

## Chapter 5

# The Q-tensor model with uniaxial constraint

Juan Pablo Borthagaray<sup>a,b,d</sup> and Shawn W. Walker<sup>c,\*,e</sup>

<sup>a</sup>*Department of Mathematics, University of Maryland, College Park, MD, United States,*

<sup>b</sup>*Departamento de Matemática y Estadística del Litoral, Universidad de la República, Salto, Uruguay,* <sup>c</sup>*Department of Mathematics and Center for Computation and Technology (CCT), Louisiana State University, Baton Rouge, LA, United States*

\*Corresponding author: e-mail address: [walker@math.lsu.edu](mailto:walker@math.lsu.edu)

## Contents

<b>1 Physics of liquid crystals</b>	<b>314</b>	4.3 Gradient flow	349
1.1 Fundamentals	314	4.4 Gamma convergence	354
1.2 Applications	316	4.5 Numerical experiment	357
<b>2 Modeling of nematic liquid crystals</b>	<b>316</b>	<b>5 Colloidal effects</b>	<b>360</b>
2.1 Order parameters	316	5.1 Conforming non-obtuse mesh	361
2.2 Continuum mechanics	319	5.2 Immersed boundary method	364
2.3 Dynamics	324	<b>6 Electric fields</b>	<b>370</b>
<b>3 Mathematical formulation</b>	<b>325</b>	6.1 Modified energies	371
3.1 Oseen-Frank	325	6.2 Computational electric field	
3.2 Landau-de Gennes	326	example	373
3.3 Ericksen	331	<b>7 The Landau-de Gennes model with</b>	
<b>4 The uniaxially constrained Q-model</b>	<b>344</b>	<b>and without the uniaxial constraint</b>	<b>373</b>
4.1 Modeling assumptions	345	<b>8 Conclusion</b>	<b>376</b>
4.2 Discretization	347	<b>References</b>	<b>377</b>

---

## Abstract

This chapter is about the modeling of nematic liquid crystals (LCs) and their numerical simulation. We begin with an overview of the basic physics of LCs and discuss some of their many applications. Next, we delve into the modeling arguments needed to obtain macroscopic order parameters which can be used to formulate a continuum model. We then survey different continuum descriptions, namely the Oseen-Frank, Ericksen, and Landau-de Gennes (Q-tensor) models, which essentially model the LC material like an anisotropic elastic material. In particular, we review the mathematical theory underlying the three different continuum models and highlight the different trade-offs of using these models.

---

d. JPB has been supported in part by NSF grant DMS-14111808.

e. SWW has been supported in part by NSF grant DMS-1555222 (CAREER).

Next, we consider the numerical simulation of these models with a survey of various methods, with a focus on the Ericksen model. We then show how techniques from the Ericksen model can be combined with the Landau-de Gennes model to yield a  $\mathbf{Q}$ -tensor model that *exactly* enforces uniaxiality, which is relevant for modeling many nematic LC systems. This is followed by an in-depth numerical analysis, using tools from  $\Gamma$ -convergence, to justify our discrete method. We also show several numerical experiments and comparisons with the standard Landau-de Gennes model.

### Keywords

Nematic liquid crystals, Finite element method, Gamma-convergence, Landau-de Gennes, Defects

### MSC Codes

65N30, 49M25, 35J70

---

## 1 Physics of liquid crystals

### 1.1 Fundamentals

The name “liquid crystal” appears self-contradictory. A crystal has a rigid molecular structure and so is associated with being a solid. How could a crystal be liquid? Thinking more broadly, a crystal is matter that possesses some kind of macroscopic order, such as having individual molecules arranged in a lattice. On the other hand, a liquid has no macroscopic order.

Liquid crystals (LCs) are a *meso-phase* of matter, having a degree of macroscopic order that is *between* a liquid and a solid (Virga, 1994). A classic solid crystal has both translational order (points in the lattice do not move) and orientational order (neighboring molecules have similar orientation). A LC has no strong translational order, i.e. the molecules are free to slide about, but they must roughly maintain the same orientation with neighboring molecules. Thus, LCs have *partial orientational order*.

The initial, accidental discovery of LCs is classically attributed to the Austrian botanist Friedrich Reinitzer (Reinitzer, 1888, 1889), who was studying carrots. While heating cholesteryl benzoate, he saw the material exhibit an LC phase. In order to better understand this, he sought the help of German physicist, Otto Lehmann (Lehmann, 1889) who had experimental apparatus capable of better analysis. After this initial work, Lehmann continued to study LCs, while Reinitzer moved on. Further information on the history of LCs can be found in Sluckin et al. (2004), which contains translations of Reinitzer’s and Lehmann’s initial papers. Moreover, one can consult (de Gennes and Prost, 1995; de Jeu, 2012; Luckhurst and Veracini, 1994; Mottram and Newton, 2014) for more details on the basic physics of LCs.

The LC state may be obtained as a function of temperature between the crystalline and isotropic liquid phases; in such a case, the material is called a thermotropic LC. Other classes include lyotropic and metallotropic LCs, in

which concentration of the LC molecules in a solvent or the ratio between organic and inorganic molecules determine the phase transitions, respectively.

Let us consider thermotropic LCs. In a crystalline solid, molecules exhibit both long-range ordering of the positions of the centers and orientation of the molecules. As the substance is heated, the molecules gain kinetic energy and large molecular vibrations usually make these two ordering types disappear. This results in a fluid phase. In substances capable of producing an LC mesophase, the long-range orientational ordering survives until a higher temperature than the long-range positional ordering. Whenever long-range positional ordering is completely absent, but orientational order remains, the LC is regarded as *nematic*. At lower temperatures, the molecules may order along a preferred direction, forming layered structures: this is called the *smectic* phase. In turn, smectic mesophases may be classified into subclasses (such as smectics A, smectics C and hexatic smectics), depending on the type and degree of positional and orientational order (de Gennes and Prost, 1995). Some substances with chiral molecules (i.e. different from their mirror image) may give rise to a *cholesteric* mesophase, in which the structure possesses a helical distortion.

This chapter will only consider nematic LCs, and regard their molecules as rods, elongated in one direction and thin in the other two directions. Imagining a bunch of thin rods packed together, it is natural to expect the orientation of neighboring rods to be similar, but the rods are free to slide along each other. Indeed, the partial order of LCs is essentially due to the anisotropic shape of the LC molecules. Naturally, most LC molecules do not possess axial symmetry. If the molecules resemble more laths than rods, it is expected that the energy interaction can be minimized if the molecules are fully aligned; this necessarily involves a certain degree of *biaxiality*. Roughly, this was the rationale behind the prediction of the biaxial nematic phase by Freiser (1970).

Since that seminal work, empirical evidence of biaxial states in certain lyotropic LCs has been well documented (see Yu and Saupe, 1980, for example). Nevertheless, for thermotropic LCs the nematic biaxial phase remained elusive for a long period, and was first reported long after Freiser's original prediction (Acharya et al., 2004; Madsen et al., 2004; Prasad et al., 2005). As pointed out by Sonnet and Virga (2012, Section 4.1),

*The vast majority of nematic liquid crystals do not, at least in homogeneous equilibrium states, show any sign of biaxiality.*

Therefore, in this chapter, we shall focus mainly on uniaxial LCs. We discuss three models for the equilibrium configurations. Uniaxiality is naturally built into both the Oseen-Frank and the Ericksen models, as the LC orientation is modeled by a vector field. In contrast, the Landau-de Gennes model represents molecular orientation by means of a tensor field, and thus accounts for biaxiality. However, one can enforce uniaxiality and obtain a model that is not equivalent to either the Ericksen or the Oseen-Frank models. In particular, the uniaxially

constrained Landau-de Gennes model represents the LC molecule orientation by a *line field*, and thus it allows for non-orientable configurations. We will discuss this with more detail in Section 4.

## 1.2 Applications

The most well-known application of LCs is in electronic displays (Hoogboom et al., 2007; Roques-Carmes et al., 2004), which is due to an LC's birefringence property. Indeed, some LC materials polarize light (depending on the orientation of molecules) and this can be controlled through external fields, such as electric fields. This, combined with sophisticated engineering, delivers the flat panel LC display.

However, many newer uses for LCs are being found in material science, that either further build on LCs ability to manipulate light or take advantage of the mechanical properties of the material's anisotropy (Lagerwall and Scalia, 2012). For example, Humar and Muševič (2010) demonstrates three dimensional LC droplets that act as lasers, which may be used as bio-sensors. Self-assembly of rigid particles (inclusions) immersed in an LC medium (Čopar et al., 2014, 2015) has the potential to make new materials. Clever optical effects with LC droplets (Schwartz et al., 2018) provide novel means of creating secure “markers” that cannot be counterfeit.

Furthermore, LC models provide a test bed for investigating continuum models of complex fluids (e.g. Ericksen-Leslie (Cruz et al., 2013; Mottram et al., 2013; Walkington, 2011)), especially swarms of bacteria (Vicsek and Zafeiris, 2012; Marchetti et al., 2013; Giomi, 2015) which is sometimes called *active matter* (Ramaswamy, 2010, 2017; Doostmohammadi et al., 2018; Miles et al., 2019). The shape of a bacterium is very reminiscent of LC molecules (elongated rods) so it is not surprising that LC models, coupled with fluid dynamics, may be reused. Therefore, LC research is a very active field within physics, mathematics, biology, and soft-matter in general.

## 2 Modeling of nematic liquid crystals

We review three models for the equilibrium states of LCs. Specifically, we show how to obtain a continuum description by an appropriate averaging over molecules. This leads to continuum mechanics type models that derive from minimizing an energy; we refer to de Gennes and Prost (1995), Virga (1994), Mottram and Newton (2014) for more details on the modeling of LCs.

### 2.1 Order parameters

Modeling individual LC molecules is certainly viable via molecular dynamics or Monte Carlo methods (Biscarini et al., 1995; Vanzo et al., 2012; Moreno-Razo et al., 2012; Whitmer et al., 2013; Teixeira-Souza et al., 2015; Changizrezaei and Denniston, 2017) and has the advantage of being based on first principles

(i.e. completely “correct”), but is also very expensive computationally. In order to build on these models for, say, coupling to fluids or doing optimal design, we require a macroscopic description of LCs. Usually, the transition between phases of different symmetry is described in terms of an order parameter, that represents the extent to which the configuration of the more symmetric phase differs from that of the less symmetric phase.

In the following discussion, we fix the spatial dimension to be  $d = 3$ . As a first step, suppose we have an ensemble of LC molecules in a small region where the state of each molecule is defined by its orientation in  $\mathbb{R}^3$ , i.e. let  $\mathbf{p} \in \mathbb{S}^2$  be a vector in the unit sphere that indicates the orientation of an LC molecule. Let  $\rho(\mathbf{p}) \geq 0$  be the probability distribution of the orientation of LC molecules. Obviously, the distribution of LC molecules may vary in space (and in time), e.g.  $\rho \equiv \rho(\mathbf{x}, \mathbf{p})$ , but we shall omit the dependence on these variables for simplicity of notation. It is reasonable to assume that an LC molecule is just as likely to be observed with orientation  $\mathbf{p}$  as  $-\mathbf{p}$ ; hence,  $\rho$  satisfies  $\rho(\mathbf{p}) = \rho(-\mathbf{p})$ . Based on this *head-to-tail symmetry*, it must be that

$$\int_{\mathbb{S}^2} \mathbf{p} \rho(\mathbf{p}) d\mathbf{p} = \mathbf{0}, \quad (1)$$

so direct averaging does not yield a useful order parameter.

Therefore, the first nontrivial information on the molecule distribution is given by the second moments of  $\rho$ , namely:

$$\int_{\mathbb{S}^2} \mathbf{p} \mathbf{p}^T \rho(\mathbf{p}) d\mathbf{p} = \mathbf{M}, \quad (2)$$

where  $\mathbf{M} \in \mathbb{R}^{3 \times 3}$  is symmetric and effectively captures the average state of the LC molecules (i.e. this is a useful order parameter). For a uniformly random (isotropic) distribution of LC molecules,  $\rho(\mathbf{p}) = 1/(4\pi)$  and  $\mathbf{M} = \mathbf{M}_{\text{iso}} = (1/3)\mathbf{I}$ . Thus, it is convenient to define an auxiliary matrix

$$\mathbf{Q} := \mathbf{M} - \mathbf{M}_{\text{iso}}, \quad (3)$$

which is symmetric and traceless, i.e.

$$\mathbf{Q} \in \Lambda := \left\{ \mathbf{Q} \in \mathbb{R}^{3 \times 3} \mid \mathbf{Q} = \mathbf{Q}^T, \text{tr } \mathbf{Q} = 0 \right\}. \quad (4)$$

Clearly, for an isotropic distribution of LC molecules,  $\mathbf{Q} = \mathbf{0}$ . However, as pointed out in Virga (1994, Sec. 1.3.4),  $\mathbf{Q} = \mathbf{0}$  is a necessary but not sufficient condition for isotropy. For the sake of obtaining a continuum theory in which microscopic order is described by  $\mathbf{Q}$  only, we shall regard all distributions satisfying  $\mathbf{Q} = \mathbf{0}$  as isotropic.

We can further characterize  $\mathbf{Q}$  by its eigenframe and is often written in the form:

$$\mathbf{Q} = s_1(\mathbf{n}_1 \otimes \mathbf{n}_1) + s_2(\mathbf{n}_2 \otimes \mathbf{n}_2) - \frac{1}{3}(s_1 + s_2)\mathbf{I}, \quad (5)$$

where  $\mathbf{n}_1, \mathbf{n}_2$  are orthonormal eigenvectors of  $\mathbf{Q}$ , with eigenvalues given by

$$\lambda_1 = \frac{2s_1 - s_2}{3}, \quad \lambda_2 = \frac{2s_2 - s_1}{3}, \quad \lambda_3 = -\frac{s_1 + s_2}{3}, \quad (6)$$

where  $\lambda_3$  corresponds to the eigenvector  $\mathbf{n}_3 \perp \mathbf{n}_1, \mathbf{n}_2$ . The eigenvalues of  $\mathbf{Q}$  are constrained by

$$-\frac{1}{3} \leq \lambda_i = \int_{\mathbb{S}^2} (\mathbf{p} \cdot \mathbf{n}_i)^2 \rho(\mathbf{p}) d\mathbf{p} - \frac{1}{3} \leq \frac{2}{3}, \quad i = 1, 2, 3. \quad (7)$$

When all eigenvalues are equal, since  $\mathbf{Q}$  is traceless, we must have  $\lambda_1 = \lambda_2 = \lambda_3 = 0$  and  $s_1 = s_2 = 0$ , i.e. the distribution of LC molecules is isotropic. If two eigenvalues are equal, i.e.

$$\begin{aligned} \lambda_1 = \lambda_2 &\Leftrightarrow s_1 = s_2, \\ \lambda_1 = \lambda_3 &\Leftrightarrow s_1 = 0, \\ \lambda_2 = \lambda_3 &\Leftrightarrow s_2 = 0, \end{aligned} \quad (8)$$

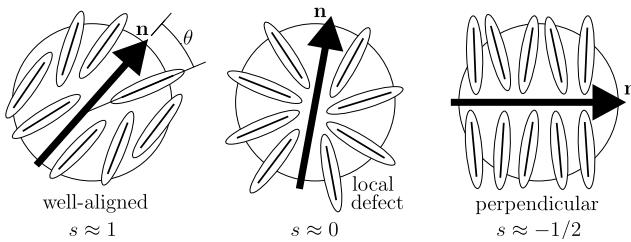
then we encounter a uniaxial state, in which either molecules prefer to orient in alignment with the simple eigenspace (in case it corresponds to a positive eigenvalue) or perpendicular to it (in case it corresponds to a negative eigenvalue). If all three eigenvalues are distinct, then the state is called biaxial. We recall that, as discussed in Section 1.1, most nematic LCs can be effectively modeled as uniaxial states.

Let us consider a uniaxial state for  $\mathbf{Q}$ , that can be written in the equivalent form

$$\mathbf{Q} = s \left( \mathbf{n} \otimes \mathbf{n} - \frac{1}{3} \mathbf{I} \right). \quad (9)$$

Above,  $\mathbf{n}$  is the main eigenvector with eigenvalue  $\lambda = 2s/3$ ; the other two eigenvalues equal  $-s/3$ . The scalar field  $s$  is called the *degree of orientation* of the LC molecules. Taking into account identity (6) and the restriction (7), it follows that the physically meaningful range is  $s \in [-1/2, 1]$ . As Fig. 1 illustrates, the  $s$  variable characterizes the local order. In case  $s = 1$ , the molecular long axes are in perfect alignment with the direction of  $\mathbf{n}$ , whereas  $s = -1/2$  represents the state in which all molecules are perpendicular to  $\mathbf{n}$ .

**Remark 1** (Problems in 2d). *The discussion above simplifies considerably when  $d = 2$ . Indeed, the fact that  $\mathbf{Q}$  is a zero-trace tensor forces it to be uniaxial. Writing  $\mathbf{Q}$  as (9), we deduce that its eigenvalues are  $\lambda_1 = s/2$ , with eigenvector  $\mathbf{n}$ , and  $\lambda_2 = -\lambda_1$ , with eigenvector  $\mathbf{n}^\perp$ . Because eigenvalues are constrained to satisfy  $\lambda_i \in (-1/2, 1/2)$ , we deduce that the physically meaningful range is  $s \in (-1, 1)$ . Actually, one can further simplify to  $s \in [0, 1]$  by noting that a state with director  $\mathbf{n}$  and degree of orientation  $s < 0$  is equivalent to a state with director  $\mathbf{m} \perp \mathbf{n}$  and degree of orientation  $-s$ .*



**FIGURE 1** The degree-of-orientation variable  $s(x)$  provides information on the probability distribution  $\rho(x, \cdot)$  over a local ensemble. The case  $s = 1$  corresponds to a Dirac delta distribution, and represents the state of perfect alignment in which all molecules in the ensemble are parallel to  $\mathbf{n}(x)$ . Likewise,  $s = -1/2$  represents the state in which  $\rho(x, \cdot)$  is a uniform distribution on the unit circle perpendicular to  $\mathbf{n}(x)$ . Finally, when  $s(x) = 0$ ,  $\rho(x, \cdot)$  is a uniform distribution on  $\mathbb{S}^2$ ; such a state is regarded as a *defect* in the LC.

## 2.2 Continuum mechanics

Given the order parameter, we still need a model to determine its state as a function of space. For modeling equilibrium states, this amounts to finding minimizers of an energy functional. A common approach from continuum mechanics (Holzapfel, 2000; Temam and Miranville, 2005; Truesdell, 1976) is to construct the “simplest” functional possible that is quadratic in gradients of the order parameter whilst obeying standard laws of physics, such as frame indifference and material symmetries.

### 2.2.1 Oseen-Frank

If we assume  $\mathbf{Q}$  has the form (9), and assume  $s$  is constant, then  $\mathbf{Q}$  is fully determined by  $\mathbf{n}$ . Hence, we may take  $\mathbf{n}$  to be the order parameter, which is usually called the *director*. If we now seek the simplest energy functional that is quadratic in gradients of  $\mathbf{n}$ , and obeying symmetry relations such as  $\mathbf{n} \equiv -\mathbf{n}$ , then we obtain the Oseen-Frank energy (Virga, 1994):

$$\begin{aligned} E_{\text{OF}}[\mathbf{n}] &:= \frac{1}{2} \int_{\Omega} \mathcal{W}_{\text{OF}}(\mathbf{n}, \nabla \mathbf{n}) d\mathbf{x}, \\ \mathcal{W}_{\text{OF}}(\mathbf{n}, \nabla \mathbf{n}) &:= k_1(\text{div } \mathbf{n})^2 + k_2(\mathbf{n} \cdot \text{curl } \mathbf{n})^2 + k_3|\mathbf{n} \times \text{curl } \mathbf{n}|^2 \\ &\quad + (k_2 + k_4)[\text{tr}((\nabla \mathbf{n})^2) - (\text{div } \mathbf{n})^2], \end{aligned} \quad (10)$$

where  $|\mathbf{n}| = 1$  and  $\{k_i\}_{i=1}^4$  are independent material parameters. Note that minimizing  $E_{\text{OF}}$  subject to  $|\mathbf{n}| = 1$  is a non-convex optimization problem. We give more discussion on mathematical issues, such as regularity of minimizers, in Section 3. The Oseen-Frank model has been used extensively in the modeling of LC-based flat panel pixel displays, so in that sense has been very successful. As a simplification, one can take  $k_1 = k_2 = k_3 = 1, k_4 = 0$  to obtain

$$E_{\text{OF,one}}[\mathbf{n}] := \frac{1}{2} \int_{\Omega} |\nabla \mathbf{n}|^2 d\mathbf{x}, \quad (11)$$

which is known as the *one-constant approximation*.

There are two main drawbacks to using (10) or (11). First, the state of the LC molecules is unaffected by the sign of  $\mathbf{n}$ . For example, if  $\mathbf{n} = \mathbf{n}(\mathbf{x})$  is a minimizer of  $E_{\text{OF}}$ , then arbitrarily changing the sign of  $\mathbf{n}$  on any subset of the domain  $\Omega$  does not affect the state of the LC molecules. However, changing the sign of  $\mathbf{n}$  at points arbitrarily close together leads to very large gradients in  $\mathbf{n}$ , which of course affects the energy. Thus, the energy in (10) does not fully respect the basic symmetry condition  $\mathbf{n} \equiv -\mathbf{n}$  of nematic LCs.

However, even allowing for the smoothest possible configurations of the director, boundary conditions may force another problem. For example, suppose  $\Omega$  is a simply connected open set containing the origin, and suppose we fix  $\mathbf{n} = \mathbf{x}/|\mathbf{x}|$  on the boundary  $\partial\Omega$ . Then, by the Poincaré-Hopf Theorem, every smooth vector field  $\mathbf{v}$  in  $\Omega$  that coincides with  $\mathbf{n}$  on  $\partial\Omega$  must have at least a zero with non-zero index. Therefore, the unit-length vector field  $\mathbf{n} = \mathbf{v}/|\mathbf{v}|$  must have a point of discontinuity. Moreover, if  $\Omega \subset \mathbb{R}^2$ , then  $E_{\text{OF,one}}[\mathbf{n}] = \infty$  by basic arguments. Since defects naturally occur in many LC systems, this is a major problem with the Oseen-Frank model.

If  $\Omega \subset \mathbb{R}^3$ , then  $E_{\text{OF,one}}[\mathbf{n}]$  for a point defect is actually finite. However, for line defects in  $\mathbb{R}^3$  such as

$$\mathbf{n}(x_1, x_2, x_3) = \frac{(x_1, x_2, 0)^T - (a_1, a_2, 0)^T}{|(x_1, x_2, 0) - (a_1, a_2, 0)|}, \quad (12)$$

which is a two dimensional point defect extruded in the  $x_3$ -direction, one has  $E_{\text{OF,one}}[\mathbf{n}] = \infty$ . The same holds true for curvilinear defects in  $\mathbb{R}^3$ .

**Remark 2.** *This discussion on defects is not merely academic; defects do occur in LC systems. For example, Gu and Abbott (2000) gives experimental evidence for the “Saturn-ring” defect which is a closed curvilinear loop in  $\mathbb{R}^3$  that surrounds a rigid inclusion. Other examples of LC defects can be found in Dierking et al. (2005), Goodby (2012), Kinderlehrer et al. (1993), Tojo et al. (2009), Čopar et al. (2015).*

## 2.2.2 Landau-de Gennes

When defects are relevant to an LC system, the Oseen-Frank model is not appropriate. A better model, using  $\mathbf{Q}$  as the order parameter, is the Landau-de Gennes energy (de Gennes and Prost, 1995; Sonnet and Virga, 2012):

$$\begin{aligned} E_{\text{LdG}}[\mathbf{Q}] &:= \int_{\Omega} \mathcal{W}_{\text{LdG}}(\mathbf{Q}, \nabla \mathbf{Q}) d\mathbf{x} + \frac{1}{\eta_B} \int_{\Omega} \psi_{\text{LdG}}(\mathbf{Q}) d\mathbf{x}, \\ \mathcal{W}_{\text{LdG}}(\mathbf{Q}, \nabla \mathbf{Q}) &:= \frac{1}{2} \left( L_1 |\nabla \mathbf{Q}|^2 + L_2 |\nabla \cdot \mathbf{Q}|^2 + L_3 (\nabla \mathbf{Q})^T : \nabla \mathbf{Q} \right), \end{aligned} \quad (13)$$

where  $\{L_i\}_{i=1}^3$ ,  $\eta_B$  are material parameters,  $\psi_{\text{LdG}}$  is a “bulk” (thermotropic) potential and

$$|\nabla \mathbf{Q}|^2 := (\partial_k Q_{ij})(\partial_k Q_{ij}), \quad |\nabla \cdot \mathbf{Q}|^2 := (\partial_j Q_{ij})^2, \quad (\nabla \mathbf{Q})^T : \nabla \mathbf{Q} := (\partial_j Q_{ik})(\partial_k Q_{ij}), \quad (14)$$

and we use the convention of summation over repeated indices. This is a relatively simple form for  $\mathcal{W}_{\text{LdG}}$ ; more complicated models can also be considered (Mottram and Newton, 2014; de Gennes and Prost, 1995; Sonnet and Virga, 2012).

The bulk potential  $\psi_{\text{LdG}}$  is a double-well type of function that confines the eigenvalues of  $\mathbf{Q}$  to the physically meaningful range  $\lambda_i \in [-1/d, 1 - 1/d]$ , where the simplest form is given by

$$\psi_{\text{LdG}}(\mathbf{Q}) = K + \frac{A}{2} \text{tr}(\mathbf{Q}^2) - \frac{B}{3} \text{tr}(\mathbf{Q}^3) + \frac{C}{4} \left( \text{tr}(\mathbf{Q}^2) \right)^2. \quad (15)$$

Above,  $A, B, C$  are material parameters such that  $A$  has no sign, and  $B, C$  are positive;  $K$  is a convenient constant.

Stationary points of  $\psi_{\text{LdG}}$  are either uniaxial or isotropic  $\mathbf{Q}$ -tensors (Majumdar, 2010); for such,  $\psi_{\text{LdG}}$  is a quartic polynomial on the degree of orientation  $s$  in (9), which has a local extremum at  $s = 0$ . A straightforward calculation shows that  $s = 0$  is a maximum if and only if  $A \leq 0$  (because  $B, C > 0$ ). Thus, in three dimensions it is typical to let  $A \leq 0$  in order to favor uniaxial states over isotropic states, so throughout this paper we assume that

$$A \leq 0, \quad B, C > 0, \quad (16)$$

which implies that  $\psi_{\text{LdG}}(\mathbf{Q}) \geq 0$  assuming  $K$  is suitably chosen. In two dimensions,  $\text{tr}(\mathbf{Q}^3) = 0$ , because  $\mathbf{Q}^2 = \frac{s^2}{4} \mathbf{I}$ . Hence,  $B$  is irrelevant when  $d = 2$ , and it is *necessary* that  $A$  be strictly negative in order to have a stable nematic phase. This also implies that  $\psi_{\text{LdG}}$  is an *even* function of  $s$  if  $\mathbf{Q}$  is uniaxial (see Remark 1).

In the same spirit as in (11), one can take  $L_1 = 1, L_2 = L_3 = 0$  to obtain a *one-constant approximation*

$$E_{\text{LdG,one}}[\mathbf{Q}] := \frac{1}{2} \int_{\Omega} |\nabla \mathbf{Q}|^2 d\mathbf{x} + \frac{1}{\eta_B} \int_{\Omega} \psi_{\text{LdG}}(\mathbf{Q}) d\mathbf{x}. \quad (17)$$

### 2.2.3 Remarks on uniaxiality

As we discussed in Section 1.1, the biaxial phase is elusive among thermotropic nematic LCs: it took about 30 years after Freiser's prediction (Freiser, 1970) to empirically observe a nematic biaxial phase (Acharya et al., 2004; Madsen et al., 2004; Prasad et al., 2005). Moreover, in the Landau-de Gennes theory, there is no *a priori* constraint on the eigenvalues of the tensor  $\mathbf{Q}$ , in contrast with the probabilistic definition from (2) and (3).

In Majumdar (2010) it is shown that, in the low-temperature regime, the Landau-de Gennes model can lead to  $\mathbf{Q}$  having physically unrealistic eigenvalues. As a remedy, Ball and Majumdar (2010) propose a continuum energy

functional that interpolates between the Landau-de Gennes energy and the mean-field Maier-Saupe energy. This gives rise to the bulk potential

$$\psi_B(\mathbf{Q}) := T \inf_{\rho \in \mathcal{A}_Q} \int_{\mathbb{S}^{d-1}} \rho(\mathbf{p}) \ln \rho(\mathbf{p}) d\mathbf{p} - \kappa |\mathbf{Q}|^2,$$

where  $T$  denotes the absolute temperature,  $\kappa$  is a constant related to the strength of intramolecular interactions and  $\mathcal{A}_Q$  is the set of probability distributions that yield the tensor  $\mathbf{Q}$ ,

$$\begin{aligned} \mathcal{A}_Q = \left\{ \rho : \mathbb{S}^{d-1} \rightarrow [0, \infty), \int_{\mathbb{S}^{d-1}} \rho(\mathbf{p}) d\mathbf{p} = 1, \right. \\ \left. \mathbf{Q} = \int_{\mathbb{S}^{d-1}} \left( \mathbf{p} \mathbf{p}^T - \frac{1}{d} \mathbf{I} \right) \rho(\mathbf{p}) d\mathbf{p} \right\}. \end{aligned}$$

This potential satisfies the key property  $\psi_B(\mathbf{Q}) \rightarrow \infty$  as any of the eigenvalues  $\lambda_i$  approaches the boundary of the physically meaningful range.

Clearly, if uniaxiality is built into the model, for instance by forcing  $\mathbf{Q}$  to have the form (9), then keeping the eigenvalues within the physically meaningful range reduces to guaranteeing that a single parameter ( $s$ ) lies in a suitable range. This is an important simplification if, for example, the energy has the form (13) or if there is a large external forcing. Clearly, another approach to enforce uniaxiality is to consider a director model. If we allow for a variable degree of orientation, then we are led to the Ericksen model, that we discuss in the next section.

#### 2.2.4 Ericksen model

Though the Landau-de Gennes model is quite general, it can be fairly expensive when  $d = 3$ . In such a case, since  $\mathbf{Q} \in \mathbb{R}^{3 \times 3}$  and symmetric, it has five independent variables. Moreover, the bulk potential  $\psi_{\text{LdG}}$  is a non-linear function of  $\mathbf{Q}$ , which couples all five variables together when seeking a minimizer of  $E_{\text{LdG}}$ .

Therefore, we present the Ericksen model of LCs, which is an intermediate model between Oseen-Frank and Landau-de Gennes. Assuming that  $\mathbf{Q}$  is uniaxial (9), we can take  $s$  and  $\mathbf{n}$  as order parameters and obtain an energy analogous to (10) (Ericksen, 1991; Virga, 1994):

$$\begin{aligned} E_{\text{erk}}[s, \mathbf{n}] := \frac{1}{2} \int_{\Omega} \mathcal{W}_{\text{erk}}(s, \nabla s, \mathbf{n}, \nabla \mathbf{n}) d\mathbf{x} + \frac{1}{\eta_B} \int_{\Omega} \psi_{\text{erk}}(s) d\mathbf{x}, \\ \mathcal{W}_{\text{erk}}(s, \nabla s, \mathbf{n}, \nabla \mathbf{n}) = k_1 s^2 (\text{div } \mathbf{n})^2 + k_2 s^2 (\mathbf{n} \cdot \text{curl } \mathbf{n})^2 + k_3 s^2 |\mathbf{n} \times \text{curl } \mathbf{n}|^2 \\ + (k_2 + k_4) s^2 [\text{tr}(\nabla \mathbf{n})^2 - (\text{div } \mathbf{n})^2] + b_1 |\nabla s|^2 + b_2 (\nabla s \cdot \mathbf{n})^2 \\ + b_3 s (\text{div } \mathbf{n}) (\nabla s \cdot \mathbf{n}) + b_4 s \nabla s \cdot [\nabla \mathbf{n}] \mathbf{n}, \end{aligned} \tag{18}$$

where  $|\mathbf{n}| = 1$ , and  $\{k_i\}_{i=1}^4$  and  $\{b_i\}_{i=1}^4$  are material constants. Moreover, we have the *one-constant* version of (18):

$$E_{\text{erk,one}}[s, \mathbf{n}] := \frac{\kappa}{2} \int_{\Omega} |\nabla s|^2 d\mathbf{x} + \frac{1}{2} \int_{\Omega} s^2 |\nabla \mathbf{n}|^2 d\mathbf{x} + \frac{1}{\eta_B} \int_{\Omega} \psi_{\text{erk}}(s) d\mathbf{x}, \quad (19)$$

where  $\kappa > 0$  is a single material parameter, and  $\psi_{\text{erk}}$  is a double-well potential like  $\psi_{\text{LdG}}$ , except acting on  $s$ .

