ENERGY MINIMIZATION FOR SKYRMIONS ON PLANAR THIN FILMS

GIOVANNI DI FRATTA, MICHAEL INNERBERGER, DIRK PRAETORIUS, AND VALERIY SLASTIKOV

Abstract. We consider an energy functional that arises in micromagnetic and liquid crystal theory on thin films. In particular, our energy comprises a non-convex term that models anti-symmetric exchange as well as an anisotropy term. We devise an algorithm for energy minimization in the continuous case and show weak convergence of a subsequence towards a solution of the corresponding Euler–Lagrange equation. Furthermore, an algorithm for numerical energy minimization is presented. We show empirically that this numerical algorithm converges to the correct solutions for a benchmark problem without the need for user-supplied parameters, and present a rigorous convergence analysis for important special cases.

1. Introduction

1.1. Model problem. We consider a two-dimensional bounded domain $\Omega \subset \mathbb{R}^2$ with Lipschitz boundary $\partial\Omega$. Given $\mu, \kappa \in \mathbb{R}$, we define the energy

$$\mathcal{E}(\boldsymbol{u}) := \int_\Omega \frac{1}{2} |\nabla \boldsymbol{u}|^2 + \kappa\, \boldsymbol{u} \cdot \operatorname{curl} \boldsymbol{u} + \frac{\mu}{2} |\boldsymbol{u} \cdot \mathbf{e}_3|^2 \,\mathrm{d}x \quad \text{for all } u \in \boldsymbol{H}^1(\Omega) := [H^1(\Omega)]^3, \tag{1}$$

where $\mathbf{e}_i \in \mathbb{R}^3$ is the i-th canonical basis vector and the 2D (vector) curl-operator is defined by

$$\operatorname{curl} \boldsymbol{u} := \sum_{i=1}^{2} \mathbf{e}_i \times \partial_i \boldsymbol{u}. \tag{2}$$

We aim to solve the following minimization problem:

$$\text{Find} \quad \boldsymbol{u} \in \mathcal{M} := \big\{ \boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \,\big|\, |\boldsymbol{u}| = 1 \text{ a.e. in } \Omega \big\} \quad \text{such that} \quad \mathcal{E}(\boldsymbol{u}) = \min_{\boldsymbol{v} \in \mathcal{M}} \mathcal{E}(\boldsymbol{v}). \tag{3}$$

The task of solving such minimization problems arises in the theories of micromagnetics and nematic liquid crystals on thin (planar) films [DDPR22].

The seminal work [Alo97] introduced a constructive algorithm for computing minimizers to the three-dimensional Dirichlet energy with Dirichlet boundary conditions, which is a simplified variant of (1). The work [Bar05] provided and analyzed a finite element version of this algorithm, which indeed involves some subtle technicalities to carry over

Acknowledgment. G. Di F., M. I., and D. P. acknowledge support through the Austrian Science Fund (FWF) through the doctoral school *Dissipation and dispersion in nonlinear PDEs* (grant W1245) and the special research program *Taming complexity in partial differential systems* (grant SFB F65). M. I. was supported by HHMI Janelia. G. Di F. and V. S. would also like to thank the Max Planck Institute for Mathematics in the Sciences in Leipzig, the Erwin Schrödinger International Institute for Mathematics and Physics in Wien, and the Vienna Center for PDEs at TU Wien for support and hospitality. G. Di F. is a member of the *Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni* (GNAMPA), which is part of the *Istituto Nazionale di Alta Matematica* (INdAM). The work of G. Di F. was partially supported by the Italian Ministry of Education and Research through the PRIN2022 project *Variational Analysis of Complex Systems in Material Science, Physics and Biology* (No. 2022HKBF5C). V. S. acknowledges support by Leverhulme grant RPG-2018-438.

the argument from the continuous case. In particular, the algorithm poses a certain angle condition to the underlying triangulation; see Definition 9 below.

1.2. Contributions of the present work. First, we present a surrogate energy functional which has the same set of minimizers as (7) while consisting only of positive contributions, regardless of the involved parameters κ and μ. This makes the surrogate energy much more convenient for analysis and numerical algorithms. In particular, the surrogate energy is subsequently used to formulate an iterative algorithm in the spirit of [Alo97] for computing minimizers of (7). Our algorithm provides a constructive way of finding a minimizing sequence (up to a subsequence) with monotonically decreasing energy starting from an arbitrary initial function. In analogy to [Alo97], we show convergence of this algorithm to solutions of the corresponding Euler–Lagrange equations. We stress that our problem deals with a model that is considerably more general than that of [Alo97], as it incorporates terms of lower order than the Dirichlet energy; in particular, the antisymmetric exchange term $\boldsymbol{u} \cdot \operatorname{curl} \boldsymbol{u}$. This poses several challenges, which we overcome by means of the surrogate energy. We note that our approach can cover an even larger class of models than (1) by permitting different kinds of anisotropy terms; see Section 2.2. Furthermore, we present a numerical algorithm in the spirit of [Bar05], which provides a convenient way of computing approximations to minimizers of $\mathcal{E}$. We give implementation details overcoming difficulties that arise from considering natural boundary conditions and that substantially enhance the algorithm given in [Bar05].

1.3. Outline. In Section 2, we introduce a surrogate energy for (1) that is easier to work with on the analytical as well as on the numerical level. We proceed by stating our energy minimization algorithm in Section 3, as well as a convergence result. Finally, we present a practical finite element discretization in Section 4 together with some implementation details, numerical experiments, and a convergence analysis for important special cases.

1.4. Notation. Throughout, we use boldface letters to denote vectors and italic boldface letters to denote vector-valued functions, whereas for components of vectors and functions we do not (e.g., u_3 is the third component of $\mathbf{u} \in \mathbb{R}^3$ or some function $\boldsymbol{u} \colon \Omega \to \mathbb{R}^3$, respectively). This is to visually distinguish components and collections of vectors. Moreover, to ensure uniformity of notation even in the discrete setting of Theorem 10, sequences of vectors generated by algorithms are denoted by upper indices, e.g., $(\mathbf{u}^n)_{n\in\mathbb{N}}$. Finally, the cross product $\mathbf{A} \times \mathbf{u}$ of a matrix $\mathbf{A} \in \mathbb{R}^{3\times 3}$ and a vector $\mathbf{u} \in \mathbb{R}^3$ is taken column-wise.

2. Surrogate energy $\mathcal{J}$

2.1. Setup. To find minimizers of $\mathcal{E}$ from (1), we seek minimizers of a surrogate energy with nicer properties for both analysis and numerics. To this end, we introduce the two-dimensional *helical derivatives* (cf. [Mel14; DIP20])

$$\partial_i^{\mathfrak{h}} \boldsymbol{v} := \partial_i \boldsymbol{v} + \kappa\, \boldsymbol{v} \times \mathbf{e}_i, \qquad \nabla_{\mathfrak{h}} \boldsymbol{v} := (\partial_1^{\mathfrak{h}} \boldsymbol{v}, \partial_2^{\mathfrak{h}} \boldsymbol{v}), \qquad \Delta_{\mathfrak{h}} \boldsymbol{v} := \sum_{i=1}^{2} \partial_i^{\mathfrak{h}} \partial_i^{\mathfrak{h}} \boldsymbol{v}, \tag{4}$$

and define, with $\gamma \in \mathbb{R}$, the surrogate energy

$$\mathcal{J}(\boldsymbol{v}) := \frac{1}{2} \int_\Omega \left[|\nabla_{\mathfrak{h}} \boldsymbol{v}|^2 + g_\gamma(\boldsymbol{v}, \boldsymbol{v}) \right] \mathrm{d}x \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}^1(\Omega), \tag{5}$$

where the bilinear form $g_\gamma\colon \mathbb{R}^3\times\mathbb{R}^3 \to \mathbb{R}$ reads

$$g_\gamma(\mathbf{w},\mathbf{v}) := \begin{cases} \gamma\,(\mathbf{w}\cdot\mathbf{e}_3)(\mathbf{v}\cdot\mathbf{e}_3) & \gamma \geqslant 0, \\ |\gamma|\,(\mathbf{w}\times\mathbf{e}_3)\cdot(\mathbf{v}\times\mathbf{e}_3) & \gamma < 0. \end{cases} \tag{6}$$

Note that all terms in $\mathcal{J}$ are positive (which is required by our analysis below) as opposed to the terms in $\mathcal{E}$ from (1). With $\mathcal{M}$ from (3), the surrogate minimization problem then reads as follows:

$$\text{Find} \quad \boldsymbol{u} \in \mathcal{M} \quad \text{such that} \quad \mathcal{J}(\boldsymbol{u}) = \min_{\boldsymbol{v}\in\mathcal{M}} \mathcal{J}(\boldsymbol{v}). \tag{7}$$

2.2. Relationship of $\mathcal{J}$ and $\mathcal{E}$. It is straightforward to verify, using the direct method of the calculus of variations, that the minimization problems for both $\mathcal{E}$ and $\mathcal{J}$ over $\mathcal{M}$ admit solutions. Moreover, as we now demonstrate, the two problems are variationally equivalent: for suitable choices of the parameters, minimizers of $\mathcal{J}$ coincide with that of $\mathcal{E}$.

Proposition 1. *For $\gamma = \mu - \kappa^2$, it holds that*

$$\left\{\boldsymbol{u}\in\mathcal{M} \;\middle|\; \mathcal{E}(\boldsymbol{u}) = \inf_{\boldsymbol{v}\in\mathcal{M}} \mathcal{E}(\boldsymbol{v})\right\} = \left\{\boldsymbol{u}\in\mathcal{M} \;\middle|\; \mathcal{J}(\boldsymbol{u}) = \inf_{\boldsymbol{v}\in\mathcal{M}} \mathcal{J}(\boldsymbol{v})\right\}.$$

Proof. We begin by examining the relation between the terms $|\boldsymbol{v}\cdot\mathbf{e}_3|^2$ and $|\boldsymbol{v}\times\mathbf{e}_3|^2$. For any $\boldsymbol{v}\in\boldsymbol{H}^1(\Omega)$, it holds that

$$|\boldsymbol{v}\times\mathbf{e}_j|^2 = \sum_{i=1,\,i\neq j}^{3} |\boldsymbol{v}\cdot\mathbf{e}_i|^2, \tag{8}$$

which can be seen by explicit computation. From the definition (6) of g_γ, we see that

$$\begin{aligned} &\gamma \geqslant 0: \quad g_\gamma(\boldsymbol{v},\boldsymbol{v}) = \gamma\,|\boldsymbol{v}\cdot\mathbf{e}_3|^2, \\ &\gamma < 0: \quad g_\gamma(\boldsymbol{v},\boldsymbol{v}) = -\gamma\,|\boldsymbol{v}\times\mathbf{e}_3|^2 \overset{(8)}{=} \gamma\,|\boldsymbol{v}\cdot\mathbf{e}_3|^2 - \gamma\,|\boldsymbol{v}|^2. \end{aligned} \tag{9}$$

For all $\gamma\in\mathbb{R}$, this yields that $g_\gamma(\boldsymbol{v},\boldsymbol{v}) = \gamma\,|\boldsymbol{v}\cdot\mathbf{e}_3|^2 + C\,|\boldsymbol{v}|^2$ with $C = |\min\{0,\gamma\}| \geq 0$. Finally, we follow [DIP20] and expand the helical term in $\mathcal{J}$ to see that

$$\begin{aligned} |\nabla_{\mathfrak{h}}\boldsymbol{v}|^2 &\overset{(4)}{=} |\nabla\boldsymbol{v}|^2 + 2\kappa\,\boldsymbol{v}\cdot\operatorname{curl}\boldsymbol{v} + \kappa^2\sum_{i=1}^{2}|\boldsymbol{v}\times\mathbf{e}_i|^2 \\ &\overset{(8)}{=} |\nabla\boldsymbol{v}|^2 + 2\kappa\,\boldsymbol{v}\cdot\operatorname{curl}\boldsymbol{v} + \kappa^2(|\boldsymbol{v}|^2 + |\boldsymbol{v}\cdot\mathbf{e}_3|^2). \end{aligned} \tag{10}$$

Combining the identities above and recalling that $\mu = \kappa^2 + \gamma$, we arrive at

$$\begin{aligned} \mathcal{J}(\boldsymbol{v}) &= \frac{1}{2}\int_\Omega \left[|\nabla_{\mathfrak{h}}\boldsymbol{v}|^2 + g_\gamma(\boldsymbol{v},\boldsymbol{v})\right]\mathrm{d}x \\ &\overset{(10)}{=} \int_\Omega \left[\frac{1}{2}|\nabla\boldsymbol{v}|^2 + \kappa\,\boldsymbol{v}\cdot\operatorname{curl}\boldsymbol{v} + \frac{\kappa^2+\gamma}{2}|\boldsymbol{v}\cdot\mathbf{e}_3|^2 + \frac{\kappa^2+C}{2}|\boldsymbol{v}|^2\right]\mathrm{d}x \\ &= \mathcal{E}(\boldsymbol{v}) + \frac{\kappa^2+C}{2}\|\boldsymbol{v}\|^2_{\boldsymbol{L}^2(\Omega)}. \end{aligned}$$

