A FINITE ELEMENT METHOD FOR THE GENERALIZED ERICKSEN MODEL OF NEMATIC LIQUID CRYSTALS

SHAWN W. WALKER*

Abstract. We consider the generalized Ericksen model of liquid crystals, which is an energy with 8 independent “elastic” constants that depends on two order parameters $n$ (director) and $s$ (variable degree of orientation). In addition, we present a new finite element discretization for this energy, that can handle the degenerate elliptic part without regularization, with the following properties: it is stable and it $\Gamma$-converges to the continuous energy. Moreover, it does not require the mesh to be weakly acute (which was an important assumption in our previous work). Furthermore, we include other effects such as weak anchoring (normal and tangential), as well as fully coupled electro-statics with flexo-electric and order-electric effects. We also present several simulations (in 2-D and 3-D) illustrating the effects of the different elastic constants and electric field parameters.

Mathematics Subject Classification. MSC 65N30, 49M25, 35J70.

Received May 7, 2019. Accepted December 13, 2019.

1. Introduction

Liquid crystals (LCs) are a classic example of anisotropic matter. Indeed, LCs are considered a *meso-phase* of matter, between a liquid and a solid, which is directly due to the anisometric shape of the LC molecules (*i.e.* LCs are an anisotropic material). The most famous application of LCs are in display devices [47, 78], but many other novel uses are being found in material science, such as self-assembly of composites, optics, and biotechnology [53].

The mathematical modeling of LCs is rather sophisticated. The Landau-deGennes macroscopic order parameter $Q$ is derived via an ensemble type of averaging [84, 87]. With this, and the tools of classical continuum mechanics, one can formulate an energy functional which the LC material minimizes at equilibrium. Mathematical analysis of the $Q$-tensor model has been done in several works; for instance, see [10–12, 59, 61, 74].

In contrast, the Oseen-Frank model is the simplest model of a nematic LC [37, 49, 87]. This model uses a unit vector field $n$ called the *director* as the order parameter. The corresponding energy functional is given by $\int_{\Omega} |\nabla n|^2$ (in the one-constant case). Much of the mathematical analysis of Oseen-Frank is related to harmonic mappings [4, 15, 16, 28, 41, 45, 46, 57]. The Oseen-Frank model is a work-horse of the display industry [42, 75, 81], however its main drawback is that defects (discontinuities in $n$) usually have infinite energy.

*Keywords and phrases.* Liquid crystals, defects, finite element method, gamma-convergence, flexo-electric.

Department of Mathematics and Center for Computation and Technology (CCT), Louisiana State University, Baton Rouge, LA 70803, USA.

*Corresponding author: walker@math.lsu.edu

Article published by EDP Sciences © EDP Sciences, SMAI 2020
The Ericksen model of LCs was developed to allow for defects with finite energy [34, 87]. Here, two order parameters are used, \( n \) and \( s \), with a corresponding (one-constant) energy functional given by (5.1). The preliminary mathematical analysis of the Ericksen model can be found in [5, 6, 55, 56] with later work in [25].

Minimizers of the Ericksen model may yield non-trivial defects [18, 20, 23, 55, 56, 80]. The variable degree-of-orientation variable \( s \) in (5.1) gives a degenerate Euler–Lagrange equation for \( n \). The advantage here is that it allows for line and plane defects (of \( n \)) in dimension \( d = 3 \) with finite energy. Defects are important in applications, especially those that lie on three dimensional space curves [7, 29, 43, 85].

Many numerical methods have been developed for simulating the statics and dynamics of liquid crystals [1–3, 14, 30, 77]; see [9] for a survey. The methods in [4, 15, 28, 57] are for harmonic mappings, i.e. nematic liquid crystals with a fixed degree of orientation. However, until recently, there has not been much numerical work on the Ericksen model, except for [14, 25].

A method was developed in [69, 70] by the author and collaborators to solve the (one-constant) Ericksen model of nematic liquid crystals (summarized in Sect. 2.1.1). A discrete form of the energy (5.2) was developed in [69, 70] and shown to Γ-converge to (5.2); in addition, a method for computing discrete minimizers was given. This method was later extended to account for colloidal particle effects and external electric fields [71], as well as simulating liquid crystal droplets with anisotropic surface tension effects [33, 63]. The two main limitations of the approach in [33, 63, 70, 71] are: (1) it is for the one-constant Ericksen model, and (2) the method requires the computational mesh to be weakly acute to guarantee convexity of the discrete energy.

Summary 1.1. In this paper, we consider a new discretization of the Ericksen model that is capable of handling the general form of Ericksen’s model, i.e. not the one-constant model (see Sect. 2.1.4). In addition, the method only requires a shape regular mesh; the weakly acute mesh condition is no longer required. This is especially important in three dimensions because generating a weakly acute mesh of a general non-trivial three dimensional domain is an open problem. The reason is that the current discretization uses a mass lumping technique, which is different than in our previous work [70, 71], where the weak acuteness condition cannot be dropped. Furthermore, we fully couple non-linear electro-statics to the Ericksen model, including flexo-electric and order electric effects [1, 24, 64]; previously only a given electric field \( E \) was considered.

Moreover, we are able to prove convergence of our finite element method using the tools of Γ-convergence [21, 31]. The Euler–Lagrange equation for the Ericksen model is not easy to analyze because the PDE for the director \( n \) is degenerate, i.e. the coefficient of the elliptic term is \( s^2 \) which can vanish. Regularizing the \( s^2 \) term, with a small positive parameter, is not desirable because it destroys the main purpose of the Ericksen model (see Rem. 2.3). Using Γ-convergence, we can avoid dealing with the Euler–Lagrangian equation entirely.

An outline is as follows. Section 2 describes the continuum equilibrium model and develops several analytic results needed in our Γ-convergence proof, and Section 3 describes our finite element discretization of the continuous problem. In Section 4, we prove that our finite element scheme Γ-converges to the continuous problem; several technical results are built up to accomplish this. Numerical results are given in Section 5, followed by some concluding remarks in Section 6. Several technical results are collected in Appendix A.

2. Equilibrium model

We describe the different (energetic) parts of the liquid crystal model. Section 2.1 gives the general Ericksen (free) energy, as well as its basic mathematical formulation. Section 2.2 describes how weak anchoring effects are modeled, and Section 2.3 gives the non-linear electro-static model with flexo-electric and order-electric effects. We conclude in Section 2.4 with some analytical results for the continuous model.

2.1. Ericksen’s model

Let \( \Omega \) be a bounded Lipschitz domain in \( \mathbb{R}^d \) with \( d = 2 \), or 3. The director field \( n : \Omega \to S^{d-1} \) is a vector-valued function with unit length. The degree-of-orientation \( s : \Omega \subset \mathbb{R}^d \to [-\frac{1}{2}, 1] \) is a real valued function. The variable \( n \), by itself, cannot properly describe a “loss of order” in the liquid crystal material because it has unit
length. The $s$ variable models the “local order” of the liquid crystal molecules. See [71,87] for a description of the meaning of $n$ and $s$.

**Remark 2.1.** Since the Ericksen model uses a director (vector) field $n$, clearly Ericksen cannot model half-integer defects, which is an obvious limitation in modeling nematic LCs in some situations. Indeed, nematics are best modeled by using a line field (essentially, vectors without orientation). If a line field is orientable, then it can be replaced by a vector field with no adverse effect [11,12], i.e., a vector field is sufficient to model the nematic state.

For non-orientable line fields, a possible remedy is to adopt the approach in [48,61,92] that enforces the equivalence of $\pm n$. Unfortunately, their method assumes $s$ is a constant parameter, which is not true for the Ericksen model. In addition, their method uses an explicit time-stepping scheme, so is not efficient. However, the main idea in [48,61,92] (also see [17]) could potentially be combined with our approach, but this is left to future work.

In this paper, we explore the general Ericksen model with the vector field approach because vector fields are adequate in some situations and the numerical realization of the generalized Ericksen model has not been done before. Moreover, nematic vector field models may be useful for other physical applications where orientation is important.

We begin by recalling the Ericksen one-constant model, followed by its theoretical framework and the more general Ericksen model. In doing so, for a generic domain $\mathcal{D}$, we will use the following $L^2(\mathcal{D})$, $[L^2(\mathcal{D})]^d$, $[L^2(\mathcal{D})]^{d \times d}$ inner products: $(u,v)_D := \int_D uv$, $(\mathbf{u},\mathbf{v})_D := \int_D \mathbf{u} \cdot \mathbf{v}$, $(\mathbf{M},\mathbf{Y})_D := \int_D \mathbf{M} : \mathbf{Y}$. For simplicity, we will write $(u,v) := (u,v)_\Omega$ when integrating over $\Omega$; integrals over co-dimension 1 subsets, e.g. $\Gamma \subset \partial \Omega$, will always have a subscript $\Gamma$.

### 2.1.1. Ericksen’s simple energy

The equilibrium state of the liquid crystal is given by a pair $(s,n)$ that minimizes a bulk free energy functional, whose simplest form is the following (dimensional) energy:

$$J(s,n) = E_s(s,n) + \int_{\Omega} \psi(s) \, dx,$$

$$E_s(s,n) := \frac{1}{2} \int_{\Omega} \left(b_0 |\nabla s|^2 + k_0 s^2 |\nabla n|^2\right) \, dx,$$  \hspace{1cm} (2.1)

where $b_0,k_0 > 0$ are model parameters. Typical physical values for $k_0$ are on the order of $10^{-13} \text{J/m}$ ([76], Tab. 1, p. 168). Unfortunately, we are unaware of available experimental data for $b_0$, thus we assume $b_0$ is of roughly the same order as $k_0$.

The double well potential $\psi$ is a $C^2$ function defined on $-1/2 < s < 1$ that satisfies [5,34,56]

(i) $\lim_{s \to -1/2} \psi(s) = \lim_{s \to -1/2} \psi(s) = \infty$,

(ii) $\psi(0) > \psi(s^*) = \min_{s \in [-1/2,1]} \psi(s) = 0$ for some $s^* \in (0,1)$,

(iii) $\psi'(0) = 0$.

**Remark 2.2.** The form of $\psi$ follows from the (uniaxial) Landau-deGennes theory of nematic LCs [32,87]. Usually, the following choice is made:

$$\psi(s) = \frac{A'}{2} s^2 - \frac{B'}{3} s^3 + \frac{C'}{4} s^4,$$  \hspace{1cm} (2.2)

where the parameters $A',B',C'$ are material dependent with $B'$, $C'$ positive and $A'$ has no definite sign. Usually, $A'$ depends on temperature $T$ [39] having the form $A' \propto (T - T^*)$, where $T^*$ is the super-cooling temperature. Physical values for $A',B',C'$ are on the order of $10^5 \text{J/m}^3$ ([76], Tab. 1, p. 168).

Property (iii) of $\psi$ is automatically satisfied by (2.2). If $A'$ is less than a sufficiently small positive number $A'_0$, then property (ii) is also satisfied; this corresponds to having a stable nematic phase. In other words, if $A'$ is too large (positive), then the only stable phase is the isotropic phase, meaning $s = 0$ everywhere. Property
(i) is not satisfied by (2.2). However, the $s^4$ term can be modified near the bounds $s = -1/2, +1$ to enforce property (i), without affecting the stability of the nematic phases.

For ease in our numerical implementation, we assume the form of (2.2) for $\psi$, but one can certainly add barrier/penalty functions to enforce property (i) numerically.

When the degree of orientation $s$ is a non-zero constant, the energy $E_{\text{one},b_0}(s,n)$ in (5.2) reduces to the Oseen-Frank free energy $\int_\Omega |\nabla n|^2$. The degree of orientation avoids singular energies when defects are present. In fact, discontinuities in $n$ (i.e. defects) have finite energy provided they occur in the singular set

$$S := \{x \in \Omega : s(x) = 0\}.$$  \hfill (2.3)

Existence of minimizers was shown in [5,56] and analytic solutions for minimizers with defects were constructed in [70]. Minimizers with other types of defect structures were discovered numerically in [70].

**Remark 2.3.** One cannot simply regularize $E_s$ by $E^\varepsilon_s(s,n) = \frac{1}{2} \int_\Omega (b_0|\nabla s|^2 + k_0(s^2 + \varepsilon^2)|\nabla n|^2)$ for some finite $\varepsilon > 0$ as was done in [14,25]. This fundamentally changes the Ericksen model into a variant of Oseen-Frank, i.e. point defects in two dimensions, and line defects in three dimensions, will give $E^\varepsilon_s(s,n) = +\infty$. If defects are important in the physical model, then regularization is not appropriate. In a sense, the finite element discretization automatically regularizes the problem without needing an extra term.

For simplicity throughout this paper, we assume the parameters have been normalized, i.e. $k_0 \equiv 1$, and $b_0$ and $\psi$ are non-dimensional (see Sect. 5.1).

### 2.1.2. Function space framework

An auxiliary variable $u := sn$ and identity was introduced in [5,56] that allows the energy $E_{\text{one},b_0}(s,n)$ to be rewritten as

$$E_{\text{one},b_0}(s,n) = \tilde{E}_{\text{one},b_0}(s,u) := \frac{1}{2} \int_\Omega \left( (b_0 - 1)|\nabla s|^2 + |\nabla u|^2 \right) dx,$$  \hfill (2.4)

which uses $\nabla u = n \otimes \nabla s + s \nabla n$ and the unit length constraint $|n| = 1$. Whence, even with $0 < b_0 < 1$, the minimization problem for $\tilde{E}_{\text{one},b_0}(s,u)$ is well-defined [5,56] over the following (closed) admissible set:

$$\mathcal{A} := \{(s,n) \in H^1(\Omega) \times [L^\infty(\Omega)]^d] ; \ (s,u,n) \text{ satisfies (2.6), with } u \in [H^1(\Omega)]^d\},$$  \hfill (2.5)

where

$$u = sn, \quad -1/2 \leq s \leq 1 \ a.e. \ in \ \Omega, \ and \ n \in S^{d-1} \ a.e. \ in \ \Omega,$$  \hfill (2.6)

is called the structural condition of $\mathcal{A}$. If we write $(s,u,n)$ in $\mathcal{A}$, we mean that $(s,n)$ in $\mathcal{A}$, $u$ in $[H^1(\Omega)]^d$, and $(s,u,n)$ satisfies (2.6). Note: the identity (2.4) only holds for $(s,u,n)$ in $\mathcal{A}$.

### 2.1.3. Boundary conditions

Boundary conditions are captured by functions $g : \mathbb{R}^d \to \mathbb{R}$, $r, q : \mathbb{R}^d \to \mathbb{R}^d$ that satisfy the following.

**Assumption 2.4** (Boundary data is regular). There exists $g \in W^{1,\infty}(\mathbb{R}^d)$, $r \in [W^{1,\infty}(\mathbb{R}^d)]^d$, $q \in [L^\infty(\mathbb{R}^d)]^d$, such that $(g,r,q)$ satisfies (2.6) on $\mathbb{R}^d$, i.e. $r = gq$ and $q \in S^{d-1} a.e. in \mathbb{R}^d$. Furthermore, we assume there is a fixed $\rho_0 > 0$ such that

$$-1/2 + \rho_0 \leq g \leq 1 - \rho_0.$$  \hfill (2.7)

Note that $q \in [W^{1,\infty}(|\{g| > \epsilon\}|)]^d$, for all $\epsilon > 0$.

Next, set $\Gamma := \partial \Omega$ and let $\Gamma_s \subset \Gamma$ be the open set on which we set $s = g$; further assume $\Gamma_s$ decomposes as:

$$\Gamma_s = \text{int} (\Gamma_{|s| \geq \delta_0} \cup \Gamma_{|s| \leq \delta_0}) ; \quad \Gamma_{|s| \geq \delta_0} := \{|s| \geq \delta_0\} ; \quad \Gamma_{|s| \leq \delta_0} := \{|s| \leq \delta_0\},$$  \hfill (2.8)
A FEM FOR THE GENERALIZED ERICKSEN MODEL

Figure 1. Illustration of liquid crystal domain $\Omega$ and boundary conditions on $\Gamma := \partial \Omega$ with unit outer normal vector $\nu$. Note that $\Gamma_{|s| \geq \delta_0} := \Gamma_{s > \delta_1} \cup \Gamma_{s < -\delta_2}$, where $\delta_1, \delta_2 > 0$ and $\delta_0 := \min\{\delta_1, \delta_2\}$ (refer to main text for notation). Moreover, $\Gamma_n := \Gamma_{|s| \geq \delta_0}$.

for some fixed $\delta_0 > 0$. Next, let $\Gamma_n \subset \Gamma$ be the open set on which we set $n = q$. For simplicity, we demand that $\Gamma_n \subset \Gamma_{|s| \geq \delta_0}$, which implies that $q$ is $W^{1,\infty}$ in a neighborhood of $\Gamma_n$ and $n$ is $H^1$ in a neighborhood of $\Gamma_n$ (see Fig. 1 for an illustration). So setting boundary conditions for $(s, n)$ is meaningful. Thus, the admissible class, with boundary conditions, is given by

$$\mathcal{A}(g, q) := \{(s, n) \in \mathcal{A} : s|_{\Gamma_s} = g, \quad n|_{\Gamma_n} = q\}, \quad (2.9)$$

Note: we use a similar abuse of notation as above when writing $(s, u, n)$ in $\mathcal{A}(g, q)$.

In proving our $\Gamma$-convergence result in Section 4, we require the following technical assumption regarding boundary data.

**Assumption 2.5 (Multiple boundary pieces).** Suppose $\partial \Omega \equiv \Gamma = \bigcup_{i=1}^M \Gamma_i$ decomposes into $M \geq 1$ disconnected components, where each component $\Gamma_i$ is connected. We assume that $\Gamma_s = \Gamma_{|s| \geq \delta_0} = \Gamma_n = \bigcup_{k=1}^{\overline{M}} i_k$, where $\overline{M} \leq M$ and $i_k \in \{1, ..., M\}$ for all $1 \leq k \leq \overline{M}$. Moreover, we further assume that $|g| > \delta_0$ on $\Gamma_s \subset \Gamma$, for some $\delta_0 > 0$.

Note that (2.10) implies that $\Gamma_{|s| \leq \delta_0} = \emptyset$ (recall Fig. 1).

2.1.4. Ericksen’s general energy

The general form of Ericksen’s free energy can be found in [34, 87]. Starting from [87] page 325, we have

$$E_{\text{erk}}(s, n) = \frac{1}{2} \int_{\Omega} \mathcal{W}(s, \nabla s, n, \nabla n) \, dx, \quad (2.11)$$

where the free energy density $\mathcal{W} : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is given by:

$$\mathcal{W}(s, g, n, M) := k_1 s^2 \text{tr}(M)^2 + k_2 s^2 (|n| : M)^2 + k_3 s^2 |Mn|^2 + (k_2 + k_4) s^2 [(M^T : M) - \text{tr}(M)^2] + b_1 |g|^2 + b_2 (g \cdot n)^2 + b_3 s (g \cdot n) \text{tr}(M) + b_4 s g \cdot Mn, \quad (2.12)$$
where \([a]_x \in \mathbb{R}^{d \times d}\) is the anti-symmetric matrix defined by \([a]_x b := a \times b\) (if \(a, b \in \mathbb{R}^d\)), and \(\{k_i\}_{i=1}^4\) and \(\{b_i\}_{i=1}^4\) are bounded constants. More specifically, one can show that

\[
\mathcal{W}(s, \nabla s, n, \nabla n) = k_1 s^2 (\text{div} n)^2 + k_2 s^2 (n \cdot \text{curl} n)^2 + k_3 s^2 n \times \text{curl} n^2 \\
+ (k_2 + k_3) s^2 [\text{tr}(\nabla n^2) - (\text{div} n)^2] + b_1 |\nabla s|^2 + b_2 (\nabla s \cdot n)^2 + b_3 s (\text{div} n)(\nabla s \cdot n) \\
+ b_4 s \nabla s \cdot (\nabla n)n, 
\]

(2.13)

where we use the identity \(\text{tr}(M^T Y) = M : Y = \sum_{i,j} m_{ij} y_{ij}\), and the identities \(n \times \text{curl} n = [\nabla n] \times : [\nabla n]\), which hold when \(|n| = 1\). Note that the coefficients can be generalized \([34, 87]\), where, for instance, \(k_i s^2\) is replaced by \(\bar{k}_i = \tilde{k}_i(s)\), \(i.e., a general function of s\). However, for simplicity, we take (2.12) as our model. Note that a derivative of \(n\) is always paired with a factor of \(s\). For simplicity, we assume the coefficients are non-dimensional (see Sect. 5.1).