We point out that both (18) and (19) are *degenerate* in the sense that  $s$  may vanish, which allows for  $\mathbf{n}$  to have discontinuities (i.e. defects) with finite energy. Indeed, the hallmark of this model is to regularize defects using  $s$ , but still retain part of the Oseen-Frank model. Discontinuities in  $\mathbf{n}$  may still occur in the singular set

$$\mathcal{S} := \{x \in \Omega : s(x) = 0\}. \quad (20)$$

For problems in  $\mathbb{R}^3$ , because  $\mathbf{n} \in \mathbb{S}^2$ , it is uniquely defined by two parameters. Thus, in such a case the Ericksen model only has three scalar order parameters, as opposed to five in the Landau-de Gennes model. Another advantage of the Ericksen model is that  $s$  and  $\mathbf{n}$  are easy to decouple when searching for a minimizer numerically.

Additionally, the parameter  $\kappa$  in (19) plays a major role in the occurrence of defects. Assuming that  $s$  equals a sufficiently large positive constant on  $\partial\Omega$ , if  $\kappa$  is large, then  $\int_{\Omega} \kappa |\nabla s|^2 dx$  dominates the energy and  $s$  stays close to such a positive constant within the domain  $\Omega$ . Thus, defects are less likely to occur. If  $\kappa$  is small (say  $\kappa < 1$ ), then  $\int_{\Omega} s^2 |\nabla \mathbf{n}|^2 dx$  dominates the energy, and  $s$  may vanish in regions of  $\Omega$  and induce a defect. This is confirmed by the numerical experiments in Nochetto et al. (2015, 2017).

**Remark 3** (Orientability). *Director field models –either Oseen-Frank or Ericksen– are more than adequate in some situations, although in general they introduce a nonphysical orientational bias into the problem. Even though LC molecules may be polar, in nematics one always finds that the states with  $\mathbf{n}$  and  $-\mathbf{n}$  are equivalent (Gramsbergen et al., 1986). At the molecular level, this means that the same number of molecules point “up” and “down”. Therefore, line-fields are more appropriate for modeling nematic LCs.*

*Another issue with the use of the vector field  $\mathbf{n}$  as an order parameter instead of the matrix  $\mathbf{Q}$  is that the only allowable defects in such a case are integer-order defects. On the other hand  $\mathbf{Q}$ , specifically  $\mathbf{n} \otimes \mathbf{n}$  in (9), is able to represent line fields having half-integer defects. These have been largely observed and documented in experiments, see for example (Brinkman and Cladis, 1982; Ohzono et al., 2017) and references therein. We point out that, if a line field is orientable, then a vector field representation is essentially equivalent (Ball and Zarnescu, 2007, 2011).*

## 2.3 Dynamics

Dynamic LC models become more complicated than the ones discussed in Section 2.2. The simplest setting is to assume the dynamics are dictated by a gradient flow (Burger, 2002; Doğan et al., 2007; Braides, 2014). Let  $t$  represent “time” and suppose that  $u \equiv u(\mathbf{x}, t)$  is an evolving solution such that  $\lim_{t \rightarrow \infty} u(\cdot, t) =: u_*$  is a local minimizer of some energy functional  $J(u)$  that is bounded below by a constant. If  $u(\mathbf{x}, 0) := u_0$  is the initial guess, then the energy minimizing evolution can be found via (steepest) gradient descent:

$$(\partial_t u(\cdot, t), v) = -\delta_u J(u; v), \quad (21)$$

for all perturbations  $v$  in an appropriate space, where  $(\cdot, \cdot)$  is the  $L^2(\Omega)$  inner product, and  $\delta_u J(u; v)$  is the variational derivative of  $J(u)$  with respect to  $u$  in the direction  $v$ . Formally, the solution of (21) will converge to a local minimizer of  $J$  depending on the initial guess  $u_0$  (Doğan et al., 2007).

We can approximate this evolution by a time semi-discrete scheme known as minimizing movements (see De Giorgi, 2006, Ch. “New problems on minimizing movement”, Braides, 2014, Ch. 7). Discretizing in time, we let  $u_k(\mathbf{x}) \approx u(\mathbf{x}, k\delta t)$ , where  $\delta t > 0$  is a finite time step and  $k$  is the time index. Next, define an auxiliary functional

$$F_k(u) := J(u) + \frac{1}{2\delta t} \|u - u_k\|_{\Omega}^2. \quad (22)$$

Treating  $u_0$  as given, setting  $\delta t > 0$ , and initializing  $k = 0$ , we iterate the following scheme:

1. Let  $u_{k+1} := \arg \min F_k(u)$ .
2. Update  $k := k + 1$ .

At each iteration, we solve the variational problem  $\delta_u F_k(u; v) = 0$  for all  $v$ , i.e.

$$\left( \frac{u_{k+1} - u_k}{\delta t}, v \right) = -\delta_u J(u_{k+1}; v), \quad \forall v, \quad (23)$$

which is a backward Euler discretization of (21). The following properties of this scheme are immediate:

$$\begin{aligned} J(u_{k+1}) &\leq J(u_{k+1}) + \frac{1}{2\delta t} \|u_{k+1} - u_k\|_{\Omega}^2 = \min_u F_k(u) \leq J(u_k), \\ \|u_{k+1} - u_k\|_{\Omega}^2 &\leq -2\delta t (J(u_{k+1}) - J(u_k)). \end{aligned} \quad (24)$$

Both (21) and the minimizing movement scheme may be applied to any of the energy functionals we have discussed.

More realistic dynamics can be derived by generalizing the  $L^2(\Omega)$  inner product and coupling in other PDE constraints (such as Stokes flow) using Onsager’s variational framework of *minimum energy dissipation* (Onsager,

1931a,b; Lin and Liu, 1995, 2001; Qian et al., 2006; Hyon et al., 2010). We refer the interested reader to Liu and Walkington (2000), Walkington (2011), Cruz et al. (2013), González and Gutiérrez-Santacreu (2013), Yang et al. (2013), Zhao et al. (2016) for more information on dynamical theories for LCs.

### 3 Mathematical formulation

We revisit the LC models discussed in Section 2.2 with an emphasis on their mathematical formulation. Functional minimization must include an admissible set in which to find the minimizer. In other words, we describe the function spaces over which to minimize the energies given in Section 2.2, as well as other mathematical issues. Although often overlooked, the function space is an important part of a continuum mathematical model (Ball, 2017; Ball et al., 2015). For example, Lavrentiev gap phenomena between Sobolev and special bounded variation (SBV) functions in certain nematic LC models are examined in Bedford (2016).

#### 3.1 Oseen-Frank

Taking the first variation of  $E_{\text{OF}}$  and setting to zero yields the first order optimality conditions, i.e. the PDE satisfied by a minimizer. For simplicity, let us consider the one-constant energy  $E_{\text{OF},\text{one}}$  whose minimization problem is as follows

$$\min_{\mathbf{v} \in H_{\mathbf{g}}^1(\Omega)} E_{\text{OF},\text{one}}[\mathbf{v}], \quad \text{subject to } |\mathbf{v}| = 1, \text{ a.e.}, \quad (25)$$

where  $H_{\mathbf{g}}^1(\Omega) = \{\mathbf{v} \in H^1(\Omega) \mid \mathbf{v}|_{\partial\Omega} = \mathbf{g}\}$ . This is an instance of the *harmonic map* problem (Schoen and Yau, 1994; Schoen and Uhlenbeck, 1982; Brezis et al., 1986; Brezis, 2003; Jost, 2017; Struwe, 2008). The Euler-Lagrange equation for a minimizer of (25), in strong form, is given by

$$-\Delta \mathbf{n} - |\nabla \mathbf{n}|^2 \mathbf{n} = \mathbf{0}, \text{ in } \Omega, \quad \mathbf{n} = \mathbf{g}, \text{ on } \partial\Omega. \quad (26)$$

The existence of minimizers is complicated if defects are present, unless the boundary data  $\mathbf{g}$  is sufficiently restricted (recall the discussion in Section 2.2.1).

There is an extensive literature on numerical methods to find a minimizer of (25), e.g. Hu et al. (2009), Cohen et al. (1989), Lin and Luskin (1989), Alouges (1997), Bartels (2006, 2015), all of which solve the non-convex minimization problem iteratively. Their main contribution is in how they handle the unit length constraint,  $|\mathbf{n}| = 1$ , at each iteration. For instance, Lagrange multipliers may be used by solving the linearized Euler-Lagrange equation in a saddle point framework (Hu et al., 2009). Alternatively, one can project the current (iterative) solution onto the constraint manifold (see the algorithm in Section 3.3.4 for an example), which usually takes advantage of a discrete maximum principle that

is built into the method (Cohen et al., 1989; Alouges, 1997; Bartels, 2015). One can find more recent methods for Oseen-Frank type models coupled to other physics in Adler et al. (2015a,b, 2016), Gartland and Ramage (2015).

For this paper, our main interest is in modeling defects such that they have *finite* energy, so we will not discuss more on the extensive literature of harmonic maps and their numerical approximation.

## 3.2 Landau-de Gennes

We follow Davis and Gartland (1998) to outline the basic theory of the Landau-de Gennes model. We also write down a numerical method to simulate Landau-de Gennes motivated by Bajc et al. (2016), Davis and Gartland (1998), Zhao and Wang (2016).

### 3.2.1 Theoretical background

First, we define the function space for  $\mathbf{Q}$  when seeking a minimizer:

$$\mathbb{V}(\mathbf{P}) := \left\{ \mathbf{Q} \in H^1(\Omega) \mid \mathbf{Q} \in \Lambda \text{ a.e. in } \Omega, \mathbf{Q}|_{\Gamma_D} = \mathbf{P} \right\}, \quad (27)$$

where  $\Lambda$  is defined by (4),  $\Gamma_D \subset \Gamma$  and  $\mathbf{P} \in H^1(\Omega)$  is arbitrary such that  $\mathbf{P}(\mathbf{x}) \in \Lambda$  for a.e.  $\mathbf{x} \in \Omega$ . For the sake of generality, we slightly modify the energy  $E_{\text{LdG}}$  in (13):

$$\begin{aligned} E_{\text{LdG}}[\mathbf{Q}] := & \int_{\Omega} \mathcal{W}_{\text{LdG}}(\mathbf{Q}, \nabla \mathbf{Q}) d\mathbf{x} + \frac{1}{\eta_B} \int_{\Omega} \psi_{\text{LdG}}(\mathbf{Q}) d\mathbf{x} \\ & + \eta_{\Gamma} \int_{\Gamma} f_{\Gamma}(\mathbf{Q}) dS(\mathbf{x}) - \int_{\Omega} \chi_{\text{LdG}}(\mathbf{Q}) d\mathbf{x}, \end{aligned} \quad (28)$$

where  $\mathcal{W}_{\text{LdG}}$  is given in (13),  $\psi_{\text{LdG}}$  is given in (15),  $\eta_{\Gamma} \geq 0$ , and a Rapini-Papoulier type anchoring energy (Barbero and Durand, 1986) is used:

$$f_{\Gamma}(\mathbf{Q}) = \frac{1}{2} \text{tr}(\mathbf{Q} - \mathbf{Q}_{\Gamma})^2 \equiv \frac{1}{2} |\mathbf{Q} - \mathbf{Q}_{\Gamma}|^2, \quad (29)$$

where  $\mathbf{Q}_{\Gamma} \in H^1(\Omega)$  is given and  $\mathbf{Q}_{\Gamma}(\mathbf{x}) \in \Lambda$  for all  $\mathbf{x} \in \Omega$ . The extra term involving  $f_{\Gamma}$  gives an energetic way to penalize boundary conditions, provided  $\eta_{\Gamma}$  is large enough.

The functional  $\chi_{\text{LdG}}(\cdot)$  accounts for external forcing effects, e.g. from an electric field. For example, the energy density of a dielectric with fixed boundary potential is given by  $-1/2 \mathbf{D} \cdot \mathbf{E}$  (Walker, 2018), where the electric displacement  $\mathbf{D}$  is related to the electric field  $\mathbf{E}$  by the linear constitutive law (Feynman et al., 1964; de Gennes and Prost, 1995; Biscari and Cesana, 2007):

$$\mathbf{D} = \boldsymbol{\epsilon} \mathbf{E} = \bar{\epsilon} \mathbf{E} + \epsilon_a \mathbf{Q} \mathbf{E}, \quad \boldsymbol{\epsilon}(\mathbf{Q}) = \bar{\epsilon} \mathbf{I} + \epsilon_a \mathbf{Q}, \quad (30)$$

where  $\mathbf{e}$  is the LC material's dielectric tensor and  $\bar{\varepsilon}$ ,  $\varepsilon_a$  are constitutive dielectric permittivities. Thus, in the presence of an electric field,  $\chi_{\text{LdG}}(\cdot)$  becomes

$$\chi_{\text{LdG}}(\mathbf{Q}) = -\frac{1}{2}\mathbf{D} \cdot \mathbf{E} = -\frac{1}{2} \left[ \bar{\varepsilon}|\mathbf{E}|^2 + \varepsilon_a \mathbf{E} \cdot \mathbf{Q} \mathbf{E} \right]. \quad (31)$$

The minimization problem for the Landau-de Gennes energy functional (28) is as follows

$$\min_{\mathbf{Q} \in \mathbb{V}(\mathbf{Q}_D)} E_{\text{LdG}}[\mathbf{Q}], \quad (32)$$

where  $\mathbf{Q}_D \in H^1(\Omega)$  is given and  $\mathbf{Q}_D(\mathbf{x}) \in \Lambda$  for a.e.  $\mathbf{x} \in \Omega$ . This minimization problem is not as delicate as (25); for instance, there is no non-convex, unit-length constraint. Existence of a minimizer is guaranteed by the following results (taken from Davis and Gartland, 1998, Lem. 4.1).

**Theorem 1** (Coercivity). *Let  $a_{\text{LdG}}(\cdot, \cdot) : H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$  be the symmetric bilinear form defined by*

$$a_{\text{LdG}}(\mathbf{Q}, \mathbf{P}) = \int_{\Omega} L_1 \nabla \mathbf{Q} : \nabla \mathbf{P} + L_2 (\nabla \cdot \mathbf{Q}) \cdot (\nabla \cdot \mathbf{P}) + L_3 (\nabla \mathbf{Q})^T : \nabla \mathbf{P} d\mathbf{x}. \quad (33)$$

*Then  $a_{\text{LdG}}(\cdot, \cdot)$  is bounded. If  $L_1, L_2, L_3$  satisfy*

$$0 < L_1, \quad -L_1 < L_3 < 2L_1, \quad -\frac{3}{5}L_1 - \frac{1}{10}L_3 < L_2, \quad (34)$$

*then there is a constant  $C > 0$  such that  $a_{\text{LdG}}(\mathbf{Q}, \mathbf{Q}) \geq C|\mathbf{Q}|_{H^1(\Omega)}^2$  for all  $\mathbf{Q} \in H^1(\Omega)$ . Moreover, if  $|\Gamma_D| > 0$ , then there is a constant  $C' > 0$  such that  $a_{\text{LdG}}(\mathbf{Q}, \mathbf{Q}) \geq C' \|\mathbf{Q}\|_{H^1(\Omega)}^2$  for all  $\mathbf{Q} \in \mathbb{V}(\mathbf{0})$ .*

Combining Theorem 1 with the form of the energy in (28) and other basic results (see Davis and Gartland, 1998, Lem. 4.2, Thm. 4.3) we arrive at the following result.

**Theorem 2** (Existence of a minimizer). *Let  $E_{\text{LdG}}$  be of the form (28), and assume that  $\mathbf{Q}_D$  and  $\Gamma_D$  are defined as above and that  $\chi_{\text{LdG}}$  is a bounded linear functional on  $\mathbb{V}(\mathbf{Q}_D)$ . Then (32) has at least one minimizer.*

The Euler-Lagrange equation for a minimizer of (32), in weak form, is as follows. Find  $\mathbf{Q} \in \mathbb{V}(\mathbf{Q}_D)$  such that

$$\begin{aligned} \delta_{\mathbf{Q}} E_{\text{LdG}}[\mathbf{Q}; \mathbf{P}] &:= a_{\text{LdG}}(\mathbf{Q}, \mathbf{P}) \\ &+ \frac{1}{\eta_B} \int_{\Omega} \frac{\partial \psi_{\text{LdG}}(\mathbf{Q})}{\partial \mathbf{Q}} : \mathbf{P} d\mathbf{x} + \eta_{\Gamma} \int_{\Gamma} \frac{\partial f_{\Gamma}(\mathbf{Q})}{\partial \mathbf{Q}} : \mathbf{P} dS(\mathbf{x}) \\ &- \int_{\Omega} \frac{\partial \chi_{\text{LdG}}(\mathbf{Q})}{\partial \mathbf{Q}} : \mathbf{P} d\mathbf{x} = 0, \quad \forall \mathbf{P} \in \mathbb{V}(\mathbf{0}), \end{aligned} \quad (35)$$

where the variational derivatives are given by

$$\begin{aligned}\frac{\partial \psi_{\text{LdG}}(\mathbf{Q})}{\partial \mathbf{Q}} : \mathbf{P} &= \left[ A \mathbf{Q} - B \mathbf{Q}^2 + C \text{tr}(\mathbf{Q}^2) \mathbf{Q} \right] : \mathbf{P}, \\ \frac{\partial f_{\Gamma}(\mathbf{Q})}{\partial \mathbf{Q}} : \mathbf{P} &= [\mathbf{Q} - \mathbf{Q}_{\Gamma}] : \mathbf{P} \\ \frac{\partial \chi_{\text{LdG}}(\mathbf{Q})}{\partial \mathbf{Q}} : \mathbf{P} &= -\frac{1}{2} \varepsilon_a \mathbf{E} \cdot \mathbf{P} \mathbf{E},\end{aligned}\tag{36}$$

where we used

$$\begin{aligned}\frac{\partial \text{tr}(\mathbf{Q}^2)}{\partial \mathbf{Q}} : \mathbf{P} &= 2 \mathbf{Q} : \mathbf{P}, \quad \frac{\partial \text{tr}(\mathbf{Q}^3)}{\partial \mathbf{Q}} : \mathbf{P} = 3 \mathbf{Q}^2 : \mathbf{P}, \\ \frac{\partial (\text{tr}(\mathbf{Q}^2))^2}{\partial \mathbf{Q}} : \mathbf{P} &= 4 \text{tr}(\mathbf{Q}^2) \mathbf{Q} : \mathbf{P}.\end{aligned}\tag{37}$$

The strong form of the Euler-Lagrange equation is given by

$$\begin{aligned}- (L_1 \partial_k \partial_k Q_{ij} + (L_2 + L_3) \partial_j \partial_k Q_{ik}) \\ + \frac{1}{\eta_B} \left[ A Q_{ij} - B Q_{ik} Q_{kj} + C \text{tr}(\mathbf{Q}^2) Q_{ij} \right] &= -\frac{1}{2} \varepsilon_a E_i E_j, \text{ in } \Omega, \\ Q_{ij} &= Q_{0,ij}, \text{ on } \Gamma_D, \\ - (L_1 v_k \partial_k Q_{ij} + L_2 v_j \partial_k Q_{ik} + L_3 v_k \partial_j Q_{ik}) &= Q_{ij} - Q_{\Gamma,ij}, \text{ on } \Gamma \setminus \Gamma_D,\end{aligned}\tag{38}$$

for  $1 \leq i, j \leq 3$ , where  $\mathbf{v} \equiv [v_k]_{k=1}^3$  is the unit outer normal of  $\Gamma$ . If  $L_2 = L_3 = 0$ ,  $\mathbf{E} = \mathbf{0}$ , and  $\Gamma_D = \Gamma$ , then the strong form simplifies to

$$-\Delta \mathbf{Q} + \frac{1}{\eta_B} \left[ A \mathbf{Q} - B \mathbf{Q}^2 + C \text{tr}(\mathbf{Q}^2) \mathbf{Q} \right] = \mathbf{0}, \text{ in } \Omega, \quad \mathbf{Q} = \mathbf{Q}_D, \text{ on } \Gamma_D,\tag{39}$$

which is an elliptic Dirichlet problem with non-linear lower order term due to the bulk potential. From Davis and Gartland (1998, Thm 6.3), we have

**Theorem 3** (Regularity). *Let  $\Omega$  be a bounded, open, connected set, and assume  $\Omega$  is either convex or  $C^{1,1}$ , and assume  $\Gamma_D \equiv \Gamma := \partial\Omega$ . Moreover, let  $F(\mathbf{Q}) := \int_{\Omega} \chi_{\text{LdG}}(\mathbf{Q}) \mathbf{d}x$  be a bounded linear functional on  $L^2(\Omega)$ . Then any solution of (35) is in  $H^2(\Omega) \cap H_0^1(\Omega)$ .*

### 3.2.2 Numerical method

One can try to solve (35) either directly (Davis and Gartland, 1998; Gartland and Ramage, 2015), or look for an energy minimizer (Lee et al., 2002; Ravnik and Žumer, 2009; Zhao and Wang, 2016; Bajc et al., 2016). To better compare with

our method for the uniaxially constrained Landau-de Gennes model (Section 4), we adopt the later approach and state a simple gradient flow method to find a minimizer of (32).

Let  $t$  represent “time” and suppose that  $\mathbf{Q} \equiv \mathbf{Q}(\mathbf{x}, t)$  is an evolving solution such that  $\lim_{t \rightarrow \infty} \mathbf{Q}(\cdot, t) =: \mathbf{Q}_*$  is a local minimizer of  $E_{\text{LdG}}$ , where  $\mathbf{Q}(\mathbf{x}, 0) = \mathbf{Q}_0$ , and  $\mathbf{Q}_0 \in \mathbb{V}(\mathbf{Q}_D)$  is the initial guess for the minimizer. Next, we evolve  $\mathbf{Q}(\cdot, t)$  according to the following  $L^2(\Omega)$  gradient flow:

$$(\partial_t \mathbf{Q}(\cdot, t), \mathbf{P}) = -\delta_{\mathbf{Q}} E_{\text{LdG}}[\mathbf{Q}; \mathbf{P}], \quad \forall \mathbf{P} \in \mathbb{V}(\mathbf{0}), \quad (40)$$

where the variational derivative is given in (35). Formally, the solution of (40) will converge to  $\mathbf{Q}_*$ .

We derive a numerical scheme for approximating (40) by first discretizing in time by minimizing movements (see Section 2.3). Let  $\mathbf{Q}_k(\mathbf{x}) \approx \mathbf{Q}(\mathbf{x}, k\delta t)$ , where  $\delta t > 0$  is a finite time-step, and  $k$  is the time index. Then (40) becomes a sequence of variational problems. Given  $\mathbf{Q}_k$ , find  $\mathbf{Q}_{k+1} \in \mathbb{V}(\mathbf{Q}_D)$  such that

$$\left( \frac{\mathbf{Q}_{k+1} - \mathbf{Q}_k}{\delta t}, \mathbf{P} \right) = -\delta_{\mathbf{Q}} E_{\text{LdG}}[\mathbf{Q}_{k+1}; \mathbf{P}], \quad \forall \mathbf{P} \in \mathbb{V}(\mathbf{0}), \quad (41)$$

which is equivalent to

$$\mathbf{Q}_{k+1} = \arg \min_{\mathbf{Q} \in \mathbb{V}(\mathbf{Q}_D)} F(\mathbf{Q}), \quad F(\mathbf{Q}) := \frac{1}{2\delta t} \|\mathbf{Q} - \mathbf{Q}_k\|_{L^2(\Omega)}^2 + E_{\text{LdG}}[\mathbf{Q}], \quad (42)$$

and yields the useful property  $F(\mathbf{Q}_{k+1}) \leq F(\mathbf{Q}_k)$ . However, (41) is a *fully-implicit* equation and requires an iterative solution because of the non-linearity in  $\psi_{\text{LdG}}(\mathbf{Q})$ . For convenience, we shall, instead, use a *semi-implicit* approach via convex splitting (Wise et al., 2009; Zhao and Wang, 2016; Xu et al., 2019). Let us define the following split of (15):

$$\begin{aligned} \psi_c(\mathbf{Q}) &= K + \frac{A + D}{2} \text{tr}(\mathbf{Q}^2), \\ \psi_e(\mathbf{Q}) &= \frac{D}{2} \text{tr}(\mathbf{Q}^2) + \frac{B}{3} \text{tr}(\mathbf{Q}^3) - \frac{C}{4} \left( \text{tr}(\mathbf{Q}^2) \right)^2, \\ \Rightarrow \quad \psi_{\text{LdG}}(\mathbf{Q}) &\equiv \psi_c(\mathbf{Q}) - \psi_e(\mathbf{Q}), \end{aligned} \quad (43)$$

where  $D > 0$  is chosen sufficiently large. Indeed, for all  $\mathbf{Q}$  that satisfy the physical eigenvalue ranges (7),  $\psi_c$  and  $\psi_e$  are both convex functions if  $D > 0$  is large enough.

Therefore, referring to (35), we obtain the following semi-implicit weak formulation of (41). Given  $\mathbf{Q}_k$ , find  $\mathbf{Q}_{k+1} \in \mathbb{V}(\mathbf{Q}_D)$  such that

$$\begin{aligned} & \left( \frac{\mathbf{Q}_{k+1} - \mathbf{Q}_k}{\delta t}, \mathbf{P} \right) + a_{\text{LdG}}(\mathbf{Q}_{k+1}, \mathbf{P}) + \frac{1}{\eta_B} \int_{\Omega} \frac{\partial \psi_c(\mathbf{Q}_{k+1})}{\partial \mathbf{Q}} : \mathbf{P} dx \\ & + \eta_{\Gamma} \int_{\Gamma} \frac{\partial f_{\Gamma}(\mathbf{Q}_{k+1})}{\partial \mathbf{Q}} : \mathbf{P} dS(\mathbf{x}) = \frac{1}{\eta_B} \int_{\Omega} \frac{\partial \psi_e(\mathbf{Q}_k)}{\partial \mathbf{Q}} : \mathbf{P} dx \\ & + \int_{\Omega} \frac{\partial \chi_{\text{LdG}}(\mathbf{Q}_k)}{\partial \mathbf{Q}} : \mathbf{P} dx, \quad \forall \mathbf{P} \in \mathbb{V}(\mathbf{0}), \end{aligned} \quad (44)$$

where the left-hand-side of (44) is linear in  $\mathbf{Q}_{k+1}$  and the right-hand-side is explicitly known at each iteration.

Next, we approximate (44) by a finite element method, so we introduce some basic notation and assumptions in that regard. We assume that  $\Omega \subset \mathbb{R}^3$  is discretized by a conforming shape regular triangulation  $\mathcal{T}_h = \{\mathcal{T}_i\}$  consisting of simplices, i.e. we define  $\Omega_h := \cup_{T \in \mathcal{T}_h} T$ . Furthermore, we define the space of continuous piecewise linear functions on  $\Omega_h$ :

$$\mathbb{V}_h(\Omega_h) := \left\{ v \in C^0(\Omega_h) \mid v|_T \in \mathcal{P}_1(T), \forall T \in \mathcal{T}_h \right\}, \quad (45)$$

where  $\mathcal{P}_k(T)$  is the space of polynomials of degree  $\leq k$  on  $T$ .

Next, we discretize (44) by a  $\mathcal{P}_1$  approximation of the  $\mathbf{Q}$  variable denoted  $\mathbf{Q}_h$ . With the following notation

$$\mathbf{Q} \in \Lambda \Leftrightarrow \mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{12} & q_{22} & q_{23} \\ q_{13} & q_{23} & q_{33} \end{bmatrix}, \quad \text{where } q_{33} := -q_{11} - q_{22}, \quad (46)$$

we approximate  $\mathbf{Q}$  by  $\mathbf{Q}_h \in \mathbb{Q}_h$ :

$$\begin{aligned} \mathbb{Q}_h(\Omega_h) := \left\{ \mathbf{P} \in C^0(\Omega_h) \mid \mathbf{P} \in \Lambda, p_{ij}|_T \in \mathcal{P}_1(T), \right. \\ \left. 1 \leq i, j, \leq 3, \forall T \in \mathcal{T}_h \right\}. \end{aligned} \quad (47)$$

We point out that enforcing (46) at the mesh nodes guarantees that  $\mathbf{Q}_h(x) \in \Lambda$  for all  $x \in \Omega$ . Additionally, other bases can be used to represent  $\mathbf{Q}$  (Gartland et al., 1991). Therefore, the minimization problem (28) becomes

$$\min_{\mathbf{Q}_h \in \mathbb{Q}_h \cap \mathbb{V}(\mathbf{Q}_{D,h})} E_{\text{LdG}}[\mathbf{Q}_h], \quad (48)$$

where  $\mathbf{Q}_{D,h} := \mathcal{I}_h \mathbf{Q}_D$  and  $\mathcal{I}_h$  denotes the Lagrange interpolation operator. We find a local minimizer of (48) by solving a finite element approximation of (44), i.e. given  $\mathbf{Q}_{h,k}$ , find  $\mathbf{Q}_{h,k+1} \in \mathbb{Q}_h \cap \mathbb{V}(\mathbf{Q}_{D,h})$  such that

$$\begin{aligned} & \left( \frac{\mathbf{Q}_{h,k+1} - \mathbf{Q}_{h,k}}{\delta t}, \mathbf{P}_h \right) + a_{\text{LdG}}(\mathbf{Q}_{h,k+1}, \mathbf{P}_h) + \frac{A + D}{\eta_B} (\mathbf{Q}_{h,k+1}, \mathbf{P}_h) \\ & + \eta_{\Gamma} (\mathbf{Q}_{h,k+1}, \mathbf{P}_h)_{\Gamma} = \eta_{\Gamma} (\mathcal{I}_h \mathbf{Q}_{\Gamma}, \mathbf{P}_h)_{\Gamma} - \frac{1}{2} \varepsilon_a (\mathcal{I}_h (\mathbf{E} \otimes \mathbf{E}), \mathbf{P}_h) \\ & \frac{1}{\eta_B} \left( \left[ D \mathbf{Q}_{h,k} + B \mathbf{Q}_{h,k}^2 - C \text{tr}(\mathbf{Q}_{h,k}^2) \mathbf{Q}_{h,k} \right], \mathbf{P}_h \right), \quad \forall \mathbf{P}_h \in \mathbb{Q}_h \cap \mathbb{V}(\mathbf{0}), \end{aligned} \quad (49)$$

where we have written the scheme more explicitly. We iterate this procedure until some stopping criterium is achieved. Numerical results for the standard LdG model are given in Section 4.5.2 and Section 7.

### 3.3 Ericksen

The (general) Ericksen model was originally proposed in Ericksen (1991); see also Virga (1994) for another description. In this section, we concentrate on the *one-constant model* of Ericksen and review its theoretical aspects, which can be found in Ambrosio (1990a,b), Lin (1991). Moreover, we describe a robust numerical method for finding local minimizers of the one-constant Ericksen model (see Nochetto et al., 2017, 2018 for more details).

#### 3.3.1 Energy minimization framework

We review a few hypotheses required to have a well-posed energy minimization problem, and some key features of the one-constant Ericksen energy. For convenience, we re-state (19) here:

$$\begin{aligned} E_{\text{erk}}[s, \mathbf{n}] &:= \int_{\Omega} s^2 |\nabla \mathbf{n}|^2 d\mathbf{x} \equiv (s \nabla \mathbf{n}, s \nabla \mathbf{n}), \\ E_{\text{erk-m}}[s, \mathbf{n}] &:= \frac{\kappa}{2} (\nabla s, \nabla s) + \frac{1}{2} E_{\text{erk}}[s, \mathbf{n}], \\ E_{\text{erk,bulk}}[s] &:= \frac{1}{\eta_B} (\psi_{\text{erk}}(s), 1) \\ E_{\text{erk,one}}[s, \mathbf{n}] &:= E_{\text{erk-m}}[s, \mathbf{n}] + E_{\text{erk,bulk}}[s], \end{aligned} \quad (50)$$

where  $E_{\text{erk-m}}$  is the “main” part of Ericksen’s energy. Note that the double well potential  $\psi_{\text{erk}} : (-1/2, 1) \rightarrow \mathbb{R}$  is a  $C^2$  function that satisfies (Ericksen, 1991; Ambrosio, 1990a; Lin, 1991):

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

where  $s = 0$  is a local maximum of  $\psi_{\text{erk}}$ .