For $\boldsymbol{v}\in\mathcal{M}$, we have that $|\boldsymbol{v}| = 1$ a.e. in Ω and hence $\|\boldsymbol{v}\|^2_{\boldsymbol{L}^2(\Omega)} = |\Omega|$. Thus, $\mathcal{J}(\boldsymbol{v}) = \mathcal{E}(\boldsymbol{v}) + \text{const}$ for all $\boldsymbol{v}\in\mathcal{M}$ so that the sets of minimizers of $\mathcal{J}$ and $\mathcal{E}$ coincide. □

Remark 2. *It is also of practical interest to consider anisotropy that favors out-of-plane configurations instead of in-plane ones, which is modeled by the modified energy functional*

$$\widetilde{\mathcal{E}}(\boldsymbol{u}) := \int_\Omega \Big[\frac{1}{2}\,|\nabla \boldsymbol{u}|^2 + \kappa\,\boldsymbol{u}\cdot\operatorname{curl}\boldsymbol{u} + \frac{\mu}{2}\,|\boldsymbol{u}\times\mathbf{e}_3|^2\Big]\,\mathrm{d}x.$$

If $\gamma = -(\mu+\kappa^2)$, in analogy to the above computations, we see that

$$\big\{\boldsymbol{u}\in\mathcal{M}\;\big|\;\widetilde{\mathcal{E}}(\boldsymbol{u}) = \inf_{\boldsymbol{v}\in\mathcal{M}}\widetilde{\mathcal{E}}(\boldsymbol{v})\big\} = \big\{\boldsymbol{u}\in\mathcal{M}\;\big|\;\mathcal{J}(\boldsymbol{u}) = \inf_{\boldsymbol{v}\in\mathcal{M}}\mathcal{J}(\boldsymbol{v})\big\}.$$

Note that, thanks to (8) and the unit-length constraint, any mixture of out-of-plane and in-plane anisotropies is encompassed by our g_γ term for an appropriate choice of γ. Such combinations arise, for instance, in the thin-film limit of three-dimensional models with a stray-field contribution; see [DMRS20; Di 20; DDPR22].

2.3. Euler–Lagrange equations of $\mathcal{J}$. For fixed $\boldsymbol{v}\in\boldsymbol{H}^1(\Omega)$, we denote by $\boldsymbol{g}_\gamma(\boldsymbol{v})\in\boldsymbol{L}^2(\Omega)$ the Riesz representative of the linear form induced by $g_\gamma(\boldsymbol{v},\cdot)$ from (6), i.e.,

$$\int_\Omega \boldsymbol{g}_\gamma(\boldsymbol{v})\cdot\boldsymbol{w}\,\mathrm{d}x = \int_\Omega g_\gamma(\boldsymbol{v},\boldsymbol{w})\,\mathrm{d}x \quad\text{with}\quad \boldsymbol{g}_\gamma(\boldsymbol{v}) := \begin{cases} \gamma\, v_3\mathbf{e}_3 & \gamma\geqslant 0,\\ |\gamma|\,(v_1\mathbf{e}_1+v_2\mathbf{e}_2) & \gamma<0.\end{cases}$$

Lemma 3. *In strong form, the Euler–Lagrange equations associated with the minimization problem for $\mathcal{J}$ in (7) are*

$$-\Delta_{\mathfrak{h}}\boldsymbol{u} + \boldsymbol{g}_\gamma(\boldsymbol{u}) = \big(|\nabla_{\mathfrak{h}}\boldsymbol{u}|^2 + g_\gamma(\boldsymbol{u},\boldsymbol{u})\big)\boldsymbol{u} \quad \text{in } \Omega, \tag{11}$$

subject to the pointwise constraint $|\boldsymbol{u}|=1$ in Ω and the natural boundary condition

$$\partial_{\boldsymbol{n}}\boldsymbol{u} = \kappa\boldsymbol{n}\times\boldsymbol{u}, \quad \text{on } \partial\Omega, \tag{12}$$

where $\boldsymbol{n}$ denotes the unit outward normal on $\partial\Omega$.

Moreover, a function $\boldsymbol{u}\in\mathcal{M}$ is a weak solution of (11) if and only if it satisfies the variational identity

$$\int_\Omega \big[(\boldsymbol{u}\times\nabla_{\mathfrak{h}}\boldsymbol{u}) : \nabla_{\mathfrak{h}}\boldsymbol{v} + g_\gamma(\boldsymbol{u},\boldsymbol{v}\times\boldsymbol{u})\big]\,\mathrm{d}x = 0 \quad \text{for all } \boldsymbol{v}\in\boldsymbol{H}^1(\Omega)\cap\boldsymbol{L}^\infty(\Omega). \tag{13}$$

Sketch of proof. The argument parallels the harmonic-map case. First, consider an appropriate variation of the energy to derive the strong Euler–Lagrange equations (11). Next, insert the vector field $\boldsymbol{v}\times\boldsymbol{u}$ as a test function in the weak formulation and employ standard vector identities to show that any weak solution of the Euler–Lagrange equations satisfies (13). The converse implication is obtained by the fundamental lemma of the calculus of variations. □

3. Energy minimization

3.1. Computation of minimizers. Finding (numerical) solutions of (7) is challenging for two main reasons. First, the pointwise constraint $|\boldsymbol{u}(x)| = 1$ a.e. in Ω is non-convex, so standard minimization algorithms designed for convex constraints do not apply. Second, minimizers may fail to be unique (for instance, when $\kappa = 0$ and $\gamma \geq 0$ the energy is invariant under orthogonal transformations that fix the $\mathbf{e}_3$-axis). For these reasons, a direct attack on the Euler–Lagrange equations is generally not viable.

To handle the unit-length constraint, we use the nearest-point projection onto the unit sphere $\mathbb{S}^2$, i.e.,

$$\Pi\colon \mathcal{M}^+ \to \mathcal{M},\ \Pi\boldsymbol{v} \coloneqq \frac{\boldsymbol{v}}{|\boldsymbol{v}|}, \quad \text{where } \mathcal{M}^+ \coloneqq \big\{\boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \;\big|\; |\boldsymbol{v}| \geqslant 1 \text{ a.e. in } \Omega\big\}. \tag{14}$$

For variational and numerical purposes, it is convenient to denote by $\mathcal{K}[\boldsymbol{u}]$ the tangent space to a given function $\boldsymbol{u} \in \mathcal{M}$:

$$\mathcal{K}[\boldsymbol{u}] := \{\boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \mid \boldsymbol{v} \cdot \boldsymbol{u} = 0 \text{ a.e. in } \Omega\}. \tag{15}$$

While $\mathcal{M}$ is non-convex (though weakly closed) in $\boldsymbol{H}^1(\Omega)$, $\mathcal{K}[\boldsymbol{u}]$ is a closed linear subspace of $\boldsymbol{H}^1(\Omega)$ and hence a Hilbert space for each $\boldsymbol{u} \in \mathcal{M}$; this makes $\mathcal{K}[\boldsymbol{u}]$ a natural space for linearized computations around $\boldsymbol{u}$.

Finally, we introduce the symmetric bilinear form

$$a(\boldsymbol{w}, \boldsymbol{v}) := \int_\Omega \big[\nabla_\mathfrak{h}\boldsymbol{w} : \nabla_\mathfrak{h}\boldsymbol{v} + g_\gamma(\boldsymbol{w}, \boldsymbol{v})\big]\,\mathrm{d}x \quad \text{for all } \boldsymbol{w}, \boldsymbol{v} \in \boldsymbol{H}^1(\Omega). \tag{16}$$

With these ingredients (the projection Π, the tangent spaces $\mathcal{K}[\boldsymbol{u}]$, and the bilinear form $a(\cdot,\cdot)$) we adapt and generalize the numerical scheme from [Alo97] to compute minimizers of (5).

Algorithm A (Continuous energy minimization). ***Input:*** *Initial guess* $\boldsymbol{u}^0 \in \mathcal{M}$.
Loop: *For all* $n = 0, 1, 2, \ldots$, *do*

(i) *Find* $\boldsymbol{w}^n \in \mathcal{K}[\boldsymbol{u}^n]$ *such that*

$$a(\boldsymbol{w}^n, \boldsymbol{v}) = a(\boldsymbol{u}^n, \boldsymbol{v}) \qquad \text{for all } \boldsymbol{v} \in \mathcal{K}[\boldsymbol{u}^n]. \tag{17}$$

(ii) *Set* $\boldsymbol{u}^{n+1} := \Pi(\boldsymbol{u}^n - \boldsymbol{w}^n)$.

Output: *Sequence of functions* $(\boldsymbol{u}^n)_{n\in\mathbb{N}} \subset \mathcal{M}$.

Our main result states well-posedness and convergence of Algorithm A to a critical point of $\mathcal{J}$. The proof is postponed to Subsection 3.2 below.

Theorem 4. *Algorithm A is well-defined, i.e., for any* $\boldsymbol{u}^n \in \mathcal{M}$, *the variational formulation* (17) *admits at least one solution* $\boldsymbol{w}^n \in \mathcal{K}[\boldsymbol{u}^n]$ *and* $\boldsymbol{u}^n - \boldsymbol{w}^n \in \mathcal{M}^+$. *For any sequence* $(\boldsymbol{u}^n)_{n\in\mathbb{N}}$ *generated by Algorithm A, there exists a subsequence (not relabeled) and a function* $\boldsymbol{u} \in \mathcal{M}$ *such that, along this subsequence,*

$$\boldsymbol{u}^n \rightharpoonup \boldsymbol{u} \quad \text{weakly in } \boldsymbol{H}^1(\Omega) \text{ as } n \to \infty \tag{18}$$

and $\boldsymbol{u}$ *is a solution of the Euler–Lagrange equations* (11) *of* $\mathcal{J}$.

While the existence of a solution $\boldsymbol{w}^n \in \mathcal{K}[\boldsymbol{u}^n]$ to (17) is ensured by Theorem 4, uniqueness might fail in general. Clearly, uniqueness of the solution to (17) is guaranteed when Dirichlet boundary conditions are imposed on $\partial\Omega$ (or at least on a nontrivial portion of $\partial\Omega$), as considered in [Alo97; Bar05].

As our analysis is concerned with natural boundary conditions (cf. (12)), the uniqueness of step (i) in Algorithm A cannot be taken for granted. In fact, for certain parameter regimes, additional constraints on $\boldsymbol{w}^n$ are required to secure uniqueness. The precise circumstances under which non-uniqueness may occur are characterized in point (a) of the proposition below.

Proposition 5. *Suppose that* Ω *is connected. Let* $\boldsymbol{u} \in \mathcal{M}$ *and* $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathcal{K}[\boldsymbol{u}]$ *be two solutions of the Euler–Lagrange equations* (17). *Then, there hold the following assertions:*

(a) *For* $\kappa = 0$, *the difference* $\boldsymbol{w}_1 - \boldsymbol{w}_2 \in \mathcal{K}[\boldsymbol{u}]$ *is constant and satisfies*

- $\boldsymbol{w}_1 - \boldsymbol{w}_2 \perp \mathbf{e}_3$ *a.e. in* Ω *if* $\gamma > 0$,
- $\boldsymbol{w}_1 - \boldsymbol{w}_2 \in \operatorname{span}\{\mathbf{e}_3\}$ *a.e. in* Ω *if* $\gamma < 0$.

