystals strongly interact with externally applied electric fields and are capable of producing internal electric fields due to flexoelectric effects, sometimes also referred to as piezoelectricity in the context of certain materials. Free-energy models were discussed for applied electric fields and flexoelectric effects in Sections 2.2 and 2.3, respectively. Here, we first derive the energy-minimization finite-element approach with Lagrange multipliers for the applied electric field case and then move to the flexoelectric model.

5.1 Applied Electric Fields

Recall that the functional to be minimized, in the presence of an applied electric field, is

$$\begin{aligned}\mathcal{F}_3(\mathbf{n},\phi) = (K_1 - K_2 - K_4)\|\nabla\cdot\mathbf{n}\|_0^2 + K_3\langle\mathbf{Z}\nabla\times\mathbf{n},\nabla\times\mathbf{n}\rangle_0 \\ + (K_2+K_4)\big(\langle\nabla n_1,\frac{\partial\mathbf{n}}{\partial x}\rangle_0 + \langle\nabla n_2,\frac{\partial\mathbf{n}}{\partial y}\rangle_0 + \langle\nabla n_3,\frac{\partial\mathbf{n}}{\partial z}\rangle_0\big) \\ - \epsilon_0\epsilon_\perp\langle\nabla\phi,\nabla\phi\rangle_0 - \epsilon_0\epsilon_a\langle\mathbf{n}\cdot\nabla\phi,\mathbf{n}\cdot\nabla\phi\rangle_0, \qquad (5.1)\end{aligned}$$

where $\mathbf{E} = -\nabla\phi$. As discussed in Section 2.2, using a potential function guarantees that Faraday's law is trivially satisfied.

In the presence of full Dirichlet boundary conditions or a rectangular domain

with mixed Dirichlet and periodic boundary conditions, the functional to be minimized is significantly simplified to

$$\begin{aligned}\mathcal{F}_5(\mathbf{n},\phi) = K_1\|\nabla\cdot\mathbf{n}\|_0^2 + K_3\langle\mathbf{Z}\nabla\times\mathbf{n},\nabla\times\mathbf{n}\rangle_0 \\ - \epsilon_0\epsilon_\perp\langle\nabla\phi,\nabla\phi\rangle_0 - \epsilon_0\epsilon_a\langle\mathbf{n}\cdot\nabla\phi,\mathbf{n}\cdot\nabla\phi\rangle_0,\end{aligned} \tag{5.2}$$

by the application of (2.4). However, the functional still contains nonlinear terms introduced by, for instance, the presence of $\mathbf{Z} = \mathbf{Z}(\mathbf{n})$.

We proceed with the functional in (5.1) in building a framework for minimization under general boundary conditions. However, in the treatment of existence and uniqueness theory, we assume the application of full Dirichlet or mixed Dirichlet and periodic boundary conditions and, therefore, appeal to the simplified form in Equation (5.2).

Let

$$H^{1,g}(\Omega) = \{f \in H^1(\Omega) : B_1(f) = g\},$$

where $H^1(\Omega)$ represents the classical Sobolev space and $B_1(f) = g$ is an appropriate boundary condition expression for ϕ. Using Functional (5.1), the desired minimization becomes

$$\mathbf{n}_0,\phi_0 = \underset{\mathbf{n},\phi\in(\mathcal{S}^2\cap\mathcal{H}^{DC}(\Omega))\times H^{1,g}(\Omega)}{\operatorname{argmin}} \mathcal{F}_3(\mathbf{n},\phi).$$

5.1.1 First-Order Optimality and Newton Linearization

Constructing the Lagrange multiplier formulation in a similar fashion to Section 3.3, the Lagrangian is written

$$\mathcal{L}(\mathbf{n},\phi,\lambda) = \mathcal{F}_3(\mathbf{n},\phi) + \int_\Omega \lambda(\mathbf{x})((\mathbf{n},\mathbf{n}) - 1)\,dV,$$

where $\lambda \in L^2(\Omega)$. In order to minimize (5.1), we compute the Gâteaux derivatives of $\mathcal{L}$ with respect to $\mathbf{n}$, ϕ, and λ in the directions $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$, $\psi \in H^{1,0}(\Omega)$, and $\gamma \in L^2(\Omega)$, respectively. Hence, necessary continuum first-order optimality

conditions are derived as

$$\begin{aligned}
\mathcal{L}_{\mathbf{n}}[\mathbf{v}] &= \frac{\partial}{\partial \mathbf{n}}\mathcal{L}(\mathbf{n},\phi,\lambda)[\mathbf{v}] = 0, && \forall \mathbf{v} \in \mathcal{H}_0^{DC}(\Omega),\\
\mathcal{L}_{\phi}[\psi] &= \frac{\partial}{\partial \phi}\mathcal{L}(\mathbf{n},\phi,\lambda)[\psi] = 0, && \forall \psi \in H^{1,0}(\Omega),\\
\mathcal{L}_{\lambda}[\gamma] &= \frac{\partial}{\partial \lambda}\mathcal{L}(\mathbf{n},\phi,\lambda)[\gamma] = 0, && \forall \gamma \in L^2(\Omega).
\end{aligned}$$

Computing these derivatives yields the variational system,

$$\begin{aligned}
\mathcal{L}_{\mathbf{n}}[\mathbf{v}] &= 2(K_1 - K_2 - K_4)\langle \nabla\cdot\mathbf{n}, \nabla\cdot\mathbf{v}\rangle_0 + 2K_3\langle \mathbf{Z}\nabla\times\mathbf{n}, \nabla\times\mathbf{v}\rangle_0\\
&\quad + 2(K_2 - K_3)\langle \mathbf{n}\cdot\nabla\times\mathbf{n}, \mathbf{v}\cdot\nabla\times\mathbf{n}\rangle_0 + 2(K_2 + K_4)\Big(\langle \nabla n_1, \frac{\partial \mathbf{v}}{\partial x}\rangle_0\\
&\quad + \langle \nabla n_2, \frac{\partial \mathbf{v}}{\partial y}\rangle_0 + \langle \nabla n_3, \frac{\partial \mathbf{v}}{\partial z}\rangle_0\Big) - 2\epsilon_0\epsilon_a\langle \mathbf{n}\cdot\nabla\phi, \mathbf{v}\cdot\nabla\phi\rangle_0\\
&\quad + 2\int_\Omega \lambda(\mathbf{n},\mathbf{v})\, dV = 0, && \forall \mathbf{v} \in \mathcal{H}_0^{DC}(\Omega),\\
\mathcal{L}_{\phi}[\psi] &= -2\epsilon_0\epsilon_\perp\langle \nabla\phi, \nabla\psi\rangle_0 - 2\epsilon_0\epsilon_a\langle \mathbf{n}\cdot\nabla\phi, \mathbf{n}\cdot\nabla\psi\rangle_0 = 0, && \forall \psi \in H^{1,0}(\Omega),\\
\mathcal{L}_{\lambda}[\gamma] &= \int_\Omega \gamma((\mathbf{n},\mathbf{n}) - 1)\, dV = 0, && \forall \gamma \in L^2(\Omega).
\end{aligned}$$

Note that $\mathcal{L}_{\phi}[\psi] = 0$, in the system above, is, in fact, the weak form of Gauss' law. Therefore, at the functional minimum both Gauss' and Faraday's laws are satisfied.

The system above is nonlinear; therefore, Newton iterations are again employed by computing a generalized first-order Taylor series expansion. Let $\mathbf{n}_k$, ϕ_k, and λ_k be the current approximations for $\mathbf{n}$, ϕ, and λ, respectively. Additionally, let $\delta\mathbf{n} = \mathbf{n}_{k+1} - \mathbf{n}_k$, $\delta\phi = \phi_{k+1} - \phi_k$, and $\delta\lambda = \lambda_{k+1} - \lambda_k$ be updates to the current approximations that we seek to compute. Then, the Newton iterations are denoted

$$\begin{bmatrix} \mathcal{L}_{\mathbf{nn}} & \mathcal{L}_{\mathbf{n}\phi} & \mathcal{L}_{\mathbf{n}\lambda} \\ \mathcal{L}_{\phi\mathbf{n}} & \mathcal{L}_{\phi\phi} & \mathcal{L}_{\phi\lambda} \\ \mathcal{L}_{\lambda\mathbf{n}} & \mathcal{L}_{\lambda\phi} & \mathcal{L}_{\lambda\lambda} \end{bmatrix} \begin{bmatrix} \delta\mathbf{n} \\ \delta\phi \\ \delta\lambda \end{bmatrix} = - \begin{bmatrix} \mathcal{L}_{\mathbf{n}} \\ \mathcal{L}_{\phi} \\ \mathcal{L}_{\lambda} \end{bmatrix}, \tag{5.3}$$

where each of the system components are evaluated at $\mathbf{n}_k$, ϕ_k, and λ_k. As above, the matrix-vector multiplication indicates the direction that the derivatives in the

Hessian are taken. For instance, $\mathcal{L}_{\phi\mathbf{n}}[\psi]\cdot\delta\mathbf{n} = \frac{\partial}{\partial\mathbf{n}}\left(\mathcal{L}_\phi(\mathbf{n}_k,\lambda_k)[\psi]\right)[\delta\mathbf{n}]$, where the partials indicate Gâteaux derivatives in the respective variables. Note that $\mathcal{L}_{\lambda\lambda} = \mathcal{L}_{\lambda\phi} = \mathcal{L}_{\phi\lambda} = 0$. Hence, the Hessian in (5.3) simplifies to a 3×3 saddle-point system given by

$$\begin{bmatrix} \mathcal{L}_{\mathbf{nn}} & \mathcal{L}_{\mathbf{n}\phi} & \mathcal{L}_{\mathbf{n}\lambda} \\ \mathcal{L}_{\phi\mathbf{n}} & \mathcal{L}_{\phi\phi} & \mathbf{0} \\ \mathcal{L}_{\lambda\mathbf{n}} & \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \delta\mathbf{n} \\ \delta\phi \\ \delta\lambda \end{bmatrix} = -\begin{bmatrix} \mathcal{L}_{\mathbf{n}} \\ \mathcal{L}_{\phi} \\ \mathcal{L}_{\lambda} \end{bmatrix}. \tag{5.4}$$

This block structure is common in constrained problems involving electric and magnetic fields. For instance, certain systems arising in the context of magnetohydrodynamics (MHD) have similar saddle-point structures [2, 28, 98]. In Chapter 6, multigrid approaches tailored to the 3×3 block matrices arising from a finite-element discretization of the Hessian in (5.4) are discussed and implemented.

Considering the remaining six components of the Hessian, the derivatives involving λ are

$$\mathcal{L}_{\lambda\mathbf{n}}[\gamma]\cdot\delta\mathbf{n} = 2\int_\Omega \gamma(\mathbf{n}_k,\delta\mathbf{n})\,dV, \qquad \mathcal{L}_{\mathbf{n}\lambda}[\mathbf{v}]\cdot\delta\lambda = 2\int_\Omega \delta\lambda(\mathbf{n}_k,\mathbf{v})\,dV.$$

The second-order terms involving ϕ are

$$\begin{aligned}
\mathcal{L}_{\phi\phi}[\psi]\cdot\delta\phi &= -2\epsilon_0\epsilon_\perp\langle\nabla\delta\phi,\nabla\psi\rangle_0 - 2\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\delta\phi,\mathbf{n}_k\cdot\nabla\psi\rangle_0,\\
\mathcal{L}_{\phi\mathbf{n}}[\psi]\cdot\delta\mathbf{n} &= -2\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\phi_k,\delta\mathbf{n}\cdot\nabla\psi\rangle_0 - 2\epsilon_0\epsilon_a\langle\delta\mathbf{n}\cdot\nabla\phi_k,\mathbf{n}_k\cdot\nabla\psi\rangle_0,\\
\mathcal{L}_{\mathbf{n}\phi}[\mathbf{v}]\cdot\delta\phi &= -2\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\phi_k,\mathbf{v}\cdot\nabla\delta\phi\rangle_0 - 2\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\delta\phi,\mathbf{v}\cdot\nabla\phi_k\rangle_0.
\end{aligned}$$

Finally, the second-order derivative with respect to $\mathbf{n}$ is

$$\begin{aligned}
\mathcal{L}_{\mathbf{nn}}[\mathbf{v}]\cdot\delta\mathbf{n} = {}& 2(K_1-K_2-K_4)\langle\nabla\cdot\delta\mathbf{n},\nabla\cdot\mathbf{v}\rangle_0 + 2K_3\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\delta\mathbf{n},\nabla\times\mathbf{v}\rangle_0\\
&+ 2(K_2-K_3)\Big(\langle\delta\mathbf{n}\cdot\nabla\times\mathbf{v},\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0\\
&+ \langle\mathbf{n}_k\cdot\nabla\times\mathbf{v},\delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k\rangle_0 + \langle\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\delta\mathbf{n}\rangle_0\\
&+ \langle\mathbf{n}_k\cdot\nabla\times\delta\mathbf{n},\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 + \langle\delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big)
\end{aligned}$$

$$+ 2(K_2 + K_4)\big(\langle \nabla \delta n_1, \frac{\partial \mathbf{v}}{\partial x}\rangle_0 + \langle \nabla \delta n_2, \frac{\partial \mathbf{v}}{\partial y}\rangle_0 + \langle \nabla \delta n_3, \frac{\partial \mathbf{v}}{\partial z}\rangle_0\big)$$
$$- 2\epsilon_0\epsilon_a\langle \delta\mathbf{n}\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0 + 2\int_\Omega \lambda_k(\delta\mathbf{n}, \mathbf{v})\, dV.$$

Completing (5.4) with the above Hessian computations yields a linearized variational system. For these iterations, we compute $\delta\mathbf{n}$, $\delta\phi$, and $\delta\lambda$ satisfying (5.4) for all $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$, $\psi \in H^{1,0}(\Omega)$, and $\gamma \in L^2(\Omega)$ with the current approximations $\mathbf{n}_k$, ϕ_k, and λ_k. If we are considering a system with full or mixed Dirichlet boundary conditions, as described above, we eliminate the $(K_2 + K_4)$ terms from (5.4). This produces a simplified, but non-trivial, linearization. Both the full and simplified electric linearized variational systems are fully expanded in Appendix A.2.

5.1.2 Well-Posedness of the Discrete Systems

Performing the outlined Newton iterations necessitates solving the above linearized systems for the update functions $\delta\mathbf{n}$, $\delta\phi$, and $\delta\lambda$. Finite elements are used to numerically approximate these updates as $\delta\mathbf{n}_h$, $\delta\phi_h$, and $\delta\lambda_h$. Throughout this section, we assume that full Dirichlet boundary conditions are enforced for $\mathbf{n}$ and ϕ. However, the theory to follow is also applicable for a rectangular domain with mixed Dirichlet and periodic boundary conditions. Such a domain is considered for the numerical experiments presented in the next chapter. Furthermore, the developed theory exclusively concerns discrete forms. Hence, we forgo the h notation.

We write the bilinear form defined by $-\mathcal{L}_{\phi\phi}[\psi]\cdot\delta\phi$ as $c(\delta\phi, \psi) = \epsilon_0\epsilon_\perp\langle\nabla\delta\phi, \nabla\psi\rangle_0 + \epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\delta\phi, \mathbf{n}_k\cdot\nabla\psi\rangle_0$ and the form associated with $\mathcal{L}_{\lambda\mathbf{n}}[\gamma]\cdot\delta\mathbf{n}$ as $b(\delta\mathbf{n}, \gamma)$. Further, we decompose the bilinear form defined by $\mathcal{L}_{\mathbf{nn}}[\mathbf{v}]\cdot\delta\mathbf{n}$ into a free-elastic term, $\tilde{a}(\delta\mathbf{n}, \mathbf{v})$, and an electric component as

$$a(\delta\mathbf{n}, \mathbf{v}) = \tilde{a}(\delta\mathbf{n}, \mathbf{v}) - \epsilon_0\epsilon_a\langle\delta\mathbf{n}\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0.$$

Note then that $\tilde{a}(\delta\mathbf{n}, \mathbf{v})$ represents the discrete version of the free-elastic bilinear form in (3.12).

Lemma 5.1.1 *Let Ω be a connected, open, bounded domain. If $\epsilon_a \geq 0$, then $c(\delta\phi, \psi)$ is a coercive bilinear form. For $\epsilon_a < 0$, if $|\mathbf{n}_k|^2 \leq \beta < \epsilon_\perp/|\epsilon_a|$, then $c(\delta\phi, \psi)$ is a coercive bilinear form.*

Proof: The proof is split into two cases.

Case 1. $\epsilon_a \geq 0$.

Note that $\delta\phi, \psi \in H^{1,0}(\Omega)$, with homogeneous Dirichlet boundary conditions. By the classical Poincaré-Friedrichs' inequality, there exists a $C_e > 0$ such that for all $\xi \in H_0^1(\Omega)$, $\|\xi\|_0^2 \leq C_e \|\nabla\xi\|_0^2$. Therefore,

$$\|\xi\|_1^2 \leq (C_e + 1)\|\nabla\xi\|_0^2.$$

This implies that, for $\xi \neq 0$,

$$\begin{aligned} c(\xi, \xi) &= \epsilon_0\epsilon_\perp \langle \nabla\xi, \nabla\xi\rangle_0 + \epsilon_0\epsilon_a \langle \mathbf{n}_k \cdot \nabla\xi, \mathbf{n}_k \cdot \nabla\xi\rangle_0 \\ &\geq \frac{\epsilon_0\epsilon_\perp}{C_e + 1}\|\xi\|_1^2 > 0. \end{aligned}$$

Case 2. $\epsilon_a < 0$.

Observe that pointwise,

$$(\mathbf{n}_k \cdot \nabla\xi)^2 \leq |\mathbf{n}_k|^2 |\nabla\xi|^2 \leq \beta |\nabla\xi|^2.$$

This implies that $\langle \mathbf{n}_k \cdot \nabla\xi, \mathbf{n}_k \cdot \nabla\xi\rangle_0 \leq \beta \langle \nabla\xi, \nabla\xi\rangle_0$. Therefore,

$$c(\xi, \xi) \geq \epsilon_0(\epsilon_\perp - \beta|\epsilon_a|)\langle \nabla\xi, \nabla\xi\rangle_0.$$

Recall that $\epsilon_\perp > 0$. Therefore, $\beta < \epsilon_\perp/|\epsilon_a|$ implies that $\epsilon_\perp - \beta|\epsilon_a| > 0$. Thus, again applying the Poincaré-Friedrichs' inequality above for $\xi \neq 0$,

$$c(\xi, \xi) \geq \frac{\epsilon_0(\epsilon_\perp - \beta|\epsilon_a|)}{C_e + 1}\|\xi\|_1^2 > 0.$$

In either case, $c(\cdot, \cdot)$ is a coercive bilinear form. □

There are a number of discretization space triples commonly used to discretize systems such as the one defined in (5.4), including equal order or mixed finite elements. Discretizing the Hessian in (5.4) with finite elements leads to the 3×3 block matrix

$$M_e = \begin{bmatrix} A & B_1 & B_2 \\ B_1^T & -\tilde{C} & \mathbf{0} \\ B_2^T & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{5.5}$$

Lemma 5.1.2 *Under the assumptions in Lemma 5.1.1, if the bilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$, defined above, are coercive and weakly coercive, respectively, on the relevant discrete spaces, the matrix in* (5.5) *is invertible.*

Proof: Denoting $B = \begin{bmatrix} B_1 & B_2 \end{bmatrix}$ (where B_2 is associated with $b(\cdot,\cdot)$), and $C = \begin{bmatrix} \tilde{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, the matrix in (5.5) is written as

$$\begin{bmatrix} A & B \\ B^T & -C \end{bmatrix}.$$

By assumption, $a(\cdot,\cdot)$ is coercive, and it is clearly symmetric. Therefore, the associated discretization block, A, is symmetric and positive definite. By Lemma 5.1.1, $\tilde{C}$ is symmetric and positive definite, and, therefore, $-C$ is symmetric and negative semi-definite. Thus, by [10, Theorem 3.1], if $\ker C \cap \ker B = \{\mathbf{0}\}$, then the matrix in (5.5) is invertible. Observe that

$$\begin{bmatrix} \tilde{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y} \\ \mathbf{z} \end{bmatrix} = \begin{bmatrix} \tilde{C}\mathbf{y} \\ \mathbf{0} \end{bmatrix} = \mathbf{0}$$

if and only if $\mathbf{y} = \mathbf{0}$. Then, if $\begin{bmatrix} \mathbf{y} & \mathbf{z} \end{bmatrix}^T \in \ker C \cap \ker B$, $\mathbf{y} = \mathbf{0}$. However, note that

$$\begin{bmatrix} B_1 & B_2 \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{z} \end{bmatrix} = B_2\mathbf{z}.$$

Since $b(\cdot,\cdot)$ is weakly coercive, $B_2\mathbf{z} = \mathbf{0}$ if and only if $\mathbf{z} = \mathbf{0}$. So $\ker C \cap \ker B = \{\mathbf{0}\}$.□

Let $C_\phi = \sup_{\mathbf{x}\in\Omega} |\nabla\phi_k|$. Using the spaces and assumptions established in Section 3.5, the following theorem is formulated.

Theorem 5.1.3 *Under the assumptions of Lemmas 3.5.8 or 3.5.9, for $\kappa = 1$ or κ satisfying the small data assumptions in Lemma 3.5.9, respectively, let $\alpha_0 > 0$ be such that $\tilde{a}(\mathbf{v},\mathbf{v}) \geq \alpha_0 \|\mathbf{v}\|_{DC}^2$. With the assumptions of Lemma 3.5.14 and those of Lemma 5.1.1, if $\epsilon_a \leq 0$ or $(\alpha_0 - \epsilon_0\epsilon_a C_\phi^2) > 0$, then the matrix defined by (5.5) is invertible.*

Proof: If $\kappa = 1$, Lemma 3.5.8 implies that such an $\alpha_0 > 0$ exists. Similarly, if κ satisfies the small data assumptions of Lemma 3.5.9, then such an $\alpha_0 > 0$ also exists. If $\epsilon_a \leq 0$, clearly this implies that $a(\cdot,\cdot)$ is coercive. For $\epsilon_a > 0$, note that

$$\begin{aligned}\langle \mathbf{v}\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0 = \int_\Omega (\mathbf{v}\cdot\nabla\phi_k)^2\, dV &\leq \int_\Omega |\mathbf{v}|^2|\nabla\phi_k|^2\, dV \\ &\leq C_\phi^2 \int_\Omega |\mathbf{v}|^2\, dV \\ &\leq C_\phi^2 \|\mathbf{v}\|_{DC}^2. \end{aligned} \tag{5.6}$$

Hence,

$$|\epsilon_0\epsilon_a\langle \mathbf{v}\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0| \leq \epsilon_0\epsilon_a C_\phi^2 \|\mathbf{v}\|_{DC}^2. \tag{5.7}$$

Therefore,

$$\begin{aligned} a(\mathbf{v},\mathbf{v}) &\geq \alpha_0\|\mathbf{v}\|_{DC}^2 - \epsilon_0\epsilon_a C_\phi^2\|\mathbf{v}\|_{DC}^2 \\ &= (\alpha_0 - \epsilon_0\epsilon_a C_\phi^2)\|\mathbf{v}\|_{DC}^2. \end{aligned}$$

Thus, if $(\alpha_0 - \epsilon_0\epsilon_a C_\phi^2) > 0$, $a(\cdot,\cdot)$ is coercive.

Finally, Lemma 3.5.14 asserts that $b(\cdot,\cdot)$ is weakly coercive. Hence, Lemma 5.1.2 implies that M_e, as defined in (5.5), is invertible. □

Theorem 5.1.3 implies that no additional inf-sup condition for ϕ is necessary to guarantee uniqueness of the solution to the system in (5.4). Moreover, the discretization space for ϕ may be freely chosen without concern for stability.

5.2 Flexoelectricity

The flexoelectric effect in liquid crystals continues to generate new and innovative research [30, 62, 74]. We now derive and analyze the variational systems arising within our energy-minimization framework in the context flexoelectric effects. The full form of the free-energy functional for flexoelectric free energy discussed in Section 2.3 is

$$\begin{aligned}\mathcal{F}_4(\mathbf{n},\phi) = {} & (K_1 - K_2 - K_4)\|\nabla\cdot\mathbf{n}\|_0^2 + K_3\langle \mathbf{Z}\nabla\times\mathbf{n}, \nabla\times\mathbf{n}\rangle_0 \\ & + (K_2+K_4)\Big(\langle\nabla n_1, \frac{\partial\mathbf{n}}{\partial x}\rangle_0 + \langle\nabla n_2, \frac{\partial\mathbf{n}}{\partial y}\rangle_0 + \langle\nabla n_3, \frac{\partial\mathbf{n}}{\partial z}\rangle_0\Big) \\ & - \epsilon_0\epsilon_\perp\langle\nabla\phi,\nabla\phi\rangle_0 - \epsilon_0\epsilon_a\langle\mathbf{n}\cdot\nabla\phi, \mathbf{n}\cdot\nabla\phi\rangle_0 + 2e_s\langle\nabla\cdot\mathbf{n}, \mathbf{n}\cdot\nabla\phi\rangle_0 \\ & + 2e_b\langle\mathbf{n}\times\nabla\times\mathbf{n}, \nabla\phi\rangle_0, \end{aligned} \tag{5.8}$$

where, again, $\mathbf{E} = -\nabla\phi$ to satisfy Faraday's law. In the presence of appropriate boundary conditions, using (2.4) implies that this functional simplifies to

$$\begin{aligned}\mathcal{F}_6(\mathbf{n},\phi) = {} & K_1\|\nabla\cdot\mathbf{n}\|_0^2 + K_3\langle\mathbf{Z}\nabla\times\mathbf{n},\nabla\times\mathbf{n}\rangle_0 - \epsilon_0\epsilon_\perp\langle\nabla\phi,\nabla\phi\rangle_0 \\ & - \epsilon_0\epsilon_a\langle\mathbf{n}\cdot\nabla\phi,\mathbf{n}\cdot\nabla\phi\rangle_0 + 2e_s\langle\nabla\cdot\mathbf{n},\mathbf{n}\cdot\nabla\phi\rangle_0 + 2e_b\langle\mathbf{n}\times\nabla\times\mathbf{n},\nabla\phi\rangle_0. \end{aligned} \tag{5.9}$$

As in the previous cases, we use Functional (5.8) to build a framework for minimization under general boundary conditions, while in the treatment of existence and uniqueness theory, we assume the application of full Dirichlet or mixed Dirichlet and periodic boundary conditions and utilize the functional in (5.9).