If  $s \neq 0$  and constant, then  $E_{\text{erk,one}}[s, \mathbf{n}]$  effectively reduces to the Oseen-Frank (one-constant) energy  $\int_{\Omega} |\nabla \mathbf{n}|^2$ . When  $s$  is variable,  $E_{\text{erk,one}}[s, \mathbf{n}]$  avoids singular energies when defects (discontinuities in  $\mathbf{n}$ ) are present by allowing  $s$  to vanish wherever there are defects. Hence, all defects must be contained in the *singular set* (cf. (20))

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

Existence of minimizers was shown in Ambrosio (1990a), Lin (1991) through the following clever trick. By introducing the auxiliary variable  $\mathbf{u} := s\mathbf{n}$ , one can rewrite  $E_{\text{erk-m}}[s, \mathbf{n}]$  as

$$E_{\text{erk-m}}[s, \mathbf{n}] = \tilde{E}_{\text{erk-m}}[s, \mathbf{u}] := \frac{1}{2} \int_{\Omega} \left[ (\kappa - 1) |\nabla s|^2 + |\nabla \mathbf{u}|^2 \right] d\mathbf{x}, \quad (52)$$

which uses  $\nabla \mathbf{u} = \mathbf{n} \otimes \nabla s + s \nabla \mathbf{n}$  and the unit length constraint  $|\mathbf{n}| = 1$ . Thus, the total energy in terms of  $s$  and  $\mathbf{u}$  is

$$\tilde{E}_{\text{erk-one}}[s, \mathbf{u}] = \tilde{E}_{\text{erk-m}}[s, \mathbf{u}] + E_{\text{erk,bulk}}[s]. \quad (53)$$

The advantage of (52) is that it is quadratic in terms of  $s$  and  $\mathbf{u}$ , which makes the (closed) admissible set of minimizers straightforward to define (Ambrosio, 1990a; Lin, 1991):

$$\mathcal{A}_{\text{erk}} := \{(s, \mathbf{n}) \in H^1(\Omega) \times [L^\infty(\Omega)]^d : (s, \mathbf{u}, \mathbf{n}) \text{ satisfies (55),} \\ \text{with } \mathbf{u} \in [H^1(\Omega)]^d\}, \quad (54)$$

where

$$\mathbf{u} = s\mathbf{n}, \quad -1/2 \leq s \leq 1 \text{ a.e. in } \Omega, \text{ and } \mathbf{n} \in \mathbb{S}^{d-1} \text{ a.e. in } \Omega, \quad (55)$$

is called the *structural condition* of  $\mathcal{A}_{\text{erk}}$ . If we write  $(s, \mathbf{u}, \mathbf{n})$  in  $\mathcal{A}_{\text{erk}}$ , then this is equivalent to  $(s, \mathbf{n})$  in  $\mathcal{A}_{\text{erk}}$ ,  $\mathbf{u}$  in  $[H^1(\Omega)]^d$ , and  $(s, \mathbf{u}, \mathbf{n})$  satisfies (55). Indeed, (52) only holds for  $(s, \mathbf{u}, \mathbf{n})$  in  $\mathcal{A}_{\text{erk}}$ . Sometimes, we refer to the identity  $\mathbf{u} = s\mathbf{n}$  in (55) as the *cone constraint* for obvious reasons.

Boundary conditions are accounted for by functions  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $\mathbf{r}, \mathbf{q} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  so that the following is satisfied.

**Hypothesis 1** (Regularity of boundary data). *There exists  $g \in W^{1,\infty}(\mathbb{R}^d)$ ,  $\mathbf{r} \in [W^{1,\infty}(\mathbb{R}^d)]^d$ ,  $\mathbf{q} \in [L^\infty(\mathbb{R}^d)]^d$ , such that  $(g, \mathbf{r}, \mathbf{q})$  satisfies (55) on  $\mathbb{R}^d$ , i.e.  $\mathbf{r} = g\mathbf{q}$  and  $\mathbf{q} \in \mathbb{S}^{d-1}$  a.e. in  $\mathbb{R}^d$ . Furthermore, we assume there is a fixed  $c_0 > 0$  (small) such that*

$$c_0 \leq g \leq 1 - c_0, \quad (56)$$

which implies that  $\mathbf{q} \in [W^{1,\infty}(\mathbb{R}^d)]^d$ . Moreover, let  $\Gamma_s$ ,  $\Gamma_{\mathbf{u}}$ ,  $\Gamma_{\mathbf{n}}$  be open subsets of  $\partial\Omega$  on which to enforce Dirichlet conditions for  $s$ ,  $\mathbf{u}$ ,  $\mathbf{n}$  (respectively), and assume that  $\Gamma_{\mathbf{u}} = \Gamma_{\mathbf{n}} \subset \Gamma_s$ .

The admissible class, with boundary conditions, is given by

$$\mathcal{A}_{\text{erk}}(g, \mathbf{q}) := \{(s, \mathbf{n}) \in \mathcal{A}_{\text{erk}} : s|_{\Gamma_s} = g, \quad \mathbf{n}|_{\Gamma_{\mathbf{n}}} = \mathbf{q}\}, \quad (57)$$

and Hypothesis 1 guarantees that setting boundary conditions for  $(s, \mathbf{n})$  is meaningful.

For technical reasons, we require the following assumption on  $\psi_{\text{erk}}$ .

**Hypothesis 2** (Growth of potential). *Let  $c_0 > 0$  be taken from Hypothesis 1. The bulk potential  $\psi_{\text{erk}}$  satisfies*

$$\begin{aligned}\psi_{\text{erk}}(s) &\geq \psi_{\text{erk}}(1 - c_0) \quad \text{for } s \geq 1 - c_0, \\ \psi_{\text{erk}}(s) &\geq \psi_{\text{erk}}\left(-\frac{1}{2} + c_0\right) \quad \text{for } s \leq -\frac{1}{2} + c_0,\end{aligned}\tag{58}$$

which is consistent with property 1 of  $\psi_{\text{erk}}$  in (51).

The existence of a minimizer  $(s, \mathbf{n}) \in \mathcal{A}_{\text{erk}}(g, \mathbf{q})$  of  $E_{\text{erk,one}}[s, \mathbf{n}]$  is shown in Ambrosio (1990a), Lin (1991), but is also a consequence of the  $\Gamma$ -convergence theory that we review in Section 3.3.6.

### 3.3.2 Finite element discretization

The Ericksen model is degenerate in the director field  $\mathbf{n}$ . This feature, that makes it capable of capturing non-trivial defects, also makes its numerical analysis difficult. Consequently, references on numerical methods for such a model are scarce. We refer to Barrett et al. (2006), Calderer et al. (2002) and to Nochetto et al. (2017, 2018) for finite element approximations to dynamics and equilibrium configurations, respectively.

In this section we review (Nochetto et al., 2017, Sec. 2.2). First, discretize  $\Omega$  as we did in Section 3.2.2, i.e.  $\Omega \subset \mathbb{R}^d$  is approximated by  $\Omega_h$  which comes from a conforming shape-regular mesh  $\mathcal{T}_h = \{T_i\}$  consisting of simplices. For simplicity, we assume that  $\Omega \equiv \Omega_h$ , i.e. that there is no geometric error caused by the triangulation. Furthermore, let  $\mathcal{N}_h$  be the set of nodes (vertices) of  $\mathcal{T}_h$  and, with some abuse of notation, let  $N$  be the cardinality of  $\mathcal{N}_h$ .

Next, define continuous piecewise linear finite element spaces on  $\Omega$ :

$$\begin{aligned}\mathbb{S}_h &:= \{s_h \in H^1(\Omega) : s_h|_T \in \mathcal{P}_1(T), \forall T \in \mathcal{T}_h\}, \\ \mathbb{U}_h &:= \{\mathbf{u}_h \in [H^1(\Omega)]^d : \mathbf{u}_h|_T \in \mathcal{P}_1(T), \forall T \in \mathcal{T}_h\}, \\ \mathbb{N}_h &:= \{\mathbf{n}_h \in \mathbb{U}_h : |\mathbf{n}_h(\mathbf{x}_i)| = 1, \forall \mathbf{x}_i \in \mathcal{N}_h\},\end{aligned}\tag{59}$$

where the unit length constraint is enforced in  $\mathbb{N}_h$  at the nodes (vertices) of the mesh. Dirichlet boundary conditions are included via the following discrete spaces (recall Hypothesis 1):

$$\begin{aligned}\mathbb{S}_h(\Gamma_s, g_h) &:= \{s_h \in \mathbb{S}_h : s_h|_{\Gamma_s} = g_h\}, \\ \mathbb{U}_h(\Gamma_{\mathbf{u}}, \mathbf{r}_h) &:= \{\mathbf{u}_h \in \mathbb{U}_h : \mathbf{u}_h|_{\Gamma_{\mathbf{u}}} = \mathbf{r}_h\}, \\ \mathbb{N}_h(\Gamma_{\mathbf{n}}, \mathbf{q}_h) &:= \{\mathbf{n}_h \in \mathbb{N}_h : \mathbf{n}_h|_{\Gamma_{\mathbf{n}}} = \mathbf{q}_h\},\end{aligned}\tag{60}$$

where  $g_h := \mathcal{I}_h g$ ,  $\mathbf{r}_h := \mathcal{I}_h \mathbf{r}$ , and  $\mathbf{q}_h := \mathcal{I}_h \mathbf{q}$  is the discrete Dirichlet data. This leads to the following discrete admissible class with boundary conditions:

$$\begin{aligned}\mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h) &:= \{(s_h, \mathbf{n}_h) \in \mathbb{S}_h(\Gamma_s, g_h) \times \mathbb{N}_h(\Gamma_{\mathbf{n}}, \mathbf{q}_h) : \\ &\quad (s_h, \mathbf{u}_h, \mathbf{n}_h) \text{ satisfies (62), with } \mathbf{u}_h \in \mathbb{U}_h(\Gamma_{\mathbf{u}}, \mathbf{r}_h)\},\end{aligned}\tag{61}$$

where

$$\mathbf{u}_h = \mathcal{I}_h(s_h \mathbf{n}_h), \quad -1/2 \leq s_h \leq 1 \text{ in } \Omega, \quad \text{and } |\mathbf{n}_h(\mathbf{x}_i)| = 1, \forall \mathbf{x}_i \in \mathcal{N}_h, \quad (62)$$

is called the *discrete structural condition* of  $\mathcal{A}_{\text{erk}}^h$ . If we write  $(s_h, \mathbf{u}_h, \mathbf{n}_h)$  in  $\mathcal{A}_{\text{erk}}^h$ , then this is equivalent to  $(s_h, \mathbf{n}_h)$  in  $\mathcal{A}_{\text{erk}}^h$ ,  $\mathbf{u}_h$  in  $\mathbb{U}_h$ , and  $(s_h, \mathbf{u}_h, \mathbf{n}_h)$  satisfies (62). We emphasize that the approximation we are considering is not conforming. Indeed, the inclusion  $\mathcal{A}_{\text{erk}}^h \subset \mathcal{A}_{\text{erk}}$  fails because, at the discrete level, we only impose the structural condition  $\mathbf{u} = s\mathbf{n}$  at the mesh nodes.

The discretization of  $E_{\text{erk}}[s, \mathbf{n}]$  in (50) is non-standard because of the delicate nature of the degenerate term  $s^2 |\nabla \mathbf{n}|^2$ . In fact, this requires us to make an additional assumption on the meshes. We shall denote by  $\phi_i$  the standard piecewise linear “hat” function associated with a node  $\mathbf{x}_i \in \mathcal{N}_h$ , so that  $\{\phi_i\}$  are basis functions of the spaces in (59). Moreover, we indicate with  $\omega_i = \text{supp } \phi_i$  the patch of a node  $\mathbf{x}_i$  (i.e. the “star” of elements in  $\mathcal{T}_h$  that contain the vertex  $\mathbf{x}_i$ ).

**Hypothesis 3** (Weak acuteness). *For all  $h > 0$ , the mesh  $\mathcal{T}_h$  is weakly acute:*

$$k_{ij} := - \int_{\Omega} \nabla \phi_i \cdot \nabla \phi_j dx \geq 0 \quad \text{for all } i \neq j. \quad (63)$$

Condition (63) imposes a severe geometric restriction on  $\mathcal{T}_h$  (Ciarlet and Raviart, 1973; Strang and Fix, 2008). We recall the following characterization of (63) for  $d = 2$ .

**Proposition 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 (63) holds.*

Generalizations of Proposition 1 to three dimensions involve the interior dihedral angles of tetrahedra (Brandts et al., 2008; Korotov et al., 2001). We also point out that a non-obtuse tetrahedral mesh is automatically weakly acute.

We now motivate our discretization of  $E_{\text{erk}}[s, \mathbf{n}]$ . Note that for all  $\mathbf{x}_i \in \mathcal{N}_h$

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

because  $\sum_{j=1}^N \phi_j = 1$  in the domain  $\Omega$  (i.e.  $\{\phi_j\}_{j=1}^N$  is a partition of unity). So, if  $\mathbb{S}_h \ni s_h = \sum_{i=1}^N s_h(\mathbf{x}_i) \phi_i$ , then

$$\int_{\Omega} |\nabla s_h|^2 dx = - \sum_{i=1}^N k_{ii} [s_h(\mathbf{x}_i)]^2 - \sum_{i,j=1, i \neq j}^N k_{ij} s_h(\mathbf{x}_i) s_h(\mathbf{x}_j),$$

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

$$\begin{aligned} \int_{\Omega} |\nabla s_h|^2 dx &= \sum_{i,j=1}^N k_{ij} s_h(\mathbf{x}_i) (s_h(\mathbf{x}_i) - s_h(\mathbf{x}_j)) \\ &= \frac{1}{2} \sum_{i,j=1}^N k_{ij} (s_h(\mathbf{x}_i) - s_h(\mathbf{x}_j))^2 = \frac{1}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2, \end{aligned} \quad (64)$$

where we define

$$\delta_{ij} s_h := s_h(\mathbf{x}_i) - s_h(\mathbf{x}_j), \quad \delta_{ij} \mathbf{n}_h := \mathbf{n}_h(\mathbf{x}_i) - \mathbf{n}_h(\mathbf{x}_j). \quad (65)$$

On the other hand, we discretize  $E_{\text{erk}}[s, \mathbf{n}]$  by

$$E_{\text{erk}}^h[s_h, \mathbf{n}_h] := \frac{1}{2} \sum_{i,j=1}^N k_{ij} \left( \frac{s_h(\mathbf{x}_i)^2 + s_h(\mathbf{x}_j)^2}{2} \right) |\delta_{ij} \mathbf{n}_h|^2, \quad (66)$$

and the main part of the Ericksen energy by

$$\begin{aligned} E_{\text{erk-m}}^h[s_h, \mathbf{n}_h] &:= \frac{\kappa}{2} (\nabla s_h, \nabla s_h) + \frac{1}{2} E_{\text{erk}}^h[s_h, \mathbf{n}_h], \\ &= \frac{\kappa}{2} \left( \frac{1}{2} \sum_{i,j=1}^N k_{ij} (\delta_{ij} s_h)^2 \right) + \frac{1}{2} E_{\text{erk}}^h[s_h, \mathbf{n}_h]. \end{aligned} \quad (67)$$

Eq. (66) does *not* come from applying the standard discretization of  $\int_{\Omega} s^2 |\nabla \mathbf{n}|^2 dx$  by piecewise linear finite elements (though it is a first order approximation). This special discretization of the energy preserves an important energy inequality (see Lemma 1), which is crucial to proving the  $\Gamma$ -convergence of our discrete energy with the degenerate coefficient  $s^2$  *without* regularization.

The double-well energy is discretized in the usual way,

$$E_{\text{erk,bulk}}^h[s_h] := \frac{1}{\eta_B} (\psi_{\text{erk}}(s_h), 1) = \frac{1}{\eta_B} \int_{\Omega} \psi_{\text{erk}}(s_h(x)) dx. \quad (68)$$

Therefore, our discrete minimization problem for the Ericksen model is as follows. Find  $(s_h^*, \mathbf{n}_h^*) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h)$  such that

$$(s_h^*, \mathbf{n}_h^*) = \arg \min_{(s_h, \mathbf{n}_h) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h)} E_{\text{erk,one}}^h[s_h, \mathbf{n}_h], \quad (69)$$

where

$$E_{\text{erk,one}}^h[s_h, \mathbf{n}_h] := E_{\text{erk-m}}^h[s_h, \mathbf{n}_h] + E_{\text{erk,bulk}}^h[s_h]. \quad (70)$$

We close with a result showing that (67) preserves the key structure of (52) at the discrete level, and is a key component of the  $\Gamma$ -convergence of the method. First, we recall that  $\mathcal{I}_h$  denotes the Lagrange interpolation operator and introduce  $\tilde{s}_h := \mathcal{I}_h|s_h|$  and two discrete versions of the vector field  $\mathbf{u}$ ,

$$\mathbf{u}_h := \mathcal{I}_h(s_h \mathbf{n}_h) \in \mathbb{U}_h, \quad \tilde{\mathbf{u}}_h := \mathcal{I}_h(\tilde{s}_h \mathbf{n}_h) \in \mathbb{U}_h. \quad (71)$$

Note that both triplets  $(s_h, \mathbf{u}_h, \mathbf{n}_h), (\tilde{s}_h, \tilde{\mathbf{u}}_h, \mathbf{n}_h) \in \mathbb{S}_h \times \mathbb{U}_h \times \mathbb{N}_h$  satisfy (62). The following is taken from Nocchetto et al. (2017, Lem. 2.2).

**Lemma 1** (Energy inequalities). *Let the mesh  $\mathcal{T}_h$  satisfy (63). If  $(s_h, \mathbf{n}_h) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h)$ , then, for any  $\kappa > 0$ , the discrete energy (67) satisfies*

$$E_{\text{erk-m}}^h[s_h, \mathbf{n}_h] \geq (\kappa - 1) \int_{\Omega} |\nabla s_h|^2 dx + \int_{\Omega} |\nabla \mathbf{u}_h|^2 dx =: \tilde{E}_{\text{erk-m}}^h[s_h, \mathbf{u}_h], \quad (72)$$

and

$$E_{\text{erk-m}}^h[s_h, \mathbf{n}_h] \geq (\kappa - 1) \int_{\Omega} |\nabla \tilde{s}_h|^2 dx + \int_{\Omega} |\nabla \tilde{\mathbf{u}}_h|^2 dx =: \tilde{E}_{\text{erk-m}}^h[\tilde{s}_h, \tilde{\mathbf{u}}_h]. \quad (73)$$

### 3.3.3 Continuous gradient flow

We begin with a formal derivation of a gradient flow to find a local minimizer of  $E_{\text{erk,one}}[s, \mathbf{n}]$  in (50). Since  $s$  and  $\mathbf{n}$  are coupled variables, we will have two coupled gradient flows.

First, let  $a_s(\cdot, \cdot) : H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$  be a bounded bilinear form (inner product) for  $s$ . For the sake of exposition, assume an  $L^2(\Omega)$  inner product, e.g.  $a_s(s, z) := (s, z)$ , but other choices can be made. Similar to (40), we define a gradient flow for  $s$ :

$$a_s(\partial_t s(\cdot, t), z) = -\delta_s E_{\text{erk,one}}[s, \mathbf{n}; z], \quad \forall z \in H_{\Gamma_s}^1(\Omega), \quad (74)$$

where  $H_{\Gamma_s}^1(\Omega) := \{z \in H^1(\Omega) : z|_{\Gamma_s} = 0\}$  preserves the boundary condition for  $s$ , and the first variation is given by

$$\delta_s E_{\text{erk,one}}[s, \mathbf{n}; z] = \kappa \int_{\Omega} \nabla s \cdot \nabla z dx + \int_{\Omega} s z |\nabla \mathbf{n}|^2 dx + \frac{1}{\eta_B} \int_{\Omega} \psi'_{\text{erk}}(s) z dx. \quad (75)$$

Applying a formal integration by parts to (74) gives

$$\int_{\Omega} \partial_t s z dx = - \int_{\Omega} \left( -\kappa \Delta s + |\nabla \mathbf{n}|^2 s + \frac{1}{\eta_B} \psi'_{\text{erk}}(s) \right) z dx, \quad \text{for all } z \in H_{\Gamma_s}^1(\Omega), \quad (76)$$

where we use the implicit Neumann condition  $\mathbf{v} \cdot \nabla s = 0$  on  $\partial\Omega \setminus \Gamma_s$ , with  $\mathbf{v}$  being the outer unit normal of  $\partial\Omega$ . Hence,  $s$  satisfies the (nonlinear) parabolic

PDE:

$$\begin{aligned}\partial_t s - \kappa \Delta s + |\nabla \mathbf{n}|^2 s + \frac{1}{\eta_B} \psi'_{\text{erk}}(s) &= 0, \quad \text{in } \Omega, \\ s = g, \quad \mathbf{v} \cdot \nabla s &= 0, \quad \text{on } \partial\Omega \setminus \Gamma_s.\end{aligned}\tag{77}$$

Next, given  $(s, \mathbf{n}) \in \mathcal{A}_{\text{erk}}(g, \mathbf{q})$ , we consider the space of tangential variations of  $\mathbf{n}$ :

$$\mathbb{V}^\perp(s, \mathbf{n}) := \left\{ \mathbf{v} \in [L^2(\Omega)]^d : \mathbf{v} \cdot \mathbf{n} = 0 \text{ a.e. in } \Omega, \text{ and } s\mathbf{v} \in [H^1(\Omega)]^d \right\}, \tag{78}$$

which is connected with the constraint  $\mathbf{n} \in \mathbb{S}^{d-1}$  in the following sense. If  $\mathbf{n} \equiv \mathbf{n}(\mathbf{x}, t)$  is evolving director field such that  $|\mathbf{n}|^2 = 1$ , then

$$0 = \partial_t |\mathbf{n}|^2 = 2\partial_t \mathbf{n} \cdot \mathbf{n} \Rightarrow \partial_t \mathbf{n} \in \mathbb{V}^\perp(s, \mathbf{n}).$$

Indeed, introducing a tangential perturbation of  $\mathbf{n}$ :  $\mathbf{p}(t) = \mathbf{n} + t\mathbf{v}$  where  $\mathbf{v} \in \mathbb{V}^\perp(s, \mathbf{n})$ , we have that  $|\mathbf{p}|^2 = 1 + t^2|\mathbf{v}|$ , which preserves  $|\mathbf{n}| = 1$  up to second order in  $t$ . We remark that if  $s \geq c_0 > 0$  in  $\Omega$ , then  $\mathbf{v} \in \mathbb{V}^\perp(\mathbf{n})$  is necessarily in  $[H^1(\Omega)]^d$ .

Let  $a_{\mathbf{n}}(\cdot, \cdot) : \mathbb{V}^\perp(s, \mathbf{n}) \times \mathbb{V}^\perp(s, \mathbf{n}) \rightarrow \mathbb{R}$  be a bounded bilinear form (inner product) for tangential variations of  $\mathbf{n}$ . For the sake of exposition, assume an  $L^2(\Omega)$  inner product, e.g.  $a_{\mathbf{n}}(\mathbf{t}, \mathbf{v}) := (\mathbf{t}, \mathbf{v})$ , but other choices can be made. Similar to (74), we define a gradient flow for  $\mathbf{n}$ :

$$a_{\mathbf{n}}(\partial_t \mathbf{n}(\cdot, t), \mathbf{v}) = -\delta_{\mathbf{n}} E_{\text{erk,one}}[s, \mathbf{n}; \mathbf{v}], \quad \forall \mathbf{v} \in \mathbb{V}^\perp(s, \mathbf{n}) \cap [H_{\Gamma_{\mathbf{n}}}^1(\Omega)]^d, \tag{79}$$

where  $\partial_t \mathbf{n} = 0$  on  $\Gamma_{\mathbf{n}}$  preserves the boundary condition for  $\mathbf{n}$ , and the first variation is given by

$$\delta_{\mathbf{n}} E_{\text{erk,one}}[s, \mathbf{n}; \mathbf{v}] = \int_{\Omega} s^2 \nabla \mathbf{n} \cdot \nabla \mathbf{v} d\mathbf{x}. \tag{80}$$

Applying a formal integration by parts to (79) gives

$$\int_{\Omega} \partial_t \mathbf{n} \cdot \mathbf{v} d\mathbf{x} = - \int_{\Omega} -\nabla \cdot (s^2 \nabla \mathbf{n}) \cdot \mathbf{v} d\mathbf{x}, \quad \forall \mathbf{v} \in \mathbb{V}^\perp(s, \mathbf{n}) \cap [H_{\Gamma_{\mathbf{n}}}^1(\Omega)]^d, \tag{81}$$

where we use the implicit Neumann condition  $\mathbf{v} \cdot \nabla \mathbf{n} = \mathbf{0}$  on  $\partial\Omega \setminus \Gamma_{\mathbf{n}}$ . Hence,  $\mathbf{n}$  satisfies the (nonlinear) *degenerate* parabolic PDE:

$$\begin{aligned}\partial_t \mathbf{n} - \nabla \cdot (s^2 \nabla \mathbf{n}) &= \mathbf{0}, \quad \text{in } \Omega, \\ \mathbf{n} = \mathbf{q}, \quad \mathbf{v} \cdot \nabla \mathbf{n} &= \mathbf{0}, \quad \text{on } \partial\Omega \setminus \Gamma_{\mathbf{n}}.\end{aligned}\tag{82}$$

Assuming  $(s, \mathbf{n})$  evolve according to (77), (82), we have that

$$\begin{aligned}\partial_t E_{\text{erk,one}}[s, \mathbf{n}] &= \delta_s E_{\text{erk,one}}[s, \mathbf{n}; \partial_t s] + \delta_{\mathbf{n}} E_{\text{erk,one}}[s, \mathbf{n}; \partial_t \mathbf{n}], \\ &= -a_s(\partial_t s, \partial_t s) - a_{\mathbf{n}}(\partial_t \mathbf{n}, \partial_t \mathbf{n}) \leq 0,\end{aligned}\tag{83}$$

and therefore the energy is monotonically decreasing.

### 3.3.4 Discrete gradient flow

Here we discuss a discrete quasi-gradient flow algorithm, as described in Nochetto et al. (2017, Section 4.2.2). Let  $(s_h^k, \mathbf{n}_h^k) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h)$  where  $k$  indicates a “time-step” index. To simplify notation, we write

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

Using a fully implicit, backward Euler time discretization for  $\partial_t s$ , and applying the finite element space discretization in Section 3.3.2, we discretize (74) by

$$a_s \left( \frac{s_h^{k+1} - s_h^k}{\delta t}, z_h \right) = -\delta_s E_{\text{erk,one}}^h[s_h^{k+1}, \mathbf{n}_h^{k+1}; z_h], \quad \forall z_h \in \mathbb{S}_h(\Gamma_s, 0), \quad (84)$$

where  $\delta t > 0$  is a finite time step. The discrete variational derivative is given by

$$\begin{aligned} \delta_s E_{\text{erk,one}}^h[s_h^k, \mathbf{n}_h^k; z_h] &= \kappa \left( \nabla s_h^k, \nabla z_h \right) + \frac{1}{2} \delta_s E_{\text{erk}}^h[s_h^k, \mathbf{n}_h^k; z_h] \\ &\quad + \frac{1}{\eta_B} \left( \psi'_{\text{erk}}(s_h^k), z_h \right), \\ \delta_s E_{\text{erk}}^h[s_h^k, \mathbf{n}_h^k; z_h] &= \sum_{i,j=1}^N k_{ij} |\delta_{ij} \mathbf{n}_h^k|^2 \left( \frac{s_i^k z_i + s_j^k z_j}{2} \right). \end{aligned} \quad (85)$$

Next, we define the discrete version of (78):

$$\mathbb{V}_h^\perp(\mathbf{n}_h) = \{ \mathbf{v}_h \in \mathbb{U}_h : \mathbf{v}_h(\mathbf{x}_i) \cdot \mathbf{n}_h(\mathbf{x}_i) = 0, \text{ for all nodes } \mathbf{x}_i \in \mathcal{N}_h \}. \quad (86)$$

Using a “linearized” backward Euler time discretization for  $\partial_t \mathbf{n}$ , and the finite element space discretization in Section 3.3.2, we discretize (79) by

$$\begin{aligned} a_{\mathbf{n}} \left( \frac{\mathbf{n}_h^{k+1} - \mathbf{n}_h^k}{\delta t}, \mathbf{v}_h \right) &= -\delta_{\mathbf{n}} E_{\text{erk,one}}^h[s_h^{k+1}, \mathbf{n}_h^{k+1}; \mathbf{v}_h], \\ \forall \mathbf{v}_h \in \mathbb{V}_h^\perp(\mathbf{n}_h^k) \cap \mathbb{U}_h(\Gamma_{\mathbf{u}}, \mathbf{0}), \\ \text{subject to } |\mathbf{n}_h^{k+1}(\mathbf{x}_i)| &= 1, \text{ for all nodes } \mathbf{x}_i \in \mathcal{N}_h, \end{aligned} \quad (87)$$

where the discrete variational derivative is given by

$$\begin{aligned} \delta_{\mathbf{n}} E_{\text{erk,one}}^h[s_h^k, \mathbf{n}_h^k; \mathbf{v}_h] &= \frac{1}{2} \delta_{\mathbf{n}} E_{\text{erk}}^h[s_h^k, \mathbf{n}_h^k; \mathbf{v}_h], \\ \delta_{\mathbf{n}} E_{\text{erk}}^h[s_h^k, \mathbf{n}_h^k; \mathbf{v}_h] &= \sum_{i,j=1}^N k_{ij} \left( \frac{(s_i^k)^2 + (s_j^k)^2}{2} \right) (\delta_{ij} \mathbf{n}_h^k) \cdot (\delta_{ij} \mathbf{v}_h). \end{aligned} \quad (88)$$

Thus, a *possible* algorithm is the following. Given  $(s_h^k, \mathbf{n}_h^k) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h)$ , solve (84), (87) simultaneously to obtain  $(s_h^{k+1}, \mathbf{n}_h^{k+1}) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h)$ . Starting from an initial guess  $(s_h^0, \mathbf{n}_h^0) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h)$ , we iterate this until some convergence criterium is achieved.

Unfortunately, this is a fully coupled non-linear system of equations with a non-convex constraint  $|\mathbf{n}_h(\mathbf{x}_i)| = 1$ . Therefore, we adopt to split the gradient flow iteration into three sequential steps. In order to obtain a monotone, energy decreasing scheme, we first employ a convex splitting approach (Wise et al., 2009; Shen and Yang, 2010a,b) for the double well potential  $\psi_{\text{erk}}$ , i.e. we write it as a difference of two convex functions  $\psi_c, \psi_e$ :

$$\psi_{\text{erk}}(s) = \psi_c(s) - \psi_e(s). \quad (89)$$

With this, and recalling (84), (85), we make the following approximation

$$(\psi'_{\text{erk}}(s_h^{k+1}), z_h) := (\psi'_c(s_h^{k+1}), z_h) - (\psi'_e(s_h^k), z_h). \quad (90)$$

The following result is from Nochetto et al. (2017, Lem. 4.1).

**Lemma 2** (Convex-concave splitting). *For any  $s_h^k$  and  $s_h^{k+1}$  in  $\mathbb{S}_h$ , (90) implies*

$$(\psi_{\text{erk}}(s_h^{k+1}), 1) - (\psi_{\text{erk}}(s_h^k), 1) \leq (\psi'_{\text{erk}}(s_h^{k+1}), s_h^{k+1} - s_h^k). \quad (91)$$

We can now formulate our alternating direction, discrete gradient flow algorithm. Given  $(s_h^0, \mathbf{n}_h^0) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h)$ , iterate Steps 1-3 for  $k \geq 0$ :

1. **Tangential flow for  $\mathbf{n}_h$ .** First, linearize  $\mathbf{n}_h^{k+1}$  by  $\tilde{\mathbf{n}}_h^{k+1} := \mathbf{n}_h^k + \mathbf{t}_h^k$ , for some  $\mathbf{t}_h^k \in \mathbb{V}_h^\perp(\mathbf{n}_h^k) \cap H_{\Gamma_{\mathbf{n}}}^1(\Omega)$  to be determined. Note that  $\tilde{\mathbf{n}}_h^{k+1} \notin \mathbb{N}_h$ . Next, choose  $a_{\mathbf{n}}^k(\mathbf{t}_h, \mathbf{v}_h) := \delta_{\mathbf{n}} E_{\text{erk}}^h[s_h^k, \mathbf{t}_h; \mathbf{v}_h]$  which is an effective (discrete) inner product on  $\mathbb{V}_h^\perp(\mathbf{n}_h^k)$ .

Then, assuming time step is unity, we replace (87) by: find  $\mathbf{t}_h^k \in \mathbb{V}_h^\perp(\mathbf{n}_h^k) \cap H_{\Gamma_{\mathbf{n}}}^1(\Omega)$  such that

$$a_{\mathbf{n}}^k(\mathbf{t}_h^k, \mathbf{v}_h) = -\delta_{\mathbf{n}} E_{\text{erk}}^h[s_h^k, \mathbf{n}_h^k; \mathbf{v}_h], \quad \forall \mathbf{v}_h \in \mathbb{V}_h^\perp(\mathbf{n}_h^k) \cap \mathbb{U}_h(\Gamma_{\mathbf{u}}, \mathbf{0}). \quad (92)$$

2. **Projection.** Define  $\mathbf{n}_h^{k+1} \in \mathbb{N}_h(\Gamma_{\mathbf{n}}, \mathbf{q}_h)$  by

$$\mathbf{n}_i^{k+1} := \frac{\tilde{\mathbf{n}}_i^{k+1}}{|\tilde{\mathbf{n}}_i^{k+1}|} \equiv \frac{\mathbf{n}_i^k + \mathbf{t}_i^k}{|\mathbf{n}_i^k + \mathbf{t}_i^k|}, \quad \text{at all nodes } x_i \in \mathcal{N}_h. \quad (93)$$

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

$$\begin{aligned} \left( \frac{s_h^{k+1} - s_h^k}{\delta t}, z_h \right) &= -\delta_s E_{\text{erk,one}}^h[s_h^{k+1}, \mathbf{n}_h^{k+1}; z_h], \quad \forall z_h \in \mathbb{S}_h(\Gamma_s, 0), \\ &= -\kappa \left( \nabla s_h^{k+1}, \nabla z_h \right) - \frac{1}{2} \delta_s E_{\text{erk}}^h[s_h^{k+1}, \mathbf{n}_h^{k+1}; z_h] \quad (94) \\ &\quad - \frac{1}{\eta_B} \left( \psi'_{\text{erk}}(s_h^{k+1}), z_h \right), \quad \forall z_h \in \mathbb{S}_h(\Gamma_s, 0). \end{aligned}$$

The following result, taken from Nochetto et al. (2017, Thm. 4.2), shows the robustness of this algorithm.