(b) *For* $\kappa \neq 0$*, i.e., in the presence of the anti-symmetric exchange term, the difference* $\boldsymbol{w}_1 - \boldsymbol{w}_2$ *vanishes. In particular, the solution of* (17) *is unique.*

Proof. Since $\mathcal{K}[\boldsymbol{u}]$ is a linear space, there holds $\boldsymbol{w}_1 - \boldsymbol{w}_2 \in \mathcal{K}[\boldsymbol{u}]$. By assumption,

$$a(\boldsymbol{w}_1 - \boldsymbol{w}_2, \boldsymbol{v}) \overset{(17)}{=} 0 \qquad \text{for all } \boldsymbol{v} \in \mathcal{K}[\boldsymbol{u}].$$

For $\boldsymbol{v} := \boldsymbol{w}_1 - \boldsymbol{w}_2$, we obtain

$$0 = a(\boldsymbol{v}, \boldsymbol{v}) = \int_\Omega \big[|\nabla_\mathfrak{h} \boldsymbol{v}|^2 + g_\gamma(\boldsymbol{v}, \boldsymbol{v})\big] \,\mathrm{d}x. \tag{19}$$

Non-negativity of $g_\gamma(\boldsymbol{v}, \boldsymbol{v})$ implies $\nabla_\mathfrak{h} \boldsymbol{v} = 0 = g_\gamma(\boldsymbol{v}, \boldsymbol{v})$ a.e. in Ω. In particular, we have

$$\partial_i \boldsymbol{v} + \kappa\, \boldsymbol{v} \times \mathbf{e}_i = 0 \qquad \text{for } i = 1, 2. \tag{20}$$

Taking the dot product of (20) with $\mathbf{e}_i$ for $i = 1, 2, 3$, we obtain that

$$\begin{aligned} &\text{(a)}\ \ \partial_1 v_1 = 0, \quad \text{(c)}\ \ \partial_1 v_2 + \kappa\, \boldsymbol{v} \cdot \mathbf{e}_3 = 0, \quad \text{(e)}\ \ \partial_1 v_3 - \kappa\, \boldsymbol{v} \cdot \mathbf{e}_2 = 0, \\ &\text{(b)}\ \ \partial_2 v_1 - \kappa\, \boldsymbol{v} \cdot \mathbf{e}_3 = 0, \quad \text{(d)}\ \ \partial_2 v_2 = 0, \quad \text{(f)}\ \ \partial_2 v_3 + \kappa\, \boldsymbol{v} \cdot \mathbf{e}_1 = 0. \end{aligned} \tag{21}$$

Next, we now consider the assertions (a)–(b) separately and exploit $g_\gamma(\boldsymbol{v}, \boldsymbol{v}) = 0$. Recall that $g_\gamma(\boldsymbol{v}, \boldsymbol{v}) \simeq |\boldsymbol{v} \cdot \mathbf{e}_3|^2$ for $\gamma > 0$ and $g_\gamma(\boldsymbol{v}, \boldsymbol{v}) \simeq |\boldsymbol{v} \times \mathbf{e}_3|^2$ for $\gamma < 0$.

(a): For $\kappa = 0$, (21) simplifies to $\nabla \boldsymbol{v} = \mathbf{0}$; hence, $\boldsymbol{v}$ is constant. For $\gamma > 0$, we have that $v_3 = 0$ a.e. in Ω and hence $\boldsymbol{v} \perp \mathbf{e}_3$. For $\gamma < 0$, we have $v_1 = 0 = v_2$ a.e. in Ω and hence $\boldsymbol{v} \in \operatorname{span}\{\mathbf{e}_3\}$.

(b): For $\kappa \neq 0$, we examine the sign of γ. For $\gamma > 0$, there holds $v_3 = 0$. Column (21e–f) implies that also $v_1 = v_2 = 0$, and hence $\boldsymbol{v} = 0$. For $\gamma < 0$, we have $v_1 = 0 = v_2$. By (21a–d), there follows $v_3 = 0$ and hence $\boldsymbol{v} = 0$. For $\gamma = 0$, the relations $\partial_1 v_1 = \partial_2 v_2 = 0$ prove that $v_1(x_1, x_2) = v_1(x_2)$ and $v_2(x_1, x_2) = v_2(x_1)$, i.e., v_1 is independent of x_1 while v_2 is independent of x_2. Adding (21b–c), we see $\partial_1 v_2(x_1) + \partial_2 v_1(x_2) = 0$ for every $(x_1, x_2) \in \Omega$. This implies that $\partial_1 v_2(x_1)$ and $\partial_2 v_1(x_2)$ are both constant. Hence, (21b) yields that $\kappa\, v_3 = \partial_2 v_1(x_2)$ is constant. From $\kappa \neq 0$, we thus obtain that v_3 is constant in Ω. Column (21e–f) then yields that both $v_1 = 0 = v_2$. From (21b), we finally conclude that also v_3 is zero and hence $\boldsymbol{v} = 0$ also in this case. □

3.2. Proof of Theorem 4. First, we ensure that both steps of Algorithm A lead to a decrease in the energy $\mathcal{J}$. We start by showing that subtracting the tangential update obtained in step (i) decreases the energy (although this intermediate update lies in $\mathcal{M}^+$ from (14) rather than in $\mathcal{M}$ from (3)).

Lemma 6. *Let* $\boldsymbol{u} \in \boldsymbol{H}^1(\Omega)$ *and* $\boldsymbol{w} \in \mathcal{K}[\boldsymbol{u}]$ *such that* $a(\boldsymbol{w}, \boldsymbol{v}) = a(\boldsymbol{u}, \boldsymbol{v})$ *for all* $\boldsymbol{v} \in \mathcal{K}[\boldsymbol{u}]$*. Then, it holds that*

$$\mathcal{J}(\boldsymbol{u} - \boldsymbol{w}) = \mathcal{J}(\boldsymbol{u}) - \mathcal{J}(\boldsymbol{w}) \leq \mathcal{J}(\boldsymbol{u}). \tag{22}$$

Proof. We choose $\boldsymbol{v} = 2\boldsymbol{w} \in \mathcal{K}[\boldsymbol{u}]$ to obtain that

$$\begin{aligned} 0 = a(\boldsymbol{u} - \boldsymbol{w}, 2\boldsymbol{w}) &= a(\boldsymbol{u}, \boldsymbol{u}) - a(\boldsymbol{u}, \boldsymbol{u}) + 2a(\boldsymbol{u}, \boldsymbol{w}) - 2a(\boldsymbol{w}, \boldsymbol{w}) \\ &= a(\boldsymbol{u}, \boldsymbol{u}) - a(\boldsymbol{u} - \boldsymbol{w}, \boldsymbol{u} - \boldsymbol{w}) - a(\boldsymbol{w}, \boldsymbol{w}) \\ &= 2\mathcal{J}(\boldsymbol{u}) - 2\mathcal{J}(\boldsymbol{u} - \boldsymbol{w}) - 2\mathcal{J}(\boldsymbol{w}). \end{aligned}$$

By non-negativity of $\mathcal{J}$, it follows that

$$\mathcal{J}(\boldsymbol{u}-\boldsymbol{w}) = \mathcal{J}(\boldsymbol{u}) - \mathcal{J}(\boldsymbol{w}) \leq \mathcal{J}(\boldsymbol{u}).$$

This concludes the proof. □

Also, the projection step (ii) decreases the energy.

Lemma 7. *Let $\boldsymbol{u} \in \mathcal{M}^+$. Then, it holds that*

$$\mathcal{J}(\Pi(\boldsymbol{u})) \leqslant \mathcal{J}(\boldsymbol{u}). \tag{23}$$

Proof. It follows from the definition (14) of the projection $\Pi(\boldsymbol{u})$ that

$$|\Pi(\boldsymbol{u})\cdot\mathbf{e}_i|^2 = \frac{|\boldsymbol{u}\cdot\mathbf{e}_i|^2}{|\boldsymbol{u}|^2} \leqslant |\boldsymbol{u}\cdot\mathbf{e}_i|^2 \quad \text{for all } i=1,2,3,$$

where, here and in the following, all estimates hold a.e. in Ω. From (9), we then see that

$$g_\gamma(\Pi(\boldsymbol{u}),\Pi(\boldsymbol{u})) \leqslant g_\gamma(\boldsymbol{u},\boldsymbol{u}). \tag{24}$$

Thus, it remains to inspect the helical term in the energy density. Note that the derivative of the projection can be computed explicitly to be

$$\big(\nabla\Pi\big)(\boldsymbol{u}) = \frac{1}{|\boldsymbol{u}|}\Big(\mathbf{1} - \frac{\boldsymbol{u}\boldsymbol{u}^\intercal}{|\boldsymbol{u}|^2}\Big). \tag{25}$$

Inserting this into the helical energy density, the chain rule proves that

$$\begin{aligned}
|\nabla_{\mathfrak{h}}\Pi(\boldsymbol{u})|^2 &\overset{(4)}{=} \sum_{i=1}^2 \big|\big(\nabla\Pi\big)(\boldsymbol{u})\partial_i\boldsymbol{u} + \kappa\,\Pi(\boldsymbol{u})\times\mathbf{e}_i\big|^2 \\
&\overset{(25)}{=} \sum_{i=1}^2 \Big|\frac{1}{|\boldsymbol{u}|}\partial_i\boldsymbol{u} - \frac{1}{|\boldsymbol{u}|^3}\boldsymbol{u}\boldsymbol{u}^\intercal\partial_i\boldsymbol{u} + \frac{\kappa}{|\boldsymbol{u}|}\boldsymbol{u}\times\mathbf{e}_i\Big|^2 \\
&= \sum_{i=1}^2 \Big|\frac{1}{|\boldsymbol{u}|}(\partial_i\boldsymbol{u} + \kappa\,\boldsymbol{u}\times\mathbf{e}_i) - \frac{1}{|\boldsymbol{u}|^3}(\boldsymbol{u}\cdot\partial_i\boldsymbol{u})\boldsymbol{u}\Big|^2 \\
&= \sum_{i=1}^2 \frac{1}{|\boldsymbol{u}|^2}|\partial_i\boldsymbol{u} + \kappa\,\boldsymbol{u}\times\mathbf{e}_i|^2 + \frac{1}{|\boldsymbol{u}|^4}|\boldsymbol{u}\cdot\partial_i\boldsymbol{u}|^2 - \frac{2}{|\boldsymbol{u}|^4}(\partial_i\boldsymbol{u} + \kappa\,\boldsymbol{u}\times\mathbf{e}_i)\cdot(\boldsymbol{u}\cdot\partial_i\boldsymbol{u})\boldsymbol{u}.
\end{aligned}$$

Note that $\boldsymbol{u}\times\mathbf{e}_i \perp (\boldsymbol{u}\cdot\partial_i\boldsymbol{u})\boldsymbol{u}$ and $\partial_i\boldsymbol{u}\cdot(\boldsymbol{u}\cdot\partial_i\boldsymbol{u})\boldsymbol{u} = |\boldsymbol{u}\cdot\partial_i\boldsymbol{u}|^2$. With $|\boldsymbol{u}| \geqslant 1$ a.e. in Ω, we thus get

$$\begin{aligned}
|\nabla_{\mathfrak{h}}\Pi(\boldsymbol{u})|^2 &= \sum_{i=1}^2 \frac{1}{|\boldsymbol{u}|^2}|\partial_i\boldsymbol{u} + \kappa\,\boldsymbol{u}\times\mathbf{e}_i|^2 - \frac{1}{|\boldsymbol{u}|^4}|\boldsymbol{u}\cdot\partial_i\boldsymbol{u}|^2 \\
&\leqslant \sum_{i=1}^2 |\partial_i\boldsymbol{u} + \kappa\,\boldsymbol{u}\times\mathbf{e}_i|^2 = |\nabla_{\mathfrak{h}}\boldsymbol{u}|^2.
\end{aligned} \tag{26}$$

Integrating (24) and (26) over Ω, we conclude the proof. □

The next lemma provides quantitative two-sided estimates for the surrogate energy $\mathcal{J}$. From these estimates, we deduce that $\mathcal{J}$ controls the $\boldsymbol{H}^1$-seminorm and hence is coercive on $\mathcal{M}$.

Lemma 8. *For $\boldsymbol{u} \in \boldsymbol{H}^1(\Omega)$, we have that*

$$\|\nabla \boldsymbol{u}\|^2_{\boldsymbol{L}^2(\Omega)} \leqslant 4\mathcal{J}(\boldsymbol{u}) + 4\kappa^2 \|\boldsymbol{u}\|^2_{\boldsymbol{L}^2(\Omega)}, \tag{27}$$

$$\mathcal{J}(\boldsymbol{u}) \leqslant \|\nabla \boldsymbol{u}\|^2_{\boldsymbol{L}^2(\Omega)} + \Big(\kappa^2 + \frac{|\gamma|}{2}\Big) \|\boldsymbol{u}\|^2_{\boldsymbol{L}^2(\Omega)}. \tag{28}$$

Proof. We first show the estimate (27). For $\kappa = 0$, the estimate (27) is obvious by definition of $\mathcal{J}$ (and holds even with a factor 2 instead of 4). Let $\kappa \neq 0$. The Young inequality $2ab \leqslant a^2/2 + 2b^2$ shows that

$$\begin{aligned} |\nabla_{\mathfrak{h}} \boldsymbol{u}|^2 &\overset{(4)}{\geqslant} \sum_{i=1}^{2} \Big(|\partial_i \boldsymbol{u}|^2 + \kappa^2 |\boldsymbol{u} \times \mathbf{e}_i|^2 - 2|\kappa\, \partial_i \boldsymbol{u} \cdot (\boldsymbol{u} \times \mathbf{e}_i)| \Big) \\ &\geqslant \big(1 - 1/2\big) |\nabla \boldsymbol{u}|^2 + \big(\kappa^2 - 2\kappa^2\big) \sum_{i=1}^{2} |\boldsymbol{u} \times \mathbf{e}_i|^2. \end{aligned}$$

Hence, we get that

$$\frac{1}{2} |\nabla \boldsymbol{u}|^2 \leqslant |\nabla_{\mathfrak{h}} \boldsymbol{u}|^2 + \kappa^2 \sum_{i=1}^{2} |\boldsymbol{u} \times \mathbf{e}_i|^2 \leqslant |\nabla_{\mathfrak{h}} \boldsymbol{u}|^2 + 2\kappa^2 |\boldsymbol{u}|^2.$$

Adding the term $g_\gamma(\boldsymbol{u}, \boldsymbol{u}) \geqslant 0$ on the right-hand side and integrating over Ω, this shows (27). The estimate (28) follows from $g_\gamma(\boldsymbol{u}, \boldsymbol{u}) \leqslant |\gamma| \, \|\boldsymbol{u}\|^2_{\boldsymbol{L}^2(\Omega)}$ as well as

$$\|\nabla_{\mathfrak{h}} \boldsymbol{u}\|^2_{\boldsymbol{L}^2(\Omega)} \leqslant 2\big(\|\nabla \boldsymbol{u}\|^2_{\boldsymbol{L}^2(\Omega)} + \kappa^2 \|\boldsymbol{u}\|^2_{\boldsymbol{L}^2(\Omega)}\big).$$

Combining these estimates, we conclude the proof. □

Finally, we can prove the convergence of Algorithm A.