5.2.1 First-Order Optimality and Newton Linearization

The flexoelectric Lagrangian is constructed as

$$\hat{\mathcal{L}}(\mathbf{n},\phi,\lambda) = \mathcal{F}_4(\mathbf{n},\phi) + \int_\Omega \lambda((\mathbf{n},\mathbf{n}) - 1)\, dV. \tag{5.10}$$

As in the sections above, in order to minimize (5.10), Gâteaux derivatives for $\hat{\mathcal{L}}(\mathbf{n},\phi,\lambda)$ must be computed. Derivation of this variational system is identical to

that of the applied electric case in Section 5.1.1, with the exception of the derivative calculations for the additional flexoelectric energy terms. Therefore, the flexoelectric variational system is written compactly as

$$\begin{aligned}
\hat{\mathcal{L}}_{\mathbf{n}}[\mathbf{v}] &= \mathcal{L}_{\mathbf{n}}[\mathbf{v}] + 2e_s\big(\langle \nabla\cdot\mathbf{n}, \mathbf{v}\cdot\nabla\phi\rangle_0 + \langle \nabla\cdot\mathbf{v}, \mathbf{n}\cdot\nabla\phi\rangle_0\big) \\
&\quad + 2e_b\big(\langle \mathbf{n}\times\nabla\times\mathbf{v}, \nabla\phi\rangle_0 + \langle \mathbf{v}\times\nabla\times\mathbf{n}, \nabla\phi\rangle_0\big) = 0, && \forall \mathbf{v}\in\mathcal{H}_0^{DC}(\Omega), \\
\hat{\mathcal{L}}_{\phi}[\psi] &= \mathcal{L}_{\phi}[\psi] + 2e_s\langle \nabla\cdot\mathbf{n}, \mathbf{n}\cdot\nabla\psi\rangle_0 + 2e_b\langle \mathbf{n}\times\nabla\times\mathbf{n}, \nabla\psi\rangle_0 = 0, && \forall \psi\in H^{1,0}(\Omega), \\
\hat{\mathcal{L}}_{\lambda}[\gamma] &= \mathcal{L}_{\lambda}[\gamma] = 0, && \forall \gamma\in L^2(\Omega).
\end{aligned}$$

Constructing the Newton iterations to address the nonlinearities, as above, yields a Newton linearization system with the same saddle-point structure as the electric field case. These Newton iterations are denoted

$$\begin{bmatrix} \hat{\mathcal{L}}_{\mathbf{nn}} & \hat{\mathcal{L}}_{\mathbf{n}\phi} & \hat{\mathcal{L}}_{\mathbf{n}\lambda} \\ \hat{\mathcal{L}}_{\phi\mathbf{n}} & \hat{\mathcal{L}}_{\phi\phi} & \mathbf{0} \\ \hat{\mathcal{L}}_{\lambda\mathbf{n}} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \delta\mathbf{n} \\ \delta\phi \\ \delta\lambda \end{bmatrix} = - \begin{bmatrix} \hat{\mathcal{L}}_{\mathbf{n}} \\ \hat{\mathcal{L}}_{\phi} \\ \hat{\mathcal{L}}_{\lambda} \end{bmatrix}. \tag{5.11}$$

Since the flexoelectric energy terms are first-order with respect to ϕ and do not depend on λ, many of the second-order derivatives are the same as the simple electric case. On the other hand, the mixed partial derivatives involving ϕ contain additional terms,

$$\begin{aligned}
\hat{\mathcal{L}}_{\phi\mathbf{n}}[\psi]\cdot\delta\mathbf{n} &= \mathcal{L}_{\phi\mathbf{n}}[\psi]\cdot\delta\mathbf{n} + 2e_s\big(\langle \nabla\cdot\delta\mathbf{n}, \mathbf{n}_k\cdot\nabla\psi\rangle_0 + \langle \nabla\cdot\mathbf{n}_k, \delta\mathbf{n}\cdot\nabla\psi\rangle_0\big) \\
&\quad + 2e_b\big(\langle \mathbf{n}_k\times\nabla\times\delta\mathbf{n}, \nabla\psi\rangle_0 + \langle \delta\mathbf{n}\times\nabla\times\mathbf{n}_k, \nabla\psi\rangle_0\big), \\
\hat{\mathcal{L}}_{\mathbf{n}\phi}[\mathbf{v}]\cdot\delta\phi &= \mathcal{L}_{\mathbf{n}\phi}[\mathbf{v}]\cdot\delta\phi + 2e_s\big(\langle \nabla\cdot\mathbf{n}_k, \mathbf{v}\cdot\nabla\delta\phi\rangle_0 + \langle \nabla\cdot\mathbf{v}, \mathbf{n}_k\cdot\nabla\delta\phi\rangle_0\big) \\
&\quad + 2e_b\big(\langle \mathbf{n}_k\times\nabla\times\mathbf{v}, \nabla\delta\phi\rangle_0 + \langle \mathbf{v}\times\nabla\times\mathbf{n}_k, \nabla\delta\phi\rangle_0\big).
\end{aligned}$$

Finally, the second-order derivative with respect to $\mathbf{n}$ also contains additional terms,

$$\begin{aligned}
\hat{\mathcal{L}}_{\mathbf{nn}}[\mathbf{v}]\cdot\delta\mathbf{n} &= \mathcal{L}_{\mathbf{nn}}[\mathbf{v}]\cdot\delta\mathbf{n} + 2e_s\big(\langle \nabla\cdot\delta\mathbf{n}, \mathbf{v}\cdot\nabla\phi_k\rangle_0 + \langle \nabla\cdot\mathbf{v}, \delta\mathbf{n}\cdot\nabla\phi_k\rangle_0\big) \\
&\quad + 2e_b\big(\langle \delta\mathbf{n}\times\nabla\times\mathbf{v}, \nabla\phi_k\rangle_0 + \langle \mathbf{v}\times\nabla\times\delta\mathbf{n}, \nabla\phi_k\rangle_0\big).
\end{aligned}$$

Note that $\mathcal{L}_{\phi\mathbf{n}}[\psi]\cdot\delta\mathbf{n}$, $\mathcal{L}_{\mathbf{n}\phi}[\mathbf{v}]\cdot\delta\phi$, and $\mathcal{L}_{\mathbf{nn}}[\mathbf{v}]\cdot\delta\mathbf{n}$ are second-order derivatives from the applied electric field Hessian computed in Section 5.1.1.

Completing the system in (5.11) with the above Hessian and right-hand-side computations yields the flexoelectric linearized variational system. For these iterations, we again compute $\delta\mathbf{n}$, $\delta\phi$, and $\delta\lambda$ satisfying (5.11) for all $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$, $\psi \in H^{1,0}(\Omega)$, and $\gamma \in L^2(\Omega)$ with the current approximations $\mathbf{n}_k$, ϕ_k, and λ_k. If we are considering a system with full or mixed Dirichlet boundary conditions, as described above, we eliminate the $(K_2 + K_4)$ terms from (5.11). This produces a moderately simplified flexoelectric linearization. Appendix A.3 expands both the full and simplified flexoelectric linearized variational systems.

5.2.2 Well-Posedness of the Discrete Systems

As with the simple electric linearization, finite elements are used to numerically approximate the updates as $\delta\mathbf{n}_h$, $\delta\phi_h$, and $\delta\lambda_h$. For simplicity, throughout this section we assume that full Dirichlet boundary conditions are enforced for $\mathbf{n}$ and ϕ. However, the theory is, as above, also applicable for a rectangular domain with mixed Dirichlet and periodic boundary conditions. As in the simple electric case, we define bilinear forms to represent relevant components of the computed Hessian. The bilinear forms associated with $-\hat{\mathcal{L}}_{\phi\phi}[\psi]\cdot\delta\phi$ and $\hat{\mathcal{L}}_{\lambda\mathbf{n}}[\gamma]\cdot\delta\mathbf{n}$ are denoted $c(\delta\phi,\psi)$ and $b(\delta\mathbf{n},\gamma)$, respectively, and are identical to the corresponding components of the simple electric case above. We again decompose the bilinear form defined by $\hat{\mathcal{L}}_{\mathbf{nn}}[\mathbf{v}]\cdot\delta\mathbf{n}$ into a free elastic term, $\tilde{a}(\delta\mathbf{n},\mathbf{v})$, and a flexoelectric component as

$$\begin{aligned} a(\delta\mathbf{n},\mathbf{v}) = \tilde{a}(\delta\mathbf{n},\mathbf{v}) &- \epsilon_0\epsilon_a\langle\delta\mathbf{n}\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0 \\ &+ e_s\big(\langle\nabla\cdot\delta\mathbf{n}, \mathbf{v}\cdot\nabla\phi_k\rangle_0 + \langle\nabla\cdot\mathbf{v}, \delta\mathbf{n}\cdot\nabla\phi_k\rangle_0\big) \\ &+ e_b\big(\langle\delta\mathbf{n}\times\nabla\times\mathbf{v}, \nabla\phi_k\rangle_0 + \langle\mathbf{v}\times\nabla\times\delta\mathbf{n}, \nabla\phi_k\rangle_0\big). \end{aligned}$$

Recalling that $C_\phi = \sup_{\mathbf{x}\in\Omega}|\nabla\phi_k|$, we formulate the following lemma.

Lemma 5.2.1 *Under the assumptions of Lemma 3.5.8 or 3.5.9, let $\alpha_0 > 0$ be such that $\tilde{a}(\mathbf{v},\mathbf{v}) \geq \alpha_0\|\mathbf{v}\|_{DC}^2$. If $\epsilon_a \leq 0$ and $\alpha_0 > 2C_\phi(|e_b| + |e_s|)$ or $\epsilon_a > 0$ and $\alpha_0 >$*

$\epsilon_0\epsilon_a C_\phi^2 + 2C_\phi(|e_b| + |e_s|)$, *then there exists an* $\alpha_1 > 0$ *such that* $a(\mathbf{v},\mathbf{v}) \geq \alpha_1 \|\mathbf{v}\|_{DC}^2$.

Proof: The proof is split into two cases.

Case 1. $\epsilon_a \leq 0$.

Since $\epsilon_0 > 0$ and $\langle \mathbf{v}\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0$ is clearly positive definite,

$$\tilde{a}(\mathbf{v},\mathbf{v}) - \epsilon_0\epsilon_a\langle \mathbf{v}\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0 \geq \alpha_0\|\mathbf{v}\|_{DC}^2. \tag{5.12}$$

Note that

$$\begin{aligned} |2e_s\langle \nabla\cdot\mathbf{v}, \mathbf{v}\cdot\nabla\phi_k\rangle_0| &\leq 2|e_s|\|\nabla\cdot\mathbf{v}\|_0\|\mathbf{v}\cdot\nabla\phi_k\|_0 \\ &\leq 2|e_s|\|\mathbf{v}\|_{DC}\|\mathbf{v}\cdot\nabla\phi_k\|_0. \end{aligned}$$

Furthermore, from (5.6),

$$\|\mathbf{v}\cdot\nabla\phi_k\|_0^2 \leq C_\phi^2\|\mathbf{v}\|_{DC}^2.$$

Hence,

$$|2e_s\langle \nabla\cdot\mathbf{v}, \mathbf{v}\cdot\nabla\phi_k\rangle_0| \leq 2C_\phi|e_s|\|\mathbf{v}\|_{DC}^2. \tag{5.13}$$

Bounding the second relevant term,

$$\begin{aligned} |2e_b\langle \mathbf{v}\times\nabla\times\mathbf{v}, \nabla\phi_k\rangle_0| &\leq 2|e_b||\langle \mathbf{v}, (\nabla\times\mathbf{v})\times\nabla\phi_k\rangle_0| \\ &\leq 2|e_b|\|\mathbf{v}\|_0\|(\nabla\times\mathbf{v})\times\nabla\phi_k\|_0. \end{aligned}$$

Pointwise,

$$|(\nabla\times\mathbf{v})\times\nabla\phi_k|^2 \leq |\nabla\times\mathbf{v}|^2|\nabla\phi_k|^2.$$

Therefore,

$$\|(\nabla\times\mathbf{v})\times\nabla\phi_k\|_0^2 = \int_\Omega |(\nabla\times\mathbf{v})\times\nabla\phi_k|^2\,dV \leq \int_\Omega |\nabla\times\mathbf{v}|^2|\nabla\phi_k|^2\,dV$$

$$\leq C_\phi^2 \int_\Omega |\nabla \times \mathbf{v}|^2 \, dV$$
$$\leq C_\phi^2 \|\nabla \times \mathbf{v}\|_0^2 \leq C_\phi^2 \|\mathbf{v}\|_{DC}^2.$$

Thus,

$$|2e_b\langle \mathbf{v} \times \nabla \times \mathbf{v}, \nabla\phi_k\rangle_0| \leq 2C_\phi|e_b|\|\mathbf{v}\|_0\|\nabla \times \mathbf{v}\|_0 \leq 2C_\phi|e_b|\|\mathbf{v}\|_{DC}^2. \tag{5.14}$$

Gathering the bounds in (5.13)-(5.14),

$$\begin{aligned} a(\mathbf{v},\mathbf{v}) &\geq \alpha_0\|\mathbf{v}\|_{DC}^2 - 2|e_b|C_\phi\|\mathbf{v}\|_{DC}^2 - 2|e_s|C_\phi\|\mathbf{v}\|_{DC}^2 \\ &= \big(\alpha_0 - 2C_\phi(|e_b| + |e_s|)\big)\|\mathbf{v}\|_{DC}^2. \end{aligned}$$

Then, set $\alpha_1 = \alpha_0 - 2C_\phi(|e_b| + |e_s|) > 0$.

Case 2. $\epsilon_a > 0$.

In this case the additional term, $\langle \mathbf{v}\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0$, is important. Recall, from (5.7), that

$$|\epsilon_0\epsilon_a\langle \mathbf{v}\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0| \leq \epsilon_0\epsilon_a C_\phi^2\|\mathbf{v}\|_{DC}^2. \tag{5.15}$$

Employing the bounds in (5.13)-(5.15),

$$\begin{aligned} a(\mathbf{v},\mathbf{v}) &\geq \alpha_0\|\mathbf{v}\|_{DC}^2 - \epsilon_0\epsilon_a C_\phi^2\|\mathbf{v}\|_{DC}^2 - 2C_\phi(|e_b| + |e_s|)\|\mathbf{v}\|_{DC}^2 \\ &= \big(\alpha_0 - (\epsilon_0\epsilon_a C_\phi^2 + 2C_\phi(|e_b| + |e_s|))\big)\|\mathbf{v}\|_{DC}^2. \end{aligned}$$

Thus, let $\alpha_1 = \alpha_0 - \big(\epsilon_0\epsilon_a C_\phi^2 + 2C_\phi(|e_b| + |e_s|)\big) > 0$. □

When discretizing the flexoelectric linearization, the 3×3 saddle-point block structure,

$$M_f = \begin{bmatrix} \bar{A} & \bar{B}_1 & B_2 \\ \bar{B}_1^T & -\tilde{C} & \mathbf{0} \\ B_2^T & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{5.16}$$

described in (5.5) resurfaces. Blocks B_2 and $\tilde{C}$ are identical to those in (5.5) as they are discretizations of the same bilinear forms in the simple electric case. Again, making use of the discretization spaces defined in (3.37) and (3.38) above, the following theorem holds.

Theorem 5.2.2 *Under the assumptions of Lemma 3.5.14 and Lemmas 5.1.1 and 5.2.1, M_f is invertible.*

Proof: Lemma 3.5.14 implies that the bilinear form $b(\delta\mathbf{n}, \gamma)$, associated with B_2, is weakly coercive and Lemma 5.2.1 implies that $a(\delta\mathbf{n}, \mathbf{v})$ is coercive. Therefore, Lemma 5.1.2 implies that M_f is invertible. □

Therefore, as in the simple electric case above, Theorem 5.2.2 implies that no additional inf-sup condition for ϕ is necessary to guarantee uniqueness of the solution to the system in (5.11), and the discretization space for ϕ may be freely chosen without concern for stability.

We postpone discussion of numerical implementation and results for the electrically augmented problems discussed above until Section 6.3 in Chapter 6. Chapter 6 begins by discussing the construction and implementation of a number of multigrid relaxation techniques tailored to the electrically-coupled linear systems described in (5.5) and (5.16). Such methods are incorporated into the algorithms designed to carry out the energy minimization in the presence of electric fields.

Chapter 6

Multigrid

To efficiently solve large-scale linear systems, geometric multigrid methods utilize complementary techniques consisting of fine-grid relaxation schemes to reduce highly oscillatory error and coarse-grid corrections to eliminate smooth error modes. Relaxation techniques are iterative methods designed to repeatedly reduce solution error each time they are applied. The Jacobi [65] and Gauss-Seidel methods [112], as well as the many methods branching from these two approaches, are widely applied in multigrid frameworks as iterative relaxation schemes. These schemes aggressively reduce highly oscillatory error. Figure 6.1 depicts error in the solution approximation for the linear system arising from a finite-element discretization of the Laplace equation. The error shown in Figure 6.1a, after 5 relaxation iterations of Red-Black Gauss-Seidel, remains relatively oscillatory while the error in Figure 6.1b, after 30 iterations, is quite smooth. As the error becomes smoother, however, convergence slows for many stationary iterative methods [120].

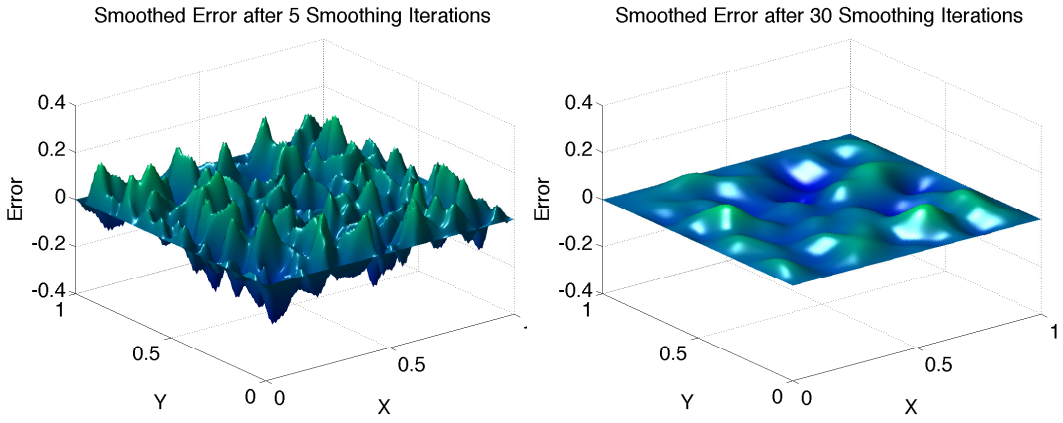


Figure 6.1: The solution error resulting from (a) 5 and (b) 30 iterations of Red-Black Gauss-Seidel on a finite-element discretization of the Laplace equation.

To alleviate this stagnating convergence, the problem is transferred to a coarser grid where the smooth errors appear more oscillatory. Relaxation is then performed

on the coarser problem in order to compute a correction to the current approximation on the finer grid. This process may be performed recursively, forming the basis of a multigrid framework. These recursive coarsening and correction procedures take a number of forms. Figure 6.2 depicts a standard V-cycle approach, which is used in our multigrid implementations to be discussed below, and Figure 6.3 outlines a typical F-cycle. For more thorough overviews of multigrid methods, see [17, 120]

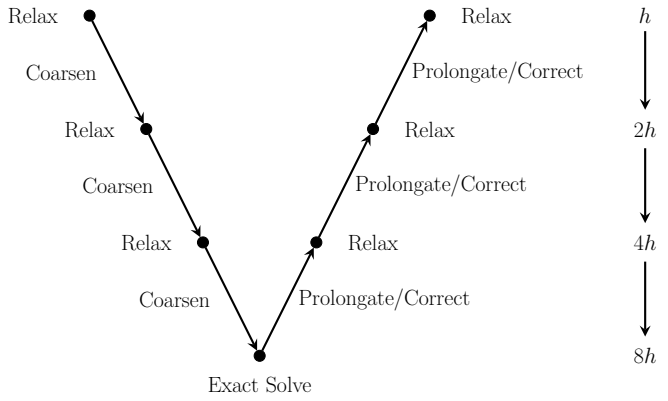


Figure 6.2: Geometric multigrid V-cycle with four grid levels and an exact solve on the coarsest grid.

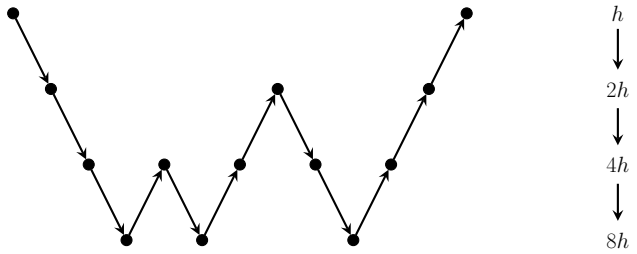


Figure 6.3: Geometric multigrid F-cycle with four grid levels and exact solves on the coarsest grid. Arrows traveling down grids imply solution restriction. Arrows traveling up grids indicate solution interpolation and coarse-grid correction.

Significant research surrounding multigrid methods for saddle-point linear systems has been pursued [10, 42, 121]. The design of proper relaxation procedures for fluid systems arising from the discretization of the Stokes and Navier-Stokes problems is challenging [15, 122], as traditional relaxation techniques, such as pure Gauss-Seidel or Jacobi, cannot be applied due to the non-positive-definite nature of the matrices resulting from discretization of the problems. Therefore, a number

of efficient relaxation schemes specifically addressing the fluid systems' saddle-point block structure have been designed and implemented [69, 70, 121].

In the sections below, we discuss generalizations of Vanka- and Braess-Sarazin-type relaxation schemes to the electric systems in (5.5) and (5.16). This choice of schemes is motivated by the performance and robustness studies of such methods for the block linear systems pertaining to incompressible flows in [68, 75], as well as the numerical success of similar generalizations for the incompressible, resistive MHD equations in [2]. In this case, we are investigating the performance of these multigrid methods as preconditioners for Krylov subspace methods, specifically GMRES.

In the multigrid methods to be discussed, we assume that Q_2–Q_2–P_0 finite elements are used to approximate $\delta\mathbf{n}_h$, $\delta\phi_h$, and $\delta\lambda_h$, respectively, on each grid. Note that the lemmas proved in Chapter 5 demonstrate that the discretization space for $\delta\phi$, in both the electric and flexoelectric models, may be arbitrarily chosen without regard for stability.

6.1 Vanka-type Relaxation

In this section, we discuss the implementation and associated studies of a coupled multigrid method with Vanka-type relaxation. Numerical studies have demonstrated the accuracy and efficiency of Vanka-type relaxation schemes for fluid systems. Furthermore, the relaxation and convergence properties of element-wise Vanka-type relaxation techniques have been studied analytically for the Poisson, Stokes, and Navier-Stokes equations in [89, 91, 97, 111, 114]. Finally, these methods have been shown to achieve desirable convergence rates for coupled-physics systems with saddle-point structures such as those in (5.5) and (5.16) [2]. In this section, we write the general system to be solved as

$$\mathcal{M}\begin{bmatrix}\mathbf{n}\\ \phi\\ \lambda\end{bmatrix} = \begin{bmatrix} A & B_1 & B_2\\ B_1^T & -\tilde{C} & \mathbf{0}\\ B_2^T & \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{n}\\ \phi\\ \lambda\end{bmatrix} = \begin{bmatrix} f_{\mathbf{n}}\\ f_\phi\\ f_\lambda\end{bmatrix},$$

where $\mathcal{M}$ represents a matrix arising for either the electric or flexoelectric models.

Vanka-type relaxation schemes aim at forming small, easily inverted matrices based on the local coupling of the finite-element nodes. The idea is to relax nodes based on cheaply computed solutions to local problems. Due to the use of cell-centered, discontinuous finite elements for the Lagrange multiplier, the Vanka-type relaxation techniques herein, originally formulated in [122] for finite-difference discretizations, are *mesh-cell oriented*. Therefore, in the construction of the Vanka-type relaxation block associated with each Lagrange multiplier degree of freedom, all director and electric potential degrees of freedom associated with the same cell are considered. Let $\mathscr{N}_h$, $\mathscr{E}_h$, and $\mathscr{Q}_h$ denote the director, electric potential, and Lagrange multiplier degrees of freedom, respectively. Define $\mathscr{V}_{hj} = \mathscr{N}_{hj} \cup \mathscr{E}_{hj} \cup \mathscr{Q}_{hj}$ to be the set of degrees of freedom associated with mesh cell j. Let $\mathcal{M}_j$ be the block of matrix $\mathcal{M}$ formed by extracting the rows and columns of $\mathcal{M}$ corresponding to the degrees of freedom in $\mathscr{V}_{hj}$. Hence,

$$\mathcal{M}_j = \begin{bmatrix} A_j & B_{1,j} & B_{2,j} \\ B_{1,j}^T & -\tilde{C}_j & \mathbf{0} \\ B_{2,j}^T & \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{6.1}$$

with dimension $|\mathscr{V}_{hj}|\times|\mathscr{V}_{hj}|$. Solution values for degrees of freedom in $\mathscr{V}_{hj}$ are updated as

$$\begin{bmatrix} \mathbf{n}_{i+1} \\ \phi_{i+1} \\ \lambda_{i+1} \end{bmatrix}_j = \begin{bmatrix} \mathbf{n}_i \\ \phi_i \\ \lambda_i \end{bmatrix}_j + \xi \mathcal{M}_j^{-1} \left(\begin{bmatrix} f_{\mathbf{n}} \\ f_\phi \\ f_\lambda \end{bmatrix} - \mathcal{M} \begin{bmatrix} \mathbf{n}_i \\ \phi_i \\ \lambda_i \end{bmatrix} \right)_j, \tag{6.2}$$

where the subscript j restricts the vectors to the appropriate rows. Thus, a single relaxation step consists of a loop over all mesh elements in the domain. With the use of Q_2 elements for the components of $\delta\mathbf{n}$ and $\delta\phi$ and P_0 elements for $\delta\lambda$, the Vanka blocks, $\mathcal{M}_j$, are matrices of dimension 37×37 and, while relatively dense, quite fast to invert. In this thesis, we refer to this relaxation scheme as full Vanka-type relaxation.

Within the underlying multigrid method, we use standard finite-element interpolation operators and Galerkin coarsening. The multigrid cycles used are $V(1,1)$. The numerical problems to which the multigrid methods presented herein are applied have periodically constrained boundary conditions. Therefore, special care must be taken in the construction of the prolongation operators and coarsening to preserve the periodic structure. For additional details on the numerical implementation of the multigrid method and associated relaxation schemes, see [2].

In addition to the coupled, mesh-cell oriented Vanka-type relaxation technique above, a so-called coupled, economy Vanka-type relaxation approach is considered. This relaxation approach is an extension of the diagonal Vanka smoothers, discussed in [68, 75], aimed at preserving the electric coupling effects present in the problems under consideration, while further reducing the cost of the matrix inversions. The modification results in a block-diagonal relaxation technique rather than a strictly diagonal relaxation matrix as in [68, 75]. In order to formulate the economy Vanka-type relaxation method, the matrix, $\mathcal{M}_j$, in (6.1) is further reduced.

As discussed above, Q_2 finite elements are used for the components of $\mathbf{n}$ and the electric potential, ϕ. Therefore, at each Q_2 finite-element node on mesh cell j, there are four collocated degrees of freedom corresponding to the three components of $\mathbf{n}$ and the scalar component ϕ. Thus, 4×4 blocks are constructed from the matrix entries associated with the interaction of these degrees of freedom. Permuting the matrix entries appropriately yields a block-diagonal Vanka-type relaxation matrix

$$\mathcal{M}_j^E = \begin{bmatrix} A_j^{(1)} & \mathbf{0} & \dots & B_{2,j}^{(1)} \\ \mathbf{0} & A_j^{(2)} & \dots & B_{2,j}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ (B_{2,j}^{(1)})^T & (B_{2,j}^{(2)})^T & \dots & \mathbf{0} \end{bmatrix},$$

where $A_j^{(l)}$ represents the 4×4 block associated with the degrees of freedom grouping for node l on mesh cell j. Note that the blocks $B_{2,j}$ and $B_{2,j}^T$ in (6.1) are not modified with the exception of proper permutations. This economy Vanka-type

relaxation preserves a portion of the coupling between the director and the electric field, which is lost in a strictly diagonal relaxation implementation, while reducing the relaxation matrix structure to a block-diagonal form. The iterative relaxation steps discussed in (6.2) remain valid for the economy relaxation technique with the appropriate permutations included to reflect those performed in the construction of $\mathcal{M}_j^E$. This economy Vanka-type relaxation approach has been shown to provide effective relaxation and improved time per iteration compared with the full Vanka-type relaxation method for the MHD equations [2].

6.1.1 Parameter and Timing Studies

In the following section, we perform parameter studies to determine the optimal value of ξ in (6.2) for both full and economy Vanka-type relaxation. The studies are performed using a flexoelectrically coupled nano-patterened boundary condition problem. Using the same problem, the performance of the multigrid methods using the full and economy Vanka-type relaxation are compared against that of applying UMFPACK LU decomposition [32–35], linked through the deal.II library, as an exact solver.

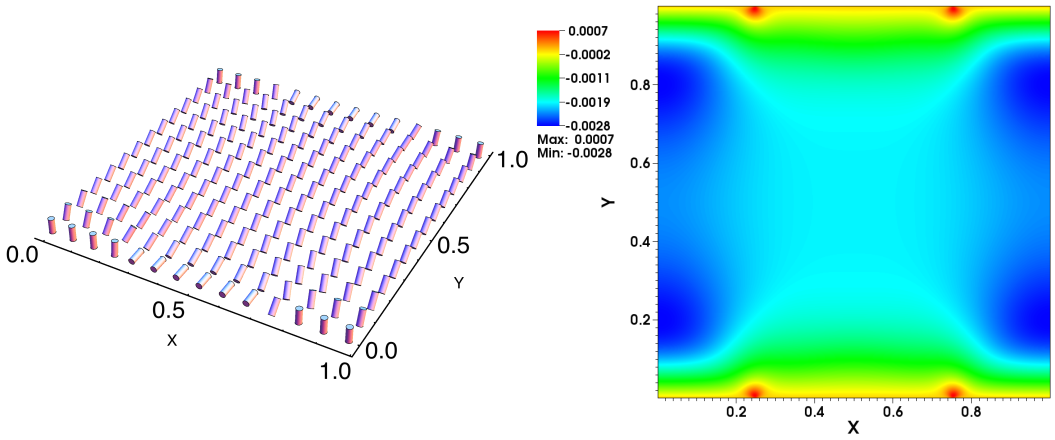
The test problem for these studies considers the same slab domain described in Section 3.7.2. Therefore the domain remains 2-D with $\Omega = \{(x, y) \mid 0 \leq x, y \leq 1\}$. As before, the problem assumes periodic boundary conditions at the edges $x = 0$ and $x = 1$. Dirichlet boundary conditions are enforced on the y-boundaries.

Elastic Constants	$K_1 = 1$	$K_2 = 4$	$K_3 = 1$	$\kappa = 4$	–
Electric Constants	$\epsilon_{\parallel} = 7$	$\epsilon_{\perp} = 7$	$\epsilon_0 = 1.42809$	$e_s = 0.5$	$e_b = 0.5$

Table 6.1: Relevant liquid crystal constants for the Vanka-type relaxation studies.

