

## A FINITE ELEMENT METHOD FOR NEMATIC LIQUID CRYSTALS WITH VARIABLE DEGREE OF ORIENTATION\*

RICARDO H. NOCHETTO<sup>†</sup>, SHAWN W. WALKER<sup>‡</sup>, AND WUJUN ZHANG<sup>§</sup>

**Abstract.** We consider the simplest one-constant model, put forward by J. Ericksen, for nematic liquid crystals with variable degree of orientation. The equilibrium state is described by a director field  $\mathbf{n}$  and its degree of orientation  $s$ , where the pair  $(s, \mathbf{n})$  minimizes a sum of Frank-like energies and a double well potential. In particular, the Euler–Lagrange equations for the minimizer contain a degenerate elliptic equation for  $\mathbf{n}$ , which allows for line and plane defects to have finite energy. We present a structure preserving discretization of the liquid crystal energy with piecewise linear finite elements that can handle the degenerate elliptic part without regularization, and we show that it is consistent and stable. We prove  $\Gamma$ -convergence of discrete global minimizers to continuous ones as the mesh size goes to zero. We develop a quasi-gradient flow scheme for computing discrete equilibrium solutions and prove that it has a strictly monotone energy decreasing property. We present simulations in two and three dimensions to illustrate the method’s ability to handle nontrivial defects. The following online video illustrates the role of numerical analysis for the simulation of liquid crystal phenomena: [Walker, Shawn. “Mathematical Modeling and Simulation of Nematic Liquid Crystals (A Montage).” YouTube video, 04:13. Posted March 6, 2016. <http://www.youtube.com/watch?v=pWWw7.6cQ-U>].

**Key words.** liquid crystals, finite element method,  $\Gamma$ -convergence, gradient flow, line defect, plane defect

**AMS subject classifications.** 65N30, 49M25, 35J70

**DOI.** 10.1137/15M103844X

**1. Introduction.** Complex fluids are ubiquitous in nature and industrial processes and are critical for modern engineering systems [32, 41, 15]. A critical difficulty in modeling and simulating complex fluids is their inherent microstructure. Manipulating the microstructure via external forces can enable control of the mechanical, chemical, optical, or thermal properties of the material. Liquid crystals [47, 25, 21, 4, 3, 13, 7, 33, 34, 5, 46] are a relatively simple example of a material with microstructure that may be immersed in a fluid with a free interface [53, 52].

Several numerical methods for liquid crystals have been proposed in [10, 29, 23, 35, 2] for harmonic mappings and liquid crystals with fixed degree of orientation; i.e., a unit vector field  $\mathbf{n}(x)$  (called the director field) is used to represent the orientation of liquid crystal molecules. See [28, 36, 49] for methods that couple liquid crystals to Stokes flow. We also refer the reader to the survey paper [6] for more numerical methods.

In this paper, we consider the one-constant model for liquid crystals with variable

---

\*Received by the editors September 8, 2015; accepted for publication (in revised form) February 7, 2017; published electronically June 7, 2017.

<http://www.siam.org/journals/sinum/55-3/M103844.html>

**Funding:** The first and third authors were partially supported by NSF grant DMS-1411808. The second author was partially supported by NSF grants DMS-1418994 and DMS-1555222. The first author was also supported by the Institut Henri Poincaré (Paris). The third author was also supported by the University of Maryland through the Brin postdoctoral fellowship.

<sup>†</sup>Department of Mathematics and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742 (rhn@math.umd.edu).

<sup>‡</sup>Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918 (walker@math.lsu.edu).

<sup>§</sup>Department of Mathematics, Rutgers University, Piscataway, NJ 08854 (wujun@math.rutgers.edu).

degree of orientation [26, 25, 47]. The state of the liquid crystal is described by a director field  $\mathbf{n}(x)$  and a scalar function  $s(x)$ ,  $-1/2 < s < 1$ , that represents the degree of alignment that molecules have with respect to  $\mathbf{n}$ . The equilibrium state is given by  $(s, \mathbf{n})$ , which minimizes the so-called one-constant Ericksen's energy (2.1).

Despite the simple form of the one-constant Ericksen's model, its minimizer may have nontrivial defects. If  $s$  is a nonvanishing constant, then the energy reduces to the Oseen–Frank energy, whose minimizers are harmonic maps that may exhibit point defects (depending on boundary conditions) [14, 16, 20, 34, 33, 42]. If  $s$  is part of the minimization of (2.1), then  $s$  may vanish to allow for line (and plane) defects in dimension  $d = 3$  [5, 46], and the resulting Euler–Lagrange equation for  $\mathbf{n}$  is degenerate. However, in [34] it was shown that both  $s$  and  $\mathbf{u} = s\mathbf{n}$  have strong limits, which enabled the study of regularity properties of minimizers and the size of defects. This inspired the study of dynamics [21] and corresponding numerics [8], which are most relevant to our paper. However, in both studies the model was regularized to avoid the degeneracy introduced by the  $s$  parameter.

We design a finite element method (FEM) *without* any regularization. We prove stability and convergence properties and explore equilibrium configurations of liquid crystals via quasi-gradient flows. Our method builds on [12, 9, 11] and consists of a structure preserving discretization of (2.1). Given a weakly acute mesh  $\mathcal{T}_h$  with mesh size  $h$  (see section 2.2), we use the subscript  $h$  to denote continuous piecewise linear functions defined over  $\mathcal{T}_h$ , e.g.,  $(s_h, \mathbf{n}_h)$  is a discrete approximation of  $(s, \mathbf{n})$ .

Our discretization of the energy is defined in (2.18) and requires that  $\mathcal{T}_h$  be *weakly acute*. This discretization preserves the underlying structure and converges to the continuous energy in the sense of  $\Gamma$ -convergence [17] as  $h$  goes to zero. Next, we develop a quasi-gradient flow scheme for computing discrete equilibrium solutions. We prove that this scheme has a strictly monotone energy decreasing property. Finally, we carry out numerical experiments and show that our FEM and gradient flow allow for computing minimizers that exhibit line and plane defects.

The paper is organized as follows. In section 2, we describe Ericksen's model for liquid crystals with variable degree of orientation and give details of our discretization. Section 3 shows the  $\Gamma$ -convergence of our numerical method. A quasi-gradient flow scheme is given in section 4, where we also prove a strictly monotone energy decreasing property. Section 5 presents simulations in two and three dimensions that exhibit nontrivial defects in order to illustrate the method's capabilities.

**2. Discretization of Ericksen's model.** We review the model [26] and relevant analysis results from the literature. We then develop our discretization strategy and show that it is stable. The space dimension  $d \geq 2$  can be arbitrary.

**2.1. Ericksen's one-constant model.** Let the director field  $\mathbf{n} : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{S}^{d-1}$  be a vector-valued function with unit length, and let the degree of orientation  $s : \Omega \subset \mathbb{R}^d \rightarrow [-\frac{1}{2}, 1]$  be a real-valued function. The case  $s = 1$  represents the state of perfect alignment in which all molecules are parallel to  $\mathbf{n}$ . Likewise,  $s = -1/2$  represents the state of microscopic order in which all molecules are orthogonal to the orientation  $\mathbf{n}$ . When  $s = 0$ , the molecules do not lie along any preferred direction, which represents the state of an isotropic distribution of molecules.

The equilibrium state of the liquid crystals is described by the pair  $(s, \mathbf{n})$  minimizing a bulk-energy functional, which in the simplest one-constant model reduces to

$$(2.1) \quad E[s, \mathbf{n}] := \underbrace{\int_{\Omega} (\kappa |\nabla s|^2 + s^2 |\nabla \mathbf{n}|^2) dx}_{=: E_1[s, \mathbf{n}]} + \underbrace{\int_{\Omega} \psi(s) dx}_{=: E_2[s]},$$

with  $\kappa > 0$  and double well potential  $\psi$ , which is a  $C^2$  function defined on  $-1/2 < s < 1$  that satisfies the following properties:

1.  $\lim_{s \rightarrow 1} \psi(s) = \lim_{s \rightarrow -1/2} \psi(s) = \infty$ ,
2.  $\psi(0) > \psi(s^*) = \min_{s \in [-1/2, 1]} \psi(s) = 0$  for some  $s^* \in (0, 1)$ ,
3.  $\psi'(0) = 0$ ;

see [26]. Note that when the degree of orientation  $s$  equals a nonzero constant, the energy (2.1) effectively reduces to the Oseen–Frank energy  $\int_{\Omega} |\nabla \mathbf{n}|^2$ . The degree of orientation  $s$  relaxes the energy of defects (i.e., discontinuities in  $\mathbf{n}$ ), which may still have finite energy  $E[s, \mathbf{n}]$  if the singular set

$$(2.2) \quad \mathcal{S} := \{x \in \Omega, s(x) = 0\}$$

is nonempty; in this case,  $\mathbf{n} \notin H^1(\Omega)$ .

By introducing an auxiliary variable  $\mathbf{u} = s\mathbf{n}$  [34, 3], we rewrite the energy as

$$(2.3) \quad E_1[s, \mathbf{n}] = \tilde{E}_1[s, \mathbf{u}] := \int_{\Omega} ((\kappa - 1)|\nabla s|^2 + |\nabla \mathbf{u}|^2) dx,$$

which follows from the orthogonal splitting  $\nabla \mathbf{u} = \mathbf{n} \otimes \nabla s + s \nabla \mathbf{n}$  due to the constraint  $|\mathbf{n}| = 1$ . Accordingly, we define the admissible class

$$(2.4) \quad \mathbb{A} := \{(s, \mathbf{u}) : \Omega \rightarrow (-1/2, 1) \times \mathbb{R}^d : (s, \mathbf{u}) \in [H^1(\Omega)]^{d+1}, \mathbf{u} = s\mathbf{n}, \mathbf{n} \in \mathbb{S}^{d-1}\}.$$

We say that the pair  $(s, \mathbf{u})$  satisfies the *structural condition* for Ericksen’s energy if

$$(2.5) \quad \mathbf{u} = s\mathbf{n}, \quad -1/2 < s < 1 \quad \text{a.e. in } \Omega, \quad \text{and } \mathbf{n} \in \mathbb{S}^{d-1} \quad \text{a.e. in } \Omega.$$

Moreover, we may enforce boundary conditions on  $(s, \mathbf{u})$ —possibly on different parts of the boundary. Let  $(\Gamma_s, \Gamma_{\mathbf{u}})$  be open subsets of  $\partial\Omega$ , where we set Dirichlet boundary conditions for  $(s, \mathbf{u})$ . Then we have the restricted admissible class

$$(2.6) \quad \mathbb{A}(g, \mathbf{r}) := \{(s, \mathbf{u}) \in \mathbb{A} : s|_{\Gamma_s} = g, \quad \mathbf{u}|_{\Gamma_{\mathbf{u}}} = \mathbf{r}\}$$

for some given functions  $(g, \mathbf{r}) \in [W_{\infty}^1(\mathbb{R}^d)]^{d+1}$  that satisfy the structural condition (2.5) on  $\partial\Omega$ . We assume the existence of  $\delta_0 > 0$  sufficiently small such that

$$(2.7) \quad -\frac{1}{2} + \delta_0 \leq g(x), \quad \mathbf{r}(x) \cdot \xi \leq 1 - \delta_0 \quad \text{for all } x \in \mathbb{R}^d, \xi \in \mathbb{R}^d, |\xi| = 1,$$

and the potential  $\psi$  satisfies

$$(2.8) \quad \psi(s) \geq \psi(1 - \delta_0) \quad \text{for } s \geq 1 - \delta_0, \quad \psi(s) \geq \psi\left(-\frac{1}{2} + \delta_0\right) \quad \text{for } s \leq -\frac{1}{2} + \delta_0.$$

This is consistent with property 1 of  $\psi$ . If we further assume that

$$(2.9) \quad g \geq \delta_0 \quad \text{on } \partial\Omega,$$

then the function  $\mathbf{n}$  is  $H^1$  in a neighborhood of  $\partial\Omega$  and satisfies  $\mathbf{n} = g^{-1}\mathbf{r}$  on  $\partial\Omega$ .

The existence of a minimizer  $(s, \mathbf{u}) \in \mathbb{A}(g, \mathbf{r})$  is shown in [34, 3], but this is also a consequence of our  $\Gamma$ -convergence theory. It is worth mentioning that the constant  $\kappa$  in  $E[s, \mathbf{n}]$  of (2.1) plays a significant role in the occurrence of defects. Roughly speaking, if  $\kappa$  is large, then  $\int_{\Omega} \kappa |\nabla s|^2 dx$  dominates the energy and  $s$  is close to a constant. In this case, defects with finite energy are less likely to occur. But if  $\kappa$  is small, then  $\int_{\Omega} s^2 |\nabla \mathbf{n}|^2 dx$  dominates the energy, and  $s$  may become zero. In this case, defects are more likely to occur. (This heuristic argument is later confirmed in the numerical experiments.) Since the investigation of defects is of primary interest in this paper, we consider the most significant case to be  $0 < \kappa < 1$ .

We now describe our finite element discretization  $E_h[s_h, \mathbf{n}_h]$  of the energy (2.1) and its minimizer  $(s_h, \mathbf{n}_h)$ .

**2.2. Discretization of the energy.** Let  $\mathcal{T}_h = \{T\}$  be a conforming simplicial triangulation of the domain  $\Omega$ . We denote by  $\mathcal{N}_h$  the set of nodes (vertices) of  $\mathcal{T}_h$  and by  $N$  the cardinality of  $\mathcal{N}_h$  (with some abuse of notation). We demand that  $\mathcal{T}_h$  be *weakly acute*, namely,

$$(2.10) \quad k_{ij} := - \int_{\Omega} \nabla \phi_i \cdot \nabla \phi_j dx \geq 0 \quad \text{for all } i \neq j,$$

where  $\phi_i$  is the standard “hat” function associated with node  $x_i \in \mathcal{N}_h$ . We indicate with  $\omega_i = \text{supp } \phi_i$  the patch of a node  $x_i$  (i.e., the “star” of elements in  $\mathcal{T}_h$  that contain the vertex  $x_i$ ). Condition (2.10) imposes a severe geometric restriction on  $\mathcal{T}_h$  [22, 45]. We recall the following characterization of (2.10) for  $d = 2$ .

**PROPOSITION 2.1** (weak acuteness in two dimensions). *For any pair of triangles  $T_1, T_2$  in  $\mathcal{T}_h$  that share a common edge  $e$ , let  $\alpha_i$  be the angle in  $T_i$  opposite to  $e$  (for  $i = 1, 2$ ). If  $\alpha_1 + \alpha_2 \leq \pi$  for every edge  $e$ , then (2.10) holds.*

Generalizations of Proposition 2.1 to three dimensions, involving interior dihedral angles of tetrahedra, can be found in [30, 19].

We construct continuous piecewise affine spaces associated with the mesh, i.e.,

$$(2.11) \quad \begin{aligned} \mathbb{S}_h &:= \{s_h \in H^1(\Omega) : s_h|_T \text{ is affine for all } T \in \mathcal{T}_h\}, \\ \mathbb{U}_h &:= \{\mathbf{u}_h \in H^1(\Omega)^d : \mathbf{u}_h|_T \text{ is affine in each component for all } T \in \mathcal{T}_h\}, \\ \mathbb{N}_h &:= \{\mathbf{n}_h \in \mathbb{U}_h : |\mathbf{n}_h(x_i)| = 1 \text{ for all nodes } x_i \in \mathcal{N}_h\}. \end{aligned}$$

Let  $I_h$  denote the piecewise linear Lagrange interpolation operator on mesh  $\mathcal{T}_h$  with values in either  $\mathbb{S}_h$  or  $\mathbb{U}_h$ . We say that a pair  $(s_h, \mathbf{u}_h) \in \mathbb{S}_h \times \mathbb{U}_h$  satisfies the *discrete structural condition* for Ericksen’s energy if there exists  $\mathbf{n}_h \in \mathbb{N}_h$  such that

$$(2.12) \quad \mathbf{u}_h = I_h[s_h \mathbf{n}_h], \quad -\frac{1}{2} < s_h < 1 \quad \text{in } \Omega.$$

We then let  $g_h := I_h g$  and  $\mathbf{r}_h := I_h \mathbf{r}$  be the discrete Dirichlet data, and we introduce the discrete spaces that include (Dirichlet) boundary conditions

$$\mathbb{S}_h(\Gamma_s, g_h) := \{s_h \in \mathbb{S}_h : s_h|_{\Gamma_s} = g_h\}, \quad \mathbb{U}_h(\Gamma_{\mathbf{u}}, \mathbf{r}_h) := \{\mathbf{u}_h \in \mathbb{U}_h : \mathbf{u}_h|_{\Gamma_{\mathbf{u}}} = \mathbf{r}_h\},$$

along with the discrete admissible class

$$(2.13) \quad \mathbb{A}_h(g_h, \mathbf{r}_h) := \left\{ (s_h, \mathbf{u}_h) \in \mathbb{S}_h(\Gamma_s, g_h) \times \mathbb{U}_h(\Gamma_{\mathbf{u}}, \mathbf{r}_h) : (2.12) \text{ holds} \right\}.$$

In view of (2.9), we can also impose the Dirichlet condition  $\mathbf{n}_h = I_h[g_h^{-1} \mathbf{r}_h]$  on  $\partial\Omega$ .

In order to motivate our discrete version of  $E_1[s, \mathbf{n}]$ , note that for all  $x_i \in \mathcal{N}_h$ ,

$$\sum_{j=1}^N k_{ij} = - \sum_{j=1}^N \int_{\Omega} \nabla \phi_i \cdot \nabla \phi_j dx = 0$$

because  $\sum_{j=1}^N \phi_j = 1$  in the domain  $\Omega$ ; the set of hat functions  $\{\phi_j\}_{j=1}^N$  is a partition of unity. Therefore, for piecewise linear  $s_h = \sum_{i=1}^N s_h(x_i)\phi_i$ , we have

$$\int_{\Omega} |\nabla s_h|^2 dx = - \sum_{i=1}^N k_{ii} s_h(x_i)^2 - \sum_{i,j=1, i \neq j}^N k_{ij} s_h(x_i) s_h(x_j),$$

whence, exploiting  $k_{ii} = - \sum_{j \neq i} k_{ij}$  and the symmetry  $k_{ij} = k_{ji}$ , we get

$$\begin{aligned} \int_{\Omega} |\nabla s_h|^2 dx &= \sum_{i,j=1}^N k_{ij} s_h(x_i) (s_h(x_i) - s_h(x_j)) \\ (2.14) \qquad &= \frac{1}{2} \sum_{i,j=1}^N k_{ij} (s_h(x_i) - s_h(x_j))^2 = \frac{1}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2, \end{aligned}$$

where we define

$$(2.15) \qquad \delta_{ij} s_h := s_h(x_i) - s_h(x_j), \quad \delta_{ij} \mathbf{n}_h := \mathbf{n}_h(x_i) - \mathbf{n}_h(x_j).$$

With this in mind, we define the discrete energies to be

$$(2.16) \quad E_1^h[s_h, \mathbf{n}_h] := \frac{\kappa}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2 + \frac{1}{2} \sum_{i,j=1}^N k_{ij} \left( \frac{s_h(x_i)^2 + s_h(x_j)^2}{2} \right) |\delta_{ij} \mathbf{n}_h|^2$$

and

$$(2.17) \qquad E_2^h[s_h] := \int_{\Omega} \psi(s_h(x)) dx,$$

for  $(s_h, \mathbf{u}_h) \in \mathbb{A}_h(g_h, \mathbf{u}_h)$ . The second summation in (2.16) does *not* come from applying the standard discretization of  $\int_{\Omega} s^2 |\nabla \mathbf{n}|^2 dx$  by piecewise linear elements. It turns out that this special form of the discrete energy preserves the key energy inequality (Lemma 2.2), which allows us to establish our  $\Gamma$ -convergence analysis for the degenerate coefficient  $s^2$  *without* regularization. Eventually, we seek an approximation  $(s_h, \mathbf{u}_h) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$  of the pair  $(s, \mathbf{u})$  such that the discrete pair  $(s_h, \mathbf{n}_h)$  minimizes the discrete version of the bulk energy (2.1) given by

$$(2.18) \qquad E_h[s_h, \mathbf{n}_h] := E_1^h[s_h, \mathbf{n}_h] + E_2^h[s_h].$$

The following result shows that definition (2.16) preserves the key structure (2.3) of [3, 34] at the discrete level, which turns out to be crucial for our analysis as well.