Proof of Theorem 4. The proof is split into five steps.

Step 1 (Algorithm A is well-defined): We note that $\mathcal{J}(\boldsymbol{u}) \geqslant 0$ is convex, continuous, and coercive. By convexity of $\mathcal{J}$, (17) is equivalent to finding minimizers of $\mathcal{J}(\boldsymbol{u}^n - (\cdot))$ over $\mathcal{K}[\boldsymbol{u}^n]$ by means of the Euler–Lagrange equations of the corresponding energy. Since $\mathcal{J}$ is convex and continuous, it is also sequentially lower semicontinuous, and $\mathcal{K}[\boldsymbol{u}^n]$ is a reflexive Banach space. Therefore, the existence of at least one solution to (17) follows by the direct method of calculus of variations, which means that step (i) in Algorithm A is well-defined. For all $n \in \mathbb{N}$ there holds pointwise orthogonality $\boldsymbol{u}^n \perp \boldsymbol{w}^n$ and thus

$$|\boldsymbol{u}^n - \boldsymbol{w}^n|^2 = |\boldsymbol{u}^n|^2 + |\boldsymbol{w}^n|^2 \geqslant 1 \quad \text{a.e. in } \Omega.$$

Therefore, $\boldsymbol{u}^n - \boldsymbol{w}^n \in \mathcal{M}^+$ and step (ii) in Algorithm A is always well-defined.

Step 2 (Energy decrease along computed sequence): It is an immediate consequence of Lemma 6 and Lemma 7 that

$$\mathcal{J}(\boldsymbol{u}^{n+1}) = \mathcal{J}(\Pi(\boldsymbol{u}^n - \boldsymbol{w}^n)) \overset{(22)}{\leqslant} \mathcal{J}(\boldsymbol{u}^n - \boldsymbol{w}^n) \overset{(23)}{\leqslant} \mathcal{J}(\boldsymbol{u}^n) \quad \text{for all } n \in \mathbb{N}_0. \tag{29}$$

Step 3 (Energy of tangential updates vanishes in the limit): Together with $\boldsymbol{u}^{n+1} = \Pi(\boldsymbol{u}^n - \boldsymbol{w}^n)$, Lemma 6 and Lemma 7 yield

$$0 \leqslant \mathcal{J}(\boldsymbol{w}^n) \overset{(22)}{=} \mathcal{J}(\boldsymbol{u}^n) - \mathcal{J}(\boldsymbol{u}^n - \boldsymbol{w}^n) \overset{(23)}{\leqslant} \mathcal{J}(\boldsymbol{u}^n) - \mathcal{J}(\boldsymbol{u}^{n+1}).$$

Summing the last inequality over $n = 0, \dots, N$, we get that

$$0 \leqslant \sum_{n=0}^{N} \mathcal{J}(\boldsymbol{w}^n) \leqslant \mathcal{J}(\boldsymbol{u}^0) - \mathcal{J}(\boldsymbol{u}^{N+1}) \leqslant \mathcal{J}(\boldsymbol{u}^0) \qquad \text{for all } N \in \mathbb{N}.$$

In particular, this yields that

$$\mathcal{J}(\boldsymbol{w}^n) \to 0 \quad \text{as } n \to \infty. \tag{30}$$

Step 4 (Existence of limit $\boldsymbol{u}$ of weakly convergent subsequence): We infer from Lemma 8, (29), and the fact that $|\boldsymbol{u}^n| = 1$ a.e. in Ω that

$$\|\boldsymbol{u}^n\|_{\boldsymbol{H}^1(\Omega)}^2 \overset{(27)}{\lesssim} \mathcal{J}(\boldsymbol{u}^n) + \|\boldsymbol{u}^n\|_{\boldsymbol{L}^2(\Omega)} \overset{(29)}{\leqslant} \mathcal{J}(\boldsymbol{u}^0) + |\Omega|. \tag{31}$$

In particular, $(\boldsymbol{u}^n)_{n\in\mathbb{N}}$ is uniformly bounded in $\boldsymbol{H}^1(\Omega)$. Hence there exists a subsequence of $(\boldsymbol{u}^n)_{n\in\mathbb{N}}$ (which is not relabeled) and a limit function $\boldsymbol{u} \in \boldsymbol{H}^1(\Omega)$ such that $\boldsymbol{u}^n \rightharpoonup \boldsymbol{u}$ in $\boldsymbol{H}^1(\Omega)$. Since $\mathcal{M}$ is weakly closed in $\boldsymbol{H}^1(\Omega)$, we conclude that $\boldsymbol{u} \in \mathcal{M}$.

Step 5 ($\boldsymbol{u}$ solves Euler–Lagrange equations (13) of $\mathcal{J}$): Let $\boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \cap \boldsymbol{L}^\infty(\Omega)$. Note that $\boldsymbol{v} \times \boldsymbol{u}^n \in \mathcal{K}[\boldsymbol{u}^n]$. The properties of the triple product imply that

$$\nabla_{\!\mathfrak{h}} \boldsymbol{u}^n : \nabla_{\!\mathfrak{h}} (\boldsymbol{v} \times \boldsymbol{u}^n) = \boldsymbol{u} \times \nabla_{\!\mathfrak{h}} \boldsymbol{u}^n : \nabla_{\!\mathfrak{h}} \boldsymbol{v}.$$

Using this identity, we derive

$$\begin{aligned}
&\int_\Omega \big[(\boldsymbol{u}^n \times \nabla_{\!\mathfrak{h}} \boldsymbol{u}^n) : \nabla_{\!\mathfrak{h}} \boldsymbol{v} + g_\gamma(\boldsymbol{u}^n, \boldsymbol{v} \times \boldsymbol{u}^n) \big] \,\mathrm{d}x = a(\boldsymbol{u}^n, \boldsymbol{v} \times \boldsymbol{u}^n) = a(\boldsymbol{w}^n, \boldsymbol{v} \times \boldsymbol{u}^n) \\
&\quad \leqslant a(\boldsymbol{w}^n, \boldsymbol{w}^n)^{1/2} a(\boldsymbol{v} \times \boldsymbol{u}^n, \boldsymbol{v} \times \boldsymbol{u}^n)^{1/2} \\
&\quad = 2\, \mathcal{J}(\boldsymbol{w}^n)^{1/2} \mathcal{J}(\boldsymbol{v} \times \boldsymbol{u}^n)^{1/2} \\
&\quad \overset{(28)}{\lesssim} \mathcal{J}(\boldsymbol{w}^n)^{1/2} \|\boldsymbol{v} \times \boldsymbol{u}^n\|_{\boldsymbol{H}^1(\Omega)} \lesssim \mathcal{J}(w^n)\, \|u^n\|_{\boldsymbol{H}^1(\Omega)} \overset{(31)}{\lesssim} \mathcal{J}(w^n).
\end{aligned} \tag{32}$$

With Step 3, the right-hand side of (32) vanishes in the limit. For the left-hand side, we can exploit weak convergence in $\boldsymbol{H}^1(\Omega)$ and strong convergence in $\boldsymbol{L}^2(\Omega)$ of $\boldsymbol{u}^n$. Overall, (32) converges to

$$\int_\Omega (\boldsymbol{u} \times \nabla_{\!\mathfrak{h}} \boldsymbol{u}) : \nabla_{\!\mathfrak{h}} \boldsymbol{v} + g_\gamma(\boldsymbol{u}, \boldsymbol{v} \times \boldsymbol{u}) \,\mathrm{d}x = 0 \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \cap \boldsymbol{L}^\infty(\Omega).$$

Together with Lemma 3, this concludes the proof. □

4. Finite element discretization

4.1. Algorithm. In this section, we extend the approach of [Bar05]. To this end, let $\mathcal{T}_h$ be a conforming triangulation of Ω into compact non-degenerate triangles. Let $\mathcal{N}_h$ be the set of nodes of $\mathcal{T}_h$. We define the lowest-order finite element space

$$\mathcal{S}^1(\mathcal{T}_h) := \{v \in H^1(\Omega) \mid v|_T \text{ is affine } \forall T \in \mathcal{T}_h\}.$$

A convenient basis of $\mathcal{S}^1(\mathcal{T}_h)$ is the so-called nodal basis $\{\varphi_{\mathbf{z}} \mid \mathbf{z} \in \mathcal{N}_h\}$ consisting of hat functions, i.e., $\mathcal{T}_h$-piecewise affine functions that satisfy $\varphi_{\mathbf{z}}(\mathbf{z}') = \delta_{\mathbf{z}\mathbf{z}'}$ for all $\mathbf{z}, \mathbf{z}' \in \mathcal{N}_h$. A vector-valued finite element function $\boldsymbol{v}_h \in \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h) := [\mathcal{S}^1(\mathcal{T}_h)]^3$ and its Jacobian can then be written as

$$\boldsymbol{v}_h = \sum_{\mathbf{z}\in\mathcal{N}_h} \boldsymbol{v}_h(\mathbf{z}) \varphi_{\mathbf{z}} \quad \text{and} \quad \nabla \boldsymbol{v}_h = \sum_{\mathbf{z}\in\mathcal{N}_h} \boldsymbol{v}_h(\mathbf{z}) \nabla \varphi_{\mathbf{z}},$$

respectively, where the column vectors $\boldsymbol{v}_h(\mathbf{z}) \in \mathbb{R}^3$ are the values at the nodes $\mathbf{z} \in \mathcal{N}_h$; note that $\nabla \boldsymbol{v}_h$ is pointwise in $\mathbb{R}^{3\times 2}$ with $(\nabla \boldsymbol{v}_h)_{jk} = \partial_k v_{h,j}$.

For $\boldsymbol{u}_h \in \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h)$, the constraint $|\boldsymbol{u}_h| = 1$ a.e. in Ω can only be satisfied by constant functions. Therefore, it is natural to prescribe the constraint only at the nodes. In analogy to Section 3, we define $\mathcal{M}_h := \{\boldsymbol{u}_h \in \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h) \mid |\boldsymbol{u}_h(\mathbf{z})| = 1 \text{ for all } \mathbf{z} \in \mathcal{N}_h\}$ and $\mathcal{K}_h[\boldsymbol{u}_h] := \{\boldsymbol{v}_h \in \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h) \mid \boldsymbol{v}_h(\mathbf{z}) \cdot \boldsymbol{u}_h(\mathbf{z}) = 0 \text{ for all } \mathbf{z} \in \mathcal{N}_h\}$ as well as the nodal unit-length projection

$$\Pi_h \colon \mathcal{M}_h^+ \to \mathcal{M}_h \colon \boldsymbol{v}_h \mapsto \sum_{\mathbf{z}\in\mathcal{N}_h} \frac{\boldsymbol{v}_h(\mathbf{z})}{|\boldsymbol{v}_h(\mathbf{z})|}\varphi_{\mathbf{z}}, \tag{33}$$
$$\text{with } \mathcal{M}_h^+ := \big\{\boldsymbol{v}_h \in \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h) \,\big|\, |\boldsymbol{v}_h(\mathbf{z})| \geqslant 1 \text{ for all } \mathbf{z} \in \mathcal{N}_h\big\}.$$

This means that, in general, the unit-length constraint as well as orthogonality cannot be achieved in the interior of triangles. However, there still holds $|\boldsymbol{u}_h| \leqslant 1$ a.e. in Ω for any $\boldsymbol{u}_h \in \mathcal{M}_h$. We define the discrete version of (5) as

$$\begin{aligned} \mathcal{J}_h(\boldsymbol{v}_h) &:= \frac{1}{2} a_h(\boldsymbol{v}_h, \boldsymbol{v}_h) \quad \text{for all } \boldsymbol{v}_h \in \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h), \text{ where} \\ a_h(\boldsymbol{v}_h, \boldsymbol{w}_h) &:= \int_\Omega \Big[\nabla_{\mathfrak{h}} \boldsymbol{v}_h : \nabla_{\mathfrak{h}} \boldsymbol{w}_h + \mathcal{I}_h\big(g_\gamma(\boldsymbol{v}_h, \boldsymbol{w}_h)\big)\Big] \,\mathrm{d}x \quad \text{for all } \boldsymbol{v}_h, \boldsymbol{w}_h \in \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h), \end{aligned} \tag{34}$$

where $\mathcal{I}_h \colon C(\overline{\Omega}) \to \mathcal{S}^1(\mathcal{T}_h)$, $\mathcal{I}_h f := \sum_{\mathbf{z}\in\mathcal{N}_h} f(\mathbf{z})\varphi_{\mathbf{z}}$ is the nodal interpolation operator. Then, the discrete minimization problem reads:

$$\text{Find } \boldsymbol{u}_h \in \mathcal{M}_h \text{ such that} \quad \mathcal{J}_h(\boldsymbol{u}_h) = \min_{\boldsymbol{v}_h \in \mathcal{M}_h} \mathcal{J}_h(\boldsymbol{v}_h). \tag{35}$$

The term $\int_\Omega \mathcal{I}_h(g_\gamma(\boldsymbol{v}_h, \boldsymbol{v}_h)) \,\mathrm{d}x$ will be referred to as *mass lumping* term. In analogy to [Bar05], the discrete problem (35) is solved via a discretized version of Algorithm A.