**Theorem 4** (Energy decrease). *Let  $\mathcal{T}_h$  satisfy (63). The iterate  $(s_h^{k+1}, \mathbf{n}_h^{k+1})$  of the above algorithm exists and satisfies*

$$E_{\text{erk,one}}^h[s_h^{k+1}, \mathbf{n}_h^{k+1}] \leq E_{\text{erk,one}}^h[s_h^k, \mathbf{n}_h^k] - \frac{1}{\delta t} \int_{\Omega} (s_h^{k+1} - s_h^k)^2 dx \quad (95)$$

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

After summing in  $k$  in (95), the estimate

$$E_{\text{erk,one}}^h[s_h^K, \mathbf{n}_h^K] \leq E_{\text{erk,one}}^h[s_h^0, \mathbf{n}_h^0] - \frac{1}{\delta t} \sum_{k=0}^{K-1} \|s_h^{k+1} - s_h^k\|_{L^2(\Omega)}^2$$

follows immediately. Therefore, if we set as termination criterion that  $\|s_h^{k+1} - s_h^k\|_{L^2(\Omega)} < \varepsilon$  for some prescribed tolerance  $\varepsilon > 0$ , the algorithm must finish in a finite number of iterations.

**Remark 4** (Projection is energy-decreasing). *In order to guarantee that Step 2 above is energy decreasing, we need to use the weak-acuteness assumption on the mesh from Hypothesis 3, cf. Alouges (1997), Bartels (2006), Nochetto et al. (2017).*

**Example 1.** We illustrate the energy monotonicity with a computational experiment. We consider the square  $\Omega = (0, 1)^2$  and set  $\kappa = 1$  and  $\eta_B = 1$  in the Ericksen energy. The double-well potential we consider is such that its convex splitting (recall (89)) is

$$\begin{aligned} \psi_{\text{erk}}(s) &= \psi_c(s) - \psi_e(s) \\ &:= 63.0s^2 - (-16s^4 + 21.3333333333333s^3 + 57s^2). \end{aligned} \quad (96)$$

We point out that  $\psi_{\text{erk}}$  has a local minimum at  $s = 0$  and a global minimum at  $s = s^* := 0.750025$ . We impose Dirichlet boundary conditions for both  $s$  and  $\mathbf{n}$  on  $\Gamma_s = \Gamma_{\mathbf{n}} = \partial\Omega$ ,

$$s = s^*, \quad \mathbf{n}(x, y) = (\cos \theta, \sin \theta)^\top, \quad \theta(x, y) = 3 \text{atan2}(x - 0.3, y - 0.6), \quad (97)$$

where  $\text{atan2} : \mathbb{R}^2 \setminus \{0\} \rightarrow [-\pi, \pi]$  is the four quadrant inverse tangent function, namely,  $\text{atan2}(x_0, y_0)$  is the angle between the positive  $x$ -axis and the ray to the point  $(x_0, y_0)$ . This gives a defect of degree +3.

We consider the gradient flow algorithm discussed in Section 3.3.4 with time step  $\delta t = 10^{-1}$ , initialized with  $(s_h^0, \mathbf{n}_h^0) \equiv (s^*, (1, 0)^\top)$  in the interior nodes, and with stopping criterion  $\|s_h^{k+1} - s_h^k\|_{L^2(\Omega)} < 10^{-8}$ . Fig. 2 illustrates the energy monotonicity property of our algorithm, that finishes in 328 steps. Fig. 3 shows the evolution of the iterates at some steps in the algorithm. We obtain an equilibrium configuration in which three point defects are present in the domain. More precisely, at the final configuration the degree of orientation  $s_h$  reaches local minima approximately at  $s_h(0.20, 0.78) \approx 8.0 \times 10^{-3}$ ,  $s_h(0.31, 0.35) \approx 4.9 \times 10^{-3}$  and  $s_h(0.63, 0.58) \approx 3.3 \times 10^{-3}$ .

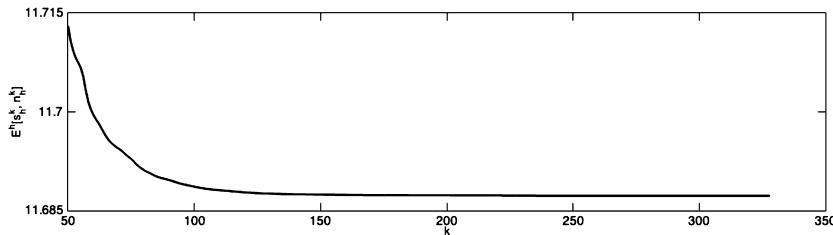


FIGURE 2 Evolution of the discrete Ericksen energy  $E_{\text{erk,one}}^h[s_h^k, \mathbf{n}_h^k]$  in Example 1.

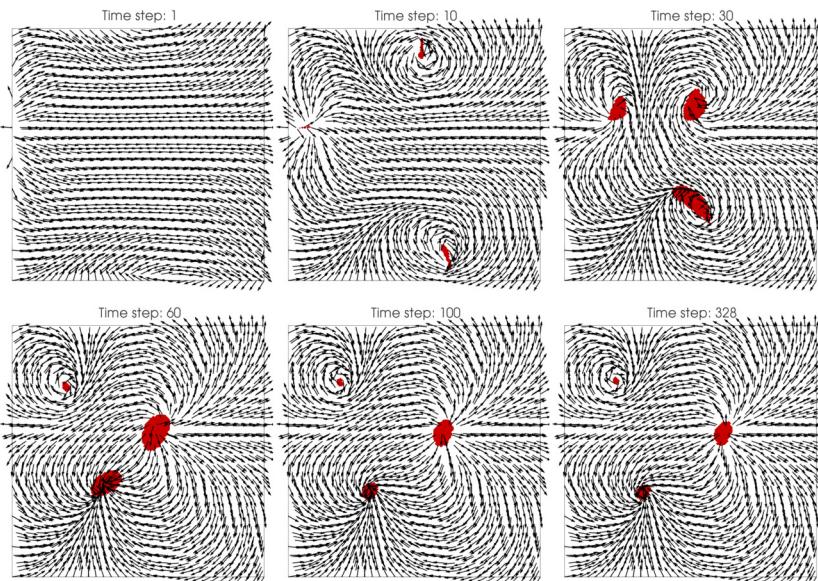


FIGURE 3 Director field  $\mathbf{n}$  in Example 1. In red, we highlight the regions where  $s < 3 \times 10^{-2}$ .

### 3.3.5 Theoretical tools

The singular set  $\mathcal{S} = \{s = 0\}$  plays a critical role in the  $\Gamma$ -convergence analysis. The following basic result from Evans (1998, Ch.5, exer. 17) is used repeatedly when dealing with the singular set.

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

Since  $\psi_{\text{erk}}$  diverges at  $s = -1/2$  and  $s = 1$ , it is useful to truncate  $s$  away from  $s = -1/2, 1$ . The next result, which is a slight modification of Nocettono et al. (2017, Lem. 3.1), clarifies this.

**Lemma 4** (Truncation). *Assume  $(g, \mathbf{q})$  satisfies Hypothesis 1 (recall  $c_0 > 0$ ). Let  $(s, \mathbf{u}, \mathbf{n}) \in \mathcal{A}_{\text{erk}}(g, \mathbf{q})$  and define*

$$\lfloor s \rfloor_\rho := \max \left\{ -\frac{1}{2} + \rho, \min\{s, 1 - \rho\} \right\}, \quad (98)$$

for any  $\rho \geq 0$ , and set  $\lfloor \mathbf{u} \rfloor_\rho := \lfloor s \rfloor_\rho \mathbf{n}$ . Then,  $(\lfloor s \rfloor_\rho, \lfloor \mathbf{u} \rfloor_\rho, \mathbf{n}) \in \mathcal{A}_{\text{erk}}(g, \mathbf{q})$  for all  $\rho \leq c_0$  and

$$\|(s, \mathbf{u}) - (\lfloor s \rfloor_\rho, \lfloor \mathbf{u} \rfloor_\rho)\|_{H^1(\Omega)} \rightarrow 0, \quad \text{as } \rho \rightarrow 0.$$

This is also implies that

$$E_{\text{erk-m}}[\lfloor s \rfloor_\rho, \mathbf{n}] \leq E_{\text{erk-m}}[s, \mathbf{n}], \quad (\psi_{\text{erk}}(\lfloor s \rfloor_\rho), 1) \leq (\psi_{\text{erk}}(s), 1), \quad (99)$$

assuming Hypothesis 2 holds as well.

The same assertion holds for any  $(s_h, \mathbf{u}_h, \mathbf{n}_h) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h)$  except the truncation is defined node-wise, i.e.  $(\mathcal{I}_h \lfloor s_h \rfloor_\rho, \mathcal{I}_h \lfloor \mathbf{u}_h \rfloor_\rho, \mathbf{n}_h) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{q}_h)$  and

$$E_{\text{erk-m}}^h[\mathcal{I}_h \lfloor s_h \rfloor_\rho, \mathbf{n}_h] \leq E_{\text{erk-m}}^h[s_h, \mathbf{n}_h], \quad (\psi_{\text{erk}}(\mathcal{I}_h \lfloor s_h \rfloor_\rho), 1) \leq (\psi_{\text{erk}}(s_h), 1). \quad (100)$$

The following proposition (taken from Nocettono et al., 2017, Prop. 3.2) is needed to construct a recovery sequence (see Section 3.3.6).

**Proposition 2** (Regularization in  $\mathcal{A}_{\text{erk}}(g, \mathbf{q})$ ). *Suppose the boundary data satisfies Hypothesis 1. Let  $(s, \mathbf{u}, \mathbf{n}) \in \mathcal{A}_{\text{erk}}(g, \mathbf{q})$ , with  $-\frac{1}{2} + \rho \leq s \leq 1 - \rho$  a.e. in  $\Omega$  for any  $\rho$  such that  $0 \leq \rho \leq c_0$ . Then, given  $\delta > 0$ , there exists a triple  $(s_\delta, \mathbf{u}_\delta, \mathbf{n}_\delta) \in \mathcal{A}_{\text{erk}}(g, \mathbf{q})$ , such that  $s_\delta \in W^{1,\infty}(\Omega)$ ,  $\mathbf{u}_\delta \in [W^{1,\infty}(\Omega)]^d$ , and*

$$\begin{aligned} & \|(s, \mathbf{u}) - (s_\delta, \mathbf{u}_\delta)\|_{H^1(\Omega)} \leq \delta, \\ & -\frac{1}{2} + \rho \leq s_\delta(x) \leq 1 - \rho, \quad \forall x \in \Omega. \end{aligned}$$

Thus, there exists  $Z_\epsilon \subset \Omega$  such that  $|Z_\epsilon| < \epsilon$  and  $(s_\delta, \mathbf{u}_\delta)$  converges uniformly on  $\Omega \setminus Z_\epsilon$ .

Moreover, define  $\mathbf{n}_\delta := \mathbf{u}_\delta/s_\delta$  if  $s_\delta \neq 0$ , and take  $\mathbf{n}_\delta$  to be any unit vector if  $s_\delta = 0$ . Then,  $\mathbf{n}_\delta \rightarrow \mathbf{n}$  in  $[L^2(\Omega \setminus \mathcal{S})]^d$ . Moreover, for each fixed  $\epsilon > 0$ ,  $\mathbf{n}_\delta$  is Lipschitz on  $\Omega \setminus \{|s_\delta| \leq \epsilon\}$  with Lipschitz constant proportional to  $\epsilon^{-1}$ .

### 3.3.6 Gamma convergence

We briefly review the main results of Nochetto et al. (2017) needed to prove  $\Gamma$ -convergence of  $E_{\text{erk,one}}^h[s_h, \mathbf{n}_h]$  to  $E_{\text{erk,one}}[s, \mathbf{n}]$ . For the existence of a recovery sequence we have Nochetto et al. (2017, Lem. 3.3).

**Lemma 5** (lim-sup inequality). *Let  $(s_\epsilon, \mathbf{u}_\epsilon, \mathbf{n}_\epsilon) \in \mathcal{A}_{\text{erk}}(g, \mathbf{r}, \mathbf{q}) \cap [W^{1,\infty}(\Omega)]^{1+2d}$  be the functions constructed in Proposition 2, for any  $\epsilon > 0$ , and let  $(s_{\epsilon,h}, \mathbf{u}_{\epsilon,h}, \mathbf{n}_{\epsilon,h}) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{r}_h, \mathbf{q}_h)$  be their Lagrange interpolants. Then*

$$\begin{aligned} E_{\text{erk-m}}[s_\epsilon, \mathbf{n}_\epsilon] &= \lim_{h \rightarrow 0} E_{\text{erk-m}}^h[s_{\epsilon,h}, \mathbf{n}_{\epsilon,h}] \\ &= \lim_{h \rightarrow 0} \tilde{E}_{\text{erk-m}}^h[s_{\epsilon,h}, \mathbf{u}_{\epsilon,h}] = \tilde{E}_{\text{erk-m}}[s_\epsilon, \mathbf{u}_\epsilon]. \end{aligned} \quad (101)$$

The next result (Nochetto et al., 2017, Lem. 3.4) is needed for the lim-inf inequality.

**Lemma 6** (Weak lower semi-continuity). *The energy  $\int_\Omega L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) d\mathbf{x}$ , with*

$$L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) := (\kappa - 1)|\nabla \mathcal{I}_h(\mathbf{w}_h)|^2 + |\nabla \mathbf{w}_h|^2, \quad (102)$$

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

$$\liminf_{h \rightarrow 0} \int_\Omega L_h(\mathbf{w}_h, \nabla \mathbf{w}_h) d\mathbf{x} \geq \int_\Omega (\kappa - 1)|\nabla \mathbf{w}|^2 + |\nabla \mathbf{w}|^2 d\mathbf{x}. \quad (103)$$

A basic part of any  $\Gamma$ -convergence result is an equi-coercivity result (Nochetto et al., 2017, Lem. 3.5).

**Lemma 7** (Coercivity). *For any  $(s_h, \mathbf{u}_h, \mathbf{n}_h) \in \mathcal{A}_{\text{erk}}^h(g_h, \mathbf{r}_h, \mathbf{q}_h)$ , we have*

$$E_{\text{erk-m}}^h[s_h, \mathbf{n}_h] \geq \min\{\kappa, 1\} \max \left\{ \int_\Omega |\nabla \mathbf{u}_h|^2 d\mathbf{x}, \int_\Omega |\nabla s_h|^2 d\mathbf{x} \right\}$$

as well as (recall (71))

$$E_{\text{erk-m}}^h[s_h, \mathbf{n}_h] \geq \min\{\kappa, 1\} \max \left\{ \int_\Omega |\nabla \tilde{\mathbf{u}}_h|^2 d\mathbf{x}, \int_\Omega |\nabla I_h(s_h)|^2 d\mathbf{x} \right\}.$$

The following result is a modification of Nochetto et al. (2017, Lem. 3.6), which characterizes the limit functions in our  $\Gamma$ -convergence result.

**Lemma 8.** Let  $(s_h, \mathbf{u}_h, \mathbf{n}_h)$  in  $\mathcal{A}_{\text{erk}}^h$  and suppose  $(s_h, \mathbf{u}_h)$  converges weakly to  $(s, \mathbf{u})$  in  $[H^1(\Omega)]^{1+d}$ . Then,  $(s_h, \mathbf{u}_h)$  converges to  $(s, \mathbf{u})$  strongly in  $[L^2(\Omega)]^{1+d}$ , a.e. in  $\Omega$ , where  $-1/2 \leq s \leq 1$ ,  $|s| = |\mathbf{u}|$  a.e. in  $\Omega$ , and there exists a director field  $\mathbf{n} : \Omega \rightarrow \mathbb{S}^{d-1}$ , with  $\mathbf{n} \in [L^2(\Omega)]^d \cap [L^\infty(\Omega)]^d$ , such that  $\mathbf{u} = s\mathbf{n}$  a.e. in  $\Omega$ . Thus,  $(s, \mathbf{u}, \mathbf{n})$  in  $\mathcal{A}_{\text{erk}}$ .

Furthermore,  $\mathbf{n}_h$  converges to  $\mathbf{n}$  in  $[L^2(\Omega \setminus \mathcal{S})]^d$  and a.e. in  $\Omega \setminus \mathcal{S}$ , and for each fixed  $\epsilon > 0$ :

1. there exists  $Z'_\epsilon \subset \Omega$  such that  $|Z'_\epsilon| < \epsilon$  and  $(s_h, \mathbf{u}_h)$  converges uniformly to  $(s, \mathbf{u})$  on  $\Omega \setminus Z'_\epsilon$ ;
2.  $\mathbf{n}_h \rightarrow \mathbf{n}$  uniformly on  $\Omega \setminus (\mathcal{S}_\epsilon \cup Z'_\epsilon)$ , where  $\mathcal{S}_\epsilon = \{|s(x)| \leq \epsilon\}$ .

Note: the same results hold for  $(\tilde{s}_h, \tilde{\mathbf{u}}_h, \mathbf{n}_h)$  in  $\mathcal{A}_{\text{erk}}^h$  converging to  $(\tilde{s}, \tilde{\mathbf{u}}, \mathbf{n})$  in  $\mathcal{A}_{\text{erk}}$ , where  $\tilde{s} := |s|$ , and  $\tilde{\mathbf{u}} := \tilde{s}\mathbf{n}$  (recall (71)).

Combining the above results, Nocet et al. (2017, Thm. 3.7) demonstrates  $\Gamma$ -convergence of our discrete energy to the continuous energy.

**Theorem 5** (Convergence of global discrete minimizers). Let  $\{\mathcal{T}_h\}$  satisfy (63). If  $(s_h, \mathbf{u}_h, \mathbf{n}_h) \in \mathcal{A}_{\text{erk}}(g_h, \mathbf{r}_h, \mathbf{q}_h)$  is a sequence of global minimizers of  $E_{\text{erk,one}}^h[s_h, \mathbf{n}_h]$  in (70), then every cluster point is a global minimizer of the continuous energy  $E_{\text{erk,one}}[s, \mathbf{n}]$  in (50).

## 4 The uniaxially constrained Q-model

In this section, we address the mathematical formulation of the minimization problem for the one-constant Landau-de Gennes energy  $E_{\text{LdG,one}}$  (cf. (17)) under the uniaxiality constraint (9). For three-dimensional problems, the approach discussed in Section 3.2.2 has two drawbacks.

First, a basic argument (Sonnet and Virga, 2012) shows that minimizers of  $\mathbf{Q} \mapsto \int_\Omega \psi_{\text{LdG}}(\mathbf{Q})$  have the form of a uniaxial nematic (9). This is *false* for  $E_{\text{LdG,one}}$  in (17) with general boundary conditions. Thus, minimizers of the form (46) violate the algebraic form of (9) and exhibit a *biaxial escape* (Palffy-muhoray et al., 1994; Sonnet et al., 1995; Lamy, 2013). This is analogous to the escape to the 3rd dimension in LC director models (Virga, 1994). This is not desirable if the underlying nematic LC is guaranteed to be uniaxial, which is the case in most thermotropic materials.

In second place, the minimization problem (48) leads to a non-linear system with five coupled variables in 3-D, which is expensive to solve and possibly not robust (Lee et al., 2002; Rávnik and Žumer, 2009; Zhao and Wang, 2016; Zhao et al., 2016).

These drawbacks motivate us to enforce the uniaxiality constraint (9) in the Landau-de Gennes one-constant energy (17). The model we obtain has similarities with the Ericksen model, but it has the advantage of allowing for non-orientable minimizers that exhibit half-integer order defects. We also point out that in 2-D, this approach is equivalent to minimizing (17) because, according to Remark 1,  $\mathbf{Q}$ -tensors must be uniaxial.

The approach we pursue is based on the Ericksen model (Section 3.3). Namely, we shall use  $(s, \mathbf{n})$  as variables, where  $\mathbf{n}$  is a (possibly non-orientable) vector field, and then recover  $\mathbf{Q}$  by means of (9), namely

$$\mathbf{Q} = s \left( \mathbf{n} \otimes \mathbf{n} - \frac{1}{d} \mathbf{I} \right).$$

Compared to directly minimizing (17) using the  $\mathbf{Q}$ -tensor as a variable, this will allow us to derive an algorithm that can find a minimizer by *solving a sequence of linear systems of smaller dimension*. See Ball et al. (2015, Prop. 1, pg. 11) for a different approach to enforcing uniaxiality.

Finally, we comment that uniaxial models effectively arise in a small elastic constant limit. In Majumdar and Zarnescu (2010), Majumdar and Zarnescu studied the one-constant model (17) with a small bulk coefficient  $\eta_B$  (which is equivalent to a small elastic constant). They showed that, under suitable boundary conditions, in the limit  $\eta_B \rightarrow 0$ , Landau-de Gennes minimizers converge to minimizers for the Oseen-Frank energy. The analysis in Majumdar and Zarnescu (2010) is refined in Nguyen and Zarnescu (2013), where the dependence of the difference between the solution to both models with respect to  $\eta_B$  is analyzed.

## 4.1 Modeling assumptions

For a uniaxially constrained  $\mathbf{Q}$ -tensor as in (9), we write  $\Theta = \mathbf{n} \otimes \mathbf{n}$ , which will be treated as a control variable in minimizing (17). We introduce the set

$$L^{d-1} = \{ \mathbf{A} \in \mathbb{R}^{d \times d} : \text{there exists } \mathbf{n} \in \mathbb{S}^{d-1}, \mathbf{A} = \mathbf{n} \otimes \mathbf{n} \}, \quad (104)$$

which can be identified with the real projective space  $\mathbf{RP}^{d-1}$  through the map

$$\mathbf{n} \otimes \mathbf{n} \longmapsto \{\mathbf{n}, -\mathbf{n}\}.$$

This illustrates that the uniaxially constrained Landau-de Gennes model takes into account the molecular *direction* but not the orientation. In contrast to the Oseen-Frank and Ericksen models, the  $\mathbf{Q}$ -tensor model allows for half-integer defects.

Because  $\nabla \mathbf{Q} = \nabla s \otimes \left( \Theta - \frac{1}{d} \mathbf{I} \right) + s \nabla \Theta$ , we have

$$|\nabla \mathbf{Q}|^2 = |\nabla s|^2 \left| \Theta - \frac{1}{d} \mathbf{I} \right|^2 + s^2 |\nabla \Theta|^2 + 2s \left[ \nabla s \otimes \left( \Theta - \frac{1}{d} \mathbf{I} \right) \right] : \nabla \Theta.$$

A direct calculation gives  $\left| \Theta - \frac{1}{d} \mathbf{I} \right|^2 = \frac{d-1}{d}$  and  $\left[ \nabla s \otimes \left( \Theta - \frac{1}{d} \mathbf{I} \right) \right] : \nabla \Theta = 0$ , and therefore

$$|\nabla \mathbf{Q}|^2 = \frac{d-1}{d} |\nabla s|^2 + s^2 |\nabla \Theta|^2.$$

Also, the equalities:

$$\begin{aligned} \text{for } d=2: \quad (1/2)s^2 &= \text{tr}(\mathbf{Q}^2), \quad 0 = \text{tr}(\mathbf{Q}^3), \quad (1/4)s^4 = (\text{tr}(\mathbf{Q}^2))^2, \\ \text{for } d=3: \quad (2/3)s^2 &= \text{tr}(\mathbf{Q}^2), \quad (2/9)s^3 = \text{tr}(\mathbf{Q}^3), \quad (4/9)s^4 = (\text{tr}(\mathbf{Q}^2))^2, \end{aligned}$$

follow immediately. Therefore, in the one-constant approximation of the uniaxially constrained  $\mathbf{Q}$ -tensor model, the energy (17) becomes

$$\begin{aligned} E_{\text{LdG,one}}[\mathbf{Q}] &= E_{\text{uni}}[s, \Theta] := E_{\text{uni-m}}[s, \Theta] + E_{\text{LdG,bulk}}[s], \\ E_{\text{uni-m}}[s, \Theta] &:= \frac{1}{2} \left( \frac{d-1}{d} \int_{\Omega} |\nabla s|^2 dx + E_{\text{uni}}[s, \Theta] \right), \\ E_{\text{uni}}[s, \Theta] &:= \int_{\Omega} s^2 |\nabla \Theta|^2 dx, \\ E_{\text{LdG,bulk}}[s] &:= \frac{1}{\eta_B} \int_{\Omega} \psi_{\text{LdG}}(s) dx, \end{aligned} \tag{105}$$

where, with some abuse of notation, we write  $\psi_{\text{LdG}}(s) := \psi_{\text{LdG}}(\mathbf{Q})$ .

It is apparent that (105) has the *same form* as the Ericksen energy (19), with the only difference that  $\Theta$  replaces  $\mathbf{n}$ . Thus, we introduce a change of variable analogous to the one in the Ericksen case; we set  $\mathbf{U} = s\Theta$  and rewrite

$$E_{\text{uni-m}}[s, \Theta] = \tilde{E}_{\text{uni-m}}[s, \mathbf{U}] := \frac{1}{2} \left( -\frac{1}{d} \int_{\Omega} |\nabla s|^2 dx + \int_{\Omega} |\nabla \mathbf{U}|^2 dx \right). \tag{106}$$

From the discussion in Section 2.1, we recall that the degree of orientation needs to satisfy  $s \in [-\frac{1}{d-1}, 1]$ . In the same spirit as before, we define the admissible class as

$$\begin{aligned} \mathcal{A}_{\text{uni}} := \{ (s, \Theta) \in H^1(\Omega) \times [L^\infty(\Omega)]^{d \times d} : & (s, \mathbf{U}, \Theta) \text{ satisfies (108),} \\ & \text{with } \mathbf{u} \in [H^1(\Omega)]^d \}, \end{aligned} \tag{107}$$

with the structural condition

$$-\frac{1}{d-1} \leq s \leq 1, \quad \mathbf{U} = s\Theta, \quad \Theta \in L^{d-1} \text{ a.e. in } \Omega. \tag{108}$$

In the same fashion as we did with the Ericksen model, we shall write  $(s, \mathbf{U}, \Theta)$  in  $\mathcal{A}_{\text{uni}}$ , to denote  $(s, \Theta)$  in  $\mathcal{A}_{\text{uni}}$ ,  $\mathbf{U}$  in  $[H^1(\Omega)]^{d \times d}$ , and  $(s, \mathbf{U}, \Theta)$  satisfies (108). In order to enforce boundary conditions on  $(s, \mathbf{U})$ , possibly on different parts of the boundary, we assume the following condition (cf. Hypothesis 1).

**Hypothesis 4.** *There exist functions  $g \in W^{1,\infty}(\mathbb{R}^d)$ ,  $\mathbf{R} \in [W^{1,\infty}(\mathbb{R}^d)]^{d \times d}$ ,  $\mathbf{M} \in [L^\infty(\mathbb{R}^d)]^{d \times d}$ , such that  $(g, \mathbf{R}, \mathbf{M})$  satisfies (108) on  $\mathbb{R}^d$ , i.e.  $\mathbf{R} = g\mathbf{M}$  and  $\mathbf{M} \in$*

$L^{d-1}$  a.e. in  $\mathbb{R}^d$ . Furthermore, we assume that  $g$  satisfies (56), that is, there is a fixed  $c_0 > 0$  (small) such that

$$c_0 \leq g \leq 1 - c_0.$$

The latter implies that  $\mathbf{M}$  is of class  $H^1$  in a neighborhood of  $\partial\Omega$  and satisfies  $g^{-1}\mathbf{R} = \mathbf{M} \in L^{d-1}$  on  $\partial\Omega$ .

Moreover, let  $\Gamma_s$ ,  $\Gamma_{\mathbf{U}}$ ,  $\Gamma_{\Theta}$  be open subsets of  $\partial\Omega$  on which to enforce Dirichlet conditions for  $s$ ,  $\mathbf{U}$ ,  $\Theta$  (respectively), and assume that  $\Gamma_{\mathbf{U}} = \Gamma_{\Theta} \subset \Gamma_s$ .

With these boundary conditions, we have the following restricted admissible class,

$$\mathcal{A}_{\text{uni}}(g, \mathbf{M}) := \{(s, \Theta) \in \mathcal{A}_{\text{uni}} : s|_{\Gamma_s} = g, \quad \Theta|_{\Gamma_{\Theta}} = \mathbf{M}\}, \quad (109)$$

and Hypothesis 4 guarantees that setting boundary conditions for  $(s, \Theta)$  is meaningful.

Finally, we require the double-well potential to effectively confine the degree of orientation variable  $s$  to a meaningful range.

**Hypothesis 5** (Landau-de Gennes potential). *The coefficients  $A, B, C$  in (15) are such that*

$$\begin{aligned} \psi_{\text{LdG}}(s) &\geq \psi_{\text{LdG}}(1 - \delta_0) \text{ for } s \geq 1 - \delta_0, \\ \psi_{\text{LdG}}(s) &\geq \psi_{\text{LdG}}\left(-\frac{1}{d-1} + \delta_0\right) \text{ for } s \leq -\frac{1}{d-1} + \delta_0. \end{aligned} \quad (110)$$

Moreover, we modify  $\psi_{\text{LdG}}$  near the bounds  $s = -1/(d-1)$  and  $s = 1$  so that  $\psi_{\text{LdG}}(\cdot)$  diverges (recall (51)).

## 4.2 Discretization

We discretize  $\Omega$  in the same fashion as in Section 3.3.2. We assume  $\Omega \subset \mathbb{R}^d$  is partitioned by a conforming simplicial shape-regular triangulation  $\mathcal{T}_h = \{T_i\}$ , with no geometric error caused by domain approximation. Moreover, we maintain the weak-acuteness mesh assumption (cf. Hypothesis 3).

Next, we consider continuous linear Lagrange finite element spaces on  $\Omega$ . That is, the space for  $s_h$  is  $\mathbb{S}_h$  as in (59), while the spaces for the tensor variables  $\mathbf{U}_h$  and  $\Theta_h$  are

$$\begin{aligned} \overline{\mathbb{U}}_h &:= \{\mathbf{U}_h \in [H^1(\Omega)]^{d \times d} : \mathbf{U}_h|_T \in \mathcal{P}_1(T), \forall T \in \mathcal{T}_h\}, \\ \mathbb{T}_h &:= \{\Theta_h \in \overline{\mathbb{U}}_h : \Theta_h(x_i) \in L^{d-1}, \forall x_i \in \mathcal{N}_h\}, \end{aligned} \quad (111)$$

where  $\mathbb{T}_h$  imposes the rank-one, unit norm constraint only at the vertices of the mesh. Dirichlet boundary conditions are included via the following discrete spaces:

$$\begin{aligned}\mathbb{S}_h(\Gamma_s, g_h) &:= \{s_h \in \mathbb{S}_h : s_h|_{\Gamma_s} = g_h\}, \\ \overline{\mathbb{U}}_h(\Gamma_U, \mathbf{R}_h) &:= \{\mathbf{U}_h \in \overline{\mathbb{U}}_h : \mathbf{U}_h|_{\Gamma_U} = \mathbf{R}_h\}, \\ \mathbb{T}_h(\Gamma_\Theta, \mathbf{M}_h) &:= \{\Theta_h \in \mathbb{T}_h : \Theta_h|_{\Gamma_\Theta} = \mathbf{M}_h\},\end{aligned}\quad (112)$$

where  $g_h := I_h g$ ,  $\mathbf{R}_h := I_h \mathbf{R}$ , and  $\mathbf{M}_h := I_h \mathbf{M}$  are the discrete Dirichlet data. This leads to the following discrete admissible class with boundary conditions:

$$\begin{aligned}\mathcal{A}_{\text{uni}}^h(g_h, \mathbf{M}_h) &:= \{(s_h, \Theta_h) \in \mathbb{S}_h(\Gamma_s, g_h) \times \mathbb{T}_h(\Gamma_\Theta, \mathbf{M}_h) : \\ &\quad (s_h, \mathbf{U}_h, \Theta_h) \text{ satisfies (114), with } \mathbf{U}_h \in \overline{\mathbb{U}}_h(\Gamma_U, \mathbf{R}_h)\},\end{aligned}\quad (113)$$

where

$$\mathbf{U}_h = I_h(s_h \Theta_h), \quad -\frac{1}{d-1} \leq s_h \leq 1 \text{ in } \Omega, \quad \text{and } \Theta_h(x_i) \in L^{d-1}, \forall x_i \in \mathcal{N}_h, \quad (114)$$

is called the *discrete structural condition* of  $\mathcal{A}_{\text{uni}}^h$ . If we write  $(s_h, \mathbf{U}_h, \Theta_h) \in \mathcal{A}_{\text{uni}}^h$ , then this is equivalent to  $(s_h, \Theta_h) \in \mathcal{A}_{\text{uni}}^h$ ,  $\mathbf{U}_h \in \overline{\mathbb{U}}_h$ , and  $(s_h, \mathbf{U}_h, \Theta_h)$  satisfies (62). In view of Hypothesis 4, we can also impose the Dirichlet condition  $\Theta_h = I_h[g_h^{-1} \mathbf{R}_h]$  on  $\partial\Omega$ .