For conciseness later, we introduce the following multi-linear forms:

\[
w_{k_1}(s, z; M, Y) := k_1 (s \text{tr}(M), z \text{tr}(Y)), \quad w_{k_2}(s, z; n, v; M, Y) := k_2 (s([n] : M), z([v] : Y)), \\
w_{k_3}(s, z; n, v; M, Y) := k_3 (sMn, zYv), \quad w_{k_4}(s, z; M, Y) := (k_2 + k_3) [(sM^T, zY) - (s \text{tr}(M), z \text{tr}(Y))], \\
w_{b_1}(g, h) := b_1 (g, h), \quad w_{b_2}(g, h; n, v) := b_2 (g \cdot n, h \cdot v), \\
w_{b_3}(z; h; v; Y) := b_3 ((h \cdot v), z \text{tr}(Y)), \quad w_{b_4}(z; h; v; Y) := b_4 (h, zYv), 
\]

(2.14)

(2.15)

where we use “;” to separate disparate terms. With this, we have

\[
E_{\text{erk}}(s, n) = \frac{1}{2} [w_{k_1}(s, s; \nabla n, \nabla n) + w_{k_2}(s, s; n, n; \nabla n, \nabla n) + w_{k_3}(s, s; n, n; \nabla n, \nabla n) + w_{k_4}(s, s; \nabla n, \nabla n) \\
+ w_{b_1}(\nabla s, \nabla s) + w_{b_2}(\nabla s, \nabla s; n, n) + w_{b_3}(s; \nabla s; n, n) + w_{b_4}(s; \nabla s; n, n)]. 
\]

We will also consider a “stabilized” form of (2.12), \(i.e., let \(\theta > 0\) and define

\[
\widehat{\mathcal{W}}(s, g, n, M) := \mathcal{W}(s, g, n, M) + \theta s^2 |M^T n|^2, \quad w_\theta(s, z; n, v; M, Y) := \theta (sM^T n, zY^T v). 
\]

(2.16)

In this case, the energy functional becomes

\[
\widehat{E}_{\text{erk}}(s, n) = \frac{1}{2} \int_\Omega \widehat{\mathcal{W}}(s, \nabla s, n, \nabla n) \, dx = E_{\text{erk}}(s, n) + \frac{1}{2} w_\theta(s, s; n, n; \nabla n, \nabla n). 
\]

(2.17)

Note that if \(|n| = 1\) \(a.e.\), and \(n\) is sufficiently smooth, then \(n^T \nabla n = 0^T\); thus, \(|n^T \nabla n| \equiv 0\) and \(\widehat{E}_{\text{erk}}(s, n) = E_{\text{erk}}(s, n)\). In Section 3.3, \(\theta |n^T \nabla n|^2\) will play the role of a stabilization/consistency term.

**Proposition 2.6.** The energies (2.11), (2.17) are bounded on \(\mathcal{A}\), \(i.e.,\)

\[
E_{\text{erk}}(s, n) \leq \widehat{E}_{\text{erk}}(s, n) \leq C \left(\|\nabla s\|_{L^2(\Omega)}^2 + \|\nabla u\|_{L^2(\Omega)}^2\right) < \infty, 
\]

for all \((s, u, n)\) in \(\mathcal{A}\), where \(C > 0\) only depends on \(\{k_i\}_{i=1}^4, \{b_i\}_{i=1}^4\), and \(\theta\).

**Proof.** Follows by straightforward bounds. \(\square\)

We also need coercivity of (2.11), (2.17) over the admissible class (2.5), which requires certain inequality conditions \([34]\). To this end, define the following auxiliary coefficients:

\[
k'_1 := k_1 - \frac{b_2^2}{4[(b_1 + b_2) - 3\ell_0]}, \quad k'_3 := k_3 - \frac{b_4^2}{4[b_1 - 2\ell_0]}, 
\]

(2.18)
where $\ell_0 > 0$ is the “coercivity” constant. We assume the coefficients obey the following strict inequalities:

$$k'_1 - |k'_1 - k_2 - k_4| \geq 2\ell_0, \quad k_2 - |k_4| \geq 2\ell_0, \quad k'_3 \geq 2\ell_0, \quad b_1 > 2\ell_0, \quad b_1 + b_2 > 3\ell_0,$$

(2.19)

which implies that $k_1, k_3$ are bounded. With (2.19), we obtain the following theorem.

**Theorem 2.7.** Assume the dimension is $d = 3$ and assume (2.19) holds with a fixed constant $\ell_0 > 0$. Then,

$$\overline{W}(s, g, n, M) \geq \ell_0 (2|g|^2 + s^2|M|^2),$$

(2.20)

for all $s \in \mathbb{R}$, $n \in \mathbb{S}^{d-1}$, all $g \in \mathbb{R}^d$, and all $M \in \mathbb{R}^{d \times d}$, provided that $\theta > 0$ satisfies

$$\theta \geq \max \left\{ \ell_0^{-1}(b_1^2/4) + 2|k_1 - k_2 - k_4|^2 - 3\ell_0, \quad \ell_0^{-1}(k_2 + k_4)^2 + \ell_0 \right\}. \quad (2.21)$$

Furthermore,

$$W(s, g, n, M) \geq \ell_0 (2|g|^2 + s^2|M|^2),$$

(2.22)

for all $s \in \mathbb{R}$, $n \in \mathbb{S}^{d-1}$, all $g \in \mathbb{R}^d$, and all $M \in L(n, \mathbb{R}^d)$, where $L(n, \mathbb{R}^d) := \{ A \in \mathbb{R}^{d \times d} : n^T A = 0^T \}$.

**Proof.** The result follows by following the same arguments in [87], pages 125, 325 with some modification to account for $\theta|\mathbf{n}^T \mathbf{M}|^2$ and proving strict coercivity.

- **Corollary 2.8.** Assume the hypothesis of Theorem 2.7. Then,

$$\hat{E}_{\text{erk}}(s, n) \geq E_{\text{erk}}(s, n) \geq \ell_0 E_{\text{one}, 2}(s, n), \quad \text{for all} \quad (s, n) \in \mathcal{A}. \quad (2.23)$$

**Proof.** Let $(s, u, n) \in \mathcal{A}$. For any $\epsilon > 0$, we have that $n \in H^1(\{|s| > \epsilon\})$, which implies that $u^T(\nabla n) = 0^T$ a.e. in $\{|s| > \epsilon\}$. Hence, (2.22) is true a.e. in $\{|s| > \epsilon\}$. Integrating (2.22), we get (for all $\epsilon > 0$)

$$\int_{\{|s| > \epsilon\}} |\nabla s|^2 + |\nabla u|^2 = \ell_0 \int_{\{|s| > \epsilon\}} |\nabla s|^2 + |\nabla u|^2,$$

where we used Proposition 2.6 and (2.4). Thus, by the monotone convergence theorem,

$$E_{\text{erk}}(s, n) \geq \ell_0 \int_{\Lambda \setminus \{s = 0\}} (|\nabla s|^2 + |\nabla u|^2) = \ell_0 \bar{E}_{\text{one}, 2}(s, u) = \ell_0 E_{\text{one}, 2}(s, n),$$

where we used the fact that $|\nabla s| = 0$ a.e. on $\{s = 0\}$, as well as $|\nabla u| = 0$ a.e. on $\{u = 0\} \equiv \{s = 0\}$ (see Lem. A.3).

- **Remark 2.9** (Stabilization). In the continuous formulation, because $|n| = 1$ a.e., we have that $u^T(\nabla n) = 0^T$. In our finite element discretization (see Sect. 3.3), $u^T(\nabla n) \neq 0^T$ because $n$ only has unit length at the mesh nodes. Thus, one can think of $\theta|\mathbf{n}^T \nabla \mathbf{n}|^2$ as a “stabilization” term to handle this inconsistency.

- **Remark 2.10** (Erickson inequalities). The non-negativity of (2.13) was proved in [34,87] under the inequalities

$$k'_1 - |k'_1 - k_2 - k_4| \geq 0, \quad k_2 - |k_4| \geq 0, \quad k'_3 \geq 0, \quad b_1 > 0, \quad b_1 + b_2 > 0,$$

(2.24)

where

$$k'_1 := k_1 - \frac{b_3^2}{4(b_1 + b_2)}, \quad k'_3 := k_3 - \frac{b_4^2}{4b_1}. \quad (2.25)$$

These inequalities are less restrictive than (2.19), but they only ensure non-negativity; stronger assumptions are needed to enforce full coercivity over the admissible set (2.5). Setting $\ell_0 = 0$, we see that (2.19) reduces to (2.24). Therefore, (2.19) is a reasonable modification of (2.24) to ensure coercivity instead of just non-negativity.

Note that one can show that the pair of inequalities $k'_1 - |k'_1 - k_2 - k_4| \geq 2\ell_0$ and $k_2 - |k_4| \geq 2\ell_0$ is equivalent to $2k'_1 - k_2 - k_4 \geq 2\ell_0$ and $k_2 - |k_4| \geq 2\ell_0$. 
The following result is used in proving the weak lower semi-continuity of the discrete version of $\tilde{E}_{\text{erk}}$ (see Lem. A.21).

**Corollary 2.11.** Assume the hypothesis of Theorem 2.7 and note that the directional derivatives of $\tilde{W}$ and $W$ are given by

$$D_gW(s, g, n, M) \cdot h = D_g\tilde{W}(s, g, n, M) \cdot h = 2b_1g \cdot h + 2b_2(g \cdot n)(h \cdot n) + b_3s(h \cdot n)\text{tr}(M) + b_4sh \cdot Mn$$

$$D_MW(s, g, n, M) \cdot Y = 2s^2[k_11\text{tr}(M)\text{tr}(Y) + k_2([n]_x : M)([n]_x : Y) + k_3(Mn) : (Yn) + (k_2 + k_4)([M^T : Y] - \text{tr}(M)\text{tr}(Y))] + b_3s(g \cdot n)\text{tr}(Y) + b_4sg \cdot Yn$$

$$D_M\tilde{W}(s, g, n, M) \cdot Y = D_M\tilde{W}(s, g, n, M) \cdot Y + s^2\theta(M^Tn) \cdot (Y^Tn).$$

(2.26)

for all $(h, Y)$ in $\mathbb{R}^d \times \mathbb{R}^{d \times d}$. Then, $\tilde{W}(s, g, n, M)$ is convex with respect to $(g, M)$ in $\mathbb{R}^d \times \mathbb{R}^{d \times d}$ for all values of $s \in \mathbb{R}$ and $n \in S^{d-1}$, i.e.

$$\tilde{W}(s, g, n, M) \geq \tilde{W}(s, h, n, Y) + D_g\tilde{W}(s, h, n, Y) \cdot (g - h) + D_M\tilde{W}(s, h, n, Y) : (M - Y),$$

(2.28)

for all $(h, Y)$ in $\mathbb{R}^d \times \mathbb{R}^{d \times d}$. Similarly, $W(s, g, n, M)$ is convex with respect to $(g, M)$ in $\mathbb{R}^d \times L(n, \mathbb{R}^d)$ for all values of $s \in \mathbb{R}$ and $n \in S^{d-1}$.

**Proof.** For any given $s$ and $n$, $\tilde{W}(s, g, n, M)$ and $W(s, g, n, M)$ are quadratic functions of $g$ and $M$. Furthermore, by Theorem 2.7, $\tilde{W}$ and $W$ are non-negative. Hence, they must be convex. □

### 2.2. Weak anchoring

For LC droplets, the orientation of the LC molecules are influenced by the two-phase interface. This is usually modeled by adding a weak anchoring energy to the total energy of the system [87]. In the sharp interface setting, one adds an energy of the form $E = \int_\Gamma \gamma(\nu, n) \, dS$, where $\nu$ is the oriented unit normal vector of $\Gamma$ and $\gamma$ is a weak anchoring energy density function. One possible choice for $\gamma$ is given by [87]:

$$\gamma(\nu, n) = \frac{1}{2} (\alpha_{\perp}(\nu \cdot n)^2 + \alpha_{\parallel}[1 - (\nu \cdot n)^2]), \quad \alpha_{\perp}, \alpha_{\parallel} \geq 0,$$

(2.29)

where the first (second) term tends to make the minimizing director field $n$ perpendicular (parallel) to $\nu$. The weak anchoring energy function we take is similar and can be found in [71]. Let $E_a(s, n) := \beta_{a,n}E_{a,n}(s, n) + \beta_{a,s}E_{a,s}(s)$, where $\beta_{a,n}, \beta_{a,s} > 0$, and

$$E_{a,n}(s, n) := \frac{1}{2} \left( a_{\perp}(s, s; n, n) + a_{\parallel}(s, s; n, n) \right), \quad a_{\perp}(s, z; n, v) := (a_{\perp}s(n \cdot v), z(v \cdot v))_{\Gamma},$$

$$a_{\parallel}(s, z; n, v) := (a_{\parallel}s(n \otimes v), z(v \otimes v))_{\Gamma},$$

(2.30)

where we included the degree-of-orientation $s$ to model the loss of anisotropy when orientational order vanishes, and we add an energetic term penalizing $s$ to agree with $s_a$ on the interface:

$$E_{a,s}(s) := \frac{1}{2} \int_\Gamma \alpha_{\text{ori}}(s - s_a)^2 \, dS(x) = \frac{1}{2} \alpha_{\text{ori}}(s - s_a, s - s_a), \quad \alpha_{\text{ori}}(s, z) := (\alpha_{\text{ori}}s, z)_{\Gamma},$$

(2.31)

which is needed to ensure that $s$ does not trivially vanish on the interface, and so cause (2.30) to vanish as well [33, 63, 71]. The parameters $\alpha_{\perp}, \alpha_{\parallel}, \alpha_{\text{ori}} : \Gamma \to [0, \infty)$ allow for different weighting and the ability to model more general physical settings; throughout the paper, we assume $\alpha_{\perp}, \alpha_{\parallel}, \alpha_{\text{ori}}$ in $L^\infty(\Gamma)$. The derivation of (2.30), (2.31) (found in [71], Sect. 5.2.3) follows from the classic Rapini-Papoular type anchoring energy [13, 64] for Q-models. Note that other types of anchoring energies could be considered as well. For simplicity, we take the weight parameters $\beta_{a,n}, \beta_{a,s}$ to be non-dimensional (see Sect. 5.1).
2.3. Electro-statics

The LC can be coupled to other effects, such as external fields, which we now illustrate by incorporating electro-statics.

2.3.1. Dielectric permittivity

Due to the anisotropic nature of the LC molecules, the relative dielectric permittivity tensor of the material is modeled by [1, 19, 38, 64]

\[
\mathbf{\varepsilon}(s, n) := \mathbf{\varepsilon} + \varepsilon_a s \left( \mathbf{n} \otimes \mathbf{n} - \frac{1}{3} \mathbf{I} \right) = \left( \mathbf{\varepsilon} - s \frac{\varepsilon_a}{3} \right) \mathbf{I} + \varepsilon_a s (\mathbf{n} \otimes \mathbf{n}),
\]

which is a symmetric matrix, where \( \mathbf{\varepsilon} = (\mathbf{\varepsilon} \mathbf{\varepsilon} + 2\varepsilon_0) / 3, \varepsilon_a = \varepsilon_{\parallel} - \varepsilon_{\perp}, \gamma_a = \varepsilon_a / (3\varepsilon), \) and \( \varepsilon_{\parallel}, \varepsilon_{\perp} \) are positive. The eigenvalues of \( \mathbf{\varepsilon} \) are \( \mathbf{\varepsilon}_{1} (1-s\gamma_a), \mathbf{\varepsilon}_{3} (1-s\gamma_a) + \varepsilon_0 s, \) thus, since \(-1/2 < s < 1\), defining \( \varepsilon_{\min} = \min \{ \varepsilon_{\parallel}, \varepsilon_{\perp} \}, \varepsilon_{\max} = \max \{ \varepsilon_{\parallel}, \varepsilon_{\perp} \} \) we see that \( \mathbf{\varepsilon} \) is uniformly positive definite and satisfies

\[
\varepsilon_{\min} \leq |\mathbf{\varepsilon}(s, n)|_2 \leq \varepsilon_{\max}.
\]

2.3.2. Electro-static energy

The electric field \( \mathbf{E} \), in the LC domain, can be described by a potential function \( \varphi : \Omega \to \mathbb{R} \) [36], with \( \mathbf{E} = -\nabla \varphi \). Indeed, \( \varphi \) can be associated with an energy minimization principle [54]. Given \( (s, n) \in \mathcal{A} \) (fixed), define the dimensional electro-static energy as \( \beta_d J_d \), where \( \beta_d = \varepsilon_0 L_0 V_0^2 \), \( \varepsilon_0 \) is the permittivity of vacuum, \( V_0 \) is the voltage scale, and \( J_d \) is dimensionless [89]:

\[
J_d(\varphi; s, n) := \frac{1}{2} \int_\Omega \nabla \varphi \cdot \mathbf{\varepsilon}(s, n) \nabla \varphi \, dx - \int_\Omega \mathbf{P}(s, n) \cdot \nabla \varphi \, dx,
\]

where the (non-dimensional) polarization vector \( \mathbf{P} = \mathbf{P}(s, n) \) is given by

\[
\mathbf{P}(s, n) := \mathbf{P}_f(s, n) + \mathbf{P}_t(s, n), \quad \mathbf{P}_f(s, n) := f_1 s \nabla(\nabla n) + f_3 s(\nabla n) n, \quad \mathbf{P}_t(s, n) := r_1 (n \cdot \nabla s) n + r_2 \nabla s,
\]

where \( \mathbf{P}_f(s, n) \equiv f_1 s(\nabla n) n + f_3 s(\nabla n) n \), \( f_1, f_3 \) are relative (indefinite) flexoelectric parameters, and \( r_1, r_2 \) are relative (indefinite) order electric parameters (all non-dimensional), which models flexo electric and order electric effects induced by the LC [1, 24, 64]. The dimensional versions of \( f_1, f_3, r_1, r_2 \) are obtained by scaling with \( \varepsilon_0 V_0 \): possible physical values for \( |f_1|, |f_3| \) are on the order of \( 5 \times 10^{-12} \text{ C/m} [62] \).

Note that \( \mathbf{P} \equiv 0 \) a.e. in \( \{ s = 0 \} \subset \Omega \), i.e. \( \mathbf{P} \) vanishes when the material is isotropic. Furthermore, if \( (s, n) \in \mathcal{A} \), then

\[
\| \mathbf{P}(s, n) \|_{L^2(\Omega)} \leq C_f \| \nabla(\nabla n) \|_{L^2(\Omega)} + C_r \| \nabla s \|_{L^2(\Omega)} \leq C_p \left( \| \nabla s \|_{L^2(\Omega)} + \| \nabla \mathbf{u} \|_{L^2(\Omega)} \right) < \infty,
\]

where \( C_p > 0 \) is a uniform constant; thus, \( \mathbf{P}(s, n) \in L^2(\Omega) \) for all \( (s, n) \in \mathcal{A} \).

Let \( \varphi_0 \) in \( H^1(\Omega) \) and assume \( \varphi = \varphi_0 \) on the boundary \( \Gamma \), i.e. we fix the potential on \( \Gamma \). Then, for fixed \( (s, n) \) in \( \mathcal{A} \), the electrical potential \( \varphi \) is characterized as the unique minimizer of (2.34): \( \varphi = \arg \min_{\eta \in H^1_{\varphi_0}(\Omega)} J_d(\eta; s, n) \), where the admissible set \( H^1_{\varphi_0}(\Omega) := \{ \eta \in H^1(\Omega) : \eta = \varphi_0, \text{ on } \Gamma \} \) accounts for the boundary conditions. It is convenient to define \( \tilde{\varphi} = \varphi - \varphi_0 \) and separate the boundary condition. In this case, minimization problem is equivalent to finding \( \tilde{\varphi} \) in \( H^1(\Omega) \) such that

\[
J_d(\tilde{\varphi}; s, n) := \frac{1}{2} \epsilon (\tilde{\varphi} + \varphi_0, \tilde{\varphi} + \varphi_0; \mathbf{\varepsilon}(s, n)) - (\mathbf{P}(s, n), \nabla(\tilde{\varphi} + \varphi_0)),
\]
is minimized over $H_0^1(\Omega)$, where $e(\varphi, \eta; \varepsilon(s, n)) := (\nabla \varphi \varepsilon(s, n), \nabla \eta)$. Setting the first variation of (2.37) to zero, while holding $(s, n)$ fixed, we obtain the Euler–Lagrange equation in weak form: find $\tilde{\varphi}$ in $H_0^1(\Omega)$ such that

$$e(\tilde{\varphi}, \eta; \varepsilon(s, n)) = -e(\varphi_0, \eta; \varepsilon(s, n)) + (P(s, n), \nabla \eta), \quad \text{for all } \eta \in H_0^1(\Omega).$$

(2.38)

For later use, we let $T: \mathcal{A} \to H_0^1(\Omega)$ denote the solution operator for (2.38), i.e., $\tilde{\varphi} = T(s, n)$ solves (2.38). Note that the strong form solution of (2.38) is given by:

$$-\nabla \cdot (\varepsilon(s, n) \nabla \tilde{\varphi}) = \nabla \cdot (\varepsilon(s, n) \nabla \varphi_0) - \nabla \cdot P(s, n), \quad \text{in } \Omega, \quad \tilde{\varphi} = 0, \quad \text{on } \Gamma.$$  

(2.39)

2.3.3. Contribution to LC energy

The electrical energy contribution to the total liquid crystal energy is given by [1, 19, 38, 64]:

$$E_{el}(s, n) := -J_{el}(T(s, n); s, n) = -J_{el}(\tilde{\varphi}; s, n)$$

$$= -\frac{1}{2} e(\tilde{\varphi}, \tilde{\varphi}; \varepsilon(s, n)) - \frac{1}{2} e(\varphi_0, \varphi_0; \varepsilon(s, n)) - e(\varphi_0, \tilde{\varphi}; \varepsilon(s, n)) + (P(s, n), \nabla (\tilde{\varphi} + \varphi_0)).$$

(2.40)

Note the minus sign, which is connected to the fact that the potential $\varphi$ is fixed on the boundary [36]; see [89] for a first principles derivation.