Algorithm B (Discrete energy minimization). ***Input:*** *Conforming simplicial triangulation $\mathcal{T}_h$, initial guess $\boldsymbol{u}_h^0 \in \mathcal{M}_h$, tolerance $\mathrm{tol} > 0$.*
Loop: *For all $n = 0, 1, 2, \dots$ do*

(i) *Compute $\boldsymbol{w}_h^n \in \mathcal{K}_h[\boldsymbol{u}_h^n]$ such that*

$$a_h(\boldsymbol{w}_h^n, \boldsymbol{v}_h) = a_h(\boldsymbol{u}_h^n, \boldsymbol{v}_h) \qquad \text{for all } \boldsymbol{v}_h \in \mathcal{K}_h[\boldsymbol{u}_h^n]. \tag{36}$$

(ii) *If $\mathcal{J}_h(\boldsymbol{w}_h^n) \leqslant \mathrm{tol}$, define $\boldsymbol{u}_h := \boldsymbol{u}_h^n$ and terminate.*
(iii) *Otherwise, set $\boldsymbol{u}_h^{n+1} := \Pi_h(\boldsymbol{u}_h^n - \boldsymbol{w}_h^n)$.*

Output: *Approximate minimizer $\boldsymbol{u}_h \in \mathcal{M}_h$ of $\mathcal{J}_h$.*

Convergence of Algorithm B requires the following technical assumption on the triangulation.

Definition 9 (Angle condition). *For any two triangles $T_1, T_2 \in \mathcal{T}_h$ that share an edge (and, hence, two nodes), let α_1 and α_2 be the interior angles of the triangles T_1 and T_2 at the node which they do not share, respectively; see Figure 1 for a visualization. We say that $\mathcal{T}_h$ satisfies the angle condition, if there holds*

$$\cot \alpha_1 + \cot \alpha_2 \geqslant 0 \quad \text{for all neighboring } T_1, T_2 \in \mathcal{T}_h. \tag{37}$$

A sufficient condition for (37) to hold is $\alpha_1 + \alpha_2 \leqslant \pi$. In particular, this is clearly satisfied if all angles of $\mathcal{T}_h$ are non-obtuse.

The next theorem states convergence of Algorithm B to a stationary point of $\mathcal{J}$ for $\kappa = 0$; see also the discussion in Section 4.4 on the case $\kappa \neq 0$. The proof of Theorem 10 is postponed to Section 4.2 below.

FIGURE 1. Illustration of the angle condition with notation from Definition 9. Left: configuration satisfying the angle condition. Right: configuration *not* satisfying the angle condition.

Theorem 10. *Let $\kappa = 0$. For any $h > 0$, let $\mathcal{T}_h$ be a conforming simplicial triangulation with maximal mesh-width h that satisfies the angle condition (37). Let further $\mathrm{tol}_h > 0$ be a tolerance with $\mathrm{tol}_h \to 0$ as $h \to 0$. Then, there hold the following statements* (i)–(ii)*:*

(i) *Algorithm B is well-defined, i.e., for any $\boldsymbol{u}_h^n \in \mathcal{M}_h$, the variational problem (36) admits a solution $\boldsymbol{w}_h^n \in \mathcal{K}_h[\boldsymbol{u}_h^n]$ so that $\boldsymbol{u}_h^n - \boldsymbol{w}_h^n \in \mathcal{M}_h^+$. Moreover, the termination criterion $\mathcal{J}_h(\boldsymbol{w}_h^n) \leqslant \mathrm{tol}_h$ is reached in finite time and Algorithm B provides $\boldsymbol{u}_h = \boldsymbol{u}_h^n \in \mathcal{M}_h$ with $\mathcal{J}_h(\boldsymbol{u}_h) \leqslant \mathcal{J}_h(\boldsymbol{u}_h^0)$ for some $n \in \mathbb{N}_0$.*

(ii) *Suppose that the initial guesses $\boldsymbol{u}_h^0 \in \mathcal{M}_h$ are uniformly bounded, i.e., there exists $C_0 > 0$ such that $\mathcal{J}_h(\boldsymbol{u}_h^0) \leqslant C_0$ for all $h > 0$. Let $\boldsymbol{u}_h$ be the output of Algorithm B with $(\mathcal{T}_h, \boldsymbol{u}_h^0, \mathrm{tol}_h)$ as input. Then, there exists a weak solution $\boldsymbol{u} \in \mathcal{M}$ of the Euler–Lagrange equations (11) of $\mathcal{J}$ and a subsequence of $(\boldsymbol{u}_h)_h$ such that, along this subsequence,*

$$\boldsymbol{u}_h \rightharpoonup \boldsymbol{u} \quad \text{weakly in } \boldsymbol{H}^1(\Omega) \qquad \text{and} \qquad \mathcal{J}(\boldsymbol{u}) \leqslant \liminf_{h \to 0} \mathcal{J}_h(\boldsymbol{u}_h). \tag{38}$$

Remark 11. *Note that the proof of well-definedness of the discrete problem (36) below does not make use of $\kappa = 0$. Thus, Algorithm B is well-defined for all $\kappa \in \mathbb{R}$. However, the current proof of energy reduction and hence convergence (resp. termination) relies on $\kappa = 0$, even though Algorithm B empirically appears to converge for any $\kappa \in \mathbb{R}$; see Section 4.5 below for a related practical experiment.*

As for the continuous case in Section 3, one can prove that the discrete tangential update $\boldsymbol{w}_h^n \in \mathcal{K}[\boldsymbol{u}_h^n]$ of (36) exists, but can be non-unique in general. The next proposition comments on the uniqueness of the tangential update in the discrete setting. It suggests additional linear constraints to add to the discrete system of equations of (17) to enforce uniqueness.

Proposition 12. *Suppose that Ω is connected. Let $\boldsymbol{u}_h \in \mathcal{M}_h$ and $\boldsymbol{w}_h^1, \boldsymbol{w}_h^2 \in \mathcal{K}_h[\boldsymbol{u}_h]$ be two solutions of the discrete Euler–Lagrange equations (36), i.e.,*

$$a_h(\boldsymbol{w}_h^1, \boldsymbol{v}_h) = a_h(\boldsymbol{u}_h, \boldsymbol{v}_h) = a_h(\boldsymbol{w}_h^2, \boldsymbol{v}_h) \quad \text{for all } \boldsymbol{v}_h \in \mathcal{K}_h[\boldsymbol{u}_h].$$

Then, there hold the following assertions:

(a) *For $\kappa = 0$, the difference $\boldsymbol{w}_h^1 - \boldsymbol{w}_h^2 \in \mathcal{K}_h[\boldsymbol{u}_h]$ is constant and satisfies*
- *$\boldsymbol{w}_h^1(\mathbf{z}) - \boldsymbol{w}_h^2(\mathbf{z}) \perp \mathbf{e}_3$ for all $\mathbf{z} \in \mathcal{N}_h$ if $\gamma > 0$,*
- *$\boldsymbol{w}_h^1(\mathbf{z}) - \boldsymbol{w}_h^2(\mathbf{z}) \in \mathrm{span}\{\mathbf{e}_3\}$ for all $\mathbf{z} \in \mathcal{N}_h$ if $\gamma < 0$.*

(b) *For $\kappa \neq 0$, i.e., in the presence of the anti-symmetric exchange term, the difference $\boldsymbol{w}_h^1 - \boldsymbol{w}_h^2$ vanishes. In particular, the solution of (17) is unique.*

Moreover, let $\kappa = 0$ and let

$$B := \begin{cases} \{\boldsymbol{u}_h(\mathbf{z}) \mid \mathbf{z} \in \mathcal{N}_h\} & \text{if } \gamma = 0, \\ \{\boldsymbol{u}_h(\mathbf{z}) \mid \mathbf{z} \in \mathcal{N}_h\} \cup \{\mathbf{e}_3\} & \text{if } \gamma > 0, \\ \{\boldsymbol{u}_h(\mathbf{z}) \mid \mathbf{z} \in \mathcal{N}_h\} \cup \{\mathbf{e}_1, \mathbf{e}_2\} & \text{if } \gamma < 0. \end{cases}$$

Let further $B_\perp$ be an orthonormal basis of the orthogonal complement $(\operatorname{span} B)^\perp$. Let m be the cardinality of $B_\perp = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$. Adding the linear constraints $\int_\Omega \boldsymbol{w}_h \cdot \mathbf{b} \,\mathrm{d}x = 0$ for all $\mathbf{b} \in B_\perp$ to (36) as Lagrange multipliers yields the following saddle point problem: find $(\boldsymbol{w}_h, \boldsymbol{\lambda}) \in \mathcal{K}_h[\boldsymbol{u}_h] \times \mathbb{R}^m$ such that

$$\begin{aligned} a_h(\boldsymbol{w}_h, \boldsymbol{v}_h) + b_h(\boldsymbol{v}_h, \boldsymbol{\lambda}) &= a_h(\boldsymbol{u}_h, \boldsymbol{v}_h) && \text{for all } \boldsymbol{v}_h \in \mathcal{K}_h[\boldsymbol{u}_h], \\ b_h(\boldsymbol{w}_h, \boldsymbol{\mu}) &= 0 && \text{for all } \boldsymbol{\mu} \in \mathbb{R}^m \end{aligned} \tag{39}$$

with the bilinear form $b_h \colon \mathcal{K}_h[\boldsymbol{u}_h] \times \mathbb{R}^m \to \mathbb{R}$ defined by

$$b_h(\boldsymbol{v}_h, \boldsymbol{\mu}) := \sum_{i=1}^m \mu_i \int_\Omega \boldsymbol{v}_h \cdot \mathbf{b}_i \,\mathrm{d}x \quad \text{for all } \boldsymbol{v}_h \in \mathcal{K}_h[\boldsymbol{u}_h], \boldsymbol{\mu} \in \mathbb{R}^m.$$

The saddle point problem (39) admits a unique solution $(\boldsymbol{w}_h, \boldsymbol{\lambda}) \in \mathcal{K}_h[\boldsymbol{u}_h] \times \mathbb{R}^m$ and $\boldsymbol{w}_h$ solves (36).

Proof. The statements (a)–(b) follow analogously to the proof of Proposition 5. To show that the saddle point problem (39) admits a unique solution, we employ the Brezzi theorem [BBF13], i.e., we show that the operator $\mathcal{B} \colon \mathcal{K}_h[\boldsymbol{u}_h] \to (\mathbb{R}^m)^* : \boldsymbol{v}_h \mapsto b_h(\boldsymbol{v}_h, \cdot)$ is surjective and that $a_h(\cdot, \cdot)$ is coercive on the kernel of $\mathcal{B}$.