We first introduce  $\tilde{s}_h := I_h[s_h]$  and two discrete versions of the vector field  $\mathbf{u}$ ,

$$(2.19) \qquad \mathbf{u}_h := I_h[s_h \mathbf{n}_h] \in \mathbb{U}_h, \quad \tilde{\mathbf{u}}_h := I_h[\tilde{s}_h \mathbf{n}_h] \in \mathbb{U}_h.$$

Note that both pairs  $(s_h, \mathbf{u}_h), (\tilde{s}_h, \tilde{\mathbf{u}}_h) \in \mathbb{S}_h \times \mathbb{U}_h$  satisfy (2.12).

LEMMA 2.2 (energy inequality). *Let the mesh  $\mathcal{T}_h$  satisfy (2.10). If  $(s_h, \mathbf{u}_h) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$ , then for any  $\kappa > 0$ , the discrete energy (2.16) satisfies*

$$(2.20) \quad E_1^h[s_h, \mathbf{n}_h] \geq (\kappa - 1) \int_{\Omega} |\nabla s_h|^2 dx + \int_{\Omega} |\nabla \mathbf{u}_h|^2 dx =: \tilde{E}_1^h[s_h, \mathbf{u}_h]$$

as well as

$$(2.21) \quad E_1^h[s_h, \mathbf{n}_h] \geq (\kappa - 1) \int_{\Omega} |\nabla \tilde{s}_h|^2 dx + \int_{\Omega} |\nabla \tilde{\mathbf{u}}_h|^2 dx =: \tilde{E}_1^h[\tilde{s}_h, \tilde{\mathbf{u}}_h].$$

*Proof.* Since

$$\begin{aligned} s_h(x_i) \mathbf{n}_h(x_i) - s_h(x_j) \mathbf{n}_h(x_j) &= \frac{s_h(x_i) + s_h(x_j)}{2} (\mathbf{n}_h(x_i) - \mathbf{n}_h(x_j)) \\ &\quad + (s_h(x_i) - s_h(x_j)) \frac{\mathbf{n}_h(x_i) + \mathbf{n}_h(x_j)}{2}, \end{aligned}$$

using the orthogonality relation  $(\mathbf{n}_h(x_i) - \mathbf{n}_h(x_j)) \cdot (\mathbf{n}_h(x_i) + \mathbf{n}_h(x_j)) = |\mathbf{n}_h(x_i)|^2 - |\mathbf{n}_h(x_j)|^2 = 0$  and (2.14) yields

$$\begin{aligned} \int_{\Omega} |\nabla \mathbf{u}_h|^2 dx &= \frac{1}{2} \sum_{i,j=1}^N k_{ij} |s_h(x_i) \mathbf{n}_h(x_i) - s_h(x_j) \mathbf{n}_h(x_j)|^2 \\ &= \frac{1}{2} \sum_{i,j=1}^N k_{ij} \left( \frac{s_h(x_i) + s_h(x_j)}{2} \right)^2 |\delta_{ij} \mathbf{n}_h|^2 + \frac{1}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2 \left| \frac{\mathbf{n}_h(x_i) + \mathbf{n}_h(x_j)}{2} \right|^2. \end{aligned}$$

Exploiting the relations  $|\mathbf{n}_h(x_i) - \mathbf{n}_h(x_j)|^2 + |\mathbf{n}_h(x_i) + \mathbf{n}_h(x_j)|^2 = 4$  and  $(s_h(x_i) + s_h(x_j))^2 = 2(s_h(x_i)^2 + s_h(x_j)^2) - (s_h(x_i) - s_h(x_j))^2$ , we obtain

$$(2.22) \quad \begin{aligned} \int_{\Omega} |\nabla \mathbf{u}_h|^2 dx &= \frac{1}{2} \sum_{i,j=1}^N k_{ij} \frac{s_h(x_i)^2 + s_h(x_j)^2}{2} |\delta_{ij} \mathbf{n}_h|^2 \\ &\quad + \frac{1}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2 - \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2 \left| \frac{\mathbf{n}_h(x_i) - \mathbf{n}_h(x_j)}{2} \right|^2, \end{aligned}$$

whence we infer that

$$(2.23) \quad E_1^h[s_h, \mathbf{n}_h] = \int_{\Omega} \left( (\kappa - 1) |\nabla s_h|^2 + |\nabla \mathbf{u}_h|^2 \right) dx + \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2 \left| \frac{\delta_{ij} \mathbf{n}_h}{2} \right|^2.$$

The inequality (2.20) follows directly from  $k_{ij} \geq 0$  for  $i \neq j$ .

To prove (2.21), we note that (2.22) still holds if we replace  $(s_h, \mathbf{u}_h)$  with  $(\tilde{s}_h, \tilde{\mathbf{u}}_h)$ :

$$(2.24) \quad \begin{aligned} \int_{\Omega} |\nabla \tilde{\mathbf{u}}_h|^2 dx &= \frac{1}{2} \sum_{i,j=1}^N k_{ij} \frac{\tilde{s}_h(x_i)^2 + \tilde{s}_h(x_j)^2}{2} |\delta_{ij} \mathbf{n}_h|^2 \\ &\quad + \frac{1}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} \tilde{s}_h)^2 - \sum_{i,j=1}^N k_{ij} (\delta_{ij} \tilde{s}_h)^2 \left| \frac{\mathbf{n}_h(x_i) - \mathbf{n}_h(x_j)}{2} \right|^2. \end{aligned}$$

We finally find that

$$\begin{aligned} \tilde{E}_1^h[\tilde{s}_h, \tilde{\mathbf{u}}_h] &= \int_{\Omega} \left( |\nabla \tilde{\mathbf{u}}_h|^2 + (\kappa - 1) |\nabla \tilde{s}_h|^2 \right) dx = \frac{1}{2} \sum_{i,j=1}^N k_{ij} \frac{\tilde{s}_h(x_i)^2 + \tilde{s}_h(x_j)^2}{2} |\delta_{ij} \mathbf{n}_h|^2 \\ &\quad + \frac{\kappa}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} \tilde{s}_h)^2 - \sum_{i,j=1}^N k_{ij} (\delta_{ij} \tilde{s}_h)^2 \left| \frac{\mathbf{n}_h(x_i) - \mathbf{n}_h(x_j)}{2} \right|^2 \leq E_1^h[s_h, \mathbf{n}_h], \end{aligned}$$

where we have dropped the last term and used the triangle inequality  $|\delta_{ij} \tilde{s}_h| = |\tilde{s}_h(x_i) - \tilde{s}_h(x_j)| \leq |s_h(x_i) - s_h(x_j)| = |\delta_{ij} s_h|$  along with  $k_{ij} \geq 0$  to obtain

$$(2.25) \quad \|\nabla \tilde{s}_h\|_{L^2(\Omega)}^2 = \frac{1}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} \tilde{s}_h)^2 \leq \frac{1}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2 = \|\nabla s_h\|_{L^2(\Omega)}^2.$$

This concludes the proof. □

*Remark 2.3* (relation between (2.20) and (2.21)). Both (2.20) and (2.21) account for the *variational crime* that is committed when enforcing  $\mathbf{u}_h = s_h \mathbf{n}_h$  and  $\tilde{\mathbf{u}}_h = \tilde{s}_h \mathbf{n}_h$  only at the vertices, and that mimics (2.3). Since we have a precise control of the consistency error for (2.20), this inequality will be used later for the consistency (or lim-sup) step of  $\Gamma$ -convergence of our discrete energy (2.16) to the original continuous energy in (2.1). On the other hand, (2.21) has a suitable structure to prove the weak lower semicontinuity (or lim-inf) step of  $\Gamma$ -convergence. This property is not obvious when  $\kappa < 1$ , the most significant case for the formation of defects.

**3.  $\Gamma$ -convergence of the discrete energy.** In this section, we show that our discrete energy (2.16) converges to the continuous energy (2.1) in the sense of  $\Gamma$ -convergence. To this end, we first let the continuous and discrete spaces be

$$\mathbb{X} := L^2(\Omega) \times [L^2(\Omega)]^d, \quad \mathbb{X}_h := \mathbb{S}_h \times \mathbb{U}_h.$$

We next define  $E[s, \mathbf{n}]$  as in (2.1) for  $(s, \mathbf{u}) \in \mathbb{A}(g, \mathbf{r})$  and  $E[s, \mathbf{u}] = \infty$  for  $(s, \mathbf{n}) \in \mathbb{X} \setminus \mathbb{A}(g, \mathbf{r})$ . Likewise, we define  $E_h[s_h, \mathbf{n}_h]$  as in (2.18) for  $(s_h, \mathbf{u}_h) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$  and  $E_h[s, \mathbf{n}] = \infty$  for all  $(s, \mathbf{u}) \in \mathbb{X} \setminus \mathbb{A}_h(g_h, \mathbf{r}_h)$ .

We split the proof of  $\Gamma$ -convergence into four subsections. In subsection 3.1, we use the energy  $\tilde{E}_1^h[s_h, \mathbf{u}_h]$  to show the consistency property (recall Remark 2.3), whereas we employ the energy  $\tilde{E}_1^h[\tilde{s}_h, \tilde{\mathbf{u}}_h]$  in subsection 3.2 to derive the weak lower semicontinuity property. Furthermore, our functionals exhibit the usual equi-coercivity property for both pairs  $(s, \mathbf{u})$  and  $(\tilde{s}, \tilde{\mathbf{u}})$  but not for the director field  $\mathbf{n}$ , which is only well defined whenever the order parameter  $s \neq 0$ . We discuss these issues in subsection 3.3 and characterize the limits  $(s, \mathbf{u})$ ,  $(\tilde{s}, \tilde{\mathbf{u}})$ , and  $(s, \mathbf{n})$ . We eventually prove  $\Gamma$ -convergence in subsection 3.4 by combining these results.

**3.1. Consistency or lim-sup property.** We prove the following: if  $(s, \mathbf{u}) \in \mathbb{A}(g, \mathbf{r})$ , then there exist a sequence  $(s_h, \mathbf{u}_h) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$  converging to  $(s, \mathbf{u})$  in  $H^1(\Omega)$  and a discrete director field  $\mathbf{n}_h \in \mathbb{N}_h$  converging to  $\mathbf{n}$  in  $L^2(\Omega \setminus \mathcal{S})$  such that

$$(3.1) \quad E_1[s, \mathbf{n}] \geq \limsup_{h \rightarrow 0} E_1^h[s_h, \mathbf{n}_h] \geq \limsup_{h \rightarrow 0} \tilde{E}_1^h[s_h, \mathbf{u}_h].$$

We observe that if  $(s, \mathbf{u}) \notin \mathbb{A}(g, \mathbf{r})$ , then  $E_1[s, \mathbf{n}] = \infty$  and (3.1) is valid for any sequences  $(s_h, \mathbf{u}_h)$  and  $(s_h, \mathbf{n}_h)$  in light of (2.20).

We first show that we can always assume  $-\frac{1}{2} + \delta_o \leq s \leq 1 - \delta_o$  for  $(s, \mathbf{u}) \in \mathbb{A}(g, \mathbf{r})$ .

LEMMA 3.1 (truncation). *Given  $(s, \mathbf{u}) \in \mathbb{A}(g, \mathbf{r})$ , let  $(\hat{s}, \hat{\mathbf{u}})$  be the truncations*

$$\hat{s}(x) = \min \left\{ 1 - \delta_0, \max \left( -\frac{1}{2} + \delta_0, s(x) \right) \right\}, \quad \hat{\mathbf{u}}(x) = \hat{s}(x) \mathbf{u}(x) \quad \text{a.e. } x \in \Omega.$$

Then  $(\hat{s}, \hat{\mathbf{u}}) \in \mathbb{A}(g, \mathbf{r})$  and

$$E_1[\hat{s}, \mathbf{n}] \leq E_1[s, \mathbf{n}], \quad E_2[\hat{s}] \leq E_2[s].$$

The same assertion is true for any  $(s_h, \mathbf{u}_h) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$ , except that the truncations are defined nodewise, i.e.,  $(I_h \hat{s}_h, I_h \hat{\mathbf{u}}_h) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$ .

*Proof.* The fact that  $(\hat{s}, \hat{\mathbf{u}})$  satisfy the Dirichlet boundary conditions is a consequence of (2.7). Moreover,  $(\hat{s}, \hat{\mathbf{u}}) \in [H^1(\Omega)]^{d+1}$  and the structural property (2.5) holds by construction, whence  $(\hat{s}, \hat{\mathbf{u}}) \in \mathbb{A}(g, \mathbf{r})$ . We next observe that

$$\nabla \hat{s} = \chi_{\Omega_0} \nabla s, \quad \Omega_0 := \left\{ x \in \Omega : -\frac{1}{2} + \delta_0 \leq s(x) \leq 1 - \delta_0 \right\};$$

see [27, Chap. 5, Exercise 17]. Consequently, we obtain

$$E_1[\hat{s}, \mathbf{n}] = \int_{\Omega} \kappa |\nabla \hat{s}|^2 + |\hat{s}|^2 |\nabla \mathbf{n}|^2 \leq \int_{\Omega} \kappa |\nabla s|^2 + |s|^2 |\nabla \mathbf{n}|^2 = E_1[s, \mathbf{n}],$$

as well as

$$E_2[\hat{s}] = \int_{\Omega} \psi(\hat{s}) \leq \int_{\Omega} \psi(s) = E_2[s],$$

because of (2.8). This concludes the proof.  $\square$

To construct a recovery sequence  $(s_h, \mathbf{u}_h) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$  we need point values of  $(s, \mathbf{u})$  and thus a regularization procedure of functions in the admissible class  $\mathbb{A}(g, \mathbf{r})$ . We must enforce both the structural property (2.5) and the Dirichlet boundary conditions  $s = g$  and  $\mathbf{u} = \mathbf{r}$ ; neither is guaranteed by convolution. We are able to do this provided  $\Gamma_s = \Gamma_{\mathbf{u}} = \partial\Omega$  and the Dirichlet datum  $g$  satisfies (2.9).

PROPOSITION 3.2 (regularization of functions in  $\mathbb{A}(g, \mathbf{r})$ ). *Let  $\Gamma_s = \Gamma_{\mathbf{u}} = \partial\Omega$ ,  $(s, \mathbf{u}) \in \mathbb{A}(g, \mathbf{r})$ , and let  $g$  satisfy (2.9). Given  $\epsilon > 0$  there exists a pair  $(s_{\epsilon}, \mathbf{u}_{\epsilon}) \in \mathbb{A}(g, \mathbf{r}) \cap [W_{\infty}^1(\Omega)]^{d+1}$  such that*

$$(3.2) \quad \|(s, \mathbf{u}) - (s_{\epsilon}, \mathbf{u}_{\epsilon})\|_{H^1(\Omega)} \leq \epsilon,$$

$$(3.3) \quad -\frac{1}{2} + \delta_0 \leq s_{\epsilon}(x), \quad \mathbf{u}_{\epsilon}(x) \cdot \xi \leq 1 - \delta_0 \quad \text{for all } x \in \Omega, \xi \in \mathbb{R}^d, |\xi| = 1.$$

*Proof.* We construct a two-scale approximation with scales  $\delta < \sigma$ , which satisfies the boundary conditions exactly. We split the argument into several steps.