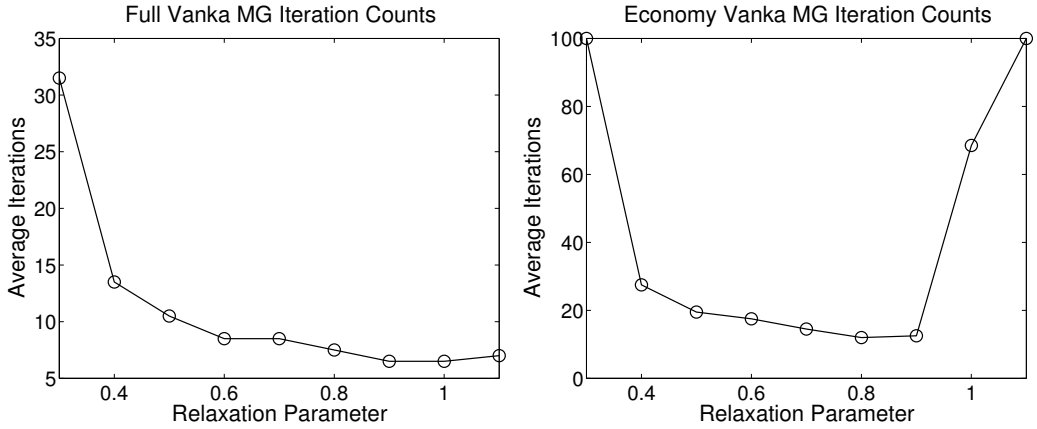
The relevant constants for the flexoelectric problem used in the studies are detailed in Table 6.1. Note that here, and in subsequent sections, we have scaled the appropriate constants by the K_1 scalar value of 5CB, a common liquid crystal, mainly for convenience in adjusting relative sizes of the parameters. We apply the same nano-patterned boundary conditions as in (3.75)-(3.77). The nonlinear residual tolerance is held constant at 10^{-4} for these studies. In Figure 6.4, the final

computed solution for the test problem is displayed alongside the flexoelectrically induced electric potential.



(a) (b)

Figure 6.4: (a) The final computed solution for the test problem on a 512×512 mesh (restricted for visualization). (b) The flexoelectrically induced electric potential.



(a) (b)

Figure 6.5: The average number of multigrid iterations for varying ξ relaxation parameters on a 512×512 grid for (a) full Vanka and (b) economy Vanka.

In this first set of studies, we focus on determining the optimal Vanka relaxation parameter ξ. For these numerical experiments, the multigrid convergence tolerance, which is based on the ratio of the current solution's residual to that of the initial guess, is 10^{-6} for each grid level and Newton step. The relaxation parameter for the full Vanka approach is varied from $\xi = 0.1$ to $\xi = 1.1$, while the economy Vanka relaxation parameter varied from $\xi = 0.1$ to $\xi = 1.0$, each in increments of 0.05. The

corresponding average multigrid iteration counts for a 512×512 grid and a selection of ξ values is displayed in Figure 6.5a and 6.5b for full and economy Vanka-type relaxation, respectively.

For the figure, relaxation parameters smaller than 0.3 are not included, as they resulted in iteration counts of over 100 before the multigrid residual tolerance was satisfied. The studies indicate that a relaxation parameter of $\xi = 1.00$ for full Vanka-type relaxation and an under-relaxation parameter $\xi = 0.85$ for economy are optimal for convergence. Note that, in all cases, economy Vanka relaxation required more iterations to reach the convergence tolerance. However, each iteration individually requires less time than an iteration of the full Vanka relaxation approach.

The second set of numerical experiments compares the system solve times for the multigrid solver with both Vanka-type relaxation techniques against the performance of the UMFPACK LU decomposition solver utilized by deal.II. The experiments compare the linear solvers, on the above problem, with full nested iteration beginning on an 8×8 grid uniformly refining to a 512×512 mesh. For all solvers considered, we report the total time to solution, including both the setup and solve phases of the algorithms, but neglect some overhead associated with converting data formats and interfacing libraries. The optimal relaxation parameters, $\xi = 1.00$ and $\xi = 0.85$, are used for the full and economy Vanka-type relaxation techniques, respectively. We consider multigrid methods using standard residual-based stopping tolerances, fixed on all grids, of reduction in the linear residual by factors of 10^{-8}, 10^{-6}, and 10^{-4}.

Table 6.2 displays the average time to solution for the linear systems arising on successive grids. In the table, each of the multigrid solve timings are scaling nearly perfectly with grid size, while the LU decomposition solve times are growing at a faster rate. For the present timings, the LU decomposition solver is approximately scaling with a factor of 5, and has an expected asymptotic scaling factor of 8. The table also displays a clear confluence of the solve time for LU decomposition and the multigrid solvers. For a multigrid residual tolerance of 10^{-4}, the time to solution for the full Vanka-type relaxation becomes nearly equal to that of the LU decomposition

Solver\Grid	8×8	16×16	32×32	64×64	128×128	256×256	512×512
LU	0.02	0.11	0.57	2.67	12.03	55.78	275.86
Full 1e-8	0.05	0.23	1.17	4.87	20.18	82.77	337.84
Full 1e-6	0.05	0.20	0.91	3.78	16.72	66.38	276.91
Full 1e-4	0.04	0.17	0.74	3.06	13.13	**54.39**	**214.14**
Econ. 1e-8	0.06	0.31	1.54	6.53	27.19	109.50	439.61
Econ. 1e-6	0.05	0.26	1.21	5.13	20.81	85.87	344.45
Econ. 1e-4	0.04	0.19	0.90	3.77	15.56	64.41	**251.56**

Table 6.2: Comparison of average time to solution (in seconds) with LU decomposition (LU), full Vanka relaxation (Full), and economy Vanka relaxation (Econ) for varying grid size. Numbers following the relaxation type indicate the multigrid residual tolerance. Bold face numbers indicate improved time to solution compared with the LU decomposition solver.

solver as early as the 128×128 grid. This occurs at the 512×512 grid for the multigrid solver with economy Vanka-type relaxation and a 10^{-4} multigrid tolerance. Moreover, though the applied Vanka-type relaxation methods yield approximate linear solvers, the number of overall Newton steps does not increase for any of the experiments compared to the direct solver. Therefore, the method is robust with respect to adjustments in the multigrid tolerance.

The results of these studies suggest that the full and economy Vanka-type relaxation methods discussed above yield effective, efficient, and scalable multigrid solvers applicable to the coupled saddle-point linear systems arising in the discretization of the electric and flexoelectric models. Furthermore, the relaxation techniques exhibit notable performance for a range of multigrid residual tolerances and relaxation parameters. Due to the superior performance of the full Vanka-type relaxation approach, when using Vanka-type relaxation in the numerical simulations, only full Vanka relaxation is applied in future studies with a relaxation parameter of 1.00 and a multigrid residual tolerance of 10^{-6} for assured accuracy.

6.2 Braess-Sarazin-type Relaxation

As with the Vanka-type relaxation schemes, Braess-Sarazin relaxation methods have been well studied for fluid systems arising from discretization of the Stokes' and Navier-Stokes' equations [70, 71, 75]. Numerical studies of Braess-Sarazin-type

schemes, generalized to the MHD equations, show convincing evidence that multigrid with Braess-Sarazin-type relaxation readily outperforms both direct solvers and Vanka-type relaxation schemes [2]. This notion is reaffirmed in the numerical tests conducted below.

As above, the general electric system under consideration is written

$$\mathcal{M}\begin{bmatrix} \mathbf{n} \\ \phi \\ \lambda \end{bmatrix} = \begin{bmatrix} A & B_1 & B_2 \\ B_1^T & -\tilde{C} & \mathbf{0} \\ B_2^T & \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{n} \\ \phi \\ \lambda \end{bmatrix} = \begin{bmatrix} f_{\mathbf{n}} \\ f_\phi \\ f_\lambda \end{bmatrix}.$$

Define blocks of $\mathcal{M}$ as

$$\hat{A} = \begin{bmatrix} A & B_1 \\ B_1^T & -\tilde{C} \end{bmatrix}, \qquad \hat{B} = \begin{bmatrix} B_2 \\ \mathbf{0} \end{bmatrix}. \tag{6.3}$$

Furthermore, let $\hat{u} = [\mathbf{n}\ \phi]^T$ and $\hat{f}_{\hat{u}} = [f_{\mathbf{n}}\ f_\phi]^T$. With these block definitions, the Braess-Sarazin update scheme, originally formulated in [15] for Stokes flows, is used and takes the form

$$\begin{bmatrix} \hat{u}_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} \hat{u}_k \\ \lambda_k \end{bmatrix} + \begin{bmatrix} \gamma_b R & \hat{B} \\ \hat{B}^T & \mathbf{0} \end{bmatrix}^{-1}\left(\begin{bmatrix} \hat{f}_{\hat{u}} \\ f_\lambda \end{bmatrix} - \begin{bmatrix} \hat{A} & \hat{B} \\ \hat{B}^T & \mathbf{0} \end{bmatrix}\begin{bmatrix} \hat{u}_k \\ \lambda_k \end{bmatrix}\right), \tag{6.4}$$

where R is an appropriate preconditioner for $\hat{A}$ and γ_b is a weighting parameter. Performing these Braess-Sarazin updates requires solving a system of the form

$$\begin{bmatrix} \gamma_b R & \hat{B} \\ \hat{B}^T & \mathbf{0} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} d \\ e \end{bmatrix}. \tag{6.5}$$

Solutions of (6.5) satisfy

$$\begin{aligned} Sy &= \hat{B}^T R^{-1} d - \gamma_b e, \\ x &= \frac{1}{\gamma_b} R^{-1}(d - \hat{B}y), \end{aligned} \tag{6.6}$$

where $S = -\hat{B}^T R^{-1} \hat{B}$. For the multigrid implementations used here, (6.6) is approximately solved using one sweep of Symmetric (point) Gauss-Seidel.

Since we use Q_2 elements for both $\mathbf{n}$ and ϕ, the degrees of freedom for the components of $\mathbf{n}$ and ϕ are collocated. As suggested in [2], we construct two possible preconditioners. The first preconditioner, R_e, is simply

$$R_e = \text{diag}(\hat{A}) = \begin{bmatrix} \text{diag}(A) & \mathbf{0} \\ \mathbf{0} & \text{diag}(-C) \end{bmatrix}.$$

Since R_e is strictly diagonal, computing R_e^{-1} and performing multiplication is quite simple.

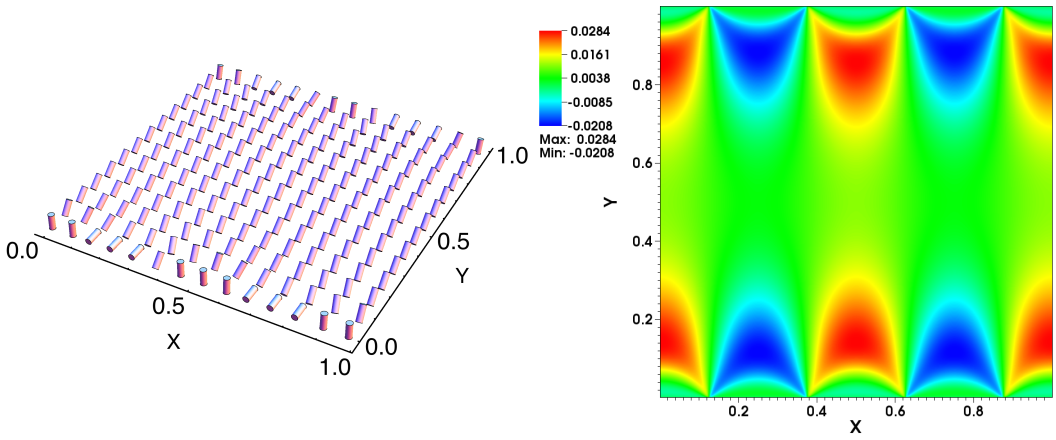
The second preconditioner, R_d, is formed by extracting 4×4-blocks of $\hat{A}$ corresponding to the nodally collocated degrees of freedom for $\mathbf{n}$ and ϕ. With careful permutation of the degrees of freedom in Equation (6.4), R_d becomes a block-diagonal matrix with a diagonal consisting of the 4×4 collocation blocks. This preconditioner maintains some of the electric coupling of the original system, $\hat{A}$, while remaining relatively easy to invert and compute with. In [2], it is observed that as the magnetic forces begin to dominate the kinetics of the MHD equations, preserving the electric coupling becomes increasingly important to solver performance.

The same multigrid framework used for the multigrid methods with Vanka-type relaxation is applied for the multigrid techniques with Braess-Sarazin-type relaxation. For this chapter, the Braess-Sarazin-type relaxation scheme using the preconditioner R_e is referred to as diagonal Braess-Sarazin relaxation, while the scheme utilizing R_d is designated as block-diagonal Braess-Sarazin relaxation.

6.2.1 Parameter and Timing Studies

A flexoelectrically coupled problem is again used for the studies to follow. However, the boundary conditions considered are a doubling of the nano-pattern described by Equations (3.75) - (3.77), such that the pattern contains a second strip parallel to the xy-plane; see Figure 6.6a. While there is no applied electric field, the curvature

induced by the nano-patterning again generates an internal electric field due to the flexoelectric properties of the liquid crystals. The computed equilibrium configuration and induced field are displayed in Figure 6.6a and b, respectively. The relevant Frank and electric constants for the timing and parameter testing are displayed in Table 6.3.



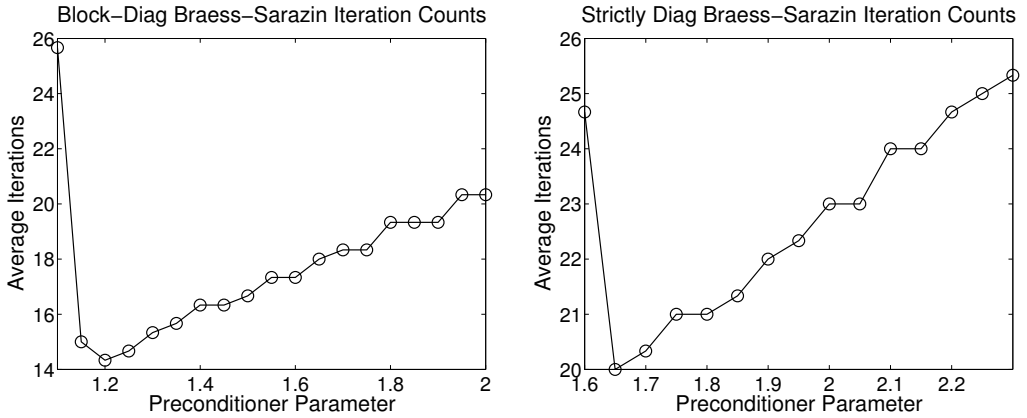
(a) (b)

Figure 6.6: (a) The computed final solution for the double nano-patterned boundary conditions with electric and flexoelectric augmentation on a 512×512 grid (restricted for visualization). (b) The flexoelectrically induced electric potential.

Elastic Constants	$K_1 = 1.0$	$K_2 = 4.0$	$K_3 = 1.0$	$\kappa = 4$	–
Electric Constants	$\epsilon_{\parallel} = 7.0$	$\epsilon_{\perp} = 7.0$	$\epsilon_0 = 1.42809$	$e_s = 1.5$	$e_b = -1.5$

Table 6.3: Relevant liquid crystal constants for the Braess-Sarazin-type relaxation studies.

We first focus on determining the optimal γ_b value for both Braess-Sarazin relaxation methods. Here, the multigrid convergence tolerance remains fixed at 10^{-6} for each grid level and Newton step, while the nonlinear tolerance was 10^{-4}. The parameter γ_b was varied from 1.10 to 2.00 for block-diagonal Braess-Sarazin relaxation and 1.60 to 2.30 for diagonal Braess-Sarazin, each in increments of 0.05. Displayed in Figures 6.7a and b are the multigrid iteration counts averaged over Newton iterations on a 512×512 mesh with respect to varying values of γ_b for block-diagonal and diagonal Braess-Sarazin relaxation, respectively.

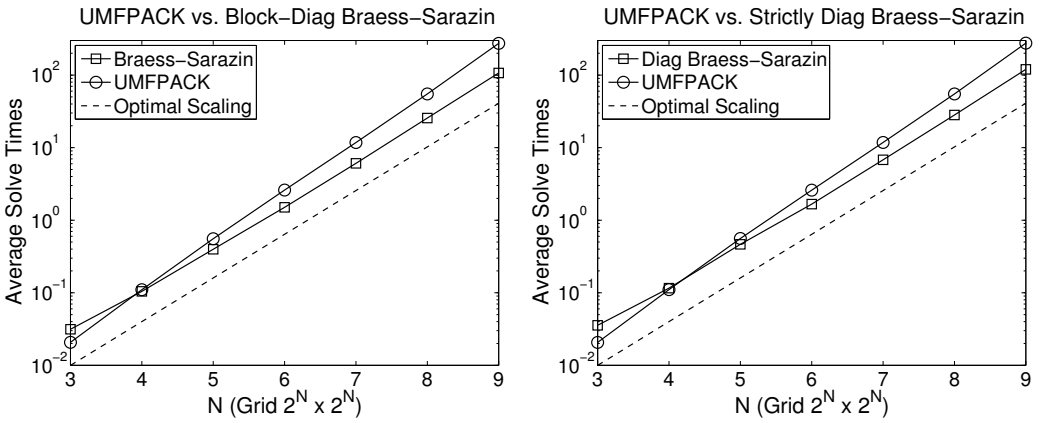


(a) (b)

Figure 6.7: The average number of multigrid iterations for varying γ_b on a 512×512 grid with (a) block-diagonal Braess-Sarazin relaxation and (b) diagonal Braess-Sarazin

The parameter studies suggest that γ_b values of 1.20 for block-diagonal Braess-Sarazin relaxation and 1.65 for Braess-Sarazin relaxation are optimal for convergence. Therefore, for the remainder of the multigrid applications, these values for γ_b are applied. It is interesting to note that iteration counts are relatively insensitive to increases in γ_b, for both schemes, above the optimum values of 1.20 and 1.65. While slightly more sensitive, the average iteration counts for diagonal Braess-Sarazin relaxation only increase by about 5 if γ_b is increased to 2.30.

Figures 6.8a and b exhibit average setup and solve times across a hierarchy of grids beginning at an 8×8 mesh ascending to a 512×512 mesh for the UMFPACK direct solver and Braess-Sarazin-type multigrid schemes. Using Q_2–Q_2–P_0 elements for $\mathbf{n}$, ϕ, and λ, respectively, our matrices are of dimension $4,464,644 \times 4,464,644$ with $286,969,900$ nonzero entries on the finest mesh. It is clear in the figure that the multigrid schemes with Braess-Sarazin-type relaxation are scaling optimally with the grid size. Furthermore, there is a clear timing crossover around the 16×16 mesh, at which point both multigrid methods with Braess-Sarazin-type relaxation become the faster solver. In agreement with the numerical findings in [2], this timing intersection occurs considerably earlier than the Vanka-type relaxation scheme discussed in Section 6.1.1.



(a) (b)

Figure 6.8: The average time to solution for the (a) block-diagonal Braess-Sarazin and (b) diagonal Braess-Sarazin schemes with a multigrid tolerance of 10^{-6} compared to the UMFPACK direct solver.

Solver\Grid	8×8	16×16	32×32	64×64	128×128	256×256	512×512
LU	0.02	0.11	0.56	2.64	11.87	55.14	273.38
Block 1e-8	0.04	0.12	**0.46**	**1.78**	**7.06**	**30.04**	**125.59**
Block 1e-6	0.03	**0.10**	**0.40**	**1.50**	**6.07**	**25.56**	**106.76**
Block 1e-4	0.03	**0.09**	**0.36**	**1.31**	**5.13**	**21.88**	**91.45**
Diag. 1e-8	0.04	0.14	**0.56**	**2.04**	**8.26**	**34.44**	**143.87**
Diag. 1e-6	0.04	0.12	**0.46**	**1.67**	**6.80**	**28.15**	**119.89**
Diag. 1e-4	0.03	**0.10**	**0.36**	**1.35**	**5.39**	**24.03**	**97.86**

Table 6.4: Comparison of average time to solution (in seconds) with LU decomposition (LU), block-diagonal Braess-Sarazin (Block), and strictly diagonal Braess-Sarazin (Diag) for varying grid size. Numbers following the relaxation type indicate the multigrid residual tolerance. Bold face numbers indicate improved time to solution compared with the LU decomposition solver.

Note that the block-diagonal Braess-Sarazin technique is performing slightly better than the diagonal Braess-Sarazin scheme in comparison to the UMFPACK solver. This difference in the solve times is reflected across grids and varying multgrid tolerances in Table 6.4. Even at a multigrid tolerance of 10^{-8}, both the block-diagonal and diagonal Braess-Sarazin based multigrid schemes outpace the UMFPACK exact solve times by the 32×32 grid. This leads to the large time to solution improvements seen in Table 6.5.

Table 6.5 details an itemized comparison of the UMFPACK direct solver's performance to that of the Braess-Sarazin-type multigrid scheme. The totals represent the time each algorithm spent performing the listed task, summed across all grids.

For the run statistics in the table, we also paired the multigrid solver with the simple trust-region method for the Lagrange multiplier formulation discussed in Chapter 4. With and without trust regions, the computed free energy between each of the solvers is identical. On the other hand, Braess-Sarazin-type relaxation reduces overall runtime by approximately 32% and 28% for the block-diagonal and diagonal approaches, respectively. This speed up is most notable when considering the fact that overall runtime for the multigrid solver experiments includes porting variables to types compatible with the Trilinos computational library [63] and computing collocation information to match the Trilinos format. Focusing on the linear setup and solve times alone, the block-diagonal and diagonal relaxation schemes yield approximately 58% and 54% reductions, respectively. Overall, the multigrid methods with Braess-Sarazin-type relaxation offer optimal scaling and exceptionally efficient timing.

	UMFPACK Solve		Block Braess-Sarazin		Diag. Braess-Sarazin	
Trust-Region	None	Simple	None	Simple	None	Simple
Free Energy	16.413	16.413	16.413	16.413	16.413	16.413
Sys. Assem.	136.1s	131.3s	136.8s	131.8s	136.5s	130.4s
Data Conv.	–	–	136.9s	132.6s	136.3s	133.8s
Lin. Setup/Solve	1053.2s	1035.2s	436.8s	425.2s	488.9s	475.6s
Mem./Output	284.7s	283.2s	305.8s	302.6s	303.6s	300.3s
Total Time	1474.0s	1449.7s	1016.3s	992.2s	1065.3s	1040.1s

Table 6.5: A comparison of computation statistics for runs using the UMFPACK direct solver or the Braess-Sarazin schemes. Each solver is run with and without trust regions. For each algorithm, the computed free energy on the finest grid and the overall run time, broken into constituent parts, are included.

Table 6.6 compares the performance of the Braess-Sarazin-type relaxation techniques to that of the full Vanka-type scheme. We compare the methods using the flexoelectric simulation outlined in Table 6.1 for the Vanka studies. Note that full Vanka-type relaxation is the clear leader in terms of iteration counts. However, repeated computation of the local residuals for the Vanka update scheme in (6.2) is relatively expensive. Thus, while the iteration counts for the Vanka-type method are much smaller, the overall time per iteration is much larger than the Braess-Sarazin-type relaxation approaches, resulting in a significant time-to-solution advantage for the Braess-Sarazin-type techniques. The table shows that both Braess-Sarazin-type

methods reduce the total time spent in the linear solve phase by approximately 500 seconds when compared to the Vanka-type scheme and block-diagonal Braess-Sarazin-type relaxation demonstrates the best performance.

Relaxation Scheme	Avg Iters	Sys. Assem.	Lin. Setup/Solve	Total Time
Block Braess-Sarazin	14.6	102.2s	328.6s	782.7s
Diagonal Braess-Sarazin	20.0	101.9s	366.8s	822.7s
Full Vanka	6.4	97.3s	847.4s	1326.0s

Table 6.6: A comparison of computation statistics for Vanka- and Braess-Sarazin-type schemes. For each multigrid approach, the average number of iterations and total time spent performing each task are reported. The multigrid tolerance for both methods was fixed at 10^{-6}, while the nonlinear tolerance was 10^{-4}.

6.3 Numerical Results

In this section, we present numerical simulations incorporating applied electric and internally induced electric fields. The algorithm to perform the minimizations discussed in Chapter 5 is similar to Algorithm 1, with the addition of the electric potential, ϕ. This algorithm, outlined in Algorithm 7, performs nested iteration and uses the first-order optimality condition stopping tolerance discussed above.

The linear system for each Newton step has the anticipated saddle-point block structure, detailed in (5.5) and (5.16). Due to its exceptional performance, the discretization matrices are inverted using the block-diagonal Braess-Sarazin based multigrid approach with a multigrid tolerance of 1e−6, in order to approximately solve for the discrete updates $\delta\mathbf{n}_h$, $\delta\phi_h$, and $\delta\lambda_h$. Finally, damped Newton corrections are performed. That is, the new iterates are given by

$$\begin{bmatrix} \mathbf{n}_{k+1} \\ \phi_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{n}_k \\ \phi_k \\ \lambda_k \end{bmatrix} + \omega \begin{bmatrix} \delta\mathbf{n}_h \\ \delta\phi_h \\ \delta\lambda_h \end{bmatrix}, \tag{6.7}$$

where $\omega \leq 1$. For this algorithm, ω is chosen to begin at 0.2 on the coarsest grid and increases by 0.2, to a maximum of 1, after each grid refinement. As mentioned above, Q_2–Q_2–P_0 discretizations are used to approximate $\delta\mathbf{n}_h$, $\delta\phi_h$, and $\delta\lambda_h$, respectively.

For all of the problems undertaken below, the slab domain of Section 3.7.2, $\Omega = \{(x, y) \mid 0 \le x, y \le 1\}$, is considered with mixed Dirichlet and periodic boundary conditions.

Algorithm 7: Minimization algorithm with NI for the electric systems

0. Initialize $(\mathbf{n}_0, \phi_0, \lambda_0)$ on coarse grid.

while *Refinement limit not reached* **do**

  while *First-order optimality conformance threshold not satisfied* **do**

    1. Set up discrete linear system (5.11) on current grid, H.
    2. Solve for $\delta\mathbf{n}_H$, $\delta\phi_H$, and $\delta\lambda_H$.
    3. Compute $\mathbf{n}_{k+1}$, ϕ_{k+1}, and λ_{k+1} as in (6.7).

  end

  4. Uniformly refine the grid.
  5. Interpolate $\mathbf{n}_H \to \mathbf{n}_h$, $\phi_H \to \phi_h$, and $\lambda_H \to \lambda_h$.

end

6.3.1 Simple Electric Freedericksz Transition

The first liquid crystal numerical experiment considers simple director boundary conditions, such that $\mathbf{n}$, along both of the substrates, lies uniformly parallel to the x-axis. The boundary conditions for the electric potential, ϕ, are such that $\phi = 0$ on the lower substrate at $y = 0$ and $\phi = 1$ at $y = 1$. The relevant constants for the problem are detailed in Table 6.7. Since the electric anisotropy constant, ϵ_a, is positive, the expected behavior for the liquid crystal configuration is a Freedericksz transition [51, 130] so long as the applied field is strong enough to overcome the inherent elastic effects of the system. That is, for an applied voltage above a critical value, known as a Freedericksz threshold [117], the liquid crystal configuration departs from uniform alignment parallel to the x-axis and instead tilts in the direction of the applied field.

Elastic Constants	$K_1 = 1$	$K_2 = 0.62903$	$K_3 = 1.32258$	$\kappa = 0.475608$
Electric Constants	$\epsilon_0 = 1.42809$	$\epsilon_\parallel = 18.5$	$\epsilon_\perp = 7$	$\epsilon_a = 11.5$

Table 6.7: Relevant liquid crystal constants for Freedericksz transition problem.

The problem considered here has an analytical solution. Using the approach outlined in [40,117], let $\mathbf{n} = (\cos\theta(y), \sin\theta(y), 0)$, and denote the voltage applied at the top of the substrate as V. Further, observe that due to the imposed Dirichlet boundary conditions, $\theta(0) = \theta(1) = 0$. Hence, symmetry across the horizontal midline between the substrates is expected. Therefore, the solution discussed below is valid for $0 \le y \le \frac{1}{2}$.