To show surjectivity, it suffices to show that, for any $\boldsymbol{\mu} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$, there exists $\boldsymbol{v}_h \in \mathcal{K}_h[\boldsymbol{u}_h]$ such that $b_h(\boldsymbol{v}_h, \boldsymbol{\mu}) \neq 0$. Given $\boldsymbol{\mu} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$, we define the constant function $\boldsymbol{v}_h := \sum_{i=1}^m \mu_i \mathbf{b}_i$. From the definition of $B_\perp$, we obtain that $\boldsymbol{v}_h \in \mathcal{K}_h[\boldsymbol{u}_h]$ and

$$b_h(\boldsymbol{v}_h, \boldsymbol{\mu}) = \sum_{i=1}^m \mu_i \int_\Omega \Big(\sum_{j=1}^m \mu_j \mathbf{b}_j \Big) \cdot \mathbf{b}_i \,\mathrm{d}x = \sum_{i=1}^m \mu_i^2 \int_\Omega 1 \,\mathrm{d}x = |\boldsymbol{\mu}|^2 |\Omega| > 0.$$

To show coercivity of $a_h(\cdot, \cdot)$ on the kernel of $\mathcal{B}$, let $\boldsymbol{v}_h \in \mathcal{K}_h[\boldsymbol{u}_h]$ with $\mathcal{B}\boldsymbol{v}_h = \mathbf{0}$ and note that $a(\boldsymbol{v}_h, \boldsymbol{v}_h) = 0$ implies that $\boldsymbol{v}_h \equiv \mathbf{c} \in \mathbb{R}^3$ is constant as well as $\mathbf{c} \perp \operatorname{span} B$ according to the statements (a)–(b) of the proposition. Since $|\Omega| \mathbf{c} \cdot \mathbf{b} = \int_\Omega \boldsymbol{v}_h \cdot \mathbf{b} \,\mathrm{d}x = b_h(\boldsymbol{v}_h, \mathbf{b}) = 0$ for all $\mathbf{b} \in B_\perp$ due to $\mathcal{B}\mathbf{c} = \mathcal{B}\boldsymbol{v}_h = \mathbf{0}$, we obtain that also $\mathbf{c} \perp \operatorname{span} B_\perp = (\operatorname{span} B)^\perp$ and, hence, $\mathbf{c} = \mathbf{0}$.

Thus, the assumptions of the Brezzi theorem are satisfied and the saddle point problem (39) admits a unique solution $(\boldsymbol{w}_h, \boldsymbol{\lambda}) \in \mathcal{K}_h[\boldsymbol{u}_h] \times \mathbb{R}^m$. To show that $\boldsymbol{w}_h$ solves (36), we test (39) with the constant functions $\boldsymbol{v}_h := \mathbf{b}_i \in \mathcal{K}_h[\boldsymbol{u}_h]$ for $i = 1, \dots, m$. This yields that $b_h(\mathbf{b}_i, \boldsymbol{\lambda}) = 0$ for all $i = 1, \dots, m$ and, hence, $\boldsymbol{\lambda} = \mathbf{0} \in \mathbb{R}^m$. Thus, we see that $b_h(\boldsymbol{v}_h, \boldsymbol{\lambda}) = 0$ for all $\boldsymbol{v}_h \in \mathcal{K}_h[\boldsymbol{u}_h]$ and, by plugging this into (39), that $\boldsymbol{w}_h$ solves (36). □

Remark 13. *Two comments on the practical implementation of the discrete system (36) with additional constraints are in order.*

(i) *First, recall from [KPP+19] how the solution to equation (36) can be computed using Householder matrices: Let $\mathbf{u} \in \mathbb{R}^3$ be a vector with $|\mathbf{u}| = 1$. Then, if $\mathbf{u} = -\mathbf{e}_3$, define the diagonal matrix $\mathbf{Q}[\mathbf{u}] := \operatorname{diag}(-1, -1, 1) \in \mathbb{R}^{3\times3}$, else define, with the identity*

matrix $\mathbf{I}_{3\times3} \in \mathbb{R}^{3\times3}$,

$$\mathbf{Q}[\mathbf{u}] := \mathbf{I}_{3\times3} - 2\mathbf{v}\mathbf{v}^\intercal \in \mathbb{R}^{3\times3}, \quad \text{where} \quad \mathbf{v} := \frac{\mathbf{u}+\mathbf{e}_3}{|\mathbf{u}+\mathbf{e}_3|} \in \mathbb{R}^3 \quad \text{with } |\mathbf{v}| = 1.$$

Note that $\mathbf{Q}[\mathbf{u}]$ *is orthogonal and* $(\mathbf{Q}[\mathbf{u}]\mathbf{e}_i)_{i=1}^3$ *is an orthonormal basis of* $\mathbb{R}^3$ *with* $\mathbf{Q}[\mathbf{u}]\mathbf{e}_3 = -\mathbf{u}$ *and, consequently,* $\mathbf{Q}[\mathbf{u}]\mathbf{e}_i \perp \mathbf{u}$ *for* $i = 1, 2$. *We then define the prolongation operator* $\mathcal{Q}[\mathbf{u}]\colon \mathbb{R}^2 \to \mathbb{R}^3$ *by*

$$\mathcal{Q}[\mathbf{u}]\widehat{\mathbf{w}} := \sum_{i=1}^{2} \widehat{w}_i \mathbf{Q}[\mathbf{u}]\mathbf{e}_i \quad \text{for all } \widehat{\mathbf{w}} \in \mathbb{R}^2$$

and, for $\boldsymbol{u}_h \in \mathcal{M}_h$, *the discrete prolongation operator*

$$\mathcal{Q}_h[\boldsymbol{u}_h]\colon [\mathcal{S}^1(\mathcal{T}_h)]^2 \to \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h)\colon \mathcal{Q}_h[\boldsymbol{u}_h]\widehat{\boldsymbol{w}}_h := \sum_{\mathbf{z}\in\mathcal{N}_h} \mathcal{Q}[\boldsymbol{u}_h(\mathbf{z})]\widehat{\boldsymbol{w}}_h(\mathbf{z})\varphi_{\mathbf{z}}.$$

These definitions yield that $\mathcal{Q}_h[\boldsymbol{u}_h]\widehat{\boldsymbol{w}}_h \in \mathcal{K}_h[\boldsymbol{u}_h]$ *for all* $\boldsymbol{w}_h \in [\mathcal{S}^1(\mathcal{T}_h)]^2$. *Thus, the discrete Euler–Lagrange equations (36) can be transformed to a system of equations in* $[\mathcal{S}^1(\mathcal{T}_h)]^2$ *via the bijective map* $\mathcal{Q}_h[\boldsymbol{u}_h]\colon [\mathcal{S}^1(\mathcal{T}_h)]^2 \to \mathcal{K}_h[\boldsymbol{u}_h]$: *find* $\widehat{\boldsymbol{w}}_h \in [\mathcal{S}^1(\mathcal{T}_h)]^2$ *such that*

$$a_h(\mathcal{Q}_h[\boldsymbol{u}_h]\widehat{\boldsymbol{w}}_h, \mathcal{Q}_h[\boldsymbol{u}_h]\widehat{\boldsymbol{v}}_h) = a_h(\boldsymbol{u}_h, \mathcal{Q}_h[\boldsymbol{u}_h]\widehat{\boldsymbol{v}}_h) \quad \text{for all } \widehat{\boldsymbol{v}}_h \in [\mathcal{S}^1(\mathcal{T}_h)]^2,$$

and set $\boldsymbol{w}_h := \mathcal{Q}_h[\boldsymbol{u}_h]\widehat{\boldsymbol{w}}_h$. *This system of* $2\#\mathcal{N}_h$ *equations can be solved efficiently by iterative solvers, e.g., the generalized minimal residual method (GMRES) from [SS86].*

(ii) *With* B *from Proposition 12, in most practical situations there holds that* $\operatorname{span} B = \mathbb{R}^3$ *so that the tangential update* $\boldsymbol{w}_h$ *is unique without additional constraints. It is easy to check the dimension of* $\operatorname{span} B$ *to determine whether additional constraints are necessary.*

E.g., if $\gamma = 0$, *take an arbitrary node* $\mathbf{z}^0 \in \mathcal{N}_h$ *and compute* $\mathbf{p}_{\mathbf{z}} := \boldsymbol{u}_h(\mathbf{z}^0) \times \boldsymbol{u}_h(\mathbf{z})$ *for all* $\mathbf{z} \in \mathcal{N}_h$. *If* $\mathbf{p}_{\mathbf{z}} = \mathbf{0}$ *for all* $\mathbf{z} \in \mathcal{N}_h$, *then* $B_\perp$ *consists of any two orthonormal vectors orthogonal to* $\boldsymbol{u}_h(\mathbf{z}^0)$ *and two constraints need to be added to (36). Otherwise, let* $\mathbf{z}^1 \in \mathcal{N}_h$ *be any node with* $\mathbf{p}_{\mathbf{z}^1} \neq \mathbf{0}$. *If* $\mathbf{p}_{\mathbf{z}^1} \cdot \boldsymbol{u}_h(\mathbf{z}) = 0$ *for all* $\mathbf{z} \in \mathcal{N}_h$, *then* $B_\perp = \{\mathbf{p}_{\mathbf{z}^1}/|\mathbf{p}_{\mathbf{z}^1}|\}$ *and one constraint needs to be added to (36). Otherwise,* $B_\perp = \{\mathbf{0}\}$ *and no additional constraint is necessary.*

The cases $\gamma > 0$ *and* $\gamma < 0$ *can be treated analogously.*

4.2. Proof of Theorem 10. To prove convergence of Algorithm B, we intend to mimic the argument for the proof of convergence of Algorithm A. First, we ensure that both steps of Algorithm B decrease the discrete energy $\mathcal{J}_h$. We start with step (i).

Lemma 14. *Let* $\boldsymbol{u}_h \in \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h)$ *and* $\boldsymbol{w}_h \in \mathcal{K}_h[\boldsymbol{u}_h]$ *such that* $a_h(\boldsymbol{w}_h, \boldsymbol{v}_h) = a_h(\boldsymbol{u}_h, \boldsymbol{v}_h)$ *for all* $\boldsymbol{v}_h \in \mathcal{K}_h[\boldsymbol{u}_h]$. *Then, it holds that*

$$\mathcal{J}_h(\boldsymbol{u}_h - \boldsymbol{w}_h) = \mathcal{J}_h(\boldsymbol{u}_h) - \mathcal{J}_h(\boldsymbol{w}_h) \le \mathcal{J}_h(\boldsymbol{u}_h). \tag{40}$$

Proof. The proof is analogous to the continuous case; see Lemma 6. □

In contrast to the continuous case, $|\boldsymbol{u}_h - \boldsymbol{w}_h| \geqslant 1$ is not guaranteed to hold true except at vertices $\mathbf{z} \in \mathcal{N}_h$. Therefore, the analysis of the projection in step (iii) requires a different approach from the continuous case. We note that the angle condition is crucial for treating the exchange energy [Bar05, Lemma 3.2]. For the convenience of the reader, we include the main argument of [Bar05] in the proof of the next lemma.

Lemma 15. *Let $\kappa = 0$ and let $\mathcal{T}_h$ satisfy the angle condition (37). Then, it holds that*

$$\mathcal{J}_h(\Pi_h(\boldsymbol{v}_h)) \leqslant \mathcal{J}_h(\boldsymbol{v}_h) \qquad \text{for all } \boldsymbol{v}_h \in \mathcal{M}_h^+. \tag{41}$$

Proof. The proof is split into two steps.

Step 1: We recall the argument of [Bar05] to show that

$$\int_\Omega |\nabla(\Pi_h \boldsymbol{v}_h)|^2 \,\mathrm{d}x \leqslant \int_\Omega |\nabla \boldsymbol{v}_h|^2 \,\mathrm{d}x. \tag{42}$$

Note that the angle condition (37) implies that the off-diagonal entries

$$A_{\mathbf{z}\mathbf{z}'} := \int_\Omega \nabla\varphi_{\mathbf{z}} \cdot \nabla\varphi_{\mathbf{z}'} \,\mathrm{d}x = -(\cot\alpha_1 + \cot\alpha_2) \leq 0$$

of the stiffness matrix A are non-positive. For any function $\boldsymbol{v}_h \in \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h)$, we have that

$$\int_\Omega |\nabla \boldsymbol{v}_h|^2 \,\mathrm{d}x = \sum_{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h} \boldsymbol{v}_h(\mathbf{z}) \cdot \boldsymbol{v}_h(\mathbf{z}') \int_\Omega \nabla\varphi_{\mathbf{z}} \cdot \nabla\varphi_{\mathbf{z}'} \,\mathrm{d}x = \sum_{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h} A_{\mathbf{z}\mathbf{z}'} \boldsymbol{v}_h(\mathbf{z}) \cdot \boldsymbol{v}_h(\mathbf{z}').$$

From the definition of the stiffness matrix A, we infer that there holds

$$\sum_{\mathbf{z}\in\mathcal{N}_h} A_{\mathbf{z}\mathbf{z}'} = \int_\Omega \nabla 1 \cdot \nabla\varphi_{\mathbf{z}'} \,\mathrm{d}x = 0 \qquad \text{for all } \mathbf{z}' \in \mathcal{N}_h.$$

This leads to

$$\int_\Omega |\nabla \boldsymbol{v}_h|^2 \,\mathrm{d}x = \sum_{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h} A_{\mathbf{z}\mathbf{z}'} \boldsymbol{v}_h(\mathbf{z}) \cdot \boldsymbol{v}_h(\mathbf{z}') - \sum_{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h} A_{\mathbf{z}\mathbf{z}'} \boldsymbol{v}_h(\mathbf{z}')^2 = \sum_{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h} A_{\mathbf{z}\mathbf{z}'} \big(\boldsymbol{v}_h(\mathbf{z}) - \boldsymbol{v}_h(\mathbf{z}')\big) \cdot \boldsymbol{v}_h(\mathbf{z}').$$