*Step 1: Regularization with Dirichlet condition.* Extend  $s - g \in H_0^1(\Omega)$  by zero to  $\mathbb{R}^d \setminus \Omega$ . Let  $\eta_{\delta}$  be a smooth and nonnegative mollifier with support contained in the ball  $B_{\delta}(0)$  centered at 0 with radius  $\delta$ . Define  $d_{\delta} : \mathbb{R}^d \rightarrow \mathbb{R}$  by

$$d_{\delta}(x) := \chi_{\Omega}(x) \min\{\delta^{-1} \text{dist}(x, \partial\Omega), 1\},$$

which is Lipschitz in  $\mathbb{R}^d$ , and observe that  $\nabla d_{\delta}$  is supported in the boundary layer

$$\omega_{\delta} := \{x \in \Omega : \text{dist}(x, \partial\Omega) \leq \delta\},$$

and that  $|\nabla d_\delta| = \delta^{-1} \chi_{\omega_\delta}$ . We consider the Lipschitz approximations of  $(s, \mathbf{u})$  given by

$$s_\delta := d_\delta (s * \eta_\delta) + (1 - d_\delta)g, \quad \mathbf{u}_\delta := d_\delta (\mathbf{u} * \eta_\delta) + (1 - d_\delta)\mathbf{r}.$$

Since  $d_\delta$  vanishes on  $\partial\Omega$ , we readily see that  $(s_\delta, \mathbf{u}_\delta) = (g, \mathbf{r})$  on  $\partial\Omega$ . Moreover, the following properties are valid:

$$(3.4) \quad s_\delta \rightarrow s, \quad \mathbf{u}_\delta \rightarrow \mathbf{u}, \quad |\mathbf{u}_\delta| \rightarrow |\mathbf{u}| \quad \text{a.e. and in } H^1(\Omega).$$

The last property is a consequence of the middle one via triangle inequality, and the first two are similar. It thus suffices to show the first property for  $s$ . We simply write

$$\nabla(s_\delta - s) = \nabla d_\delta (s - g) * \eta_\delta + \nabla d_\delta (g * \eta_\delta - g) + d_\delta \nabla (s * \eta_\delta - s) + (d_\delta - 1) \nabla (s - g).$$

Since  $s - g \in H^1(\omega_\delta)$  and  $s - g = 0$  on  $\partial\Omega$ , we apply Poincaré’s inequality to deduce

$$\|s - g\|_{L^2(\omega_\delta)} \leq C\delta \|\nabla(s - g)\|_{L^2(\omega_\delta)},$$

whence

$$\|\nabla d_\delta (s - g) * \eta_\delta\|_{L^2(\Omega)} \leq C\delta^{-1} \|s - g\|_{L^2(\omega_\delta)} \leq C \|\nabla(s - g)\|_{L^2(\omega_\delta)} \rightarrow 0 \quad \text{as } \delta \rightarrow 0.$$

Likewise, a similar argument gives, for the fourth term,

$$\|(d_\delta - 1) \nabla(s - g)\|_{L^2(\Omega)} \leq C \|\nabla(s - g)\|_{L^2(\omega_\delta)} \rightarrow 0 \quad \text{as } \delta \rightarrow 0.$$

On the other hand, the estimate  $\|g * \eta_\delta - g\|_{L^\infty(\Omega)} \leq \delta \|\nabla g\|_{L^\infty(\mathbb{R}^d)}$  yields

$$\|g * \eta_\delta - g\|_{L^2(\omega_\delta)} \leq |\omega_\delta|^{1/2} \delta \|\nabla g\|_{L^\infty(\mathbb{R}^d)} \leq C\delta^{3/2} \|\nabla g\|_{L^\infty(\mathbb{R}^d)},$$

which implies, for the second term above,

$$\|\nabla d_\delta (g * \eta_\delta - g)\|_{L^2(\Omega)} \leq C\delta^{1/2} \|\nabla g\|_{L^\infty(\mathbb{R}^d)}.$$

Finally, for the third term we recall that  $s \in H^1(\mathbb{R}^d)$  equals  $g$  outside  $\Omega$ , and we exploit the convergence  $\nabla s * \eta_\delta \rightarrow \nabla s$  in  $L^2(\Omega)$  to obtain

$$\|d_\delta \nabla (s * \eta_\delta - s)\|_{L^2(\Omega)} \rightarrow 0 \quad \text{as } \delta \rightarrow 0.$$

*Step 2: Structural condition.* Unfortunately, the pair  $(s_\delta, \mathbf{u}_\delta)$  does not satisfy the structural condition (2.5). We now construct a closely related pair that satisfies (2.5). Recall that  $(g, \mathbf{r}) \in [W^1_\infty(\mathbb{R}^d)]^{d+1}$  satisfy the bounds (2.7) in  $\mathbb{R}^d$ , and hence so do the extensions of  $(s, \mathbf{u})$  because  $s = g, \mathbf{u} = \mathbf{r}$  outside  $\Omega$ . Thus, we can show that

$$-\frac{1}{2} + \delta_0 \leq s_\delta(x), \quad \mathbf{u}_\delta(x) \cdot \xi \leq 1 - \delta_0 \quad \text{for all } x \in \Omega, \xi \in \mathbb{R}^d, |\xi| = 1;$$

we only argue with  $s_\delta$  because dealing with  $\mathbf{u}_\delta \cdot \xi$  is similar. We have  $a := -\frac{1}{2} + \delta_0 \leq s * \eta \leq 1 - \delta_0 =: b$  because  $\eta_\delta \geq 0$  and the convolution preserves constants, whence

$$s_\delta \leq d_\delta b + (1 - d_\delta)b = b, \quad s_\delta \geq d_\delta a + (1 - d_\delta)a = a \quad \text{in } \Omega.$$

We next introduce the second parameter  $\sigma > \delta$  and the Lipschitz approximation of the sign function

$$\rho_\sigma(t) = \min \{1, \max\{-1, t/\sigma\}\},$$

along with the two-scale approximation of  $(s, \mathbf{u})$ ,

$$s_{\sigma,\delta} := \rho_\sigma(s_\delta)|\mathbf{u}_\delta|, \quad \mathbf{u}_{\sigma,\delta} := |\rho_\sigma(s_\delta)|\mathbf{u}_\delta.$$

We note that  $|s_{\sigma,\delta}| = |\mathbf{u}_{\sigma,\delta}|$  by construction, whence (2.5) holds, and  $(s_{\sigma,\delta}, \mathbf{u}_{\sigma,\delta}) = (g, \mathbf{r})$  on  $\partial\Omega$  because  $\rho_\sigma(s_\delta) = 1$  on  $\partial\Omega$ , for  $\sigma \leq \delta_0$ , according to (2.9); hence  $(s_{\sigma,\delta}, \mathbf{u}_{\sigma,\delta}) \in \mathbb{A}(g, \mathbf{r}) \cap [W_\infty^1(\Omega)]^{d+1}$ . It remains to show how to choose  $\delta$  and  $\sigma$ , which we do next.

*Step 3: Convergence in  $H^1$  as  $\delta \rightarrow 0$ .* In view of (3.4) we readily deduce that

$$s_{\sigma,\delta} \rightarrow s_\sigma := \rho_\sigma(s)|s|, \quad \mathbf{u}_{\sigma,\delta} \rightarrow \mathbf{u}_\sigma := |\rho_\sigma(s)|\mathbf{u} \quad \text{a.e. and in } L^2(\Omega).$$

We now prove convergence also in  $H^1(\Omega)$ . Since  $\nabla \rho_\sigma(s_\delta) = \sigma^{-1} \chi_{\{|s_\delta| \leq \sigma\}} \nabla s_\delta$ , we get

$$\nabla \rho_\sigma(s_\delta) - \nabla \rho_\sigma(s) = \sigma^{-1} (\chi_{\{|s_\delta| \leq \sigma\}} - \chi_{\{|s| \leq \sigma\}}) \nabla s + \sigma^{-1} \chi_{\{|s_\delta| \leq \sigma\}} (\nabla s_\delta - \nabla s).$$

Applying the Lebesgue dominated convergence theorem for the first term and (3.4) for the second term yields, as  $\delta \rightarrow 0$ ,

$$(3.5) \quad \nabla \rho_\sigma(s_\delta) \rightarrow \nabla \rho_\sigma(s), \quad \nabla |\rho_\sigma(s_\delta)| \rightarrow \nabla |\rho_\sigma(s)| \quad \text{in } L^2(\Omega).$$

The second convergence result is due to the fact that  $\nabla|f| = \text{sgn}_0(f)\nabla f$  for any  $f \in W_1^1(\Omega)$ , where  $\text{sgn}_0(f)$  is the sign function that vanishes at 0 [27, Chap. 5, Exercise 17]. We next write

$$\begin{aligned} \nabla(s_{\sigma,\delta} - s_\sigma) &= \nabla(\rho_\sigma(s_\delta) - \rho_\sigma(s))(|\mathbf{u}_\delta| - |\mathbf{u}|) + \nabla \rho_\sigma(s)(|\mathbf{u}_\delta| - |\mathbf{u}|) \\ &\quad + \rho_\sigma(s_\delta)\nabla(|\mathbf{u}_\delta| - |\mathbf{u}|) + \nabla(\rho_\sigma(s_\delta) - \rho_\sigma(s))|\mathbf{u}| + (\rho_\sigma(s_\delta) - \rho_\sigma(s))\nabla|\mathbf{u}| \end{aligned}$$

and infer that  $\nabla(s_{\sigma,\delta} - s_\sigma) \rightarrow 0$  as  $\delta \rightarrow 0$  in  $L^2(\Omega)$  upon using again the Lebesgue dominated convergence theorem for the second and fifth terms, together with (3.4), (3.5), and  $|\mathbf{u}|, |\mathbf{u}_\delta|, |\rho_\sigma(s_\delta)| \leq 1$  for the other terms.

*Step 4: Convergence in  $H^1$  as  $\sigma \rightarrow 0$ .* It remains to prove

$$s_\sigma = \rho_\sigma(s)|s| \rightarrow s, \quad \mathbf{u}_\sigma = |\rho_\sigma(s)|\mathbf{u} \rightarrow \mathbf{u} \quad \text{in } H^1(\Omega).$$

To this end, we use again that  $\nabla|s| = \text{sgn}_0(s)\nabla s$  and write

$$\nabla(s_\sigma - s) = \nabla \rho_\sigma(s)|s| + (\rho_\sigma(s) \text{sgn}_0(s) - 1)\nabla s.$$

Since  $\nabla \rho_\sigma(s) = \sigma^{-1} \chi_{\{|s| < \sigma\}} \nabla s$ , we readily obtain as  $\sigma \rightarrow 0$ ,

$$\|\nabla \rho_\sigma(s)|s|\|_{L^2(\Omega)} \leq \|\nabla s\|_{L^2(\{|s| < \sigma\})} \rightarrow 0.$$

On the other hand,  $\rho_\sigma(t) \rightarrow \text{sgn}_0(t)$  for all  $t \in \mathbb{R}$ , whence

$$\|(\rho_\sigma(s) \text{sgn}_0(s) - 1)\nabla s\|_{L^2(\Omega)} \rightarrow \|\chi_{\{s=0\}}\nabla s\|_{L^2(\Omega)} = 0 \quad \text{as } \sigma \rightarrow 0$$

because  $\chi_{\{s=0\}}\nabla s = 0$  a.e. in  $\Omega$  [27, Chap. 5, Exercise 17]. Recalling that  $|\mathbf{u}| = |s|$  a.e. in  $\Omega$ , and thus  $\chi_{\{|\mathbf{u}|=0\}} = \chi_{\{s=0\}}$ , a similar argument shows that  $\mathbf{u}_\sigma \rightarrow \mathbf{u}$ .

*Step 5: Choice of  $\sigma$  and  $\delta$ .* Given  $\epsilon > 0$ , we first use Step 4 to choose  $\sigma$  such that

$$\|(s, \mathbf{u}) - (s_\sigma, \mathbf{u}_\sigma)\|_{H^1(\Omega)} \leq \frac{\epsilon}{2}.$$

We finally resort to Step 3 to select  $\delta < \sigma$ , depending on  $\sigma$ , such that

$$\|(s_\sigma, \mathbf{u}_\sigma) - (s_{\delta,\sigma}, \mathbf{u}_{\delta,\sigma})\|_{H^1(\Omega)} \leq \frac{\epsilon}{2}.$$

Therefore, we obtain the desired regularized pair, i.e.,  $s_\epsilon := s_{\delta,\sigma}$ ,  $\mathbf{u}_\epsilon := \mathbf{u}_{\delta,\sigma}$  satisfies  $(s_\epsilon, \mathbf{u}_\epsilon) \in \mathbb{A}(g, \mathbf{r}) \cap [W_\infty^1(\Omega)]^{d+1}$  along with (3.2) and (3.3). The proof is complete.  $\square$

We now fix  $\epsilon > 0$  and let  $(s_{\epsilon,h}, \mathbf{u}_{\epsilon,h}) \in \mathbb{X}_h$  be the Lagrange interpolants of  $(s_\epsilon, \mathbf{u}_\epsilon) \in \mathbb{A}(g, \mathbf{r})$  given in Proposition 3.2, which are well defined because  $(s_\epsilon, \mathbf{u}_\epsilon) \in [W_\infty^1(\Omega)]^{d+1}$  and satisfy  $(s_{\epsilon,h}, \mathbf{u}_{\epsilon,h}) = (g_h, \mathbf{r}_h)$  on  $\partial\Omega$ . For any node  $x_i$ , we set

$$\mathbf{n}_{\epsilon,h}(x_i) = \begin{cases} \mathbf{u}_{\epsilon,h}(x_i)/s_{\epsilon,h}(x_i) & \text{if } s_{\epsilon,h}(x_i) \neq 0, \\ \text{any unit vector} & \text{otherwise} \end{cases}$$

and observe that (2.12) holds, whence  $(s_{\epsilon,h}, \mathbf{u}_{\epsilon,h}) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$ . In view of the energy identity (2.23) and the property  $\|(s_{\epsilon,h}, \mathbf{u}_{\epsilon,h}) - (s_\epsilon, \mathbf{u}_\epsilon)\|_{H^1(\Omega)} \rightarrow 0$  as  $h \rightarrow 0$ , to show (3.1) it suffices to prove that the consistency term satisfies

$$(3.6) \quad C_1^h[s_{\epsilon,h}, \mathbf{n}_{\epsilon,h}] := \sum_{i,j=1}^N k_{ij}(\delta_{ij}s_{\epsilon,h})^2 |\delta_{ij}\mathbf{n}_{\epsilon,h}|^2 \rightarrow 0 \quad \text{as } h \rightarrow 0.$$

Heuristically, if  $\mathbf{n}_\epsilon = \mathbf{u}_\epsilon/s_\epsilon$  is in  $W_\infty^1(\Omega)$ , then the sum (3.6) would be of order  $h^2 \int_\Omega |\nabla s_{\epsilon,h}|^2 dx$ , which obviously converges to zero. However, such an argument fails if the director field  $\mathbf{n}_\epsilon$  lacks high regularity, which is the case with defects. Since  $\mathbf{n}_\epsilon$  is not regular in general when  $s_\epsilon$  vanishes, the proof of consistency requires a separate treatment of the region where  $\mathbf{n}_\epsilon$  is regular and the region where  $\mathbf{n}_\epsilon$  is singular. The heuristic argument carries over in the regular region, while in the singular region we appeal to basic measure theory. With this motivation in mind, we now prove the following lemma.

LEMMA 3.3 (lim-sup inequality). *Let  $(s_\epsilon, \mathbf{u}_\epsilon) \in \mathbb{A}(g, \mathbf{r}) \cap [W_\infty^1(\Omega)]^{d+1}$  be the functions constructed in Proposition 3.2, for any  $\epsilon > 0$ , and let  $(s_{\epsilon,h}, \mathbf{u}_{\epsilon,h}) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$  be their Lagrange interpolants. Then*

$$E_1[s_\epsilon, \mathbf{n}_\epsilon] = \lim_{h \rightarrow 0} E_1^h[s_{\epsilon,h}, \mathbf{n}_{\epsilon,h}] = \lim_{h \rightarrow 0} \tilde{E}_1^h[s_{\epsilon,h}, \mathbf{u}_{\epsilon,h}] = \tilde{E}_1[s_\epsilon, \mathbf{u}_\epsilon].$$

*Proof.* Since  $\epsilon$  is fixed, we simplify the notation and write  $(s_h, \mathbf{n}_h)$  instead of  $(s_{\epsilon,h}, \mathbf{n}_{\epsilon,h})$ . In order to prove that  $C_1^h[s_h, \mathbf{n}_h] \rightarrow 0$  in (3.6), we choose an arbitrary  $\delta > 0$  and divide the domain  $\Omega$  into two disjoint regions,

$$\mathcal{S}_\delta := \{x \in \Omega : |s_\epsilon(x)| < \delta\}, \quad \mathcal{K}_\delta := \bar{\Omega} \setminus \mathcal{S}_\delta,$$

and split  $C_1^h[s_h, \mathbf{n}_h]$  into two parts,

$$\mathcal{I}_h(\mathcal{K}_\delta) := \sum_{x_i, x_j \in \mathcal{K}_\delta} k_{ij}(\delta_{ij}s_h)^2 |\delta_{ij}\mathbf{n}_h|^2, \quad \mathcal{I}_h(\mathcal{S}_\delta) := \sum_{x_i \text{ or } x_j \in \mathcal{S}_\delta} k_{ij}(\delta_{ij}s_h)^2 |\delta_{ij}\mathbf{n}_h|^2.$$

*Step 1: Estimate on  $\mathcal{K}_\delta$ .* Since both  $s_\epsilon$  and  $\mathbf{u}_\epsilon$  are Lipschitz in  $\bar{\Omega}$ , the set  $\mathcal{K}_\delta$  is a compact set and the field  $\mathbf{n}_\epsilon = s_\epsilon^{-1}\mathbf{u}_\epsilon$  is also Lipschitz in  $\mathcal{K}_\delta$  with a constant that depends on  $\epsilon$  and  $\delta$ . Therefore,  $|\delta_{ij}\mathbf{n}_h| = |\mathbf{n}_h(x_i) - \mathbf{n}_h(x_j)| \leq C_{\epsilon,\delta}h$ , because  $x_i$  and  $x_j$  are connected by a single edge of the mesh, whence

$$\mathcal{I}_h(\mathcal{K}_\delta) \leq C_{\epsilon,\delta}h^2 \sum_{i,j=1}^N k_{ij}(\delta_{ij}s_h)^2 \rightarrow 0 \quad \text{as } h \rightarrow 0,$$

because  $\frac{1}{2} \sum_{i,j=1}^N k_{ij}(\delta_{ij}s_h)^2 = \|\nabla s_h\|_{L^2(\Omega)}^2 \leq C\|\nabla s_\epsilon\|_{L^2(\Omega)}^2 < \infty$ .

*Step 2: Estimate on  $\mathcal{S}_\delta$ .* If either  $x_i$  or  $x_j$  is in  $\mathcal{S}_\delta$ , without loss of generality, we assume that  $x_i \in \mathcal{S}_\delta$ . Since  $s_\epsilon$  is Lipschitz, and  $s_h = I_h s_\epsilon$  is the Lagrange interpolant

of  $s_\epsilon$ , there is a mesh size  $h$  such that for any  $x$  in the star  $\omega_i$  of  $x_i$ ,  $|s_h(x) - s_h(x_i)| \leq C_\epsilon h \leq \delta$ , which implies that  $\omega_i \subset \mathcal{S}_{2\delta}$ . Since  $|\delta_{ij}\mathbf{n}_h| \leq 2$ , we get

$$\mathcal{I}_h(\mathcal{S}_\delta) \leq 4 \sum_{x_i \text{ or } x_j \in \mathcal{S}_\delta} k_{ij}(\delta_{ij}s_h)^2 \leq 8 \int_{\cup \omega_i} |\nabla s_h|^2 dx \leq 8 \int_{\mathcal{S}_{2\delta}} |\nabla s_h|^2 dx,$$

where the union  $\cup \omega_i$  is taken over all nodes  $x_i$  in  $\mathcal{S}_\delta$ . If  $d < p < \infty$ , we infer that

$$\int_{\mathcal{S}_{2\delta}} |\nabla s_h|^2 dx \leq C \left( \int_{\mathcal{S}_{2\delta}} |\nabla I_h s_\epsilon|^p dx \right)^{\frac{2}{p}} \rightarrow C \left( \int_{\mathcal{S}_{2\delta}} |\nabla s_\epsilon|^p dx \right)^{\frac{2}{p}} \quad \text{as } h \rightarrow 0,$$

in view of the stability of the Lagrange interpolation operator  $I_h$  in  $W_p^1$  for  $p > d$ .