Given a fixed voltage V, the maximum angular displacement from elastic rest, θ_m, is given implicitly by the equation

$$V = 2\sqrt{\frac{K_1}{\epsilon_0\epsilon_a}}(1+\mu\sin^2\theta_m)^{1/2}\int_0^{\pi/2}\left(\frac{1+\kappa_0\sin^2\theta_m\sin^2\lambda}{(1+\mu\sin^2\theta_m\sin^2\lambda)(1-\sin^2\theta_m\sin^2\lambda)}\right)^{1/2}d\lambda, \quad (6.8)$$

where $\mu = \epsilon_a/\epsilon_\perp$ and $\kappa_0 = (K_3 - K_1)/K_1$. This parameter is fixed by the physical constants of the liquid crystal and the experimentally applied voltage. Using θ_m, for a given value of y, the analytical solution for $\theta(y)$ is expressed implicitly as

$$2y\int_0^{\theta_m}\left(\frac{(1+\kappa_0\sin^2\hat{\theta})(1+\mu\sin^2\hat{\theta})}{\sin^2\theta_m - \sin^2\hat{\theta}}\right)^{1/2}d\hat{\theta} = \int_0^{\theta}\left(\frac{(1+\kappa\sin^2\hat{\theta})(1+\mu\sin^2\hat{\theta})}{\sin^2\theta_m - \sin^2\hat{\theta}}\right)^{1/2}d\hat{\theta}. \quad (6.9)$$

In order to simplify the involved computations, these equations are non-dimensionalized. Let $V_c = \pi\sqrt{\frac{K_1}{\epsilon_a\epsilon_0}}$. Then (6.8) is rewritten

$$\bar{V} = \frac{V}{V_c} = \frac{2}{\pi}(1+\mu\sin^2\theta_m)^{1/2}\int_0^{\pi/2}\left(\frac{1+\kappa_0\sin^2\theta_m\sin^2\lambda}{(1+\mu\sin^2\theta_m\sin^2\lambda)(1-\sin^2\theta_m\sin^2\lambda)}\right)^{1/2}d\lambda.$$

Applying the substitution

$$\sin\hat{\theta} = \sin\theta_m\sin\lambda,$$

from [130], Equation (6.9) is rewritten for $\theta_m > 0$ as

$$2y\int_0^{\pi/2}\left(\frac{(1+\kappa_0\sin^2\theta_m\sin^2\lambda)(1+\mu\sin^2\theta_m\sin^2\lambda)}{1-\sin^2\theta_m\sin^2\lambda}\right)d\lambda = \int_0^{\Theta}\left(\frac{(1+\kappa_0\sin^2\theta_m\sin^2\lambda)(1+\mu\sin^2\theta_m\sin^2\lambda)}{1-\sin^2\theta_m\sin^2\lambda}\right)d\lambda,$$

where we solve for $\Theta = \sin^{-1}\left(\frac{\sin\theta}{\sin\theta_m}\right)$. Utilizing the symmetry discussed above, solutions for $\frac{1}{2} \le y \le 1$ are given by $\theta(y) = \theta(1-y)$. This specific rescaling of V is

chosen with the implication that if $\bar{V} > 1$, then a Freedericksz transition is expected to occur and a nonzero θ_m is involved in the free-energy minimizing configuration.

For the specific constants considered in Table 6.7, the critical voltage is V_c = 0.7752 and the angular deviation θ_m = 0.662. Thus, the anticipated solution should demonstrate a true Freedericksz transition away from uniform free-elastic alignment. Indeed, the final computed solution in Figure 6.9, displayed alongside the initial guess for the algorithm, demonstrates the expected transition. The true free energy, computed from the analytical solution, for this problem is −5.3295. This free energy is accurately captured by the minimization approach.

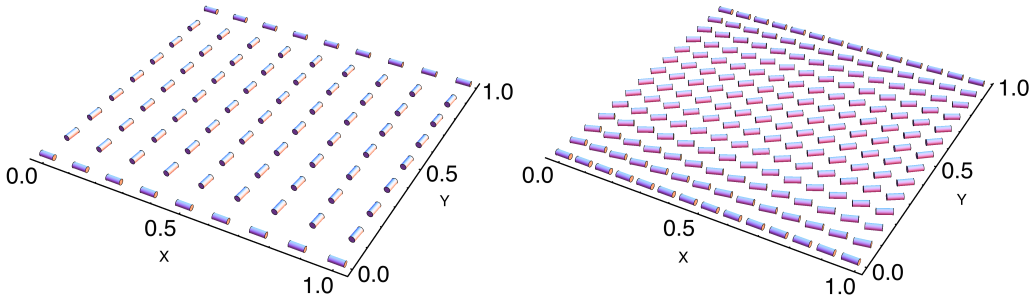
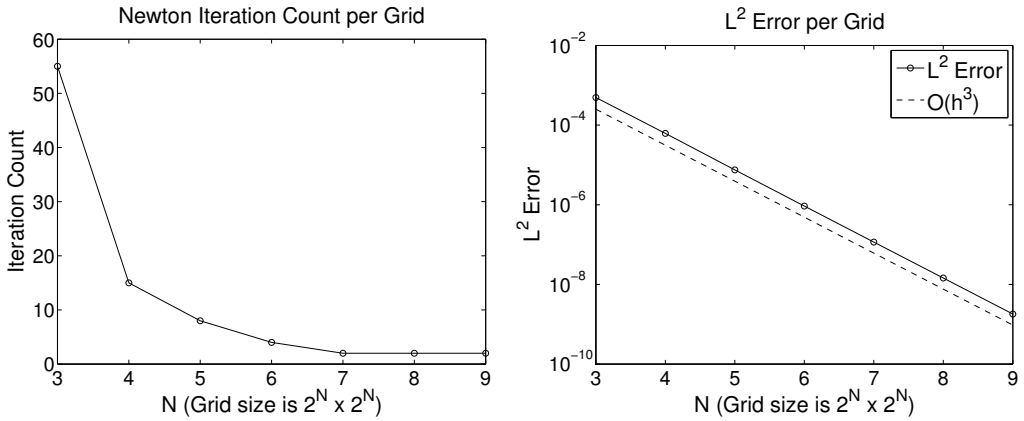


Figure 6.9: (a) Initial guess on 8×8 mesh with initial free energy of 26.767 and (b) resolved solution on a 512×512 mesh (restricted for visualization) with final free energy of -5.330 for a Freedericksz transition.

The problem is solved on an 8×8 coarse grid with six successive uniform refinements resulting in a 512×512 fine grid. The minimized functional energy is $\mathcal{F}_5 = -5.3295$, compared to the initial guess energy of 26.767. Figure 6.10a details the number of Newton iterations necessary to reduce the (nonlinear) residual below the given tolerance, 10^{-4}, on each grid. Note that a sizable majority of the Newton iteration computations are isolated to the coarsest grids, with the finest grids requiring only two Newton iteration to reach the tolerance limit. Without the use of nested iteration, the algorithm requires 63 Newton steps on the finest grid, alone, to reach a similar error measure. The nested-iteration-Newton-multigrid method achieves an accurate solution in 11.35 minutes, compared to a total run time of over 3.53 hours for standard Newton-multigrid. This corresponds to a speed up factor of 18.6 or a work requirement for the nested iterations equivalent to 3.38 times that of

assembling and solving a single linearization step on the finest grid.



(a) (b)

Figure 6.10: (a) Newton iterations and (b) L^2-error per grid for the Freedericksz transition.

Also detailed in Figure 6.10b is the reduction in overall L^2-error comparing the analytical solution to the resolved solution on each grid. Note that the error is approximately reduced by a full order of magnitude on each successive grid, corresponding to approximately $O(h^3)$ reductions in overall error. Moreover, for the finer grids, a single Newton step was sufficient to achieve such a reduction. Table 6.8 details run statistics on each grid for the Freedericksz transition problem. The algorithm matches the analytical free energy by the 16×16 grid and results in quite accurate unit-length conformance.

Grid Dim.	L^2-Error	Min Dev.	Max Dev.	Final Energy
8×8	4.92e-04	-9.87e-04	9.98e-04	-5.3297
16×16	6.19e-05	-1.64e-04	1.50e-04	-5.3295
32×32	7.50e-06	-1.56e-05	1.54e-05	-5.3295
64×64	9.28e-07	-1.94e-06	1.92e-06	-5.3295
128×128	1.16e-07	-2.39e-07	2.37e-07	-5.3295
256×256	1.45e-08	-2.98e-08	2.96e-08	-5.3295
512×512	1.81e-09	-3.72e-09	3.71e-09	-5.3295

Table 6.8: Grid and solution progression for the simple Freedericksz transition problem with L^2-error, minimum and maximum director deviations from unit length at the quadrature nodes, and final functional energy on each grid.

6.3.2 Electric Field with Patterned Boundary Conditions

In the second run, the nano-patterned boundary conditions described by (3.75) - (3.77) are applied. The same constants outlined in Table 6.7 are also used for this problem. However, a stronger voltage such that $\phi = 2$ on the substrate at $y = 1$ is applied. Along the other substrate, ϕ remains equal to 0. The final solution, as well as the initial guess, are displayed in Figure 6.11. For this problem, the grid progression again begins on an 8×8 grid ascending uniformly to a 512×512 fine grid. The minimized functional energy is $\mathcal{F}_5 = -41.960$, compared to the initial guess energy of -31.141.

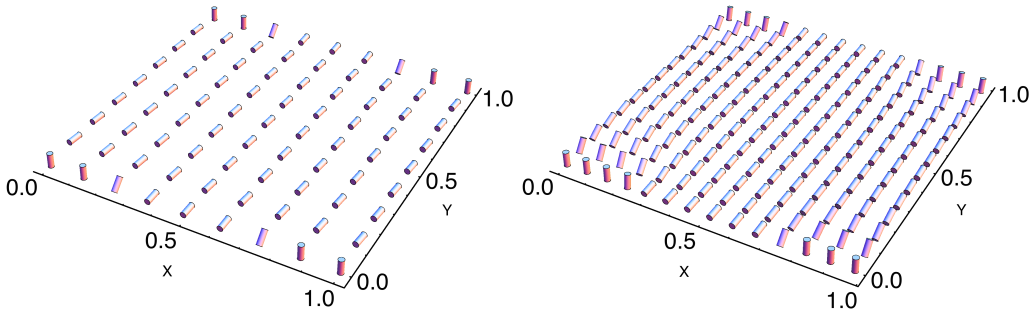


Figure 6.11: (a) Initial guess on an 8×8 mesh with initial free energy of -31.141 and (b) resolved solution on a 512×512 mesh (restricted for visualization) with final free energy of -41.960 for a nano-patterned boundary.

Grid Dim.	Newton Iter.	Init. Res.	Final Res.	Deviation in $\lvert\mathbf{n}\rvert^2$	Final Energy
8×8	54	12.27e-00	7.29e-05	-1.11e-01, 6.10e-02	-42.701
16×16	21	2.01e-00	4.47e-05	-7.64e-02, 4.24e-02	-42.170
32×32	12	9.91e-01	1.66e-05	-4.60e-02, 2.92e-02	-41.963
64×64	7	5.52e-01	7.05e-06	-1.80e-02, 1.31e-02	-41.950
128×128	3	2.36e-01	1.55e-12	-3.63e-03, 2.89e-03	-41.960
256×256	2	7.26e-02	1.64e-10	-4.92e-04, 3.62e-04	-41.960
512×512	2	1.87e-02	6.48e-12	-7.36e-05, 6.39e-05	-41.960

Table 6.9: Grid and solution progression for an electric problem and nano-patterned boundary with initial and final residuals for the first-order optimality conditions, minimum and maximum director deviations from unit length at the quadrature nodes, and final functional energy on each grid.

In Table 6.9, the number of Newton iterations per grid is detailed as well as the conformance of the solution to the first-order optimality conditions after the first and final Newton steps, respectively, on each grid. As with the previous example, much

of the computational work is relegated to the coarsest grids. Here, the total work required is 3.54 times that of assembling and solving a single linearization step on the finest grid. In contrast, without nested iteration, the algorithm requires 62 Newton steps on the 512×512 fine grid, alone, to satisfy the tolerance limit. While the nested-iteration-Newton-multigrid method achieves convergence in 11.91 minutes, the standard Newton-multigrid total run time is over 3.47 hours. Also shown in Table 6.9, the minimum and maximum director deviations from unit length at the quadrature nodes are descending towards zero.

Due to the sizable applied electric field, and the elastic influence of the central boundary condition pattern aligned with the electric field, the expected configuration is a quick transition from the boundary conditions to uniform alignment with the field. That is, the strength of the Freedericksz transition on the interior of Ω is augmented by the presence of this type of patterned boundary condition. This behavior is accurately resolved in the computed solution.

6.3.3 Flexoelectric Phenomena

In this section, we demonstrate the ability of the energy-minimization approach to capture expected physical phenomenon and predict the presence of new physics. As discussed above, internally generated electric fields due to flexoelectricity are an important physical aspect of certain liquid crystal configurations. This polarization due to curvature can significantly affect stable liquid crystal configurations in the presence of certain boundary conditions, such as patterned surfaces, that cause large distortions in the nematic. These may also cause physical phenomenon, such as bistability [5, 6, 30], that are important for display applications.

The following numerical results utilize boundary conditions similar to those in (3.75)-(3.77) with an extra parameter*, φ, which has the effect of varying the imposed azimuthal director angle along the x-axis of the outer, vertically-aligned strips

*Note that, here, φ is utilized for the azimuthal angle whereas in [5], ϕ was used.

on the boundary,

$$n_1 = \sin(\varphi)\sin\left(r(\pi + 2\tan^{-1}(X_m) - 2\tan^{-1}(X_p))\right),$$
$$n_2 = \cos\left(r(\pi + 2\tan^{-1}(X_m) - 2\tan^{-1}(X_p))\right),$$
$$n_3 = \cos(\varphi)\sin\left(r(\pi + 2\tan^{-1}(X_m) - 2\tan^{-1}(X_p))\right).$$

The NI progression from 8×8 grids to 512×512 grids persists for each of the simulations. Due to the complexity of the flexoelectric systems, the nonlinear residual stopping tolerance is decreased to 10^{-5}.

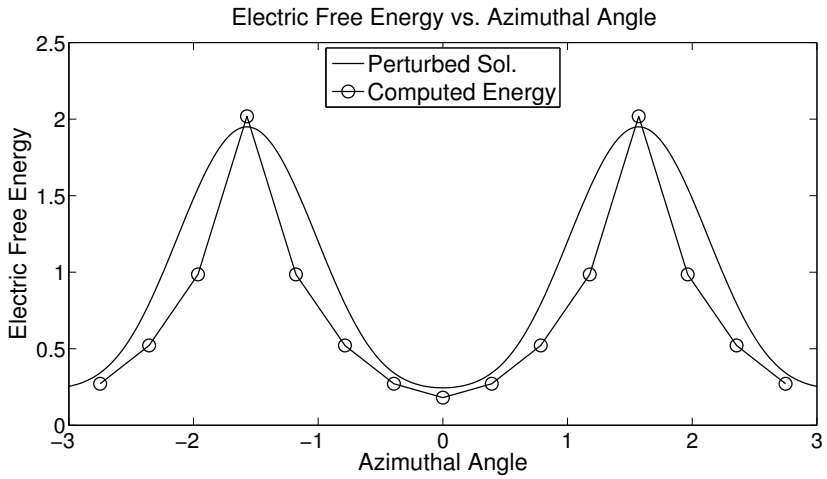


Figure 6.12: The computed final free energy of the perturbative solution with $K_1 = K_2 = K_3 = 1$, $e_s = 5$, and $e_b = -5$ for varying φ values. A perturbation solution similar to that given in [5] is overlaid

In the first experiment, we isolate the influence of flexoelectricity on the configuration by removing elastic anisotropy, setting $K_1 = K_2 = K_3 = 1$, and using a small dielectric anisotropy $\epsilon_\parallel = 7$ and $\epsilon_\perp = 6.9$. For both experiments, as above, $\epsilon_0 = 1.42809$. The computed free energy as a function of the azimuthal angle φ is shown in Figure 6.12, revealing that $\varphi = 0$ and $\varphi = \pi$ are the minima, corresponding to alignment along the length of the stripes. Hence, flexoelectricity serves as an aligning effect in the presence of the patterned surface. Also displayed in the figure is the free energy of a perturbation solution similar to the one derived in [5] (note, a different unit convention and sign error exists in [5]). There, the perturbation solution is valid for a single semi-infinite planar-vertical junction. In the numerical

computation, the director profile for the striped cell consists of four junctions per unit cell. Thus, we approximate the perturbation by adding the mirror image and doubling. If the junctions are well separated from each other, the cell thickness is larger than the penetration depth of the nematic, and the length of the surface planar-vertical transition is very small, this is a valid approximation. Even with this limitation, the computed energies trace the characteristics of the perturbation solution quite closely, verifying the alignment influence of flexoelectricity. When considering internally induced electric fields in the presence of nano-patterned boundaries, the algorithm's computed free energies capture the qualitative prediction from the perturbation solution, but do so with a quantitative accuracy that is not readily matched by perturbation techniques.

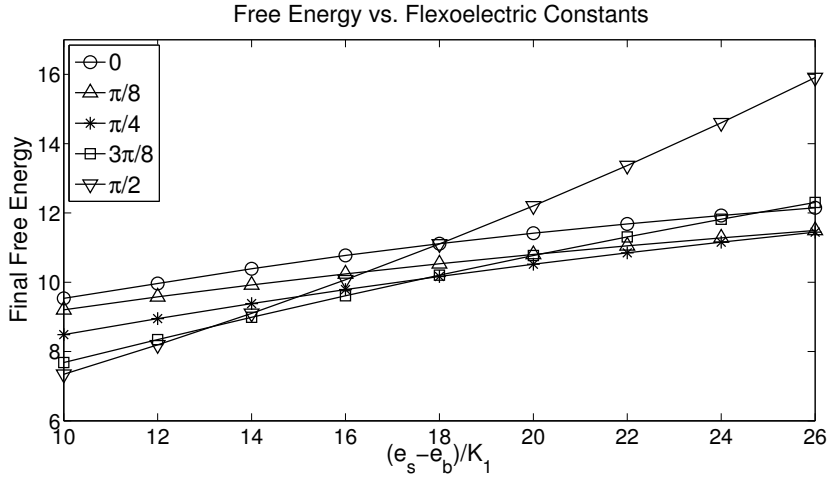


Figure 6.13: Final flexoelectric energies with nano-patterned boundary conditions for varying flexoelectric constants e_s and e_b. Each line corresponds to a different φ value.

For the second experiment, $\epsilon_{\parallel} = 7$ and $\epsilon_{\perp} = 7$. By including anisotropic elastic constants, it is possible to promote alignment perpendicular to the stripes, if $K_1, K_3 < K_2$, or parallel to the length of the stripes, if $K_1, K_3 > K_2$. We use $K_1 = K_3 = 1$ and $K_2 = 4$ to select perpendicular alignment and simulate the configurations with $\varphi \in \{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}\}$ for varying values of the flexoelectric constants; the results are displayed in Figure 6.13. As can be seen, for $(e_s - e_b)/K_1 = 10$, the overall minimum of the free energy lies at an azimuthal angle $\varphi = \pi/2$ as expected.

As the flexoelectric parameter is increased however, the configurations with different azimuthal angle increase at different rates; for example at a critical value of $(e_s - e_b)/K_1 \approx 17.5$, the solutions for $\varphi = 0$ and $\varphi = \pi/2$ become degenerate. Hence, as the strength of the flexoelectric effect is increased, the azimuthal angle corresponding to the ground state gradually rotates because flexoelectricity and elastic anisotropy favor opposing configurations. The phenomenon is important for applications because it may lead to multiple stable configurations in some regions of the parameter space, or a significant renormalization of the anchoring behavior for materials with large flexoelectric response. These phenomena allow engineers to control the ground states and, potentially, the switching response by adjusting the pattern. The energy-minimization approach offers an efficient predictive tool for identifying the parameters that lead to the desired effect.

Chapter 7

Three-Dimensional Problems with Patterned Surfaces

In addition to the development of the energy-minimization approach discussed in previous chapters, we consider solving partial differential equations (PDEs) derived through the Euler-Lagrange equations for thin liquid crystal films with geometrically patterned substrates. While the energy-minimization approach considered above can be used to efficiently model these configurations with high precision, the nonlinear PDEs associated with the particular configurations considered in this chapter are clean enough that they can be analyzed and numerically solved directly.

We specifically investigate a three-dimensional (3D) film where the substrates in the z-direction represent a lattice of circularly patterned surfaces. Thin film cells have been considered in the case of square patterning in [3, 4]. There, the patterned surfaces give rise to two energetically degenerate configurations, which suggests the potential for bistability similar to that seen in the post-aligned bistable display [26]. Such bistablilities have important energy and design implications for devices which utilize liquid crystal birefringence to control light propagation.

Throughout this chapter, the director is parameterized as a function of spherical coordinates ϕ and θ such that the components of $\mathbf{n}$ are written

$$n_1 = \cos\theta \sin\phi, \qquad n_2 = \cos\theta\cos\phi, \qquad n_3 = \sin\theta,$$

and θ represents the angular rise of the director off the xy-plane and ϕ is the clockwise rotation from the positive y-axis. This coordinate system is primarily chosen for convenience in the calculations.

In the configurations to be studied, we consider anisotropic Frank constants such that $K_1 = K_3 \neq K_2$. As noted in [3], the one-constant approximation does not

accurately capture liquid crystal alignment behaviors in the presence of patterned surfaces. On the other hand, the relationship of the Frank constants outlined above has been used successfully to study liquid crystal configurations induced by geometrically patterned substrates [3, 4, 6].

In this chapter, we introduce the micron length scale $\sigma = 10^{-6}m$ corresponding to the width of the cell substrate in the x-direction. Thus, using the change of variable, $\mathbf{x} = \sigma\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} \in \mathbb{R}^3$ is dimensionless, the Frank-Oseen elastic free energy, ignoring the null-Lagrangian, is written

$$\tilde{\mathcal{F}}_B = \frac{1}{2}\sigma \int_{\tilde{\Omega}} K_1(\nabla \cdot \tilde{\mathbf{n}})^2 + K_2(\tilde{\mathbf{n}} \cdot \nabla \times \tilde{\mathbf{n}})^2 + K_3|\tilde{\mathbf{n}} \times \nabla \times \tilde{\mathbf{n}}|^2 \, d\tilde{V}.$$

Denote $\tau = K_2/K_1$. Dividing the free energy by the value $K_1\sigma$ and dropping the tilde notation, the bulk free-energy is characterized by

$$\mathcal{F}_B = \frac{1}{2} \int_{\Omega} (\nabla \cdot \mathbf{n})^2 + \tau(\mathbf{n} \cdot \nabla \times \mathbf{n})^2 + |\mathbf{n} \times \nabla \times \mathbf{n}|^2 \, dV, \tag{7.1}$$

on a domain Ω. In the numerical experiments, we consider the rectangular cuboid domain, $\Omega = [0, 1] \times [0, 1] \times [-z_0, z_0]$. In addition, we consider a twisting ϕ-profile via the ansatz that $\phi(x, y, z) = \phi_0 + z\phi_1$, for constants ϕ_0 and ϕ_1. With this assumption and the energy expression in (7.1), we apply the relevant equilibrium equation in [117, pg. 38] to produce a partial differential equation in terms of the unknown, θ. This equation for θ is written

$$-\nabla \cdot A\nabla\theta - \frac{1}{2}\phi_1^2 \sin(2\theta)\left(\tau - \beta\cos(2\theta)\right) = 0 \qquad \text{on } \Omega, \tag{7.2}$$

where

$$A = \begin{bmatrix} \frac{1}{2}\big(\alpha - \beta\cos(2\phi(z))\big) & \frac{1}{2}\beta\sin(2\phi(z)) & 0 \\ \frac{1}{2}\beta\sin(2\phi(z)) & \frac{1}{2}\big(\alpha + \beta\cos(2\phi(z))\big) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{7.3}$$

such that $\alpha = 1 + \tau$ and $\beta = 1 - \tau$. Observe that since $\tau = K_2/K_1$, $\tau > 0$ by Ericksen's inequalities. Thus, the relevant Euler-Lagrange equation in (7.2) represents a nonlinear anisotropic reaction-diffusion equation with diffusion tensor, A, written in (7.3). This tensor is an anisotropic scaling that depends on the director angle and varies spatially with z. Note that in the special case of $\phi_1 = 0$, there is no twist in the director with variation of z and (7.2) simplifies to a linear anisotropic diffusion problem.

This approach to computing the equilibrium configuration differs from the energy-minimization method discussed in previous chapters, in that we use the Euler-Lagrange equations to produce a PDE from the Frank-Oseen elastic free energy rather than minimizing the free energy directly. This is motivated by the simplicity and clarity of the equation to be solved.

In the following sections, among other cases, we consider the outlined PDE with periodic boundary conditions such that

$$\theta(0, y, z) = \theta(1, y, z), \tag{7.4}$$

$$\theta(x, 0, z) = \theta(x, 1, z), \tag{7.5}$$

on the cuboid domain. For the boundaries in the z-direction, Dirichlet and Robin boundary conditions are considered separately. We use a finite-element method with Newton linearizations to deal with the nonlinear reaction term and show that, under certain assumptions, the intermediate discrete linearized systems are well-posed.

7.1 Variational Form and Linearization

To use finite elements, we multiply the equation by a test function, w, integrate over the domain, and use integration by parts to obtain the variational form

$$-\int_{\partial\Omega} \big((A\nabla\theta)\cdot\nu\big) w\, dS + \int_\Omega (A\nabla\theta)\cdot\nabla w\, dV - \frac{1}{2}\phi_1^2 \int_\Omega \sin(2\theta)\,(\tau - \beta\cos(2\theta))\, w\, dV = 0.$$

Here, ν is the outward facing normal for $\partial\Omega$. Denoting the boundary integral as $\langle\cdot,\cdot\rangle_{\partial\Omega}$ and the induced norm space $L^2(\partial\Omega)$, the above equation is written more compactly as

$$-\langle w, (A\nabla\theta)\cdot\nu\rangle_{\partial\Omega} + \langle A\nabla\theta, \nabla w\rangle_0 - \frac{1}{2}\phi_1^2\langle \sin(2\theta)\left(\tau - \beta\cos(2\theta)\right), w\rangle_0 = 0. \tag{7.6}$$

Since the third term in (7.6) is nonlinear in θ, we apply Newton's method and linearize the variational form around a current iterate, θ_k, with update, γ, such that $\theta_{k+1} = \theta_k + \gamma$. Then, denoting the variational form as a functional in θ, $\mathcal{F}(\theta)$, we find the Gâteaux derivative in the direction of γ. This is denoted $\mathcal{F}'(\theta_k)[\gamma]$. Setting the derivative equal to the nonlinear residual, $-\mathcal{F}(\theta_k)$, we then solve for the update γ.

In computing the functional derivative, we use the fact that the Gâteaux derivative of $\sin(2\theta)$ is $2\gamma\cos(2\theta)$ and the derivative of $\cos(2\theta)$ is $-2\gamma\sin(2\theta)$. Thus, the full Gâteaux derivative is

$$\begin{aligned}
\mathcal{F}'(\theta_k)[\gamma] &= -\langle w, (A\nabla\gamma)\cdot\nu\rangle_{\partial\Omega} + \langle A\nabla\gamma, \nabla w\rangle_0 \\
&\qquad - \frac{1}{2}\phi_1^2\langle \left(2\gamma\cos(2\theta_k)\left(\tau - \beta\cos(2\theta_k)\right) + 2\gamma\beta\sin^2(2\theta_k)\right), w\rangle_0 \\
&= -\langle w, (A\nabla\gamma)\cdot\nu\rangle_{\partial\Omega} + \langle A\nabla\gamma, \nabla w\rangle_0 \\
&\qquad - \frac{1}{2}\phi_1^2\langle 2\gamma\beta\left(\sin^2(2\theta_k) - \cos^2(2\theta_k)\right) + 2\gamma\tau\cos(2\theta_k), w\rangle_0 \\
&= -\langle w, (A\nabla\gamma)\cdot\nu\rangle_{\partial\Omega} + \langle A\nabla\gamma, \nabla w\rangle_0 \\
&\qquad + \phi_1^2\langle \left(\beta\cos(4\theta_k) - \tau\cos(2\theta_k)\right)\gamma, w\rangle_0.
\end{aligned}$$

Forming the full linearization system, $\mathcal{F}'(\theta_k)[\gamma] = -\mathcal{F}(\theta_k)$, with no boundary conditions applied yields

$$\begin{aligned}
&-\langle w, (A\nabla\gamma)\cdot\nu\rangle_{\partial\Omega} + \langle A\nabla\gamma, \nabla w\rangle_0 + \phi_1^2\langle \left(\beta\cos(4\theta_k) - \tau\cos(2\theta_k)\right)\gamma, w\rangle_0 = \\
&\qquad \langle w, (A\nabla\theta_k)\cdot\nu\rangle_{\partial\Omega} - \langle A\nabla\theta_k, \nabla w\rangle_0 + \frac{1}{2}\phi_1^2\langle \sin(2\theta_k)\left(\tau - \beta\cos(2\theta_k)\right), w\rangle_0.
\end{aligned} \tag{7.7}$$

7.1.1 Uniform Symmetric Positive Definiteness of A

Consider the coefficient matrix, A, described in (7.3) and define the constants $\eta = \min(\tau, 1)$ and $\Lambda = \max(\tau, 1)$.

Lemma 7.1.1 *The matrix, A, is USPD with upper and lower bounds Λ and η, independent of Ω, respectively.*

Proof: Clearly A is symmetric for all $\mathbf{x} \in \Omega$. Moreover, the eigenvalues of A are computed to be $\lambda_1 = 1$, $\lambda_2 = \frac{\alpha-\beta}{2} = \tau$, and $\lambda_3 = \frac{\alpha+\beta}{2} = 1$. Note that $\lambda_i > 0$ for $i = 1, 2, 3$ and are bounded as $0 < \eta \leq \lambda_i \leq \Lambda$, where η and Λ are constants, independent of Ω. Using standard functional analysis arguments, it is straightforward to show that

$$0 < \eta \leq \frac{\xi^T A(\mathbf{x})\xi}{\xi^T \xi} \leq \Lambda, \qquad \forall x \in \Omega,\ \xi \in \mathbb{R}^3.$$

Therefore, A is uniformly symmetric positive definite with lower and upper bounds η and Λ, respectively. □

Note that the USPD bounds depend only on the value of τ.