The discrete version of  $E_{\text{uni-m}}[s, \Theta]$  is derived similarly to the Ericksen case. We set

$$\delta_{ij} s_h := s_h(x_i) - s_h(x_j), \quad \delta_{ij} \Theta_h := \Theta_h(x_i) - \Theta_h(x_j), \quad (115)$$

and define the main part of the discrete energy to be

$$\begin{aligned}E_{\text{uni-m}}^h[s_h, \Theta_h] &:= \frac{d-1}{4d} \sum_{i,j=1}^n k_{ij} (\delta_{ij} s_h)^2 \\ &\quad + \frac{1}{4} \sum_{i,j=1}^n k_{ij} \left( \frac{s_h(x_i)^2 + s_h(x_j)^2}{2} \right) |\delta_{ij} \Theta_h|^2.\end{aligned}\quad (116)$$

Above, the first term corresponds to

$$\frac{1}{2} \sum_{i,j=1}^n k_{ij} (\delta_{ij} s_h)^2 = \int_{\Omega} |\nabla s_h|^2 dx,$$

while the second term is a first order approximation of  $\frac{1}{2} \int_{\Omega} s^2 |\nabla \Theta|^2 dx$ . For convenience, we shall denote

$$E_{\text{uni}}^h[s_h, \Theta_h] := \frac{1}{2} \sum_{i,j=1}^n k_{ij} \left( \frac{s_h(x_i)^2 + s_h(x_j)^2}{2} \right) |\delta_{ij} \Theta_h|^2. \quad (117)$$

The bulk energy is discretized in the same way as before,

$$E_{\text{LdG,bulk}}^h[s_h] := \frac{1}{\eta_B} \int_{\Omega} \psi_{\text{LdG}}(s_h) dx. \quad (118)$$

With the notation introduced above, the formulation of the discrete problem reads as follows. Find  $(s_h, \Theta_h) \in \mathbb{S}_h(\Gamma_s, g_h) \times \mathbb{T}_h(\Gamma_\Theta, \mathbf{M}_h)$  such that the following energy is minimized:

$$E_{\text{uni}}^h[s_h, \Theta_h] := E_{\text{uni-m}}^h[s_h, \Theta_h] + E_{\text{LdG,bulk}}^h[s_h]. \quad (119)$$

Because the discrete spaces consist of piecewise linear functions, the structural condition  $\mathbf{U}_h = s_h \Theta_h$  is only satisfied at the mesh nodes (cf. (114)). Therefore, there is a variational crime that we need to account for. Similarly to Lemma 1, the discrete Landau-de Gennes energy possesses an energy inequality property (Borthagaray et al., 2020, Lem. 1). For our analysis, we introduce the functions

$$\tilde{s}_h = I_h(|s_h|), \quad \tilde{\mathbf{U}}_h = I_h(|s_h| \Theta_h), \quad (120)$$

and remark that  $(\tilde{s}_h, \tilde{\mathbf{U}}_h, \Theta_h)$  satisfies (114).

**Lemma 9** (Energy inequality). *Let the mesh  $\mathcal{T}_h$  satisfy (63). Then, for all  $(s_h, \mathbf{U}_h, \Theta_h) \in \mathcal{A}_{\text{uni}}^h(g_h, \mathbf{R}_h, \mathbf{M}_h)$ , the discrete energy satisfies*

$$E_{\text{uni-m}}^h[s_h, \Theta_h] - \tilde{E}_{\text{uni-m}}^h[\tilde{s}_h, \tilde{\mathbf{U}}_h] = \mathcal{E}_h, \quad (121)$$

as well as

$$E_{\text{uni-m}}^h[s_h, \Theta_h] - \tilde{E}_{\text{uni-m}}^h[\tilde{s}_h, \tilde{\mathbf{U}}_h] \geq \tilde{\mathcal{E}}_h, \quad (122)$$

where

$$\tilde{E}_{\text{uni-m}}^h[s_h, \mathbf{U}_h] := \frac{1}{2} \left( -\frac{1}{d} \int_{\Omega} |\nabla s_h|^2 dx + \int_{\Omega} |\nabla \mathbf{U}_h|^2 dx \right),$$

and

$$\mathcal{E}_h := \frac{1}{8} \sum_{i,j=1}^n k_{ij} (\delta_{ij} s_h)^2 |\delta_{ij} \Theta_h|^2 \geq 0, \quad \tilde{\mathcal{E}}_h := \frac{1}{8} \sum_{i,j=1}^n k_{ij} (\delta_{ij} \tilde{s}_h)^2 |\delta_{ij} \Theta_h|^2 \geq 0. \quad (123)$$

## 4.3 Gradient flow

### 4.3.1 Continuous gradient flow

We discuss a formal gradient flow to find local minimizers of  $E_{\text{uni}}[s, \Theta]$  in (105). More precisely, we revisit (40), and impose the uniaxial constraint (9). Because our control variables are  $s$  and  $\Theta$ , we shall evolve these two quantities

separately, although the resulting gradient flows are coupled. For the sake of exposition, we consider an  $L^2$  gradient flow for both  $s$  and  $\Theta$ , although other choices can be made.

We define

$$a_s(\partial_t s(\cdot, t), z) = -\delta_s E_{\text{uni}}[s, \Theta; z], \quad \forall z \in H_{\Gamma_s}^1(\Omega), \quad (124)$$

where  $H_{\Gamma_s}^1(\Omega) := \{z \in H^1(\Omega) : z|_{\Gamma_s} = 0\}$  preserves the boundary condition for  $s$  and, as in Section 3.3.3, we shall consider  $a_s(\cdot, \cdot)$  as the standard  $L^2$ -inner product. Upon applying a formal integration by parts to this equality and using the Neumann condition  $\mathbf{v} \cdot \nabla s = 0$  on  $\partial\Omega \setminus \Gamma_s$ , it follows that  $s$  satisfies

$$\begin{aligned} \partial_t s - \frac{d-1}{d} \Delta s + |\nabla \Theta|^2 s + \frac{1}{\eta_B} \psi'_{\text{LdG}}(s) &= 0, \quad \text{in } \Omega, \\ s = g, \quad \mathbf{v} \cdot \nabla s &= 0, \quad \text{on } \partial\Omega \setminus \Gamma_s. \end{aligned} \quad (125)$$

We now consider the evolution of the  $\Theta$  variable. For that purpose, we need a characterization of the tangent space  $T_{\Theta} L^{d-1}$  at  $\Theta \in L^{d-1}$ . Following Bartels and Raisch (2014), given  $\Theta = \mathbf{n} \otimes \mathbf{n}$ , we consider a smooth curve  $\gamma : (-\delta, \delta) \rightarrow \mathbb{S}^{d-1}$  such that  $\gamma(0) = \mathbf{n}$ , and set  $\Gamma : (-\delta, \delta) \rightarrow L^{d-1}$ ,  $\Gamma(t) = \gamma(t) \otimes \gamma(t)$ . Then, setting  $\mathbf{v} := \frac{d\gamma}{dt}(0) \in T_{\mathbf{n}} \mathbb{S}^{d-1}$ , we obtain  $\mathbf{V} \in T_{\Theta} L^{d-1}$  by

$$\mathbf{V} = \frac{d\Gamma}{dt}(0) = \left( \frac{d}{dt} \gamma(t) \otimes \gamma(t) \right) \Big|_{t=0} = \mathbf{n} \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{n}.$$

Thus, at  $\Theta = \mathbf{n} \otimes \mathbf{n}$ , there is a bijection between  $T_{\Theta} L^{d-1}$  and  $T_{\mathbf{n}} \mathbb{S}^{d-1}$ .

This motivates us to define the space of tangential variations of  $\Theta = \mathbf{n} \otimes \mathbf{n}$  as

$$\begin{aligned} \mathbb{V}^{\perp}(s, \Theta) := \{ & \mathbf{V} \in [L^2(\Omega)]^{d \times d} : \mathbf{V} = \mathbf{n} \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{n}, \mathbf{v} \cdot \mathbf{n} = 0 \text{ a.e. in } \Omega, \\ & s \mathbf{V} \in [H^1(\Omega)]^{d \times d} \}. \end{aligned} \quad (126)$$

Clearly, the restriction  $\mathbf{v} \cdot \mathbf{n} = 0$  a.e. in  $\Omega$  is sufficient to guarantee orthogonality, because

$$(\mathbf{v} \otimes \mathbf{n}) : (\mathbf{n} \otimes \mathbf{n}) = (\mathbf{v} \cdot \mathbf{n})(\mathbf{n} \cdot \mathbf{n}) = 0 \quad \text{if } \mathbf{v} \cdot \mathbf{n} = 0.$$

Moreover, if  $s \geq c_0 > 0$  in  $\Omega$ , then  $\mathbf{V} \in \mathbb{V}^{\perp}(s, \Theta)$  must belong to  $[H^1(\Omega)]^d$ .

Additionally, considering a tangential perturbation of  $\Theta$ ,  $\mathbf{T} = \Theta + \delta t \mathbf{V}$  with  $\mathbf{V} \in \mathbb{V}^{\perp}(s, \Theta)$  preserves the constraint  $|\Theta| = 1$  up to second order, since  $|\mathbf{T}|^2 = 1 + \delta t^2 |\mathbf{V}|^2$ . However, our discrete gradient flow algorithm shall exploit the identification  $T_{\Theta} L^{d-1} \simeq T_{\mathbf{n}} \mathbb{S}^{d-1}$  to consider vector-valued perturbations (instead of tensor-valued). Namely, if we take tangential variations in  $\mathbf{n}$  and

set  $\mathbf{W} := (\mathbf{n} + \delta t \mathbf{v}) \otimes (\mathbf{n} + \delta t \mathbf{v})$  with  $\mathbf{v} \in T_{\mathbf{n}} \mathbb{S}^{d-1}$ , then this yields a tangential variation of  $\Theta$  up to second order,

$$\mathbf{W} = \Theta + \delta t \mathbf{V} + \delta t^2 \mathbf{v} \otimes \mathbf{v}, \quad \mathbf{V} = \mathbf{n} \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{n} \in \mathbb{V}^\perp(s, \Theta). \quad (127)$$

We set a gradient flow for  $\Theta$  as follows:

$$a_{\Theta}(\partial_t \Theta(\cdot, t), \mathbf{V}) = -\delta_{\Theta} E_{\text{uni}}[s, \Theta; \mathbf{V}], \quad \forall \mathbf{V} \in \mathbb{V}^\perp(s, \Theta) \cap [H_{\Gamma_{\Theta}}^1(\Omega)]^{d \times d}. \quad (128)$$

Assume that  $a_{\Theta}(\cdot, \cdot)$  above is the inner product in  $L^2(\Omega)$ . After an integration by parts, it follows that  $\Theta$  satisfies

$$\begin{aligned} \partial_t \Theta - \nabla \cdot (s^2 \nabla \Theta) &= \mathbf{0}, \quad \text{in } \Omega, \\ \mathbf{n} = \mathbf{M}, \quad \mathbf{v} \cdot \nabla \Theta &= \mathbf{0}, \quad \text{on } \partial \Omega \setminus \Gamma_{\Theta}. \end{aligned} \quad (129)$$

Assuming  $(s, \mathbf{n})$  evolve according to (125), (129), it follows that

$$\begin{aligned} \partial_t E_{\text{uni}}[s, \Theta] &= \delta_s E_{\text{uni}}[s, \Theta; \partial_t s] + \delta_{\Theta} E_{\text{uni}}[s, \Theta; \partial_t \Theta], \\ &= -\|\partial_t s\|_{L^2(\Omega)} - \|\partial_t \Theta\|_{L^2(\Omega)} \leq 0, \end{aligned} \quad (130)$$

and therefore the energy is monotonically decreasing.

### 4.3.2 Discrete gradient flow

Given  $k \geq 0$ , let  $(s_h^k, \Theta_h^k) \in \mathcal{A}_{\text{uni}}^h(g_h, \mathbf{M}_h)$  and we write

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

We consider the finite element discretization discussed in Section 4.2 and use a fully implicit, backward Euler time discretization for  $\partial_t s$ , to discretize (124) by

$$a_s \left( \frac{s_h^{k+1} - s_h^k}{\delta t}, z_h \right) = -\delta_s E_{\text{uni}}^h[s_h^{k+1}, \Theta_h^{k+1}; z_h], \quad \forall z_h \in \mathbb{S}_h(\Gamma_s, 0), \quad (131)$$

where  $\delta t > 0$  is a finite time step and  $\mathbb{S}_h(\Gamma_s, 0)$  is defined according to (112). The discrete variational derivative is given by

$$\begin{aligned} \delta_s E_{\text{uni}}^h[s_h^k, \Theta_h^k; z_h] &= \frac{d-1}{d} (\nabla s_h^k, \nabla z_h) + \frac{1}{2} \delta_s E_{\text{uni}}^h[s_h^k, \Theta_h^k; z_h] \\ &\quad + \frac{1}{\eta_B} (\psi'_{\text{LdG}}(s_h^k), z_h), \end{aligned} \quad (132)$$

$$\delta_s E_{\text{uni}}^h[s_h^k, \Theta_h^k; z_h] = \sum_{i,j=1}^N k_{ij} |\delta_{ij} \Theta_h^k|^2 \left( \frac{s_i^k z_i + s_j^k z_j}{2} \right).$$

The discrete version of (126), where  $\Theta_h = \mathbf{n}_h \otimes \mathbf{n}_h$  at the mesh nodes, is defined by

$$\begin{aligned} \mathbb{V}_h^\perp(\Theta_h) = \{ \mathbf{V}_h \in \overline{\mathbb{U}}_h : \mathbf{V}_h(\mathbf{x}_i) &= \mathbf{n}_h(\mathbf{x}_i) \otimes \mathbf{v}_h(\mathbf{x}_i) + \mathbf{v}_h(\mathbf{x}_i) \otimes \mathbf{n}_h(\mathbf{x}_i), \\ &\mathbf{v}_h(\mathbf{x}_i) \cdot \mathbf{n}_h(\mathbf{x}_i) = 0, \text{ for all nodes } \mathbf{x}_i \in \mathcal{N}_h \}. \end{aligned} \quad (133)$$

Thus, a natural way to discretize (128) would be

$$\begin{aligned} a_\Theta \left( \frac{\Theta_h^{k+1} - \Theta_h^k}{\delta t}, \mathbf{V}_h \right) &= -\delta_\Theta E_{\text{uni}}^h[s_h^{k+1}, \Theta_h^{k+1}; \mathbf{V}_h], \\ \forall \mathbf{V}_h \in \mathbb{V}_h^\perp(\Theta_h) \cap \overline{\mathbb{U}}_h(\Gamma_U, \mathbf{0}), \\ \text{subject to } \Theta_h^{k+1}(\mathbf{x}_i) &\in L^{d-1}, \text{ for all nodes } \mathbf{x}_i \in \mathcal{N}_h, \end{aligned} \quad (134)$$

where the discrete variational derivatives are given by

$$\begin{aligned} \delta_\Theta E_{\text{uni}}^h[s_h^k, \Theta_h^k; \mathbf{V}_h] &= \frac{1}{2} \delta_\Theta E_{\text{uni}}^h[s_h^k, \Theta_h^k; \mathbf{V}_h], \\ \delta_\Theta E_{\text{uni}}^h[s_h^k, \Theta_h^k; \mathbf{V}_h] &= \sum_{i,j=1}^N k_{ij} \left( \frac{(s_i^k)^2 + (s_j^k)^2}{2} \right) (\delta_{ij} \Theta_h^k) : (\delta_{ij} \mathbf{V}_h). \end{aligned} \quad (135)$$

Therefore, we could consider the following algorithm: given  $(s_h^k, \Theta_h^k) \in \mathcal{A}_{\text{uni}}^h(g_h, \mathbf{M}_h)$ , solve (131), (134) *simultaneously* to obtain  $(s_h^{k+1}, \Theta_h^{k+1}) \in \mathcal{A}_{\text{uni}}^h(g_h, \mathbf{M}_h)$ . Starting from an initial guess  $(s_h^0, \Theta_h^0) \in \mathcal{A}_{\text{uni}}^h(g_h, \mathbf{M}_h)$ , we iterate this until some convergence criterium is achieved. This algorithm yields a fully coupled non-linear system of equations with the constraint  $\Theta_h^{k+1}(\mathbf{x}_i) \in L^{d-1}$ .

There are two simplifications to be made. First, in the same fashion as we did for the Ericksen model in Section 3.3.4, we split the gradient flow iteration into three steps. Namely, we evolve  $\Theta_h$ , resulting in a tangential update that does not necessarily belong to  $L^{d-1}$  at the nodes; after this, we need to project this update; and finally, evolve  $s_h$  with a gradient flow step.

However, the second step in this algorithm is problematic: projecting an arbitrary tensor onto  $L^{d-1}$  is more challenging than the simple unit length normalization step in the algorithm from Section 3.3.4. Therefore, instead of looking for tensor variations of  $\Theta$ , we shall exploit the identification between the tangent spaces  $T_\Theta L^{d-1}$  and  $T_{\mathbf{n}} \mathbb{S}^{d-1}$  to obtain a *vectorial update*.

We point out that if  $\Theta_h^{k+1}$  was a tangential update, so that

$$\Theta_i^{k+1} - \Theta_i^k = \mathbf{n}_i^k \otimes \mathbf{t}_i + \mathbf{t}_i \otimes \mathbf{n}_i^k$$

for some  $\mathbf{t}_i$  such that  $\mathbf{n}_i^k \cdot \mathbf{t}_i = 0$ , then we could replace the Frobenius inner product by a vectorial one:

$$(\Theta_i^{k+1} - \Theta_i^k) : \mathbf{V}_i = 2\mathbf{t}_i^k \cdot \mathbf{v}_i, \quad \forall \mathbf{V}_i = \mathbf{n}_i^k \otimes \mathbf{v}_i + \mathbf{v}_i \otimes \mathbf{n}_i^k.$$

Therefore, instead of (134) we consider, for  $\Theta_i^k = \mathbf{n}_i^k \otimes \mathbf{n}_i^k$ ,

$$\begin{aligned} \frac{1}{\delta t} a_{\mathbf{n}} \left( \mathbf{t}_h^k, \mathbf{v}_h \right) &= -\delta_{\Theta} E_{\text{uni}}^h [s_h^{k+1}, \Theta_h^k + \mathbf{n}_h^k \otimes \mathbf{t}_h^k + \mathbf{t}_h^k \otimes \mathbf{n}_h^k; \mathbf{V}_h], \\ &\forall \mathbf{V}_h \in \mathbb{V}_h^{\perp}(\mathbf{n}_h^k) \cap \mathbb{U}_h(\Gamma_{\mathbf{u}}, \mathbf{0}). \end{aligned} \quad (136)$$

Upon taking the update  $\tilde{\mathbf{n}}_i^{k+1} = \mathbf{n}_i^k + \mathbf{t}_i^k$ , we can recover the constraint  $\Theta_i^{k+1} \in L^{d-1}$  by considering

$$\Theta_i^{k+1} = \frac{\tilde{\mathbf{n}}_i^{k+1}}{|\tilde{\mathbf{n}}_i^{k+1}|} \otimes \frac{\tilde{\mathbf{n}}_i^{k+1}}{|\tilde{\mathbf{n}}_i^{k+1}|}.$$

Because of the second-order inconsistency committed when updating  $\Theta$  with a non-tangential variation (recall (127)), we need a careful selection of the  $a_{\mathbf{n}}(\cdot, \cdot)$ -form. Moreover, near the discrete singular set, namely wherever  $s_h^k$  is small, it is critical to allow for relatively large variations  $\mathbf{t}_h^k$  in order to accelerate the algorithm. Given a function  $\omega \in L^{\infty}(\Omega)$  with  $\omega \geq 0$ , we define the weighted  $H^1$ -space

$$\|\mathbf{v}\|_{H_{\omega}^1(\Omega)} := \left( \int_{\Omega} |\mathbf{v}(x)|^2 dx + \int_{\Omega} |\nabla \mathbf{v}(x)|^2 \omega(x) dx \right)^{1/2},$$

and we write by  $(\cdot, \cdot)_{H_{\omega}^1(\Omega)}$  its inner product. In the algorithm below, we shall consider the weight  $\omega = (s_h^k)^2$ .

Finally, we point out that the double well potential can be treated in the same way as for the Ericksen model. Indeed, by using a convex-concave splitting  $\psi_{\text{LdG}} = \psi_c - \psi_e$  and considering the approximation

$$\left( \psi'_{\text{LdG}}(s_h^{k+1}), z_h \right) := \left( \psi'_c(s_h^{k+1}), z_h \right) - \left( \psi'_e(s_h^k), z_h \right), \quad (137)$$

we obtain an unconditionally stable evolution for  $E_{\text{LdG,bulk}}^h[s_h]$  (cf. Lemma 2).

Our discrete quasi-gradient flow algorithm is as follows. Given  $(s_h^0, \Theta_h^0) \in \mathcal{A}_{\text{uni}}^h(g_h, \mathbf{M}_h)$ , with  $\Theta_h^0 = \mathbf{n}_h^0 \otimes \mathbf{n}_h^0$ , and a time step  $\delta t > 0$ , iterate Steps 1–3 for  $k \geq 0$ :

- 1. (Weighted) tangent flow step for  $\Theta_h$ :** find  $\mathbf{t}_h^k \in \mathbb{V}_h^{\perp}(\mathbf{n}_h^k) \cap H_{\Gamma_{\Theta}}^1(\Omega)$  and  $\mathbf{T}_h^k = \mathbf{n}_h^k \otimes \mathbf{t}_h^k + \mathbf{t}_h^k \otimes \mathbf{n}_h^k$ , such that

$$\begin{aligned} \frac{1}{\delta t} (\mathbf{t}_h^k, \mathbf{v}_h)_{H_{(s_h^k)^2}^1(\Omega)} &= -\delta_{\Theta} E_{\text{uni}}^h [s_h^k, \Theta_h^k + \mathbf{T}_h^k; \mathbf{V}_h], \\ &\forall \mathbf{V}_h = \mathbf{n}_h^k \otimes \mathbf{v}_h + \mathbf{v}_h \otimes \mathbf{n}_h^k, \quad \mathbf{v}_h \in \mathbb{V}_h \cap H_{\Gamma_{\Theta}}^1(\Omega). \end{aligned} \quad (138)$$

**2. Projection:** update  $\Theta_h^{k+1} \in \mathbb{T}_h(\Gamma_\Theta, \mathbf{M}_h)$  by

$$\Theta_h^{k+1}(x_i) := \frac{\mathbf{n}_h^k(x_i) + \mathbf{t}_h^k(x_i)}{|\mathbf{n}_h^k(x_i) + \mathbf{t}_h^k(x_i)|} \otimes \frac{\mathbf{n}_h^k(x_i) + \mathbf{t}_h^k(x_i)}{|\mathbf{n}_h^k(x_i) + \mathbf{t}_h^k(x_i)|}, \quad \forall x_i \in \mathcal{N}_h. \quad (139)$$

**3. Gradient flow step for  $s_h$ :** find  $s_h^{k+1} \in \mathbb{S}_h(\Gamma_s, g_h)$  such that, for all  $z_h \in \mathbb{S}_h(\Gamma_s, 0)$ ,

$$\begin{aligned} \frac{1}{\delta t} (s_h^{k+1} - s_h^k, z_h) &= -\delta_s E_{\text{uni}}^h[s_h^{k+1}, \Theta_h^{k+1}; z_h], \\ &= -\frac{d-1}{d} (\nabla s_h^{k+1}, \nabla z_h) - \frac{1}{2} \delta_s E_{\text{uni}}^h[s_h^{k+1}, \Theta_h^{k+1}; z_h] \\ &\quad - \frac{1}{\eta_B} (\psi'_{\text{LdG}}(s_h^{k+1}), z_h). \end{aligned} \quad (140)$$

Under a mild time-step restriction, this algorithm is energy-decreasing (Borthagaray et al., 2020, Thm. 2).

**Theorem 6** (Energy decrease). *Assume the family of meshes is weakly acute (cf. Hypothesis 3) and  $\delta t < Ch^{d/2}$ . Then, it holds that*

$$E_{\text{uni}}^h[s_h^N, \Theta_h^N] + \frac{1}{\delta t} \left( \sum_{k=0}^{N-1} \|\mathbf{t}_h^k\|_{H_{(s_h^k)^2}^1(\Omega)}^2 + \|s_h^{k+1} - s_h^k\|_{L^2(\Omega)}^2 \right) \leq E_{\text{uni}}^h[s_h^0, \Theta_h^0]$$

$$\forall N \geq 1.$$

Thus, the discrete energy is monotonically decreasing.

**Remark 5** (CFL condition). *The use of the weighted  $H_{(s_h^k)^2}^1(\Omega)$ -norm in Step 1 is needed to bound the second-order consistency error (126). This, in turn, leads to the stability constraint  $\delta t \leq Ch^{d/2}$  because of the use of an inverse estimate between  $L^\infty(\Omega)$  and  $L^2(\Omega)$  (Borthagaray et al., 2020). However, if  $(s_h^k)^2$  is bounded away from zero, then a milder CFL condition can be obtained, namely  $\delta t \leq Ch^{d/2-1} |\log h|^{-1}$  (Bartels and Raisch, 2014).*

#### 4.4 Gamma convergence

The roadmap to prove  $\Gamma$ -convergence of the discrete energy minimization problems to the continuous one is the same as for the Ericksen model, and makes use of the general philosophy (Braides, 2002),

equi-coerciveness +  $\Gamma$ -convergence  $\Rightarrow$  convergence of minimum problems.

As a first step, we remark that truncating the double-well potential decreases energy.

**Lemma 10** (Truncation). *Assume  $(g, \mathbf{M})$  satisfies Hypothesis 4. Let  $(s, \mathbf{U}, \Theta) \in \mathcal{A}_{\text{erk}}(g, \mathbf{M})$  and, given  $\rho \geq 0$ , consider  $\lfloor s \rfloor_\rho$  as in (98), namely: define*

$$\lfloor s \rfloor_\rho := \max \left\{ -\frac{1}{2} + \rho, \min\{s, 1 - \rho\} \right\}.$$

Define  $\lfloor \mathbf{U} \rfloor_\rho := \lfloor s \rfloor_\rho \Theta$ . Then,  $(\lfloor s \rfloor_\rho, \lfloor \mathbf{U} \rfloor_\rho, \mathbf{n}) \in \mathcal{A}_{\text{erk}}(g, \mathbf{R})$  for all  $\rho \leq c_0$  and

$$\|(s, \mathbf{U}) - (\lfloor s \rfloor_\rho, \lfloor \mathbf{U} \rfloor_\rho)\|_{H^1(\Omega)} \rightarrow 0, \text{ as } \rho \rightarrow 0.$$

This is also implies that

$$E_{\text{uni-m}}[\lfloor s \rfloor_\rho, \Theta] \leq E_{\text{uni-m}}[s, \Theta], \quad E_{\text{LdG,bulk}}[\lfloor s \rfloor_\rho] \leq E_{\text{LdG,bulk}}[s],$$

where we also assume Hypothesis 5.

The same assertion holds for any  $(s_h, \mathbf{U}_h, \Theta_h) \in \mathcal{A}_{\text{uni}}^h(g_h, \mathbf{M}_h)$  if the truncation is defined node-wise. Namely, if  $(\mathcal{I}_h \lfloor s_h \rfloor_\rho, \mathcal{I}_h \lfloor \mathbf{U}_h \rfloor_\rho, \Theta_h) \in \mathcal{A}_{\text{uni}}^h(g_h, \mathbf{M}_h)$  then

$$E_{\text{uni-m}}^h[\mathcal{I}_h \lfloor s_h \rfloor_\rho, \Theta_h] \leq E_{\text{uni-m}}^h[s_h, \Theta_h], \quad E_{\text{LdG,bulk}}^h[\mathcal{I}_h \lfloor s_h \rfloor_\rho] \leq E_{\text{LdG,bulk}}^h[s_h].$$

Because our discrete admissible class is defined by enforcing the structural conditions nodewise, we use Lagrange interpolation to construct a recovery sequence (i.e., to prove the lim-sup property needed for  $\Gamma$ -convergence). However, the natural space for  $(s, \mathbf{U})$  is  $[H^1(\Omega)]^{1+d \times d}$  (cf. (107)), and thus this construction cannot be done a priori: the Lagrange interpolant of an admissible pair  $(s, \mathbf{U})$  may not be defined at all if  $d \geq 2$ . This motivates the following result, which is a counterpart of Proposition 2. Essentially, it guarantees that Lipschitz continuous functions are  $H^1$ -dense in the admissible class.

**Proposition 3** (Regularization in  $\mathcal{A}_{\text{uni}}(g, \mathbf{M})$ ). *Suppose the boundary data satisfies Hypothesis 4. Let  $(s, \mathbf{U}, \Theta) \in \mathcal{A}_{\text{uni}}(g, \mathbf{M})$ , with  $-\frac{1}{d-1} + \rho \leq s \leq 1 - \rho$  a.e. in  $\Omega$  for any  $\rho$  such that  $0 \leq \rho \leq c_0$ . Then, given  $\epsilon > 0$ , there exists a triple  $(s_\epsilon, \mathbf{U}_\epsilon, \Theta_\epsilon) \in \mathcal{A}_{\text{uni}}(g, \mathbf{M})$ , such that  $s_\epsilon \in W^{1,\infty}(\Omega)$ ,  $\mathbf{U}_\epsilon \in [W^{1,\infty}(\Omega)]^{d \times d}$ , and*

$$\begin{aligned} \|(s, \mathbf{U}) - (s_\epsilon, \mathbf{U}_\epsilon)\|_{H^1(\Omega)} &\leq \epsilon, \\ -\frac{1}{d-1} + \rho &\leq s_\epsilon(x) \leq 1 - \rho, \quad \forall x \in \Omega. \end{aligned}$$

Thus, there exists  $Z_\delta \subset \Omega$  such that  $|Z_\delta| < \epsilon$  and  $(s_\epsilon, \mathbf{U}_\epsilon)$  converges uniformly on  $\Omega \setminus Z_\delta$ .

Moreover, define  $\Theta_\epsilon := \mathbf{U}_\epsilon / s_\epsilon$  if  $s_\epsilon \neq 0$ , and take  $\Theta_\epsilon$  to be any tensor in  $L^{d-1}$  if  $s_\epsilon = 0$ . Then,  $\Theta_\epsilon \rightarrow \Theta$  in  $[L^2(\Omega \setminus \mathcal{S})]^d$ . Moreover, for each fixed  $\delta > 0$ ,  $\Theta_\epsilon$  is Lipschitz on  $\Omega \setminus \{|s_\epsilon| \leq \delta\}$  with Lipschitz constant proportional to  $\delta^{-1}$ .

The proof of the proposition above is more delicate than for the Ericksen case. Indeed, smoothening the tensor field  $\Theta$  involves convolution and thus breaks its uniaxial structure. Therefore, uniaxiality needs to be rebuilt into the regularized field. We recall that, for instance in three dimensions, the eigenvalues of the uniaxial  $\mathbf{Q}$  tensor  $\mathbf{Q} = s(\Theta - \frac{1}{3}\mathbf{I})$  are  $\{2s/3, -s/3, -s/3\}$ . Heuristically, convolution with a localized kernel should not affect much the eigenframe of  $\Theta(x)$  if  $s$  is uniformly positive in a neighborhood of  $x$ . In such a case, one can simply extract the leading eigenspace to construct a uniaxial field. However, if  $s$  is not uniformly positive the argument does not carry. To deal with this, the idea in Borthagaray et al. (2020, Prop. 7) is to regularize the positive semidefinite field  $|s|\Theta$  within a scale  $\delta$ , to rebuild the uniaxiality and, for a coarser scale  $\sigma \geq \delta$ , to recover the sign of  $s$  by using a suitably regularized sign function.

Once we know that Lipschitz continuous functions are dense among the admissible pairs  $(s, \mathbf{U})$ , we can build a recovery sequence by using Lagrange (nodal) interpolation.

**Lemma 11** (lim-sup inequality). *Let  $(s_\epsilon, \mathbf{U}_\epsilon, \Theta_\epsilon) \in \mathcal{A}_{\text{uni}}(g, \mathbf{R}, \mathbf{M})$  be the functions constructed in Proposition 3, for any  $\epsilon > 0$ , and let  $(s_{\epsilon,h}, \mathbf{U}_{\epsilon,h}, \Theta_{\epsilon,h}) \in \mathcal{A}_{\text{uni}}^h(g_h, \mathbf{R}_h, \mathbf{M}_h)$  be their Lagrange interpolants. Then*

$$\begin{aligned} E_{\text{uni-m}}[s_\epsilon, \Theta_\epsilon] &= \lim_{h \rightarrow 0} E_{\text{uni-m}}^h[s_{\epsilon,h}, \Theta_{\epsilon,h}] \\ &= \lim_{h \rightarrow 0} \tilde{E}_{\text{uni-m}}^h[s_{\epsilon,h}, \mathbf{U}_{\epsilon,h}] = \tilde{E}_{\text{uni-m}}[s_\epsilon, \mathbf{U}_\epsilon]. \end{aligned}$$

Weak lower semi-continuity follows by the same arguments as in the Ericksen case.