We emphasize that $\tilde{\varphi}$ is not an independent variable in the liquid crystal energy minimization we consider in (2.43); $\tilde{\varphi}$ is determined uniquely for any given $(s, n)$ in $\mathcal{A}$. In fact, this leads to a useful identity. Setting $\eta = \tilde{\varphi}$ in (2.38) implies $e(\tilde{\varphi}, \tilde{\varphi}; \varepsilon(s, n)) = -e(\varphi_0, \tilde{\varphi}; \varepsilon(s, n)) + (P(s, n), \nabla \tilde{\varphi})$, and plugging into (2.40) yields

$$E_{el}(s, n) = \frac{1}{2} e(\tilde{\varphi}, \tilde{\varphi}; \varepsilon(s, n)) - \frac{1}{2} e(\varphi_0, \varphi_0; \varepsilon(s, n)) + (P(s, n), \nabla \varphi_0),$$

(2.41)

which essentially states that $E_{el}$ is convex in $\nabla \tilde{\varphi}$. This is used in Section 2.4 to show that the total energy is bounded below.

2.4. Total energy

The total energy we seek to minimize is defined to be

$$E(s, n) = \beta_{erk} \left( E_{erk}(s, n) + \frac{1}{\epsilon_{dw}} E_{dw}(s) \right) + \beta_{a,n} E_{a,n}(s, n) + \beta_{a,s} E_{a,s}(s) + \beta_{el} E_{el}(s, n),$$

(2.42)

for constant weights $\beta_{erk}, \epsilon_{dw} > 0, \beta_{a,n}, \beta_{a,s}, \beta_{el} \geq 0$ defined earlier. The minimization problem for $E$ is then

$$(s^*, n^*) = \arg \min_{(s,n) \in \mathcal{A}(g, q)} E(s, n).$$

(2.43)

The energy (2.42) is bounded below by the following argument. From (2.41) and (2.33), and using a Cauchy inequality, we have

$$E_{el}(s, n) \geq \frac{1}{2} \varepsilon_{\min} \|\nabla \tilde{\varphi}\|_{L^2(\Omega)}^2 - \frac{1}{2} \varepsilon_{\max} \|\nabla \varphi_0\|_{L^2(\Omega)}^2 - \frac{1}{2\delta} \|P(s, n)\|_{L^2(\Omega)}^2 - \frac{\delta}{2} \|\nabla \varphi_0\|_{L^2(\Omega)}^2,$$

for some $\delta > 0$. And by (2.36), this reduces to

$$E_{el}(s, n) \geq \frac{1}{2} \varepsilon_{\min} \|\nabla \tilde{\varphi}\|_{L^2(\Omega)}^2 - (C_0 + \delta) \|\nabla \varphi_0\|_{L^2(\Omega)}^2 - \frac{C_{P}}{\delta} \left( \|\nabla s\|_{L^2(\Omega)}^2 + \|\nabla u\|_{L^2(\Omega)}^2 \right)^2.$$  

(2.44)

Next, since $E_{dw}, E_{a,n},$ and $E_{a,s}$ are non-negative, we bound (2.42) below by

$$E(s, n) \geq \beta_{erk} E_{erk}(s, n) + \beta_{el} E_{el}(s, n) \geq \left( \beta_{erk} \ell_0 + \beta_{el} \frac{C_{P}}{\delta} \right) \left( \|\nabla s\|_{L^2(\Omega)}^2 + \|\nabla u\|_{L^2(\Omega)}^2 \right)$$

$$- \beta_{el} (C_0 + \delta) \|\nabla \varphi_0\|_{L^2(\Omega)}^2,$$

using (2.23), (2.4) and (2.44). Choosing $\delta > 0$ sufficiently large (depending on fixed parameters), we find that the total energy is bounded below by a uniform constant $C_1 > 0$ that only depends on the fixed parameters of the problem, i.e., $E(s, n) \geq -C_1$, for all $(s, n) \in \mathcal{A}$. 


3. Finite element scheme

3.1. Domain approximation

Let \( \Omega \) be a Lipschitz domain. Moreover, we assume \( \Omega \) is polyhedral and discretize it by a conforming set of simplicial elements, denoted \( T_h = \{ T \} \), and let \( N_h \) be the set of nodes of \( T_h \) with cardinality \( |N_h| \). Moreover, the boundary \( \Gamma \) is represented by simplicial elements of co-dimension 1 that are embedded in \( T_h \). Furthermore, the mesh is assumed to be shape regular \[22,26\]. We do not assume the mesh is weakly acute, which was needed in \[70\] to prove convergence of the finite element scheme.

**Remark 3.1.** The polyhedral assumption allows us to avoid dealing with a variational crime \[22,26\] related to the approximation of the domain.

3.2. Finite element spaces

The following finite element spaces are used in discretizing the energy:

- \( S_h := \{ s_h \in H^1(\Omega) : s_h|_T \text{ is affine for all } T \in T_h \} \),
- \( U_h := \{ u_h \in [H^1(\Omega)]^d : u_h|_T \text{ is affine in each component for all } T \in T_h \} \),
- \( N_h := \{ n_h \in U_h : n_h(x_i) = 1 \text{ for all nodes } x_i \in N_h \} \),
- \( V_h := \{ v_h \in H^1(\Omega) : v_h|_T \text{ is affine for all } T \in T_h \} \),

where \( N_h \) imposes the unit length constraint at the vertices of the mesh.

Let \( I_h \) denote the piecewise linear Lagrange interpolation operator on the mesh \( T_h \) with values in either \( S_h, U_h, \) or \( V_h \). Mimicking (2.5) at the discrete level, we have

\[
\mathcal{A}_h := \{ (s_h, u_h, n_h) \in S_h \times U_h \times N_h : (s_h, u_h, n_h) \text{ satisfies (3.3), with } u_h \in U_h \},
\]

where (3.2) is called the discrete structural condition of \( \mathcal{A}_h \). Note: if we write \( (s_h, u_h, n_h) \) in \( \mathcal{A}_h \), we mean that \( (s_h, u_h) \) in \( \mathcal{A}_h \), \( u_h \) in \( U_h \), and \( (s_h, u_h, n_h) \) satisfies (3.3).

Next, let \( g_h := I_h g, r_h := I_h r \), and \( q_h := I_h q \) be the discrete Dirichlet data, where \( g_h \) automatically satisfies (2.7). Note that the interpolant of \( q \) is well defined in an open neighborhood of \( \Gamma_n \) (because \( q \in [W^{1,\infty}(\Omega)]^d \) near \( \Gamma_n \subset \Gamma_{|s| \geq h} \)). Wherever \( q \) lacks the regularity \( [W^{1,\infty}(\Omega)]^d \), set \( q_h := e_1 \). Therefore, the discrete spaces that include (Dirichlet) boundary conditions are

\[
S_h(\Gamma, g_h) := \{ s_h \in S_h : s_h|_{\Gamma} = g_h \}, \quad U_h(\Gamma_u, r_h) := \{ u_h \in U_h : u_h|_{\Gamma_u} = r_h \},
\]

\[
N_h(\Gamma_n, q_h) := \{ n_h \in N_h : n_h|_{\Gamma_n} = q_h \}.
\]

The discrete admissible class with boundary conditions is given by

\[
\mathcal{A}_h(g_h, q_h) := \{ (s_h, u_h, n_h) \in \mathcal{A}_h : s_h \in S_h(\Gamma, g_h), n_h \in N_h(\Gamma_n, q_h) \}.
\]

Note: we use a similar abuse of notation as before when writing \( (s_h, u_h, n_h) \) in \( \mathcal{A}_h(g_h, q_h) \). Boundary conditions for the electric field are enforced via the space \( V_h \cap H^1_0(\Omega) \).

3.3. Discrete Ericksen energy

We will utilize the following discrete \( L^2 \) inner products:

\[
(u, v)^h_{D_h} := \sum_{T \in T_h} \int_T I_h(uv), \quad (u, v)_D^h := \sum_{T \in T_h} \int_T I_h(u \cdot v), \quad (M, Y)_D^h := \sum_{T \in T_h} \int_I h(M \cdot Y),
\]

where \( T \in T_h \) are tetrahedral elements in the mesh of \( \Omega, \tilde{T} \subset T_h, D_h = \cup_{T \in T_h} T, \) and the function arguments are polynomial functions over each element (possibly discontinuous across element edges). We write \((u, v)^h := (u, v)^h_{\Omega}\) when integrating over \( \Omega \); integrals over subsets will have a subscript.
3.3.1. Lumping

We require a “lumped” form of the discrete Ericksen energy. Let \( s_h \in S_h, \mathbf{n}_h \in N_h \) and consider their restriction to an element \( T \in T_h \) (note that \( \nabla s_h \) and \( \nabla \mathbf{n}_h \) are discontinuous across \( \partial T \)). By Theorem 2.7 and (2.20), and setting \( s = s_h|_T, \mathbf{n} = \mathbf{n}_h|_T, \mathbf{g} = \nabla s_h|_T, \mathbf{M} = \nabla \mathbf{n}_h|_T \), we have that

\[
\begin{align*}
\widehat{\mathcal{W}}(s_h, \nabla s_h, \mathbf{n}_h, \nabla \mathbf{n}_h)(x_i) & \geq \ell_0 \left( 2|\nabla s_h|^2 + s_h^2|\nabla \mathbf{n}_h|^2 \right) \bigg|_{x=x_i} \\
\end{align*}
\]  
(3.6)

holds at each node \( x_i \in T \), because \( |\mathbf{n}_h| = 1 \) at the nodes. Therefore, we define the discrete (stabilized) Ericksen energy to be

\[
\begin{align*}
\widehat{E}_{\text{erk}}^h(s_h, \mathbf{n}_h) := \frac{1}{2} \sum_{T \in T_h} \int_T I_h \widehat{\mathcal{W}}(s_h, \nabla s_h, \mathbf{n}_h, \nabla \mathbf{n}_h) \, dx = \frac{1}{2} \left( \widehat{\mathcal{W}}(s_h, \nabla s_h, \mathbf{n}_h, \nabla \mathbf{n}_h), 1 \right)^h.
\end{align*}
\]  
(3.7)

By (3.6), (A.5), we see that

\[
\begin{align*}
\widehat{E}_{\text{erk}}^h(s_h, \mathbf{n}_h) & \geq \frac{\ell_0}{2} \int_\Omega \left( 2|\nabla s_h|^2 + s_h^2|\nabla \mathbf{n}_h|^2 \right) \, dx = \ell_0 E_{\text{one}, 2}(s_h, \mathbf{n}_h),
\end{align*}
\]  
(3.8)

where we used that \( \nabla s_h, \nabla \mathbf{n}_h \) are constant on \( T \). Clearly, (3.7) is non-negative for all \( h \). So, by finite dimensional optimization theory [68], \( \widehat{E}_{\text{erk}}^h \) has a minimizer.

It will be useful later to write (3.7) in terms of various forms. We define the discrete forms \( \{w_k^h\}_{k=1}^4 \) in the same way as (2.14), (2.15), except we use the discrete inner products (3.5). Therefore, we obtain

\[
\begin{align*}
\widehat{E}_{\text{erk}}^h(s_h, \mathbf{n}_h) = \frac{1}{2} \left[ w_{k_1}^h (s_h, s_h; \nabla \mathbf{n}_h, \nabla \mathbf{n}_h) + w_{k_2}^h (s_h, s_h; \mathbf{n}_h, \nabla \mathbf{n}_h) + w_{k_3}^h (s_h, s_h; \mathbf{n}_h, \mathbf{n}_h; \nabla \mathbf{n}_h, \nabla \mathbf{n}_h) + w_{k_4}^h (s_h, s_h; \mathbf{n}_h, \mathbf{n}_h; \nabla \mathbf{n}_h, \nabla \mathbf{n}_h) \right. \\
\left. + w_{k_4}^h(s_h, s_h; \nabla \mathbf{n}_h, \nabla \mathbf{n}_h) + w_{k_4}^h(s_h, s_h; \mathbf{n}_h, \nabla \mathbf{n}_h) + w_{k_4}^h(s_h, s_h; \mathbf{n}_h, \mathbf{n}_h; \nabla \mathbf{n}_h, \nabla \mathbf{n}_h) + w_{k_4}^h (\nabla s_h, \nabla s_h) \right].
\end{align*}
\]  
(3.9)

Next, we express each of the terms in (3.9) in a slightly modified form that will be convenient in later sections. For instance, defining \( W_h = \{v \in L^2(\Omega) : v \text{ is constant on each } T \in T_h \} \), and taking \( s_h, z_h \in S_h, \mathbf{n}_h, \mathbf{v}_h \in U_h \), and \( \mathbf{M}_h, \mathbf{Y}_h \in [W_h]^{d \times d} \), we have for \( w_{k_3}^h \):

\[
\begin{align*}
k_3^{-1} w_{k_3}^h (s_h, z_h; \mathbf{n}_h, \mathbf{v}_h; \mathbf{M}_h, \mathbf{Y}_h) &= (s_h(\mathbf{M}_h)\mathbf{n}_h, z_h(\mathbf{Y}_h)\mathbf{v}_h)^h = \sum_{T \in T_h} \int_T I_h \left\{ s_h z_h [(\mathbf{M}_h)\mathbf{n}_h] \cdot [(\mathbf{Y}_h)\mathbf{v}_h] \right\} \\
&= \sum_{T \in T_h} \int_T I_h \left\{ s_h z_h \mathbf{n}_h \otimes \mathbf{v}_h \right\} : (\mathbf{M}^T_h \mathbf{Y}_h) = \left( I_h \left\{ s_h z_h \mathbf{n}_h \otimes \mathbf{v}_h \right\}, \mathbf{M}^T_h \mathbf{Y}_h \right),
\end{align*}
\]  
(3.10)

where \( \mathbf{M}_h, \mathbf{Y}_h \) are pulled out of \( I_h \) because they are constant on each element \( T \in T_h \). Similar arguments yield the discrete versions of (2.14)–(2.16):

\[
\begin{align*}
w_{k_1}^h (s_h, z_h; \mathbf{M}_h, \mathbf{Y}_h) &= k_1 \left( I_h \left\{ s_h z_h \right\}, \text{tr}(\mathbf{M}_h)\text{tr}(\mathbf{Y}_h) \right), \\
w_{k_2}^h (s_h, z_h; \mathbf{n}_h, \mathbf{v}_h; \mathbf{M}_h, \mathbf{Y}_h) &= k_2 \left( I_h \left\{ s_h z_h \mathbf{n}_h \right\} \otimes \mathbf{v}_h, \mathbf{M}_h \otimes \mathbf{Y}_h \right), \\
w_{k_3}^h (s_h, z_h; \mathbf{n}_h, \mathbf{v}_h; \mathbf{M}_h, \mathbf{Y}_h) &= k_3 \left( I_h \left\{ s_h z_h \mathbf{n}_h \otimes \mathbf{v}_h \right\}, \mathbf{M}^T_h \mathbf{Y}_h \right), \\
w_{k_4}^h (s_h, z_h; \mathbf{M}_h, \mathbf{Y}_h) &= (k_2 + k_3) \left( I_h \left\{ s_h z_h \mathbf{n}_h \right\}, \mathbf{M}^T_h \mathbf{Y}_h \right), \\
w_{k_4}^h (s_h, z_h; \mathbf{M}_h, \mathbf{Y}_h) &= \theta \left( I_h \left\{ s_h z_h \mathbf{n}_h \otimes \mathbf{v}_h \right\}, \mathbf{M}^T_h \mathbf{Y}_h \right),
\end{align*}
\]  
(3.11)
The discrete electro-static problem is as follows. Let \( \phi \) be an energy function that is bounded below. Moreover, we have a result similar to (2.41). Setting \( \eta \) as before, the contribution to the LC energy is \( E_n(s_h, n_h) := \beta_{n,n} E_{a,n}^h(s_h, n_h) + \beta_{n,a} E_{a,a}^h(s_h) \).

### 3.3.2 Double well energy

The double well energy \( E_{dw}(\cdot) \) is discretized in the usual way: \( E_{dw}^h(s_h) := \int_{\Omega} \psi(s_h(x)) \, dx \). In our numerical minimization scheme (Sect. 5), we use a convex splitting [82,83,91] of \( E_{dw}(s_h) \).

### 3.4 Discrete weak anchoring energy

Let \( s_h \in S_h, n_h \in N_h \) and define the discrete (weak) anchoring energy for the director similarly to (2.30): \( E_{a,n}^h(s_h, n_h) := E_{a,n}(s_h, n_h) = \frac{1}{2} (a_\perp(s_h, s_h; n_h, n_h) + a_\parallel(s_h, s_h; n_h, n_h)) \). For the degree of orientation, we have \( E_{a,a}^h(s_h) := E_{a,a}(s_h) = \frac{1}{2} \int_{\Gamma} a_{ori}(s_h - s_a)^2 \, dS(x) = \frac{1}{2} a_{ori} (s_h - s_a, s_h - s_a) \). Therefore, the total anchoring energy is

\[
E_a^h(s_h, n_h) := \beta_{a,n} E_{a,n}^h(s_h, n_h) + \beta_{a,a} E_{a,a}^h(s_h).
\]

### 3.5 Discrete electric energy

We discretize the dielectric permittivity tensor in the obvious way, i.e. \( \varepsilon = \varepsilon(s_h, n_h) \) (recall (2.32)), which satisfies the same bounds in (2.33):

\[
\varepsilon_{\text{min}} \leq |\varepsilon(s_h, n_h)|_2 \leq \varepsilon_{\text{max}}, \quad \text{for all } (s_h, n_h) \in A_h.
\]

For the electro-static problem, we use a standard discretization, i.e. replace \((s, n)\) with \((s_h, n_h)\). Hence, the discrete electro-static problem is as follows. Let \( \varphi_{0,h} \in V_h \) be the elliptic projection of \( \varphi_0 \) (A.6). Given \((s_h, n_h) \in A_h \) (fixed), find \( \tilde{\varphi}_h \) in \( V_{h,0} := V_h \cap H_0^1(\Omega) \) such that

\[
J_{el}^h(\tilde{\varphi}_h; s_h, n_h) := \frac{1}{2} (\tilde{\varphi}_h + \varphi_{0,h} - \tilde{\varphi}_h + \varphi_{0,h} - \varphi(s_h, n_h)) - (P(s_h, n_h), \nabla(\tilde{\varphi}_h + \varphi_{0,h})),
\]

is minimized over \( V_{h,0} \).

The corresponding discrete version of (2.38) is: find \( \tilde{\varphi}_h \) in \( V_{h,0} \) such that

\[
\varepsilon \left( \tilde{\varphi}_h, \eta_h; \varepsilon(s_h, n_h) \right) = -\varepsilon \left( \varphi_{0,h}, \eta_h; \varepsilon(s_h, n_h) \right) - (P(s_h, n_h), \nabla \eta_h),
\]

for all \( \eta_h \in V_{h,0} \). Let \( T_h : A_h \to V_{h,0} \) denote the solution operator for (3.16), i.e. \( \tilde{\varphi}_h := T_h(s_h, n_h) \) solves (3.16).

As before, the contribution to the LC energy is \( E_{el}^h(s_h, n_h) := -J_{el}^h(T_h(s_h, n_h); s_h, n_h) = -J_{el}^h(\tilde{\varphi}_h; s_h, n_h) \). Moreover, we have a result similar to (2.41). Setting \( \eta_h = \varphi_{0,h} \) in (3.16) implies

\[
\varepsilon \left( \varphi_{0,h}, \tilde{\varphi}_h; \varepsilon(s_h, n_h) \right) = -\varepsilon \left( \varphi_{0,h}, \tilde{\varphi}_h; \varepsilon(s_h, n_h) \right) + (P(s_h, n_h), \nabla \varphi_{0,h}),
\]

and plugging into \( E_{el}^h(s_h, n_h) \) yields

\[
E_{el}^h(s_h, n_h) = \frac{1}{2} \varepsilon \left( \varphi_{0,h}, \tilde{\varphi}_h; \varepsilon(s_h, n_h) \right) - \frac{1}{2} \varepsilon \left( \varphi_{0,h}, \varphi_{0,h}; \varepsilon(s_h, n_h) \right) + (P(s_h, n_h), \nabla \varphi_{0,h}),
\]

which essentially states that \( E_{el}^h \) is convex in \( \nabla \varphi_{0,h} \). This is used in Section 4.2 to show that the total discrete energy is bounded below.
3.6. Discrete total energy

The total (discrete) energy we seek to minimize is defined to be

\[ E^h(s_h, n_h) = \beta_{\text{erk}} \left( E_{\text{erk}}^h(s_h, n_h) + \frac{1}{\epsilon_{d\text{w}}} E_{d\text{w}}^h(s_h) \right) + \beta_{a,n} E_{a,n}^h(s_h, n_h) + \beta_{a,s} E_{a,s}^h(s_h) + \beta_{\text{el}} E_{\text{el}}^h(s_h, n_h). \]  

(3.18)

The minimization problem for \( E^h \) is: \( (s_h^*, n_h^*) = \arg\min_{(s_h,n_h)\in A(g_h,q_h)} E^h(s_h, n_h) \). We show that the total discrete energy is bounded below in Section 4.2.