7.2 Well-Posedness for Dirichlet Boundary Conditions

In this section, we consider the existence and uniqueness of solutions to the linearized systems with full Dirichlet boundary conditions on a general domain or the mixed periodic and Dirichlet boundary conditions on the cuboid domain, discussed above. In the mixed boundary condition case this implies that we assume

$$\theta(x, y, -z_0) = g_1(x, y), \qquad \theta(x, y, z_0) = g_2(x, y),$$

for functions g_1 and g_2.

Considering the general linearized system in (7.7), note that in the case of full Dirichlet boundary conditions the variations, γ, and test functions, w, have zero Dirichlet boundary conditions. For the mixed conditions, w and γ have matching

periodic boundaries in the $x-$ and y-directions and zero Dirichlet boundary conditions in the z-direction. Thus, on the Dirichlet boundaries, the surface integrals in (7.7) are zero. For the periodic boundaries, the outward facing normals, ν, have opposite sign, therefore the boundary integrals cancel. Thus, for these boundary conditions, the linearization system simplifies to

$$\begin{aligned}\langle A\nabla\gamma, \nabla w\rangle_0 + \phi_1^2\langle(\beta\cos(4\theta_k) - \tau\cos(2\theta_k))\,\gamma, w\rangle_0 \\ = -\langle A\nabla\theta_k, \nabla w\rangle_0 + \frac{1}{2}\phi_1^2\langle\sin(2\theta_k)\,(\tau - \beta\cos(2\theta_k))\,, w\rangle_0.\end{aligned}$$

Let

$$\begin{aligned}a(\gamma, w) &= \langle A\nabla\gamma, \nabla w\rangle_0 + \phi_1^2\langle\big(\beta\cos(4\theta_k) - \tau\cos(2\theta_k)\big)\gamma, w\rangle_0, \\ F(w) &= \frac{1}{2}\phi_1^2\langle\sin(2\theta_k)(\tau - \beta\cos(2\theta_k)), w\rangle_0 - \langle A\nabla\theta_k, \nabla w\rangle_0.\end{aligned}$$

In the following, we assume full Dirichlet boundary conditions such that $\gamma, w \in H_0^1(\Omega)$ and $\theta_k \in H^1(\Omega)$ with proper boundary conditions. However, the proofs to follow are equally applicable to the mixed boundary conditions. For brevity, we drop the γ notation and simply address the bilinear form $a(u, w)$ in the theory.

In the first instance, we are primarily concerned with establishing tight bounds on the term $\beta\cos(4\theta_k) - \tau\cos(2\theta_k)$. Note that since $\beta\cos(4\theta_k) - \tau\cos(2\theta_k)$ is periodic and bounded, it has a periodically reoccurring global minimum and maximum. The following lemma characterizes these extrema in terms of β and τ.

Lemma 7.2.1 *If* $\frac{4}{5} < \tau < \frac{4}{3}$, *let* $\mu_1 = \max(1, |1 - 2\tau|)$ *and* $\mu_2 = 1 - 2\tau$. *Otherwise, let* $\mu_1 = \max\left(\left|\frac{8\beta^2+\tau^2}{8\beta}\right|, 1, |1 - 2\tau|\right)$ *and* $\mu_2 = \min\left(-\left(\frac{8\beta^2+\tau^2}{8\beta}\right), 1 - 2\tau\right)$. *Then*

$$\mu_2 \le \beta\cos(4\theta_k) - \tau\cos(2\theta_k), \qquad |\beta\cos(4\theta_k) - \tau\cos(2\theta_k)| \le \mu_1.$$

Proof: Computing the derivative with respect to θ_k,

$$\frac{\partial}{\partial\theta_k}(\beta\cos(4\theta_k) - \tau\cos(2\theta_k)) = 2\tau\sin(2\theta_k) - 4\beta\sin(4\theta_k).$$

Setting the derivative equal to zero implies that a set of critical points is characterized by

$$\frac{2\beta}{\tau} = \frac{\sin(2\theta_k)}{\sin(4\theta_k)} = \frac{1}{2}\sec(2\theta_k).$$

This implies that, at these critical points, $\cos(2\theta_k) = \frac{\tau}{4\beta} = \frac{\tau}{4(1-\tau)}$. These critical points are only feasible if $\left|\frac{\tau}{4(1-\tau)}\right| \leq 1$, which occurs for $0 < \tau \leq \frac{4}{5}$ and $\tau \geq \frac{4}{3}$. Observe that $\beta\cos(4\theta_k) = 2\beta\cos^2(2\theta_k) - \beta$. Thus, if τ satisfies the feasibility bounds, at the critical points

$$\begin{aligned}\beta\cos(4\theta_k) - \tau\cos(2\theta_k) &= 2\beta\cos^2(2\theta_k) - \beta - \tau\cos(2\theta_k)\\ &= 2\beta\left(\frac{\tau}{4\beta}\right)^2 - \beta - \tau\left(\frac{\tau}{4\beta}\right)\\ &= -\left(\frac{8\beta^2+\tau^2}{8\beta}\right).\end{aligned}$$

The remaining critical points are those θ_k such that

$$\sin(2\theta_k) = \sin(4\theta_k) = 0.$$

Note that $\sin(2\theta_k) = 0$ for $\theta_k = \frac{n\pi}{2}$ and $n \in \mathbb{Z}$, while $\sin(4\theta_k) = 0$ for $\theta_k = \frac{n\pi}{4}$ and $n \in \mathbb{Z}$. Therefore, the additional critical points are $\theta_k = \frac{n\pi}{2}$ for $n \in \mathbb{Z}$. At these critical points,

$$\begin{aligned}\beta\cos(4\theta_k) - \tau\cos(2\theta_k) &= \beta\cos(2n\pi) - \tau\cos(n\pi)\\ &= \beta - (-1)^n\tau\\ &= \begin{cases}1 & n \text{ odd},\\ 1-2\tau & n \text{ even}.\end{cases}\end{aligned}$$

Thus, the minimum and absolute maximum values, depending on constants β and τ, are given by μ_2 and μ_1, respectively. □

7.2.1 Continuity

For this section, we prove that $a(u,w)$ and $F(w)$ are continuous bilinear and linear forms on $H_0^1(\Omega)$, respectively.

Lemma 7.2.2 *The bilinear form is bounded as*

$$|a(u,w)| \leq (\Lambda + \phi_1^2\mu_1)\|u\|_1\|w\|_1,$$

for $u, w \in H_0^1(\Omega)$.

Proof: By the triangle inequality,

$$|a(u,w)| \leq |\langle A\nabla u, \nabla w\rangle_0| + \phi_1^2|\langle\big(\beta\cos(4\theta_k) - \tau\cos(2\theta_k)\big)u, w\rangle_0|.$$

From Lemma 7.1.1 and the Cauchy-Schwarz inequality,

$$\begin{aligned} |\langle A\nabla u, \nabla w\rangle_0| &\leq \|A\nabla u\|_0\|\nabla w\|_0 \\ &\leq \Lambda\|\nabla u\|_0\|\nabla w\|_0 \leq \Lambda\|u\|_1\|w\|_1. \end{aligned} \tag{7.8}$$

Using the bound from Lemma 7.2.1,

$$\begin{aligned} \phi_1^2|\langle\big(\beta\cos(4\theta_k) - \tau\cos(2\theta_k)\big)u, w\rangle_0| &\leq \phi_1^2\|\big(\beta\cos(4\theta_k) - \tau\cos(2\theta_k)\big)u\|_0\|w\|_0 \\ &= \phi_1^2\left(\int_\Omega \big(\beta\cos(4\theta_k) - \tau\cos(2\theta_k)\big)^2u^2\,dV\right)^{1/2}\|w\|_0 \\ &\leq \phi_1^2\mu_1\|u\|_0\|w\|_0 \\ &\leq \phi_1^2\mu_1\|u\|_1\|w\|_1. \end{aligned} \tag{7.9}$$

Using inequalities (7.8) and (7.9) this implies that $|a(u,w)| \leq (\Lambda + \phi_1^2\mu_1)\|u\|_1\|w\|_1$.□

Lemma 7.2.3 *Assume that* $\theta_k \in H^1(\Omega)$. *Letting* $\mu_3 = |\tau| + |\beta|$ *implies that*

$$|F(w)| \leq (\Lambda\|\nabla\theta_k\|_0 + \frac{1}{2}\phi_1^2\mu_3|\Omega|^{1/2})\|w\|_1.$$

Proof: Note that

$$\begin{aligned}|\sin(2\theta_k)(\tau - \beta\cos(2\theta_k))| &\leq |\sin(2\theta_k)||\tau - \beta\cos(2\theta_k)| \\ &\leq |\tau - \beta\cos(2\theta_k)| \\ &\leq |\tau| + |\beta||\cos(2\theta_k)| \\ &\leq |\tau| + |\beta| = \mu_3.\end{aligned}$$

Then via applications of the triangle and Cauchy-Schwarz inequalities

$$\begin{aligned}|F(w)| &\leq |\langle A\nabla\theta_k, \nabla w\rangle_0| + \frac{1}{2}\phi_1^2|\langle\sin(2\theta_k)(\tau - \beta\cos(2\theta_k), w\rangle_0| \\ &\leq \|A\nabla\theta_k\|_0\|\nabla w\|_0 + \frac{1}{2}\phi_1^2\|\sin(2\theta_k)(\tau - \beta\cos(2\theta_k)\|_0\|w\|_0 \\ &\leq \Lambda\|\nabla\theta_k\|_0\|\nabla w\|_0 + \frac{1}{2}\phi_1^2\mu_3|\Omega|^{1/2}\|w\|_0 \\ &\leq (\Lambda\|\nabla\theta_k\|_0 + \frac{1}{2}\phi_1^2\mu_3|\Omega|^{1/2})\|w\|_1. \qquad \square\end{aligned}$$

7.2.2 Coercivity

Here, we demonstrate that $a(u, w)$ is a coercive bilinear form on $H_0^1(\Omega)$. That is, there exists a $C_z > 0$ such that

$$a(w, w) \geq C_z\|w\|_1^2, \qquad \forall w \in H_0^1(\Omega).$$

Lemma 7.2.4 *Assume that Ω is bounded with a Lipschitz boundary and $w \in H_0^1(\Omega)$. If $\frac{\eta}{C_f+1} + \phi_1^2\mu_2 > 0$, for the Poincaré-Friedrichs' constant $C_f > 0$, then $a(w, w)$ is a coercive bilinear form.*

Proof: Lemma 7.1.1 implies that A is USPD and, therefore,

$$\langle A\nabla w, \nabla w\rangle_0 \geq \eta\langle\nabla w, \nabla w\rangle_0.$$

By the classical Poincaré-Friedrichs' inequality [56], there exists a $C_f > 0$ such that

$$\langle w, w\rangle_0 \le C_f \langle \nabla w, \nabla w\rangle_0.$$

This implies that

$$\begin{aligned}\|w\|_1^2 &= \langle w, w\rangle_0 + \langle \nabla w, \nabla w\rangle_0 \\ &\le C_f \langle \nabla w, \nabla w\rangle_0 + \langle \nabla w, \nabla w\rangle_0 \\ &= (C_f + 1)\langle \nabla w, \nabla w\rangle_0.\end{aligned}$$

Thus,

$$\langle A\nabla w, \nabla w\rangle_0 \ge \eta \langle \nabla w, \nabla w\rangle_0 \ge \frac{\eta}{C_f + 1}\|w\|_1^2. \tag{7.10}$$

Using Lemma 7.2.1 and the fact that $\phi_1^2 \ge 0$,

$$\begin{aligned}\phi_1^2\langle\big(\beta\cos(4\theta_k) - \tau\cos(2\theta_k)\big)w, w\rangle_0 &= \phi_1^2 \int_\Omega \big(\beta\cos(4\theta_k) - \tau\cos(2\theta_k)\big)w^2\, dV \\ &\ge \mu_2\phi_1^2\langle w, w\rangle_0. \end{aligned}\tag{7.11}$$

Note that from Lemma 7.2.1, $\mu_2 \le 0$ for any choice of τ. If $\mu_2 \le 0$, then $\mu_2\phi_1^2\langle w, w\rangle_0 \ge \mu_2\phi_1^2\|w\|_1^2$ and the inequalities (7.10) and (7.11) imply that

$$a(w, w) \ge \frac{\eta}{C_f + 1}\|w\|_1^2 + \phi_1^2\mu_2\|w\|_1^2 = \left(\frac{\eta}{C_f + 1} + \phi_1^2\mu_2\right)\|w\|_1^2.$$

Thus, taking $C_z = \left(\frac{\eta}{C_f+1} + \phi_1^2\mu_2\right) > 0$, we establish the coercivity of $a(\cdot, \cdot)$. □

With the continuity of $F(w)$ and $a(u, w)$ established in Lemmas 7.2.2 and 7.2.3, as well as the coercivity of $a(u, w)$ in Lemma 7.2.4, the Lax-Milgram theorem [16] implies that the linearizations are well posed for all θ_k.

7.3 Well-Posedness for Robin Boundary Conditions

In the presence of non-Dirichlet boundary conditions, a surface free energy capturing the effects of substrate interaction with the nematic rods is introduced. The equilibrium configuration of a liquid crystal sample occurs at the minimum total of the volume and surface free energies. These boundary conditions are referred to in physics literature as weak anchoring conditions [37, 117]. Since we mean to study surfaces with patterns that encourage 2D geometric structures on the substrates, we consider a harmonic anchoring potential, similar to that given in [3], such that the surface free energy is computed as

$$\mathcal{F}_S = \frac{1}{2L_\theta} \int_{\partial\Omega_1} (\theta - \theta_e)^2 \, dS,$$

where $L_\theta = \frac{K_1}{W_\theta \sigma} > 0$ is a parameter associated with the polar anchoring constant W_θ and $\partial\Omega_1$ is the boundary component on which weak anchoring applies. Note that the appearance of the constant σ corresponds to the length scaling performed above for the bulk free energy. In this section, we exclusively consider the rectangular cuboid domain $\Omega = [0,1] \times [0,1] \times [-z_0, z_0]$ with mixed periodic and weak-anchoring boundary conditions.

Computing the equilibrium boundary conditions as in [3, 67, 117], we arrive at the Robin boundary conditions

$$\pm L_\theta \frac{\partial \theta}{\partial z} + \theta = \theta_e, \qquad \text{for } z = \pm z_0.$$

Note that as L_θ tends to 0, the conditions above approach the mixed periodic and Dirichlet boundary conditions discussed in the previous sections.

With the change of boundary conditions, the surface integrals of the linearization in Equation (7.7) are no longer negligible at the boundaries in the z-direction. Consider the iterate $\theta_{k+1} = \theta_k + \gamma$ and let ν_3 denote the z-component of the outward normal vector for $\partial\Omega$ (for $z = \pm z_0$ this implies that $\nu_3 = \pm 1$). The Robin conditions

dictate that at $z = \pm z_0$,

$$\nu_3 L_\theta \frac{\partial \theta_{k+1}}{\partial z} + \theta_{k+1} = \nu_3 L_\theta \frac{\partial (\theta_k + \gamma)}{\partial z} + \theta_k + \gamma = \theta_e.$$

This implies the equivalent requirement that

$$\nu_3 \frac{\partial (\theta_k + \gamma)}{\partial z} = \frac{1}{L_\theta}\big(\theta_e - \theta_k - \gamma\big). \tag{7.12}$$

Note that

$$\langle w, (A\nabla(\theta_k + \gamma)) \cdot \nu \rangle_{\partial\Omega} = \langle w, (A\nabla\theta_k) \cdot \nu \rangle_{\partial\Omega} + \langle w, (A\nabla\gamma) \cdot \nu \rangle_{\partial\Omega}. \tag{7.13}$$

Denote the surfaces at $z = \pm z_0$ as $\partial\Omega_1$. Due to the periodic boundary conditions imposed for γ and θ_k, the surface integral in (7.13) on $\partial\Omega \backslash \partial\Omega_1$ is zero. Thus

$$\begin{aligned} \langle w, (A\nabla(\theta_k + \gamma)) \cdot \nu \rangle_{\partial\Omega} &= \int_{\partial\Omega_1} w\Big(\big(A\nabla(\theta_k + \gamma)\big) \cdot \nu\Big)\, dS \\ &= \int_{\partial\Omega_1} w \left(\frac{\partial(\theta_k + \gamma)}{\partial z}\right) \nu_3 \, dS. \end{aligned} \tag{7.14}$$

In order to weakly enforce the Robin boundary conditions for θ_{k+1}, we replace the normal derivative in (7.14) with the boundary data on the right-hand-side of (7.12) such that

$$\int_{\partial\Omega_1} w \left(\frac{\partial(\theta_k + \gamma)}{\partial z}\right) \nu_3 \, dS = \frac{1}{L_\theta} \int_{\partial\Omega_1} w(\theta_e - \theta_k - \gamma)\, dS.$$

Therefore, the variational system for the Newton iterations with Robin boundary conditions is written

$$\begin{aligned} &\frac{1}{L_\theta} \int_{\partial\Omega_1} w\gamma \, dS + \langle A\nabla\gamma, \nabla w\rangle_0 + \phi_1^2 \langle (\beta\cos(4\theta_k) - \tau\cos(2\theta_k))\,\gamma, w\rangle_0 = \\ &\frac{1}{L_\theta} \int_{\partial\Omega_1} w(\theta_e - \theta_k)\, dS - \langle A\nabla\theta_k, \nabla w\rangle_0 + \frac{1}{2}\phi_1^2 \langle \sin(2\theta_k)\,(\tau - \beta\cos(2\theta_k))\,, w\rangle_0. \end{aligned} \tag{7.15}$$

Again, we define bilinear forms

$$a(\gamma, w) = \frac{1}{L_\theta}\int_{\partial\Omega_1} w\gamma\, dS + \langle A\nabla\gamma, \nabla w\rangle_0 + \phi_1^2\langle(\beta\cos(4\theta_k) - \tau\cos(2\theta_k))\,\gamma, w\rangle_0,$$
$$F(w) = \frac{1}{L_\theta}\int_{\partial\Omega_1} w(\theta_e - \theta_k)\, dS - \langle A\nabla\theta_k, \nabla w\rangle_0 + \frac{1}{2}\phi_1^2\langle\sin(2\theta_k)\,(\tau - \beta\cos(2\theta_k))\,, w\rangle_0.$$

For these iterations, we solve for $\gamma \in H^1(\Omega)$ satisfying $a(\gamma, w) = F(w)$ for all $w \in H^1(\Omega)$, where γ and w satisfy periodic boundary conditions on the appropriate boundaries. As in the previous analysis, we drop the γ notation and simply address the bilinear for $a(u, w)$.

Lemma 7.3.1 *Say that Ω is bounded, Lipschitz domain, then $a(u, w)$ is a continuous bilinear form.*

Proof: Let

$$\tilde{a}(u, w) = \langle A\nabla\gamma, \nabla w\rangle_0 + \phi_1^2\langle(\beta\cos(4\theta_k) - \tau\cos(2\theta_k))\,\gamma, w\rangle_0.$$

Lemma 7.2.2 implies that $|\tilde{a}(u, w)| \le \left(\Lambda + \phi_1^2\mu_1\right)\|u\|_1\|w\|_1$. By the triangle inequality,

$$|a(u, w)| \le \left|\frac{1}{L_\theta}\int_{\partial\Omega_1} uw\, dS\right| + |\tilde{a}(u, w)|.$$

Furthermore,

$$\left|\frac{1}{L_\theta}\int_{\partial\Omega_1} uw\, dS\right| \le \frac{1}{L_\theta}\|u\|_{L^2(\partial\Omega_1)}\|w\|_{L^2(\partial\Omega_1)} \le \frac{1}{L_\theta}\|u\|_{L^2(\partial\Omega)}\|w\|_{L^2(\partial\Omega)},$$

By the trace theorem, [16, Theorem 1.6.6], there exists a $C_T > 0$ such that

$$\|u\|_{L^2(\partial\Omega)} \le C_T\|u\|^{1/2}_{L^2(\Omega)}\|u\|^{1/2}_1 \le C_T\|u\|_1.$$

This implies that

$$\left|\frac{1}{L_\theta}\int_{\partial\Omega_1} uw\, dS\right| \le \frac{C_T^2}{L_\theta}\|u\|_1\|w\|_1.$$

Hence, $|a(u, w)| \le \left(\frac{C_T^2}{L_\theta} + \Lambda + \phi_1^2\mu_1\right)\|u\|_1\|w\|_1$. □

Lemma 7.3.2 *Assume that Ω is a bounded, Lipschitz domain, $\theta_k \in H^1(\Omega)$, and $\theta_e \in L^2(\partial\Omega)$. Then $F(w)$ is a bounded linear functional.*

Proof: Define the linear functional,

$$\tilde{F}(w) = \langle A\nabla\theta_k, \nabla w\rangle_0 + \frac{1}{2}\phi_1^2\langle \sin(2\theta_k)\left(\tau - \beta\cos(2\theta_k)\right), w\rangle_0.$$

From Lemma 7.2.3,

$$|\tilde{F}(w)| \le (\Lambda\|\nabla\theta_k\|_0 + \frac{1}{2}\phi_1^2\mu_3|\Omega|^{1/2})\|w\|_1.$$

By the triangle inequality,

$$|F(w)| \le \frac{1}{L_\theta}\left|\int_{\partial\Omega_1} w\theta_e\, dS\right| + \frac{1}{L_\theta}\left|\int_{\partial\Omega_1} w\theta_k\, dS\right| + |\tilde{F}(w)|$$

Applying the trace theorem [16, Theorem 1.6.6], as above, we arrive at the bounds

$$\left|\int_{\partial\Omega_1} w\theta_k\, dS\right| \le C_T\|\theta_k\|_{L^2(\partial\Omega_1)}\|w\|_1,$$
$$\left|\int_{\partial\Omega_1} w\theta_e\, dS\right| \le C_T\|\theta_e\|_{L^2(\partial\Omega_1)}\|w\|_1.$$

Thus,

$$|F(w)| \le \left(C_T(\|\theta_k\|_{L^2(\partial\Omega_1)} + \|\theta_e\|_{L^2(\partial\Omega_1)}) + \Lambda\|\nabla\theta_k\|_0 + \frac{1}{2}\phi_1^2\mu_3|\Omega|^{1/2}\right)\|w\|_1. \qquad \square$$

Finally, we consider the coercivity of $a(u, w)$.

Lemma 7.3.3 *Say that Ω is a bounded, Lipschitz domain and $\phi_1 \ge 0$. If $(\alpha_0 + \phi_1^2\mu_2) > 0$, for $\alpha_0 > 0$ defined in the proof below, then $a(w, w)$ is a coercive bilinear form for all $w \in H^1(\Omega)$ with appropriate boundary conditions.*

Proof: By Lemmas 7.1.1 and 7.2.1,

$$a(w, w) \ge \frac{1}{L_\theta}\int_{\partial\Omega_1} w^2\, dS + \eta\langle\nabla w, \nabla w\rangle_0 + \phi_1^2\mu_2\|w\|_1^2 := b(w, w) + \phi_1^2\mu_2\|w\|_1^2.$$

Applying a standard compactness argument for coercivity of the variational formulation for the Poisson equation with mixed periodic and Robin boundary conditions to $b(w,w)$ [1, 16, 87], there exists an $\alpha_0 > 0$ such that $b(w,w) \geq \alpha_0 \|w\|_1^2$. Thus, if $\alpha_0 + \phi_1^2 \mu_2 > 0$, then $a(w,w)$ is a coercive bilinear form. □

7.4 Numerical Results

For the numerical computations, we consider the domain $\Omega = [0,1]\times[0,1]\times[-z_0, z_0]$. In order to solve the variational systems outlined in the sections above, we use Q_1 finite elements to discretize and solve for the update γ. In the computations to follow, $\tau = 1/2$, which fixes $\alpha = 3/2$ and $\beta = 1/2$. Observe that $\tau = 1/2$ implies that the value of K_2 is half that of K_1. Define the function θ_e such that

$$\theta_e(x,y) = \begin{cases} 0 & \left(x-\frac{1}{2}\right)^2 + \left(y-\frac{1}{2}\right)^2 > r^2, \\ \frac{\pi}{2} & \left(x-\frac{1}{2}\right)^2 + \left(y-\frac{1}{2}\right)^2 \leq r^2, \end{cases}$$

where r is a positive constant. Throughout this section, we apply the periodic boundary conditions written in (7.4) and (7.5). For the numerical experiments incorporating Dirichlet boundary conditions, we use the additional condition

$$\theta(x,y,-z_0) = \theta(x,y,z_0) = \theta_e(x,y).$$

In the case of Robin boundary conditions, we have the equations

$$\pm L_\theta \frac{\partial \theta}{\partial z} + \theta = \theta_e, \qquad \text{for } z = \pm z_0.$$

Note that, as defined, θ_e is discontinuous at the boundary of a circle with radius r, centered at the point $(1/2, 1/2)$. However, since we use Lagrangian finite elements, the discontinuity is represented by a continuous function that more closely approximates the jump as the mesh is refined. Moreover, this guarantees that for the discrete variational system, $\theta_k \in H^1(\Omega)$. Thus, the well-posedness properties of

the linearizations outlined above are applicable to the discrete variational systems considered in this section.

These boundary conditions model the presence of a surface with periodic circular patterning on substrates in the z-direction. We are interested in the energy behavior of configurations resulting from such surfaces. In order to more accurately model the circular transition, adaptive refinement is used to finely resolve the regions at the circular boundary. Mesh cells are flagged and refined based on which cells contain the highest free energy. As shown in Figure 7.1, this results in high refinement around the circular region due to the abrupt behavioral transition of the liquid crystal structure, which induces inherently elevated free energy.

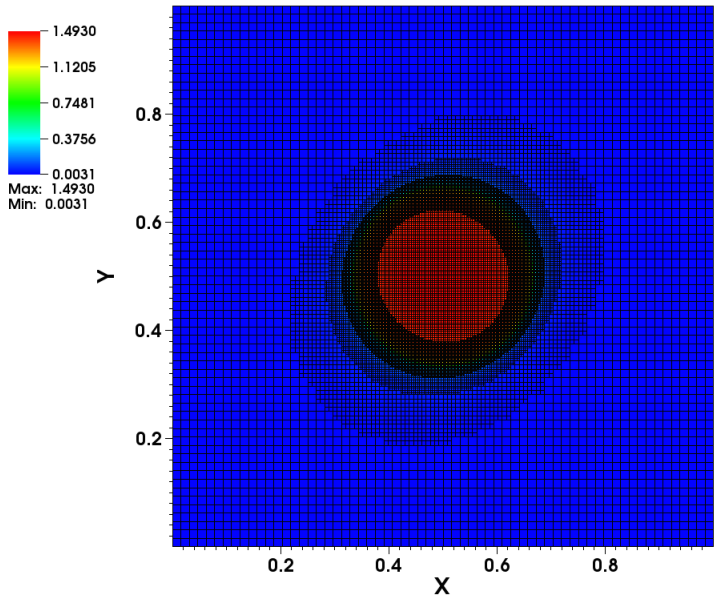


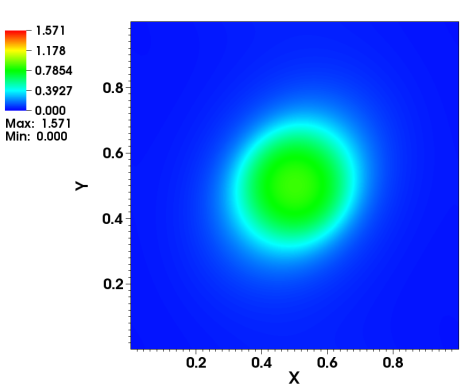
Figure 7.1: A slice perpendicular to the z-axis at the patterned boundary substrate from a 3-D computed solution.

7.4.1 Dirichlet Results

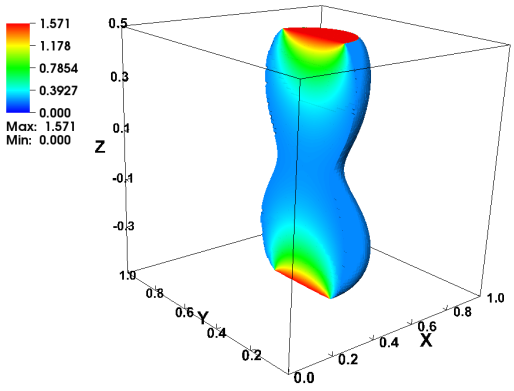
The numerical experiments in this section consider a fixed radius of $r = 1/6$ and a domain such that $z_0 = 1/2$. We consider the case that $\phi_1 = \pi/2$. Outside of the circle, the boundary conditions hold $\theta = 0$, indicating a director aligned with the xy-plane. The ansatz states that $\phi(z) = \phi_0 + \phi_1 z$. Setting $\phi_1 = \pi/2$ indicates that a full $\pi/2$

nematic twist is completed from the lower to upper substrates. The constant ϕ_0 is varied in separate experiments from 0 to $\pi/2$ in increments of $\pi/16$. On the lower substrate, the variation of ϕ_0 changes the alignment of the nematics from beginning at a ϕ-angle of $-\pi/4$ through to $\pi/4$. Nested iteration is used, starting with an initial $16 \times 16 \times 16$ grid iterating through 8 successive adaptive refinements in which the top 30% of mesh cells, in terms of computed free energy, are refined.