*Step 3: The limit  $\delta \rightarrow 0$ .* Combining Steps 1 and 2 gives, for all  $\delta > 0$ ,

$$\lim_{h \rightarrow 0} \sum_{i,j=1}^N k_{ij}(\delta_{ij}s_h)^2 |\delta_{ij}\mathbf{n}_h|^2 \leq C \left( \int_{\mathcal{S}_{2\delta}} |\nabla s_\epsilon|^p dx \right)^{\frac{2}{p}} = C \left( \int_{\Omega} |\nabla s_\epsilon|^p \chi_{\{|s_\epsilon| \leq 2\delta\}} dx \right)^{\frac{2}{p}},$$

where  $\chi_A$  is the characteristic function of the set  $A$ . By virtue of the Lebesgue dominated convergence theorem, we obtain

$$\lim_{\delta \rightarrow 0} \int_{\Omega} |\nabla s_\epsilon|^p \chi_{\{|s_\epsilon| \leq 2\delta\}} dx = \int_{\Omega} |\nabla s_\epsilon|^p \chi_{\{s_\epsilon=0\}} dx = 0,$$

because  $\nabla s_\epsilon(x) = \mathbf{0}$  for a.e.  $x \in \{s_\epsilon = 0\}$  [27, Chap. 5, Exercise 17]. This proves the lemma.  $\square$

**3.2. Weak lower semicontinuity or lim-inf property.** This property usually follows from convexity. While it is obvious that the discrete energy  $\tilde{E}_1^h[s_h, \mathbf{u}_h]$  in (2.20) is convex with respect to  $\nabla \mathbf{u}_h$  and  $\nabla s_h$  if  $\kappa \geq 1$ , the convexity is not clear if  $0 < \kappa < 1$ . It is worth mentioning that if  $\kappa < 1$ , the convexity of the continuous energy (2.3) is based on the fact that  $|\mathbf{u}| = |s|$  a.e. in  $\Omega$ , and hence the convex part  $\int_{\Omega} |\nabla \mathbf{u}|^2 dx$  controls the concave part  $(\kappa - 1) \int_{\Omega} |\nabla s|^2 dx$  [34]. However, for the discrete energy (2.20), the equality  $|\mathbf{u}_h| = |s_h|$  holds only at the vertices. Therefore, it is not obvious how to establish the weak lower semicontinuity of  $\tilde{E}_1^h[s_h, \mathbf{u}_h]$ . This is why we exploit the nodal relations  $\tilde{s}_h = |s_h| = |\mathbf{u}_h| = |\tilde{\mathbf{u}}_h|$  to derive an alternative formula for  $\tilde{E}_1^h[\tilde{s}_h, \tilde{\mathbf{u}}_h]$ . Our next lemma hinges on (2.21) and makes the convexity of  $\tilde{E}_1^h[I_h|s_h|, \tilde{\mathbf{u}}_h]$  with respect to  $\nabla \tilde{\mathbf{u}}_h$  completely explicit.

LEMMA 3.4 (weak lower semicontinuity). *The energy  $\int_{\Omega} L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) dx$ , with*

$$L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) := (\kappa - 1) |\nabla I_h |\mathbf{w}_h||^2 + |\nabla \mathbf{w}_h|^2,$$

*is well defined for any  $\mathbf{w}_h \in \mathbb{U}_h$  and is weakly lower semicontinuous in  $H^1(\Omega)$ , i.e., for any weakly convergent sequence  $\mathbf{w}_h \rightharpoonup \mathbf{w}$  in the  $H^1(\Omega)$  norm, we have*

$$\liminf_{h \rightarrow 0} \int_{\Omega} L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) dx \geq \int_{\Omega} (\kappa - 1) |\nabla |\mathbf{w}||^2 + |\nabla \mathbf{w}|^2 dx.$$

*Proof.* If  $\kappa \geq 1$ , then the assertion follows from standard arguments. Here, we only dwell upon  $0 < \kappa < 1$  and dimension  $d = 2$ , because the case  $d = 3$  is similar. After extracting a subsequence (not relabeled), we can assume that  $\mathbf{w}_h$  converges to  $\mathbf{w}$  strongly in  $L^2(\Omega)$  and pointwise a.e. in  $\Omega$ .

*Step 1: Equivalent form of  $L_h$ .* We let  $T$  be any triangle in the mesh  $\mathcal{T}_h$ , label its three vertices as  $x_0, x_1, x_2$ , and define  $\mathbf{e}_1 := x_1 - x_0$  and  $\mathbf{e}_2 := x_2 - x_0$ . After denoting  $\mathbf{w}_h^i = \mathbf{w}_h(x_i)$  for  $i = 0, 1, 2$ , a simple calculation yields

$$\begin{aligned} \nabla \mathbf{w}_h &= (\mathbf{w}_h^1 - \mathbf{w}_h^0) \otimes \mathbf{e}_1^* + (\mathbf{w}_h^2 - \mathbf{w}_h^0) \otimes \mathbf{e}_2^*, \\ \nabla I_h |\mathbf{w}_h| &= (|\mathbf{w}_h^1| - |\mathbf{w}_h^0|) \mathbf{e}_1^* + (|\mathbf{w}_h^2| - |\mathbf{w}_h^0|) \mathbf{e}_2^*, \end{aligned}$$

where  $\{\mathbf{e}_i^*\}_{i=1}^2$  is the dual basis of  $\{\mathbf{e}_i\}_{i=1}^2$ , that is,  $\mathbf{e}_i^* \cdot \mathbf{e}_j = I_{ij}$ , and  $I = (I_{ij})_{i,j=1}^2$  is the identity matrix. Assuming  $|\mathbf{w}_h^i| + |\mathbf{w}_h^0| \neq 0$ , we realize that

$$|\mathbf{w}_h^i| - |\mathbf{w}_h^0| = \frac{\mathbf{w}_h^i + \mathbf{w}_h^0}{|\mathbf{w}_h^i| + |\mathbf{w}_h^0|} \cdot (\mathbf{w}_h^i - \mathbf{w}_h^0).$$

We then obtain  $\nabla I_h |\mathbf{w}_h| = G_h(\mathbf{w}_h) : \nabla \mathbf{w}_h$ , where  $G_h(\mathbf{w}_h)$  is the 3-tensor,

$$G_h(\mathbf{w}_h) := \frac{\mathbf{w}_h^1 + \mathbf{w}_h^0}{|\mathbf{w}_h^1| + |\mathbf{w}_h^0|} \otimes \mathbf{e}_1 \otimes \mathbf{e}_1^* + \frac{\mathbf{w}_h^2 + \mathbf{w}_h^0}{|\mathbf{w}_h^2| + |\mathbf{w}_h^0|} \otimes \mathbf{e}_2 \otimes \mathbf{e}_2^*, \quad \text{on } T,$$

and the contraction between a 3-tensor and a 2-tensor in dyadic form is given by

$$(\mathbf{g}_1 \otimes \mathbf{g}_2 \otimes \mathbf{g}_3) : (\mathbf{m}_1 \otimes \mathbf{m}_2) := (\mathbf{g}_1 \cdot \mathbf{m}_1)(\mathbf{g}_2 \cdot \mathbf{m}_2)\mathbf{g}_3.$$

Therefore, we have

$$L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) = |\nabla \mathbf{w}_h|^2 + (\kappa - 1) |G_h(\mathbf{w}_h) : \nabla \mathbf{w}_h|^2,$$

which expresses  $L_h(\mathbf{w}_h, \nabla \mathbf{w}_h)$  directly in terms of  $\nabla \mathbf{w}_h$  and the nodal values of  $\mathbf{w}_h$ .

*Step 2: Convergence of  $G_h(\mathbf{w}_h)$ .* Given  $\epsilon > 0$ , Egoroff's theorem [50] asserts that

$$\mathbf{w}_h \rightarrow \mathbf{w} \quad \text{uniformly on } E_\epsilon,$$

for some subset  $E_\epsilon$  and  $|\Omega \setminus E_\epsilon| \leq \epsilon$ . We now consider the set  $A_\epsilon := \{|\mathbf{w}(x)| \geq 2\epsilon\} \cap E_\epsilon$  and observe that there exists a sufficiently small  $h_\epsilon$  such that for any  $x \in A_\epsilon$ ,

$$|\mathbf{w}_h(x)| \geq \epsilon \quad \text{for all } h \leq h_\epsilon.$$

If  $G(\mathbf{w}) := \frac{\mathbf{w}}{|\mathbf{w}|} \otimes I$ , then we claim that

$$(3.7) \quad \int_{A_\epsilon} |G_h(\mathbf{w}_h) - G(\mathbf{w})|^2 dx \rightarrow 0 \quad \text{as } h \rightarrow 0.$$

For any  $x \in A_\epsilon$ , let  $\{T_h\}$  be a sequence of triangles such that  $x \in \overline{T_h}$ . Since  $|\mathbf{w}_h(x)| \geq \epsilon$  and  $\mathbf{w}_h$  is piecewise linear, there exists a vertex of  $T_h$ , which we label as  $x_h^0$ , such that  $|\mathbf{w}_h^0| \geq \epsilon$ . To compare  $G_h(\mathbf{w}_h)$  with  $\frac{\mathbf{w}_h(x)}{|\mathbf{w}_h(x)|} \otimes I$ , we use that  $I = \mathbf{e}_1 \otimes \mathbf{e}_1^* + \mathbf{e}_2 \otimes \mathbf{e}_2^*$ :

$$G_h(\mathbf{w}_h) - \frac{\mathbf{w}_h(x)}{|\mathbf{w}_h(x)|} \otimes I = \sum_{i=1,2} \left( \frac{\mathbf{w}_h^i + \mathbf{w}_h^0}{|\mathbf{w}_h^i| + |\mathbf{w}_h^0|} - \frac{\mathbf{w}_h(x)}{|\mathbf{w}_h(x)|} \right) \otimes \mathbf{e}_i \otimes \mathbf{e}_i^*.$$

We define  $H(\mathbf{x}, \mathbf{y}) := \frac{\mathbf{x} + \mathbf{y}}{|\mathbf{x}| + |\mathbf{y}|}$  and observe that for all  $x \in A_\epsilon$ , we have

$$G_h(\mathbf{w}_h) - \frac{\mathbf{w}_h(x)}{|\mathbf{w}_h(x)|} \otimes I = \sum_{i=1,2} (H(\mathbf{w}_h^0, \mathbf{w}_h^i) - H(\mathbf{w}_h(x), \mathbf{w}_h(x))) \otimes \mathbf{e}_i \otimes \mathbf{e}_i^*.$$

Next, we estimate

$$\begin{aligned} |H(\mathbf{w}_h^0, \mathbf{w}_h^i) - H(\mathbf{w}_h(x), \mathbf{w}_h(x))| &= \left| \frac{|\mathbf{w}_h(x)|(|\mathbf{w}_h^0| + |\mathbf{w}_h^i|) - (|\mathbf{w}_h^0| + |\mathbf{w}_h^i|)|\mathbf{w}_h(x)|}{(|\mathbf{w}_h^0| + |\mathbf{w}_h^i|)|\mathbf{w}_h(x)} \right| \\ &\leq \left| \frac{|\mathbf{w}_h^0| + |\mathbf{w}_h^i| - 2|\mathbf{w}_h(x)|}{|\mathbf{w}_h^0| + |\mathbf{w}_h^i|} \right| + \left| \frac{(|\mathbf{w}_h(x)| - |\mathbf{w}_h^0|)|\mathbf{w}_h(x)}{(|\mathbf{w}_h^0| + |\mathbf{w}_h^i|)|\mathbf{w}_h(x)} \right| + \left| \frac{(|\mathbf{w}_h(x)| - |\mathbf{w}_h^i|)|\mathbf{w}_h(x)}{(|\mathbf{w}_h^0| + |\mathbf{w}_h^i|)|\mathbf{w}_h(x)} \right|. \end{aligned}$$

Since  $|\mathbf{w}_h^0|, |\mathbf{w}_h(x)| \geq \epsilon$ , and  $\mathbf{w}_h(x) - \mathbf{w}_h(x_h^i) = \nabla \mathbf{w}_h \cdot (x - x_h^i)$  for all  $x \in \overline{T_h}$ , we have

$$|H(\mathbf{w}_h^0, \mathbf{w}_h^i) - H(\mathbf{w}_h(x), \mathbf{w}_h(x))| \leq C \frac{h}{\epsilon} |\nabla \mathbf{w}_h| \quad \text{for all } x \in A_\epsilon \cap \overline{T_h}.$$

Integrating on  $A_\epsilon$ , we obtain

$$\int_{A_\epsilon} \left| G_h(\mathbf{w}_h) - \frac{\mathbf{w}_h(x)}{|\mathbf{w}_h(x)|} \otimes I \right|^2 dx \leq C \frac{h^2}{\epsilon^2} \int_{A_\epsilon} |\nabla \mathbf{w}_h(x)|^2 dx \rightarrow 0 \quad \text{as } h \rightarrow 0.$$

Since  $\mathbf{w}_h \rightarrow \mathbf{w}$  a.e. in  $\Omega$  and  $\frac{\mathbf{w}_h}{|\mathbf{w}_h|} - \frac{\mathbf{w}}{|\mathbf{w}|}$  is bounded, applying the dominated convergence theorem, we infer that

$$\int_{A_\epsilon} \left| \frac{\mathbf{w}_h}{|\mathbf{w}_h|} - \frac{\mathbf{w}}{|\mathbf{w}|} \right|^2 \rightarrow 0 \quad \text{as } h \rightarrow 0.$$

Combining these two limits, we deduce (3.7).

*Step 3: Convexity.* We now prove that the energy density

$$L(\mathbf{w}, M) := |M|^2 + (\kappa - 1)|G(\mathbf{w}) : M|^2$$

is convex with respect to any matrix  $M$  for any vector  $\mathbf{w}$ ; hereafter,  $G(\mathbf{w}) = \mathbf{z} \otimes I$  with  $\mathbf{z} = \frac{\mathbf{w}}{|\mathbf{w}|}$  provided  $\mathbf{w} \neq \mathbf{0}$ , or  $|\mathbf{z}| \leq 1$  otherwise. Note that  $L(\mathbf{w}, M)$  is a quadratic function of  $M$ , so we need only show that  $L(\mathbf{w}, M) \geq 0$  for any  $M$  and  $\mathbf{w}$ . Thus, it suffices to show that  $|G(\mathbf{w}) : M| \leq |M|$ .

Assume that  $M = \sum_{i,j} m_{ij} \mathbf{v}_i \otimes \mathbf{v}_j$ , where  $\{\mathbf{v}_i\}_{i=1}^2$  is the canonical basis on  $\mathbb{R}^2$ . Then we have  $|M|^2 = \sum_{i,j=1}^2 m_{ij}^2$ , and a simple calculation yields

$$\begin{aligned} G(\mathbf{w}) : M &= \sum_i z_i \mathbf{v}_i \otimes (\mathbf{v}_1 \otimes \mathbf{v}_1 + \mathbf{v}_2 \otimes \mathbf{v}_2) : \left( \sum_{k,l} m_{kl} \mathbf{v}_k \otimes \mathbf{v}_l \right) \\ &= \sum_{i,k,l} z_i m_{kl} \delta_{ik} \mathbf{v}_l = \sum_{i,l} z_i m_{il} \mathbf{v}_l, \end{aligned}$$

where  $\mathbf{z} = \sum_{i=1}^2 z_i \mathbf{v}_i$ . Therefore, we obtain

$$|G(\mathbf{w}) : M|^2 = \sum_{j=1}^2 \left( \sum_{i=1}^2 z_i m_{ij} \right)^2.$$

The Cauchy–Schwarz inequality yields

$$\left( \sum_{i=1}^2 z_i m_{ij} \right)^2 \leq \left( \sum_{i=1}^2 z_i^2 \right) \left( \sum_{i=1}^2 m_{ij}^2 \right) \leq \left( \sum_{i=1}^2 m_{ij}^2 \right),$$

which implies  $|G(\mathbf{w}) : M|^2 \leq |M|^2$  and  $L(\mathbf{w}, M) \geq 0$  for any matrix  $M$  and vector  $\mathbf{w}$ . A similar argument shows that  $L_h(\mathbf{w}_h, M) \geq 0$  for any matrix  $M$  and vector  $\mathbf{w}_h$ .

*Step 4: Weak lower semicontinuity.* Since  $G_h(\mathbf{w}_h) \rightarrow G(\mathbf{w})$  in  $L^2(A_\epsilon)$  according to (3.7), Egoroff's theorem yields

$$G_h(\mathbf{w}_h) \rightarrow G(\mathbf{w}) \quad \text{uniformly on } B_\epsilon,$$

where  $B_\epsilon \subset A_\epsilon$  and  $|A_\epsilon \setminus B_\epsilon| \leq \epsilon$ . We claim that

$$(3.8) \quad \liminf_{h \rightarrow 0} \int_{\Omega} L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) dx \geq \int_{B_\epsilon} L(\mathbf{w}, \nabla \mathbf{w}) dx.$$

Step 3 implies  $L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) \geq 0$  for all  $x \in \Omega$ . Hence,

$$\int_{\Omega} L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) dx \geq \int_{B_\epsilon} \left( |\nabla \mathbf{w}_h|^2 + (\kappa - 1) |G_h(\mathbf{w}_h) : \nabla \mathbf{w}_h|^2 \right) dx.$$

A simple calculation yields

$$\int_{\Omega} L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) dx \geq \int_{B_\epsilon} L(\mathbf{w}, \nabla \mathbf{w}_h) dx + (\kappa - 1) Q_h(\mathbf{w}, \mathbf{w}_h),$$

where

$$Q_h(\mathbf{w}, \mathbf{w}_h) := \int_{B_\epsilon} \left( [(G_h(\mathbf{w}_h) - G(\mathbf{w})) : \nabla \mathbf{w}_h]^t [G_h(\mathbf{w}_h) : \nabla \mathbf{w}_h] + (G(\mathbf{w}) : \nabla \mathbf{w}_h)^t [(G_h(\mathbf{w}_h) - G(\mathbf{w})) : \nabla \mathbf{w}_h] \right) dx.$$

Since  $L(\mathbf{w}, \nabla \mathbf{w}_h)$  is convex with respect to  $\nabla \mathbf{w}_h$  (Step 3), we have [27, sect. 8.2.2, p. 446]

$$\liminf_{h \rightarrow 0} \int_{B_\epsilon} L(\mathbf{w}, \nabla \mathbf{w}_h) dx \geq \int_{B_\epsilon} L(\mathbf{w}, \nabla \mathbf{w}) dx.$$

To prove (3.8), it remains to show that  $Q_h(\mathbf{w}, \mathbf{w}_h) \rightarrow 0$  as  $h \rightarrow 0$ . Since  $G(\mathbf{w})$  and  $G_h(\mathbf{w}_h)$  are bounded and  $\int_{\Omega} |\nabla \mathbf{w}_h(x)|^2 dx$  is uniformly bounded, we have

$$\begin{aligned} Q_h(\mathbf{w}, \mathbf{w}_h) &\leq C \int_{B_\epsilon} |G_h(\mathbf{w}_h) - G(\mathbf{w})| |\nabla \mathbf{w}_h|^2 dx \\ &\leq C \max_{B_\epsilon} |G_h(\mathbf{w}_h) - G(\mathbf{w})| \int_{B_\epsilon} |\nabla \mathbf{w}_h|^2 dx \rightarrow 0 \quad \text{as } h \rightarrow 0, \end{aligned}$$

due to the uniform convergence of  $G_h(\mathbf{w}_h)$  to  $G(\mathbf{w})$  in  $B_\epsilon$ . Therefore, we infer that  $\liminf_{h \rightarrow 0} \int_{\Omega} L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) dx \geq \int_{B_\epsilon} L(\mathbf{w}, \nabla \mathbf{w}) dx$ .