**Lemma 12** (Weak lower semi-continuity). *The energy  $\int_\Omega L_h(\mathbf{W}_h, \nabla \mathbf{W}_h) d\mathbf{x}$ , with*

$$L_h(\mathbf{W}_h, \nabla \mathbf{W}_h) := -\frac{1}{d} |\nabla \mathcal{I}_h(\mathbf{W}_h)|^2 + |\nabla \mathbf{W}_h|^2,$$

is well defined for any  $\mathbf{W}_h \in \overline{\mathbb{U}}_h$  and is weakly lower semi-continuous in  $H^1(\Omega)$ , i.e. for any weakly convergent sequence  $\mathbf{W}_h \rightharpoonup \mathbf{W}$  in  $H^1(\Omega)$ , we have

$$\liminf_{h \rightarrow 0} \int_\Omega L_h(\mathbf{W}_h, \nabla \mathbf{W}_h) d\mathbf{x} \geq \int_\Omega -\frac{1}{d} |\nabla \mathbf{W}|^2 + |\nabla \mathbf{W}|^2 d\mathbf{x}. \quad (141)$$

*Proof.* Indeed, “flattening” the matrix  $\mathbf{W} \in \mathbb{R}^{d \times d}$  to a vector  $\mathbf{w} \in \mathbb{R}^{d^2}$ , we can use the same proof from Lemma 6 to prove the result because the norm of the gradient of the flattened matrix equals the Fröbenius norm of  $\nabla \mathbf{U}_h$ ; recall that  $\kappa = (d-1)/d$ .  $\square$

The next result shows that the discrete energy controls the  $H^1$  norms of both  $s_h$  and  $\mathbf{U}_h$ . This gives us the compactness needed to prove convergence of discrete minimizers towards minimizers of the uniaxially constrained Landau-de Gennes energy (105).

**Lemma 13** (Coercivity). *For any  $(s_h, \mathbf{U}_h, \Theta_h) \in \mathcal{A}_{\text{uni}}^h(g_h, \mathbf{R}_h, \mathbf{M}_h)$ , we have*

$$E_{\text{uni-m}}^h[s_h, \Theta_h] \geq \frac{d-1}{d} \max \left\{ \int_{\Omega} |\nabla \mathbf{U}_h|^2 dx, \int_{\Omega} |\nabla s_h|^2 dx \right\}$$

as well as (recall (71))

$$E_{\text{uni-m}}^h[s_h, \mathbf{n}_h] \geq \frac{d-1}{d} \max \left\{ \int_{\Omega} |\nabla \tilde{\mathbf{U}}_h|^2 dx, \int_{\Omega} |\nabla I_h| s_h|^2 dx \right\}.$$

Next, we prove that the limit functions satisfy the Landau-de Gennes admissibility condition (109) (cf. Borthagaray et al., 2020, Lem. 9).

**Lemma 14.** *Let  $(s_h, \mathbf{U}_h, \Theta_h)$  in  $\mathcal{A}_{\text{uni}}^h$  and suppose  $(s_h, \mathbf{U}_h)$  converges weakly to  $(s, \mathbf{U})$  in  $[H^1(\Omega)]^{1+d \times d}$ . Then,  $(s_h, \mathbf{U}_h)$  converges to  $(s, \mathbf{U})$  strongly in  $[L^2(\Omega)]^{1+d \times d}$ , a.e. in  $\Omega$ , where  $-1/(d-1) \leq s \leq 1$ ,  $|s| = |\mathbf{U}|$  a.e. in  $\Omega$ , and there exists a field  $\Theta : \Omega \rightarrow L^{d-1}$ , so that  $\Theta \in [L^\infty(\Omega)]^{d \times d}$ , such that  $\mathbf{U} = s\Theta$  a.e. in  $\Omega$ . Thus,  $(s, \mathbf{U}, \Theta)$  in  $\mathcal{A}_{\text{uni}}$ .*

Furthermore,  $\Theta_h$  converges to  $\Theta$  in  $[L^2(\Omega \setminus \mathcal{S})]^{d \times d}$  and a.e. in  $\Omega \setminus \mathcal{S}$ ,  $\Theta$  admits a Lebesgue gradient on  $\Omega \setminus \mathcal{S}$ , that satisfies the identity  $|\nabla \mathbf{U}|^2 = |\nabla s|^2 + s^2 |\nabla \Theta|^2$  a.e. in  $\Omega \setminus \mathcal{S}$ , and for each fixed  $\epsilon > 0$ :

1. there exists  $Z'_\epsilon \subset \Omega$  such that  $|Z'_\epsilon| < \epsilon$  and  $(s_h, \mathbf{U}_h)$  converges uniformly to  $(s, \mathbf{U})$  on  $\Omega \setminus Z'_\epsilon$ ;
2.  $\Theta_h \rightarrow \Theta$  uniformly on  $\Omega \setminus (\mathcal{S}_\epsilon \cup Z'_\epsilon)$ , where  $\mathcal{S}_\epsilon = \{|s(x)| \leq \epsilon\}$ .

Note: the same results hold for  $(\tilde{s}_h, \tilde{\mathbf{U}}_h, \Theta_h)$  in  $\mathcal{A}_{\text{erk}}^h$  converging to  $(\tilde{s}, \tilde{\mathbf{U}}, \Theta)$  in  $\mathcal{A}_{\text{uni}}$ , where  $\tilde{s} := |s|$ , and  $\tilde{\mathbf{U}} := \tilde{s}\Theta$  (recall (120)).

Collecting Lemmas 11–14, a standard argument yields the  $\Gamma$ -convergence of our discrete energy to the continuous energy.

**Theorem 7** (Convergence of global discrete minimizers). *Let  $\{\mathcal{T}_h\}$  satisfy (63). If  $(s_h, \mathbf{U}_h, \Theta_h) \in \mathcal{A}_{\text{uni}}(g_h, \mathbf{R}_h, \mathbf{M}_h)$  is a sequence of global minimizers of  $E_{\text{uni}}^h[s_h, \Theta_h]$  in (119), then every cluster point is a global minimizer of the continuous energy  $E_{\text{uni}}[s, \Theta]$  in (105).*

## 4.5 Numerical experiment

We simulate a curved line defect in the unit cube  $(0, 1)^3$  that exhibits a  $+1/2$  degree “point” defect in each horizontal plane of the cube; hence, the line field is non-orientable. We first simulate the uniaxially constrained model, then the standard LdG model.

### 4.5.1 Uniaxially constrained model

The double-well potential with a convex splitting is given by

$$\begin{aligned}\psi_{\text{LdG}}(s) &= \psi_c(s) - \psi_e(s) \\ &:= (36.7709s^2 + 1) - (-7.39101s^4 + 4.51673s^3 + 39.27161s^2),\end{aligned}\quad (142)$$

with  $\eta_B = 1/16$ , and note that  $\psi_{\text{LdG}}$  has a local maximum at  $s = 0$  and a global minimum at  $s = s^* := 0.700005531$  with  $\psi_{\text{LdG}}(s^*) = 0$ .

The boundary conditions for  $\Theta$  were constructed in the following way. Let  $\theta_0(x, y)$  define a  $+1/2$  degree defect in the plane, located at  $(0.3, 0.3)$  by

$$\theta_0(x, y) = \frac{1}{2} \text{atan2}\left(\frac{y - 0.3}{x - 0.3}\right), \quad (143)$$

where  $\text{atan2}$  is the four-quadrant inverse tangent function (analogous to (97)). Likewise, let  $\theta_1(x, y)$  define a  $+1/2$  degree defect in the plane, located at  $(0.7, 0.7)$ . Next, define the Dirichlet boundary  $\Gamma_s = \Gamma_\Theta = \partial\Omega \setminus \Gamma_o$ , where  $\Gamma_o := \overline{\Omega} \cap (\{z = 0\} \cup \{z = 1\})$ . Then, the Dirichlet conditions are

$$\begin{aligned}s &= s^*, \quad \mathbf{n}(x, y) = (\cos\theta, \sin\theta, 0), \quad \Theta = \mathbf{n} \otimes \mathbf{n}, \\ \theta(x, y, z) &= (1 - z)\theta_0(x, y) + z\theta_1(x, y) + \pi z,\end{aligned}\quad (144)$$

with vanishing Neumann condition on  $\Gamma_o$ . Basically, the boundary conditions consist of rotating a planar  $+1/2$  degree point defect as a function of  $z$ . The solution is computed with the gradient flow approach in Section 4.3.2 and time step  $\delta t = 10^{-3}$ , and initialized with

$$s = s^*, \quad \mathbf{n} = (\cos\alpha, \sin\alpha, 0), \quad \Theta = \mathbf{n} \otimes \mathbf{n}, \quad \alpha(x, y, z) = \theta_2(x, y) + \pi z,$$

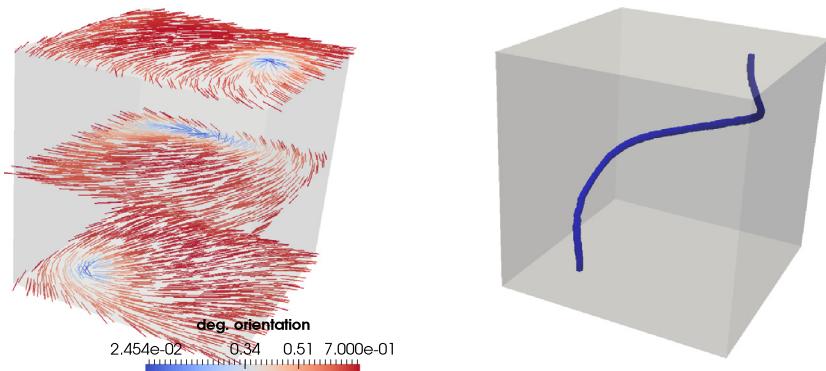
where  $\theta_2(x, y)$  corresponds to a  $+1/2$  degree defect centered at  $(0.5, 0.5)$ ; this configuration has an initial energy of  $E_{\text{uni}}^h[s_h, \Theta_h] = 10.013214$ .

Fig. 4 shows three dimensional views of the minimizing configuration, where as Fig. 5 shows four horizontal slices of the solution. A non-orientable line defect is observed, with final energy  $E_{\text{uni}}^h[s_h, \Theta_h] = 5.2042593769$  and  $\min(s_h) \approx 2.145 \times 10^{-2}$ .

### 4.5.2 The standard LdG model

Next, we simulate the model in Section 3.2.1. We use the boundary conditions in (144) and the double-well potential in (142). In terms of the standard LdG model, the Dirichlet boundary conditions on  $\Gamma_D := \Gamma_s \equiv \Gamma_\Theta$  are

$$\mathbf{Q} = s^* \left( \Theta - \frac{1}{3} \mathbf{I} \right) \text{ on } \Gamma_D, \quad (145)$$



**FIGURE 4** A  $+1/2$  degree line defect in a 3-D cube domain (Section 4.5.1). Left: line field  $\Theta$  is shown at levels  $z = 0.0, 0.5, 1.0$  (colored by  $s$ ). Right: The  $s = 0.05$  iso-surface is shown that contains the line defect. In each horizontal plane, the line field exhibits a  $+1/2$  degree *point* defect in 2-D. The twisting of the line defect is due to the choice of boundary conditions.

where  $\Theta$  is taken from (144), with vanishing Neumann condition on  $\Gamma_o$ ; this is consistent with the boundary conditions in (144). Moreover, the double-well potential is given by (15), (43), where

$$\begin{aligned} K &= 1.0, & A &= -7.502104, & B &= 60.975813, \\ C &= 66.519069, & D &= 552.230967, \end{aligned} \quad (146)$$

which is consistent with the double well potential (142). The minimizer is computed using the gradient flow approach in Section 3.2.2, with time step  $\delta t = 0.01$ , and initialized with the minimizer from the uniaxial model. All other parameters are the same.

For visualizing the solution, we shall use the *biaxiality parameter* (Majumdar and Zarnescu, 2010, eqn. (25)), given by

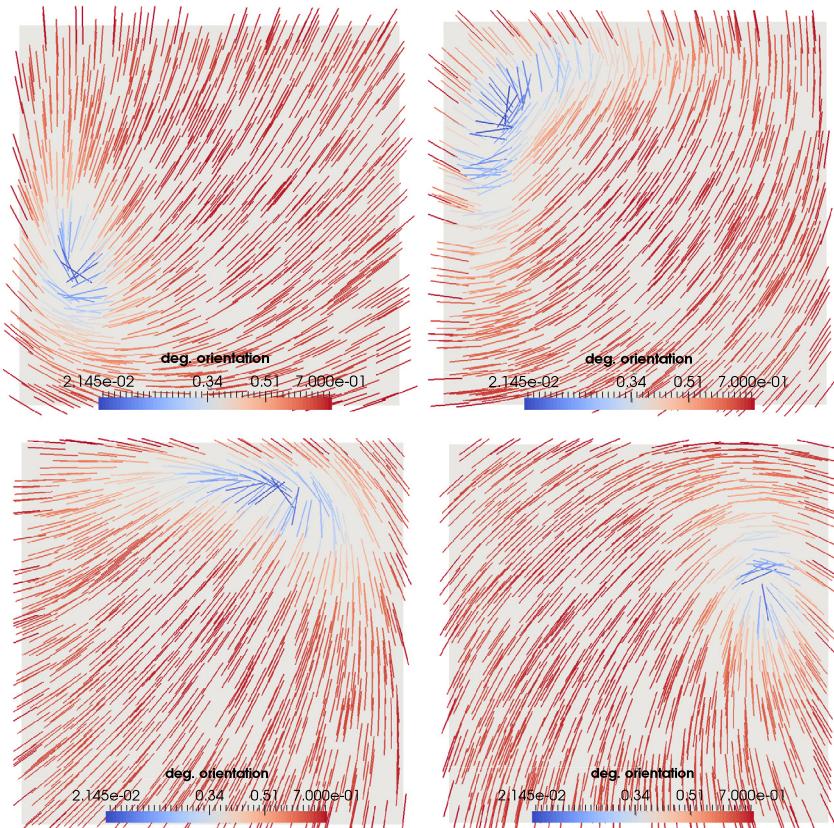
$$\beta(\mathbf{Q}) = 1 - 6 \frac{(\text{tr}(\mathbf{Q}^3))^2}{(\text{tr}(\mathbf{Q}^2))^3}, \quad (147)$$

where  $0 \leq \beta(\mathbf{Q}) \leq 1$  and has the properties:

1.  $\beta(\mathbf{Q}) = 0$  if and only if  $\mathbf{Q}$  is uniaxial, i.e.  $\mathbf{Q}$  has the form (9);
2.  $\beta(\mathbf{Q}) = 1$  if and only if  $s_1/s_2 = 2$ , where  $s_1, s_2$  appear in the biaxial form (5).

In other words,  $\beta(\mathbf{Q})$  provides a simple measure of uniaxiality versus biaxiality.

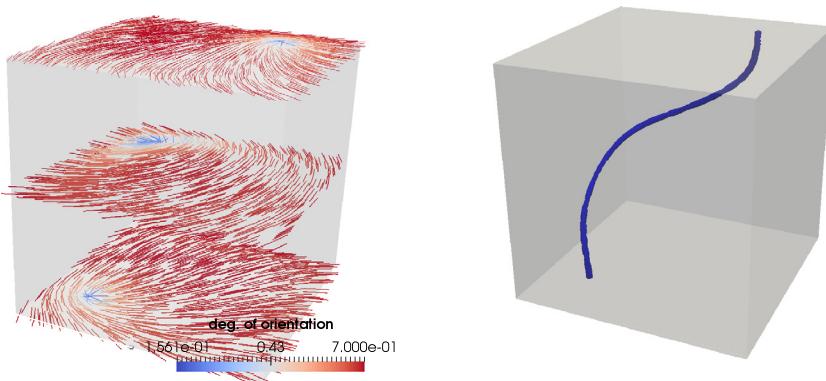
Fig. 6 shows three dimensional views of the minimizing configuration, whereas Fig. 7 shows the biaxiality and point-wise  $l^2$  error  $|\langle \mathbf{Q}_{\text{uni}} - \mathbf{Q}_{\text{LdG}} \rangle(\mathbf{x})|$  between the uniaxial (uni) and standard LdG solutions. A non-orientable line defect is observed, with final energy  $E_{\text{uni}}^h[s_h, \Theta_h] = 4.5533587$  and achieves a maximum biaxiality of 1.0.



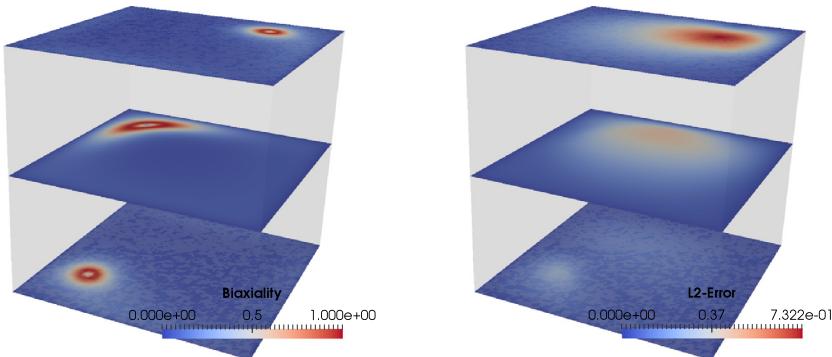
**FIGURE 5** Horizontal slices of the  $+1/2$  degree line defect in a 3-D cube domain shown in Fig. 4 (Section 4.5.1). Top: left is  $z = 0.2$ , right is  $z = 0.4$ . Bottom: left is  $z = 0.6$ , right is  $z = 0.8$ . The location of the point defect in each plane rotates with the boundary conditions.

## 5 Colloidal effects

The presence of a colloidal particle in suspension in a LC material modifies the topology of the domain. This, in turn, can induce interesting equilibrium states with non-trivial defect configurations. A famous example is the so-called *Saturn ring defect* (Alama et al., 2016; Gu and Abbott, 2000), which is a circular ring of defect surrounding a spherical hole inside the LC domain (see Fig. 8). Fig. 9 shows more detail on the director configuration for the Saturn ring defect. The boundary conditions on the spherical inclusion are  $\mathbf{n} = \mathbf{v}$  (the unit normal of the spherical hole) on  $\Gamma_i$  and  $\mathbf{n} = (0, 0, 1)^T$  on  $\Gamma_o$ . Note that the disclination ring can have an alternate configuration (see right plot in Fig. 9), which depends on the size of the particle (Wang et al., 2016; Fukuda et al., 2004; Ravnik and Žumer, 2009). Either way, we emphasize that the presence of the hole can *force* a defect in the LC.



**FIGURE 6** A  $+1/2$  degree line defect in a 3-D cube domain (Section 4.5.2). Left: line field (taken as the dominant eigenvector of  $\mathbf{Q}$ ) is shown at levels  $z = 0.0, 0.5, 1.0$  (colored by the effective  $s := (3/2)\lambda$ , where  $\lambda$  is the dominant eigenvalue). Right: The  $s = 0.22$  iso-surface is shown that contains the line defect. In each horizontal plane, the line field exhibits a  $+1/2$  degree *point* defect in 2-D. The solution looks qualitatively a little different from Fig. 4.

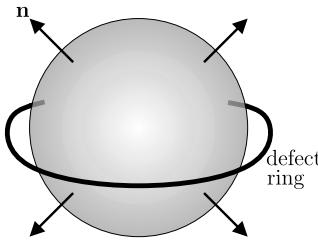


**FIGURE 7** View of the biaxiality and  $l^2$  error of the solution in Fig. 6 (Section 4.5.2). Left: clearly, there is a high degree of biaxiality near the defect. Right: plots of the  $l^2$  error  $|(Q_{\text{uni}} - Q_{\text{LDG}})(\mathbf{x})|$  are shown. The error is larger near the defect and, interestingly, it increases as a function of  $z$ .

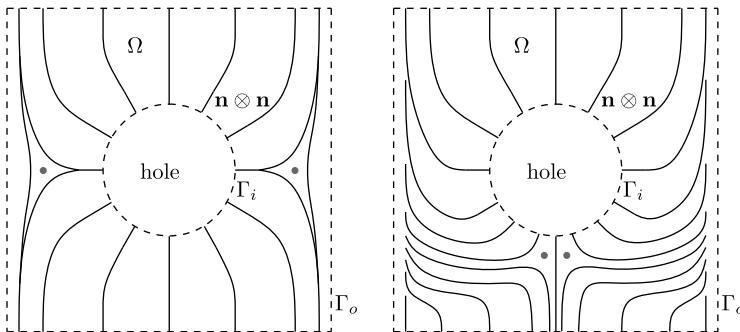
This section discusses the capabilities of the Ericksen and uniaxially constrained Landau-de Gennes models, and the corresponding numerical methods described in sections 3.3.2 and 4.2, to capture defects in the presence of colloids. We shall model colloids as spherical inclusions inside the LC domain.

## 5.1 Conforming non-obtuse mesh

Given an arbitrary domain, it may be quite difficult to generate a conforming, non-obtuse, tetrahedral mesh. As far as we know, the question of whether it is possible to generate a non-obtuse tetrahedral mesh of a general three dimensional domain remains open.



**FIGURE 8** Saturn ring defect in a director field model. A spherical colloidal particle is shown with normal anchoring conditions on its boundary (i.e. the director field  $\mathbf{n}$  is normal to the sphere). The singular set  $\mathcal{S}$  (where  $s = 0$ ) is marked by the thick curve and occurs depending on the outer boundary conditions (away from the sphere) imposed on  $\mathbf{n}$ .



**FIGURE 9** Illustration of Saturn ring defect pattern. Left: a two dimensional vertical slice of the domain shown in Fig. 8; thick lines show the line field  $\mathbf{n} \otimes \mathbf{n}$ . The defect region is marked by the two grayed circles. Right: another possible defect configuration (Wang et al., 2016; Fukuda et al., 2004; Ravnik and Žumer, 2009). The ring has a much smaller radius and is below the spherical inclusion.

Here we report on numerical results over a certain non-obtuse mesh of a cylindrical domain with a hole cut out. We refer to Nochetto et al. (2018, Sec. 5.1.1) for details about the mesh construction. For the simulations in this section, the domain  $\Omega$  is a “prism” type of cylindrical domain with square cross-section  $[-0.25\sqrt{2}, 0.75\sqrt{2}]^2$ , is centered about the  $z = 0$  plane, and has height 6. It contains a spherical inclusion, with boundary  $\Gamma_i$ , centered at  $(\sqrt{2}/4, \sqrt{2}/4, 0)$  with radius  $0.283/\sqrt{2}$ .

For the Ericksen model, one could in principle consider the strong anchoring conditions

$$\mathbf{n} = \mathbf{v} \text{ on } \Gamma_i, \quad \mathbf{n} = (0, 0, 1)^\top \text{ on } \Gamma_o = \partial\Omega \setminus \Gamma_i, \quad s = s^* \text{ on } \partial\Omega, \quad (148)$$

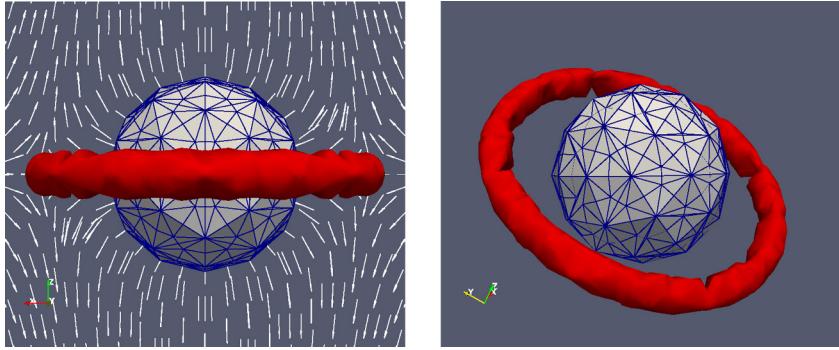
where  $\mathbf{v}$  is the outer unit normal of the spherical inclusion and  $s^*$  is the global minimum of the double well potential (96). These boundary conditions *do not* lead to a ring-like defect, but rather to disperse/point defects, depending on the value of  $\kappa$  in  $E_{\text{erk-m}}$  (cf. Nochetto et al., 2018, Sec. 5.1.2).

Instead of (148), we consider the following boundary conditions:

$$\begin{aligned}\mathbf{n} &= \mathbf{v} \text{ on } \Gamma_i, \quad s = s^* \text{ on } \partial\Omega, \\ \mathbf{n} &\text{ interpolates between } (0, 0, -1)^\top \text{ and } (0, 0, 1)^\top \text{ on } \Gamma_o.\end{aligned}\quad (149)$$

Fig. 10 shows the outcome of a numerical simulation with  $\kappa = 1$ , and a gradient flow with initial conditions  $s = s^*$ ,

$$\begin{aligned}\mathbf{n}(x, y, z) &= (0, 0, -1)^\top \text{ if } z < 0, \\ \mathbf{n}(x, y, z) &= (0, 0, 1)^\top \text{ if } z \geq 0.\end{aligned}$$



**FIGURE 10** Simulation results for the Ericksen model with boundary conditions (149). The surface mesh of the internal hole is shown and the defect region is indicated by the  $s = 0.12$  iso-surface (plotted in red). The director field  $\mathbf{n}$  is depicted with white arrows.

Importantly, the structure of the director field does not coincide with the one expected from the Landau-de Gennes model (Alama et al., 2016). Here, at every vertical slice, the defect in the director field has degree  $-1$ , while in Alama et al. (2016) the degree of the defect is  $-1/2$ . The Ericksen model imposes an orientability constraint that is *not* part of the physical problem.

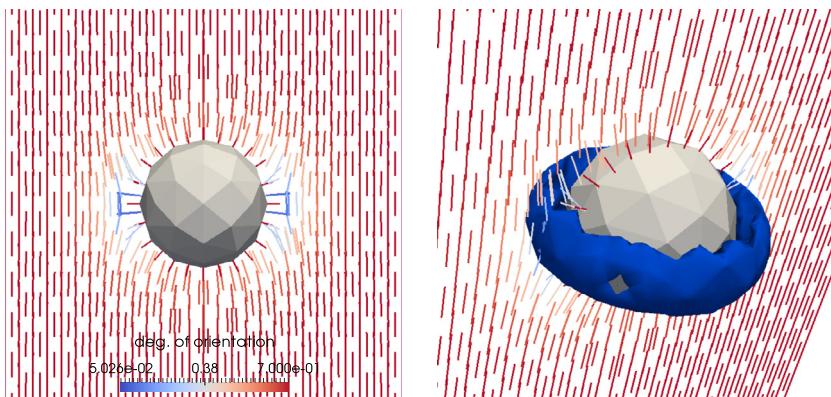
The uniaxially-constrained Landau-de Gennes model is capable of capturing such a non-orientable configuration. We impose the strong anchoring conditions

$$\mathbf{n} = \mathbf{v} \text{ on } \Gamma_i, \quad \mathbf{n} = (0, 0, 1)^\top \text{ on } \Gamma_o, \quad \Theta = \mathbf{n} \otimes \mathbf{n} \text{ on } \partial\Omega, \quad s = s^* \text{ on } \partial\Omega, \quad (150)$$

where now  $s^* = 0.7$  is the global minimum of the double-well potential

$$\begin{aligned}\psi_{\text{LdG}}(s) &= \psi_c(s) - \psi_e(s) \\ &:= (36.770913s^2 + 1) \\ &\quad - (-7.3910077s^4 + 4.5167269s^3 + 39.271614s^2).\end{aligned}\quad (151)$$

We take a time step  $\delta t = 10^{-3}$  for the gradient flow, which is initiated with  $s = s^*$  and  $\mathbf{n} = (0, 0, 1)^\top$ . Fig. 11 displays the final configuration of  $(s, \Theta)$ . A cross-section shows the non-orientability of the resulting line field.



**FIGURE 11** Simulation results for the uniaxially-constrained Landau-de Gennes model with boundary conditions (150). In the left panel, the line field is plotted with color scale based on  $s$ , and the  $-1/2$  degree defect are visible on the left and right sides of the spherical inclusion. The right panel shows the  $s = 0.25$  isosurface in blue.

## 5.2 Immersed boundary method

Mesh weak-acuteness imposes a hard geometric constraint on the meshes, and can be extremely difficult to satisfy in implementations in three dimensions. As an alternative to it, Nochetto et al. (2018, Section 5.2) proposes an immersed boundary approach to deal with general colloid shapes. This approach consists in representing the LC domain by using a phase field function and to incorporate a penalty term into the energies to weakly enforce boundary conditions on the colloid's boundary.

### 5.2.1 Colloid representation

Assume the colloid is given by an open set  $\Omega_c \subset\subset \Omega$ , and let  $\hat{\Omega}_c \subset \mathbb{R}^d$  be a reference shape such that there is an affine parametrization  $\mathbf{F} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ,

$$\hat{\mathbf{x}} = \mathbf{F}(\mathbf{x}) = \mathbf{Z}\mathbf{x} + \mathbf{b}, \quad \hat{\Omega}_c = \mathbf{F}(\Omega_c).$$

Above,  $\mathbf{Z}$  is a rotation matrix and  $\mathbf{b}$  a translation vector. We also use the signed distance functions to  $\partial\hat{\Omega}_c$  and  $\partial\Omega_c$ , that we denote by  $d$  and  $\hat{d}$  respectively, and are related by

$$d(\mathbf{x}) = d(\mathbf{F}(\hat{\mathbf{x}})) = \hat{d}(\hat{\mathbf{x}}), \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

Applying the chain rule, we also deduce the identity

$$\nabla_{\mathbf{x}} d(\mathbf{x}) = \nabla_{\hat{\mathbf{x}}} \hat{d}(\hat{\mathbf{x}}) \mathbf{Z}.$$

Next, we introduce a phase field function to approximate the colloidal domain. Given  $\epsilon > 0$ , that will represent the thickness of the transition, we consider

$$\phi_\epsilon^{\text{ref}} : \mathbb{R} \rightarrow (-1, 1), \quad \phi_\epsilon^{\text{ref}}(t) = \frac{1}{2} \left( \frac{2}{\pi} \arctan \left( -\frac{t}{\epsilon} \right) + 1 \right).$$

Using this reference phase field function, we define

$$\phi_\epsilon(\mathbf{x}) = \phi_\epsilon^{\text{ref}}(d(\mathbf{x})) = \phi_\epsilon^{\text{ref}}(\hat{d}(\hat{\mathbf{x}})),$$

that yields

$$|\nabla_{\mathbf{x}} \phi_\epsilon(\mathbf{x})|^2 = \left( \frac{1}{\pi \epsilon} \right)^2 \frac{1}{\left( 1 + \left( \frac{\hat{d}(\mathbf{F}(\mathbf{x}))}{\epsilon} \right)^2 \right)^2} |\nabla_{\hat{\mathbf{x}}} \hat{d}(\mathbf{F}(\mathbf{x}))|^2.$$

In order to motivate the penalty term that will account for the colloidal inclusion, we note a relation between bulk and surface integrals. Given  $f \in C(\overline{\Omega})$ , let

$$J_\epsilon(f) := \epsilon \frac{|\mathbb{S}^{d-1}|}{2} \int_{\Omega} f(\mathbf{x}) |\nabla \phi_\epsilon(\mathbf{x})|^2 d\mathbf{x}. \quad (152)$$

Then, in the limit  $\epsilon \rightarrow 0$ ,  $J_\epsilon$  recovers the surface integral of  $f$ ,

$$\lim_{\epsilon \rightarrow 0} J_\epsilon(f) = \int_{\partial\Omega_c} f(\mathbf{x}) dS(\mathbf{x}). \quad (153)$$

### 5.2.2 Weak anchoring

Boundary conditions can either be imposed by a Dirichlet condition (*strong anchoring*) or by an energetic penalization term (*weak anchoring*). Indeed, in some physical situations, weak anchoring is a better reflection of the physics (de Gennes and Prost, 1995; Virga, 1994). We take advantage of this for modeling colloids (Araki and Tanaka, 2006; Conradi et al., 2009; Čopar et al., 2014, 2015).

Specifically, we incorporate penalization terms  $E_{\text{erk,a}}$  and  $E_{\text{uni,a}}$  into either  $E_{\text{erk,one}}$  or  $E_{\text{uni}}$ , and corresponding terms in the discrete energies. In the Q-tensor model, a standard approach is to add the energy term

$$J_v(\mathbf{Q}) = \frac{K_a^v}{2} \int_{\partial\Omega_c} |\mathbf{Q} - \mathbf{Q}^v|^2 dS(\mathbf{x}), \quad (154)$$

where  $\mathbf{Q}^v$  is the preferred state for  $\mathbf{Q}$  on the boundary of the colloid, which is imposed by an energetic penalization with  $K_a^v$  as the weighting term. For example,  $\mathbf{Q}^v$  may have the form (Mottram and Newton, 2014)

$$\mathbf{Q}^v = s^* \left( \mathbf{v} \otimes \mathbf{v} - \frac{1}{d} \mathbf{I} \right), \quad (155)$$

which is a uniaxial tensor, where  $\mathbf{v}$  is the unit vector normal to  $\partial\Omega_c$ ; this is called a uniaxial, homeotropic (normal anchoring) condition.