4. Γ-convergence of the FEM

We show that the finite element approximation of the discrete energy (3.18) Γ-converges to the continuous energy (2.42). The result presented here is not the same as the result shown in [70,71] or in [33], all of which used a special discretization of the Ericksen energy that is limited to the one constant approximation (5.2). Furthermore, their discretization requires the underlying mesh to be weakly acute in order to prove Γ-convergence of their method; the weakly acute assumption is quite severe for three-dimensional meshes [51,52,86].

In contrast, our method has the following advantages: (a) no assumption is made on the mesh structure (other than being shape regular); (b) the Ericksen energy can be very general (not just the one-constant approximation); (c) the method non-linearly couples full electro-statics, which was not done previously. Therefore, our result is more general than in [33,70,71].

4.1. Main result

We begin with some preliminaries before stating the main Γ-convergence result. The discrete energy \( E^h(s_h, n_h) \) is defined on \( \mathbb{Z}_h := S_h \times N_h \), but convergence cannot be insured for a sequence \( (s_h, n_h) \in \mathbb{Z}_h \), because \( n_h \) will not (in general) converge on the singular set \( S \). However, we can guarantee convergence for \( (s_h, u_h) \in \mathbb{X}_h := S_h \times U_h \), i.e. \( u_h \) is well-behaved. Thus, Theorem 4.1 is a minor modification of the usual definition of Γ-convergence [21,31].

To this end, we define the continuous space to be \( \mathbb{X} := L^2(\Omega) \times [L^2(\Omega)]^d \), and note that \( \mathbb{X}_h \subset \mathbb{X} \) and \( \mathbb{Z}_h \subset \mathbb{X} \). Next, the continuous energy \( E : \mathbb{X} \rightarrow \mathbb{R} \) is defined as follows: \( E(s,n) \) is given by (2.42) if \( (s,n) \in \mathbb{A}(g,q) \), and set \( E(s,n) = \infty \) if \( (s,n) \in \mathbb{X} \setminus \mathbb{A}(g,q) \). Likewise, define the discrete energy \( E^h(s_h, n_h) \) by (3.18) if \( (s_h, n_h) \in \mathbb{A}(g_h,q_h) \), and set \( E^h(s,n) = \infty \) if \( (s,n) \in \mathbb{X} \setminus \mathbb{A}(g_h,q_h) \).

**Theorem 4.1 (Γ-convergence).** Given \((s,n) \in \mathbb{X}, \) where \(|n| = 1\) a.e., define the corresponding element \((s,u) \in \mathbb{X}, \) where \( u := sn. \) In addition, given \((s_h,n_h) \in \mathbb{Z}_h, \) define the corresponding element \((s_h,u_h) \in \mathbb{X}_h, \) where \( u_h := I_h(s_h,n_h). \) Let \( \{T_h\} \) be a sequence of shape regular meshes. Then, under Assumptions 2.4 and 2.5, the following properties hold:

- Lim-inf inequality. For every sequence \((s_h,n_h) \in \mathbb{Z}_h \subset \mathbb{X}, \) such that the corresponding sequence \((s_h,u_h) \in \mathbb{X}_h \subset \mathbb{X} \) converges strongly to the corresponding pair \((s,u), \) we have

\[ E(s,n) \leq \liminf_{h \to 0} E^h(s_h, n_h); \]  

(4.1)

- Lim-sup inequality. There exists a sequence \((s_h,n_h) \in \mathbb{Z}_h \subset \mathbb{X} \) such that the corresponding sequence \((s_h,u_h) \in \mathbb{X}_h \subset \mathbb{X} \) converges strongly to the corresponding pair \((s,u), \) and

\[ E(s,n) \geq \limsup_{h \to 0} E^h(s_h, n_h). \]  

(4.2)

In the following sections, we build up several intermediate results which are used to prove Theorem 4.1.
4.2. Bounded below

Lemma 4.2 (Coercivity). Adopt the hypothesis of Lemma A.13. Then,

$$E_{\text{one},1}(s_h, n_h) \geq \|\nabla s_h\|^2_{L^2(\Omega)} + \|\nabla u_h\|^2_{L^2(\Omega)}, \quad \text{for all } (s_h, n_h) \in \mathcal{A}_h,$$

(4.3)

where \(\gamma_0 > 0\) only depends on the shape regularity of the mesh \(T_h\).

Proof. The first inequality is trivial. For the second, we use (A.8) to get

$$\|\nabla u_h\|_{L^2(\Omega)} \leq \|\nabla (u_h - s_h n_h)\|_{L^2(\Omega)} + \|\nabla (s_h n_h)\|_{L^2(\Omega)} \leq C \|\nabla s_h\|_{L^2(\Omega)} + \|\nabla s_h \otimes n_h\|_{L^2(\Omega)} + \|s_h \nabla n_h\|_{L^2(\Omega)} \leq (C + 1)\|\nabla s_h\|_{L^2(\Omega)} + \|s_h \nabla n_h\|_{L^2(\Omega)}.$$  

Since \(E_{\text{one},1}(s_h, n_h) = \frac{1}{2} (\|\nabla s_h\|^2_{L^2(\Omega)} + \|s_h \nabla n_h\|^2_{L^2(\Omega)})\), we obtain the assertion with \(\gamma_0 = 1/(4(C + 1)^2)\). \qed

The discrete energy (3.18) is bounded below by the following argument. From (3.17) and (2.33), and using a Cauchy inequality, we have

$$E_{\text{el}}^h(s_h, n_h) \geq \varepsilon_{\min} \|\nabla \tilde{\varphi}_h\|^2_{L^2(\Omega)} - \frac{1}{2} \varepsilon_{\max} \|\nabla \varphi_{0,h}\|^2_{L^2(\Omega)} - \frac{1}{2} \|P(s_h, n_h)\|^2_{L^2(\Omega)} - \frac{\delta}{2} \|\nabla \varphi_{0,h}\|^2_{L^2(\Omega)},$$

for some \(\delta > 0\). And by the discrete version of (2.36), this reduces to

$$E_{\text{el}}^h(s_h, n_h) \geq \varepsilon_{\min} \|\nabla \tilde{\varphi}_h\|^2_{L^2(\Omega)} - (C_0 + \delta) \|\nabla \varphi_{0,h}\|^2_{L^2(\Omega)} - \frac{C_0}{\delta} \left(\|\nabla s_h\|^2_{L^2(\Omega)} + \|\nabla u_h\|^2_{L^2(\Omega)}\right)^2.$$  

(4.4)

Next, since \(E_{\text{dw}}^h, E_{a,n}^h, E_{n,n}^h\) are non-negative, we bound (3.18) below by

$$E^h(s_h, n_h) \geq \beta_{\text{er}} \tilde{E}_{\text{er}}^h(s_h, n_h) + \beta_\text{el} E_{\text{el}}^h(s_h, n_h) \geq \left(\frac{\beta_{\text{er}} \tilde{A}}{2} - \beta_\text{el} \frac{C_0}{\delta}\right) \left(\|\nabla s_h\|^2_{L^2(\Omega)} + \|\nabla u_h\|^2_{L^2(\Omega)}\right) - \beta_\text{el}(C_0 + \delta) \|\nabla \varphi_{0,h}\|^2_{L^2(\Omega)},$$

using (3.8), Lemma 4.2, and (4.4); note: \(\tilde{A} > 0\) is a uniform constant independent of \(h > 0\). Choosing \(\delta > 0\) sufficiently large (depending on fixed parameters), and noting that \(\|\nabla \varphi_{0,h}\|_{L^2(\Omega)} \leq C \|\nabla \varphi_0\|^2_{L^2(\Omega)}\), we find that the total discrete energy is bounded below: \(E^h(s_h, n_h) \geq -\tilde{C}_1\), for all \((s_h, n_h) \in \mathcal{A}_h\), where \(\tilde{C}_1 > 0\) is a uniform constant independent of \(h\).

4.3. Recovery sequence

In proving the lim-sup part of Theorem 4.1, we break it up into the following lemmas. The existence of a discrete sequence is given by Lagrange interpolation, which is then shown to deliver a recovery sequence for the Ericksen energy, double-well energy, weak anchoring energy, and the electrical energy.

Lemma 4.3. Assume the hypothesis of Lemma A.17. Moreover, assume that \((s, u, n) \in \mathcal{A}(g, q)\) also satisfies

$$-1/2 + 1/k \leq s \leq 1 - 1/k \text{ for some } k \geq 1.$$  

Then there exists a sequence \((s_h, u_h, n_h) \in \mathcal{A}_h(g_h, q_h)\), converging in the sense of Lemma A.17, such that

$$E_{\text{er}}(s, n) = \tilde{E}_{\text{er}}(s, n) = \lim_{h \to 0} \tilde{E}_{\text{er}}^h(s_h, n_h), \quad E_{\text{dw}}(s) = \lim_{h \to 0} E_{\text{dw}}^h(s_h), \quad E_a(s, n) = \lim_{h \to 0} E_a^h(s_h, n_h).$$

Proof. First, we show that \(\lim_{h \to 0} \tilde{E}_{\text{er}}^h(s_h, n_h) = \tilde{E}_{\text{er}}(s, n)\). By Lemma A.18, we only need to show that

$$\lim_{h \to 0} \left(\tilde{W}(s_h, \nabla s_h, n_h, \nabla n_h), 1\right) - \left(\tilde{W}(s, \nabla s, n, \nabla n), 1\right) \to 0.$$  

(4.5)
We demonstrate this for one of the terms in (3.9); the other terms follow by similar arguments. First, we consider $w^h_{k_3}$ and show that $G^h_{k_3} := |(\phi_h(\nabla \eta_h) \mathbf{n}_h) - (\phi(\nabla \mathbf{n}) s(\nabla \mathbf{n}))| \to 0$. Fix $\varepsilon > 0$. Since $s_h \to s$, $\mathbf{n}_h \to \mathbf{n}$ in $W^{1, \infty}(\Omega \setminus \mathcal{S}_c)$, it is clear that $\int_{\Omega \setminus \mathcal{S}_c} s^2_h |(\nabla \eta_h) \mathbf{n}_h|^2 \to \int_{\Omega \setminus \mathcal{S}_c} s^2 |(\nabla \mathbf{n}) \mathbf{n}|^2$. On the other hand, using (A.8), for $h > 0$ sufficiently small, we have

$$\int_{\mathcal{S}_c} s^2_h |(\nabla \eta_h) \mathbf{n}_h|^2 \leq \|s_h \nabla \eta_h\|_{L^2(\mathcal{S}_c)}^2 \leq C \left( \|s \nabla \eta\|_{L^2(\mathcal{S}_c)}^2 + \|s \nabla \mathbf{n}\|_{L^2(\mathcal{S}_c)}^2 \right) \leq C \left( \|s \nabla \eta\|_{L^2(\mathcal{S}_c)}^2 + \|s \nabla \mathbf{n}\|_{L^2(\mathcal{S}_c)}^2 \right),$$

for all $\varepsilon > 0$. Ergo, $\lim_{h \to 0} \|s_h \nabla \eta_h \mathbf{n}_h\|_{L^2(\mathcal{S}_c)}^2 \leq C \left( \|s \nabla \eta\|_{L^2(\mathcal{S}_c)}^2 + \|s \nabla \mathbf{n}\|_{L^2(\mathcal{S}_c)}^2 \right)$, for all $\varepsilon > 0$. So, taking $\varepsilon \to 0$ and using the monotone convergence theorem, we get

$$\lim_{h \to 0} G^h_{k_3} \leq C \left( \int_{\{s = 0\}} |s|^2 + \int_{\{|u = 0\}} |u|^2 \right) = 0,$$

where we used Lemma A.3. Therefore, this shows that

$$w^h_{k_3} (s_h, s_h, \mathbf{n}_h, \mathbf{n}_h; \nabla \eta_h, \nabla \mathbf{n}_h) \to w_{k_3} (s, s, \mathbf{n}, \mathbf{n}; \nabla \mathbf{n}, \nabla \mathbf{n}), \quad \text{as } h \to 0.$$  

By similar reasoning, we get that $w^h_{i} \to w_{i}$, for $1 \leq i \leq 4$, $w^h_{\vartheta} \to w_{\vartheta}$, and $w^h_{b} \to w_{b}$, for $1 \leq i \leq 4$.

Thus, we have shown that $E^h_{erk}(s_h, \mathbf{n}_h) \to E_{erk}(s, \mathbf{n})$ as $h \to 0$. Furthermore, note that $w_{\vartheta} (s, s, \mathbf{n}, \mathbf{n}; \nabla \mathbf{n}, \nabla \mathbf{n}) = 0$, because $(s, \mathbf{u}, \mathbf{n}) \in \mathcal{A}(g, q)$, which implies that $E_{erk}(s, \mathbf{n}) = E_{erk}(s, \mathbf{n})$.

Next, we show that $E^h_{\text{el}}(s_h) \to E_{\text{el}}(s)$ as $h \to 0$, i.e. $\int_{\Omega} \psi(s_h) \to \int_{\Omega} \psi(s)$, as $h \to 0$. Since $s_h$ is piecewise linear, by hypothesis $-1/2 + 1/k \leq s_h \leq 1 - 1/k$ for all $h > 0$. Thus, $\psi(s_h)$ is bounded uniformly in $h$, and $\psi(s)$ is also bounded. Since $\psi(s_h) \to \psi(s)$ a.e. in $\Omega$, the dominated convergence theorem implies that $\int_{\Omega} \psi(s_h) \to \int_{\Omega} \psi(s)$.

Finally, taking advantage of strong convergence in $L^2(\Gamma)$, we get convergence of the anchoring energy:

$$\lim_{h \to 0} E^h_{a}(s_h, \mathbf{n}_h) = E_{a}(s, \mathbf{n}).$$

Lemma 4.4 (Recovery of electrical energy). Assume the hypothesis of Lemma A.17. Moreover, assume that $(s, \mathbf{u}, \mathbf{n}) \in \mathcal{A}(g, q)$ also satisfies $-1/2 + 1/k \leq s \leq 1 - 1/k$ for some $k \geq 2$. Then there exists a sequence $(s_h, \mathbf{u}_h, \mathbf{n}_h) \in \mathcal{A}_h(g_h, q_h)$, converging in the sense of Lemma A.17, such that $E_{\text{el}}(s, \mathbf{n}) = \lim_{h \to 0} E^h_{\text{el}}(s_h, \mathbf{n}_h)$.

Proof. First, we must show that the sequence of solutions to (3.16) $\{\tilde{\varphi}_h\}_{h>0}$ converges as $h \to 0$, and that the limit solves the electro-static problem. Let $\eta_h = \eta_h(\eta)$, where $\eta \in C^\infty_0(\Omega)$; clearly $\eta_h \to \eta$ in $H^1_0(\Omega)$. Next, we show that $\mathbf{P}_l(s_h, \mathbf{u}_h, \nabla \eta_h) \to \mathbf{P}_l(s, \mathbf{u}, \nabla \eta)$ and $\mathbf{P}_b(s_h, \mathbf{u}_h, \nabla \eta_h) \to \mathbf{P}_b(s, \mathbf{u}, \nabla \eta)$. The arguments are similar to the proof of Lemma 4.3, so we will focus on one term in $\mathbf{P}_l$, i.e. show that

$$G^h_{l_1}(\Omega) := (\nabla s_h \cdot \mathbf{n}_h, \nabla \eta_h \cdot \mathbf{n}_h) - (\nabla s \cdot \mathbf{n}, \nabla \eta \cdot \mathbf{n}) \to 0, \quad \text{as } h \to 0.$$  

Fix $\varepsilon > 0$. Since $s_h \to s$, $\mathbf{n}_h \to \mathbf{n}$ in $W^{1, \infty}(\Omega \setminus \mathcal{S}_c)$, it is clear that $G^h_{l_1}(\Omega \setminus \mathcal{S}_c) \to 0$ as $h \to 0$. On the other hand, by the stability of the interpolant, we have

$$\int_{\mathcal{S}_c} (\nabla s_h \cdot \mathbf{n}_h) \nabla \eta_h \cdot \mathbf{n}_h \leq \|\nabla s_h\|_{L^2(\mathcal{S}_c)} \|\nabla \eta_h\|_{L^2(\mathcal{S}_c)} \leq C_1 \|\nabla s\|_{L^2(\mathcal{S}_c)} \|\nabla \eta\|_{L^2(\mathcal{S}_c)}.$$  

Ergo, $\lim_{h \to 0} |G^h_{l_1}(\Omega)| \leq (C_1 + 1) \|\nabla s\|_{L^2(\mathcal{S}_c)} \|\nabla \eta\|_{L^2(\mathcal{S}_c)}$, for all $\varepsilon > 0$. So, taking $\varepsilon \to 0$ and using the monotone convergence theorem, we get $\lim_{h \to 0} |G^h_{l_1}(\Omega)| \leq (C_1 + 1) \left( \int_{\{s = 0\}} |\nabla s|^2 \right)^{1/2} \|\nabla \eta\|_{L^2(\Omega)} = 0$, because $\nabla s = 0$ a.e. in $\{s = 0\}$ (see Lem. A.3).
Note that the permittivity tensor $\varepsilon(s_h, n_h)$ converges to $\varepsilon(s, n)$ a.e. in $\Omega$, using similar arguments as in Lemma 4.3. Next, choosing $\eta_h = \tilde{\varphi}_h$ in (3.16) and using (3.14), we find that
\[
\varepsilon_{\min} \|\nabla \tilde{\varphi}_h\|^2_{L^2(\Omega)} \leq \varepsilon_{\max} \left( \frac{c_1}{2} \|\nabla \varphi_0, h\|^2_{L^2(\Omega)} + \frac{1}{2c_1} \|\nabla \tilde{\varphi}_h\|^2_{L^2(\Omega)} \right) + \frac{c_2}{2} \|\mathbf{P}(s_h, n_h)\|^2_{L^2(\Omega)} + \frac{1}{2c_2} \|\nabla \tilde{\varphi}_h\|^2_{L^2(\Omega)},
\]
where $c_1, c_2 > 0$ are to be chosen. Upon recalling (2.36), and using the stability of the interpolant, we have that $\|\mathbf{P}(s_h, n_h)\|_{L^2(\Omega)}$ is uniformly bounded for all $h > 0$. Choosing $c_1, c_2$ sufficiently large, we find that $\|\nabla \tilde{\varphi}_h\|_{L^2(\Omega)} \leq C < \infty$, for all $h > 0$ for some fixed constant $C > 0$. Thus, $\tilde{\varphi}_h \rightharpoonup \varphi$ in $H^1_0(\Omega)$.

Furthermore, $\varepsilon(s_h, n_h)\nabla \eta_h^T \rightarrow \varepsilon(s, n)\nabla \eta^T$ in $L^2(\Omega)$ by Lebesgue’s dominated convergence theorem. So, combining with the weak convergence of $\tilde{\varphi}_h$, we see that $\int_{\Omega} \nabla \tilde{\varphi}_h \varepsilon(s_h, n_h)\nabla \eta_h^T \rightarrow \int_{\Omega} \nabla \tilde{\varphi} \varepsilon(s, n)\nabla \eta^T$. Thus, combining with the convergence of the other terms in (3.16), we see that $\tilde{\varphi} = T(s, n)$ solves (2.38) with data $(s, n)$.