Since the inequality above holds for arbitrarily small  $\epsilon$ , taking  $\epsilon \rightarrow 0$  yields

$$\liminf_{h \rightarrow 0} \int_{\Omega} L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) dx \geq \int_{\Omega \setminus \{\mathbf{w}(x)=0\}} L(\mathbf{w}, \nabla \mathbf{w}) dx = \int_{\Omega} L(\mathbf{w}, \nabla \mathbf{w}) dx,$$

where the last equality follows from  $\nabla \mathbf{w} = \mathbf{0}$  a.e. in the set  $\{\mathbf{w}(x) = \mathbf{0}\}$  [27, Chap. 5, Exercise 17, p. 292]. Finally, noting that  $G(\mathbf{w}) : \nabla \mathbf{w} = \nabla |\mathbf{w}|$ , we get the assertion.  $\square$

**3.3. Equi-coercivity.** We now prove uniform  $H^1$ -bounds for the pairs  $(s_h, \mathbf{u}_h)$  and  $(\tilde{s}_h, \tilde{\mathbf{u}}_h)$ , which enables us to extract convergence subsequences in  $L^2(\Omega)$  and pointwise a.e. in  $\Omega$ . We then characterize and relate the limits of such sequences.

LEMMA 3.5 (coercivity). *For any  $(s_h, \mathbf{u}_h) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$ , we have*

$$E_1^h[s_h, \mathbf{n}_h] \geq \min\{\kappa, 1\} \max \left\{ \int_{\Omega} |\nabla \mathbf{u}_h|^2 dx, \int_{\Omega} |\nabla s_h|^2 dx \right\}$$

as well as

$$E_1^h[s_h, \mathbf{n}_h] \geq \min\{\kappa, 1\} \max \left\{ \int_{\Omega} |\nabla \tilde{\mathbf{u}}_h|^2 dx, \int_{\Omega} |\nabla I_h |s_h||^2 dx \right\}.$$

*Proof.* Inequality (2.20) of Lemma 2.2 shows that

$$(3.9) \quad E_1^h[s_h, \mathbf{n}_h] \geq (\kappa - 1) \int_{\Omega} |\nabla s_h|^2 dx + \int_{\Omega} |\nabla \mathbf{u}_h|^2 dx.$$

If  $\kappa \geq 1$ , then  $E_1^h[s_h, \mathbf{n}_h]$  obviously controls the  $H^1$ -norm of  $\mathbf{u}_h$  with constant 1.

If  $0 < \kappa < 1$ , then combining (2.16) with (2.22) yields

$$\begin{aligned} E_1^h[s_h, \mathbf{n}_h] &\geq \frac{\kappa}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2 \\ &\quad + \frac{\kappa}{2} \sum_{i,j=1}^N k_{ij} \left( \frac{s_h(x_i)^2 + s_h(x_j)^2}{2} \right) |\delta_{ij} \mathbf{n}_h|^2 \geq \kappa \int_{\Omega} |\nabla \mathbf{u}_h|^2 dx, \end{aligned}$$

whence  $E_1^h[s_h, \mathbf{n}_h] \geq \min\{\kappa, 1\} \int_{\Omega} |\nabla \mathbf{u}_h|^2 dx$  as asserted. The same argument, but invoking (2.21) and (2.24), leads to a similar estimate for  $\int_{\Omega} |\nabla \tilde{\mathbf{u}}_h|^2 dx$ .

Finally, we note that (2.16) implies

$$E_1^h[s_h, \mathbf{n}_h] \geq \frac{\kappa}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2 = \kappa \int_{\Omega} |\nabla s_h|^2 dx.$$

Upon recalling  $\tilde{s}_h = I_h |s_h|$  and noting  $|\delta_{ij} s_h| \geq |\delta_{ij} \tilde{s}_h|$  and  $k_{ij} \geq 0$ , we deduce  $\|\nabla s_h\|_{L^2(\Omega)} \geq \|\nabla \tilde{s}_h\|_{L^2(\Omega)}$  and complete the proof.  $\square$

LEMMA 3.6 (characterizing limits). *Let  $\{\mathcal{T}_h\}$  satisfy (2.10), and let a sequence  $(s_h, \mathbf{u}_h) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$  satisfy*

$$(3.10) \quad E_1^h[s_h, \mathbf{n}_h] \leq \Lambda \quad \text{for all } h > 0,$$

with a constant  $\Lambda > 0$  independent of  $h$ . Then there exist subsequences (not relabeled)  $(s_h, \mathbf{u}_h) \in \mathbb{X}_h$  and  $(\tilde{s}_h, \tilde{\mathbf{u}}_h) \in \tilde{\mathbb{X}}_h$  weakly converging in  $[H^1(\Omega)]^{d+1}$  such that

- $(s_h, \mathbf{u}_h)$  converges to  $(s, \mathbf{u}) \in [H^1(\Omega)]^{d+1}$  in  $L^2(\Omega)$  and a.e. in  $\Omega$ ;
- $(\tilde{s}_h, \tilde{\mathbf{u}}_h)$  converges to  $(\tilde{s}, \tilde{\mathbf{u}}) \in [H^1(\Omega)]^{d+1}$  in  $L^2(\Omega)$  and a.e. in  $\Omega$ ;
- the limits satisfy  $\tilde{s} = |s| = |\mathbf{u}| = |\tilde{\mathbf{u}}|$  a.e. in  $\Omega$ ;
- there exists a director field  $\mathbf{n}$  defined in  $\Omega$  such that  $\mathbf{n}_h$  converges to  $\mathbf{n}$  in  $L^2(\Omega \setminus \mathcal{S})$  and a.e. in  $\Omega \setminus \mathcal{S}$  and  $\mathbf{u} = s\mathbf{n}$ ,  $\tilde{\mathbf{u}} = \tilde{s}\mathbf{n}$  a.e. in  $\Omega$ .

*Proof.* The sequences  $(s_h, \mathbf{u}_h)$  and  $(\tilde{s}_h, \tilde{\mathbf{u}}_h)$  are uniformly bounded in  $H^1(\Omega)$  according to Lemma 3.5 (coercivity). Therefore, since  $H^1(\Omega)$  is compactly embedded

in  $L^2(\Omega)$  [1], there exist subsequences (not relabeled) that converge in  $L^2(\Omega)$  and a.e. in  $\Omega$  to pairs  $(s, \mathbf{u}) \in [H^1(\Omega)]^{d+1}$  and  $(\tilde{s}, \tilde{\mathbf{u}}) \in [H^1(\Omega)]^{d+1}$ , respectively.

Since  $s_h \rightarrow s$  and  $\tilde{s}_h \rightarrow \tilde{s}$  as  $h \rightarrow 0$ , invoking the triangle inequality yields

$$|\tilde{s} - |s|| \leq |\tilde{s} - \tilde{s}_h| + |\tilde{s}_h - |s_h|| + ||s_h| - |s|| \rightarrow 0 \quad \text{as } h \rightarrow 0,$$

which is a consequence of interpolation theory and (2.25), namely,

$$\|\tilde{s}_h - |s_h|\|_{L^2(\Omega)} = \|I_h|s_h| - |s_h|\|_{L^2(\Omega)} \leq Ch\|\nabla|s_h|\|_{L^2(\Omega)} \leq Ch\|\nabla s_h\|_{L^2(\Omega)} \leq C\Lambda h.$$

A similar argument shows

$$\|\tilde{\mathbf{u}}_h - I_h\tilde{\mathbf{u}}_h\|_{L^2(\Omega)} \leq Ch\|\nabla\tilde{\mathbf{u}}_h\|_{L^2(\Omega)} \leq Ch\|\nabla\tilde{\mathbf{u}}\|_{L^2(\Omega)} \leq C\Lambda h.$$

Since  $I_h|\tilde{\mathbf{u}}_h| = \tilde{s}_h \rightarrow \tilde{s}$  and  $|\tilde{\mathbf{u}}_h| \rightarrow |\tilde{\mathbf{u}}|$  as  $h \rightarrow 0$ , we deduce  $|\tilde{\mathbf{u}}| = \tilde{s}$  a.e. in  $\Omega$ . Likewise, arguing instead with the pair  $(s_h, \mathbf{u}_h)$ , we infer that  $|\mathbf{u}| = |s|$  a.e. in  $\Omega$ .

We now define the limiting director field  $\mathbf{n}$  in  $\Omega \setminus \mathcal{S}$  to be  $\mathbf{n} = s^{-1}\mathbf{u}$  and see that  $|\mathbf{n}| = 1$  a.e. in  $\Omega \setminus \mathcal{S}$ ; we define  $\mathbf{n}$  in  $\mathcal{S}$  to be an arbitrary unit vector. In order to relate  $\mathbf{n}$  with  $\mathbf{n}_h$ , we observe that both  $s_h$  and  $\mathbf{n}_h$  are piecewise linear. Applying the classical interpolation theory on each element  $T$  of  $\mathcal{T}_h$ , we obtain  $\|s_h\mathbf{n}_h - I_h[s_h\mathbf{n}_h]\|_{L^1(T)} \leq Ch^2\|\nabla s_h \otimes \nabla \mathbf{n}_h\|_{L^1(T)}$ . Summing over all  $T \in \mathcal{T}_h$ , we get

$$\|s_h\mathbf{n}_h - I_h[s_h\mathbf{n}_h]\|_{L^1(\Omega)} \leq Ch^2\|\nabla s_h \otimes \nabla \mathbf{n}_h\|_{L^1(\Omega)} \leq Ch^2\|\nabla s_h\|_{L^2(\Omega)}\|\nabla \mathbf{n}_h\|_{L^2(\Omega)}.$$

An inverse estimate gives  $\|\nabla \mathbf{n}_h\|_{L^2(\Omega)} \leq Ch^{-1}$ , because  $|\mathbf{n}_h| \leq 1$ . Hence,

$$\|s_h\mathbf{n}_h - I_h[s_h\mathbf{n}_h]\|_{L^1(\Omega)} \leq C\Lambda h \rightarrow 0 \quad \text{as } h \rightarrow 0.$$

Since  $\mathbf{u}_h = I_h[s_h\mathbf{n}_h] \rightarrow \mathbf{u}$  as  $h \rightarrow 0$ , we discover that also  $s_h\mathbf{n}_h \rightarrow \mathbf{u}$  a.e. in  $\Omega$  as  $h \rightarrow 0$ . Consequently, for a.e.  $x \in \Omega \setminus \mathcal{S}$  we have  $s_h(x) \rightarrow s(x) \neq 0$ , whence  $s_h(x)^{-1}$  is well defined for  $h$  small and

$$\mathbf{n}_h(x) = \frac{s_h(x)\mathbf{n}_h(x)}{s_h(x)} \rightarrow \frac{\mathbf{u}(x)}{s(x)} = \mathbf{n}(x) \quad \text{as } h \rightarrow 0.$$

Since  $|\mathbf{n}_h| \leq 1$ , the Lebesgue dominated convergence theorem yields

$$\|\mathbf{n}_h\chi_{\Omega \setminus \mathcal{S}} - \mathbf{n}\chi_{\Omega \setminus \mathcal{S}}\|_{L^2(\Omega)} \rightarrow 0 \quad \text{as } h \rightarrow 0.$$

It remains only to prove  $\tilde{\mathbf{u}} = \tilde{s}\mathbf{n}$  a.e. in  $\Omega$ . The same argument employed above gives

$$\|\tilde{s}_h\mathbf{n}_h - I_h[\tilde{s}_h\mathbf{n}_h]\|_{L^1(\Omega)} \leq C\Lambda h \rightarrow 0 \quad \text{as } h \rightarrow 0,$$

whence  $\tilde{s}_h\mathbf{n}_h \rightarrow \tilde{\mathbf{u}}$ . This implies that  $\tilde{s}_h(x)^{-1}$  is well defined for a.e.  $x \in \Omega \setminus \mathcal{S}$  and

$$\mathbf{n}_h(x) = \frac{\tilde{s}_h(x)\mathbf{n}_h(x)}{\tilde{s}_h(x)} \rightarrow \frac{\tilde{\mathbf{u}}(x)}{\tilde{s}(x)} = \mathbf{n}(x) \quad \text{as } h \rightarrow 0.$$

This completes the proof. □

**3.4.  $\Gamma$ -convergence.** We are now in the position to prove the main result, namely the convergence of global discrete minimizers. The proof is a minor variation of the standard one [18, 24].

**THEOREM 3.7** (convergence of global discrete minimizers). *Let  $\{\mathcal{T}_h\}$  satisfy (2.10). If  $(s_h, \mathbf{u}_h) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$  is a sequence of global minimizers of  $E_h[s_h, \mathbf{n}_h]$  in (2.18), then every cluster point is a global minimizer of the continuous energy  $E[s, \mathbf{n}]$  in (2.1).*

*Proof.* In view of (2.18), assume there is a constant  $\Lambda > 0$  such that

$$\liminf_{h \rightarrow 0} E_h[s_h, \mathbf{n}_h] = \liminf_{h \rightarrow 0} (E_1^h[s_h, \mathbf{n}_h] + E_2^h[s_h]) \leq \Lambda;$$

otherwise, there is nothing to prove. Applying Lemma 3.6 yields subsequences (not relabeled)  $(\tilde{s}_h, \tilde{\mathbf{u}}_h) \rightarrow (\tilde{s}, \tilde{\mathbf{u}})$  and  $(s_h, \mathbf{u}_h) \rightarrow (s, \mathbf{u})$  converging weakly in  $[H^1(\Omega)]^{d+1}$ , strongly in  $[L^2(\Omega)]^{d+1}$ , and a.e. in  $\Omega$ . Using Lemma 3.4, we deduce

$$\tilde{E}_1[\tilde{s}, \tilde{\mathbf{u}}] = \int_{\Omega} (\kappa - 1)|\nabla \tilde{s}|^2 + |\nabla \tilde{\mathbf{u}}|^2 dx \leq \liminf_{h \rightarrow 0} \tilde{E}_1^h[\tilde{s}_h, \tilde{\mathbf{u}}_h] \leq \liminf_{h \rightarrow 0} E_1^h[s_h, \mathbf{n}_h],$$

where the last inequality is a consequence of (2.21). Since  $s_h$  converges a.e. in  $\Omega$  to  $s$ , so does  $\psi(s_h)$  to  $\psi(s)$ . Now apply Fatou’s lemma to arrive at

$$E_2[s] = \int_{\Omega} \psi(s) = \int_{\Omega} \lim_{h \rightarrow 0} \psi(s_h) \leq \liminf_{h \rightarrow 0} \int_{\Omega} \psi(s_h) = \liminf_{h \rightarrow 0} E_2^h[s_h].$$

Consequently, we obtain

$$\tilde{E}_1[\tilde{s}, \tilde{\mathbf{u}}] + E_2[s] \leq \liminf_{h \rightarrow 0} E_h[s_h, \mathbf{n}_h] \leq \limsup_{h \rightarrow 0} E_h[s_h, \mathbf{n}_h].$$

Moreover, the triple  $(s, \mathbf{u}, \mathbf{n})$  given by Lemma 3.6 satisfies the structure property (2.5).

In view of Proposition 3.2, given  $\epsilon > 0$  arbitrary, we can always find a pair  $(t_\epsilon, \mathbf{v}_\epsilon) \in \mathbb{A}(g, \mathbf{r}) \cap [W_\infty^1(\Omega)]^{d+1}$  such that

$$\tilde{E}_1[t_\epsilon, \mathbf{v}_\epsilon] + E_2[t_\epsilon] = E_1[t_\epsilon, \mathbf{m}_\epsilon] + E_2[t_\epsilon] \leq \inf_{(t, \mathbf{m}) \in \mathbb{A}(g, \mathbf{r})} E[t, \mathbf{m}] + \epsilon \leq E[s, \mathbf{n}] + \epsilon,$$

where  $\mathbf{m}_\epsilon := t_\epsilon^{-1} \mathbf{v}_\epsilon$  if  $t_\epsilon \neq 0$ ; otherwise,  $\mathbf{m}_\epsilon$  is an arbitrary unit vector. Apply Lemma 3.3 to  $(t_\epsilon, \mathbf{v}_\epsilon)$  and  $\mathbf{m}_\epsilon$  to find  $(t_{\epsilon, h}, \mathbf{v}_{\epsilon, h}) \in \mathbb{A}_h(g_h, \mathbf{r}_h)$ ,  $\mathbf{m}_{\epsilon, h} \in \mathbb{N}_h$  such that

$$E_1[t_\epsilon, \mathbf{m}_\epsilon] = \lim_{h \rightarrow 0} E_1^h[t_{\epsilon, h}, \mathbf{m}_{\epsilon, h}].$$

On the other hand, (2.8) and (3.3) imply that  $0 \leq \psi(t_{\epsilon, h}) \leq \max\{\psi(-\frac{1}{2} + \delta_0), \psi(1 - \delta_0)\}$ , and we can invoke the Lebesgue dominated convergence theorem to infer that

$$E_2[t_\epsilon] = \int_{\Omega} \lim_{h \rightarrow 0} \psi(t_{\epsilon, h}) = \lim_{h \rightarrow 0} \int_{\Omega} \psi(t_{\epsilon, h}) = \lim_{h \rightarrow 0} E_2^h[t_{\epsilon, h}].$$

Therefore, collecting the preceding estimates, we arrive at

$$\tilde{E}_1[\tilde{s}, \tilde{\mathbf{u}}] + E_2[s] \leq \limsup_{h \rightarrow 0} E_h[s_h, \mathbf{n}_h] \leq \lim_{h \rightarrow 0} E_h[t_{\epsilon, h}, \mathbf{m}_{\epsilon, h}] \leq E[s, \mathbf{n}] + \epsilon.$$

We now prove that  $\tilde{E}_1[\tilde{s}, \tilde{\mathbf{u}}] = E_1[s, \mathbf{n}]$ . We exploit the relation  $\tilde{\mathbf{u}} = \tilde{s} \mathbf{n}$  a.e. in  $\Omega$  with  $|\mathbf{n}| = 1$ , together with the fact that  $\mathbf{n}$  admits a weak gradient in  $\Omega \setminus \mathcal{S}$ , to find the orthogonal decomposition  $\nabla \tilde{\mathbf{u}} = \nabla \tilde{s} \otimes \mathbf{n} + \tilde{s} \nabla \mathbf{n}$  a.e. in  $\Omega \setminus \mathcal{S}$ . Hence,

$$\begin{aligned} \tilde{E}_1[\tilde{s}, \tilde{\mathbf{u}}] &= \int_{\Omega \setminus \mathcal{S}} (\kappa - 1)|\nabla \tilde{s}|^2 + |\nabla \tilde{\mathbf{u}}|^2 dx = \int_{\Omega \setminus \mathcal{S}} \kappa |\nabla \tilde{s}|^2 + \tilde{s}^2 |\nabla \mathbf{n}|^2 dx \\ &= \int_{\Omega \setminus \mathcal{S}} \kappa |\nabla s|^2 + s^2 |\nabla \mathbf{n}|^2 dx \equiv E_1[s, \mathbf{n}], \end{aligned}$$

because  $\tilde{s} = |s|$  and  $\|\nabla|s|\|_{L^2(\Omega \setminus \mathcal{S})} = \|\nabla s\|_{L^2(\Omega \setminus \mathcal{S})}$ . Note that the singular set  $\mathcal{S}$  does not contribute, because  $\|\nabla s\|_{L^2(\mathcal{S})} = \|s\nabla \mathbf{n}\|_{L^2(\mathcal{S})} = 0$ . Finally, letting  $\epsilon \rightarrow 0$ , we see that the pair  $(s, \mathbf{n})$  is a global minimizer of  $E$  as asserted.  $\square$

If the global minimizer of the continuous energy  $E[s, \mathbf{n}]$  is unique, then Theorem 3.7 readily implies that the discrete energy minimizer  $(s_h, \mathbf{n}_h)$  converges to the unique minimizer of  $E[s, \mathbf{n}]$ . This theorem is about global minimizers only, both discrete and continuous. In the next section, we design a quasi-gradient flow to compute discrete local minimizers and show its convergence (see Theorem 4.2). In general, convergence to a global minimizer is not available, and neither are rates of convergence, due to the lack of continuous dependence results. However, if local minimizers of  $E[s, \mathbf{n}]$  are isolated, then there exist local minimizers of  $E_h[s_h, \mathbf{n}_h]$  that  $\Gamma$ -converge to  $(s, \mathbf{n})$  [18, 24].