Another popular weak anchoring condition is called *planar degenerate anchoring*, whose purpose is to enforce a uniaxial state at the boundary with the

director orthogonal to  $\mathbf{v}$  (Fournier and Galatola, 2005; Ravnik and Žumer, 2009; Changizrezaei and Denniston, 2017). Let

$$\tilde{\mathbf{Q}} := \mathbf{Q} + \frac{s}{d} \mathbf{I}, \quad \tilde{\mathbf{Q}}^\perp := [\mathbf{I} - \mathbf{v} \otimes \mathbf{v}] \tilde{\mathbf{Q}} [\mathbf{I} - \mathbf{v} \otimes \mathbf{v}]; \quad (156)$$

we point out that, with our notation,  $\tilde{\mathbf{Q}} = s\Theta = \mathbf{U}$ . Thus, we include the following energy term

$$J_\perp(\mathbf{Q}) = \frac{K_{a,1}^\perp}{2} \int_{\partial\Omega_c} |\tilde{\mathbf{Q}} - \tilde{\mathbf{Q}}^\perp|^2 dS(\mathbf{x}) + \frac{K_{a,2}^\perp}{2} \int_{\partial\Omega_c} \left( |\tilde{\mathbf{Q}}|^2 - (s^*)^2 \right)^2 dS(\mathbf{x}), \quad (157)$$

where the quartic term is necessary in the standard LdG model to fully enforce a uniaxial state (Fournier and Galatola, 2005, eqn. (4)) when  $d = 3$ .

**Remark 6.** *For some LC materials, in certain specialized experimental conditions, some biaxiality can be observed near the boundary despite using a uniaxial boundary condition (Sluckin and Poniewierski, 1985; Fournier and Galatola, 2005).*

Let us now consider the effect of imposing the uniaxial constraint  $\mathbf{Q} = s(\Theta - \frac{1}{d}\mathbf{I})$  on the weak anchoring energies. Starting with normal anchoring (154), (155), we expand  $|\mathbf{Q} - \mathbf{Q}^\mathbf{v}|^2$ , exploiting that  $\mathbf{Q}, \mathbf{Q}^\mathbf{v}$  are uniaxial (cf. (9)), symmetric, and that  $|\Theta| = 1, |\mathbf{v}| = 1$ , and so obtain

$$|\mathbf{Q} - \mathbf{Q}^\mathbf{v}|^2 = 2ss^* \left( |\Theta|^2 |\mathbf{v}|^2 - \Theta \mathbf{v} \cdot \Theta \mathbf{v} \right) + \frac{d-1}{d} (s - s^*)^2 |\Theta|^2. \quad (158)$$

Since  $\mathbf{v} = \nabla\phi_\epsilon/|\nabla\phi_\epsilon|$  on  $\partial\Omega_c$ , we combine the identity above with (152) and (153) to introduce the continuous weak normal anchoring energy for the uniaxially constrained Landau-de Gennes model:

$$\begin{aligned} J_\mathbf{v}[s, \Theta] := & \frac{K_a^\mathbf{v}}{2} |\mathbb{S}^{d-1}| \epsilon \int_{\Omega} 2ss^* \left( |\Theta|^2 |\nabla\phi_\epsilon|^2 - \Theta \nabla\phi_\epsilon \cdot \Theta \nabla\phi_\epsilon \right) \\ & + \frac{K_a^\mathbf{v}}{2} |\mathbb{S}^{d-1}| \epsilon \int_{\Omega} |\nabla\phi_\epsilon|^2 \frac{d-1}{d} (s - s^*)^2 |\Theta|^2. \end{aligned} \quad (159)$$

To better see the structure of (158), we write  $\Theta = \mathbf{n} \otimes \mathbf{n}$ , use that  $|\Theta|^2 = |\mathbf{n}|^2$ , and get

$$\begin{aligned} |\mathbf{Q} - \mathbf{Q}^\mathbf{v}|^2 &= 2ss^* \left( |\mathbf{n}|^2 |\mathbf{v}|^2 - (\mathbf{n} \cdot \mathbf{v})^2 \right) + \frac{d-1}{d} (s - s^*)^2 |\mathbf{n}|^2 \\ &= \mathbf{n}^T \left[ 2ss^* (\mathbf{I} - \mathbf{v} \otimes \mathbf{v}) + \frac{d-1}{d} (s - s^*)^2 \mathbf{I} \right] \mathbf{n} \\ &= \mathbf{n}^T \left[ \left( 2ss^* + \frac{d-1}{d} (s - s^*)^2 \right) (\mathbf{I} - \mathbf{v} \otimes \mathbf{v}) \right] \mathbf{n} \end{aligned} \quad (160)$$

$$+ \frac{d-1}{d}(s-s^*)^2(\mathbf{v} \otimes \mathbf{v}) \Bigg] \mathbf{n} =: \mathbf{n}^T H^{\mathbf{v}} \mathbf{n}.$$

It follows immediately from this identity that the matrix  $H^{\mathbf{v}}$  is uniformly positive semi-definite. Therefore, for the Ericksen model, this motivates to consider the weak normal anchoring energy

$$\begin{aligned} J_{\mathbf{v}}[s, \mathbf{n}] := & \frac{K_a^{\mathbf{v}}}{2} |\mathbb{S}^{d-1}| \epsilon \int_{\Omega} 2ss^* \left( |\mathbf{n}|^2 |\nabla \phi_{\epsilon}|^2 - (\mathbf{n} \cdot \nabla \phi_{\epsilon})^2 \right) \\ & + \frac{K_a^{\mathbf{v}}}{2} |\mathbb{S}^{d-1}| \epsilon \int_{\Omega} |\nabla \phi_{\epsilon}|^2 \frac{d-1}{d} (s-s^*)^2 |\mathbf{n}|^2. \end{aligned} \quad (161)$$

Next, we proceed similarly for the weak planar degenerate anchoring (156), (157). Expanding, and using that  $\Theta^2 = \Theta$ , we get

$$\begin{aligned} |\tilde{\mathbf{Q}} - \tilde{\mathbf{Q}}^{\perp}|^2 &= s^2 \left| \Theta - [\mathbf{I} - \mathbf{v} \mathbf{v}^T] [\Theta - (\Theta \mathbf{v}) \mathbf{v}^T] \right|^2 \\ &= s^2 \left[ 2|\Theta \mathbf{v}|^2 - (\mathbf{v}^T \Theta \mathbf{v})^2 \right] \\ &= s^2 (\Theta : \mathbf{v} \otimes \mathbf{v}) [2 - (\Theta : \mathbf{v} \otimes \mathbf{v})], \end{aligned} \quad (162)$$

and

$$(|\tilde{\mathbf{Q}}|^2 - (s^*)^2)^2 = (s^2 |\Theta|^2 - (s^*)^2)^2 = (s - s^*)^2 (s + s^*)^2, \quad (163)$$

which yields a slightly complicated energy functional for imposing planar anchoring with a desired degree of orientation,  $s^*$ . At this point, it is worthwhile to revisit the modeling assumptions made in posing (157). The main motivation for choosing (157) is to enforce planar degenerate anchoring *with a uniaxiality constraint*. However, our approach enforces uniaxiality in a more explicit way, so other energy penalization terms may be used to achieve planar anchoring.

Indeed,  $(\Theta : \mathbf{v} \otimes \mathbf{v})^2 \ll (\Theta : \mathbf{v} \otimes \mathbf{v})$  when  $|\Theta : \mathbf{v} \otimes \mathbf{v}|$  is small, e.g. when planar anchoring is achieved. Hence, it is reasonable to make the following approximation

$$|\tilde{\mathbf{Q}} - \tilde{\mathbf{Q}}^{\perp}|^2 \approx 2s^2 (\Theta : \mathbf{v} \otimes \mathbf{v}). \quad (164)$$

Moreover, we can replace (163) by  $(2s^*)^2 (s - s^*)^2$  as a simpler way to enforce the degree of orientation on the surface. Therefore, combining with the phase-field approach, we assume the following continuous weak planar degenerate anchoring energy for the uniaxially constrained Landau-de Gennes model:

$$\begin{aligned} J_{\perp}[s, \Theta] := & \frac{K_{a,1}^{\perp}}{2} |\mathbb{S}^{d-1}| \epsilon \int_{\Omega} 2s^2 (\nabla \phi_{\epsilon} \cdot \Theta \nabla \phi_{\epsilon}) \\ & + \frac{K_{a,2}^{\perp}}{2} |\mathbb{S}^{d-1}| \epsilon \int_{\Omega} |\nabla \phi_{\epsilon}|^2 (2s^*)^2 (s - s^*)^2 |\Theta|^2. \end{aligned} \quad (165)$$

Furthermore, writing  $\Theta = \mathbf{n} \otimes \mathbf{n}$ , we have

$$\begin{aligned} J_{\perp}[s, \mathbf{n}] := & \frac{K_{a,1}^{\perp}}{2} |\mathbb{S}^{d-1}| \epsilon \int_{\Omega} 2s^2 (\mathbf{n} \cdot \nabla \phi_{\epsilon})^2 \\ & + \frac{K_{a,2}^{\perp}}{2} |\mathbb{S}^{d-1}| \epsilon \int_{\Omega} |\nabla \phi_{\epsilon}|^2 (2s^*)^2 (s - s^*)^2 |\mathbf{n}|^2. \end{aligned} \quad (166)$$

Then, we can define the anchoring energies for the uniaxially constrained Landau-de Gennes and the Ericksen models respectively by  $E_{\text{uni},a}[s, \Theta] := J_{\nu}[s, \Theta] + J_{\perp}[s, \Theta]$  and  $E_{\text{erk},a}[s, \mathbf{n}] := J_{\nu}[s, \mathbf{n}] + J_{\perp}[s, \mathbf{n}]$ . In case  $K_{a,1}^{\perp} = K_{a,2}^{\perp} = 0$  (resp.  $K_a^{\nu} = 0$ ), this yields weak normal (resp. weak planar degenerate) anchoring; otherwise, it gives rise to a weak oblique anchoring.

Clearly, if  $\Theta = \mathbf{n} \otimes \mathbf{n}$  in  $\Omega$ , then  $E_{\text{uni},a}[s, \Theta] \equiv E_{\text{erk},a}[s, \mathbf{n}]$ . Note that the energies (161) and (166) are insensitive to changes in the sign of  $\mathbf{n}$ . With this, we seek to minimize the total energies

$$\begin{aligned} E_{\text{erk},\text{one}}[s, \mathbf{n}] &:= E_{\text{erk}-m}[s, \mathbf{n}] + E_{\text{erk},\text{bulk}}[s] + E_{\text{erk},a}[s, \mathbf{n}], \\ E_{\text{uni}}[s, \Theta] &:= E_{\text{uni}-m}[s, \Theta] + E_{\text{LdG},\text{bulk}}[s] + E_{\text{uni},a}[s, \Theta], \end{aligned}$$

under suitable boundary conditions.

Next, we give a discrete counterpart of  $E_{\text{erk},a}[s, \mathbf{n}]$ . For convenience, we define the following discrete bilinear forms:

$$\begin{aligned} a_h^{\mathbf{n}}(\mathbf{n}_h, \mathbf{v}_h) &:= K_a^{\nu} \int_{\Omega} \mathcal{I}_h \left\{ 2s_h s^* \left( (\mathbf{n}_h \cdot \mathbf{v}_h) |\nabla \phi_{\epsilon}|^2 - (\nabla \phi_{\epsilon} \cdot \mathbf{n}_h) (\nabla \phi_{\epsilon} \cdot \mathbf{v}_h) \right) \right. \\ &\quad \left. + |\nabla \phi_{\epsilon}|^2 \frac{d-1}{d} (s_h - s^*)^2 (\mathbf{n}_h \cdot \mathbf{v}_h) \right\} \\ &\quad + K_{a,1}^{\perp} \int_{\Omega} \mathcal{I}_h \left\{ 2s_h^2 (\mathbf{n}_h \cdot \nabla \phi_{\epsilon}) (\mathbf{v}_h \cdot \nabla \phi_{\epsilon}) \right\} \\ &\quad + K_{a,2}^{\perp} \int_{\Omega} \mathcal{I}_h \left\{ |\nabla \phi_{\epsilon}|^2 4(s^*)^2 (s_h - s^*)^2 (\mathbf{n}_h \cdot \mathbf{v}_h) \right\}, \\ a_h^s(s_h, z_h) &:= K_a^{\nu} \int_{\Omega} \mathcal{I}_h \left\{ s_h z_h |\nabla \phi_{\epsilon}|^2 \frac{d-1}{d} |\mathbf{n}_h|^2 \right\} \\ &\quad + K_{a,1}^{\perp} \int_{\Omega} \mathcal{I}_h \left\{ s_h z_h 2 (\mathbf{n}_h \cdot \nabla \phi_{\epsilon})^2 \right\} \\ &\quad + K_{a,2}^{\perp} \int_{\Omega} \mathcal{I}_h \left\{ s_h z_h |\nabla \phi_{\epsilon}|^2 4(s^*)^2 |\mathbf{n}_h|^2 \right\}, \\ \omega_h^s(z_h) &:= K_a^{\nu} \int_{\Omega} \mathcal{I}_h \left\{ 2z_h s^* \left( |\mathbf{n}_h|^2 |\nabla \phi_{\epsilon}|^2 - (\mathbf{n}_h \cdot \nabla \phi_{\epsilon})^2 \right) \right\} \\ \zeta_h^s(z_h) &:= K_a^{\nu} \int_{\Omega} \mathcal{I}_h \left\{ z_h s^* |\nabla \phi_{\epsilon}|^2 \frac{d-1}{d} |\mathbf{n}_h|^2 \right\} \\ &\quad + K_{a,2}^{\perp} \int_{\Omega} \mathcal{I}_h \left\{ z_h s^* |\nabla \phi_{\epsilon}|^2 4(s^*)^2 |\mathbf{n}_h|^2 \right\}, \end{aligned} \quad (167)$$

where  $\mathcal{I}_h$  is the Lagrange interpolant. These expressions correspond to using the so-called mass lumping quadrature which, for all  $f \in C^0(\bar{\Omega})$ , reads

$$\int_{\Omega} I_h f = \sum_{T \in \mathcal{T}_h} \int_T I_h f = \sum_{T \in \mathcal{T}_h} \frac{|T|}{d+1} \sum_{i=1}^{d+1} f(x_T^i), \quad (168)$$

where  $\{x_T^i\}_{i=1}^{d+1}$  are the vertices of  $T$ . This quadrature rule is exact for piecewise linear polynomials and has the advantage that the finite element realization of (167) is a *diagonal* matrix, which induces the following monotonicity result (proved in Nocet et al., 2018, Lem. 6).

**Lemma 15** (Monotone property for lumped mass matrix). *Let  $m_h : \mathbb{U}_h \times \mathbb{U}_h \rightarrow \mathbb{R}$  be a bilinear form defined by*

$$m_h(\mathbf{n}_h, \mathbf{v}_h) := \int_{\Omega} I_h [\mathbf{n}_h \cdot H(x) \mathbf{v}_h] dx,$$

where  $H$  is a continuous  $d \times d$  symmetric positive semi-definite matrix. If  $|\mathbf{n}_h(x_i)| \geq 1$  at all nodes  $x_i$  in  $\mathcal{N}_h$ , then

$$m_h(\mathbf{n}_h, \mathbf{n}_h) \geq m_h \left( \frac{\mathbf{n}_h}{|\mathbf{n}_h|}, \frac{\mathbf{n}_h}{|\mathbf{n}_h|} \right).$$

To apply Lemma 15 to the first bilinear form in (167) we observe that  $H = H^v + H_1^\perp + H_2^\perp$ , where  $H^v$  is given in (160), and

$$H_1^\perp = 2s_h^2 \nabla \phi_\epsilon \otimes \nabla \phi_\epsilon, \quad H_2^\perp = |\nabla \phi_\epsilon|^2 4(s^*)^2 (s_h - s^*)^2 \mathbf{I}.$$

Since  $H^v$ ,  $H_1^\perp$ ,  $H_2^\perp$  are all positive semi-definite,  $H$  is symmetric positive semi-definite, thus

$$a_h^{\mathbf{n}}(\mathbf{n}_h, \mathbf{n}_h) \geq a_h^{\mathbf{n}} \left( \frac{\mathbf{n}_h}{|\mathbf{n}_h|}, \frac{\mathbf{n}_h}{|\mathbf{n}_h|} \right). \quad (169)$$

Therefore, we take the discrete weak anchoring energy to be

$$E_{\text{erk,a}}^h[s_h, \mathbf{n}_h] := |\mathbb{S}^{d-1}| \epsilon \frac{a_h^{\mathbf{n}}(\mathbf{n}_h, \mathbf{n}_h)}{2}, \quad (170)$$

and the discrete total energy is then given by

$$\begin{aligned} E_{\text{erk,one}}^h[s_h, \mathbf{n}_h] &:= E_{\text{erk-m}}^h[s_h, \mathbf{n}_h] + E_{\text{erk,bulk}}^h[s_h] + E_{\text{erk,a}}^h[s_h, \mathbf{n}_h], \\ E_{\text{uni}}^h[s_h, \Theta_h] &:= E_{\text{uni-m}}^h[s_h, \Theta_h] + E_{\text{LdG,bulk}}^h[s_h] + E_{\text{uni,a}}^h[s_h, \Theta_h], \end{aligned}$$

again noting that  $E_{\text{uni,a}}^h[s_h, \Theta_h] \equiv E_{\text{erk,a}}^h[s_h, \mathbf{n}_h]$ .

Because  $a_h^{\mathbf{n}}(\mathbf{n}_h, \mathbf{n}_h) = a_h^s(s_h, s_h) + \omega_h^s(s_h) + \zeta_h^s(s^* - 2s_h)$ , a straightforward calculation yields, for  $z_h \in \mathbb{S}_h(\Gamma_s, 0)$  and  $\mathbf{v}_h \in \mathbb{V}_h^\perp(\mathbf{n}_h) \cap \mathbb{U}_h(\Gamma_{\mathbf{u}}, \mathbf{0})$ ,

$$\begin{aligned}\delta_{s_h} E_{\text{erk,a}}^h[s_h, \mathbf{n}_h; z_h] &= |\mathbb{S}^{d-1}| \epsilon \left( a_h^s(s_h, z_h) + \frac{\omega_h^s(z_h)}{2} - \zeta_h^s(z_h) \right), \\ \delta_{\mathbf{n}_h} E_{\text{erk,a}}^h[s_h, \mathbf{n}_h; \mathbf{v}_h] &= |\mathbb{S}^{d-1}| \epsilon a_h^{\mathbf{n}}(\mathbf{n}_h, \mathbf{v}_h),\end{aligned}\quad (171)$$

where  $\mathbb{S}_h(\Gamma_s, 0)$  and  $\mathbb{U}_h(\Gamma_{\mathbf{u}}, \mathbf{0})$  are defined in (60), and  $\mathbb{V}_h^\perp(\mathbf{n}_h)$  is given by (86). Thus, for the computation of discrete minimizers, the first variation formulas (171) must be incorporated into the algorithm described in Section 3.3.3.

**Remark 7.** Since  $E_{\text{uni,a}}^h[s_h, \Theta_h] \equiv E_{\text{erk,a}}^h[s_h, \mathbf{n}_h]$  (because  $\Theta_h = \mathcal{I}_h \mathbf{n}_h \otimes \mathbf{n}_h$ ), proving  $\Gamma$ -convergence for the discrete energy with weak anchoring  $E_{\text{erk,a}}^h[s_h, \mathbf{n}_h]$  is exactly the same as in Nocettono *et al.* (2018, Sec. 8).

### 5.2.3 Computational colloid example

We simulate a Saturn-ring defect by using the phase field approach described in Section 5.2.2. More precisely, we consider the double-well potential (151) with  $\eta_B = 1/16$ , and represent a spherical colloidal inclusion centered at  $(0.5, 0.5, 0.5)$  with radius 0.2 by means of a phase field function with  $\epsilon = 6 \times 10^{-2}$ . The domain is  $\Omega = (0, 1)^3$ , and we set homogeneous Neumann conditions on  $\Gamma_o := \overline{\Omega} \cap (\{z = 0\} \cup \{z = 1\})$ , and the Dirichlet boundary conditions

$$s = s^*, \quad \mathbf{n}(x, y) = (0, 0, 1), \quad \Theta = \mathbf{n} \otimes \mathbf{n}$$

on  $\Gamma_s = \Gamma_\Theta = \partial\Omega \setminus \Gamma_o$ .

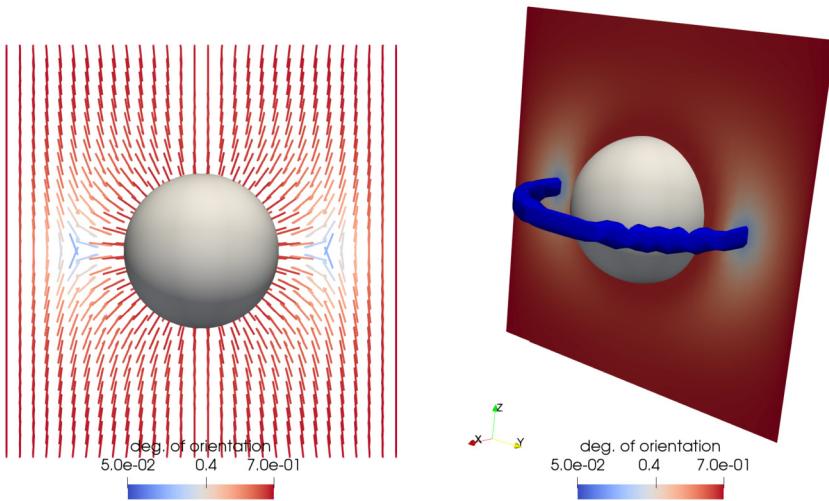
Fig. 12 shows the result of the gradient flow algorithm described in Section 5.2.2 with time-step  $\delta t = 10^{-2}$  and initialized with

$$s = s^*, \quad \mathbf{n}(x, y) = (0, 0, 1), \quad \Theta = \mathbf{n} \otimes \mathbf{n}.$$

The double-well potential and boundary conditions on  $\partial\Omega$  are essentially the same as in the experiment described in Section 5.1 for the uniaxially-constrained Landau-de Gennes model; therefore, it is no surprise that the results are similar to those illustrated in Fig. 11.

## 6 Electric fields

The LC models can be augmented by considering external forces acting on them. Here we discuss the incorporation of an electric field into the Ericksen (resp. Landau-de Gennes) model. This is achieved by adding another term to the energies  $E_{\text{erk,one}}$  (resp.  $E_{\text{uni}}$ ).



**FIGURE 12** Simulation results for the uniaxially-constrained Landau-de Gennes model under the setting described in Section 5.2.3. The left panel displays the line field  $\Theta$  on the plane  $\{x = 0.5\}$ , with color scale based on the degree of orientation  $s$ . Two defects of order  $-1/2$  are visible on the sides of the colloid. The right panel shows the degree of orientation on the same plane, and the isosurface  $s = 0.25$  is depicted in blue.

## 6.1 Modified energies

We now consider energies of the form

$$\begin{aligned} E_{\text{erk,one}}[s, \mathbf{n}] &= E_{\text{erk-m}}[s, \mathbf{n}] + E_{\text{erk,bulk}}[s] + E_{\text{erk,ext}}[s, \mathbf{n}] \\ E_{\text{uni}}[s, \Theta] &= E_{\text{uni-m}}[s, \Theta] + E_{\text{LdG,bulk}}[s] + E_{\text{uni,ext}}[s, \Theta], \end{aligned} \quad (172)$$

where we represent the external field energies in either the Ericksen and Landau-de Gennes models by  $E_{\text{erk,ext}}[s, \mathbf{n}]$  and  $E_{\text{uni,ext}}[s, \Theta]$ , respectively. Given an electric field  $\mathbf{E}$ , we consider (Biscari and Cesana, 2007; de Gennes and Prost, 1995)

$$E_{\text{erk,ext}}[s, \mathbf{n}] := -\frac{K_{\text{ext}}}{2} \left( \bar{\varepsilon} \int_{\Omega} (1 - s\gamma_a) |\mathbf{E}|^2 + \varepsilon_a \int_{\Omega} s (\mathbf{E} \cdot \mathbf{n})^2 \right), \quad (173)$$

$$E_{\text{uni,ext}}[s, \Theta] := -\frac{K_{\text{ext}}}{2} \left( \bar{\varepsilon} \int_{\Omega} (1 - s\gamma_a) |\mathbf{E}|^2 + \varepsilon_a \int_{\Omega} s \Theta \mathbf{E} \cdot \mathbf{E} \right). \quad (174)$$

Above, the constant  $K_{\text{ext}}$  is a weighting parameter. If we let  $\varepsilon_{\parallel}$ ,  $\varepsilon_{\perp}$  be the dielectric permittivities in the directions parallel and orthogonal to the LC molecules, then  $\bar{\varepsilon} = (\varepsilon_{\parallel} + (d-1)\varepsilon_{\perp})/d$  is the average dielectric permittivity and  $\varepsilon_a = \varepsilon_{\parallel} - \varepsilon_{\perp}$  is the dielectric anisotropy. Finally,  $\gamma_a = \varepsilon_a/(d\bar{\varepsilon})$  is a dimensionless ratio; whenever  $0 \leq \varepsilon_{\perp} \leq \varepsilon_{\parallel}$ , it must be  $0 \leq \gamma_a \leq 1$ . Note that the definition of the dielectric constants here accounts for the dimension  $d$ .

From (173) and (174), it is evident that, independently of  $s$  and the electric constants, if  $\Theta = \mathbf{n} \otimes \mathbf{n}$  then  $E_{\text{erk,ext}}[s, \mathbf{n}] \equiv E_{\text{uni,ext}}[s, \Theta]$ . Thus, our treatment of both energies follows the same pattern.

We point out that, although the second integrals in (173) and (174) are bounded, they may be negative. Hence, some care is required in discretizing the electric energy in order to preserve our energy decreasing minimization scheme. First, define a discrete bilinear form analogous to (167):

$$e_h(s_h, \mathbf{n}_h, \mathbf{v}_h) = \int_{\Omega} \mathcal{I}_h \left[ |\varepsilon_a| |\mathbf{E}|^2 (\mathbf{n}_h \cdot \mathbf{v}_h) - \varepsilon_a s_h (\mathbf{E} \cdot \mathbf{n}_h) (\mathbf{E} \cdot \mathbf{v}_h) \right]. \quad (175)$$

To apply Lemma 15, we see that the matrix  $H$  reads

$$H = |\varepsilon_a| |\mathbf{E}|^2 \mathbf{I} - \varepsilon_a s_h \mathbf{E} \otimes \mathbf{E},$$

and is therefore symmetric and positive semi-definite since  $|s_h| \leq 1$ . Consequently, whenever  $|\mathbf{n}_h| \geq 1$ ,

$$e_h(s_h, \mathbf{n}_h, \mathbf{n}_h) \geq e_h \left( s_h, \frac{\mathbf{n}_h}{|\mathbf{n}_h|}, \frac{\mathbf{n}_h}{|\mathbf{n}_h|} \right). \quad (176)$$

We now define the discrete counterpart of (172) to be

$$\begin{aligned} E_{\text{erk,one}}^h[s_h, \mathbf{n}_h] &:= E_{\text{erk-m}}^h[s_h, \mathbf{n}_h] + E_{\text{erk,bulk}}^h[s_h] \\ &\quad + E_{\text{erk,a}}^h[s_h, \mathbf{n}_h] + E_{\text{erk,ext}}^h[s_h, \mathbf{n}_h], \end{aligned} \quad (177)$$

where the discrete electric energy is similar to (173) and is given by

$$E_{\text{erk,ext}}^h[s_h, \mathbf{n}_h] = \frac{K_{\text{ext}}}{2} \left( -\bar{\varepsilon} \int_{\Omega} (1 - s_h \gamma_a) |\mathbf{E}|^2 + e_h(s_h, \mathbf{n}_h, \mathbf{n}_h) - |\varepsilon_a| \int_{\Omega} |\mathbf{E}|^2 \right). \quad (178)$$

Observe that (178) is an approximation of

$$\begin{aligned} E_{\text{erk,ext}}^h[s_h, \mathbf{n}_h] &= \frac{K_{\text{ext}}}{2} \left( -\bar{\varepsilon} \int_{\Omega} (1 - s_h \gamma_a) |\mathbf{E}|^2 - \varepsilon_a \int_{\Omega} s_h (\mathbf{E} \cdot \mathbf{n}_h)^2 \right. \\ &\quad \left. + |\varepsilon_a| \int_{\Omega} |\mathbf{E}|^2 (|\mathbf{n}_h|^2 - 1) \right), \end{aligned} \quad (179)$$

where the ‘‘extra’’ term is non-positive and *consistent* (i.e. it vanishes as  $h \rightarrow 0$  provided the singular set  $\mathcal{S}$  has zero Lebesgue measure). Moreover,  $\int_{\Omega} |\mathbf{E}|^2 |\mathbf{n}_h|^2$  is constant at the continuous level, whence the extra term does not fundamentally change the energy. However, it is needed to ensure the projection step in the algorithm decreases the (discrete) energy, which is guaranteed by (176).

We take first order variations of  $E_{\text{erk,ext}}^h$  in the directions  $z_h \in \mathbb{S}_h(\Gamma_s, 0)$  and  $\mathbf{v}_h \in \mathbb{V}_h^\perp(\mathbf{n}_h) \cap \mathbb{U}_h(\Gamma_u, \mathbf{0})$ , to obtain

$$\begin{aligned}\delta_{s_h} E_{\text{erk,ext}}^h[s_h, \mathbf{n}_h; z_h] &= \frac{K_{\text{ext}}}{2} \left( \bar{\varepsilon} \int_{\Omega} z_h \gamma_a |\mathbf{E}|^2 - \varepsilon_a \int_{\Omega} \mathcal{I}_h [z_h (\mathbf{E} \cdot \mathbf{n}_h)^2] \right), \\ \delta_{\mathbf{n}_h} E_{\text{erk,ext}}^h[s_h, \mathbf{n}_h; \mathbf{v}_h] &= K_{\text{ext}} e_h(s_h, \mathbf{n}_h, \mathbf{v}_h).\end{aligned}$$

**Remark 8.** Since  $E_{\text{uni,ext}}^h[s_h, \Theta_h] \equiv E_{\text{erk,ext}}^h[s_h, \mathbf{n}_h]$  (because  $\Theta_h = \mathcal{I}_h \mathbf{n}_h \otimes \mathbf{n}_h$ ), proving  $\Gamma$ -convergence for the discrete energy with the electric field contribution  $E_{\text{erk,ext}}^h[s_h, \mathbf{n}_h]$  is exactly the same as in Nochetto et al. (2018, Sec. 8).

## 6.2 Computational electric field example

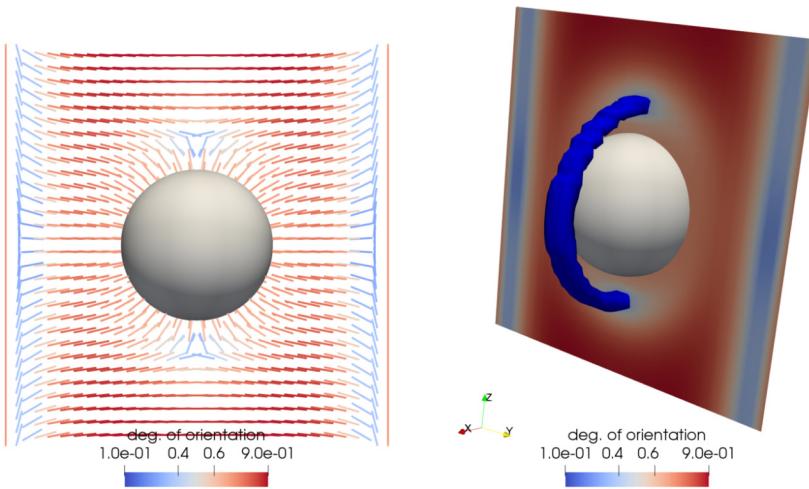
We illustrate the effect of an electric field on the same configuration as in Section 5.2.3. Namely, with the same colloidal inclusion and boundary conditions as there, we incorporate the effect of a constant electric field  $\mathbf{E} = (0, 1, 0)$ . We set the parameter  $K_{\text{ext}} = 160.0$ , and the material constants  $\varepsilon_{\parallel} = 7/3$ ,  $\varepsilon_{\perp} = 1/3$ , that yield  $\bar{\varepsilon} = 1$ ,  $\varepsilon_a = 2$ ,  $\gamma_a = 2/3$  in (178).

The results of our simulation, with the same gradient flow setting as in Section 5.2.3, are shown in Figs. 13 and 14. The presence of a strong electric force creates two noticeable effects. Clearly, the electric energy (174) is minimized whenever the field  $\Theta$  is aligned with  $\mathbf{E}$ ; thus, the LC molecules tend to deflect to the  $y$ -axis in the domain. This creates a Freedericksz-type transition (Biscari and Cesana, 2007; Hoogboom et al., 2007; Nochetto et al., 2018), in which the director field deflects towards the  $y$ -axis to better align with the electric field and this, in turn, gives rise to a defect region near the sides of the cube, on which  $\Theta$  is set to be vertical. Secondly, the Saturn-ring defect observed in Fig. 13 is rotated. Instead of having a rotation axis parallel to the  $z$ -axis, the ring has a rotation axis parallel to the  $y$ -axis. Fig. 14 shows alternative views of the simulation.