Next, we must show that $J_{el}(\tilde{\varphi}_h; s_h, n_h) \rightarrow J_{el}(\tilde{\varphi}; s, n)$. For this, we must show that $\tilde{\varphi}_h \rightharpoonup \varphi$ in $H^1_0(\Omega)$ (strong convergence). Let $P_h \varphi \in V_h$ be the elliptic projection of $\varphi$ (A.6). Similar to the previous inequality, we have
\[
\varepsilon_{\min} \|\nabla \varphi - \tilde{\varphi}\|^2_{L^2(\Omega)} \leq \int_{\Omega} \nabla (\tilde{\varphi} - \varphi) \varepsilon(s_h, n_h) \nabla (\tilde{\varphi} - \varphi)^T = \int_{\Omega} \nabla \tilde{\varphi} \varepsilon(s_h, n_h) \nabla (P_h \varphi - \varphi)^T + \int_{\Omega} \nabla \tilde{\varphi} \varepsilon(s_h, n_h) \nabla (\varphi - P_h \varphi)^T
\]
\[
+ \int_{\Omega} \nabla \tilde{\varphi} \varepsilon(s_h, n_h) \nabla (\varphi - P_h \varphi)^T = T_1^h + T_2^h + T_3^h.
\]

Since $P_h \varphi \rightarrow \varphi$ in $H^1_0(\Omega)$, and $\varepsilon(s_h, n_h)$ is uniformly bounded, $\lim_{h \rightarrow 0} T_1^h = 0$. For $T_2^h$, use the discrete problem (3.16) with data $(s_h, n_h)$:
\[
\int_{\Omega} \nabla \tilde{\varphi} \varepsilon(s_h, n_h) \nabla (\varphi - P_h \varphi)^T = - \int_{\Omega} \nabla \varphi_0, h \varepsilon(s_h, n_h) \nabla (\varphi - P_h \varphi)^T
\]
\[
+ (P_f(s_h, n_h), \nabla (\varphi - P_h \varphi)) + (P_f(s_h, n_h), \nabla (\varphi - P_h \varphi)) \rightarrow 0,
\]
by utilizing both weak and strong convergence, i.e. $\mathbf{P}(s_h, n_h) \rightarrow \mathbf{P}(s, n)$ strongly in $L^2(\Omega)$. Lastly, $T_3^h \rightarrow 0$ because $\nabla \tilde{\varphi} \varepsilon(s_h, n_h) \rightarrow \nabla \varphi \varepsilon(s, n)$ strongly in $L^2(\Omega)$, and $\nabla (\varphi - P_h \varphi) \rightarrow 0$ weakly in $L^2(\Omega)$.

Therefore, we find that $\nabla \tilde{\varphi}_h \rightarrow \nabla \varphi$ strongly in $L^2(\Omega)$. From this, we obtain that $J_{el}(\tilde{\varphi}_h; s_h, n_h) \rightarrow J_{el}(\tilde{\varphi}; s, n)$, which of course implies $E_{el}(\tilde{\varphi}_h; s_h, n_h) \rightarrow E_{el}(\tilde{\varphi}; s, n)$.

\begin{theorem}[Recovery sequence] Suppose Assumptions 2.4 and 2.5 hold. Let $(s, n, u) \in A(g, q)$. Then there exists a sequence $(s_h, u_h, n_h) \in \mathcal{A}(g_h, q_h)$, such that $(s_h, u_h)$ converges to $(s, u)$ in $H^1(\Omega)$, as well as $n_h \in N_h$ converging to $n$ in $L^2(\Omega \setminus S)$, such that
\[
E(s, n) = \lim_{h \rightarrow 0} E^h(s_h, n_h).
\]
\end{theorem}

\begin{proof}
This proof follows by combining Lemmas A.17, 4.3, 4.4, with Lemma A.9. First, note that we can assume $E(s, n) < \infty$ (otherwise, the result is trivial). Given $k \geq 1$, by Lemma A.9, there exists $(s_{\delta_k}, u_{\delta_k}, n_{\delta_k}) \in \mathcal{A}(g, q)$, with $\delta_k > 0$ sufficiently small, so that $|E(s_{\delta_k}, u_{\delta_k}) - E(s, n)| \leq \frac{1}{k}$, and moreover $(s_{\delta_k}, u_{\delta_k}) \rightarrow (s, u)$ in $[H^1(\Omega)]^{d+1}$, and $n_{\delta_k} \rightarrow n$ in $[L^2(\Omega \setminus S)]^d$. Thus, with $k > 0$ being a given integer, one can choose $\delta_k > 0$ sufficiently small so that $\| (s, u) - (s_{\delta_k}, u_{\delta_k}) \|_{H^1(\Omega)} < k^{-1}$, $\|n - n_{\delta_k}\|_{L^2(\Omega \setminus S)} < k^{-1}$.

Next, by Lemma A.17, for each fixed $k$ there exists discrete functions $(s_h, u_h, n_h) \in \mathcal{A}(g_h, q_h)$ such that $(s_h, u_h) \rightarrow (s_{\delta_k}, u_{\delta_k})$ in $[H^1(\Omega)]^{d+1}$, and $n_h \rightarrow n_{\delta_k}$ in $[L^2(\Omega \setminus S)]^d$ as $h \rightarrow 0$. Moreover, Lemmas 4.3, 4.4 imply that
\[
\lim_{h \rightarrow 0} E^h(s_h, n_h) = E(s_{\delta_k}, n_{\delta_k}).
\]
Whence, for each $\delta_k$, we may choose $h_k$ sufficiently small so that $|E(s_{h_k}, n_{h_k}) - E(s_{\delta_k}, n_{\delta_k})| \leq k^{-1}$, and $\| (s_{\delta_k}, u_{\delta_k}) - (s_{h_k}, u_{h_k}) \|_{H^1(\Omega)} < k^{-1}$, $\|n_{\delta_k} - n_{h_k}\|_{L^2(\Omega \setminus S)} < k^{-1}$. The assertion then follows by applying the triangle inequality.
\end{proof}
4.4. Proof of main result

Proof of Theorem 4.1. Lim-inf. Let \((s_h, u_h, n_h) \in \mathcal{A}_h(g_h, q_h)\) be any sequence. Without loss of generality, assume there is a constant \(\Lambda > 0\) such that \(\liminf_{h \to 0} E^h(s_h, n_h) \leq \Lambda\), for otherwise there is nothing to prove.

Combining (3.8) with Lemma 4.2, yields that \(\|s_h\|_{H^1(\Omega)}, \|u_h\|_{H^1(\Omega)}\) are uniformly bounded with respect to \(h > 0\). Whence, there is a subsequence (not relabeled) \((s_k, u_k)\) that converges weakly to \((s, u) \in \mathcal{A}\). By Lemma A.16, there exists a \(n \in L^2(\Omega)\) such that \([n] = 1\) a.e. in \(\Omega\) and \(u = sn\) a.e. in \(\Omega\). Furthermore, by a trace Sobolev embedding, we have that \(s = g\) on \(\Gamma_s\), and \(n = q\) on \(\Gamma_n\), ergo \((s, u, n) \in \mathcal{A}(g, q)\).

Note that Fatou’s lemma implies that \(\liminf_{h \to 0} E_{\text{dw}}(s_h) \geq E_{\text{dw}}(s)\), because \(s_h \to s\) a.e. in \(\Omega\). Therefore, combining Lemmas A.21–A.23, we obtain

\[
\liminf_{h \to 0} E^h(s_h, n_h) \geq E(s, n).
\]

Lim-sup. Let \((s, u, n) \in \mathcal{A}(g, q)\), for otherwise \(E(s, n) = +\infty\) so the result is trivial. The existence of a convergent sequence satisfying the necessary properties follows by Theorem 4.5.

\[\square\]

Corollary 4.6 (convergence of global discrete minimizers). Let \(\{T_h\}\) be a sequence of conforming shape-regular triangulations. If \((s_h, n_h) \in \mathcal{A}_h(g_h, q_h)\) is a sequence of global minimizers of \(E^h(s_h, n_h)\) in (3.18), then every cluster point is a global minimizer of the continuous energy \(E(s, n)\) in (2.42).

Proof. Follows from the usual \(\Gamma\)-convergence arguments \([21,31]\). \[\square\]

This implies existence of global minimizers of (2.42), and convergence of global minimizers of (3.18) to global minimizers of (2.42), along with convergence of the discrete energy to the continuous energy. Note that this result does not yield a rate of convergence, though first order is expected for \((s_h, u_h)\) in most situations (see [71] for an example).

5. Numerical results

We use an alternating direction minimization algorithm, similar to what is in \([33,63,70,71]\), for finding discrete (local) minimizers of \(E^h\). In addition, we use a line search to ensure that the energy decreases at each step. This is due to two reasons: the lack of monotonicity when projecting (normalizing) \(n\) to unit length \((c.f. \ [71], \ Thm. 8)\) and the presence of the electro-static PDE-constraint. An alternative method could be to use a Newton iteration, as described in \([40,77]\).

We implemented our method using the MATLAB/C++ finite element toolbox FELICITY \([88]\). For all 3-D simulations, we used the algebraic multi-grid solver (AGMG) \([65,66,72,73]\) to solve the linear systems for updating \(n\) and \(s\), as well as solving the electro-static equation (3.16). In 2-D, we simply used the “backslash” command in MATLAB.

5.1. Non-dimensionalization

We assume the following dimensional scales in the numerical experiments: \(k_0 = 1.5 \times 10^{-11} \text{ J/m}\) and \(L_0 = 77.5 \times 10^{-9} \text{ m}\), which gives \(\beta_{\text{erk}} = 1.1625 \times 10^{-18} \text{ J}\). The other constants are \(A'_0 = 10^4 \text{ J/m}^3\), which gives the (dimensionless) double well coefficient \(\epsilon_{\text{dw}} = (0.5)^{-2}, \alpha_0 = 9.5 \times 10^{-3} \text{ J/m}^2\), \(V_0 = 1.84\) or 2.9 Volts, and recall that \(\epsilon_0 = 8.854187817 \times 10^{-12} \text{ C/(V m)}\).

Next, we non-dimensionalize the simple Ericksen energy in (2.1) following a similar procedure as in [39]. Note that \(s\) and \(n\) are already non-dimensional. Let \(A_0'\) be the characteristic scale for the double well (see Rem. 2.2), and define \(\epsilon_{\text{dw}} := \sqrt{k_0/(A_0' L_0^2)},\) where \(L_0 = \text{diam}(\Omega)\) is the length scale. Then, (2.1) can be written as

\[
J(s, n) = k_0L_0 \left( E_{\text{one}, k_0}(s, n) + \frac{1}{\epsilon_{\text{dw}}^2} E_{\text{dw}}(s) \right) , \quad E_{\text{dw}}(s) := \int_{\Omega} \tilde{\psi}(s) \, dx = (\tilde{\psi}(s), 1) , \quad \tilde{\psi}(s) = \frac{1}{A_0} \psi(s), \quad (5.1)
\]
Figure 2. Disk domain with two holes (Sect. 5.2.1). Arrows depict the director \( n \). Color scale is based on the degree-of-orientation parameter \( s \). Weak (normal) anchoring is imposed on all boundaries. Some “defect” regions can be seen around the upper right hole. (A) Erk. coefs: \( k_1 = k_2 = k_3 = 1 \). (B) Erk. coefs: \( k_1 = 1, k_2 = k_3 = 0.25 \).

\[
E_{\text{one}, \bar{b}_0}(s, n) := \frac{1}{2} \int_\Omega \left( \bar{b}_0 |\nabla s|^2 + s^2 |\nabla n|^2 \right) \, dx = \frac{1}{2} \left[ \bar{b}_0 (\nabla s, \nabla s) + (s \nabla n, s \nabla n) \right],
\]

where \( \bar{b}_0 = \frac{b_0}{k_0} \), \( \bar{\psi}(s) \), \( E_{\text{one}, \bar{b}_0} \), and \( E_{\text{dw}} \) are non-dimensional, as well as the domains. The general energy density (2.13) is non-dimensionalized in a similar way, i.e. define \( k_0 := \max(k_1, k_2, k_3) \) and set \( \bar{k}_i := k_i/k_0 \), and \( \bar{b}_i := b_i/k_0 \), for \( 1 \leq i \leq 4 \). For the weak anchoring, \( \beta_{a,n} = \beta_{a,s} = \alpha_0 L_0^2 \), where \( \alpha_0 \) has units of J/m\(^2\), and \( \Gamma \) and \( \alpha_\perp, \alpha_\parallel, \alpha_{ori} \) are already non-dimensional. We normalize \( \beta_{a,n}, \beta_{a,s}, \) and \( \beta_{el} \) by \( \beta_{erk} \); hence, the non-dimensional value for \( \beta_{erk} \) is always unity. For each experiment, we list dimensionless values for \( \beta_{a,n}, \beta_{a,s}, \) and \( \beta_{el} \). All domains are (at least approximately) unit size. For simplicity of notation, we drop the “bar” from the non-dimensional quantities.

5.2. Disk with holes

5.2.1. Normal anchoring

The domain is taken to be a disk (of radius 0.6) with two holes (see Fig. 2). Weak anchoring is used on all boundaries with parameters given by

\[
\beta_{a,n} = \beta_{a,s} = 50, \quad \alpha_\parallel = 1, \quad \alpha_\perp = 0, \quad \alpha_{ori} = 1,
\]

which yields normal (homeotropic) anchoring. The (non-dimensional) double well potential \( \psi(s) \), for \( -\frac{1}{2} < s < 1 \), is

\[
\psi(s) := 5.2403 - 11.6667 s^2 - 27.7778 s^3 + 41.6667 s^4,
\]
Figure 3. Disk domain with two holes (Sect. 5.2.2); similar format to Figure 2. Weak (planar) anchoring is imposed on all boundaries. Some “defect” regions can be seen around both holes in (a); only the lower left hole has a decreases order parameter in (b). (A) Erk. coefs: $k_1 = k_2 = k_3 = 1$. (B) Erk. coefs: $k_1 = 1, k_2 = k_3 = 0.25$.

Figure 4. Disk domain with two holes (Sect. 5.2.3); similar format to Figure 2. Electro-static effects are turned on and weak (normal) anchoring is imposed on all boundaries. Some “defect” regions can be seen around the lower left hole. (A) Erk. coefs: $k_1 = k_2 = k_3 = 1$. (B) Erk. coefs: $k_1 = 1, k_2 = k_3 = 0.25$. 
with a local maximum at \( s = 0 \) and global minimum at \( s = s_a := 0.7 \). The initial conditions in \( \Omega \) for the gradient flow are: \( s = s_a \) and \( n \) given by a point defect at \((0.552, 0.46)^T\).

The first set of values for the Ericksen constants are

\[
    k_1 = 1, \quad k_2 = 1, \quad k_3 = 1, \quad k_4 = 0, \quad b_1 = 1, \quad b_2 = b_3 = b_4 = 0,
\]

with stabilization parameter \( \theta = 3.3341 \), effective coercivity constant is \( \ell_0 = 0.3332 \), and \( \beta_{\text{erk}} = 1 \). The results of this simulation are shown in Figure 2a.

The next simulation changes two parameters only: \( k_2 = k_3 = 0.25 \); the rest are identical. This yields a stabilization parameter \( \theta = 8.6316 \) and effective coercivity constant \( \ell_0 = 0.1249 \). The results are shown in Figure 2b which is not very different from Figure 2a.

5.2.2. Planar anchoring

In this numerical experiment, the exact same setup is used as in Figure 2a, except the weak anchoring coefficients are

\[
    \beta_{a,n} = \beta_{a,s} = 50, \quad \alpha_{\parallel} = 0, \quad \alpha_{\perp} = 1, \quad \alpha_{\text{ori}} = 1,
\]

which yields planar anchoring. The double well potential is the same as in (5.4). Same initial conditions are used.

The first set of values for the Ericksen constants are the same as in (5.5). The results of this simulation are shown in Figure 3a.

The next simulation changes two parameters only: \( k_2 = k_3 = 0.25 \); the rest are identical. The results are shown in Figure 3b which vary significantly from Figure 3a. The director field “swirls” more because \( k_2, k_3 \) are lower so bending is not penalized as much.
5.2.3. Normal anchoring with electric effect

This example uses the exact same setup as in Section 5.2.1, except now the electric field is turned on. The electro-static parameters are

\[
\beta_1 = 2, \quad \varepsilon_\parallel = 5, \quad \varepsilon_\perp = 1, \quad f_1 = 1, \quad f_3 = -1, \quad r_1 = r_2 = 0,
\]

with the boundary condition given by: \(\varphi_0 = x + y\). We start with the “one-constant” approximation, i.e. \(k_1 = k_2 = k_3 = 1\). The results of this simulation are shown in Figure 4a.

The next simulation changes two parameters only: \(k_2 = k_3 = 0.25\); the rest are identical. The results are shown in Figure 4b which is not very different from Figure 4a.

5.2.4. Planar anchoring with electric effect

This example uses the exact same setup as in Section 5.2.2, except now the electric field is turned on. The electro-static parameters are the same as in (5.7). We start with the “one-constant” approximation, i.e. \(k_1 = k_2 = k_3 = 1\). The results of this simulation are shown in Figure 5a.

The next simulation changes two parameters only: \(k_2 = k_3 = 0.25\); the rest are identical. The results are shown in Figure 5b which vary somewhat from Figure 5a. The anisotropic electric field parameters drastically affect the solution relative to no electric field in Figure 3.

5.3. Freedericksz transition

5.3.1. Off and On

The domain is taken to be a unit cube: \(\Omega := [0, 1]^3\) (see Fig. 6). Weak anchoring is not used; the boundary conditions are:
Figure 7. Cube domain with electric field (Sect. 5.3.1). Similar format to Figure 6. The director field $\mathbf{n}$ is driven to point vertically because of the electric effect, which demonstrates the classic Freedericksz Transition. Again, the $s$ variable is nearly constant here. (A) View of the $y$-$z$ plane. (B) Oblique view.

Figure 8. Cube domain with flexo-electric effect $f_1 = 1$ (Sect. 5.3.2). Similar format to Figure 7. The director field is drastically affected by the flexo-electric effect. There are no defect regions. (A) View of the $y$-$z$ plane. (B) Oblique view.
Figure 9. Cube domain with flexo-electric effect $f_3 = 1$ (Sect. 5.3.2). Similar format to Figure 7. The director field is again drastically affected by the flexo-electric effect. There are no defect regions. (A) View of the $y$-$z$ plane. (B) Oblique view.

\[
\mathbf{n} = (1, 0, 0)^T, \quad s = s_a, \quad \text{on } [0, 1]^2 \times \{0\}, \quad \mathbf{n} = (0, 1, 0)^T, \quad s = s_a, \quad \text{on } [0, 1]^2 \times \{1\},
\]

(5.8)

with a vanishing Neumann condition on the other sides of the cube. The double well potential is the same as in (5.4) with $\epsilon_{dw} := (0.5)^{-2}$. The initial conditions in $\Omega$ for the gradient flow are: $s = s_a$ and $\mathbf{n} = (0, 1, 0)^T$ constant.

The Ericksen constants are

\[
k_1 = k_2 = k_3 = 1, \quad b_1 = 1, \quad b_2 = b_3 = b_4 = 0,
\]

(5.9)

with stabilization parameter $\theta = 3.3341$, effective coercivity constant is $\ell_0 = 0.3332$, and $\beta_{erk} = 1$. The results of this simulation are shown in Figure 6. Essentially, $\mathbf{n} \cdot \mathbf{e}_z = 0$ throughout, with a smooth rotation from the bottom plane to the top plane.

Next, we turn the electric field on with parameters given by

\[
\beta_{el} = 5, \quad \varepsilon_\parallel = 5, \quad \varepsilon_\perp = 1, \quad f_1 = 0, \quad f_3 = 0, \quad r_1 = r_2 = 0,
\]

(5.10)

i.e. no flexo-electric effects are present. The boundary condition is given by: $\varphi_0 = z$. The results of this simulation are shown in Figure 7.

5.3.2. Flexo-electric

In this numerical experiment, we use the same conditions as in Section 5.3.1, except that the flexo-electric parameters are

\[
f_1 = 1, \quad f_3 = 0, \quad r_1 = r_2 = 0.
\]

(5.11)
Figure 10. Torus domain (Sect. 5.4.1), with different views (A) and (B). Arrows depict the director $n$. Color scale is based on the degree-of-orientation parameter $s$. Weak (planar) anchoring is imposed on the boundary. No defects are present. Erk. coefs: $k_1 = k_2 = k_3 = 1$.

The results of this simulation are shown in Figure 8. The director field is significantly affected by the flexo-electric effect. Recall (2.35), where $f_1$ is connected with $\text{div } n$.

Next, we change the flexo-electric parameters to

$$f_1 = 0, \quad f_3 = 1, \quad r_1 = r_2 = 0. \quad (5.12)$$

The results of this simulation are shown in Figure 9. The director field exhibits a twisting motion with axis aligned along the $x$ direction. Note that $f_3$ is connected with $n \times \text{curl } n$.

5.4. Torus

5.4.1. Planar anchoring

The domain is taken to be a torus with two radii 0.155 and 0.3 (see Fig. 10). Weak anchoring is used on all boundaries with parameters given by

$$\beta_{n,n} = \beta_{n,s} = 50, \quad \alpha_\parallel = 0, \quad \alpha_\perp = 1, \quad \alpha_{\text{ori}} = 1, \quad (5.13)$$

which yields planar anchoring. The double well potential is the same as in (5.4). The initial conditions in $\Omega$ for the gradient flow are: $s = s_a$ and $n$ a perturbed rotating vector field.

The first set of values for the Ericksen constants are

$$k_1 = 1, \quad k_2 = 1, \quad k_3 = 1, \quad k_4 = 0, \quad b_1 = 1, \quad b_2 = b_3 = b_4 = 0, \quad (5.14)$$

with stabilization parameter $\theta = 3.3341$, effective coercivity constant is $\ell_0 = 0.3332$, and $\beta_{\text{erk}} = 1$. The results of this simulation are shown in Figure 10.
Figure 11. Torus domain (Sect. 5.4.1), with different views (A) and (B). Similar format as in Figure 10. The director field twists along the torus in order to avoid bending, which is more heavily penalized by the $k_3$ term. No defects are present. Erk. coefs: $k_1 = k_2 = 0.1$, $k_3 = 1$.