**4. Quasi-gradient flow.** We consider a gradient flow methodology consisting of a gradient flow in  $s$  and a minimization in  $\mathbf{n}$  as a way to compute minimizers of (2.1) and (2.18). We begin with its description for the continuous system and verify that it has a monotone energy decreasing property. We then do the same for the discrete system.

**4.1. Continuous case.** We introduce the following subspace to enforce Dirichlet boundary conditions on open subsets  $\Gamma$  of  $\partial\Omega$ :

$$(4.1) \quad H^1_\Gamma(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma\}.$$

Let the sets  $\Gamma_s, \Gamma_{\mathbf{n}}$  satisfy  $\Gamma_{\mathbf{n}} = \Gamma_{\mathbf{u}} \subset \Gamma_s \subset \partial\Omega$ , and let (2.9) be valid on  $\Gamma_s$ . Therefore, the traces  $\mathbf{n} = \mathbf{q} := g^{-1}\mathbf{r}$  and  $\mathbf{n}_h = \mathbf{q}_h := I_h[g_h^{-1}\mathbf{r}_h]$  are well defined on  $\Gamma_{\mathbf{n}}$ .

**4.1.1. First order variation.** Consider the bulk energy  $E[s, \mathbf{n}]$ , where the pair  $(s, \mathbf{u})$ , with  $\mathbf{u} = s\mathbf{n}$ , is in the admissible class  $\mathbb{A}(g, \mathbf{r})$  defined in (2.4). We take a variation  $z \in H^1_0(\Omega)$  of  $s$  and obtain  $\delta_s E[s, \mathbf{n}; z] = \delta_s E_1[s, \mathbf{n}; z] + \delta_s E_2[s; z]$ , the first variation of  $E$  in the direction  $z$ , where

$$\delta_s E_1[s, \mathbf{n}; z] = 2 \int_{\Omega} (\nabla s \cdot \nabla z + |\nabla \mathbf{n}|^2 s z) \, dx \quad \text{and} \quad \delta_s E_2[s; z] = \int_{\Omega} \psi'(s) z \, dx.$$

Next, we introduce the space of tangential variations of  $\mathbf{n}$ :

$$(4.2) \quad V^\perp(\mathbf{n}) = \{\mathbf{v} \in H^1(\Omega)^d : \mathbf{v} \cdot \mathbf{n} = 0 \text{ a.e. in } \Omega\}.$$

In order to satisfy the constraint  $|\mathbf{n}| = 1$ , we take a variation  $\mathbf{v} \in V^\perp(\mathbf{n})$  of  $\mathbf{n}$  and get

$$\delta_{\mathbf{n}} E[s, \mathbf{n}; \mathbf{v}] = \delta_{\mathbf{n}} E_1[s, \mathbf{n}; \mathbf{v}] = 2 \int_{\Omega} s^2 (\nabla \mathbf{n} \cdot \nabla \mathbf{v}) \, dx.$$

Note that variations in  $V^\perp(\mathbf{n})$  preserve the unit length constraint up to second order accuracy [47]:  $|\mathbf{n} + t\mathbf{v}|^2 = 1 + t^2|\mathbf{v}|^2$  and  $|\mathbf{n} + t\mathbf{v}| \geq 1$  for all  $t \in \mathbb{R}$ .

**4.1.2. Quasi-gradient flow.** We consider an  $L^2$ -gradient flow for  $E$  with respect to the scalar variable  $s$ :

$$\int_{\Omega} \partial_t s z \, dx := -\delta_s E_1[s, \mathbf{n}; z] - \delta_s E_2[s; z] \quad \text{for all } z \in H^1_{\Gamma_s}(\Omega);$$

here, we enforce stationary Dirichlet boundary conditions for  $s$  on the set  $\Gamma_s \subset \partial\Omega$ , whence  $z = 0$  on  $\Gamma_s$ . A simple but formal integration by parts yields

$$\int_{\Omega} \partial_t s z \, dx = - \int_{\Omega} (-2\Delta s + 2|\nabla \mathbf{n}|^2 s + \psi'(s)) z \, dx \quad \text{for all } z \in H_{\Gamma_s}^1(\Omega),$$

where we use the implicit Neumann condition  $\boldsymbol{\nu} \cdot \nabla s = 0$  on  $\partial\Omega \setminus \Gamma_s$ , and  $\boldsymbol{\nu}$  is the outer unit normal on  $\partial\Omega$ . Therefore,  $s$  satisfies the (nonlinear) parabolic PDE:

$$(4.3) \quad \partial_t s = 2\Delta s - 2|\nabla \mathbf{n}|^2 s - \psi'(s).$$

Given  $s$  satisfying (2.9) on  $\Gamma_s$ , let  $\mathbf{n}$  satisfy  $|\mathbf{n}| = 1$  a.e. in  $\Omega$ , the stationary Dirichlet boundary condition  $\mathbf{n} = \mathbf{q}$  on the open set  $\Gamma_{\mathbf{n}} \subset \partial\Omega$ , and the following degenerate minimization problem:

$$E[s, \mathbf{n}] \leq E[s, \mathbf{m}] \quad \text{for all } |\mathbf{m}| = 1 \text{ a.e. } \Omega,$$

with the same boundary condition as  $\mathbf{n}$ . This implies

$$(4.4) \quad \delta_{\mathbf{n}} E[s, \mathbf{n}; \mathbf{v}] = 0 \quad \text{for all } \mathbf{v} \in V^{\perp}(\mathbf{n}) \cap H_{\Gamma_{\mathbf{n}}}^1(\Omega)^d.$$

**4.1.3. Formal energy decreasing property.** Differentiating the energy with respect to time, we obtain

$$\partial_t E[s, \mathbf{n}] = \delta_s E[s, \mathbf{n}; \partial_t s] + \delta_{\mathbf{n}} E[s, \mathbf{n}; \partial_t \mathbf{n}].$$

By virtue of (4.3) and (4.4), we deduce that

$$(4.5) \quad \partial_t E[s, \mathbf{n}] = -\delta_s E[s, \mathbf{n}; \partial_t s] = - \int_{\Omega} |\partial_t s|^2 \, dx.$$

Hence, the bulk energy  $E$  is monotonically decreasing for our quasi-gradient flow.

**4.2. Discrete case.** Let  $s_h^k \in \mathbb{S}_h(\Gamma_s, g_h)$  and  $\mathbf{n}_h^k \in \mathbb{N}_h(\Gamma_{\mathbf{n}}, \mathbf{q}_h)$  denote finite element functions with Dirichlet conditions  $s_h^k = g_h$  on  $\Gamma_s$  and  $\mathbf{n}_h^k = \mathbf{q}_h$  on  $\Gamma_{\mathbf{n}}$ , where  $k$  indicates a “time-step” index (see section 4.2.2 for the discrete gradient flow algorithm). To simplify notation, we use the following:

$$s_i^k := s_h^k(x_i), \quad \mathbf{n}_i^k := \mathbf{n}_h^k(x_i), \quad z_i := z_h(x_i), \quad \mathbf{v}_i := \mathbf{v}_h(x_i).$$

**4.2.1. First order variation.** First, we introduce the discrete version of (4.2),

$$(4.6) \quad V_h^{\perp}(\mathbf{n}_h) = \{\mathbf{v}_h \in \mathbb{U}_h : \mathbf{v}_h(x_i) \cdot \mathbf{n}_h(x_i) = 0 \text{ for all nodes } x_i \in \mathcal{N}_h\}.$$

Next, the first order variation of  $E_1^h$  in the direction  $\mathbf{v}_h \in V_h^{\perp}(\mathbf{n}_h^k) \cap H_{\Gamma_{\mathbf{n}}}^1(\Omega)$  at the director variable  $\mathbf{n}_h^k$  reads

$$(4.7) \quad \delta_{\mathbf{n}_h} E_1^h[s_h^k, \mathbf{n}_h^k; \mathbf{v}_h] = \sum_{i,j=1}^N k_{ij} \left( \frac{(s_i^k)^2 + (s_j^k)^2}{2} \right) (\delta_{ij} \mathbf{n}_h^k) \cdot (\delta_{ij} \mathbf{v}_h),$$

whereas the first order variation of  $E_1^h$  in the direction  $z_h \in \mathbb{S}_h \cap H_{\Gamma_s}^1(\Omega)$  at the degree of orientation variable  $s_h^k$  consists of two terms,

$$(4.8) \quad \delta_{s_h} E_1^h[s_h^k, \mathbf{n}_h^k; z_h] = \kappa \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h^k) (\delta_{ij} z_h) + \sum_{i,j=1}^N k_{ij} |\delta_{ij} \mathbf{n}_h^k|^2 \left( \frac{s_i^k z_i + s_j^k z_j}{2} \right).$$

To design an unconditionally stable scheme for the discrete gradient flow of  $E_2^h[s_h]$ , we employ the convex splitting technique in [51, 43, 44]. We split the double well potential into a convex and concave part: let  $\psi_c$  and  $\psi_e$  be both convex for all  $s \in (-1/2, 1)$  so that  $\psi(s) = \psi_c(s) - \psi_e(s)$ , and set

$$(4.9) \quad \delta_{s_h} E_2^h[s_h^{k+1}; z_h] := \int_{\Omega} [\psi'_c(s_h^{k+1}) - \psi'_e(s_h^k)] z_h dx.$$

LEMMA 4.1 (convex-concave splitting). *For any  $s_h^k$  and  $s_h^{k+1}$  in  $\mathbb{S}_h$ , we have*

$$\int_{\Omega} \psi(s_h^{k+1}) dx - \int_{\Omega} \psi(s_h^k) dx \leq \delta_{s_h} E_2^h[s_h^{k+1}; s_h^{k+1} - s_h^k].$$

*Proof.* A simple calculation, based on the mean-value theorem and the convex splitting  $\psi = \psi_c - \psi_e$ , yields

$$\int_{\Omega} (\psi(s_h^{k+1}) - \psi(s_h^k)) dx = \delta_{s_h} E_2^h[s_h^{k+1}; s_h^{k+1} - s_h^k] + T,$$

where

$$T = \int_{\Omega} \int_0^1 [\psi'_c(s_h^k + \theta(s_h^{k+1} - s_h^k)) - \psi'_c(s_h^{k+1})] (s_h^{k+1} - s_h^k) d\theta dx \\ + \int_{\Omega} \int_0^1 [\psi'_e(s_h^k) - \psi'_e(s_h^k + \theta(s_h^{k+1} - s_h^k))] (s_h^{k+1} - s_h^k) d\theta dx.$$

The convexity of both  $\psi_c$  and  $\psi_e$  implies  $T \leq 0$ , as desired. □

**4.2.2. Discrete quasi-gradient flow algorithm.** Our scheme for minimizing the discrete energy  $E_h[s_h, \mathbf{n}_h]$  is an alternating direction method, which minimizes with respect to  $\mathbf{n}_h$  and evolves  $s_h$  separately in the steepest descent direction during each iteration. Therefore, this algorithm is not a standard gradient flow.

**Algorithm** (discrete quasi-gradient flow): Given  $(s_h^0, \mathbf{n}_h^0)$  in  $\mathbb{S}_h(\Gamma_s, g_h) \times \mathbb{N}_h(\Gamma_{\mathbf{n}}, \mathbf{q}_h)$ , iterate Steps (a)–(c) for  $k \geq 0$ .

**Step (a): Minimization.** Find  $\mathbf{t}_h^k \in V_h^{\perp}(\mathbf{n}_h^k) \cap H_{\Gamma_{\mathbf{n}}}^1(\Omega)$  such that  $\mathbf{n}_h^k + \mathbf{t}_h^k$  minimizes the energy  $E_1^h[s_h^k, \mathbf{n}_h^k + \mathbf{v}_h]$  for all  $\mathbf{v}_h \in V_h^{\perp}(\mathbf{n}_h^k) \cap H_{\Gamma_{\mathbf{n}}}^1(\Omega)$ , i.e.,  $\mathbf{t}_h^k$  satisfies

$$\delta_{\mathbf{n}_h} E_1^h[s_h^k, \mathbf{n}_h^k + \mathbf{t}_h^k; \mathbf{v}_h] = 0 \quad \text{for all } \mathbf{v}_h \in V_h^{\perp}(\mathbf{n}_h^k) \cap H_{\Gamma_{\mathbf{n}}}^1(\Omega).$$

**Step (b): Projection.** Normalize  $\mathbf{n}_i^{k+1} := \frac{\mathbf{n}_i^k + \mathbf{t}_i^k}{|\mathbf{n}_i^k + \mathbf{t}_i^k|}$  at all nodes  $x_i \in \mathcal{N}_h$ .

**Step (c): Gradient flow.** Using  $(s_h^k, \mathbf{n}_h^{k+1})$ , find  $s_h^{k+1}$  in  $\mathbb{S}_h(\Gamma_s, g_h)$  such that

$$\int_{\Omega} \frac{s_h^{k+1} - s_h^k}{\delta t} z_h dx = -\delta_{s_h} E_1^h[s_h^{k+1}, \mathbf{n}_h^{k+1}; z_h] - \delta_{s_h} E_2^h[s_h^{k+1}; z_h] \quad \text{for all } z_h \in \mathbb{S}_h \cap H_{\Gamma_s}^1(\Omega).$$

We impose Dirichlet boundary conditions on both  $s_h^k$  and  $\mathbf{n}_h^k$ . Note that the scheme has no restriction on the time step thanks to the implicit Euler method in Step (c).

**4.3. Energy decreasing property.** The quasi-gradient flow scheme in section 4.2.2 has a monotone energy decreasing property, which is a discrete version of (4.5), provided the mesh  $\mathcal{T}_h$  is weakly acute, namely it satisfies (2.10) [22, 45].

**THEOREM 4.2** (energy decrease). *Let  $\mathcal{T}_h$  satisfy (2.10). The iterate  $(s_h^{k+1}, \mathbf{n}_h^{k+1})$  of the algorithm (discrete quasi-gradient flow) of section 4.2.2 exists and satisfies*

$$E^h[s_h^{k+1}, \mathbf{n}_h^{k+1}] \leq E^h[s_h^k, \mathbf{n}_h^k] - \frac{1}{\delta t} \int_{\Omega} (s_h^{k+1} - s_h^k)^2 dx.$$

Equality holds if and only if  $(s_h^{k+1}, \mathbf{n}_h^{k+1}) = (s_h^k, \mathbf{n}_h^k)$  (equilibrium state).

*Proof.* Steps (a) and (b) are monotone, whereas Step (c) decreases the energy.

*Step (a): Minimization.* Since  $E_1^h$  is convex in  $\mathbf{n}_h^k$  for fixed  $s_h^k$ , there exists a tangential variation  $\mathbf{t}_h^k$  which minimizes  $E_1^h[s_h^k, \mathbf{n}_h^k + \mathbf{v}_h^k]$  among all tangential variations  $\mathbf{v}_h^k$ . The fact that  $E_2^h$  is independent of the director field  $\mathbf{n}_h^k$  implies

$$E^h[s_h^k, \mathbf{n}_h^k + \mathbf{t}_h^k] \leq E^h[s_h^k, \mathbf{n}_h^k].$$

*Step (b): Projection.* Since the mesh  $\mathcal{T}_h$  is weakly acute, we claim that

$$\mathbf{n}_h^{k+1} = \frac{\mathbf{n}_h^k + \mathbf{t}_h^k}{|\mathbf{n}_h^k + \mathbf{t}_h^k|} \Rightarrow E_1^h[s_h^k, \mathbf{n}_h^{k+1}] \leq E_1^h[s_h^k, \mathbf{n}_h^k + \mathbf{t}_h^k].$$

We follow [2, 9]. Let  $\mathbf{v}_h = \mathbf{n}_h^k + \mathbf{t}_h^k$  and  $\mathbf{w}_h = \frac{\mathbf{v}_h}{|\mathbf{v}_h|}$  and observe that  $|\mathbf{v}_h| \geq 1$  and  $\mathbf{w}_h$  is well defined. By (2.16) (definition of discrete energy), we need only show that

$$k_{ij} \frac{(s_i^k)^2 + (s_j^k)^2}{2} |\mathbf{w}_h(x_i) - \mathbf{w}_h(x_j)|^2 \leq k_{ij} \frac{(s_i^k)^2 + (s_j^k)^2}{2} |\mathbf{v}_h(x_i) - \mathbf{v}_h(x_j)|^2$$

for all  $x_i, x_j \in \mathcal{N}_h$ . Because  $k_{ij} \geq 0$  for  $i \neq j$ , this is equivalent to showing that  $|\mathbf{w}_h(x_i) - \mathbf{w}_h(x_j)| \leq |\mathbf{v}_h(x_i) - \mathbf{v}_h(x_j)|$ . This follows from the fact that the mapping  $\mathbf{a} \mapsto \mathbf{a}/|\mathbf{a}|$  defined on  $\{\mathbf{a} \in \mathbb{R}^d : |\mathbf{a}| \geq 1\}$  is Lipschitz continuous with constant 1. Note that equality above holds if and only if  $\mathbf{n}_h^{k+1} = \mathbf{n}_h^k$  or, equivalently,  $\mathbf{t}_h^k = \mathbf{0}$ .