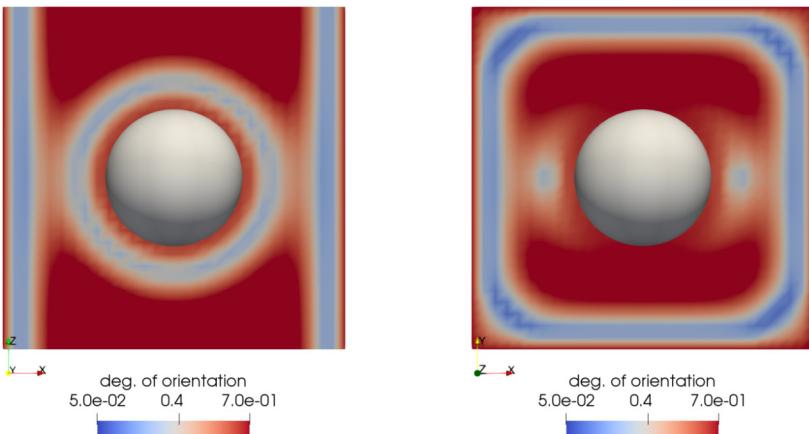
## 7 The Landau-de Gennes model with and without the uniaxial constraint

Our uniaxially constrained LdG model allows us to probe the fundamental modeling issue raised earlier in Section 1.1. Does uniaxiality significantly affect the minimizing configuration? To the best of our knowledge, our method is the first to simulate the LdG model with *uniaxiality enforced as a hard constraint*. Thus, we can do a direct *quantitative* comparison of the “standard” LdG approach against the uniaxially constrained case.

We revisit the Saturn-ring example in Section 5.1. In particular, we use the boundary conditions in (150) and the double-well potential in (151) for the uniaxially constrained model in (105). For the standard (one-constant) LdG model in (17), we use the following boundary conditions



**FIGURE 13** Computational results for the uniaxially-constrained Landau-de Gennes model under the presence of an electric field and a colloidal inclusion (represented using a phase field approach). The left panel shows the line field  $\Theta$  on the slice  $\{x = 0.5\}$ , while the right panel depicts the degree of orientation on the same plane and the isosurface  $s = 0.4$ . As opposed to Fig. 12, the two  $-1/2$ -degree defects are situated on top and bottom of the colloidal particle. The strong electric field also creates a large defect region near the sides of the cube.



**FIGURE 14** Another visualization of the experiment in Fig. 13. The left panel shows the degree of orientation in the  $\{y = 0.5\}$  plane, while the right panel shows the  $\{z = 0.5\}$  plane. The Saturn-ring is clearly aligned with the  $y$ -axis, and one can see a *secondary* line defect near the walls of the domain aligned with the  $z$ -axis.

$$\begin{aligned} \mathbf{Q} &= s^* \left( \mathbf{v} \otimes \mathbf{v} - \frac{1}{3} \mathbf{I} \right) \text{ on } \Gamma_i, \\ \mathbf{Q} &= s^* \left( (0, 0, 1) \otimes (0, 0, 1) - \frac{1}{3} \mathbf{I} \right) \text{ on } \Gamma_o, \end{aligned} \quad (180)$$

which is consistent with the boundary conditions in (150). Moreover, the double-well potential is given by (15), (43), where

$$\begin{aligned} K &= 1.0, & A &= -7.502104, & B &= 60.975813, \\ C &= 66.519069, & D &= 552.230967, \end{aligned} \quad (181)$$

which is consistent with the double well potential (151). The initial guess for the standard LdG model is chosen to be the minimizer  $\mathbf{Q}_{\text{uni}}$  of the uniaxial model. Both models were simulated using the following set of values for  $\eta_B$ :  $\{0.25, 0.16, 0.09, 0.04\}$ .

Table 1 shows a comparison of the energy  $E_{\text{uni}}[\mathbf{Q}_{\text{uni}}]$  with  $E_{\text{LdG,one}}[\mathbf{Q}_{\text{uni}}]$ . The relative error is small, but not zero, because the two numerical models are different, i.e. the error is purely due to numerical discretization and a finite mesh size. This table illustrates that the two numerical models are consistently implemented.

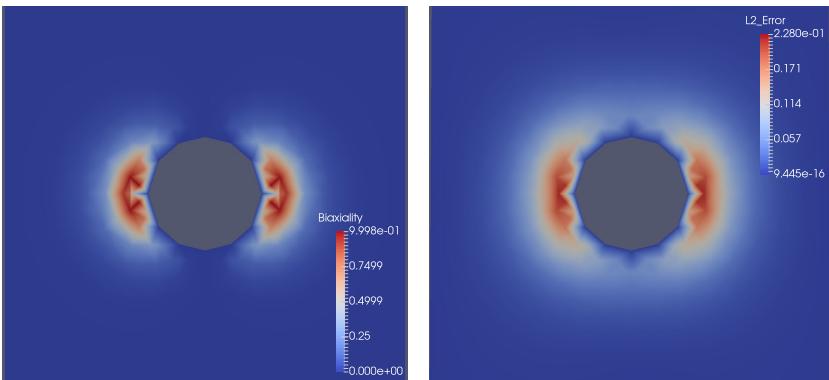
**TABLE 1** Comparison of the standard LdG model with the uniaxially constrained model: model difference error. The relative error between  $E_{\text{uni}}[\mathbf{Q}_{\text{uni}}]$  and  $E_{\text{LdG,one}}[\mathbf{Q}_{\text{uni}}]$  is shown.

$\eta_B$	$E_{\text{uni}}[\cdot]$ (initial)	$E_{\text{uni}}[\mathbf{Q}_{\text{uni}}]$ (final)	$E_{\text{LdG,one}}[\mathbf{Q}_{\text{uni}}]$ (initial)	rel. error
0.25	7.5990605	2.6644532	2.6164206	0.018358114
0.16	7.5990605	2.8031773	2.7497279	0.019438097
0.09	7.5990605	3.0018994	2.9413466	0.020586758
0.04	7.5990605	3.2711983	3.2176374	0.016646014
0.0225	7.5990605	3.5063179	3.5156034	-0.002641225

The uniaxial solution  $\mathbf{Q}_{\text{uni}}$ , for  $\eta_B = 0.25$ , is depicted in Fig. 11, in Section 5.1. In Fig. 15, we show a direct numerical comparison of the minimizer of the standard LdG model  $\mathbf{Q}_{\text{LdG}}$  with  $\mathbf{Q}_{\text{uni}}$ . On the left, we plot the biaxiality parameter given in (147). Fig. 15 shows that  $\mathbf{Q}_{\text{LdG}}$  achieves maximum biaxiality near the defect.

On the right of Fig. 15, we plot the pointwise quantity  $|\mathbf{Q}_{\text{LdG}}(\mathbf{x}) - \mathbf{Q}_{\text{uni}}(\mathbf{x})|$  with a maximum value approximately 0.228 (note that  $\mathbf{Q}_{\text{LdG}}$  and  $\mathbf{Q}_{\text{uni}}$  are  $O(1)$  tensors). Fig. 15 clearly shows that the two solutions are quite different near the defect.

Moreover, the energy of the uniaxial minimizer is significantly higher than the LdG minimizer, as shown in Table 2. The fact that it is higher is not surprising –the uniaxial model is more constrained– but it is significantly higher, which suggests that the two models could behave quite differently when other physical effects (e.g. electric/magnetic fields) are present.



**FIGURE 15** Comparison of the standard LdG model with *uniaxially* constrained version, using the Saturn-ring defect as example (Gu and Abbott, 2000) (Section 7). Left: the “standard” one-constant LdG model exhibits maximum “biaxiality” near the defect. Right: the pointwise  $L^2$  error between the minimizer from the standard model and the minimizer with uniaxial constraint.

**TABLE 2** Comparison of the standard LdG model with the uniaxially constrained model: final energy error. The relative error between  $E_{\text{uni}}[\mathbf{Q}_{\text{uni}}]$  and  $E_{\text{LdG,one}}[\mathbf{Q}_{\text{LdG}}]$  is shown.

$\eta_B$	$E_{\text{uni}}[\mathbf{Q}_{\text{uni}}]$ (final)	$E_{\text{LdG,one}}[\mathbf{Q}_{\text{LdG}}]$ (final)	rel. error
0.25	2.6644532	2.1808923	0.22172614
0.16	2.8031773	2.2931235	0.22242754
0.09	3.0018994	2.4689709	0.21585046
0.04	3.2711983	2.7850147	0.17457129
0.0225	3.5063179	3.0643099	0.14424392

## 8 Conclusion

We discussed the modeling of nematic LCs and their numerical simulation. We compared three models (namely, Oseen-Frank, Ericksen and Landau-de Gennes) for the equilibrium state of LCs. Because most thermotropic LCs do not exhibit any biaxiality, we focus on uniaxial LCs and compare Ericksen’s model with a uniaxially-constrained Landau-de Gennes model. For these, we present robust finite element schemes, which  $\Gamma$ -converge to the continuous problem as the mesh size tends to zero. For the solution of the resulting non-linear equations, we design gradient flow-type algorithms that are proven to be energy-decreasing.

We presented a variety of numerical experiments, illustrating the discretizations’ ability to capture non-trivial orientable and (for the Landau-de Gennes model) non-orientable defects. Moreover, we incorporated additional energy terms to model colloidal effects and the effect of external fields, such as electric fields. Finally, we gave a detailed numerical study of the effect of imposing the

uniaxial constraint (exactly) in the classic Landau-de Gennes model, which is a major highlight of this work.

## References

- Acharya, B.R., Primak, A., Kumar, S., 2004. Biaxial nematic phase in bent-core thermotropic mesogens. *Physical Review Letters* 92, 145506.
- Adler, J.H., Atherton, T.J., Benson, T.R., Emerson, D.B., MacLachlan, S.P., 2015a. Energy minimization for liquid crystal equilibrium with electric and flexoelectric effects. *SIAM Journal on Scientific Computing* 37 (5), S157–S176.
- Adler, J.H., Atherton, T.J., Emerson, D.B., MacLachlan, S.P., 2015b. An energy-minimization finite-element approach for the Frank–Oseen model of nematic liquid crystals. *SIAM Journal on Numerical Analysis* 53 (5), 2226–2254.
- Adler, J.H., Emerson, D.B., MacLachlan, S.P., Manteuffel, T.A., 2016. Constrained optimization for liquid crystal equilibria. *SIAM Journal on Scientific Computing* 38 (1), B50–B76.
- Alama, S., Bronsard, L., Lamy, X., 2016. Analytical description of the Saturn-ring defect in nematic colloids. *Physical Review E* 93, 012705.
- Alouges, F., 1997. A new algorithm for computing liquid crystal stable configurations: the harmonic mapping case. *SIAM Journal on Numerical Analysis* 34 (5), 1708–1726.
- Ambrosio, L., 1990a. Existence of minimal energy configurations of nematic liquid crystals with variable degree of orientation. *Manuscripta Mathematica* 68 (1), 215–228.
- Ambrosio, L., 1990b. Regularity of solutions of a degenerate elliptic variational problem. *Manuscripta Mathematica* 68 (1), 309–326.
- Araki, T., Tanaka, H., 2006. Colloidal aggregation in a nematic liquid crystal: topological arrest of particles by a single-stroke disclination line. *Physical Review Letters* 97, 127801.
- Bajc, I., Hecht, F., Žumer, S., 2016. A mesh adaptivity scheme on the Landau-de Gennes functional minimization case in 3d, and its driving efficiency. *Journal of Computational Physics* 321, 981–996.
- Ball, J.M., 2017. Mathematics and liquid crystals. *Molecular Crystals and Liquid Crystals* 647 (1), 1–27.
- Ball, J.M., Feireisl, E., Otto, F., 2015. *Mathematical Thermodynamics of Complex Fluids*. Lecture Notes in Mathematics Book Series, vol. 2200. Springer.
- Ball, J.M., Majumdar, A., 2010. Nematic liquid crystals: from Maier-Saupe to a continuum theory. *Molecular Crystals and Liquid Crystals* 525 (1), 1–11.
- Ball, J.M., Zarnescu, A., 2007. Orientable and non-orientable director fields for liquid crystals. *Proceedings in Applied Mathematics and Mechanics (PAMM)* 7 (1), 1050701–1050704.
- Ball, J.M., Zarnescu, A., 2011. Orientability and energy minimization in liquid crystal models. *Archive for Rational Mechanics and Analysis* 202 (2), 493–535.
- Barbero, G., Durand, G., 1986. On the validity of the Rapini-Papoulier surface anchoring energy form in nematic liquid crystals. *Journal de Physique France* 47 (12), 2129–2134.
- Barrett, J.W., Feng, X., Prohl, A., 2006. Convergence of a fully discrete finite element method for a degenerate parabolic system modelling nematic liquid crystals with variable degree of orientation. *ESAIM: Mathematical Modelling and Numerical Analysis* 40, 175–199.
- Bartels, S., 2006. Stability and convergence of finite-element approximation schemes for harmonic maps. *SIAM Journal on Numerical Analysis* 43 (1), 220–238.
- Bartels, S., 2015. *Numerical Methods for Nonlinear Partial Differential Equations*, 1st edition. Springer Series in Computational Mathematics. Springer.
- Bartels, S., Raisch, A., 2014. Simulation of q-tensor fields with constant orientational order parameter in the theory of uniaxial nematic liquid crystals. In: Griebel, M. (Ed.), *Singular Phenomena and Scaling in Mathematical Models*. Springer International Publishing, pp. 383–412.
- Bedford, S., 2016. Function spaces for liquid crystals. *Archive for Rational Mechanics and Analysis* 219 (2), 937–984.

- Biscari, P., Cesana, P., 2007. Ordering effects in electric splay Freedericksz transitions. *Continuum Mechanics and Thermodynamics* 19 (5), 285–298.
- Biscarini, F., Chiccoli, C., Pasini, P., Semeria, F., Zannoni, C., 1995. Phase diagram and orientational order in a biaxial lattice model: a Monte Carlo study. *Physical Review Letters* 75, 1803–1806.
- Borthagaray, J.-P., Nochetto, R.H., Walker, S.W., 2020. A structure-preserving FEM for the uniaxially constrained  $\mathbf{Q}$ -tensor model of nematic liquid crystals. *Numerische Mathematik* 145 (4), 837–881.
- Braides, A., 2002. *Gamma-Convergence for Beginners*. Oxford Lecture Series in Mathematics and Its Applications, vol. 22. Oxford Scholarship.
- Braides, A., 2014. *Local Minimization, Variational Evolution and  $\Gamma$ -Convergence*. Lecture Notes in Mathematics, vol. 2094. Springer.
- Brandts, J.H., Korotov, S., Krížek, M., 2008. The discrete maximum principle for linear simplicial finite element approximations of a reaction-diffusion problem. *Linear Algebra and Its Applications* 429 (10), 2344–2357. Special Issue in honor of Richard S. Varga.
- Brezis, H., 2003. The interplay between analysis and topology in some nonlinear pde problems. *Bulletin of the American Mathematical Society* 40, 179–201.
- Brezis, H., Coron, J.-M., Lieb, E.H., 1986. Harmonic maps with defects. *Communications in Mathematical Physics* 107 (4), 649–705.
- Brinkman, W.F., Cladis, P.E., 1982. Defects in liquid crystals. *Physics Today* 35, 48–56.
- Burger, M., 2002. A framework for the construction of level set methods for shape optimization and reconstruction. *Interfaces and Free Boundaries* 5, 301–329.
- Calderer, M., Golovaty, D., Lin, F., Liu, C., 2002. Time evolution of nematic liquid crystals with variable degree of orientation. *SIAM Journal on Mathematical Analysis* 33 (5), 1033–1047.
- Changizrezaei, S., Denniston, C., 2017. Heterogeneous colloidal particles immersed in a liquid crystal. *Physical Review E* 95, 052703.
- Ciarlet, P., Raviart, P.-A., 1973. Maximum principle and uniform convergence for the finite element method. *Computer Methods in Applied Mechanics and Engineering* 2 (1), 17–31.
- Cohen, R., Lin, S.-Y., Luskin, M., 1989. Relaxation and gradient methods for molecular orientation in liquid crystals. *Computer Physics Communications* 53 (1–3), 455–465.
- Conradi, M., Ravnik, M., Bele, M., Zorko, M., Žumer, S., Mušević, I., 2009. Janus nematic colloids. *Soft Matter* 5, 3905–3912.
- Cruz, P.A., Tomé, M.F., Stewart, I.W., McKee, S., 2013. Numerical solution of the Ericksen-Leslie dynamic equations for two-dimensional nematic liquid crystal flows. *Journal of Computational Physics* 247, 109–136.
- Čopar, S., Ravnik, M., Žumer, S., 2014. Janus nematic colloids with designable valence. *Materials* 7 (6), 4272.
- Čopar, S., Tkalec, U., Mušević, I., Žumer, S., 2015. Knot theory realizations in nematic colloids. *Proceedings of the National Academy of Sciences* 112 (6), 1675–1680.
- Davis, T., Gartland, E., 1998. Finite element analysis of the Landau-de Gennes minimization problem for liquid crystals. *SIAM Journal on Numerical Analysis* 35 (1), 336–362.
- de Gennes, P.G., Prost, J., 1995. *The Physics of Liquid Crystals*, 2nd edition. International Series of Monographs on Physics, vol. 83. Oxford Science Publication, Oxford, UK.
- De Giorgi, E., 2006. In: Maso, G.D., Forti, M., Miranda, M., Spagnolo, S.A., Ambrosio, L. (Eds.), *Selected Papers*. In: *Springer Collected Works in Mathematics*.
- de Jeu, W.H. (Ed.), 2012. *Liquid Crystal Elastomers: Materials and Applications*. Advances in Polymer Science. Springer.
- Dierking, I., Marshall, O., Wright, J., Bulleid, N., 2005. Annihilation dynamics of umbilical defects in nematic liquid crystals under applied electric fields. *Physical Review E* 71, 061709.
- Doğan, G., Morin, P., Nochetto, R.H., Verani, M., 2007. Discrete gradient flows for shape optimization and applications. *Computer Methods in Applied Mechanics and Engineering* 196, 3898–3914.
- Doostmohammadi, A., Ignés-Mullol, J., Yeomans, J.M., Sagués, F., 2018. Active nematics. *Nature Communications* 9, 3246.

- Ericksen, J., 1991. Liquid crystals with variable degree of orientation. *Archive for Rational Mechanics and Analysis* 113 (2), 97–120.
- Evans, L.C., 1998. *Partial Differential Equations*. American Mathematical Society, Providence, Rhode Island.
- Feynman, R.P., Leighton, R.B., Sands, M., 1964. *The Feynman Lectures on Physics*. Addison-Wesley Publishing Company.
- Fournier, J.-B., Galatola, P., 2005. Modeling planar degenerate wetting and anchoring in nematic liquid crystals. *Europhysics Letters (EPL)* 72 (3), 403–409.
- Freiser, M., 1970. Ordered states of a nematic liquid. *Physical Review Letters* 24 (19), 1041.
- Fukuda, J.-i., Stark, H., Yoneya, M., Yokoyama, H., 2004. Dynamics of a nematic liquid crystal around a spherical particle. *Journal of Physics: Condensed Matter* 16 (19), S1957–S1968.
- Gartland Jr, E.C., Palfy-Muhoray, P., Varga, R.S., 1991. Numerical minimization of the Landau-de Gennes free energy: defects in cylindrical capillaries. *Molecular Crystals and Liquid Crystals* 199 (1), 429–452.
- Gartland, J.E.C., Ramage, A., 2015. A renormalized Newton method for liquid crystal director modeling. *SIAM Journal on Numerical Analysis* 53 (1), 251–278.
- Giomi, L., 2015. Geometry and topology of turbulence in active nematics. *Physical Review X* 5, 031003.
- González, F.M.G., Gutiérrez-Santacreu, J.V., 2013. A linear mixed finite element scheme for a nematic Ericksen-Leslie liquid crystal model. *ESAIM: Mathematical Modelling and Numerical Analysis* 47, 1433–1464.
- Goodby, J.W., 2012. Introduction to defect textures in liquid crystals. In: Chen, J., Cranton, W., Fihn, M. (Eds.), *Handbook of Visual Display Technology*. Springer, pp. 1290–1314.
- Gramsbergen, E.F., Longa, L., de Jeu, W.H., 1986. Landau theory of the nematic-isotropic phase transition. *Physics Reports* 135 (4), 195–257.
- Gu, Y., Abbott, N.L., 2000. Observation of Saturn-ring defects around solid microspheres in nematic liquid crystals. *Physical Review Letters* 85, 4719–4722.
- Holzapfel, G.A., 2000. *Nonlinear Solid Mechanics: A Continuum Approach for Engineering*. John Wiley & Sons, Inc..
- Hoogboom, J., Elemans, J.A., Rowan, A.E., Rasing, T.H., Nolte, R.J., 2007. The development of self-assembled liquid crystal display alignment layers. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 365 (1855), 1553–1576.
- Hu, Q., Tai, X.-C., Winther, R., 2009. A saddle point approach to the computation of harmonic maps. *SIAM Journal on Numerical Analysis* 47 (2), 1500–1523.
- Humar, M., Muševič, I., 2010. 3d microlasers from self-assembled cholesteric liquid-crystal micro-droplets. *Optics Express* 18 (26), 26995–27003.
- Hyon, Y., Kwak, D.Y., Liu, C., 2010. Energetic variational approach in complex fluids: maximum dissipation principle. *Discrete and Continuous Dynamical Systems - Series A* 26 (4), 1291–1304.
- Jost, J., 2017. *Riemannian Geometry and Geometric Analysis*, 7th edition. Springer.
- Kinderlehrer, D., Walkington, N., Ou, B., 1993. The Elementary Defects of the Oseen-Frank Energy for a Liquid Crystal. Research report. Carnegie Mellon University, Department of Mathematics, Center for Nonlinear Analysis.
- Korotov, S., Křížek, M., Neittaanmäki, P., 2001. Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle. *Mathematics of Computation* 70 (233), 107–119.
- Lagerwall, J.P., Scalnia, G., 2012. A new era for liquid crystal research: applications of liquid crystals in soft matter nano-, bio- and microtechnology. *Current Applied Physics* 12 (6), 1387–1412.
- Lamy, X., 2013. A new light on the breaking of uniaxial symmetry in nematics. *ArXiv e-prints*.
- Lee, G.-D., Anderson, J., Bos, P.J., 2002. Fast q-tensor method for modeling liquid crystal director configurations with defects. *Applied Physics Letters* 81 (21), 3951–3953.

- Lehmann, O., 1889. Über fliessende krystalle (On flowing crystals). *Zeitschrift für Physikalische Chemie* 4U, 462–472.
- Lin, F.H., 1991. On nematic liquid crystals with variable degree of orientation. *Communications on Pure and Applied Mathematics* 44 (4), 453–468.
- Lin, F.-H., Liu, C., 1995. Nonparabolic dissipative systems modeling the flow of liquid crystals. *Communications on Pure and Applied Mathematics* 48 (5), 501–537.
- Lin, F.-H., Liu, C., 2001. Static and dynamic theories of liquid crystals. *Journal of Partial Differential Equations* 14 (4), 289–330.
- Lin, S.-Y., Luskin, M., 1989. Relaxation methods for liquid crystal problems. *SIAM Journal on Numerical Analysis* 26 (6), 1310–1324.
- Liu, C., Walkington, N., 2000. Approximation of liquid crystal flows. *SIAM Journal on Numerical Analysis* 37 (3), 725–741.
- Luckhurst, G., Veracini, C. (Eds.), 1994. *The Molecular Dynamics of Liquid Crystals*. Nato Science Series C, vol. 431. Springer.
- Madsen, L.A., Dingemans, T.J., Nakata, M., Samulski, E.T., 2004. Thermotropic biaxial nematic liquid crystals. *Physical Review Letters* 92, 145505.
- Majumdar, A., 2010. Equilibrium order parameters of nematic liquid crystals in the Landau-de Gennes theory. *European Journal of Applied Mathematics* 21 (2), 181–203.
- Majumdar, A., Zarnescu, A., 2010. Landau-de Gennes theory of nematic liquid crystals: the Oseen-Frank limit and beyond. *Archive for Rational Mechanics and Analysis* 196 (1), 227–280.
- Marchetti, M.C., Joanny, J.F., Ramaswamy, S., Liverpool, T.B., Prost, J., Rao, M., Simha, R.A., 2013. Hydrodynamics of soft active matter. *Reviews of Modern Physics* 85, 1143–1189.
- Miles, C.J., Evans, A.A., Shelley, M.J., Spagnolie, S.E., 2019. Active matter invasion of a viscous fluid: unstable sheets and a no-flow theorem. *Physical Review Letters* 122, 098002.
- Moreno-Razo, J.A., Sambriski, E.J., Abbott, N.L., Hernández-Ortiz, J.P., de Pablo, J.J., 2012. Liquid-crystal-mediated self-assembly at nanodroplet interfaces. *Nature* 485 (7396), 86–89.
- Mottram, N.J., Newton, C.J.P., 2014. Introduction to Q-tensor theory. *ArXiv e-prints*.
- Mottram, N.J., Pinto, J.T., McKay, G., 2013. Decoupling of the Ericksen-Leslie equations. *Liquid Crystals* 40 (6), 787–798.
- Nguyen, L., Zarnescu, A., 2013. Refined approximation for minimizers of a Landau-de Gennes energy functional. *Calculus of Variations and Partial Differential Equations* 47 (1–2), 383–432.
- Nochetto, R.H., Walker, S.W., Zhang, W., 2015. Numerics for liquid crystals with variable degree of orientation. In: *Symposium NN - Mathematical and Computational Aspects of Materials Science*. In: *MRS Proceedings*, vol. 1753.
- Nochetto, R.H., Walker, S.W., Zhang, W., 2017. A finite element method for nematic liquid crystals with variable degree of orientation. *SIAM Journal on Numerical Analysis* 55 (3), 1357–1386.
- Nochetto, R.H., Walker, S.W., Zhang, W., 2018. The Ericksen model of liquid crystals with colloidal and electric effects. *Journal of Computational Physics* 352, 568–601.
- Ohzono, T., Katoh, K., Wang, C., Fukazawa, A., Yamaguchi, S., Fukuda, J.-i., 2017. Uncovering different states of topological defects in schlieren textures of a nematic liquid crystal. *Scientific Reports* 7 (1), 16814.
- Onsager, L., 1931a. Reciprocal relations in irreversible processes. I. *Physical Review* 37, 405–426.
- Onsager, L., 1931b. Reciprocal relations in irreversible processes. II. *Physical Review* 38, 2265–2279.
- Palffy-muhoray, P., Gartland, E.C., Kelly, J.R., 1994. A new configurational transition in inhomogeneous nematics. *Liquid Crystals* 16 (4), 713–718.
- Prasad, V., Kang, S.-W., Suresh, K., Joshi, L., Wang, Q., Kumar, S., 2005. Thermotropic uniaxial and biaxial nematic and smectic phases in bent-core mesogens. *Journal of the American Chemical Society* 127 (49), 17224–17227.
- Qian, T., Wang, X.-P., Sheng, P., 2006. A variational approach to moving contact line hydrodynamics. *Journal of Fluid Mechanics* 564, 333–360.
- Ramaswamy, S., 2010. The mechanics and statistics of active matter. *Annual Review of Condensed Matter Physics* 1 (1), 323–345.

- Ramaswamy, S., 2017. Active matter. *Journal of Statistical Mechanics: Theory and Experiment* 2017 (5), 054002.
- Ravnik, M., Žumer, S., 2009. Landau-de Gennes modelling of nematic liquid crystal colloids. *Liquid Crystals* 36 (10–11), 1201–1214.
- Reinitzer, F., 1888. Beiträge zur kenntnis des cholesterins. *Monatshefte für Chemie und Verwandte Teile Anderer Wissenschaften* 9 (1), 421–441.
- Reinitzer, F., 1889. Contributions to the knowledge of cholesterol. *Liquid Crystals* 5 (1), 7–18.
- Roques-Carmes, T., Hayes, R.A., Feenstra, B.J., Schlangen, L.J.M., 2004. Liquid behavior inside a reflective display pixel based on electrowetting. *Journal of Applied Physics* 95 (8), 4389–4396.
- Schoen, R., Uhlenbeck, K., 1982. A regularity theory for harmonic maps. *Journal of Differential Geometry* 17 (2), 307–335.
- Schoen, R., Yau, S.-T., 1994. *Lectures on Harmonic Maps*. International Press of Boston.
- Schwartz, M., Lenzini, G., Geng, Y., Rønne, P.B., Ryan, P.Y.A., Lagerwall, J.P.F., 2018. Cholesteric liquid crystal shells as enabling material for information-rich design and architecture. *Advanced Materials* 30 (30), 1707382.
- Shen, J., Yang, X., 2010a. Numerical approximations of Allen-Cahn and Cahn-Hilliard equations. *Discrete and Continuous Dynamical Systems* 28 (4), 1669–1691.
- Shen, J., Yang, X., 2010b. A phase-field model and its numerical approximation for two-phase incompressible flows with different densities and viscosities. *SIAM Journal of Scientific Computing* 32 (3), 1159–1179.
- Sluckin, T.J., Dunmur, D.A., Stegemeyer, H. (Eds.), 2004. *Crystals That Flow: Classic Papers From the History of Liquid Crystals*, 1st edition. *Liquid Crystals Book Series*. CRC.
- Sluckin, T.J., Poniewierski, A., 1985. Novel surface phase transition in nematic liquid crystals: wetting and the Kosterlitz-Thouless transition. *Physical Review Letters* 55, 2907–2910.
- Sonnet, A., Kilian, A., Hess, S., 1995. Alignment tensor versus director: description of defects in nematic liquid crystals. *Physical Review E* 52, 718–722.
- Sonnet, A.M., Virga, E., 2012. *Dissipative Ordered Fluids: Theories for Liquid Crystals*. Springer.
- Strang, G., Fix, G., 2008. *An Analysis of the Finite Element Method*, 2nd edition. Wellesley-Cambridge.
- Struwe, M., 2008. *Variational Methods: Applications to Nonlinear Partial Differential Equations and Hamiltonian Systems*, 4th edition. Springer.
- Teixeira-Souza, R.T., Chiccoli, C., Pasini, P., Evangelista, L.R., Zannoni, C., 2015. Nematic liquid crystals in planar and cylindrical hybrid cells: role of elastic anisotropy on the director deformations. *Physical Review E* 92, 012501.
- Temam, R.M., Miranville, A.M., 2005. *Mathematical Modeling in Continuum Mechanics*, 2nd edition. Cambridge University Press.
- Tojo, K., Furukawa, A., Araki, T., Onuki, A., 2009. Defect structures in nematic liquid crystals around charged particles. *The European Physical Journal E* 30 (1), 55–64.
- Truesdell, C.A., 1976. *A First Course in Rational Continuum Mechanics*. Pure and Applied Mathematics, a Series of Monographs and Textbooks. Academic Press.
- Vanzo, D., Ricci, M., Berardi, R., Zannoni, C., 2012. Shape, chirality and internal order of freely suspended nematic nanodroplets. *Soft Matter* 8, 11790–11800.
- Vicsek, T., Zafeiris, A., 2012. Collective motion. *Physics Reports* 517 (3), 71–140. Collective motion.
- Virga, E.G., 1994. *Variational Theories for Liquid Crystals*, vol. 8, 1st edition. Chapman and Hall, London.
- Walker, S.W., 2018. On the Correct Thermo-dynamic Potential for Electro-static Dielectric Energy. ArXiv e-prints.
- Walkington, N.J., 2011. Numerical approximation of nematic liquid crystal flows governed by the Ericksen-Leslie equations. *ESAIM: Mathematical Modelling and Numerical Analysis* 45, 523–540.
- Wang, X., Kim, Y.-K., Bokusoglu, E., Zhang, B., Miller, D.S., Abbott, N.L., 2016. Experimental insights into the nanostructure of the cores of topological defects in liquid crystals. *Physical Review Letters* 116, 147801.

- Whitmer, J.K., Wang, X., Mondiot, F., Miller, D.S., Abbott, N.L., de Pablo, J.J., 2013. Nematic-field-driven positioning of particles in liquid crystal droplets. *Physical Review Letters* 111, 227801.
- Wise, S.M., Wang, C., Lowengrub, J.S., 2009. An energy-stable and convergent finite-difference scheme for the phase field crystal equation. *SIAM Journal on Numerical Analysis* 47 (3), 2269–2288.
- Xu, J., Li, Y., Wu, S., Bousquet, A., 2019. On the stability and accuracy of partially and fully implicit schemes for phase field modeling. *Computer Methods in Applied Mechanics and Engineering* 345, 826–853.
- Yang, X., Forest, M.G., Li, H., Liu, C., Shen, J., Wang, Q., Chen, F., 2013. Modeling and simulations of drop pinch-off from liquid crystal filaments and the leaky liquid crystal faucet immersed in viscous fluids. *Journal of Computational Physics* 236, 1–14.
- Yu, L., Saupe, A., 1980. Observation of a biaxial nematic phase in potassium laurate-1-decanol-water mixtures. *Physical Review Letters* 45 (12), 1000.
- Zhao, J., Wang, Q., 2016. Semi-discrete energy-stable schemes for a tensor-based hydrodynamic model of nematic liquid crystal flows. *Journal of Scientific Computing* 68 (3), 1241–1266.
- Zhao, J., Yang, X., Shen, J., Wang, Q., 2016. A decoupled energy stable scheme for a hydrodynamic phase-field model of mixtures of nematic liquid crystals and viscous fluids. *Journal of Computational Physics* 305, 539–556.