Figure 7.2a displays a slice of the computed solution at $z = 0.4$ for $\phi_0 = 0$. Departing from the substrate, we see an elliptic rotating profile for θ which tracks the rotation of ϕ through the twist angle of $\pi/2$. This behavior is also readily observable in Figure 7.2b. This figure displays an isovolume for $\theta \geq 0.21$. Note that the vertically aligned rods on the boundary quickly relax under the influence of the outer, planar aligned rods.



(a) A slice perpendicular to the z-axis at $z = 0.4$ for $\phi_0 = 0$.



(b) An isosurface for $\theta \geq 0.21$ and slice at $x = 0.5$ with $\phi_0 = 0$.

Figure 7.2: Plots of the solution for θ with $\phi_0 = 0$, $\phi_1 = \frac{\pi}{2}$ and Dirichlet boundary conditions.

Table 7.1 reports the computed energies with adaptive refinement broken into the constituent distortion components. The energy computations indicate that with the twisting ϕ-profile, the energetically preferred ϕ_0 value is $\pi/4$. It is important to note that as refinement around the discontinuity at the boundary increases so does the computed free energy. In fact, for this Dirichlet case, the true solution is not expected to have finite free energy in the continuous limit due to the forced discontinuity at the boundary. This energy divergence is manifest in the behavior

of the increasing computed energy after each refinement around the circular discontinuity. The circular surface disclination considered here is of similar character to those studied in [125] where certain surface disclinations are demonstrated to induce infinite free energy. Moreover, analytical solutions to the linear equations arising when $\phi_1 = 0$ have shown divergent free energy.

ϕ_0	0	$\frac{\pi}{16}$	$\frac{2\pi}{16}$	$\frac{3\pi}{16}$	$\frac{4\pi}{16}$	$\frac{5\pi}{16}$	$\frac{6\pi}{16}$	$\frac{7\pi}{16}$	$\frac{8\pi}{16}$
K_1	2.1120	2.1116	2.1108	2.1099	2.1096	2.1099	2.1108	2.1116	2.1120
K_2	1.7053	1.7059	1.7073	1.7086	1.7091	1.7086	1.7073	1.7059	1.7053
K_3	2.6545	2.6540	2.6529	2.6520	2.6516	2.6520	2.6529	2.6540	2.6545
Bulk	6.4718	6.4716	6.4711	6.4705	6.4703	6.4705	6.4711	6.4716	6.4718

Table 7.1: Free energies for variation of ϕ_0 with $\phi_1 = \frac{\pi}{2}$, Dirichlet boundary conditions, and adaptive refinement.

7.4.2 Robin Results

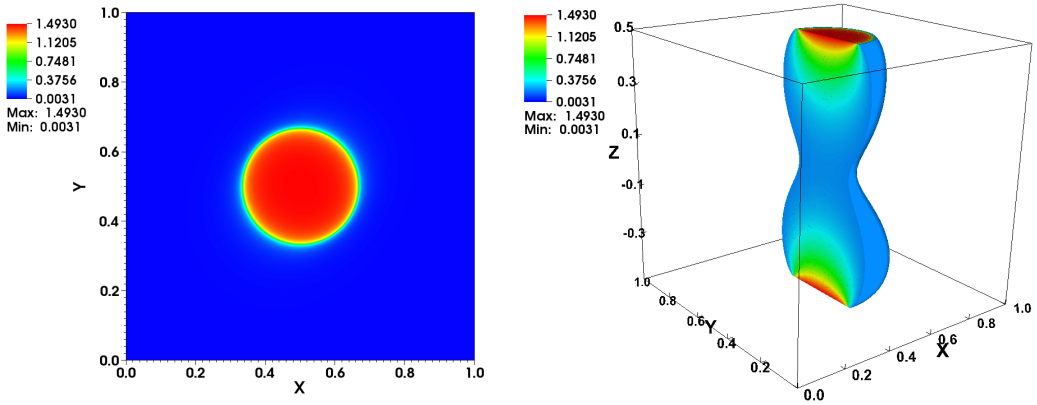
In contrast to the Dirichlet problem, the true free energy for the Robin boundary conditions is expected to converge to a finite value since the nematic rods are free to continuously transition across the circular boundary in order to minimize free energy. For these types of patterned substrates, Robin boundary conditions represent a more accurate physical model of the liquid crystal system. For the simulations considered in the section, the anchoring parameter is $L_\theta = 0.01$. The same adaptive refinement strategy and nested iteration grid hierarchy discussed in the previous section are employed here.

ϕ_0	0	$\frac{\pi}{16}$	$\frac{2\pi}{16}$	$\frac{3\pi}{16}$	$\frac{4\pi}{16}$	$\frac{5\pi}{16}$	$\frac{6\pi}{16}$	$\frac{7\pi}{16}$	$\frac{8\pi}{16}$
K_1	0.5869	0.5870	0.5874	0.5877	0.5878	0.5877	0.5874	0.5870	0.5869
K_2	0.9416	0.9415	0.9412	0.9408	0.9406	0.9408	0.9412	0.9415	0.9416
K_3	0.9130	0.9129	0.9129	0.9130	0.9131	0.9130	0.9129	0.9129	0.9130
Surface	0.7016	0.7016	0.7016	0.7017	0.7017	0.7017	0.7016	0.7016	0.7016
Bulk	2.4414	2.4414	2.4414	2.4415	2.4416	2.4415	2.4414	2.4414	2.4414
Total	3.1430	3.1430	3.1430	3.1432	3.1433	3.1432	3.1430	3.1430	3.1430

Table 7.2: Free energies for variation of ϕ_0 with $\phi_1 = \frac{\pi}{2}$, $r = \frac{1}{6}$, Robin boundary conditions, and adaptive refinement.

For the first set of experiments, $\phi_1 = \pi/2$, $z_0 = 1/2$, $r = 1/6$, and ϕ_0 is again varied from 0 to $\pi/2$ in increments of $\pi/16$. Table 7.2 reports the bulk and surface free energy for the computed solutions. The computed free energies converge quickly for each experiment.

Figure 7.3a shows a slice perpendicular to the z-axis at $z = 0.5$. This reveals the behavior of θ at the top substrate. Note that the Robin boundary conditions allow a continuous transition across the pattern boundary. The elliptic profile seen on the interior of Ω for the Dirichlet boundary conditions persists for this simulation. Figure 7.3b displays an isovolume, sliced by the plane $x = 0.5$, for $\theta \geq 0.21$.



(a) A slice perpendicular to the z-axis at $z = 0.5$ for $\phi_0 = 0$.

(b) An isosurface for $\theta \geq 0.21$ and slice at $x = 0.5$ with $\phi_0 = 0$.

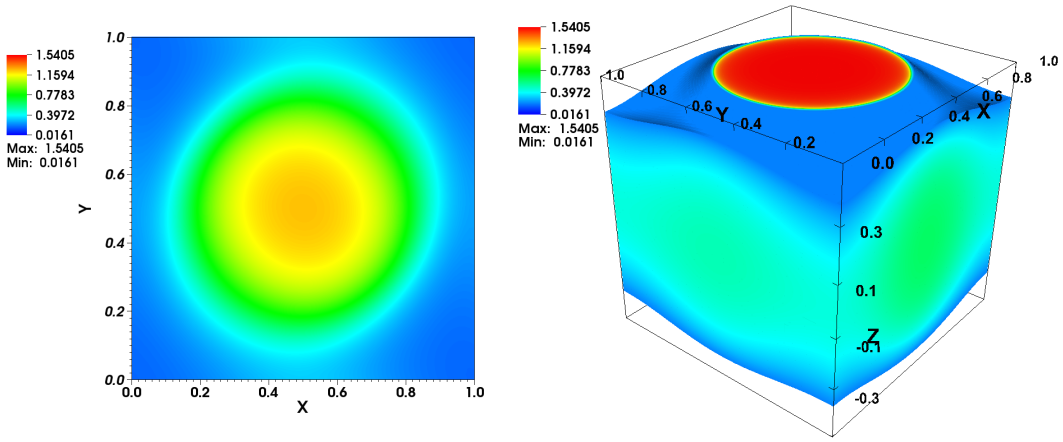
Figure 7.3: Plots of the solution for θ with $\phi_0 = 0$, $\phi_1 = \frac{\pi}{2}$, $r = 1/6$, and Robin boundary conditions.

ϕ_0	0	$\frac{\pi}{16}$	$\frac{2\pi}{16}$	$\frac{3\pi}{16}$	$\frac{4\pi}{16}$	$\frac{5\pi}{16}$	$\frac{6\pi}{16}$	$\frac{7\pi}{16}$	$\frac{8\pi}{16}$
K_1	1.5842	1.5852	1.5876	1.5902	1.5913	1.5902	1.5876	1.5852	1.5842
K_2	1.1599	1.1601	1.1603	1.1600	1.1598	1.1600	1.1603	1.1601	1.1599
K_3	2.0145	2.0128	2.0090	2.0056	2.0043	2.0056	2.0090	2.0128	2.0145
Surface	1.4070	1.4071	1.4074	1.4078	1.4079	1.4078	1.4074	1.4071	1.4070
Bulk	4.7586	4.7581	4.7569	4.7558	4.7554	4.7558	4.7569	4.7581	4.7586
Total	6.1656	6.1652	6.1643	6.1636	6.1633	6.1636	6.1643	6.1652	6.1656

Table 7.3: Free energies for variation of ϕ_0 with $\phi_1 = \frac{\pi}{2}$, $r = \frac{1}{3}$, Robin boundary conditions, and adaptive refinement.

Table 7.3 details the computed free energies for varying ϕ_0 when the radius of the circle at the boundary is increased to $r = 1/3$. With the larger radius, the influence

of the patterning reaches deeper into the cell interior, resulting in more distortion throughout the bulk. This increased ingress is reflected in larger computed free energy and can be seen in Figures 7.4a and 7.4b. These figures show that the patterning significantly alters the angular rise of the director throughout the full interior of the cell.



(a) A slice perpendicular to the z-axis at $z = 0.4$ for $\phi_0 = 0$.

(b) An isosurface for $\theta \geq 0.21$ with $\phi_0 = 0$.

Figure 7.4: Plots of the solution for θ with $\phi_0 = 0$, $\phi_1 = \frac{\pi}{2}$, $r = 1/3$, and Robin boundary conditions.

The subject of our current work is an investigation of the free-energy behavior, for fixed τ values, of these twist solutions compared to solutions with constant ϕ profiles as the circular pattern radius and values of z_0, corresponding to the thickness of the cell, vary. The constant ϕ case is simply captured by taking $\phi_1 = 0$. As noted above, this leads to a linear anisotropic diffusion PDE with Robin boundary conditions.

In our numerical simulations, we have seen evidence that there may exist a regime of radii and cell thicknesses over which the twisting and constant ϕ solutions become energetically degenerate, suggesting the presence of bistable configurations. This would imply that the interior configuration could be switched from a constant ϕ structure to a twisting structure by simply manipulating the radius of the circular pattern or the cell thickness. This would possibly be advantageous, for instance, in the control of switching behaviors in display devices. The numerical simulations

discussed above enable accelerated and precise exploration of the parameter space over which such bistabilities might exist and open the possibility of simulations considering alternative geometric patterns.

Chapter 8

Conclusions and Future Work

This thesis focuses on the development of theoretically supported numerical approaches for the modeling of nematic liquid crystal equilibrium configurations. These approaches consider configurations resulting from free-elastic effects as well as those arising in the presence of applied electric fields and flexoelectricity. An energy-minimization framework incorporating Lagrange multipliers is derived, producing linearized variational systems which are discretized with finite-elements methods. The computed linearization systems are shown to be well-posed when discretized by appropriate combinations of finite elements. Nested iteration and trust regions are shown to dramatically improve robustness and efficiency in the context of the Newton iterations.

In considering the discrete linear systems associated with the electrically coupled simulations, two multigrid relaxation techniques were investigated and implemented as part of a monolithic multigrid approach tailored to the arising saddle-point matrices. The relaxation approaches scale optimally with mesh size and offer clear time-to-solution advantages.

Additionally, an important physical problem investigating the effects of geometrically patterned substrates on three-dimensional liquid crystal configurations is presented. The problem represents a search for bistable configurations induced by circularly patterned substrates of varying radii. Modeling of the behavior of the nematics produces a nonlinear reaction-diffusion PDE for which well-posedness analysis is conducted and a numerical approach is presented.

8.1 Thesis Contributions

Numerical simulations are an indispensable part of the study of liquid crystal equilibrium configurations. Simulations are used to confirm theory, analyze experiments,

and suggest the presence of new physical phenomenon. Many current technologies and experiments, including bistable devices [30, 90], require simulations with anisotropic physical constants on two- and three-dimensional domains. As discussed in Section 3.1, there are a number of approaches to liquid crystal static and dynamic problems that make use of the one-constant approximation in both analysis and simulation. While this approach is useful, particularly in the instance that the Frank constants are not known, there are many scenarios where the use of unequal elastic constants is required to accurately capture liquid crystal behaviors [3–6, 30, 90]. The energy-minimization approach enforcing the pointwise unit-length constraint with Lagrange multipliers developed in this thesis is applicable and effective for anistropic Frank constants and is constructed for use on domains of general dimension. Numerical experiments demonstrate its accuracy in capturing configurations resulting from unequal physical parameters in the presence of elastic, electric, and flexoelectric effects. The algorithm is used to investigate a parameter regime over which a nano-patterned boundary condition induces possible bistabilities due to flexoelectricity.

As part of the energy-minimization framework, Newton linearizations were derived for free-elastic effects, as well as electric responses. This thesis presents novel theory establishing the existence and uniqueness of discrete solutions to the linearizations when discretized by certain finite elements. The associated variational system shares structural similarities to the Stokes equations but presents unique difficulties for establishing weak coercivity of the bilinear form associated with the unit-length constraint. While theory surrounding the incompressibility constraint of the Stokes' equations is well-establish, theory addressing the nonlinear unit-length constraint considered herein is less well understood. Therefore, a novel approach to proving weak coercivity for the bilinear form associated with the unit-length constraint was necessary and is presented here. Moreover, error analysis shows that the linearization systems are convergent with mesh refinement.

The discrete form of the electrically coupled linearization systems have a block saddle-point structure which presents unique challenges for the design of fast solvers.

We have proposed and numerically vetted a collection of relaxation techniques as part of an optimally-scaling monolithic multigrid method. These relaxation schemes are based on unique extensions of the Vanka and Braess-Sarazin relaxation techniques for fluid flow [15, 122]. In addition to their optimal-scaling, the multigrid solvers outperform, often significantly, the application of direct solvers. Moreover, the number of overall Newton steps and solution error remains the same regardless of the solver used.

In order to improve speed and efficiency, we have investigated the performance of nested iteration and trust-region methods. Trust-region approaches, designed in the context of finite-element discretizations, improve both robustness and overall time to solution. Nested iteration is productively used to greatly reduce overall work by isolating a significant portion of the computational cost to the coarsest grids. This results in marked improvements in algorithm efficiency for all simulations considered. Pairing these techniques with the multigrid methods for the linear solve stages yields a robust and efficient algorithm for the computational modeling of static liquid crystal configurations

Finally, a three-dimensional liquid crystal configuration problem of interest in ongoing physics research is presented. The experiments investigate the behavior of equilibrium configurations in the presence of geometrically patterned substrates and gives rise to a nonlinear anistropic reaction-diffusion equation. We have developed theory establishing the well-posedness of the associated Newton linearizations in the presence of both Dirichlet and Robin boundary conditions and presented numerical results detailing the performance of the finite-element and Newton linearization techniques applied to solve the PDE. The ongoing parameter studies represent a search for new bistable configurations induced by variations in cell-thickness and the radius of the circular nano-patterning at the substrates.

8.2 Future Work

The work presented herein reveals a number of interesting questions and opportunities for future work, both in the context of liquid crystal problems and beyond. Below, we outline some of the opportunities we plan to consider in the context of this work. In addition to the theoretical and algorithmic possibilities delineated in this section, there are many interesting physical applications and specific cases that may be investigated using the current and to be developed research.

8.2.1 Liquid Crystal Dynamics

The computational techniques and theory of this thesis focus on the resolution of equilibrium configurations for liquid crystals. However, accurately modeling the fluid flow properties of liquid crystals is also an important endeavor. For instance, dynamic effects play an important role in the manipulation of pixels in an LCD display. In changing the liquid crystal configuration via an electric field, for example, the nematic rods can briefly over-rotate due to angular momentum in an effect known as kickback [24, 82, 117]. Minimizing this effect is desirable in the design of display technologies. Accurate computational modeling enables a more efficient and adaptable search for such designs.

The flow of nematic liquid crystals is governed by an intricate system of PDEs known as the Ericksen-Leslie equations, first proposed by J. L. Ericksen [45] and later refined by F. M. Leslie [79, 80]. A thorough discussion of these equations can be found in [117]. The equations couple the free-elastic effects of liquid crystal microstructures with the constitutive laws governing fluid flow. We have begun preliminary work on a first-order system least squares (FOSLS) formulation [20, 21] for the incompressible Ericksen-Leslie system, as well as an energy-minimization framework. This work requires the derivation of first-principle free-energy laws for the coupled elastic effects of the nematic rods and the fluid properties of the sample. The theoretical work will incorporate finite-element methods from complicated fluid problems and Navier-Stokes theory.

The research aims to accurately solve the Ericksen-Leslie system with unequal Frank constants and full viscosity parameters. Anisotropic constants are often extremely important for effective simulation of physical phenomena but have not been fully studied due to the complexity that anisotropy brings to the Ericksen-Leslie system. The targeted simulations will be used, for instance, to improve the understanding and analysis of optical distortion data produced by liquid crystal experiments.

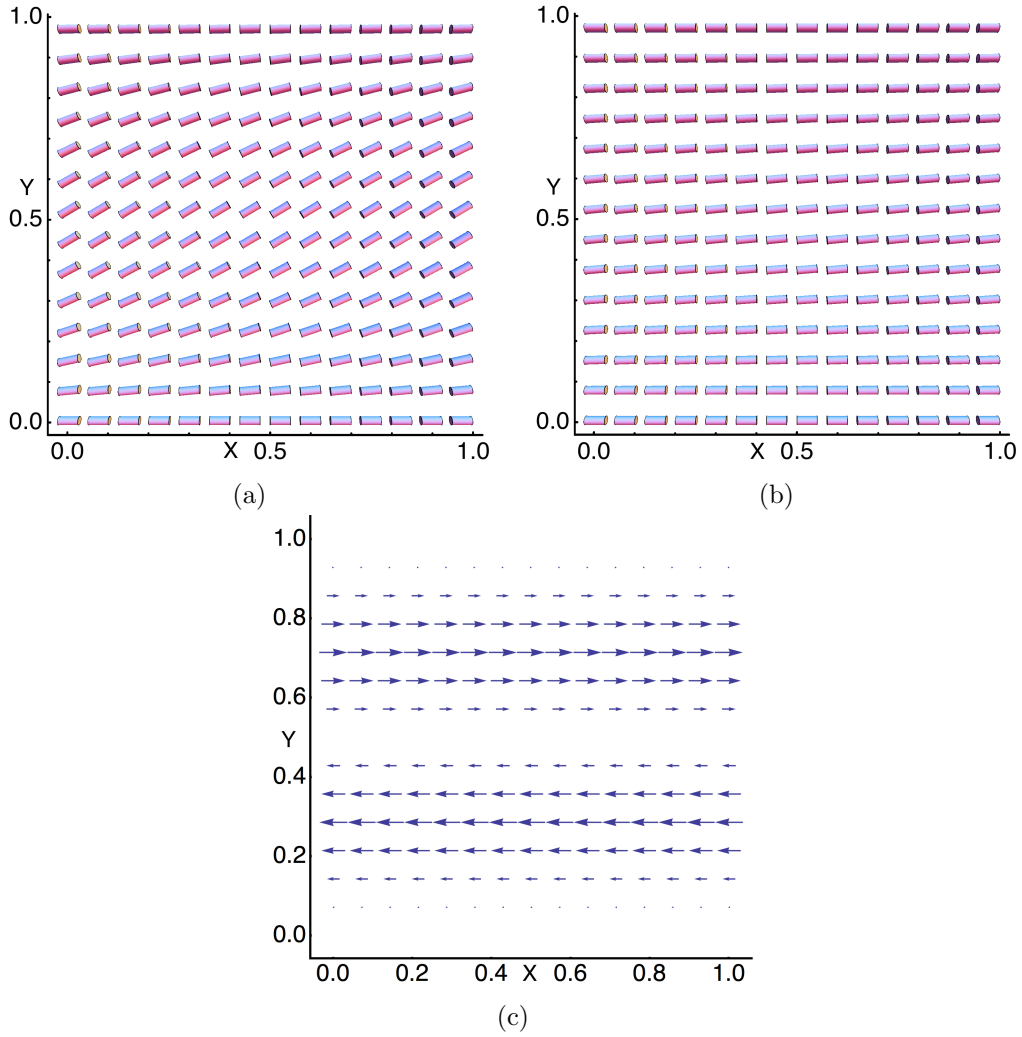


Figure 8.1: (a) Liquid crystal configuration at $t = 1$ relaxing to uniform alignment. (b) Liquid crystal configuration at $t = 3$ near fully relaxed equilibrium state. (c) Induced backflow at $t = 3$ where vectors the indicate fluid velocity field.

Preliminary results using the FOSPACK computational package [110] have shown promise in accurately capturing backflow problems and fluid effects due to Freedericksz transitions in the presence of applied electric fields. Figure 8.1 displays

the computed results of a simple backflow problem in which the nematics begin in a state offset from equilibrium and are allowed to relax to a uniform equilibrium state. This relaxation is expected to engender a mild shear flow in the fluid known as backflow [117].

Figures 8.1a and 8.1b show the liquid crystal configuration at time steps $t = 1$ and $t = 3$, respectively, computed in FOSPACK. The relaxation of the liquid crystal structure induces the computed flow field pictured in Figure 8.1c at $t = 3$, which matches the expected shear flow.

8.2.2 Unit-length Constrained Problems

In the theory developed in this thesis for static liquid crystal configurations, novel techniques to address the bilinear forms associated with the nonlinear, pointwise unit-length constraint were needed. While the mixed-method variational system shared certain similarities with the Stokes' problem, methods aimed at theoretical developments surrounding unit-length constraints are not as comprehensively studied.

Pointwise length constraints exist in many other applications including the modeling of ferromagnetic materials [73]. The use of Lagrange multiplier techniques for these problems results in discrete systems of similar structure to those discussed above. Well-posedness theory for these systems remains a relatively open question due to the challenges of the nonlinear constraint. We plan to investigate extending the theoretical schema developed herein for liquid crystal equilibrium systems to those arising in the context of ferromagnetic modeling and other problems involving pointwise length constraints.

8.2.3 Multigrid

In future research, we plan to specifically consider improvements to the multigrid relaxation techniques discussed above. The relaxation methods were implemented as serial computations. However, significant portions of these relaxation approaches lend themselves to parallel implementation. Therefore, we aim to investigate the

extent to which these methods may be parallelized and evaluate the performance of such parallelization.

In addition, we intend to implement full approximation scheme (FAS) multigrid methods [60, 107] for the nonlinear variational system in (7.6) in the context of both Dirichlet and Robin boundary conditions. FAS multigrid methods have been applied successfully in many contexts to treat nonlinear systems directly using multigrid principles. Such schemes have the potential to accurately and efficiently solve the nonlinear variational problem in (7.6) without the need for linearization of the variational form.

8.2.4 Adaptive Refinement

Finally, the nested-iteration grid hierarchies built throughout this thesis, with the exception of the final chapter, strictly use uniform grid refinements. While current implementations are quite efficient, theoretically supported adaptive refinement techniques could greatly increase efficiency and accuracy by targeting cells with high relative error for refinement, thereby allowing greater refinement in specific regions than that achievable with uniform refinement. In preliminary work, we have examined two strategies aimed at tagging cells for refinement. The first is based on cells with the highest free energy, while the second focuses on cells with the largest nonlinear residual values.

Figure 8.2 displays the distribution of free-elastic energy per cell on a 128×128 grid for the nano-patterned boundary conditions outline in (3.75)-(3.77). Note that a majority of the energy is pooled around the pattern switching junctions where the largest director distortions are forced. Figure 8.3 shows the adaptive refinement pattern resulting from 4 successive refinements of the top 30% of cells in terms of free energy, beginning on an 8×8 mesh. The adaptive refinement solution's computed free energy is 3.8904 using only 820 cells. This free energy value compares favorably with that of the 64×64 grid from Table 3.3, which contains 4096 elements.

While the use of energy as a marker for refinement performs well in the case of nano-patterned boundary conditions, areas of highest energy do not always coincide

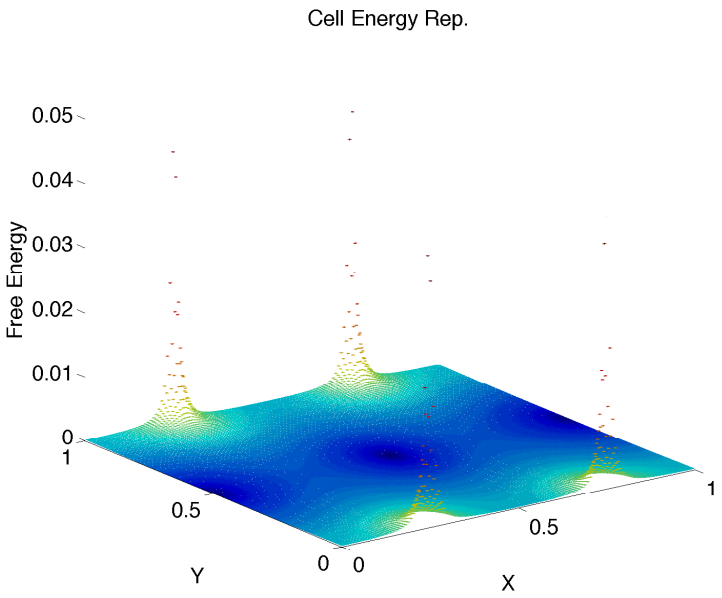


Figure 8.2: A representation of the free-elastic energy contained in each cell of a 128×128 mesh for the free-elastic nano-patterned boundary condition problem.

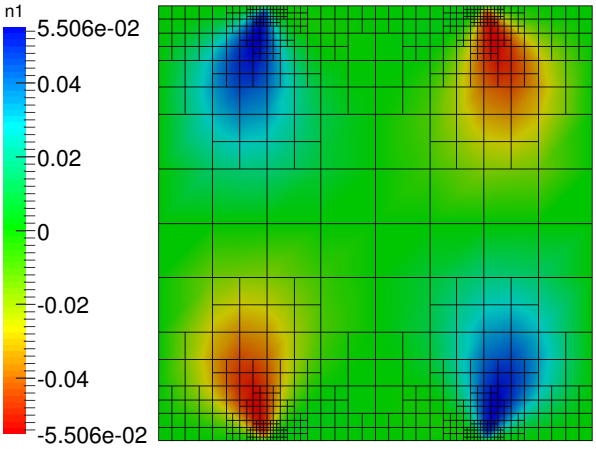


Figure 8.3: The adaptively refined mesh for the free-elastic nano-patterned boundary problem after 4 adaptive refinements based on cell contained free energy.

with the highest error. Consider the energy and error plots in Figures 8.4a and b, respectively, after 6 Newton iterations on a 128×128 grid for the tilt-twist free-elastic problem described in Chapter 4. From the free energy profile, cell refinement would occur near the substrate boundaries, $x = 0$ and $x = 1$. However, the bulk of error lies closer to the middle of the domain interior. Residual-based refinement behaves in a similar way and depends heavily on the initial guess and shape of the true solution.

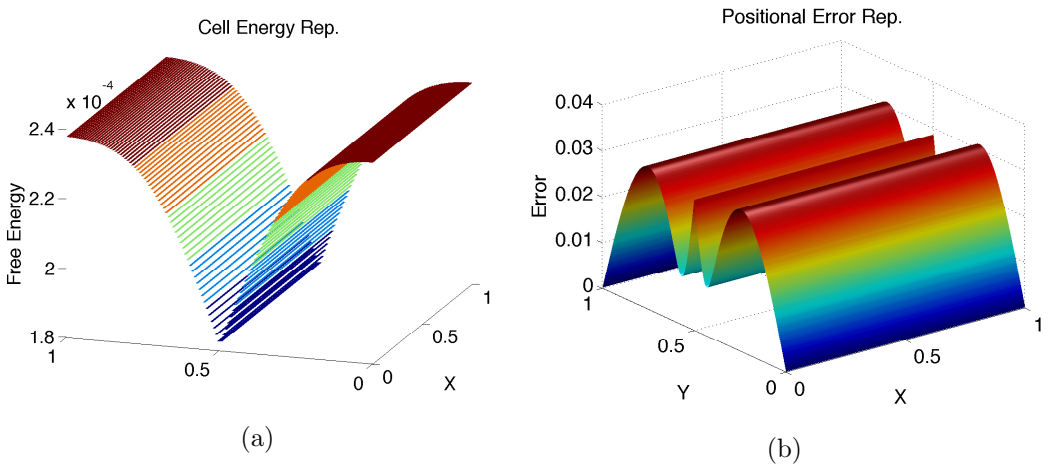


Figure 8.4: (a) Cell energy (b) Tilt-Twist error.