*Step (c): Gradient flow.* Since  $E_1^h$  is quadratic in terms of  $s_h^k$ , and since

$$2s_h^{k+1}(s_h^{k+1} - s_h^k) = (s_h^{k+1} - s_h^k)^2 + |s_h^{k+1}|^2 - |s_h^k|^2,$$

reordering terms gives

$$E_1^h[s_h^{k+1}, \mathbf{n}_h^{k+1}] - E_1^h[s_h^k, \mathbf{n}_h^{k+1}] = R_1 - E_1^h[s_h^{k+1} - s_h^k, \mathbf{n}_h^{k+1}] \leq R_1,$$

where

$$R_1 := \delta_{s_h} E_1^h[s_h^{k+1}, \mathbf{n}_h^{k+1}; s_h^{k+1} - s_h^k].$$

On the other hand, Lemma 4.1 implies

$$E_2^h[s_h^{k+1}] - E_2^h[s_h^k] = \int_{\Omega} \psi(s_h^{k+1}) dx - \int_{\Omega} \psi(s_h^k) dx \leq R_2 := \delta_{s_h} E_2^h[s_h^{k+1}; s_h^{k+1} - s_h^k].$$

Combining both estimates and invoking Step (c) of the algorithm yields

$$E^h[s_h^{k+1}, \mathbf{n}_h^{k+1}] - E^h[s_h^k, \mathbf{n}_h^{k+1}] \leq R_1 + R_2 = -\frac{1}{\delta t} \int_{\Omega} (s_h^{k+1} - s_h^k)^2 dx \leq 0,$$

which is the assertion. Note finally that equality occurs if and only if  $s_h^{k+1} = s_h^k$  and  $\mathbf{n}_h^{k+1} = \mathbf{n}_h^k$ , which corresponds to an equilibrium state. This completes the proof.  $\square$

**5. Numerical experiments.** We present computational experiments to illustrate our method, which was implemented with the MATLAB/C++ finite element toolbox FELICITY [48]. For all three-dimensional simulations, we used the algebraic multigrid solver AGMG [39, 37, 38, 40] to solve the linear systems in Steps (a) and (c) of the quasi-gradient flow algorithm. In two dimensions, we simply used the “backslash” command in MATLAB.

**5.1. Tangential variations.** Solving Step (a) of the algorithm requires a tangential basis for the test function and the solution. However, forming the matrix system is easily done by first ignoring the tangential variation constraint (i.e., arbitrary variations), followed by a simple modification of the matrix system.

Let  $\mathbf{A}t_h^k = B$  represent the linear system in Step (a), and suppose  $d = 3$ . Multiplying by a discrete test function  $\mathbf{v}_h$ , we have

$$\mathbf{v}_h^T \mathbf{A}t_h^k = \mathbf{v}_h^T B \quad \text{for all } \mathbf{v}_h \in \mathbb{R}^{dN}.$$

Next, using  $\mathbf{n}_h^k$ , find  $\mathbf{r}_1, \mathbf{r}_2$  such that  $\{\mathbf{n}_h^k, \mathbf{r}_1, \mathbf{r}_2\}$  forms an orthonormal basis of  $\mathbb{R}^3$  at each node  $x_i$ , i.e., find an orthonormal basis of  $V_h^\perp(\mathbf{n}_h^k)$ . Next, expand  $\mathbf{t}_h^k = \Phi_1 \mathbf{r}_1 + \Phi_2 \mathbf{r}_2$  and make a similar expansion for  $\mathbf{v}_h$ . After a simple rearrangement and partitioning of the linear system, one finds it decouples into two smaller systems: one for  $\Phi_1$  and one for  $\Phi_2$ . After solving for  $\Phi_1, \Phi_2$ , define the nodal values of  $\mathbf{t}_h^k$  by the formula  $\mathbf{t}_h^k = \Phi_1 \mathbf{r}_1 + \Phi_2 \mathbf{r}_2$ .

**5.2. Point defect in two dimensions.** For the classic Frank energy  $\int_\Omega |\nabla \mathbf{n}|^2$ , a point defect in two dimensions has infinite energy [47]. This is not the case for the energy (2.1), because  $s$  can go to zero at the location of the point defect, so the term  $\int_\Omega s^2 |\nabla \mathbf{n}|^2$  remains finite.

We simulate the gradient flow evolution of a point defect moving to the center of the domain ( $\Omega$  is the unit square). We set  $\kappa = 2$  and take the double well potential to have the following splitting:

$$\begin{aligned} \psi(s) &= \psi_c(s) - \psi_e(s) \\ &= 63.0s^2 - (-16.0s^4 + 21.333333333333s^3 + 57.0s^2), \end{aligned}$$

with local minimum at  $s = 0$  and global minimum at  $s = s^* := 0.750025$  (see section 2.1, and note that a vertical shift makes  $\psi(s^*) = 0$  without affecting the gradient flow). We impose the following Dirichlet boundary conditions for  $s$  and  $\mathbf{n}$  on  $\Gamma_s = \Gamma_{\mathbf{n}} = \partial\Omega$ :

$$(5.1) \quad s = s^*, \quad \mathbf{n} = \frac{(x, y) - (0.5, 0.5)}{|(x, y) - (0.5, 0.5)|}.$$

Initial conditions on  $\Omega$  for the gradient flow are  $s = s^*$  and a regularized point defect away from the center.

Figure 1 shows the evolution of the director field  $\mathbf{n}$  and the scalar degree of orientation parameter  $s$ . One can see the regularizing effect of  $s$ . We note that an  $L^2$ -gradient flow scheme, instead of the quasi- (weighted) gradient flow we use, yields a much slower evolution to equilibrium.

**5.3. Plane defect in three dimensions.** Next, we simulate the gradient flow evolution of the liquid crystal director field toward a *plane* defect in the unit cube  $\Omega = (0, 1)^3$ . This is motivated by an exact solution found in [47, sect. 6.4]. We set  $\kappa = 0.2$  and *remove* the double well potential. We impose the following mixed boundary

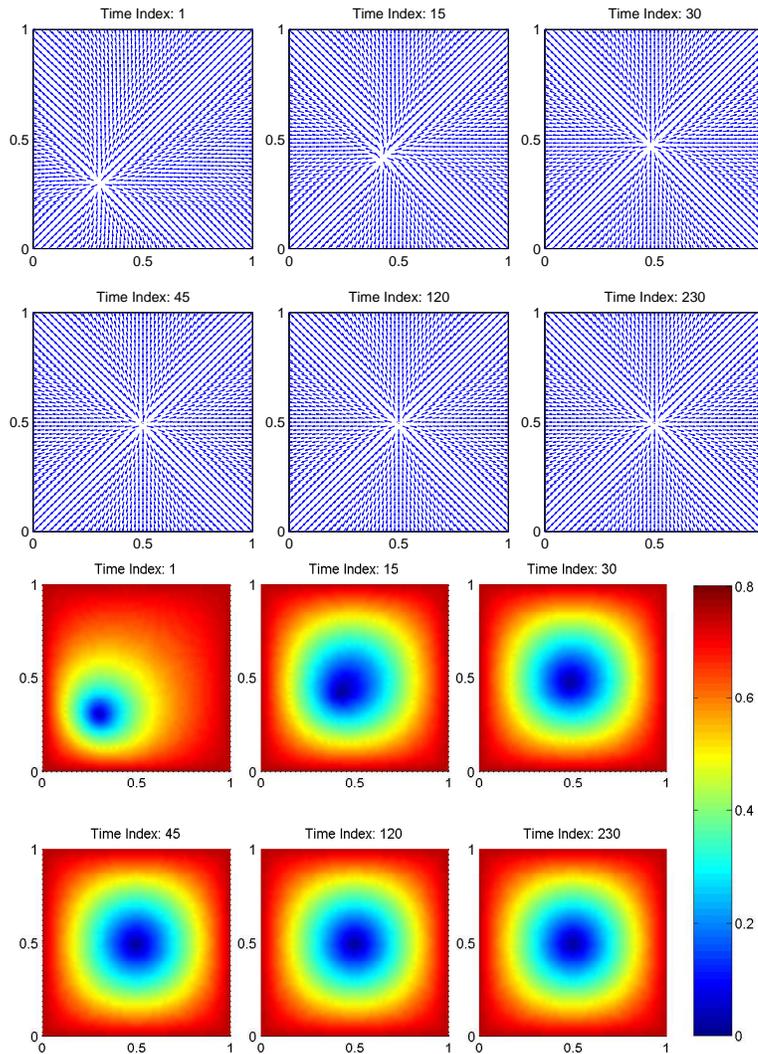


FIG. 1. Evolution of a point defect toward its equilibrium state (section 5.2). Time step is  $\delta t = 0.02$ . The minimum value of  $s$ , at time index 230, is  $2.0226 \cdot 10^{-2}$ .

conditions for  $(s, \mathbf{n})$ : Dirichlet conditions on  $\Gamma_s = \Gamma_{\mathbf{n}} = \overline{\partial\Omega} \cap (\{z = 0\} \cup \{z = 1\})$ :

$$(5.2) \quad \begin{aligned} z = 0 : \quad s &= s^*, & \mathbf{n} &= (1, 0, 0), \\ z = 1 : \quad s &= s^*, & \mathbf{n} &= (0, 1, 0) \end{aligned}$$

and Neumann conditions  $\boldsymbol{\nu} \cdot \nabla s = 0$  and  $\boldsymbol{\nu} \cdot \nabla \mathbf{n}$  on the remaining part of  $\partial\Omega$ . These conditions are not covered in section 3, but we explore them computationally. The exact solution  $(s, \mathbf{n})$  (at equilibrium) only depends on  $z$  and is given by

$$(5.3) \quad \begin{aligned} \mathbf{n}(z) &= (1, 0, 0) \text{ for } z < 0.5, & \mathbf{n}(z) &= (0, 1, 0) \text{ for } z > 0.5, \\ s(z) &= 0 \text{ at } z = 0.5, \text{ and } s(z) \text{ is linear for } z \in (0, 0.5) \cup (0.5, 1.0). \end{aligned}$$

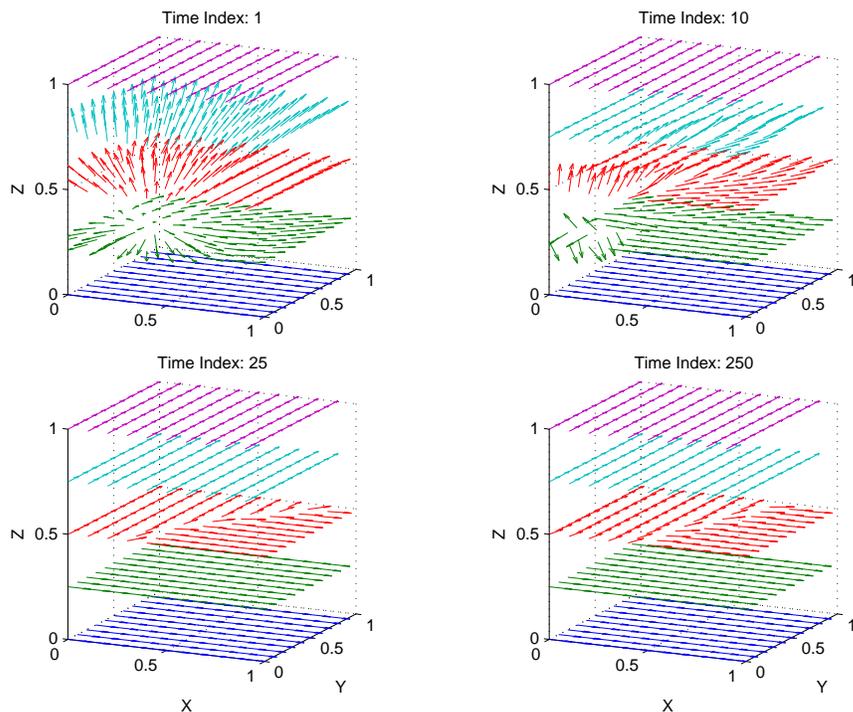


FIG. 2. Evolution toward an (equilibrium) plane defect (section 5.3). The director field  $\mathbf{n}$  is shown at five different horizontal slices. The time step used was  $\delta t = 0.02$ .

Initial conditions on  $\Omega$  for the gradient flow are  $s = s^*$  and a regularized point defect away from the center of the cube.

Figure 2 shows the evolution of the director field  $\mathbf{n}$  toward the plane defect. Only a few slices are shown in Figure 2 because of the simple form of the equilibrium solution.

Figure 3 (left) shows the components of  $\mathbf{n}$  evaluated along a one-dimensional vertical slice. Clearly, the numerical solution approximates the exact solution well, except at the narrow transition region near  $z = 0.5$ . Furthermore, Figure 3 (right) shows the corresponding evolution of the degree of orientation parameter  $s$  (evaluated along the same one-dimensional vertical slice). One can see the regularizing effect of  $s$ , i.e., at equilibrium,  $s \approx 0.008$  at the  $z = 0.5$  plane (the defect plane of  $\mathbf{n}$ ). Our numerical experiments suggest that  $s|_{z=0.5} \rightarrow 0$  as the mesh size goes to zero.

**5.4. Fluting effect and propeller defect.** This example further investigates the effect of  $\kappa$  on the presence of defects. An exact solution of a line defect in a right circular cylinder is given in [47, sect. 6.5], where it is shown that for  $\kappa$  sufficiently large (say  $\kappa > 1$ ), the director field is smooth, but if  $\kappa$  is sufficiently small, then a line defect in  $\mathbf{n}$  appears along the axis of the cylinder. Our numerical experiments confirm this.

To further illustrate this effect, we conducted a similar experiment for a unit cube domain  $\Omega = (0, 1)^3$ . Again, for simplicity we *remove* the double well potential. We set Dirichlet boundary conditions for  $(s, \mathbf{n})$  on the vertical sides of the cube  $\Gamma_s =$

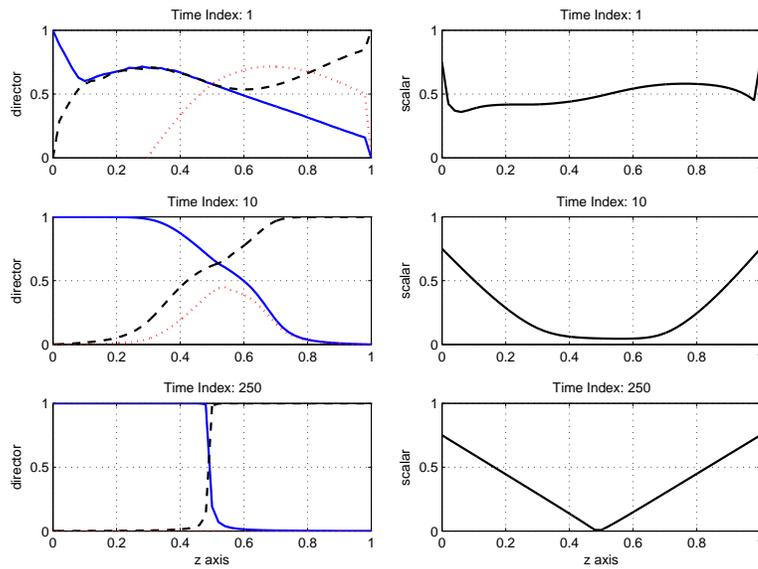


FIG. 3. Evolution toward an (equilibrium) plane defect (section 5.3); time step is  $\delta t = 0.02$ . Left: plots of the three components of  $\mathbf{n}$ , evaluated along the vertical line  $x = 0.5, y = 0.5$ , are shown at three time indices (solid blue curve:  $\mathbf{n} \cdot \mathbf{e}_1$ , dashed black curve:  $\mathbf{n} \cdot \mathbf{e}_2$ , dotted red curve:  $\mathbf{n} \cdot \mathbf{e}_3$ ). At equilibrium,  $\mathbf{n}$  is nearly piecewise constant with a narrow transition region around  $z = 0.5$ . Right: plots of the degree of orientation  $s$ , corresponding to  $\mathbf{n}$ , are shown. The equilibrium solution is piecewise linear, with a kink at  $z = 0.5$ , where  $s \approx 0.008$ .

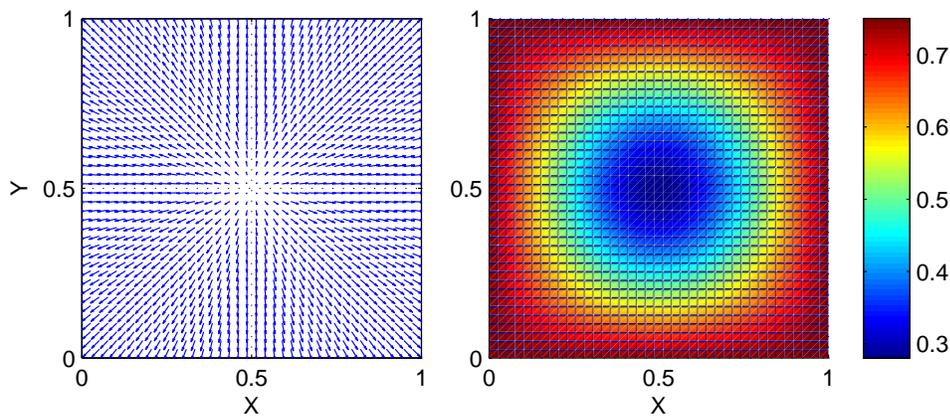


FIG. 4. Equilibrium state (section 5.4) of  $\mathbf{n}$  and  $s$ . One horizontal slice ( $z = 0.5$ ) is plotted:  $\mathbf{n}$  on the left,  $s$  on the right ( $\mathbf{n}$  and  $s$  are approximately independent of  $z$ ). The director field points out of the plane (i.e.,  $\mathbf{n} \cdot \mathbf{e}_3 \neq 0$ ), and  $s > 0.278$ , so there is no defect.