Figure 12. Torus domain with electric field (Sect. 5.4.2), with different views (A) and (B). Similar format as in Figure 10. The electric field has no significant effect relative to Figure 10. Erk. coefs: $k_1 = k_2 = k_3 = 1$.

The next simulation changes two parameters only: $k_1 = k_2 = 0.1$; the rest are identical. This yields a stabilization parameter $\theta = 0.2503$ and effective coercivity constant $\ell_0 = 0.049905$. The results are shown in
Figure 13. Torus domain with electric field (Sect. 5.4.2), with different views (A) and (B). Similar format as in Figure 10. The director field twists along the torus in order to avoid bending, which is more heavily penalized by the $k_3$ term. No defects are present, but $s$ varies more than the previous cases. Erk. coefs: $k_1 = k_2 = 0.1$, $k_3 = 1$.

Figure 11 which shows the director field developing a “twist” along the torus. This is understandable since $k_1$ and $k_2$ are much smaller than $k_3$, i.e. it is energetically favorable for the director field to develop a twist in order to avoid bending around the torus.

5.4.2. Planar anchoring with electric field

In this case, everything is the same as in Figure 10, except now we turn the electric field on with parameters given by

$$\beta_{el} = 5, \; \varepsilon_\parallel = 5, \; \varepsilon_\perp = 1, \; f_1 = 1, \; f_3 = 0, \; r_1 = r_2 = 0,$$

(5.15)

The boundary condition is given by: $\varphi_0 = z$. The results of this simulation are shown in Figure 12. Note that the flexo-electric term $f_1$ does not really play a role here because $|\text{div} \mathbf{n}| \approx 0$.

The next simulation changes the following parameters only: $k_1 = k_2 = 0.1$, and $f_1 = 0$, $f_3 = 1$; the rest are identical. The results are shown in Figure 13. Similar to Figure 11, the director field twists along the torus in order to avoid pure bending. However, the $f_3$ flexo-electric term causes the director field to distort further (note the lighter colored areas in Fig. 13b).

6. Conclusions

We have presented a finite element method for the generalized Ericksen model of liquid crystals, which can account for electro-static effects and weak anchoring conditions. The method is shown to converge in the sense of $\Gamma$-convergence for global minimizers, without requiring the mesh to be weakly acute. A key part of the method uses mass lumping (different from what is in [70,71]) to give stability.
Using a simple iterative minimization scheme with line search, we computed discrete minimizers for three different examples to illustrate the method. The numerical experiments illustrate the effect of varying the Ericksen constants; this has a direct effect on the form of the minimizers. Furthermore, the electric field can augment the director field considerably if $\beta_{el}$ is large enough.

The main advantage of the method is that it does not need a weakly acute mesh. This allows for modeling LCs with the Ericksen system on general geometries, without the need for a separate treatment of the boundary (such as in [33, 63]). This has the potential for enabling shape optimization problems related to liquid crystals, such as optimizing colloidal particles interacting with LCs.

One future direction of our method is to extend it to handle line fields, i.e. enforce the equivalence of $\pm n$ (see Rem. 2.1). Moreover, our approach could be extended to modeling and simulating the packing of DNA strands inside viral capsids [8, 60]. The idea here is to treat the DNA strand like an anisotropic material and model the material state with a director field [50]. There are many applications for this kind of modeling, from basic science [58, 79] to more practical applications [44].

**Appendix A. Auxiliary results**

**A.1. Elementary analysis**

The following convergence result is basic to everything that follows.

**Lemma A.1.** Let $(s, u, n) \in \mathcal{A}$, and suppose $\{(s_\delta, u_\delta, n_\delta)\}_{\delta > 0} \subset \mathcal{A}$ is a sequence such that $(s_\delta, u_\delta) \to (s, u)$ in $[H^1(\Omega)]^{d+1}$, as $\delta \to 0$. Then, for any subset $D$ of $\Omega$, we have

$$\int_D s_\delta^2|\nabla n_\delta|^2 \to \int_D s^2|\nabla n|^2, \text{ as } \delta \to 0.$$  

**Proof.** This follows easily from the identity (2.4). □

We note a basic compactness result regarding traces (see [67], Cor. 7.2, [27], Thm. 6.6-3, and [27], Thm. 6.6-5).

**Theorem A.2.** Let $\Omega$ be a bounded Lipschitz domain in $\mathbb{R}^d$. Then, for $d = 2$ or $3$,

$$\|u\|_{L^2(\Gamma)} \leq C\|u\|_{W^{1, p}(\Omega)}, \text{ for all } 2 \leq p \leq \infty,$$

where $C > 0$ only depends on $\Omega$ and $\Gamma$. Moreover, the trace operator on $\Gamma$, as a map from $W^{1, 2}(\Omega) \to L^2(\Gamma)$, is compact.

The singular set $\{s = 0\}$ plays a critical role in the analysis throughout the paper. The following basic result from [35], Chapter 5, Exercise 17 is used repeatedly to handle the singular set.

**Lemma A.3.** Let $u \in H^1(\Omega)$. Then, $\nabla u = 0$ a.e. on the set $\{u = c\}$, where $c \in \mathbb{R}$.

The following lemma is used to handle the vanishing set $\mathcal{Z}_\epsilon$ in Proposition A.6 (see [90]) during the proof of Lemma A.7.

**Lemma A.4.** Let $f \in L^1(\Omega)$ be non-negative, and suppose that for each $\epsilon > 0$ the set $B_\epsilon \subset \Omega$ satisfies $|B_\epsilon| < \epsilon$. Then, $\lim_{\epsilon \to 0} \int_{B_\epsilon} f = 0$. 
A.2. Truncation and regularization in the admissible set

Since the double well function \( \psi \) diverges at \( s = -1/2 \) and \( s = 1 \), it is convenient to truncate \( s \) away from \( s = -1/2, 1 \). The next result, from Lemma 3.1 of [70], indicates that this is only a small perturbation (also see Lem. A.8).

**Lemma A.5 (Truncate \( s \)).** Assume \((g, r, q)\) satisfies Assumption 2.4. Let \((s, u, n) \in \mathcal{A}(g, q)\) and define \( s_\rho := \max\{-\frac{1}{2} + \rho, \min\{s, 1 - \rho\}\}, \) for any \( \rho \geq 0 \), and set \( u_\rho := s_\rho n \). Then, \((s_\rho, u_\rho, n) \in \mathcal{A}(g, q)\) for all \( \rho \leq \rho_0 \) and \( \|(s, u) - (s_\rho, u_\rho)\|_{H^1(\Omega)} \to 0 \), as \( \rho \to 0 \).

The following proposition is a variant of Proposition 3.2 from [70] and is needed to construct a recovery sequence (see Lem. A.17 and Thm. 4.5).

**Proposition A.6 (Regularization in \( \mathcal{A}(g, q) \)).** Suppose the boundary data satisfies Assumptions 2.4 and 2.5. Let \((s, u, n) \in \mathcal{A}(g, q)\), with \(-\frac{1}{2} + \rho \leq s \leq 1 - \rho \) a.e. in \( \Omega \) for any \( \rho \) such that \( 0 \leq \rho \leq \rho_0 \). Then, given \( \delta > 0 \), there exists a triple \((s_\delta, u_\delta, n_\delta) \in \mathcal{A}(g, q)\), such that \( s_\delta \in W^{1,\infty}(\Omega) \), \( u_\delta \in [W^{1,\infty}(\Omega)]^d \), and

\[
\|(s, u) - (s_\delta, u_\delta)\|_{H^1(\Omega)} \leq \delta, \quad \frac{1}{2} + \rho \leq s_\delta(x) \leq 1 - \rho, \quad \forall x \in \Omega.
\]

This implies there exists \( Z_{\epsilon} \subset \Omega \) such that \( |Z_{\epsilon}| < \epsilon \) and \((s_\delta, u_\delta)\) converges uniformly on \( \Omega \setminus Z_{\epsilon} \).

In addition, \( n_\delta \equiv u_\delta / s_\delta \) if \( s_\delta \neq 0 \), and \( n_\delta \) can be taken to be any unit vector if \( s_\delta = 0 \). Then, \( n_\delta \to n \) in \([L^2(\Omega \setminus Z_{\epsilon})]^d\) (recall (2.3)). Moreover, for each fixed \( \epsilon > 0 \):

(i) \( n_\delta \) is Lipschitz on \( \Omega \setminus \{|s| \leq \epsilon\} \) with Lipschitz constant proportional to \( \epsilon^{-1} \);

(ii) \( n_\delta \to n \) in \([H^1(\Omega \setminus Z_{\epsilon})]^d\), as \( \delta \to 0 \), where \( Z_{\epsilon} := \{|s| \leq \epsilon\} \cup Z_{\epsilon} \).

A.3. Perturbing the energy

The following results show that we can perturb the energy (2.42) within the admissible set. This is used to construct a recovery sequence in Theorem 4.5.

**Lemma A.7.** Assume the hypothesis of Proposition A.6. Then,

\[
E_{\text{erk}}(s_\delta, n_\delta) \to E_{\text{erk}}(s, n), \quad E_a(s_\delta, n_\delta) \to E_a(s, n), \quad E_{\text{el}}(s_\delta, n_\delta) \to E_{\text{el}}(s, n),
\]

as \( \delta \to 0 \).

**Proof.** First note that \( s_\delta \to s \), a.e. in \( \Omega \), \( n_\delta \to n \) a.e. in \( \Omega \setminus \{s = 0\} \), and \( \nabla n_\delta \to \nabla n \) in \( L^2(\Omega_\epsilon) \), where \( \Omega_\epsilon := \Omega \setminus Y_\epsilon \) for any fixed \( \epsilon > 0 \), where \( Y_\epsilon := \{|s| \leq \epsilon\} \cup Z_{\epsilon} \) and \( Z_{\epsilon} \) is taken from Proposition A.6. Now consider the k3 term \( I_{k3}^\delta(D) := \int_D s_\delta^2 |n_\delta \times \text{curl} n_\delta|^2 = \int_D s_\delta^2 |[n_\delta \times \text{curl} n_\delta]|^2 \) in \( E_{\text{erk}} \), for any subset \( D \subset \Omega \), and estimate the difference

\[
|I_{k3}^\delta(\Omega_\epsilon) - \int_{\Omega_\epsilon} s_\delta^2 |n_\delta \times \text{curl} n_\delta|^2| \leq \left( \|\text{curl} n\|_{L^2(\Omega_\epsilon)} + \|\text{curl} n_\delta\|_{L^2(\Omega_\epsilon)} \right) \|\nabla n_\delta - \nabla n\|_{L^2(\Omega_\epsilon)} \leq C \|\nabla n\|_{L^2(\Omega_\epsilon)} \|\nabla n_\delta - \nabla n\|_{L^2(\Omega_\epsilon)},
\]

which clearly goes to zero as \( \delta \to 0 \). Moreover, we have that \( s_\delta^2 |[n_\delta \times \text{curl} n_\delta]|^2 \to s^2 |[n \times \text{curl} n]|^2 \) a.e. in \( \Omega_\epsilon \) and \( s_\delta^2 |[n_\delta \times \text{curl} n_\delta]|^2 \leq C |\nabla n_\delta|^2 \leq L^1(\Omega_\epsilon) \) a.e. in \( \Omega_\epsilon \), which implies that \( s_\delta^2 |[n_\delta \times \text{curl} n_\delta]|^2 \to s^2 |[n \times \text{curl} n]|^2 \) in \( L^1(\Omega_\epsilon) \) by the Lebesgue dominated convergence theorem. Therefore, \( \lim_{\delta \to 0} |I_{k3}^\delta(\Omega_\epsilon) - I_{k3}^0(\Omega_\epsilon)| = 0 \), for all \( \epsilon > 0 \).

Next, using \( u = sn \), note that

\[
I_{k3}^\delta(Y_\epsilon) \leq C \int_{Y_\epsilon} s_\delta^2 |\nabla n_\delta|^2 \leq C \int_{Y_\epsilon} |\nabla s_\delta|^2 + |\nabla u_\delta|^2, \quad \text{and so}
\]

\[
\lim_{\delta \to 0} |I_{k3}^\delta(Y_\epsilon) - I_{k3}^0(Y_\epsilon)| \leq 2C \int_{\Omega} \chi_{Y_\epsilon} (|\nabla s|^2 + |\nabla u|^2) =: I_{k3}(Y_\epsilon),
\]

Therefore, \( \lim_{\delta \to 0} E_{\text{erk}}(s_\delta, n_\delta) \to E_{\text{erk}}(s, n) \).
for some independent constant $C > 0$. Taking $\epsilon \to 0$, we have that
\[ I^*_k(\mathbf{Y}_e) \to \int_{\{s = 0\}} (|\nabla s|^2 + |\nabla \mathbf{u}|^2) + \lim_{\epsilon \to 0} \int_{Z_\epsilon} (|\nabla s|^2 + |\nabla \mathbf{u}|^2) = 0, \]
because $\nabla s = 0 = \nabla \mathbf{u}$ a.e. on the set $\{s = 0\}$ (see Lem. A.3), and $\lim_{\epsilon \to 0} |Z_\epsilon| = 0$ (see Lem. A.4). Hence, $\lim_{\epsilon \to 0} \int_{\{s = 0\}} (|\nabla^2_s(\Omega) - \nabla^2_k(\Omega)| = 0$, so then $\int_{\Omega} \nabla^2_s \mathbf{n} \times \text{curl} \mathbf{n}^2 \to \int_{\Omega} \nabla^2 \mathbf{u} \times \text{curl} \mathbf{u}^2$. The other terms in $E_{\text{erk}}$ can be handled similarly.

For the weak anchoring energy $E_a$, we only consider $\int \mathbf{s}^2 (\mathbf{v} \cdot \mathbf{n})^2 \, ds(x)$; the other terms are handled similarly. By Theorem A.2, $s_\delta \mathbf{n}_\delta \equiv \mathbf{u}_\delta \equiv \mathbf{u} \equiv \mathbf{s}_\mathbf{n}$ in $L^2(\Gamma)$. Thus, $\int \mathbf{s}^2 (\mathbf{v} \cdot \mathbf{n})^2 \, ds(x) \to \int \mathbf{s}^2 (\mathbf{v} \cdot \mathbf{n})^2 \, ds(x)$.

For the electric field, we first note that the polarization vector $\mathbf{P}(s_\delta, \mathbf{n}_\delta) \to \mathbf{P}(s, \mathbf{n})$ in $L^2(\Omega)$ by arguments similar to the above. Next, let $\tilde{\varphi}_\delta$ solve (2.38) given the data $(s_\delta, \mathbf{n}_\delta)$, i.e. $\tilde{\varphi}_\delta \equiv T(s_\delta, \mathbf{n}_\delta)$. We must now show that $\tilde{\varphi}_\delta$ converges to a limit $\tilde{\varphi} = T(s, \mathbf{n})$.

Clearly, $\epsilon(s_\delta, \mathbf{n}_\delta)$ converges to $\epsilon(s, \mathbf{n})$ a.e. in $\Omega \setminus \mathcal{S}$. Moreover, $\epsilon(s_\delta, \mathbf{n}_\delta) \to \epsilon \mathbf{l}$ a.e. in $\mathcal{S}$, and $\epsilon(s, \mathbf{n}) = \epsilon \mathbf{l}$ in $\mathcal{S}$. Thus, $\epsilon(s_\delta, \mathbf{n}_\delta)$ converges to $\epsilon(s, \mathbf{n})$ a.e. in $\Omega$. Furthermore, choosing $\eta = \tilde{\varphi}_\delta$ in (2.38) and using (2.33), we find that
\[ \min_{s, \mathbf{n}} \int_{\Omega} \nabla \tilde{\varphi}_\delta \epsilon(s_\delta, \mathbf{n}_\delta) \nabla \eta^T = \int_{\Omega} \nabla \tilde{\varphi} \epsilon(s, \mathbf{n}) \nabla \eta^T. \]

Thus, combining with the convergence of the other terms in (2.38), we see that $\tilde{\varphi}$ solves (2.38) with data $(s, \mathbf{n})$.

The following argument shows that $\tilde{\varphi}_\delta \to \tilde{\varphi}$ in $H^1_0(\Omega)$ (strong convergence). Similar to the previous inequality, we have
\[ \min_{s, \mathbf{n}} \int_{\Omega} \nabla \tilde{\varphi}_\delta - \nabla \tilde{\varphi}^T \leq \int_{\Omega} \nabla (\tilde{\varphi}_\delta - \tilde{\varphi}) \epsilon(s_\delta, \mathbf{n}_\delta) \nabla (\tilde{\varphi}_\delta - \tilde{\varphi})^T \]

where we used the PDE constraint (2.38) with data $(s_\delta, \mathbf{n}_\delta)$. By Lebesgue’s dominated convergence theorem, we have that $\nabla \varphi_\delta \epsilon(s_\delta, \mathbf{n}_\delta) \to \nabla \varphi_\epsilon \epsilon(s, \mathbf{n})$ and $\nabla \tilde{\varphi} \epsilon(s_\delta, \mathbf{n}_\delta) \to \nabla \tilde{\varphi} \epsilon(s, \mathbf{n})$ strongly in $L^2(\Omega)$. Moreover, we know that $\mathbf{P}(s_\delta, \mathbf{n}_\delta) \to \mathbf{P}(s, \mathbf{n})$ strongly in $L^2(\Omega)$. So combining with the weak convergence of $\nabla \tilde{\varphi}_\delta$ implies that the right-hand-side of (A.2) goes to zero. Therefore, we find that $\nabla \tilde{\varphi}_\delta \to \nabla \tilde{\varphi}$ strongly in $L^2(\Omega)$.

This then implies that $E_{\text{el}}(\tilde{\varphi}_\delta, s_\delta, \mathbf{n}_\delta) \to E_{\text{el}}(\tilde{\varphi}; s, \mathbf{n})$ and that the limit $\tilde{\varphi}$ solves the electro-static equation (2.38) with data $(s, \mathbf{n})$. We have thus proven (A.1).

The next result shows the effect of truncating $s$ on the total energy $E$.

**Lemma A.8.** Assume the hypothesis of Lemma A.5. Then, $E(s_\rho, \mathbf{n}) \to E(s, \mathbf{n})$, as $\rho \to 0$, where $s_\rho$ is given in Lemma A.5.
Proof. The result follows by the monotone convergence theorem, and similar techniques as in the proof of Lemma A.7.

We now combine Proposition A.6 and Lemmas A.7, A.8 to obtain the following energy perturbation result.

Lemma A.9. Suppose the boundary data satisfies Assumptions 2.4 and 2.5. Let \((s, u, n) \in \mathcal{A}(g, q)\). Then, given \(k \geq 1\), there exists a triple \((s_k, u_k, n_k) \in \mathcal{A}(g, q)\), such that \(s_k \in W^{1,\infty}(\Omega)\), \(u_k \in [W^{1,\infty}(\Omega)]^d\), \(||(s_k, u_k) - (s, u)||_{H^1(\Omega)} \to 0\) as \(k \to \infty\), and \(-\frac{1}{2} + \frac{1}{k} \leq s_k \leq \frac{1}{2} - \frac{1}{k}\) a.e. in \(\Omega\). Moreover, \(||n_k - n||_{L^2(\Omega; s=0)} \to 0\), \(n_k \in [W^{1,\infty}(\Omega \setminus \{|s_k| \leq \epsilon\})]^d\), and \(||n_k - n||_{H^1(\Omega \setminus \Gamma_e)} \to 0\) as \(k \to \infty\) for any fixed \(\epsilon > 0\), where \(\Gamma_e\) is taken from Proposition A.6, and

\[
E(s_k, n_k) \to E(s, n), \text{ as } k \to \infty.
\] (A.3)

Proof. Let \(k \geq 1\), such that \(0 < 1/k \leq \rho_0\) and define \(\tilde{s}_k := \max\{-1/2 + 1/k, \min\{s, 1 - 1/k\}\}\), and \(\tilde{u}_k := \tilde{s}_k n\). Then the hypothesis of Lemma A.5 is satisfied, so \((\tilde{s}_k, \tilde{u}_k, n) \in \mathcal{A}(g, q)\) for all \(k\) such that \(0 < 1/k \leq \rho_0\), and there exist numbers \(\{a_k\}_{k=1}^\infty\) such that \(||(\tilde{s}_k, \tilde{u}_k) - (s, u)||_{H^1(\Omega)} = a_k\) and \(\lim_{k \to \infty} a_k = 0\). Furthermore, Lemma A.8 implies that there exists numbers \(\{c_{nk}\}_{k=1}^\infty\) such that

\[
|E(\tilde{s}_k, n) - E(s, n)| = c_{nk}, \quad \text{and} \quad \lim_{k \to \infty} c_{nk} = 0.
\] (A.4)

Next, apply Proposition A.6 to \((\tilde{s}_k, \tilde{u}_k, n)\), i.e. given \(\delta < 1/k\), there exists a triple: \((s_{\delta}, u_{\delta}, n_{\delta}) \in \mathcal{A}(g, q)\), such that \((s_{\delta}, u_{\delta}) \in [W^{1,\infty}(\Omega)]^{1+d}\), \(||(s_{\delta}, u_{\delta}) - (\tilde{s}_k, \tilde{u}_k)||_{H^1(\Omega)} \leq \delta\), and \(-1/2 + 1/k \leq s_{\delta} \leq 1 - 1/k\) in \(\Omega\). Moreover, \(||n_{\delta} - n||_{L^2(\Omega; s=0)} \to 0\), \(n_{\delta} \in [W^{1,\infty}(\Omega \setminus \{|s_{\delta}| \leq \epsilon\})]^d\), and \(||n_{\delta} - n||_{H^1(\Omega \setminus \Gamma_e)} \to 0\) as \(\delta \to 0\) for any fixed \(\epsilon > 0\).