Another concern with energy-based adaptive mesh refinement is the fact that certain regions of equilibrium configurations inherently contain larger free energy. For instance, the four free-energy pools seen in Figure 8.2 persist on a highly refined mesh. Though we expect these areas to contain features which are more challenging to approximate, at a certain level of refinement it becomes advantageous to focus on other, less refined, areas. Decisions of this type are difficult to make, in a non-heuristic way, when exclusively considering free energy as a refinement strategy. Therefore, free energy and residual-based adaptive refinement strategies do not fully constitute sharp error-estimation methods.

Theoretically supported, accurate error and cost estimators exist for a number of PDE methodologies. For example, the Accuracy-per-Computational cost approach, called ACE, is used in the first-order system least squares (FOSLS) framework. It is an efficiency-based refinement method, originally developed in [38, 101], which estimates error reduction and costs resulting from computed refinement patterns. Our current free-energy formulation does not yield a clear intrinsic, a posteriori error estimator, as is seen in the FOSLS approach. Therefore, we aim to develop new techniques to accurately and efficiently estimate error and, thereby, precisely flag cells for refinement within the energy-minimization framework outlined herein.

Appendix A

Linearized Variational Systems

In this appendix, we present the fully expanded forms of the linearized variational systems derived in Chapters 3 and 5. These systems represent the intermediate variational systems discretized and solved at each Newton step in the minimization algorithms outlined in this thesis.

A.1 Free-Elastic Systems

For configurations considering only free-elastic effects, the linearized variational system is written

$$
\begin{aligned}
&(K_1 - K_2 - K_4)\langle \nabla \cdot \delta\mathbf{n}, \nabla \cdot \mathbf{v}\rangle_0 + K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla \times \delta\mathbf{n}, \nabla \times \mathbf{v}\rangle_0 \\
&\quad + (K_2 - K_3)\Big(\langle \delta\mathbf{n} \cdot \nabla \times \mathbf{v}, \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k\rangle_0 + \langle \mathbf{n}_k \cdot \nabla \times \mathbf{v}, \delta\mathbf{n} \cdot \nabla \times \mathbf{n}_k\rangle_0 \\
&\quad + \langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \delta\mathbf{n}\rangle_0 + \langle \mathbf{n}_k \cdot \nabla \times \delta\mathbf{n}, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0 \\
&\quad + \langle \delta\mathbf{n} \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0\Big) + (K_2 + K_4)\Big(\langle \nabla\delta n_1, \frac{\partial \mathbf{v}}{\partial x}\rangle_0 + \langle \nabla\delta n_2, \frac{\partial \mathbf{v}}{\partial y}\rangle_0 \\
&\quad + \langle \nabla\delta n_3, \frac{\partial \mathbf{v}}{\partial z}\rangle_0\Big) + \int_\Omega \lambda_k(\delta\mathbf{n}, \mathbf{v})\, dV + \int_\Omega \delta\lambda(\mathbf{n}_k, \mathbf{v})\, dV \\
&= -\Big((K_1 - K_2 - K_4)\langle \nabla \cdot \mathbf{n}_k, \nabla \cdot \mathbf{v}\rangle_0 + K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla \times \mathbf{n}_k, \nabla \times \mathbf{v}\rangle_0 \\
&\quad + (K_2 - K_3)\langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k\rangle_0 + (K_2 + \mathbf{Z}K_4)\Big(\langle \nabla n_{k_1}, \frac{\partial \mathbf{v}}{\partial x}\rangle_0 \\
&\quad + \langle \nabla n_{k_2}, \frac{\partial \mathbf{v}}{\partial y}\rangle_0 + \langle \nabla n_{k_3}, \frac{\partial \mathbf{v}}{\partial z}\rangle_0\Big) + \int_\Omega \lambda_k(\mathbf{n}_k, \mathbf{v})\, dV\Big),
\end{aligned}
\tag{A.1}
$$

$$
\int_\Omega \gamma(\mathbf{n}_k, \delta\mathbf{n})\, dV = -\frac{1}{2}\int_\Omega \gamma((\mathbf{n}_k, \mathbf{n}_k) - 1)\, dV. \tag{A.2}
$$

We seek to compute $\delta\mathbf{n}$ and $\delta\lambda$ satisfying the system in (A.1) and (A.2) for all $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$ and $\gamma \in L^2(\Omega)$ with the current approximations $\mathbf{n}_k$ and λ_k.

If we are considering a system with Dirichlet or mixed periodic and Dirichlet

boundary conditions, the linearized system simplifies to

$$
\begin{aligned}
&K_1\langle\nabla\cdot\delta\mathbf{n},\nabla\cdot\mathbf{v}\rangle_0 + K_3\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\delta\mathbf{n},\nabla\times\mathbf{v}\rangle_0\\
&\quad+(K_2-K_3)\Big(\langle\delta\mathbf{n}\cdot\nabla\times\mathbf{v},\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0+\langle\mathbf{n}_k\cdot\nabla\times\mathbf{v},\delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k\rangle_0\\
&\quad+\langle\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\delta\mathbf{n}\rangle_0+\langle\mathbf{n}_k\cdot\nabla\times\delta\mathbf{n},\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\\
&\quad+\langle\delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big)+\int_\Omega\lambda_k(\delta\mathbf{n},\mathbf{v})\,dV+\int_\Omega\delta\lambda(\mathbf{n}_k,\mathbf{v})\,dV\\
&=-\Big(K_1\langle\nabla\cdot\mathbf{n}_k,\nabla\cdot\mathbf{v}\rangle_0+K_3\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{n}_k,\nabla\times\mathbf{v}\rangle_0\\
&\quad+(K_2-K_3)\langle\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0+\int_\Omega\lambda_k(\mathbf{n}_k,\mathbf{v})\,dV\Big),
\end{aligned}
$$

$$
\int_\Omega\gamma(\mathbf{n}_k,\delta\mathbf{n})\,dV=-\frac{1}{2}\int_\Omega\gamma((\mathbf{n}_k,\mathbf{n}_k)-1)\,dV.
$$

A.2 Applied Electric Fields

In the presence of applied electric fields, the linearized variational system is written

$$
\begin{aligned}
&(K_1-K_2-K_4)\langle\nabla\cdot\delta\mathbf{n},\nabla\cdot\mathbf{v}\rangle_0+K_3\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\delta\mathbf{n},\nabla\times\mathbf{v}\rangle_0\\
&\quad+(K_2-K_3)\Big(\langle\delta\mathbf{n}\cdot\nabla\times\mathbf{v},\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0+\langle\mathbf{n}_k\cdot\nabla\times\mathbf{v},\delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k\rangle_0\\
&\quad+\langle\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\delta\mathbf{n}\rangle_0+\langle\mathbf{n}_k\cdot\nabla\times\delta\mathbf{n},\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\\
&\quad+\langle\delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big)+(K_2+K_4)\Big(\langle\nabla\delta n_1,\frac{\partial\mathbf{v}}{\partial x}\rangle_0+\langle\nabla\delta n_2,\frac{\partial\mathbf{v}}{\partial y}\rangle_0\\
&\quad+\langle\nabla\delta n_3,\frac{\partial\mathbf{v}}{\partial z}\rangle_0\Big)-\epsilon_0\epsilon_a\langle\delta\mathbf{n}\cdot\nabla\phi_k,\mathbf{v}\cdot\nabla\phi_k\rangle_0+\int_\Omega\lambda_k(\delta\mathbf{n},\mathbf{v})\,dV\\
&\quad-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\phi_k,\mathbf{v}\cdot\nabla\delta\phi\rangle_0-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\delta\phi,\mathbf{v}\cdot\nabla\phi_k\rangle_0+\int_\Omega\delta\lambda(\mathbf{n}_k,\mathbf{v})\,dV\\
&=-\Big((K_1-K_2-K_4)\langle\nabla\cdot\mathbf{n}_k,\nabla\cdot\mathbf{v}\rangle_0+K_3\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{n}_k,\nabla\times\mathbf{v}\rangle_0\\
&\quad+(K_2-K_3)\langle\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0+(K_2+K_4)\Big(\langle\nabla n_{k_1},\frac{\partial\mathbf{v}}{\partial x}\rangle_0+\langle\nabla n_{k_2},\frac{\partial\mathbf{v}}{\partial y}\rangle_0\\
&\quad+\langle\nabla n_{k_3},\frac{\partial\mathbf{v}}{\partial z}\rangle_0\Big)-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\phi_k,\mathbf{v}\cdot\nabla\phi_k\rangle_0+\int_\Omega\lambda_k(\mathbf{n}_k,\mathbf{v})\,dV\Big),
\end{aligned}
\tag{A.3}
$$

$$
\begin{aligned}
&-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\phi_k,\delta\mathbf{n}\cdot\nabla\psi\rangle_0-\epsilon_0\epsilon_a\langle\delta\mathbf{n}\cdot\nabla\phi_k,\mathbf{n}_k\cdot\nabla\psi\rangle_0-\epsilon_0\epsilon_\perp\langle\nabla\delta\phi,\nabla\psi\rangle_0\\
&\quad-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\delta\phi,\mathbf{n}_k\cdot\nabla\psi\rangle_0\\
&=\epsilon_0\epsilon_\perp\langle\nabla\phi_k,\nabla\psi\rangle_0+\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\phi_k,\mathbf{n}_k\cdot\nabla\psi\rangle_0,
\end{aligned}
\tag{A.4}
$$

$$\int_\Omega \gamma(\mathbf{n}_k, \delta\mathbf{n})\, dV = -\frac{1}{2}\int_\Omega \gamma((\mathbf{n}_k, \mathbf{n}_k) - 1)\, dV. \tag{A.5}$$

We compute $\delta\mathbf{n}$, $\delta\phi$, and $\delta\lambda$ satisfying (A.3)-(A.5) for all $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$, $\psi \in H^{1,0}(\Omega)$, and $\gamma \in L^2(\Omega)$ with the current approximations $\mathbf{n}_k$, ϕ_k, and λ_k.

If the system considered has full or mixed Dirichlet boundary conditions, as described throughout this thesis, the simplified linearization is

$$\begin{aligned}
&K_1\langle \nabla\cdot\delta\mathbf{n}, \nabla\cdot\mathbf{v}\rangle_0 + K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla\times\delta\mathbf{n}, \nabla\times\mathbf{v}\rangle_0 \\
&\quad + (K_2 - K_3)\Big(\langle \delta\mathbf{n}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0 + \langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k\rangle_0 \\
&\quad + \langle \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\delta\mathbf{n}\rangle_0 + \langle \mathbf{n}_k\cdot\nabla\times\delta\mathbf{n}, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 \\
&\quad + \langle \delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big) - \epsilon_0\epsilon_a\langle \delta\mathbf{n}\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0 + \int_\Omega \lambda_k(\delta\mathbf{n}, \mathbf{v})\, dV \\
&\quad - \epsilon_0\epsilon_a\langle \mathbf{n}_k\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\delta\phi\rangle_0 - \epsilon_0\epsilon_a\langle \mathbf{n}_k\cdot\nabla\delta\phi, \mathbf{v}\cdot\nabla\phi_k\rangle_0 + \int_\Omega \delta\lambda(\mathbf{n}_k, \mathbf{v})\, dV \\
&= -\Big(K_1\langle \nabla\cdot\mathbf{n}_k, \nabla\cdot\mathbf{v}\rangle_0 + K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{n}_k, \nabla\times\mathbf{v}\rangle_0 \\
&\quad + (K_2 - K_3)\langle \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k, \mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0 - \epsilon_0\epsilon_a\langle \mathbf{n}_k\cdot\nabla\phi_k, \mathbf{v}\cdot\nabla\phi_k\rangle_0 \\
&\quad + \int_\Omega \lambda_k(\mathbf{n}_k, \mathbf{v})\, dV\Big),
\end{aligned}$$

$$\begin{aligned}
&- \epsilon_0\epsilon_a\langle \mathbf{n}_k\cdot\nabla\phi_k, \delta\mathbf{n}\cdot\nabla\psi\rangle_0 - \epsilon_0\epsilon_a\langle \delta\mathbf{n}\cdot\nabla\phi_k, \mathbf{n}_k\cdot\nabla\psi\rangle_0 - \epsilon_0\epsilon_\perp\langle \nabla\delta\phi, \nabla\psi\rangle_0 \\
&\quad - \epsilon_0\epsilon_a\langle \mathbf{n}_k\cdot\nabla\delta\phi, \mathbf{n}_k\cdot\nabla\psi\rangle_0 \\
&= \epsilon_0\epsilon_\perp\langle \nabla\phi_k, \nabla\psi\rangle_0 + \epsilon_0\epsilon_a\langle \mathbf{n}_k\cdot\nabla\phi_k, \mathbf{n}_k\cdot\nabla\psi\rangle_0,
\end{aligned}$$

$$\int_\Omega \gamma(\mathbf{n}_k, \delta\mathbf{n})\, dV = -\frac{1}{2}\int_\Omega \gamma((\mathbf{n}_k, \mathbf{n}_k) - 1)\, dV.$$

A.3 Flexoelectric Augmentation

With the addition of flexoelectric effects, additional terms are added to the applied electric field system to produce the full linearized variational system

$$\begin{aligned}
&(K_1 - K_2 - K_4)\langle \nabla\cdot\delta\mathbf{n}, \nabla\cdot\mathbf{v}\rangle_0 + K_3\langle \mathbf{Z}(\mathbf{n}_k)\nabla\times\delta\mathbf{n}, \nabla\times\mathbf{v}\rangle_0 \\
&\quad + (K_2 - K_3)\Big(\langle \delta\mathbf{n}\cdot\nabla\times\mathbf{v}, \mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0 + \langle \mathbf{n}_k\cdot\nabla\times\mathbf{v}, \delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k\rangle_0
\end{aligned}$$

$$
\begin{aligned}
&+ \langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \delta\mathbf{n} \rangle_0 + \langle \mathbf{n}_k \cdot \nabla \times \delta\mathbf{n}, \mathbf{v} \cdot \nabla \times \mathbf{n}_k \rangle_0 \\
&+ \langle \delta\mathbf{n} \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k \rangle_0 \Big) + (K_2 + K_4)\Big(\langle \nabla \delta n_1, \frac{\partial \mathbf{v}}{\partial x} \rangle_0 + \langle \nabla \delta n_2, \frac{\partial \mathbf{v}}{\partial y} \rangle_0 \\
&+ \langle \nabla \delta n_3, \frac{\partial \mathbf{v}}{\partial z} \rangle_0 \Big) - \epsilon_0 \epsilon_a \langle \delta\mathbf{n} \cdot \nabla \phi_k, \mathbf{v} \cdot \nabla \phi_k \rangle_0 \\
&+ e_s\Big(\langle \nabla \cdot \delta\mathbf{n}, \mathbf{v} \cdot \nabla \phi_k \rangle_0 + \langle \nabla \cdot \mathbf{v}, \delta\mathbf{n} \cdot \nabla \phi_k \rangle_0 \Big) \\
&+ e_b\Big(\langle \delta\mathbf{n} \times \nabla \times \mathbf{v}, \nabla \phi_k \rangle_0 + \langle \mathbf{v} \times \nabla \times \delta\mathbf{n}, \nabla \phi_k \rangle_0 \Big) + \int_\Omega \lambda_k(\delta\mathbf{n}, \mathbf{v})\, dV \\
&- \epsilon_0 \epsilon_a \langle \mathbf{n}_k \cdot \nabla \phi_k, \mathbf{v} \cdot \nabla \delta\phi \rangle_0 - \epsilon_0 \epsilon_a \langle \mathbf{n}_k \cdot \nabla \delta\phi, \mathbf{v} \cdot \nabla \phi_k \rangle_0 \\
&+ e_s\Big(\langle \nabla \cdot \mathbf{n}_k, \mathbf{v} \cdot \nabla \delta\phi \rangle_0 + \langle \nabla \cdot \mathbf{v}, \mathbf{n}_k \cdot \nabla \delta\phi \rangle_0 \Big) \\
&+ e_b\Big(\langle \mathbf{n}_k \times \nabla \times \mathbf{v}, \nabla \delta\phi \rangle_0 + \langle \mathbf{v} \times \nabla \times \mathbf{n}_k, \nabla \delta\phi \rangle_0 \Big) + \int_\Omega \delta\lambda(\mathbf{n}_k, \mathbf{v})\, dV \\
= &-\Big((K_1 - K_2 - K_4) \langle \nabla \cdot \mathbf{n}_k, \nabla \cdot \mathbf{v} \rangle_0 + K_3 \langle \mathbf{Z}(\mathbf{n}_k) \nabla \times \mathbf{n}_k, \nabla \times \mathbf{v} \rangle_0 \\
&+ (K_2 - K_3) \langle \mathbf{n}_k \cdot \nabla \times \mathbf{n}_k, \mathbf{v} \cdot \nabla \times \mathbf{n}_k \rangle_0 + (K_2 + K_4)\Big(\langle \nabla n_{k_1}, \frac{\partial \mathbf{v}}{\partial x} \rangle_0 + \langle \nabla n_{k_2}, \frac{\partial \mathbf{v}}{\partial y} \rangle_0 \\
&+ \langle \nabla n_{k_3}, \frac{\partial \mathbf{v}}{\partial z} \rangle_0 \Big) - \epsilon_0 \epsilon_a \langle \mathbf{n}_k \cdot \nabla \phi_k, \mathbf{v} \cdot \nabla \phi_k \rangle_0 + e_s\Big(\langle \nabla \cdot \mathbf{n}_k, \mathbf{v} \cdot \nabla \phi_k \rangle_0 \\
&+ \langle \nabla \cdot \mathbf{v}, \mathbf{n}_k \cdot \nabla \phi_k \rangle_0 \Big) + e_b\Big(\langle \mathbf{n}_k \times \nabla \times \mathbf{v}, \nabla \phi_k \rangle_0 + \langle \mathbf{v} \times \nabla \times \mathbf{n}_k, \nabla \phi_k \rangle_0 \Big) \\
&+ \int_\Omega \lambda_k(\mathbf{n}_k, \mathbf{v})\, dV \Big),
\end{aligned} \tag{A.6}
$$

$$
\begin{aligned}
&- \epsilon_0 \epsilon_a \langle \mathbf{n}_k \cdot \nabla \phi_k, \delta\mathbf{n} \cdot \nabla \psi \rangle_0 - \epsilon_0 \epsilon_a \langle \delta\mathbf{n} \cdot \nabla \phi_k, \mathbf{n}_k \cdot \nabla \psi \rangle_0 \\
&+ e_s\Big(\langle \nabla \cdot \delta\mathbf{n}, \mathbf{n}_k \cdot \nabla \psi \rangle_0 + \langle \nabla \cdot \mathbf{n}_k, \delta\mathbf{n} \cdot \nabla \psi \rangle_0 \Big) \\
&+ e_b\Big(\langle \mathbf{n}_k \times \nabla \times \delta\mathbf{n}, \nabla \psi \rangle_0 + \langle \delta\mathbf{n} \times \nabla \times \mathbf{n}_k, \nabla \psi \rangle_0 \Big) - \epsilon_0 \epsilon_\perp \langle \nabla \delta\phi, \nabla \psi \rangle_0 \\
&- \epsilon_0 \epsilon_a \langle \mathbf{n}_k \cdot \nabla \delta\phi, \mathbf{n}_k \cdot \nabla \psi \rangle_0 \\
= &-\Big(- \epsilon_0 \epsilon_\perp \langle \nabla \phi_k, \nabla \psi \rangle_0 - \epsilon_0 \epsilon_a \langle \mathbf{n}_k \cdot \nabla \phi_k, \mathbf{n}_k \cdot \nabla \psi \rangle_0 + e_s \langle \nabla \cdot \mathbf{n}_k, \mathbf{n}_k \cdot \nabla \psi \rangle_0 \\
&+ e_b \langle \mathbf{n}_k \times \nabla \times \mathbf{n}_k, \nabla \psi \rangle_0 \Big),
\end{aligned} \tag{A.7}
$$

$$
\int_\Omega \gamma(\mathbf{n}_k, \delta\mathbf{n})\, dV = -\frac{1}{2} \int_\Omega \gamma((\mathbf{n}_k, \mathbf{n}_k) - 1)\, dV. \tag{A.8}
$$

At each iteration, we compute $\delta\mathbf{n}$, $\delta\phi$, and $\delta\lambda$ satisfying (A.6)-(A.8) for all $\mathbf{v} \in \mathcal{H}_0^{DC}(\Omega)$, $\psi \in H^{1,0}(\Omega)$, and $\gamma \in L^2(\Omega)$ with the current approximations $\mathbf{n}_k$, ϕ_k, and λ_k.

If we are considering a system with full or mixed Dirichlet boundary conditions,

as described above, the linearized system is simplified to

$$
\begin{aligned}
&K_1\langle\nabla\cdot\delta\mathbf{n},\nabla\cdot\mathbf{v}\rangle_0 + K_3\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\delta\mathbf{n},\nabla\times\mathbf{v}\rangle_0\\
&\quad+(K_2-K_3)\Big(\langle\delta\mathbf{n}\cdot\nabla\times\mathbf{v},\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k\rangle_0+\langle\mathbf{n}_k\cdot\nabla\times\mathbf{v},\delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k\rangle_0\\
&\quad+\langle\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\delta\mathbf{n}\rangle_0+\langle\mathbf{n}_k\cdot\nabla\times\delta\mathbf{n},\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\\
&\quad+\langle\delta\mathbf{n}\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0\Big)-\epsilon_0\epsilon_a\langle\delta\mathbf{n}\cdot\nabla\phi_k,\mathbf{v}\cdot\nabla\phi_k\rangle_0\\
&\quad+e_s\Big(\langle\nabla\cdot\delta\mathbf{n},\mathbf{v}\cdot\nabla\phi_k\rangle_0+\langle\nabla\cdot\mathbf{v},\delta\mathbf{n}\cdot\nabla\phi_k\rangle_0\Big)\\
&\quad+e_b\Big(\langle\delta\mathbf{n}\times\nabla\times\mathbf{v},\nabla\phi_k\rangle_0+\langle\mathbf{v}\times\nabla\times\delta\mathbf{n},\nabla\phi_k\rangle_0\Big)+\int_\Omega\lambda_k(\delta\mathbf{n},\mathbf{v})\,dV\\
&\quad-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\phi_k,\mathbf{v}\cdot\nabla\delta\phi\rangle_0-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\delta\phi,\mathbf{v}\cdot\nabla\phi_k\rangle_0\\
&\quad+e_s\Big(\langle\nabla\cdot\mathbf{n}_k,\mathbf{v}\cdot\nabla\delta\phi\rangle_0+\langle\nabla\cdot\mathbf{v},\mathbf{n}_k\cdot\nabla\delta\phi\rangle_0\Big)\\
&\quad+e_b\Big(\langle\mathbf{n}_k\times\nabla\times\mathbf{v},\nabla\delta\phi\rangle_0+\langle\mathbf{v}\times\nabla\times\mathbf{n}_k,\nabla\delta\phi\rangle_0\Big)+\int_\Omega\delta\lambda(\mathbf{n}_k,\mathbf{v})\,dV\\
&=-\Big(K_1\langle\nabla\cdot\mathbf{n}_k,\nabla\cdot\mathbf{v}\rangle_0+K_3\langle\mathbf{Z}(\mathbf{n}_k)\nabla\times\mathbf{n}_k,\nabla\times\mathbf{v}\rangle_0\\
&\quad+(K_2-K_3)\langle\mathbf{n}_k\cdot\nabla\times\mathbf{n}_k,\mathbf{v}\cdot\nabla\times\mathbf{n}_k\rangle_0-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\phi_k,\mathbf{v}\cdot\nabla\phi_k\rangle_0\\
&\quad+e_s\Big(\langle\nabla\cdot\mathbf{n}_k,\mathbf{v}\cdot\nabla\phi_k\rangle_0+\langle\nabla\cdot\mathbf{v},\mathbf{n}_k\cdot\nabla\phi_k\rangle_0\Big)\\
&\quad+e_b\Big(\langle\mathbf{n}_k\times\nabla\times\mathbf{v},\nabla\phi_k\rangle_0+\langle\mathbf{v}\times\nabla\times\mathbf{n}_k,\nabla\phi_k\rangle_0\Big)+\int_\Omega\lambda_k(\mathbf{n}_k,\mathbf{v})\,dV\Big),
\end{aligned}
$$

$$
\begin{aligned}
&-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\phi_k,\delta\mathbf{n}\cdot\nabla\psi\rangle_0-\epsilon_0\epsilon_a\langle\delta\mathbf{n}\cdot\nabla\phi_k,\mathbf{n}_k\cdot\nabla\psi\rangle_0\\
&\quad+e_s\Big(\langle\nabla\cdot\delta\mathbf{n},\mathbf{n}_k\cdot\nabla\psi\rangle_0+\langle\nabla\cdot\mathbf{n}_k,\delta\mathbf{n}\cdot\nabla\psi\rangle_0\Big)\\
&\quad+e_b\Big(\langle\mathbf{n}_k\times\nabla\times\delta\mathbf{n},\nabla\psi\rangle_0+\langle\delta\mathbf{n}\times\nabla\times\mathbf{n}_k,\nabla\psi\rangle_0\Big)-\epsilon_0\epsilon_\perp\langle\nabla\delta\phi,\nabla\psi\rangle_0\\
&\quad-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\delta\phi,\mathbf{n}_k\cdot\nabla\psi\rangle_0\\
&=-\Big(-\epsilon_0\epsilon_\perp\langle\nabla\phi_k,\nabla\psi\rangle_0-\epsilon_0\epsilon_a\langle\mathbf{n}_k\cdot\nabla\phi_k,\mathbf{n}_k\cdot\nabla\psi\rangle_0+e_s\langle\nabla\cdot\mathbf{n}_k,\mathbf{n}_k\cdot\nabla\psi\rangle_0\\
&\quad+e_b\langle\mathbf{n}_k\times\nabla\times\mathbf{n}_k,\nabla\psi\rangle_0\Big),
\end{aligned}
$$

$$
\int_\Omega\gamma(\mathbf{n}_k,\delta\mathbf{n})\,dV=-\frac{1}{2}\int_\Omega\gamma((\mathbf{n}_k,\mathbf{n}_k)-1)\,dV.
$$

Appendix B

An Inf-Sup Result

During the course of research into the well-posedness of the linearization systems derived in Chapters 3 and 5, the following inf-sup condition was considered and proven.

Lemma B.0.1 *For Assumption 3.5.1, there exists a constant $\zeta > 0$ such that*

$$\sup_{\mathbf{v}\in\mathcal{H}_0^{DC}(\Omega)} \frac{b(\mathbf{v},\gamma)}{\|\mathbf{v}\|_0} \geq \zeta\|\gamma\|_0, \qquad \forall\gamma \in L^2(\Omega), \tag{B.1}$$

where $b(\mathbf{v},\gamma) = \int_\Omega \gamma(\mathbf{v},\mathbf{n}_k)\,dV$. It turned out that the context in which this condition was examined did not lead to the desired result. However, because the literature concerning well-posedness for unit-length constrained variational problems is not as well developed as the body of research surrounding divergence constraints, such as those arising in the context of the Stokes' problem, we include the proof here in the hope that it informs further research.