$\Gamma_{\mathbf{n}} = \overline{\partial\Omega} \cap (\{x = 0\} \cup \{x = 1\} \cup \{y = 0\} \cup \{y = 1\})$ , with

$$(5.4) \quad s = s^*, \quad \mathbf{n}(x, y, z) = \frac{(x, y) - (0.5, 0.5)}{|(x, y) - (0.5, 0.5)|},$$

and Neumann conditions  $\nu \cdot \nabla s = 0$  and  $\nu \cdot \nabla \mathbf{n} = 0$  on the top and bottom parts of  $\partial\Omega$ ; this situation is not covered in section 3. Figure 4 shows the equilibrium solution

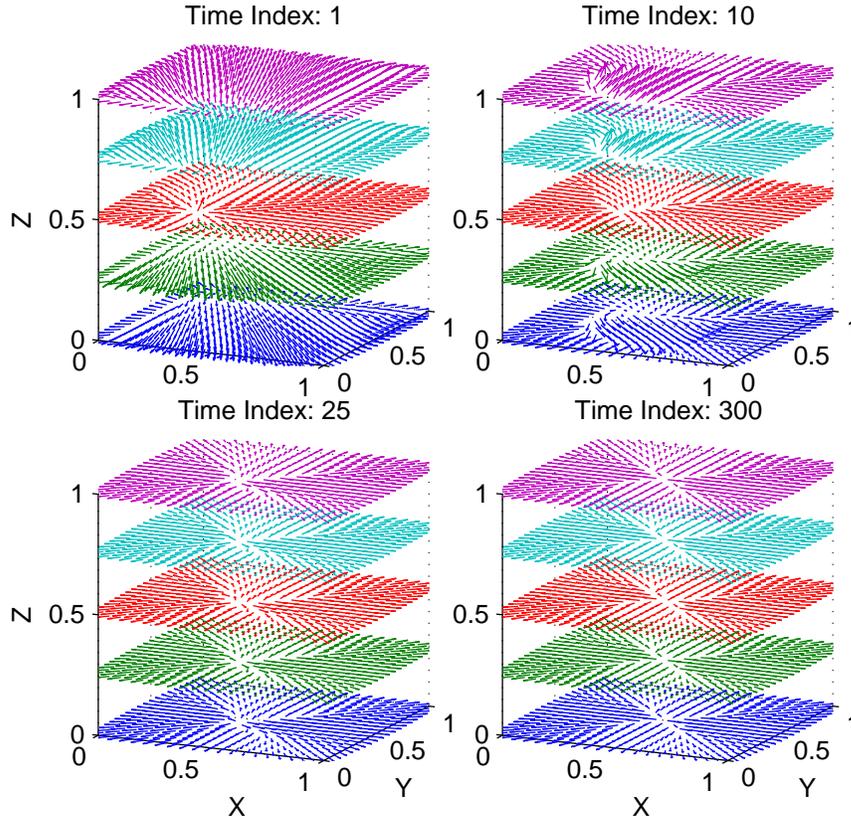


FIG. 5. Evolution toward an (equilibrium) “propeller” defect (section 5.4). Director field  $\mathbf{n}$  is shown at five different horizontal slices through the cube. The time step used was  $\delta t = 0.02$ .

when  $\kappa = 2$ . The  $z$ -component of  $\mathbf{n}$  is *not* zero, i.e., it points out of the plane of the horizontal slice that we plot. This is referred to as the “fluting effect” (or escape to the third dimension [47]). In this case, the degree of orientation parameter  $s$  is bounded well away from zero, so the director field is smooth (i.e., no defect).

Next, we choose  $\kappa = 0.1$  and initialize our gradient flow scheme with  $s = s^*$  and a regularized point defect away from the center of the cube for  $\mathbf{n}$ .

Figure 5 shows the evolution of the director field  $\mathbf{n}$  toward a “propeller” defect (two plane defects intersecting). Figure 6 shows  $\mathbf{n}$  and  $s$  in their final equilibrium state at the  $z = 0.5$  plane. Both  $\mathbf{n}$  and  $s$  are nearly uniform with respect to the  $z$  variable. The regularizing effect of  $s$  is apparent, i.e.,  $s \approx 2 \times 10^{-5}$  near where  $\mathbf{n}$  has a discontinuity. The three-dimensional shape of the defect resembles two planes intersecting near the  $x = 0.5, y = 0.5$  vertical line, i.e., the defect looks like an “X” extruded in the  $z$  direction.

**5.5. Floating plane defect.** This example investigates the effect of the domain shape on the defect. The setup here is essentially the same as that in section 5.4, with  $\kappa = 0.1$ , except the domain is the rectangular box  $\Omega = (0, 1) \times (0, 0.7143) \times (0, 1)$ . Figure 7 shows  $\mathbf{n}$  and  $s$  in their final equilibrium state at the  $z = 0.5$  plane. Both  $\mathbf{n}$  and  $s$  are approximately uniform with respect to the  $z$  variable. Instead of the propeller defect, we get a “floating” plane defect aligned with the major axis of the

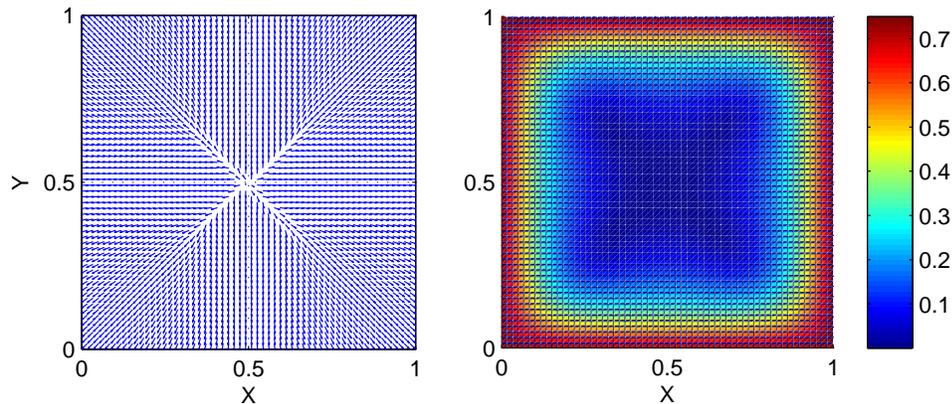


FIG. 6. Equilibrium state of a “propeller” defect (section 5.4). One horizontal slice ( $z = 0.5$ ) is plotted:  $\mathbf{n}$  on the left,  $s$  on the right ( $\mathbf{n}$  and  $s$  are nearly independent of  $z$ ). The  $z$ -component of  $\mathbf{n}$  is zero, and  $s \approx 2 \times 10^{-5}$  near the discontinuity in  $\mathbf{n}$ .

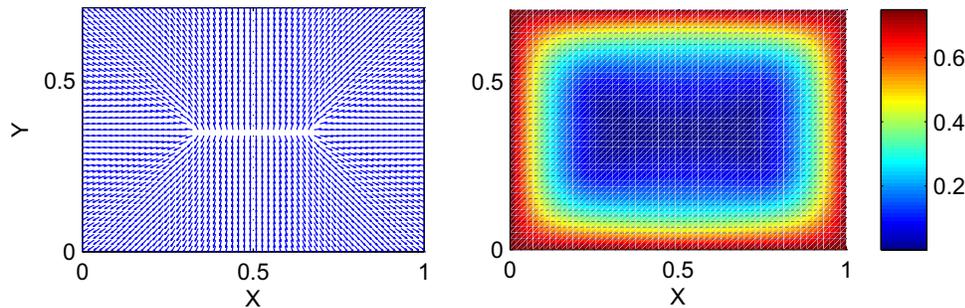


FIG. 7. Equilibrium state of a floating plane defect on a rectangular domain (section 5.5). One horizontal slice ( $z = 0.5$ ) is plotted:  $\mathbf{n}$  on the left,  $s$  on the right ( $\mathbf{n}$  and  $s$  are approximately independent of  $z$ ). The  $z$ -component of  $\mathbf{n}$  is zero, and  $s > 0$  with  $s \approx 7 \times 10^{-5}$  near the discontinuity in  $\mathbf{n}$ .

box. Again, the regularizing effect of  $s$  is apparent, i.e.,  $s \approx 7 \times 10^{-5}$  near where  $\mathbf{n}$  has a discontinuity.

**6. Conclusion.** We introduced and analyzed a robust finite element method (FEM) for a degenerate energy functional that models nematic liquid crystals with variable degree of orientation. We also developed a quasi-gradient flow scheme for computing energy minimizers, with a strict monotone energy decreasing property. The numerical experiments show a variety of defect structures that Ericksen’s model exhibits. Some of the defect structures are high dimensional with surprising shapes (see Figure 6). We mention that [31] also found a “propeller” (or “X”) shaped defect within a two-dimensional Landau–deGennes ( $\mathbf{Q}$ -tensor) model. An interesting extension of this work would be to couple the effect of external fields (e.g., magnetic and electric fields) to the liquid crystal as a way of driving and manipulating the defect structures.

**Acknowledgment.** We thank L. Ambrosio for useful discussions regarding the regularization argument in Proposition 3.2.

## REFERENCES

- [1] R. A. ADAMS AND J. J. F. FOURNIER, *Sobolev Spaces*, 2nd ed., Pure Appl. Math. (Amst.) 140, Elsevier/Academic Press, Amsterdam, 2003.
- [2] F. ALOUGES, *A new algorithm for computing liquid crystal stable configurations: The harmonic mapping case*, SIAM J. Numer. Anal., 34 (1997), pp. 1708–1726, <https://doi.org/10.1137/S0036142994264249>.
- [3] L. AMBROSIO, *Existence of minimal energy configurations of nematic liquid crystals with variable degree of orientation*, Manuscripta Math., 68 (1990), pp. 215–228.
- [4] L. AMBROSIO, *Regularity of solutions of a degenerate elliptic variational problem*, Manuscripta Math., 68 (1990), pp. 309–326.
- [5] T. ARAKI AND H. TANAKA, *Colloidal aggregation in a nematic liquid crystal: Topological arrest of particles by a single-stroke disclination line*, Phys. Rev. Lett., 97 (2006), 127801.
- [6] S. BADIA, F. M. GUILLÉN-GONZÁLEZ, AND J. V. GUTIÉRREZ-SANTACREU, *An overview on numerical analyses of nematic liquid crystal flows*, Arch. Comput. Methods Engrg., 18 (2011), pp. 285–313.
- [7] J. M. BALL AND A. ZARNESCU, *Orientability and energy minimization in liquid crystal models*, Arch. Ration. Mech. Anal., 202 (2011), pp. 493–535.
- [8] J. W. BARRETT, X. FENG, AND A. PROHL, *Convergence of a fully discrete finite element method for a degenerate parabolic system modelling nematic liquid crystals with variable degree of orientation*, ESAIM Math. Model. Numer. Anal., 40 (2006), pp. 175–199.
- [9] S. BARTELS, *Stability and convergence of finite-element approximation schemes for harmonic maps*, SIAM J. Numer. Anal., 43 (2005), pp. 220–238, <https://doi.org/10.1137/040606594>.
- [10] S. BARTELS, *Numerical analysis of a finite element scheme for the approximation of harmonic maps into surfaces*, Math. Comp., 79 (2010), pp. 1263–1301.
- [11] S. BARTELS, G. DOLZMANN, AND R. H. NOCHETTO, *A finite element scheme for the evolution of orientational order in fluid membranes*, ESAIM Math. Model. Numer. Anal., 44 (2010), pp. 1–31.
- [12] S. BARTELS, G. DOLZMANN, AND R. H. NOCHETTO, *Finite element methods for director fields on flexible surfaces*, Interfaces Free Bound., 14 (2012), pp. 231–272.
- [13] P. BAUMAN, M. C. CALDERER, C. LIU, AND D. PHILLIPS, *The phase transition between chiral nematic and smectic  $a^*$  liquid crystals*, Arch. Ration. Mech. Anal., 165 (2002), pp. 161–186.
- [14] F. BETHUEL, H. BREZIS, AND F. HÉLEIN, *Ginzburg-Landau Vortices*, Progr. Nonlinear Differential Equations Appl. 13, Birkhäuser Boston, Inc., Boston, MA, 1994.
- [15] R. B. BIRD, R. C. ARMSTRONG, AND O. HASSAGER, *Dynamics of Polymeric Liquids—Volume 1: Fluid Mechanics*, 2nd ed., Wiley Interscience Publ. 1, John Wiley & Sons, New York, 1987.
- [16] L. M. BLINOV, *Electro-Optical and Magneto-Optical Properties of Liquid Crystals*, John Wiley & Sons, New York, 1983.
- [17] A. BRAIDES,  *$\Gamma$ -Convergence for Beginners*, Oxford Lecture Ser. Math. Appl. 22, Oxford University Press, Oxford, UK, 2002.
- [18] A. BRAIDES, *Local Minimization, Variational Evolution and  $\Gamma$ -Convergence*, Lecture Notes in Math. 2094, Springer, Berlin, 2014.
- [19] J. H. BRANDTS, S. KOROTOV, AND M. KRÍŽEK, *The discrete maximum principle for linear simplicial finite element approximations of a reaction-diffusion problem*, Linear Algebra Appl., 429 (2008), pp. 2344–2357.
- [20] H. BREZIS, J.-M. CORON, AND E. H. LIEB, *Harmonic maps with defects*, Comm. Math. Phys., 107 (1986), pp. 649–705.
- [21] M. C. CALDERER, D. GOLOVATY, F.-H. LIN, AND C. LIU, *Time evolution of nematic liquid crystals with variable degree of orientation*, SIAM J. Math. Anal., 33 (2002), pp. 1033–1047, <https://doi.org/10.1137/S0036141099362086>.
- [22] P. G. CIARLET AND P.-A. RAVIART, *Maximum principle and uniform convergence for the finite element method*, Comput. Methods Appl. Mech. Engrg., 2 (1973), pp. 17–31.
- [23] R. COHEN, S.-Y. LIN, AND M. LUSKIN, *Relaxation and gradient methods for molecular orientation in liquid crystals*, Comput. Phys. Comm., 53 (1989), pp. 455–465.
- [24] G. DAL MASO, *An introduction to  $\Gamma$ -Convergence*, Progr. Nonlinear Differential Equations Appl. 8, Birkhäuser Boston, Inc., Boston, MA, 1993.
- [25] P. G. DE GENNES AND J. PROST, *The Physics of Liquid Crystals*, 2nd ed., Int. Ser. Monogr. Phys. 83, Oxford University Press, Oxford, UK, 1995.
- [26] J. L. ERICKSEN, *Liquid crystals with variable degree of orientation*, Arch. Ration. Mech. Anal., 113 (1991), pp. 97–120.
- [27] L. C. EVANS, *Partial Differential Equations*, American Mathematical Society, Providence, RI, 1998.

- [28] F. M. GUILLÉN GONZÁLEZ AND J. V. GUTIÉRREZ-SANTACREU, *A linear mixed finite element scheme for a nematic Ericksen-Leslie liquid crystal model*, ESAIM Math. Model. Numer. Anal., 47 (2013), pp. 1433–1464.
- [29] R. HARDT, D. KINDERLEHRER, AND M. LUSKIN, *Remarks about the mathematical theory of liquid crystals*, in Calculus of Variations and Partial Differential Equations, S. Hildebrandt, D. Kinderlehrer, and M. Miranda, eds., Lecture Notes in Math. 1340, Springer Berlin, Heidelberg, 1988, pp. 123–138.
- [30] S. KOROTOV, M. KRÍŽEK, AND P. NEITTAANMÄKIA, *Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle*, Math. Comp., 70 (2001), pp. 107–119.
- [31] S. KRALJ AND A. MAJUMDAR, *Order reconstruction patterns in nematic liquid crystal wells*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 470 (2014), 2169.
- [32] R. G. LARSON, *The Structure and Rheology of Complex Fluids*, Oxford University Press, Oxford, UK, 1999.
- [33] F.-H. LIN, *Nonlinear theory of defects in nematic liquid crystals; phase transition and flow phenomena*, Comm. Pure Appl. Math., 42 (1989), pp. 789–814.
- [34] F.-H. LIN, *On nematic liquid crystals with variable degree of orientation*, Comm. Pure Appl. Math., 44 (1991), pp. 453–468.
- [35] S.-Y. LIN AND M. LUSKIN, *Relaxation methods for liquid crystal problems*, SIAM J. Numer. Anal., 26 (1989), pp. 1310–1324, <https://doi.org/10.1137/0726076>.
- [36] C. LIU AND N. J. WALKINGTON, *Approximation of liquid crystal flows*, SIAM J. Numer. Anal., 37 (2000), pp. 725–741, <https://doi.org/10.1137/S0036142997327282>.
- [37] A. NAPOV AND Y. NOTAY, *Algebraic analysis of aggregation-based multigrid*, Numer. Linear Algebra Appl., 18 (2011), pp. 539–564.
- [38] A. NAPOV AND Y. NOTAY, *An algebraic multigrid method with guaranteed convergence rate*, SIAM J. Sci. Comput., 34 (2012), pp. A1079–A1109, <https://doi.org/10.1137/100818509>.
- [39] Y. NOTAY, *An aggregation-based algebraic multigrid method*, Electron. Trans. Numer. Anal., 37 (2010), pp. 123–146.
- [40] Y. NOTAY, *Aggregation-based algebraic multigrid for convection-diffusion equations*, SIAM J. Sci. Comput., 34 (2012), pp. A2288–A2316, <https://doi.org/10.1137/110835347>.
- [41] R. G. OWENS AND T. N. PHILLIPS, *Computational Rheology*, World Scientific, River Edge, NJ, 2002.
- [42] R. SCHOEN AND K. UHLENBECK, *A regularity theory for harmonic maps*, J. Differential Geom., 17 (1982), pp. 307–335.
- [43] J. SHEN AND X. YANG, *Numerical approximations of Allen-Cahn and Cahn-Hilliard equations*, Discrete Contin. Dyn. Syst., 28 (2010), pp. 1669–1691.
- [44] J. SHEN AND X. YANG, *A phase-field model and its numerical approximation for two-phase incompressible flows with different densities and viscosities*, SIAM J. Sci. Comput., 32 (2010), pp. 1159–1179, <https://doi.org/10.1137/09075860X>.
- [45] G. STRANG AND G. FIX, *An Analysis of the Finite Element Method*, 2nd ed., Wellesley-Cambridge Press, Wellesley, MA, 2008.
- [46] K. TOJO, A. FURUKAWA, T. ARAKI, AND A. ONUKI, *Defect structures in nematic liquid crystals around charged particles*, Eur. Phys. J. E, 30 (2009), pp. 55–64.
- [47] E. G. VIRGA, *Variational Theories for Liquid Crystals*, Appl. Math. Math. Comput. 8, Chapman and Hall, London, 1994.
- [48] S. W. WALKER, *FELICITY: Finite Element Implementation and Computational Interface Tool for You*, <http://www.mathworks.com/matlabcentral/fileexchange/31141-felicity>, 2017.
- [49] N. J. WALKINGTON, *Numerical approximation of nematic liquid crystal flows governed by the Ericksen-Leslie equations*, ESAIM Math. Model. Numer. Anal., 45 (2011), pp. 523–540.
- [50] R. L. WHEEDEN AND A. ZYGMUND, *Measure and Integral: An Introduction to Real Analysis*, Marcel-Dekker, New York, 1977.
- [51] S. M. WISE, C. WANG, AND J. S. LOWENGRUB, *An energy-stable and convergent finite-difference scheme for the phase field crystal equation*, SIAM J. Numer. Anal., 47 (2009), pp. 2269–2288, <https://doi.org/10.1137/080738143>.
- [52] X. YANG, M. G. FOREST, H. LI, C. LIU, J. SHEN, Q. WANG, AND F. CHEN, *Modeling and simulations of drop pinch-off from liquid crystal filaments and the leaky liquid crystal faucet immersed in viscous fluids*, J. Comput. Phys., 236 (2013), pp. 1–14.
- [53] X. YANG, M. G. FOREST, C. LIU, AND J. SHEN, *Shear cell rupture of nematic liquid crystal droplets in viscous fluids*, J. Non-Newton. Fluid Mech., 166 (2011), pp. 487–499.