Thus, the hypothesis of Lemma A.7 is fulfilled. In addition, to see the convergence of \(E_{dw}\), note that for fixed \(\delta_k\), \(\psi(\tilde{s}_k)\) is bounded on \(\Omega\) and \(\psi(s_{\delta})\) is uniformly bounded for all \(\delta\). Hence, by Lebesgue’s dominated convergence theorem, we see that \(E_{dw}(s_{\delta}) \to E_{dw}(\tilde{s}_k)\) as \(\delta \to 0\). Therefore, \(|E(s_{\delta}, n_{\delta}) - E(s_{\delta}, n)| = c_{\delta_k}\), where \(c_{\delta} \to 0\) as \(\delta \to 0\).

Now, choose \(\delta = \delta_k < 1/k\) sufficiently small so that \(\delta_k < a_k, c_{\delta_k} < c_{nk}\), and define \(s_k := s_{\delta_k}, u_k := u_{\delta_k}, n_k := n_{\delta_k}\). Whence, \(||(s_k, u_k) - (s, u)||_{H^1(\Omega)} \leq ||(s_{\delta_k}, u_{\delta_k}) - (\tilde{s}_k, \tilde{u}_k)||_{H^1(\Omega)} + ||(\tilde{s}_k, \tilde{u}_k) - (s, u)||_{H^1(\Omega)} = \delta_k + a_k\) and \(|E(s_k, n_k) - E(s, n)| \leq |E(s_{\delta_k}, n_{\delta_k}) - E(s, n)| + |E(\tilde{s}_k, n) - E(\tilde{s}_k, n)| = c_{\delta_k} + c_{nk}\). Taking \(k \to \infty\), we obtain the assertion.

A.4. Interpolation estimates

The next basic result is used in Section 3.3.1.

Proposition A.10 (Lagrange interpolant inequality). Let \(p\) be a linear function over a \(d\)-dimensional simplex \(T\), where \(1 \leq d \leq 3\). Then,

\[
\int_T (p(x))^2 \, dx \leq \int_T I_h(p^2)(x) \, dx \leq d!(d + 2) \int_T (p(x))^2 \, dx.
\] (A.5)

The following result is useful throughout the paper.

Proposition A.11 (Elliptic projection). Define the bilinear form \(a(s, z) = (s, z) + (\nabla s, \nabla z)\) and let \(P_h : H^1(\Omega) \to S_h\) be the elliptic projection defined by

\[
a(P_hz, z_h) = a(s, z_h), \quad \text{for all} \quad z_h \in S_h.
\] (A.6)

Then \(||P_hz - s||_{H^1(\Omega)} \to 0\) as \(h \to 0\); similar results hold for the elliptic projections onto \(U_h\) and \(V_h\).

We collect here several interpolation and inverse type inequalities, all of which follow by basic finite element theory.
Lemma A.12. Let $v_h : T \to \mathbb{R}$ be a polynomial on an element $T \in T_h$, of dimension $d$, where $d = 2$ or $3$. Then, the following trace estimate holds
\[
\|v_h\|_{L^2(\partial T)} \leq C_2 h^{d(1/2 - 1/p)} \left( h^{-1/2} \|v_h\|_{L^p(T)} + h^{1/2} \|\nabla v_h\|_{L^p(T)} \right), \quad \text{for all } 2 \leq p \leq \infty,
\]
where $h_T = \text{diam}(T)$ (for any $T \in T_h$), $h = \max_T h_T$, and $C > 0$ only depends on the shape regularity of the mesh $T_h$.

Lemma A.13. Let $(s_h, u_h, n_h)$ in $A_h$ and let $D = \bigcup_{T \in T_h} \tilde{T}_h \subset \Omega$, where $\tilde{T}_h$ is any subset of elements of $T_h$. Then, for $1 \leq p \leq \infty$, the following error estimates hold
\[
\|s_h n_h - u_h\|_{L^p(D)} + h \|\nabla(s_h n_h - u_h)\|_{L^p(D)} \leq Ch\|\nabla s_h\|_{L^p(D)},
\]
\[
\|s_h^{-1} n_h - u_h\|_{L^p(D)} + h \|\nabla (s_h^{-1} n_h - u_h)\|_{L^p(D)} \leq C h \|s_h^{-2}\|_{L^\infty(D)} \left( \|\nabla s_h\|_{L^p(D)} + \|\nabla u_h\|_{L^p(D)} \right),
\]
where $h_T = \text{diam}(T)$ (for any $T \in T_h$), $h = \max_T h_T$, and $C > 0$ only depends on the shape regularity of the mesh $T_h$.

Lemma A.14. Assume the hypothesis of Lemma A.13 and let $w_h \in U_h$. Then,
\[
\|s_h^2 - I_h\{s_h^2\}\|_{L^p(D)} \leq Ch \|\nabla s_h\|_{L^p(D)}, \quad \text{or} \quad Ch^2 \|\nabla s_h\|^2_{L^2(D)},
\]
\[
\|w_h \otimes w_h - I_h\{w_h \otimes w_h\}\|_{L^p(D)} \leq Ch \|w_h \otimes \nabla w_h\|_{L^p(D)}, \quad \text{or} \quad Ch^2 \|\nabla w_h\|^2_{L^2(D)},
\]
\[
\|s_h^2 (n_h \otimes n_h) - I_h\{s_h^2 (n_h \otimes n_h)\}\|_{L^p(D)} \leq Ch \|s_h\|_{L^\infty(D)} K_1, \quad \text{or} \quad Ch^2 K_2,
\]
\[
\|s_h^2 (|n_h| \times |n_h|) - I_h\{s_h^2 (|n_h| \times |n_h|)\}\|_{L^p(D)} \leq Ch \|s_h\|_{L^\infty(D)} K_1, \quad \text{or} \quad Ch^2 K_2,
\]
\[
K_1 := (\|\nabla s_h\|_{L^p(D)} + \|\nabla u_h\|_{L^p(D)}), \quad K_2 := (\|\nabla s_h\|^2_{L^2(D)} + \|\nabla u_h\|^2_{L^2(D)}),
\]
where $C > 0$ only depends on the shape regularity of the mesh $T_h$.

Lemma A.15. Let $(s_h, u_h, n_h)$ in $A_h$ and let $\Sigma = \bigcup_{F \in F_h^T} F \subset \Gamma$, where $\tilde{F}_h$ is any subset of $F_h$, which is the set of all face elements contained in $T_h$. Then, for $d = 2$ or $3$, the following estimate holds
\[
\|s_h n_h - u_h\|_{L^2(\Sigma)} \leq Ch^{1/2} \|\nabla s_h\|_{L^2(D)},
\]
where $D = \bigcup_{F \in F_h^T} T \subset \Omega$, with $\tilde{T}_h := \{T \in T_h : T \cap \Sigma \neq \emptyset\}$, and $C > 0$ only depends on the shape regularity of the mesh $T_h$.

A.5. \Gamma-convergence intermediate results

A.5.1. Characterizing limits

The following result is taken from Lemma 3.6 of [70]. Note that we only get convergence (in general) for $s_h$ and $u_h$; the convergence of $n_h$ is somewhat limited.

Lemma A.16 (Characterizing limits). Let $(s_h, u_h)$ in $A_h$ converge weakly to $(s, u)$ in $[H^1(\Omega)]^{1+d}$. Then, $(s_h, u_h)$ converges to $(s, u)$ strongly in $[L^2(\Omega)]^{1+d}$, a.e. in $\Omega$, and the limit $(s, u)$ satisfies $|s| = |u|$ a.e. in $\Omega$ (i.e. $(s, u) \in A$). In addition, there exists $Z'_e \subset \Omega$ such that $|Z'_e| < \epsilon$ and $(s_h, u_h)$ converges uniformly to $(s, u)$ on $\Omega \setminus Z'_e$.

Furthermore, the associated sequence $n_h$ in $N_h$, defined by $u_h = I_h\{s_h n_h\}$, satisfies the following properties for each fixed $\epsilon > 0$.

(i) There exists a director field $n : \Omega \to S^{d-1}$, with $n \in [L^2(\Omega)]^{d} \cap [L^\infty(\Omega)]^{d}$, $|n| = 1$ a.e., such that $n_h$ converges to $n$ in $[L^2(\Omega \setminus S)]^{d}$ and a.e. in $\Omega \setminus S$ and $u = sn$ a.e. in $\Omega$. In addition, $n_h \rightarrow n$ uniformly on $\Omega \setminus (S_e \cup Z'_e)$, where $S_e = \{|s(x)| \leq \epsilon\}$. 

\[
\frac{\|f_h - f\|_{L^2(\Omega)}}{\|f_h\|_{L^2(\Omega)}} \leq C (h^{-1/2} \|f_h\|_{L^p(\Omega)} + h^{1/2} \|\nabla f_h\|_{L^p(\Omega)}),
\]
where $h_T = \text{diam}(T)$ (for any $T \in T_h$), $h = \max_T h_T$, and $C > 0$ only depends on the shape regularity of the mesh $T_h$. 

Lemma A.15. Let $(s_h, u_h, n_h)$ in $A_h$ and let $\Sigma = \bigcup_{F \in F_h^T} F \subset \Gamma$, where $\tilde{F}_h$ is any subset of $F_h$, which is the set of all face elements contained in $T_h$. Then, for $d = 2$ or $3$, the following estimate holds
(ii) \( n_h \) converges weakly to \( n \) in \( [H^1(\Omega \setminus \mathcal{S}_c)]^d \), where \( \mathcal{S}_c := \mathcal{S} \cup \mathcal{Z}_c \) (c.f. Prop. A.6).

**Lemma A.17.** Suppose Assumptions 2.4 and 2.5 hold. Let \( (s, u, n) \in \mathcal{A}(g, q) \) such that \( (s, u) \in [W^{1,\infty}(\Omega)]^{d+1} \).

Then there exists a sequence \( (s_h, u_h, n_h) \in \mathcal{A}_h(g_h, q_h) \), such that \( (s_h, u_h) \) converges to \( (s, u) \) in \( [W^{1,\infty}(\Omega)]^{d+1} \), as well as \( n_h \in N_h \) converging to \( n \) in \( L^2(\Omega \setminus \mathcal{S}) \), and \( n_h \) converging to \( n \) in \( [W^{1,\infty}(\Omega \setminus \mathcal{S}_c)]^{d+1} \), for every fixed \( \epsilon > 0 \).

*Proof.* We introduce the Lagrange interpolants \( s_h := I_h(s) \), \( u_h := I_h(u) \); moreover, define

\[
n_h(x_i) = u_h(x_i)/s_h(x_i), \quad \text{if } s_h(x_i) \neq 0, \quad \text{otherwise } n_h(x_i) = \text{any unit vector}.
\]

for each \( x_i \in \mathcal{N}_h \). Thus, \( (s_h, n_h) \in \mathcal{A}_h(g_h, q_h) \).

Let \( s_\delta = s * \phi_\delta \), where \( \phi_\delta \) is a mollifier; hence, \( s_\delta \in C^\infty \) and \( \|s_\delta - s\|_{H^1(\Omega)} \to 0 \) as \( \delta \to 0 \). Next, use interpolation theory, and the triangle inequality:

\[
\|I_h(s) - s\|_{H^1(\Omega)} \leq \|I_h(s - s_\delta)\|_{H^1(\Omega)} + \|I_h(s_\delta) - s_\delta\|_{H^1(\Omega)} + \|s_\delta - s\|_{H^1(\Omega)} \\
\leq C_1\|s - s_\delta\|_{H^1(\Omega)} + C_2h\|D^2s_\delta\|_{L^2(\Omega)},
\]

where we used the stability of the interpolant. Taking the limit as \( h \to 0 \), we have \( \lim_{h \to 0} \|I_h(s) - s\|_{H^1(\Omega)} \leq C_1\|s - s_\delta\|_{H^1(\Omega)} \), for all \( \delta > 0 \). So, taking \( \delta \to 0 \), we see that \( \|s_h - s\|_{H^1(\Omega)} \to 0 \) as \( h \to 0 \). Similarly, \( \|u_h - u\|_{H^1(\Omega)} \to 0 \).

Next, we check \( n_h \). Let \( \Omega_\epsilon := \Omega \setminus \mathcal{S}_c \), and note that \( n \in [W^{1,\infty}(\Omega_\epsilon)]^d \) for every fixed \( \epsilon > 0 \). Since \( n_h = I_h(s^{-1}u) = I_h(n) \) on \( \Omega_\epsilon \), again by interpolation theory, we have that \( n_h \to n \) in \( H^1(\Omega_\epsilon) \). To prove the convergence in \( L^2(\Omega \setminus \mathcal{S}) \), one can follow the argument in the proof of Proposition A.6. \( \square \)

**A.5.2. Estimates for mass-lumping**

**Lemma A.18** (Remove \( I_h \) for lim-sup.). Recall (3.7) and (3.9). Let \( (s_h, u_h, n_h) \in \mathcal{A}_h \) such that \( (s_h, u_h) \) converges strongly to \( (s, u) \) in \( [W^{1,\infty}(\Omega)]^{1+d} \). Moreover, assume \( n_h \to n \) in \( [W^{1,\infty}(\Omega_\epsilon)]^d \) for every fixed \( \epsilon > 0 \), where \( \Omega_\epsilon := \Omega \setminus \mathcal{S}_c \). Then,

\[
\lim_{h \to 0} \left( \frac{\sqrt{\nabla(s_h, \nabla s_h, n_h, \nabla n_h)} - \sqrt{\nabla(s, \nabla s, n, \nabla n)}}{h} \right) \to 0. \tag{A.12}
\]

*Proof.* Note that the hypothesis implies \( \|(s_h, u_h)\|_{H^1(\Omega)} \leq A_0 \), for some constant \( A_0 \), for all \( h \). We demonstrate the result for one of the terms in (3.9); the other terms follow by similar arguments. We first show that \( |w_{k_3}^h - w_{k_3}| \to 0 \) as \( h \to 0 \). After recalling (3.11), consider the difference:

\[
G_{k_3}^h(\Omega) := (I_h(s_h^2 n_h \otimes n_h), \nabla n_h^T \nabla n_h) - (s_h^2 n_h \otimes n_h, \nabla n_h^T \nabla n_h). \tag{A.13}
\]

Throughout, we let \( C > 0 \) denote a generic constant. Now fix \( \epsilon > 0 \) and note that, for \( h \) sufficiently small, we have

\[
|G_{k_3}^h(\Omega \setminus \mathcal{S}_c)| = \int_{\Omega \setminus \mathcal{S}_c} \left| I_h(s_h^2 n_h \otimes n_h) - (s_h^2 n_h \otimes n_h) \right| : (\nabla n_h^T \nabla n_h) \\
\leq \|I_h(s_h^2 n_h \otimes n_h) - (s_h^2 n_h \otimes n_h)\|_{L^2(\Omega)} \|\nabla n_h\|_{L^2(\Omega)}, \\
\leq C \|\nabla n_h\|_{L^2(\Omega)} \|\nabla n_h\|_{L^2(\Omega)} \leq C A_0 \|\nabla n\|_{L^2(\Omega)},
\]

where we used (A.10).

Next, we examine the difference over \( \mathcal{S}_c \). For all \( h > 0 \), let \( T_h := \{ T \in \mathcal{T}_h : T \cap \mathcal{S}_c \neq \emptyset \} \subset \mathcal{T}_h \), and note that for \( h \) sufficiently small, we have \( \mathcal{S}_c \subset \mathcal{D}_c := \cup_{T \in T_h} T \subset \mathcal{S}_c \), because \( \mathcal{S}_c \) and \( \mathcal{S}_c \) are disjoint compact sets, so they are a positive distance apart.
We obtain for sufficiently small $h$:
\[
|G^h_{k_3}(S_h)| = \int_{S_h} \left| I_h \left\{ s_h^2 n_h \otimes n_h \right\} - (s_h^2 n_h \otimes n_h) : [(\nabla n_h^2) \nabla n_h] \right| \\
\leq C \| I_h \left\{ s_h^2 n_h \otimes n_h \right\} - (s_h^2 n_h \otimes n_h) \|_{L^2(\Omega)} \tilde{C}^2 h^{-2} \\
\leq CC^2 \left( \| \nabla s_h \|_{L^2(S_{2r})} + \| \nabla u_h \|_{L^2(S_{2r})} \right) \leq C \left( \| \nabla s \|_{L^2(S_{2r})} + \| \nabla u \|_{L^2(S_{2r})} \right),
\]
where we used an inverse inequality $\| \nabla n_h \|_{L^\infty(\Omega)} \leq \tilde{C} \| n_h \|_{L^\infty(\Omega)}$, as well as (A.10). Thus,
\[
\lim_{h \to 0} |G^h_{k_3}(\Omega)| \leq C \left( \| \nabla s \|_{L^2(S_{2r})} + \| \nabla u \|_{L^2(S_{2r})} \right), \quad \forall \epsilon > 0.
\]
Therefore, taking $\epsilon \to 0$ and using the monotone convergence theorem, we get
\[
\lim_{h \to 0} \left| G^h_{k_3}(\Omega) \right| = 0,
\]
because $\nabla s = 0$ and $\nabla u = 0$ a.e. in $\{s = 0\}$ (see Lem. A.3).

Therefore, $w^h_{k_3} (s_h, s_h; n_h, n_h; \nabla n_h, \nabla n_h) \to w_{k_3} (s_h, s_h; n_h, n_h; \nabla n_h, \nabla n_h) \to 0$, as $h \to 0$. The convergence of the remaining terms follows by similar arguments.

\[\square\]

Lemma A.19 (Remove $I_h$ for lim-inf.). Recall (3.7) and (3.9). Let $(s_h, u_h, n_h) \in A_h$ such that $(s_h, u_h)$ converges weakly to $(s, u)$ in $[H^1(\Omega)]^d$. Moreover, let $(\hat{s}_h, \hat{n}_h) \in S_h \times U_h$ such that $(\hat{s}_h, \hat{n}_h) \to (\hat{s}, \hat{n})$ in $[W^{1,\infty}(\Omega)]^d$, and $\| \hat{n}_h \|_{L^\infty(\Omega)} \leq C$ for all $h$. Let $F'_e := \Omega \setminus \mathcal{T}_e$ where $\mathcal{T}_e$ is taken from Lemma A.16. Then, for every fixed $\epsilon > 0$,
\[
\lim_{h \to 0} \left| \hat{W}(s_h, \nabla \hat{s}_h, n_h, \nabla \hat{n}_h, 1)_{F'_e} - \hat{W}(s_h, \nabla \hat{s}_h, n_h, \nabla \hat{n}_h, 1)_{F'_e} \right| \to 0. \tag{A.15}
\]

Proof. Without loss of generality, the hypothesis implies that $\| (s_h, u_h) \|_{H^1(\Omega)} \leq A_0$, for some constant $A_0$, for all $h$. We demonstrate the result for one of the terms in (3.9); the other terms follow by similar arguments.