Proof: Assume that $\gamma \neq 0$, since that case is trivial. For a fixed, arbitrary, nonzero $\gamma \in L^2(\Omega)$

$$\sup_{\mathbf{v}\in\mathcal{H}_0^{DC}(\Omega)} \frac{b(\mathbf{v},\gamma)}{\|\mathbf{v}\|_0} \geq \frac{\int_\Omega \gamma(\varphi,\mathbf{n}_k)\,dV}{\|\varphi\|_0},$$

for any $\varphi \in \mathcal{H}_0^{DC}(\Omega)$. Let $\mathbf{v}_1 = \left(\frac{\gamma}{|\mathbf{n}_k|}\right)\mathbf{n}_k$. Clearly $\mathbf{v}_1$ is an element of $L^2(\Omega)^3$, and

$$\|\mathbf{v}_1\|_0^2 = \left\|\left(\frac{\gamma}{|\mathbf{n}_k|}\right)\mathbf{n}_k\right\|_0^2 = \|\gamma\|_0^2. \tag{B.2}$$

Note then that

$$\frac{\int_\Omega \gamma(\mathbf{v}_1,\mathbf{n}_k)\,dV}{\|\mathbf{v}_1\|_0} = \frac{\int_\Omega \gamma^2|\mathbf{n}_k|\,dV}{\|\gamma\|_0}.$$

With the assumption of control over the director iterate length in (3.9),

$$\frac{\int_\Omega \gamma^2 |\mathbf{n}_k|\, dV}{\|\gamma\|_0} \geq \frac{\sqrt{\alpha}\int_\Omega \gamma^2\, dV}{\|\gamma\|_0} = \sqrt{\alpha}\|\gamma\|_0.$$

Observe that $\mathbf{v}_1 \in L^2(\Omega)^3$ but is not necessarily in $\mathcal{H}_0^{DC}(\Omega)$. However, $C_c^\infty(\Omega)^3 \subset H_0(\mathrm{div},\Omega) \cap H_0(\mathrm{curl},\Omega)$, where $C_c^\infty(\Omega)$ denotes the set of compactly supported smooth functions on Ω. Moreover, $C_c^\infty(\Omega)^3$ is dense in $L^2(\Omega)^3$ [59]. Thus, for any $\epsilon > 0$, there exists a $\varphi \in C_c^\infty(\Omega)^3$ such that

$$\|\mathbf{v}_1 - \varphi\|_0 \leq \epsilon.$$

Next, the objective is to show that there exists a $C_a > 0$ and a $\varphi \in C_c^\infty(\Omega)^3$ such that

$$\frac{\int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV}{\|\varphi\|_0} \geq C_a \frac{\int_\Omega \gamma(\mathbf{v}_1, \mathbf{n}_k)\, dV}{\|\mathbf{v}_1\|_0}. \tag{B.3}$$

Consider

$$\begin{aligned}\left|\int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV - \int_\Omega \gamma(\mathbf{v}_1, \mathbf{n}_k)\, dV\right| = \left|\int_\Omega \gamma(\varphi - \mathbf{v}_1, \mathbf{n}_k)\, dV\right| &\leq \|\varphi - \mathbf{v}_1\|_0 \|\gamma \mathbf{n}_k\|_0 \\ &\leq \epsilon \|\gamma \mathbf{n}_k\|_0. \end{aligned} \tag{B.4}$$

Note first that

$$\int_\Omega \gamma(\mathbf{v}_1, \mathbf{n}_k)\, dV = \int_\Omega \gamma^2 |\mathbf{n}_k|\, dV > 0.$$

Since γ is fixed and $\mathbf{n}_k$ is known, $\mathbf{v}_1$ is fixed as well. Combining this with (B.4) implies that there exists a φ and an $\epsilon_1 > 0$ such that

$$0 < \int_\Omega \gamma(\mathbf{v}_1, \mathbf{n}_k)\, dV - \epsilon_1 \leq \int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV \leq \int_\Omega \gamma(\mathbf{v}_1, \mathbf{n}_k)\, dV + \epsilon_1. \tag{B.5}$$

Hence, it is assumed that $\int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV$ is positive. This implies that (B.3) is equivalent to the inequality

$$\frac{\|\mathbf{v}_1\|_0}{\|\varphi\|_0} \geq C_a \frac{\int_\Omega \gamma(\mathbf{v}_1, \mathbf{n}_k)\, dV}{\int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV}, \tag{B.6}$$

so long as φ is sufficiently close such that (B.5) holds. By the reverse triangle inequality,

$$|\|\varphi\|_0 - \|\mathbf{v}_1\|_0| \leq \|\varphi - \mathbf{v}_1\|_0 \leq \epsilon.$$

This implies that

$$\|\varphi\|_0 \leq \|\mathbf{v}_1\|_0 + \epsilon. \tag{B.7}$$

Observe that (B.2) implies that (B.7) becomes

$$\|\varphi\|_0 \leq \|\gamma\|_0 + \epsilon. \tag{B.8}$$

Note also that (B.4) implies that

$$\int_\Omega \gamma(\mathbf{v}_1, \mathbf{n}_k)\, dV \leq \epsilon\|\gamma\mathbf{n}_k\|_0 + \int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV. \tag{B.9}$$

It is possible to formulate a sufficient condition to (B.6) as finding a $C_a > 0$ and a $\varphi \in C_c^\infty(\Omega)^3$ such that

$$\frac{\|\gamma\|_0}{\|\gamma\|_0 + \epsilon} \geq C_a \frac{\int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV + \epsilon\|\gamma\mathbf{n}_k\|_0}{\int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV}. \tag{B.10}$$

Inequality (B.10) is sufficient since, combining results (B.8) and (B.9),

$$\frac{\|\mathbf{v}_1\|_0}{\|\varphi\|_0} \geq \frac{\|\gamma\|_0}{\|\gamma\|_0 + \epsilon} \geq C_a \frac{\int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV + \epsilon\|\gamma\mathbf{n}_k\|_0}{\int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV} \geq C_a \frac{\int_\Omega \gamma(\mathbf{v}_1, \mathbf{n}_k)\, dV}{\int_\Omega \gamma(\varphi, \mathbf{n}_k)\, dV}.$$

Thus, proving (B.10) will imply (B.6). Note that the left side of (B.10) is strictly less than one and the right side, excluding C_a, is strictly greater than one. Further, note that as ϵ goes to zero these quantities approach one from below and above respectively. Because $\mathbf{v}_1$ may be approximated by φ with arbitrary accuracy, one may freely choose $\epsilon > 0$. Define

$$\epsilon_2 = \frac{\int_\Omega \gamma(\mathbf{v}_1, \mathbf{n}_k)\, dV - \epsilon_1}{\|\gamma\mathbf{n}_k\|_0}, \qquad \epsilon_3 = \|\gamma\|_0.$$

Now take $\varphi_* \in C_c^\infty(\Omega)^3$ approximating $\mathbf{v}_1$ such that

$$\|\varphi_* - \mathbf{v}_1\|_0 \leq \epsilon = \min(\epsilon_1, \epsilon_2, \epsilon_3).$$

Thus,

$$\frac{1}{2} < \frac{\|\gamma\|_0}{\|\gamma\|_0 + \epsilon} < 1.$$

Moreover,

$$1 < \frac{\int_\Omega \gamma(\varphi_*, \mathbf{n}_k)\, dV + \epsilon\|\gamma\mathbf{n}_k\|_0}{\int_\Omega \gamma(\varphi_*, \mathbf{n}_k)\, dV} = 1 + \epsilon\frac{\|\gamma\mathbf{n}_k\|_0}{\int_\Omega \gamma(\varphi_*, \mathbf{n}_k)\, dV} < 2,$$

by (B.5). Hence, letting $C_a = 1/4$, (B.10) is satisfied and, therefore, (B.6) is satisfied. Thus,

$$\begin{aligned}
\sup_{\mathbf{v}\in\mathcal{H}_0^{DC}(\Omega)} \frac{b(\mathbf{v},\gamma)}{\|\mathbf{v}\|_0} &\geq \frac{\int_\Omega \gamma(\varphi_*, \mathbf{n}_k)\, dV}{\|\varphi_*\|_0} \\
&\geq \frac{1}{4}\left(\frac{\int_\Omega \gamma(\mathbf{v}_1, \mathbf{n}_k)\, dV}{\|\mathbf{v}_1\|_0}\right) \\
&\geq \frac{\sqrt{\alpha}}{4}\|\gamma\|_0. \qquad \square
\end{aligned}$$

Bibliography

[1] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, Elsevier Ltd., 2nd ed., 2003.

[2] J. H. Adler, T. R. Benson, E. C. Cyr, S. P. MacLachlan, and R. S. Tuminaro, *Monolithic multigrid methods for 2D resistive magnetohydrodynamics*, Under Review, (2014).

[3] C. Anquetil-Deck, D. J. Cleaver, and T. J. Atherton, *Competing alignments of nematic liquid crystals on square-patterned substrates*, Phys. Rev. E, 86 (2012).

[4] C. Anquetil-Deck, D. J. Cleaver, T. J. Atherton, and J. P. Bramble, *Independent control of polar and azimuthal anchoring*, Phys. Rev. E, 88 (2013).

[5] T. J. Atherton and J. H. Adler, *Competition of elasticity and flexoelectricity for bistable alignment of nematic liquid crystals on patterned surfaces*, Phys. Rev. E, 86 (2012).

[6] T. J. Atherton and J. R. Sambles, *Orientational transition in a nematic liquid crystal at a patterned surface*, Phys. Rev. E, 74 (2006).

[7] I. Babuska, *Error-bounds for finite element methods*, Numer. Math., 16 (1971), pp. 322–333.

[8] W. Bangerth, R. Hartmann, and G. Kanschat, *deal.II – a general purpose object oriented finite element library*, ACM Trans. Math. Softw., 33 (2007), pp. 24/1–24/27.

[9] W. Bangerth, T. Heister, G. Kanschat, et al., `deal.II` *Differential Equations Analysis Library, Technical Reference*, `http://www.dealii.org`.

[10] M. Benzi, G. H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numerica, (2005), pp. 1–137.

[11] M. Benzi, E. Haber, and L. Taralli, *A preconditioning technique for a class of PDE-constrained optimization problems*, Adv. Comput. Math., 35 (2011), pp. 149–173.

[12] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*, Springer, 2013.

[13] F. K. Bogner, R. L. Fox, and L. A. Schmit, *The generation of interelement compatible stiffness and mass matrices by the use of interpolation formulas*, in Proceedings Conference on Matrix Methods in Structural Mechanics, Dayton, OH, 1965, Wright Patterson A.F.B, pp. 397–444.

[14] D. Braess, *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*, Cambridge University Press, 1997.

[15] D. Braess and R. Sarazin, *An efficient smoother for the Stokes problem*, Appl. Numer. Math., 23 (1997), pp. 3–19.

[16] S. C. Brenner and L. Scott, *The Mathematical Theory of Finite Element Methods*, Springer-Verlag, 1996.

[17] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, SIAM Books, Philadelphia, 2nd ed., 2000.

[18] R. H. Byrd, R. B. Schnabel, and G. A. Shultz, *A trust region algorithm for nonlinearly constrained optimization*, SIAM J. Numer. Anal., 24 (1987), pp. 1152–1170.

[19] ——, *Approximate solution of the trust region problem by minimization over two-dimensional subspaces*, Math. Programming, 40 (1988), pp. 247–263.

[20] Z. Cai, R. Lazarov, T. Manteuffel, and S. McCormick, *First-order system least squares for second-order partial differential equations*, SIAM J. Numer. Anal., 31 (1994), pp. 1785–1799.

[21] Z. Cai, T. Manteuffel, and S. McCormick, *First-order system least squares for second-order partial differential equations. II*, SIAM J. Numer. Anal., 34 (1997), pp. 425–454.

[22] S. Chandrasekhar, *Liquid Crystals*, Cambridge University Press, Cambridge, UK, 2nd ed., 1992.

[23] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, vol. 4, North Holland, 1978.

[24] M. G. Clark and F. M. Leslie, *A calculation of orientational relaxation in nematic liquid crystals*, Proc. R. Soc. Lond., 361 (1978), pp. 463–485.

[25] R. Cohen, R. Hardt, D. Kinderlehrer, S. Lin, and M. Luskin, *Minimum energy configurations for liquid crystals: Computational results*, in Theory and Applications of Liquid Crystals, vol. 5 of The IMA Volumes in Mathematics and Its Applications, Springer-Verlag, New York, 1987, pp. 99–121.

[26] S. Cornford and C. J. P. Newton, *An adaptive hierarchical finite element method for modelling liquid crystal devices*, Tech. Rep. HPL-2011-143, Hewlett-Packard Laboratories, 2011.

[27] J. Curie and P. Curie, *Contractions et dilatation produites par des tensions dans les cristaux hémièdres à faces inclinées*, C. R. Acad. Sci. Gen., 93 (1880).

[28] E. Cyr, J. Shadid, R. Tuminaro, R. Pawlowski, and L. Chacón, *A new approximate block factorization preconditioner for two-dimensional incompressible (reduced) resistive MHD*, SIAM J. Sci. Comput., 35 (2013), pp. B701–B730.

[29] K. R. Daly, G. D'Allessandro, and M. Kaczmarek, *Regime independent coupled-wave equations in anisotropic photorefractive media*, Appl. Phys. B: Lasers Opt., 95 (2009), pp. 589–596.

[30] A. J. Davidson and N. J. Mottram, *Flexoelectric switching in a bistable nematic device*, Phys. Rev. E, 65 (2002).

[31] T. A. Davis, *Finite Element Analysis of the Landau-de Gennes Minimization Problem for Liquid Crystals in Confinement*, PhD thesis, Case Western Reserve University, Cleveland, Ohio, 1994.

[32] ——, *Algorithm 832: UMFPACK, an unsymmetric-pattern multifrontal method*, ACM Trans. Math. Softw., 30 (2004), pp. 196–199.

[33] ——, *A column pre-ordering strategy for the unsymmetric-pattern multifrontal method*, ACM Trans. Math. Softw., 30 (2004), pp. 165–195.

[34] T. A. Davis and I. S. Duff, *An unsymmetric-pattern multifrontal method for sparse LU factorization*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 140–158.

[35] ——, *A combined unifrontal/multifrontal method for unsymmetric sparse matrices*, ACM Trans. Math. Softw., 25 (1999), pp. 1–19.

[36] T. A. Davis and E. C. Gartland-Jr., *Finite element analysis of the Landau-de Gennes minimization problem for liquid crystals*, SIAM J. Numer. Anal., 35 (1998), pp. 336–362.

[37] P. G. de Gennes and J. Prost, *The Physics of Liquid Crystals*, Clarendon Press, Oxford, UK, 2nd ed., 1993.

[38] H. DeSterck, T. Manteuffel, S. McCormick, J. Nolting, J. Ruge, and L. Tang, *Efficiency-based h- and hp-refinement strategies for finite element methods*, J. Num. Lin. Alg. Appl., 15 (2008), pp. 249–270.

[39] P. Deuflhard, *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*, Springer, Berlin, 2004.

[40] H. J. Deuling, *Deformation of nematic liquid crystals in an electric field*, Mol. Cryst. Liq. Cryst., 19 (1972), pp. 123–131.

[41] C. R. Dohrmann and P. B. Bochev, *A stabilized finite element method for the Stokes problem based on polynomial pressure projections*, Int. J. Numer. Meth. Fluids, (2000).

[42] H. C. Elman, D. J. Silvester, and A. J. Wathen, *Finite Elements and Fast Iterative Solvers: with applications in incompressible fluid dynamics*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2005.

[43] S. J. Elston, *Flexoelectricity in nematic domain walls*, Phys. Rev. E, 78 (2008).

[44] J. Ericksen, *Continuum theory of nematic liquid crystals*, Res. Mechanica, 21 (1987), pp. 381–392.

[45] J. L. Ericksen, *Conservation laws for liquid crystals*, Trans. Soc. Rheol., 5 (1961), pp. 23–34.

[46] ——, *Hydrostatic theory of liquid crystals*, Arch. Rat. Mech. Anal., 9 (1962), pp. 371–378.

[47] ——, *Inequalities in liquid crystal theory*, Phys. Fluids, 9 (1966), pp. 1205–1207.

[48] R. Fletcher, *Practical Methods of Optimization*, vol. 1, John Wiley and Sons, Inc., New York, Brisbane, and Toronto, 1980.

[49] L. P. Franca and S. L. Frey, *Stabilized finite element methods: II. the incompressible Navier-Stokes equations*, Comput. Methods Appl. Mech. Engrg., 99 (1992), pp. 209–233.

[50] F. C. Frank, *On the theory of liquid crystals*, Discuss. Faraday Soc., 25 (1958), pp. 19–28.

[51] V. Freedericksz and V. Zolina, *Forces causing the orientation of an anisotropic liquid*, Trans. Faraday Soc., 29 (1933), pp. 919–930.

[52] H. Fujita and T. Kato, *On the Navier-Stokes initial values problem I*, Arch. Rat. Mech. Anal., 16 (1964), pp. 269–315.

[53] E. Gartland-Jr. and A. Ramage, *A renormalized Newton method for liquid crystal director models with pointwise unit-vector constraints*, SIAM J. Numer. Anal., 53 (2015), pp. 251–278.

[54] E. C. Gartland-Jr. and A. Ramage, *Local stability and a renormalized Newton method for equilibrium liquid crystal director modeling*, working paper, University of Strathclyde, 2012.

[55] L. Gattermann and A. Ritschke, *Über azoxyphenoläther*, Ber. Deutsche. Chem. Ges., 23 (1890).

[56] V. Girault and P. Raviart, *Finite Element Methods for Navier-Stokes Equations: Theory and Algorithms*, Springer-Verlag, Germany, 1986.

[57] R. Glowinski, P. Lin, and X. B. Pan, *An operator-splitting method for a liquid crystal model*, Comput. Phys. Comm., 152 (2003), pp. 242–252.

[58] M. Goh, S. Matsushita, and K. Akagi, *From helical polyacetylene to helical graphite: Synthesis in the chiral nematic liquid crystal field and morphology-retaining carbonisation*, Chem. Soc. Rev., 39 (2010), pp. 2466–2476.

[59] D. H. Griffel, *Applied Functional Analysis*, Dover Publications, 2002.

[60] W. Hackbusch, *Multigrid Methods and Applications*, Springer, Berlin, Heidelberg, New York, Tokyo, 1985.

[61] I. Haller, *Elastic constants of the nematic liquid crystalline phase of p-Methoxybenzylidene-p-n-Butylaniline (MBBA)*, J. Chem. Phys., 57 (1972), pp. 1400–1405.

[62] J. Harden, M. Chambers, R. Verduzco, P. Luchette, J. T. Gleeson, S. Sprunt, and A. Jákli, *Giant flexoelectricity in bent-core nematic liuqid crystal elastomers*, Appl. Phys. Lett., 96 (2010).

[63] M. A. Heroux, R. A. Bartlett, V. E. Howle, R. J. Hoekstra, J. J. Hu, T. G. Kolda, R. B. Lehoucq, K. R. Long, R. P. Pawlowski, E. T. Phipps, A. G. Salinger, H. K. Thornquist, R. S. Tuminaro, J. M. Willenbring, A. Williams, and K. S. Stanley, *An overview of the trilinos project*, ACM Trans. Math. Softw., 31 (2005), pp. 397–423.

[64] Q. Hu and L. Yuan, *A Newton-penalty method for a simplified liquid crystal model*, Adv. Comput. Math., 40 (2014), pp. 201–244.

[65] C. G. J. Jacobi, *Über eine neue auflösungsart der bei der methode der kleinsten quadrate vorkommenden linearen gleichungen*, Astronomische Nachrichten, 22 (1845), pp. 297–306.

[66] A. Jákli, *Electro-mechanical effects in liquid crystals*, Liquid Crystals, 37 (2010), pp. 825–837.

[67] J. T. Jenkins and P. J. Barratt, *Interfacial effects in the static theory of nematic liquid crystals*, Q. Jl. Mech. Appl. Math., 27 (1974), pp. 111–127.

[68] V. John and G. Matthies, *Higher order finite element discretizations in a benchmark problem for incompressible flows*, Internat. J. Numer. Methods Fluids, 37 (2001), pp. 885–903.

[69] V. John, G. Matthies, T. I. Mitkova, L. Tobiska, and P. S. Vassilevski, *A comparison of three solvers for the incompressible Navier-Stokes equations*, in Large-scale Scientific Computations of Engineering and Environmental Problem II, vol. 73 of Notes on Numerical Fluid Mechanics, 2000, pp. 215–222.

[70] V. John and L. Tobiska, *A coupled multigrid method for nonconforming finite element discretizations of the 2d-Stokes equation*, Computing, 64 (2000), pp. 307–321.

[71] ——, *Smoothers in coupled multigrid methods for the parallel solution of the incompressible Navier-Stokes equations*, Internat. J. Numer. Methods Fluids, 33 (2000).

[72] H. Kelker and B. Scheurle, *A liquid-crystalline (nematic) phase with a particularly low solidification point*, Angew. Chem. Int., 8 (1969), pp. 884–885.

[73] M. Kružík and A. Prohl, *Recent developments in the modeling, analysis, and numerics of ferromagnetism*, SIAM Rev., 48 (2006), pp. 439–483.

[74] J. P. F. Lagerwall and G. Scalia, *A new era for liquid crystal research: Applications of liquid crystals in soft matter, nano-, bio- and microtechnology*, Curr. Appl. Phys., 12 (2012), pp. 1387–1412.

[75] M. Larin and A. Reusken, *A comparative study of efficient iterative solvers for generalized Stokes equations*, Numer. Linear Algebra Appl., 15 (2008), pp. 13–34.

[76] B. W. Lee and N. A. Clark, *Alignment of liquid crystals with patterned isotropic surfaces*, Science, 291 (2001), pp. 2576–2580.

[77] J. Leray, *Essai sur le mouvement d'un liquide visqueux emplissant l'espace*, Acta Matematica, 63 (1933), pp. 193–248.

[78] F. Leslie, *Theory of flow phenomenon in liquid crystals*, in The Theory of Liquid Crystals, vol. 4, Academic Press, 1979, pp. 1–81.

[79] F. M. Leslie, *Some constitutive equations for anisotropic fluids*, Q. Jl. Mech. Appl. Math., 19 (1966), pp. 357–370.

[80] ——, *Some constitutive equations for liquid crystals*, Arch. Rat. Mech. Anal., 28 (1968), pp. 265–283.

[81] ——, *Distorted twisted orientation patterns in nematic liquid crystals*, Pramana, Suppl. No., 1 (1975), pp. 41–55.

[82] ——, *Some topics in equilibrium theory of liquid crystals*, in Theory and Applications of Liquid Crystals, J. Ericksen and D. Kinderlehrer, eds., Springer-Verlag, New York, 1987, pp. 211–234.

[83] P. Lin, C. Liu, and H. Zhang, *An energy law preserving C0 finite element scheme for simulating the kinematic effects in liquid crystal flow dynamics*, J. Comp. Phys., (2007), pp. 1411–1427.

[84] C. Liu and H. Sun, *On energetic variational approaches in modeling the nematic liquid crystal flows*, Discrete Contin. Dyn. Syst., 23 (2009), pp. 455–475.

[85] C. Liu and N. J. Walkington, *Mixed methods for the approximation of liquid crystal flows*, ESAIM Math. Model. Numer. Anal., 36 (2002), pp. 205–222.

[86] C. Liu, H. Zhang, and S. Zhang, *Numerical simulations of hydrodynamics of nematic liquid crystals: Effects of kinematic transports*, Commun. Comput. Phys., 9 (2011), pp. 974–993.

[87] J. D. Logan, *An Introduction to Nonlinear Partial Differential Equations*, John Wiley and Sons, Inc., 2nd ed., 2008.

[88] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley and Sons, Inc., New York, 1969.

[89] S. P. MacLachlan and C. W. Oosterlee, *Local Fourier analysis for multigrid with overlapping smoothers applied to systems of PDEs*, Numer. Linear Algebra Appl., 18 (2011), pp. 751–774.

[90] A. Majumdar, C. J. P. Newton, J. M. Robbins, and M. Zyskin, *Topology and bistability in liquid crystal devices*, Phys. Rev. E, 75 (2007).

[91] S. Manservisi, *Numerical analysis of Vanka-type solvers for steady Stokes and Navier-Stokes flows*, SIAM J. Numer. Anal., 44 (2006), pp. 2025–2056.

[92] N. Maratos, *Exact Penalty Function Algorithms for Finite Dimensional and Control Optimization*, PhD thesis, Univ. of London, London, 1978.

[93] E. Marusic-Paloka, *Solvability of the Navier-Stokes system with L^2 boundary data*, Appl. Math. Optim., 41 (2000), pp. 365–375.

[94] S. McCormick, *A mesh refinement method for $Ax = \lambda Bx$*, Math. Comp., 36 (1981), pp. 485–498.

[95] R. B. Meyer, *Piezoelectric effects in liquid crystals*, Phys. Rev. Lett., 22 (1969), pp. 918–921.

[96] V. Milisic and U. Razafison, *Weighted sobolev spaces for the Laplace equation in periodic infinite strips*, Preprint, arXiv:1302.4253, (2013).

[97] J. Molenaar, *A two-grid analysis of the combination of mixed finite elements and Vanka-type relaxation*, in Multigrid Methods, III, W. Hackbusch and U. Trottenberg, eds., Birkhäuser Verlag: Basel, 1991, pp. 313–323.

[98] R. J. Moreau, *Magnetohydrodynamics*, Springer Science and Business Media, illustrated ed., 1990.

[99] H. M. Mourad, J. Dolbow, and I. Harari, *A bubble-stabilized finite element method for Dirichlet constraints on embedded interfaces*, Int. J. Numer. Meth. Engng, 69 (2006), pp. 1–21.

[100] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[101] J. Nolting, *Efficency-based Local Adaptive Refinement for FOSLS Finite Elements*, PhD thesis, University of Colorado at Boulder, 2008.

[102] E. O. Omojokun, *Trust Region Algorithms for Optimization with Nonlinear Equality and Inequality Constraints*, PhD thesis, University of Colorado, Boulder, 1989.

[103] A. Pandolfi and G. Napoli, *A numerical investigation of configurational distortions in nematic liquid crystals*, J. Nonlinear Sci., 21 (2011), pp. 785–809.

[104] R. Pierre, *Simple C^0 approximations for the computation of incompressible flows*, Comput. Methods Appl. Mech. Engrg, 68 (1988), pp. 205–227.

[105] A. Ramage and E. C. Gartland-Jr., *A preconditioned nullspace method for liquid crystal director modeling*, SIAM J. Sci. Comput., 35 (2013), pp. B226–B247.

[106] F. Reinitzer, *Beitrage zur kenntnis des cholesterins*, Monatsh. Chem., 9 (1888), pp. 421–441.

[107] A. Reusken, *Convergence of the multigrid full approximation scheme for a class of elliptic mildly nonlinear boundary value problems*, Numer. Math., 52 (1987/88), pp. 251–277.

[108] W. Rudin, *Real and Complex Analysis*, Mathematics Series, McGraw-Hill, 1987.

[109] P. Rudquist and S. T. Lagerwall, *On the flexoelectric effect in nematics*, Liq. Cryst., 23 (1997), pp. 503–510.

[110] J. Ruge, *Fospack users manual*. Version 1.0, 2000.

[111] J. Schöberl and W. Zulehner, *On Schwarz-type smoothers for saddle point problems*, Numer. Math., 95 (2003), pp. 377–399.

[112] P. L. Seidel, *Über ein verfahren die gleichungen, auf welche die methode der kleinsten quadrate führt, sowie lineare gleichungen überhaupt durch successive annäherung aufzulösen*, Abhandlungen der Bayrischen Akademie, 11 (1873), pp. 81–108.

[113] G. A. Shultz, R. B. Schnabel, and R. H. Byrd, *A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties*, SIAM J. Numer. Anal., 22 (1985), pp. 47–67.

[114] S. Sivaloganathan, *The use of local mode analysis in the design and comparison of multigrid methods*, Comput. Phys. Comm., 65 (1991), pp. 246–252.

[115] D. C. Sorensen, *Newton's method with a model trust-region modification*, SIAM J. Numer. Anal., 19 (1982), pp. 409–426.

[116] G. Starke, *Gauss-Newton multilevel methods for least-squares finite element computations of variably saturated subsurface flow*, Computing, 64 (2000), pp. 323–338.

[117] I. W. Stewart, *The Static and Dynamic Continuum Theory of Liquid Crystals: A Mathematical Introduction*, Taylor and Francis, London, 2004.

[118] K. Su and D. Pu, *A nonmonotone filter trust region method for nonlinear constrained optimization*, J. Comput. Appl. Math., 223 (2009), pp. 230–239.

[119] D. Thomsen, P. Keller, J. Naciri, R. Pink, H. Jeon, D. Shenoy, and B. Ratna, *Liquid crystal elastomers with mechanical properties of a muscle*, Macromolecules, 34 (2001), pp. 5868–5875.

[120] U. Trottenberg, C. W. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, London, 2001.

[121] S. Turek, *Efficient solvers for incompressible flow problems: An algorithmic and computational approach*, in Lecture Notes in Computational Science and Engineering, vol. 6, Berlin, 1999, Springer.

[122] S. P. Vanka, *Block-implicit multigrid calculation of two-dimensional recirculating flows*, Comput. Methods Appl. Mech. Engrg., 59 (1986), pp. 29–48.

[123] A. Vardi, *A trust region algorithm for equality constrained minimization: Convergence properties and implementation*, SIAM J. Numer. Anal., 22 (1985), pp. 575–591.

[124] E. G. Virga, *Variational Theories for Liquid Crystals*, Chapman and Hall, London, 1994.

[125] V. Vitek and M. Kléman, *Surface disclinations in nematic liquid crystals*, J. Phys., 36 (1975), pp. 59–67.

[126] Y. Wan and D. Zhao, *On the controllable soft-templating approach to mesoporous silicates*, Chem. Rev., 107 (2007), pp. 2821–2860.

[127] H. Wu, X. Xu, and C. Liu, *On the general Ericksen-Leslie system: Parodi's relation, well-posedness and stability*, Arch. Rat. Mech. Anal., (2013), pp. 59–107.

[128] M. Yamada, M. Kondo, J. Mamiya, Y. Yu, M. Kinoshita, C. Barrett, and T. Ikeda, *Photomobile polymer materials: Towards light-driven plastic motors*, Angew. Chem. Int., 47 (2008), pp. 4986–4988.

[129] H. Zhang and Q. Bai, *Numerical investigation of tumbling phenomena based on a macroscopic model for hydrodynamic nematic liquid crystals*, Commun. Comput. Phys., 7 (2010), pp. 317–332.

[130] H. Zocher, *Über die Einwirkung magnetischer, elektrischer und mechanischer Kräfte auf Mesophasen*, Physik. Zietschr., 28 (1927), pp. 790–796.