Therefore, symmetry of the stiffness matrix leads to

$$\begin{aligned}\int_\Omega |\nabla \boldsymbol{v}_h|^2 \,\mathrm{d}x &= \frac{1}{2} \sum_{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h} A_{\mathbf{z}\mathbf{z}'} \big(\boldsymbol{v}_h(\mathbf{z}) - \boldsymbol{v}_h(\mathbf{z}')\big) \cdot \boldsymbol{v}_h(\mathbf{z}') + \frac{1}{2} \sum_{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h} A_{\mathbf{z}\mathbf{z}'} \big(\boldsymbol{v}_h(\mathbf{z}') - \boldsymbol{v}_h(\mathbf{z})\big) \cdot \boldsymbol{v}_h(\mathbf{z}). \\ &= -\frac{1}{2} \sum_{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h} A_{\mathbf{z}\mathbf{z}'} \big|\boldsymbol{v}_h(\mathbf{z}) - \boldsymbol{v}_h(\mathbf{z}')\big|^2 = -\frac{1}{2} \sum_{\substack{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h \\ \mathbf{z}\neq\mathbf{z}'}} A_{\mathbf{z}\mathbf{z}'} \big|\boldsymbol{v}_h(\mathbf{z}) - \boldsymbol{v}_h(\mathbf{z}')\big|^2\end{aligned}$$

Note that the mapping $\boldsymbol{x} \in \{\tilde{\boldsymbol{x}} \in \mathbb{R}^3 \mid |\tilde{\boldsymbol{x}}| \geqslant 1\} \mapsto \boldsymbol{x}/|\boldsymbol{x}|$ is Lipschitz continuous with constant 1. With $A_{\mathbf{z}\mathbf{z}'} \leqslant 0$ for all $\mathbf{z} \neq \mathbf{z}'$ and $|\boldsymbol{v}_h(\mathbf{z})| \geqslant 1$ for all $\mathbf{z} \in \mathcal{N}_h$, we thus have

$$\begin{aligned}\int_\Omega |\nabla(\Pi_h \boldsymbol{v}_h)|^2 \,\mathrm{d}x &= -\frac{1}{2} \sum_{\substack{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h \\ \mathbf{z}\neq\mathbf{z}'}} A_{\mathbf{z}\mathbf{z}'} \left| \frac{\boldsymbol{v}_h(\mathbf{z})}{|\boldsymbol{v}_h(\mathbf{z})|} - \frac{\boldsymbol{v}_h(\mathbf{z}')}{|\boldsymbol{v}_h(\mathbf{z}')|} \right|^2 \\ &\leqslant -\frac{1}{2} \sum_{\substack{\mathbf{z},\mathbf{z}'\in\mathcal{N}_h \\ \mathbf{z}\neq\mathbf{z}'}} A_{\mathbf{z}\mathbf{z}'} \left|\boldsymbol{v}_h(\mathbf{z}) - \boldsymbol{v}_h(\mathbf{z}')\right|^2 = \int_\Omega |\nabla \boldsymbol{v}_h|^2 \,\mathrm{d}x\end{aligned}$$

This shows (42).

Step 2: We show that

$$\int_\Omega \mathcal{I}_h\big(g_\gamma(\Pi_h \boldsymbol{v}_h, \Pi_h \boldsymbol{v}_h)\big) \,\mathrm{d}x \leqslant \int_\Omega \mathcal{I}_h\big(g_\gamma(\boldsymbol{v}_h, \boldsymbol{v}_h)\big) \,\mathrm{d}x. \tag{43}$$

Since $|\boldsymbol{v}_h(\mathbf{z})| \geqslant 1$ for all $\mathbf{z} \in \mathcal{N}_h$ and since $g_\gamma(\cdot,\cdot)$ is bilinear, we have that

$$\begin{aligned}\int_\Omega \mathcal{I}_h\big(g_\gamma(\Pi_h \boldsymbol{v}_h, \Pi_h \boldsymbol{v}_h)\big)\,\mathrm{d}x &= \sum_{\mathbf{z}\in\mathcal{N}_h} \frac{g_\gamma(\boldsymbol{v}_h(\mathbf{z}),\boldsymbol{v}_h(\mathbf{z}))}{|\boldsymbol{v}_h(\mathbf{z})|^2} \int_\Omega \varphi_{\mathbf{z}}\,\mathrm{d}x \\ &\leqslant \sum_{\mathbf{z}\in\mathcal{N}_h} g_\gamma(\boldsymbol{v}_h(\mathbf{z}),\boldsymbol{v}_h(\mathbf{z})) \int_\Omega \varphi_{\mathbf{z}}\,\mathrm{d}x = \int_\Omega \mathcal{I}_h\big(g_\gamma(\boldsymbol{v}_h,\boldsymbol{v}_h)\big)\,\mathrm{d}x.\end{aligned}$$

This shows (43). Combining (42)–(43), we conclude the claim for $\kappa = 0$. □

Remark 16. *The proof of Lemma 15 relies on two facts that are specific to the case $\kappa = 0$: The angle condition on the triangulation implies the non-positivity of the off-diagonal entries of the stiffness matrix; combined with the 1-Lipschitz property of the nodal normalization $x \mapsto x/|x|$, this yields the exchange-energy estimate*

$$\int_\Omega |\nabla(\Pi_h \boldsymbol{v}_h)|^2 \leqslant \int_\Omega |\nabla \boldsymbol{v}_h|^2.$$

By contrast, when $\kappa \neq 0$, the helical (DMI) contribution produces mixed and lower-order terms that at the discrete level are not controlled by the angle condition alone; consequently, the total discrete energy may increase after projection, as we show in Section 4.4. Determining mesh or projection conditions that restore discrete monotonicity for $\kappa \neq 0$ remains an open question.

Nevertheless, the observed increases in the energy level are very small and sporadic in our tests and Algorithm B empirically converges also for $\kappa \neq 0$; see Section 4.5 for the numerical experiments.

For the proof of Theorem 10, the next lemma collects some basic properties of the nodal projection.

Lemma 17. *Recall the nodal projection $\mathcal{I}_h\colon C^0(\overline{\Omega}) \to \mathcal{S}^1(\mathcal{T}_h)$ from the definition of $\mathcal{J}_h$ in (34). Then, there hold the following properties* (i)–(iii).

(i) *For all $v \in C^2(\overline{\Omega})$ and all $1 \leqslant p < \infty$, there holds the approximation property*

$$\|(1-\mathcal{I}_h)v\|_{L^p(\Omega)} + h\|\nabla(1-\mathcal{I}_h)v\|_{L^p(\Omega)} \leq C_{\mathrm{apx}} h^2 \|D^2 v\|_{L^p(\Omega)}. \tag{44}$$

(ii) *For all $v_h \in \mathcal{S}^q(\Omega)$ with $q \in \mathbb{N}$ and all $1 \leqslant p < \infty$, there holds the inverse inequality*

$$h\|D^2 v_h\|_{L^p(\Omega)} \leq C_{\mathrm{inv}} \|\nabla v_h\|_{L^p(\Omega)}. \tag{45}$$

(iii) *For all $v_h \in \mathcal{S}^1(\Omega)$, there holds*

$$\|v_h\|^2_{L^2(\Omega)} \leqslant \int_\Omega \mathcal{I}_h(|v_h|^2)\,\mathrm{d}x. \tag{46}$$

The constant $C_{\mathrm{apx}} > 0$ depends only on shape regularity of $\mathcal{T}_h$ and the Lebesgue index p, while $C_{\mathrm{inv}} > 0$ depends only on shape regularity of $\mathcal{T}_h$, the Lebesgue index p, and the polynomial degree q.

Proof. For the proof of properties (i)–(ii) we refer to standard literature, e.g., [BS08]. For the proof of (iii), we note that on each element $T \in \mathcal{T}_h$, there holds the representation $v_h(x) = \sum_{\mathbf{z}\in\mathcal{N}_h(T)} \varphi_{\mathbf{z}}(x) v_h(\mathbf{z})$, where $\sum_{\mathbf{z}\in\mathcal{N}_h(T)} \varphi_{\mathbf{z}}(x) = 1$ and $\varphi_{\mathbf{z}}(x) \geqslant 0$ for all $x \in T$. Thus, by Jensen's inequality, we have that

$$\|v_h\|^2_{L^2(T)} = \int_T \Big|\sum_{\mathbf{z}\in\mathcal{N}(T)} \varphi_{\mathbf{z}} v_h(\mathbf{z})\Big|^2 \mathrm{d}x \leqslant \int_T \sum_{\mathbf{z}\in\mathcal{N}(T)} \varphi_{\mathbf{z}} |v_h(\mathbf{z})|^2\,\mathrm{d}x = \int_T \mathcal{I}_h(|v_h|^2)\,\mathrm{d}x.$$

Summing over all elements $T \in \mathcal{T}_h$, we conclude the proof. □

Finally, we can prove the second part of Theorem 10, i.e., convergence of the output of Algorithm B to a weak solution of the Euler–Lagrange equations (13).

Proof of Theorem 10. We split the proof into seven steps.

Step 1 (Algorithm B is well-defined): The proof of existence of at least one solution $\boldsymbol{w}_h^n$ to the discrete variational formulation (36) in step (i) of Algorithm B follows analogously to the continuous case; see step 1 in the proof of Theorem 4. For any $n \geqslant 0$, the fact that $\boldsymbol{w}_h^n \in \mathcal{K}_h[\boldsymbol{u}_h^n]$ yields that

$$|\boldsymbol{u}_h^n(\mathbf{z}) - \boldsymbol{w}_h^n(\mathbf{z})|^2 = |\boldsymbol{u}_h^n(\mathbf{z})|^2 + |\boldsymbol{w}_h^n(\mathbf{z})|^2 \geq 1 \quad \text{for all vertices } \mathbf{z} \in \mathcal{N}_h.$$

Therefore, for any $\kappa \in \mathbb{R}$, there holds $\boldsymbol{u}_h^n - \boldsymbol{w}_h^n \in \mathcal{M}_h^+$ and Step (iii) in Algorithm B is always well-defined.

Step 2 (Energy decrease along computed sequence): From Lemma 14 and Lemma 15 (exploiting $\kappa = 0$), we obtain that

$$\mathcal{J}_h\big(\boldsymbol{u}_h^{n+1}\big) = \mathcal{J}_h\big(\Pi_h(\boldsymbol{u}_h^n - \boldsymbol{w}_h^n)\big) \overset{(40)}{\leqslant} \mathcal{J}_h\big(\boldsymbol{u}_h^n - \boldsymbol{w}_h^n\big) \leqslant \mathcal{J}_h\big(\boldsymbol{u}_h^n\big) \quad \text{for all } n \in \mathbb{N}_0$$

and thus conclude $\mathcal{J}_h(u_h) \leqslant \mathcal{J}_h(u_h^0)$ by induction.

Step 3 (Algorithm B terminates): Arguing as for (30) in step 3 of the proof of Theorem 4, we conclude that $\mathcal{J}_h(\boldsymbol{w}_h^n) \to 0$ as $n \to \infty$. In particular, we have that $\mathcal{J}_h(\boldsymbol{w}_h^n) \leqslant \mathrm{tol}_h$ for some $n \in \mathbb{N}_0$ so that Algorithm B terminates. this concludes the proof of Theorem 10(i).

Step 4 (Existence of weakly convergent subsequence): Having proved Theorem 10(i), Lemma 14 yields that

$$\|\boldsymbol{u}_h\|_{\boldsymbol{H}^1(\Omega)}^2 \lesssim \mathcal{J}_h(\boldsymbol{u}_h) + \|\boldsymbol{u}_h\|_{\boldsymbol{L}^2(\Omega)}^2 \leqslant \mathcal{J}_h(\boldsymbol{u}_h^0) + \|\boldsymbol{u}_h\|_{\boldsymbol{L}^2(\Omega)}^2 \overset{(46)}{\leqslant} C_0 + |\Omega|.$$

Therefore, there exists a subsequence of $(\boldsymbol{u}_h)_h$ (which is not relabeled) and $\boldsymbol{u} \in \boldsymbol{H}^1(\Omega)$ such that $\boldsymbol{u}_h \rightharpoonup \boldsymbol{u}$ weakly in $\boldsymbol{H}^1(\Omega)$.