We first show that $|w^h_{k_3} - w_{k_3}| \to 0$ as $h \to 0$. Consider the difference $G^h_{k_3}(F'_e)$ (as in (A.13)) where we have for each $\epsilon > 0$, and for $h$ sufficiently small,
\[
|G^h_{k_3}(F'_e)| = \left| \int_{F'_e} \left[ I_h \left\{ s_h^2 n_h \otimes n_h \right\} - (s_h^2 n_h \otimes n_h) : [(\nabla n_h^2) \nabla n_h] \right| \\
\leq \| I_h \left\{ s_h^2 n_h \otimes n_h \right\} - (s_h^2 n_h \otimes n_h) \|_{L^2(F'_e)} \| \nabla n_h \|_{L^2(F'_e)} \\
\leq C \left( \| \nabla s_h \|_{L^2(F'_e)} + \| \nabla u_h \|_{L^2(F'_e)} \right) \| \nabla n_h \|_{L^\infty(F'_e)} \leq C h \left( \| \nabla s_h \|_{L^2(\Omega)} + \| \nabla u_h \|_{L^2(\Omega)} \right) \| \nabla n_h \|_{L^\infty(F'_e)},
\]
where we used (A.10). Therefore, $w^h_{k_3} (s_h, s_h; n_h, n_h; \nabla n_h, \nabla n_h)_{F'_e} \to w_{k_3} (s_h, s_h; n_h, n_h; \nabla n_h, \nabla n_h)_{F'_e} \to 0$, as $h \to 0$. The convergence of the other terms follows similarly.

\[\square\]

Lemma A.20. Assume the hypothesis of Lemma A.19. Then,
\[
\lim_{h \to 0} \left| D_M \hat{W}(s_h, \nabla \hat{s}_h, n_h, \nabla \hat{n}_h)_{F'_e} - \hat{W}(s_h, \nabla \hat{s}_h, n_h, n_h)_{F'_e} \right| \to 0, \tag{A.17}
\]
\[
\lim_{h \to 0} \left| D_g \hat{W}(s_h, \nabla \hat{s}_h, n_h, \nabla \hat{n}_h)_{F'_e} - \hat{W}(s_h, \nabla \hat{s}_h, n_h, \nabla n_h)_{F'_e} \right| \to 0, \tag{A.18}
\]
where $(t_h, y_h) \in S_h \times U_h$ and $\| (t_h, y_h) \|_{H^1(\Omega)} \leq A_2$ for all $h$.
Proof. We follow similar arguments as in the proof of Lemma A.19. First note that

\[
D_M \widetilde{\mathcal{W}}(s_h, \nabla s_h, n_h, \nabla \hat{n}_h) = 2s_h^2 [k_1 \text{tr}(\nabla \hat{n}_h) I + k_2 (|n_h|_\chi : \nabla \hat{n}_h) n_h + k_3 (\nabla \hat{n}_h n_h) \otimes n_h + (k_2 + k_4)(\nabla_{n_h}^T \nabla \hat{n}_h) I + \theta n_h \otimes (\nabla_{n_h}^T \nabla \hat{n}_h)]
+ b_3 s_h (\nabla \hat{n}_h \cdot n_h) I + b_4 s_h \nabla \hat{s}_h \otimes n_h,
\]

(A.19)

\[
D_g \widetilde{\mathcal{W}}(s_h, \nabla s_h, n_h, \nabla \hat{n}_h) = 2b_1 \nabla \hat{s}_h + 2b_2 (\nabla \hat{s}_h \cdot n_h) n_h + b_3 s_h n_h \text{tr}(\nabla \hat{n}_h) + b_4 s_h (\nabla \hat{n}_h) n_h.
\]

We demonstrate (A.17) for the \(k_3\) term in (A.19); the other terms follow by similar arguments. Define the difference:

\[
G_{k_3}^h(\Omega) := (s_h^2 (\nabla \hat{n}_h n_h) \otimes n_h, \nabla Y_h)^h - (s_h^2 (\nabla \hat{n}_h n_h) \otimes n_h, \nabla Y_h),
\]

and note that (similar to (A.13)) \((s_h^2 (\nabla \hat{n}_h n_h) \otimes n_h, \nabla Y_h)^h = \int_\Omega I_H \{ s_h^2 n_h \otimes n_h \} : (\nabla \hat{n}_h^2 \nabla Y_h). \) Then,

\[
|G_{k_3}^h(\Omega)| = \left| \int_\Omega \left[ I_H \{ s_h^2 n_h \otimes n_h \} - (s_h^2 n_h \otimes n_h) \right] \cdot (\nabla \hat{n}_h^2 \nabla Y_h) \right|
\leq \| I_H \{ s_h^2 n_h \otimes n_h \} - (s_h^2 n_h \otimes n_h) \|_{L^2(\Omega)} \| \nabla \hat{n}_h \|_{L^\infty(\Omega)} \| \nabla Y_h \|_{L^2(\Omega)}
\leq CA_2 h \left( \| \nabla s_h \|_{L^2(\Omega)} + \| \nabla u_h \|_{L^2(\Omega)} \right) \| \nabla \hat{n}_h \|_{L^\infty(\Omega)}
\leq CA_2 h \| \nabla \hat{n}_h \|_{L^\infty(\Omega)}
\]

(A.20)

where we used (A.10). Clearly, \(\lim h \to 0 |G_{k_3}^h(\Omega)| = 0\). The other terms follow by similar arguments. \(\square\)

A.5.3. Weak lower-semicontinuity

Lemma A.21 (Weak L.S.C. for \(\hat{E}_{\text{erk}}\)). Assume the hypothesis of Lemma A.16. If (2.19) holds, then \(\hat{E}_{\text{erk}}^h\) is weakly lower semi-continuous, i.e.

\[
\liminf_{h \to 0} \hat{E}_{\text{erk}}^h(s_h, n_h) \geq \hat{E}_{\text{erk}}(s, n) = E_{\text{erk}}(s, n),
\]

(A.21)

for any sequence \((s_h, u_h, n_h) \in A_h, \) such that \((s, u, n) \in A, \) and \((s_h, u_h) \to (s, u) \) in \([H^1(\Omega)]^{1+d}\).

Proof. Step 1: Egorov. Set \(L := \liminf_{h \to 0} E_{\text{erk}}(s_h, n_h); \) we must show that \(E_{\text{erk}}(s, n) \leq L. \) Without loss of generality, we can assume \(L < \infty. \) By Lemma A.16, we have that \(s_h \to s, \ n_h \to n\) a.e. in \(\Omega, \) and \(n_h \to n\) a.e. in \(\Omega \setminus S. \) Fix \(\varepsilon > 0. \) Lemma A.16 gives that \(n_h\) converges weakly to \(n\) in \([H^1(\Omega \setminus T')^d], \) where \(T' = S \cup Z', \) \((s_h, u_h) \to (s, u)\) uniformly on \(\Omega \setminus T',\) and \(n_h \to n\) uniformly on \(T'_\varepsilon = \Omega \setminus T'\).

Step 2: convexity. Let \(s_\delta = s + \delta \phi_\delta; \) where \(\phi_\delta\) is a mollifier; hence, \(s_\delta \in C^\infty(\Omega)\) and \(\|s_\delta - s\|_{H^1(\Omega)} \to 0\) as \(\delta \to 0.\)

Next, define \(n_\delta = n + \delta \phi_\delta; \) thus, \(n_\delta \in C^\infty(\Omega)\) and \(\|n_\delta - n\|_{L^2(\Omega)} \to 0\) as \(\delta \to 0,\) and \(\|n_\delta - n\|_{H^1(\Omega \setminus S')} \to 0\) as \(\delta \to 0.\)

Next, define \(\hat{s}_h := I_h s_\delta \in S_h, \) \(\hat{n}_h := I_h n_\delta \in U_h,\) and note that \(\hat{s}_h \to s_\delta\) in \(W^{1,\infty}(\Omega), \) \(\|\hat{n}_h\|_{L^\infty(\Omega)} \leq 1,\) \(n_h \to n_\delta\) in \([W^{1,\infty}(\Omega)]^d.\)

Now combine the convexity result (2.28) with the interpolation operator \(I_h\) to obtain

\[
I_h \tilde{\mathcal{W}}(s_h, \nabla s_h, n_h, \nabla \hat{n}_h) \geq I_h \tilde{\mathcal{W}}(s_h, \nabla \hat{s}_h, n_h, \nabla \hat{n}_h) + I_h \{ \Psi_1^h \} + I_h \{ \Psi_2^h \}
\]

(A.22)

\[
\Psi_1^h := D_M \tilde{\mathcal{W}}(s_h, \nabla \hat{s}_h, n_h, \nabla \hat{n}_h) : (\nabla n_h - \nabla \hat{n}_h),
\]

\[
\Psi_2^h := D_g \tilde{\mathcal{W}}(s_h, \nabla \hat{s}_h, n_h, \nabla \hat{n}_h) \cdot (\nabla s_h - \nabla \hat{s}_h).
\]

(A.23)

on every \(T \in T_h. \) Whence, \(\hat{E}_{\text{erk}}^h(s_h, n_h) \geq \frac{1}{2} \int_{T_h} \left[ I_h \tilde{\mathcal{W}}(s_h, \nabla \hat{s}_h, n_h, \nabla \hat{n}_h) + I_h \{ \Psi_1^h \} + I_h \{ \Psi_2^h \} \right].\) In lieu of (A.15), (A.17), (A.18), we may drop the interpolation operator \(I_h.\)
Step 3: \( h \to 0 \) for lower bound. We demonstrate

\[
\lim_{h \to 0} \int_{F_h} \tilde{W}(s_h, \nabla s_h, n_h, \nabla n_h) = \int_{F_\varepsilon} \tilde{W}(s, \nabla s_\delta, n, \nabla n_\delta). \tag{A.24}
\]

We start with the \( k_3 \) term in (A.24), which has the form \( I_{k_3}^h := (s_h(\nabla n_h)n_h, s_h(\nabla n_h)n_h)_{F_\varepsilon} \). By uniform convergence in Step (1), and the strong convergence of \( \nabla n_h \) in \( L^2(\Omega) \), we have that \( s_h^2(\nabla n_h)n_h \to s^2(\nabla n_\delta)n \) in \( L^2(\Omega) \). Thus, we obtain \( \lim_{h \to 0} I_{k_3}^h = (s(\nabla n_\delta)n, s(\nabla n_\delta)n)_{F_\varepsilon} =: I_{k_3}^\delta \).

Next, we consider the \( b_2 \) term in (A.24), which has the form \( I_{b_2}^h := (\nabla s_h \cdot n_h, \nabla s_h \cdot n_h)_{F_\varepsilon} \). Again, by uniform convergence in \( F_\varepsilon \) and the strong convergence of \( \nabla s_h \) in \( L^2(\Omega) \), we have that \( (\nabla s_h \cdot n_h)^2 \to (\nabla s_\delta \cdot n)^2 \) in \( L^2(\Omega) \). Whence, \( \lim_{h \to 0} I_{b_2}^h = (\nabla s_\delta \cdot n, \nabla s_\delta \cdot n)_{F_\varepsilon} =: I_{b_2}^\delta \). The other terms are similarly dealt with. So, we have proved (A.24).

Step 4: \( h \to 0 \) for residual terms. Next, we show that

\[
\lim_{h \to 0} \int_{F_\varepsilon} \Psi_{h1}^h = \left( D_M \tilde{W}(s, \nabla s_\delta, n, \nabla n_\delta), \nabla n - \nabla n_\delta \right)_{F_\varepsilon},
\quad \lim_{h \to 0} \int_{F_\varepsilon} \Psi_{h2}^h = \left( D_M \tilde{W}(\nabla s_\delta, \nabla n_\delta, \nabla s - \nabla s_\delta) \right)_{F_\varepsilon}. \tag{A.25}
\]

We start with the \( k_3 \) term in \( \int_{F_\varepsilon} \Psi_{h1}^h \), which, after recalling (A.19), has the form

\[
R_{k_3}^h := (s_h^2(\nabla n_h)n_h) \otimes n_h, \nabla n_h - \nabla n_h)_{F_\varepsilon}.
\]

Again by uniform convergence and strong convergence of \( \nabla n_h \) in \( L^2(\Omega) \), we have that \( s_h^2(\nabla n_h)n_h \to s^2(\nabla n_\delta)n \otimes n \) in \( L^2(\Omega) \). Since \( \nabla n_h \to \nabla n_\delta \) in \( L^2(\Omega) \), and \( \nabla n_h \) converges weakly to \( \nabla n \) on \( F_\varepsilon \), we have that \( \nabla n_h - \nabla n_h \) converges weakly to \( \nabla n - \nabla n_\delta \) in \( L^2(\Omega) \). Combining strong and weak convergence, we obtain

\[
\lim_{h \to 0} R_{k_3}^h = (s^2(\nabla n_\delta)n \otimes n, \nabla n - \nabla n_\delta)_{F_\varepsilon} =: R_{k_3}^\delta.
\]

The remaining terms in (A.25) follow similarly.

Step 5: take the limit \( \delta \to 0 \). The strong \( L^2(F_\varepsilon) \) convergence of \( \nabla n_\delta \) to \( \nabla n \), together with the boundedness of \( s \) and \( n \), give \( I_{k_3}^\delta \to (s(\nabla n)n, s(\nabla n)n)_{F_\varepsilon} \), as \( \delta \to 0 \). Moreover, we have \( I_{b_2}^\delta \to (\nabla s \cdot n, \nabla s \cdot n)_{F_\varepsilon} \), as \( \delta \to 0 \) (similarly for the other terms). Furthermore, the residual terms vanish, e.g. \( R_{k_3}^\delta \to 0 \) as \( \delta \to 0 \).

Step 6: conclude. We have shown that

\[
\liminf_{h \to 0} E_{\text{crk}}(s_h, n_h) \geq \frac{1}{2} \int_{F_\varepsilon} \tilde{W}(s, \nabla s, n, \nabla n) = \frac{1}{2} \int_{F_\varepsilon} W(s, \nabla s, n, \nabla n),
\]

where the equality follows from \( (\nabla n)\n = 0 \) because \( |n| = 1 \) a.e. in \( F_\varepsilon \) and \( (s, u, n) \in A \). Taking \( \epsilon \to 0 \) yields \( F_\varepsilon \to \Omega \setminus (Z' \cup S) \), where \( Z' \subset \Omega \) is a set of measure zero. By Lemmas A.3 and A.4, \( \|\nabla s\|_{L^2(Z' \cup S)} = 0 \), \( \|\nabla u\|_{L^2(Z' \cup S)} = 0 \), \( \|\nabla s\|_{L^2(Z' \cup S)} = 0 \), \( \|\nabla u\|_{L^2(Z' \cup S)} = 0 \). Therefore,

\[
\lim_{\epsilon \to 0} \int_{F_\epsilon} W(s, \nabla s, n, \nabla n) = \int_{\Omega} W(s, \nabla s, n, \nabla n),
\]

i.e. we proved (A.21). \( \square \)

Lemma A.22 (Continuity of \( E^h_n \)). Assume the hypothesis of Lemma A.16. Then \( E^h_n \) is continuous, i.e.

\[
\lim_{h \to 0} E^h_n(s_h, n_h) = E_n(s, n), \tag{A.26}
\]

for any sequence \( (s_h, u_h, n_h) \in A_h \), such that \( (s, u, n) \in A \), and \( (s_h, u_h) \to (s, u) \) in \( [H^1(\Omega)]^{1+d} \).
Proof. The result essentially follows by strong convergence in $L^2(\Gamma)$. \hfill \Box

Lemma A.23 (Weak L.S.C. for $E_{el}$). Assume the hypothesis of Lemma A.16. Then $E^{h}_{el}$ is weakly lower semi-continuous, i.e.

$$\liminf_{h \to 0} E^{h}_{el}(s_h, u_h, n_h) \geq E_{el}(s, u)$$

(A.27)

for any sequence $(s_h, u_h, n_h) \in A_h$, such that $(s, u, n) \in A$, and $(s_h, u_h) \to (s, u)$ in $[H^1(\Omega)]^{1+d}$. 

Proof. Since $||\mathbf{(s_h, \mathbf{u}_h)}||_{H^1(\Omega)}$ is uniformly bounded, and recalling (2.36), we have that $||\mathbf{P}(s_h, n_h)||_{L^2(\Omega)}$ is uniformly bounded, thus there exists a sub-sequence (not relabeled) such that $\mathbf{P}(s_h, n_h) \to \mathbf{P}(s, n)$ in $L^2(\Omega)$. Next, let $\bar{\varphi}_h$ solve (3.16) given the data $(s_h, n_h)$, i.e. $\bar{\varphi}_h \equiv T_h(s_h, n_h)$.

Next, we show that $\bar{\varphi}_h$ converges to $\bar{\varphi} = T(s, n)$. Clearly, $\epsilon(s_h, n_h)$ converges to $\epsilon(s, n)$ a.e. in $\Omega \setminus S$. Moreover, $\epsilon(s_h, n_h) \to \epsilon \mathbf{I}$ a.e. in $S$, and $\epsilon(s, n) = \epsilon \mathbf{I}$ in $S$. Thus, $\epsilon(s_h, n_h)$ converges to $\epsilon(s, n)$ a.e. in $\Omega$. Furthermore, choosing $\eta_h = \bar{\varphi}_h$ in (3.16) and using (3.14), we find that

$$\epsilon_{\min} \|\nabla \bar{\varphi}_h\|^2_{L^2(\Omega)} \leq \epsilon_{\max} \left(\frac{c_1}{2} \|\nabla \varphi_{0, h}\|^2_{L^2(\Omega)} + \frac{1}{2c_2} \|\nabla \bar{\varphi}_h\|^2_{L^2(\Omega)} + \frac{c_2}{2} \|\mathbf{P}(s_h, n_h)\|^2_{L^2(\Omega)} + \frac{1}{2c_2} \|\nabla \bar{\varphi}_h\|^2_{L^2(\Omega)}\right),$$

where $c_1, c_2 > 0$ are to be chosen. Since $||\mathbf{P}(s_h, n_h)||_{L^2(\Omega)}$ is uniformly bounded, choosing $c_1, c_2$ sufficiently large, we find that $\|\nabla \bar{\varphi}_h\|_{L^2(\Omega)} \leq C < \infty$, for all $h > 0$ for some fixed constant $C > 0$. Thus, $\bar{\varphi}_h \to \bar{\varphi}$ in $H^1_0(\Omega)$.

Furthermore, $\epsilon(s_h, n_h)\nabla \eta^T \to \epsilon(s, n)\nabla \eta^T$ (in $L^2(\Omega)$) by Lebesgue’s dominated convergence theorem, ergo

$$\|\epsilon(s_h, n_h)\nabla \eta^T_h - \epsilon(s, n)\nabla \eta^T\|_{L^2(\Omega)} \leq \|\epsilon(s_h, n_h)\nabla \eta^T_h - \epsilon(s, n)\nabla \eta^T\|_{L^2(\Omega)} + \|\epsilon(s_h, n_h)\nabla \eta^T - \epsilon(s, n)\nabla \eta^T\|_{L^2(\Omega)} \to 0, \text{ as } h \to 0,$$

because $\nabla \eta_h \to \nabla \eta$ in $L^2(\Omega)$, where we chose $\eta \in C_0^\infty(\Omega)$ and take $\eta_h = I_h \eta$. So, combining with the weak convergence of $\nabla \bar{\varphi}_h$, we see that $\int_{\Omega} \nabla \bar{\varphi}_h \epsilon(s_h, n_h)\nabla \eta^T_h \to \int_{\Omega} \nabla \bar{\varphi} \epsilon(s, n)\nabla \eta^T$. Thus, combining with the convergence of the other terms in (3.16), we see that $\bar{\varphi}$ solves (3.16) with data $(s, n)$.

Next, we recall (2.41) and (3.17) which make the convexity of $E_{el}$ and $E^{h}_{el}$ more apparent:

$$E_{el}(s, u) = \frac{1}{2} \int_{\Omega} \nabla \bar{\varphi} \epsilon(s, n)\nabla \bar{\varphi}^T - \frac{1}{2} \int_{\Omega} \nabla \varphi_0 \epsilon(s, n)\nabla \varphi_0^T + \int_{\Omega} \mathbf{P}(s, n) \cdot \nabla \varphi_0,$$

$$E^{h}_{el}(s_h, u_h, n_h) = \frac{1}{2} \int_{\Omega} \nabla \bar{\varphi}_h \epsilon(s_h, n_h)\nabla \bar{\varphi}_h^T - \frac{1}{2} \int_{\Omega} \nabla \varphi_{0, h} \epsilon(s_h, n_h)\nabla \varphi_{0, h}^T + \int_{\Omega} \mathbf{P}(s_h, n_h) \cdot \nabla \varphi_{0, h}.$$

The convergence of the last two terms is clear, since $\varphi_{0, h}$ is the elliptic projection (see Prop. A.11), so it converges strongly in $H^1(\Omega)$.

For the first term, given $\epsilon > 0$, by Egorov’s Theorem there exists $A_\epsilon \subset \Omega$ such that $|\Omega \setminus A_\epsilon| < \epsilon$ and $\epsilon(s_h, n_h) \to \epsilon(s, n)$ uniformly on $A_\epsilon$. Ergo,

$$\int_{\Omega} \nabla \bar{\varphi}_h \epsilon(s_h, n_h)\nabla \bar{\varphi}_h^T \geq \int_{A_\epsilon} \nabla \bar{\varphi}_h \epsilon(s_h, n_h)\nabla \bar{\varphi}_h^T + \int_{A_\epsilon} \nabla \bar{\varphi}_h \epsilon(s, n)\nabla \bar{\varphi}_h^T,$$

where the first term vanishes by uniform convergence of $\epsilon(s_h, n_h)$. For the last term, we use weak lower semi-continuity to obtain

$$\liminf_{h \to 0} \int_{\Omega} \nabla \bar{\varphi}_h \epsilon(s_h, n_h)\nabla \bar{\varphi}_h^T \geq \int_{A_\epsilon} \nabla \bar{\varphi} \epsilon(s, n)\nabla \bar{\varphi}^T, \text{ for all } \epsilon > 0.$$

Taking $\epsilon \to 0$, and combining with the other convergences, we arrive at (A.27). \hfill \Box

Acknowledgements. The author acknowledges financial support by the NSF via DMS-1555222 (CAREER), as well as the IMA at the University of Minnesota (Spring 2018).
REFERENCES