Step 5 (Weak limit $\boldsymbol{u}$ satisfies modulus constraint): Since $|\boldsymbol{u}_h(\mathbf{z})| = 1$ for all $\mathbf{z} \in \mathcal{N}_h$, we have that $\mathcal{I}_h(|\boldsymbol{u}_h|^2) = 1 \in \mathcal{S}^1(\mathcal{T}_h)$. Moreover, $|\boldsymbol{u}_h|^2 = \boldsymbol{u}_h \cdot \boldsymbol{u}_h \in [\mathcal{S}^2(\mathcal{T}_h)]^3$. With Lemma 17, we get on each element $T \in \mathcal{T}_h$ that

$$\begin{aligned}
\||\boldsymbol{u}_h|^2 - 1\|_{\boldsymbol{L}^2(T)} &= \|(1 - \mathcal{I}_h)|\boldsymbol{u}_h|^2\|_{\boldsymbol{L}^2(T)} \overset{(44)}{\lesssim} h^2 \|D^2 |\boldsymbol{u}_h|^2\|_{\boldsymbol{L}^2(T)} \overset{(45)}{\lesssim} h \|\nabla |\boldsymbol{u}_h|^2\|_{\boldsymbol{L}^2(T)} \\
&= 2h \|\boldsymbol{u}_h \cdot \nabla \boldsymbol{u}_h\|_{\boldsymbol{L}^2(T)} \leqslant 2h \|\boldsymbol{u}_h\|_{\boldsymbol{L}^\infty(\Omega)} \|\nabla \boldsymbol{u}_h\|_{\boldsymbol{L}^2(T)} \\
&= 2h \|\nabla \boldsymbol{u}_h\|_{\boldsymbol{L}^2(T)} \lesssim h \mathcal{J}_h(\boldsymbol{u}_h)^{1/2} \lesssim h \to 0.
\end{aligned}$$

Since $\boldsymbol{u}_h \rightharpoonup \boldsymbol{u}$ in $\boldsymbol{H}^1(\Omega)$ and $\boldsymbol{u}_h \to \boldsymbol{u}$ in $\boldsymbol{L}^2(\Omega)$, there holds (up to a subsequence which is not relabeled) that $\boldsymbol{u}_h \to \boldsymbol{u}$ almost everywhere in Ω. This implies $|\boldsymbol{u}| = 1$ a.e. in Ω and hence $\boldsymbol{u} \in \mathcal{M}$.

Step 6 ($\boldsymbol{u}$ satisfies the Euler–Lagrange equations (13) of $\mathcal{J}$): Let $n \in \mathbb{N}_0$ be the index such that $\boldsymbol{u}_h = \boldsymbol{u}_h^n$ is returned by Algorithm B. Defining $\boldsymbol{w}_h := \boldsymbol{w}_h^n$, for all $\boldsymbol{v}_h \in \mathcal{K}_h[\boldsymbol{u}_h]$, it follows that

$$\int_\Omega \nabla(\boldsymbol{u}_h - \boldsymbol{w}_h) : \nabla \boldsymbol{v}_h + \mathcal{I}_h\big(g_\gamma(\boldsymbol{u}_h - \boldsymbol{w}_h, \boldsymbol{v}_h)\big)\,\mathrm{d}x \overset{(34)}{=} a_h(\boldsymbol{u}_h - \boldsymbol{w}_h, \boldsymbol{v}_h) \overset{(36)}{=} 0.$$

For $\boldsymbol{\varphi} \in \boldsymbol{C}^\infty(\Omega)^3$, define $\overline{\boldsymbol{v}} := \boldsymbol{\varphi} \times \boldsymbol{u}_h \in \boldsymbol{H}^1(\Omega) \cap \boldsymbol{C}(\Omega)$ and $\overline{\boldsymbol{v}}_h := \mathcal{I}_h(\overline{\boldsymbol{v}}) \in \boldsymbol{\mathcal{S}}^1(\mathcal{T}_h)$. Then, the last equation gives

$$\begin{aligned}
&\int_\Omega \Big[\nabla \boldsymbol{u}_h : \nabla \overline{\boldsymbol{v}} + \mathcal{I}_h\big(g_\gamma(\boldsymbol{u}_h, \overline{\boldsymbol{v}})\big)\Big] \,\mathrm{d}x \\
&\qquad = \int_\Omega \Big[\nabla(\boldsymbol{u}_h - \boldsymbol{w}_h) : \nabla(\overline{\boldsymbol{v}} - \overline{\boldsymbol{v}}_h) + \mathcal{I}_h\big(g_\gamma(\boldsymbol{u}_h - \boldsymbol{w}_h, \overline{\boldsymbol{v}} - \overline{\boldsymbol{v}}_h)\big)\Big] \,\mathrm{d}x \qquad (47) \\
&\qquad\quad + \int_\Omega \Big[\nabla \boldsymbol{w}_h : \nabla \overline{\boldsymbol{v}} + \mathcal{I}_h\big(g_\gamma(\boldsymbol{w}_h, \overline{\boldsymbol{v}})\big)\Big] \,\mathrm{d}x.
\end{aligned}$$

It remains to show that, as $h \to 0$, the left-hand side and right-hand side of (47) converge to the left-hand side and right-hand side of (13), respectively. Then, Lemma 3 yields that $\boldsymbol{u}$ is a weak solution of the Euler–Lagrange equations (11).

Step 6.1 (Left-hand side of (47) converges to left-hand side of (13) as $h \to 0$): We notice that

$$\int_\Omega \Big[\nabla \boldsymbol{u}_h : \nabla \overline{\boldsymbol{v}} + \mathcal{I}_h\big(g_\gamma(\boldsymbol{u}_h, \overline{\boldsymbol{v}})\big)\Big] \,\mathrm{d}x = \int_\Omega \Big[(\boldsymbol{u}_h \times \nabla \boldsymbol{u}_h) : \nabla \boldsymbol{\varphi} + \mathcal{I}_h\big(g_\gamma(\boldsymbol{u}_h, \boldsymbol{\varphi} \times \boldsymbol{u}_h)\big)\Big] \,\mathrm{d}x.$$

Recall that $\nabla \boldsymbol{u} : \nabla(\boldsymbol{\varphi} \times \boldsymbol{u}) = \boldsymbol{u} \times \nabla \boldsymbol{u} : \nabla \boldsymbol{\varphi}$. Since $\boldsymbol{u}_h \rightharpoonup \boldsymbol{u}$ in $\boldsymbol{H}^1(\Omega)$ implies $\boldsymbol{u}_h \to \boldsymbol{u}$ in $\boldsymbol{L}^2(\Omega)$ and $\nabla \boldsymbol{u}_h \rightharpoonup \nabla \boldsymbol{u}$ in $\boldsymbol{L}^2(\Omega)$, there holds for the first term that

$$\int_\Omega \nabla \boldsymbol{u}_h : \nabla \overline{\boldsymbol{v}} \,\mathrm{d}x = \int_\Omega \nabla \boldsymbol{u}_h : \nabla \boldsymbol{\varphi} \times \boldsymbol{u_h} \,\mathrm{d}x = \int_\Omega (\boldsymbol{u}_h \times \nabla \boldsymbol{u}_h) : \nabla \boldsymbol{\varphi} \,\mathrm{d}x \to \int_\Omega (\boldsymbol{u} \times \nabla \boldsymbol{u}) : \nabla \boldsymbol{\varphi} \,\mathrm{d}x.$$

For the mass lumping term, we first notice that

$$\int_\Omega \mathcal{I}_h(v^h) \,\mathrm{d}x = \int_\Omega v \,\mathrm{d}x + \int_\Omega v^h - v \,\mathrm{d}x + \int_\Omega (\mathcal{I}_h - 1) v^h \,\mathrm{d}x. \qquad (48)$$

Since $\|\boldsymbol{u}\|_{\boldsymbol{L}^\infty(\Omega)} = \|\boldsymbol{u}_h\|_{\boldsymbol{L}^\infty(\Omega)} = 1$, the choices $v^h := g_\gamma(\boldsymbol{u}_h, \boldsymbol{\varphi} \times \boldsymbol{u}_h)$ and $v := g_\gamma(\boldsymbol{u}, \boldsymbol{\varphi} \times \boldsymbol{u})$ yield that

$$\int_\Omega |v^h - v| \,\mathrm{d}x \lesssim (\|\boldsymbol{u}_h\|_{\boldsymbol{L}^\infty(\Omega)} + \|\boldsymbol{u}\|_{\boldsymbol{L}^\infty(\Omega)}) \|\boldsymbol{u}_h - \boldsymbol{u}\|_{\boldsymbol{L}^2(\Omega)} \|\boldsymbol{\varphi}\|_{\boldsymbol{L}^2(\Omega)} \lesssim \|\boldsymbol{u}_h - \boldsymbol{u}\|_{\boldsymbol{L}^2(\Omega)} \to 0.$$

The third term can be treated by the local approximation property of the nodal interpolation on each element, which yields that

$$\int_T \big|(\mathcal{I}_h - 1) v^h\big| \,\mathrm{d}x \leqslant |T|^{1/2} \|(\mathcal{I}_h - 1) v^h\|_{L^2(T)} \lesssim |T|^{1/2} h^2 \|D^2 v^h\|_{L^2(T)}.$$

Looking at $v^h = g_\gamma(\boldsymbol{u}_h, \boldsymbol{\varphi} \times \boldsymbol{u}_h)$ in detail, we see that, regardless of the sign of γ, this expression is a sum of products of the form $u_{h,i} u_{h,j} \varphi_k$ for some $i, j, k \in \{1, 2, 3\}$. For the Hessian of such a product, we obtain with $\boldsymbol{u}_h \in \mathcal{S}^1(\mathcal{T}_h)$, $\|\boldsymbol{u}_h\|_{\boldsymbol{L}^\infty(\Omega)} = 1$, and $\|\nabla \boldsymbol{u}_h\|_{\boldsymbol{L}^2(\Omega)} \lesssim 1$ that

$$\begin{aligned}
&\|D^2(u_{h,i} u_{h,j} \varphi_k)\|_{\boldsymbol{L}^2(T)} \\
&\qquad = \|u_{h,i} u_{h,j} D^2 \varphi_k + \varphi_k \nabla u_{h,i} \nabla u_{h,j}^\intercal + u_{h,j} \nabla u_{h,i} \nabla \varphi_k^\intercal + u_{h,i} \nabla u_{h,j} \nabla \varphi_k^\intercal\|_{\boldsymbol{L}^2(T)} \qquad (49) \\
&\qquad \lesssim \big(\|\boldsymbol{u}_h\|_{\boldsymbol{L}^2(\Omega)}^2 + \|\nabla \boldsymbol{u}_h\|_{\boldsymbol{L}^2(\Omega)}^2 + \|\boldsymbol{u}_h\|_{\boldsymbol{L}^2(\Omega)} \|\nabla \boldsymbol{u}_h\|_{\boldsymbol{L}^2(\Omega)}\big) \|\boldsymbol{\varphi}\|_{\boldsymbol{W}^{2,\infty}(T)} \lesssim 1.
\end{aligned}$$

Combining the estimates (48)–(49) and summing over all elements, we conclude that

$$\int_\Omega \mathcal{I}_h\big(g_\gamma(\boldsymbol{u}_h, \boldsymbol{\varphi} \times \boldsymbol{u}_h)\big) \,\mathrm{d}x \quad \xrightarrow{h \to 0} \quad \int_\Omega g_\gamma(\boldsymbol{u}, \boldsymbol{\varphi} \times \boldsymbol{u}) \,\mathrm{d}x. \qquad (50)$$

Step 6.2 (Right-hand side of (47) vanishes as $h \to 0$): We first observe that there holds $\overline{\boldsymbol{v}}(\boldsymbol{z}) - \overline{\boldsymbol{v}}_h(\boldsymbol{z}) = \boldsymbol{0}$ for all $z \in \mathcal{N}_h$ and, thus, the term $\mathcal{I}_h((u_{h,3} - w_{h,3})(\overline{v}_3 - \overline{v}_{h,3}))$

vanishes. Using the element-wise interpolation property of the nodal interpolation, we further see that

$$\|\nabla(\overline{\boldsymbol{v}} - \overline{\boldsymbol{v}}_h)\|_{\boldsymbol{L}^2(T)} = \|\nabla(I - \mathcal{I}_h)(\boldsymbol{\varphi} \times \boldsymbol{u}_h)\|_{\boldsymbol{L}^2(T)} \lesssim h \|D^2(\boldsymbol{\varphi} \times \boldsymbol{u}_h)\|_{\boldsymbol{L}^2(T)}.$$

An explicit computation in the spirit of (49) shows that $\|D^2(\boldsymbol{\varphi} \times \boldsymbol{u}_h)\|_{\boldsymbol{L}^2(T)} \lesssim \|\boldsymbol{\varphi}\|_{\boldsymbol{W}^{2,\infty}(T)}$. Thus, summing over all elements, we get that

$$\|\nabla(\overline{\boldsymbol{v}} - \overline{\boldsymbol{v}}_h)\|_{\boldsymbol{L}^2(\Omega)} \lesssim h \|\boldsymbol{\varphi}\|_{\boldsymbol{W}^{2,\infty}(\Omega)} \to 0 \quad \text{as } h \to 0.$$

Recall that

$$\|\nabla(\boldsymbol{u}_h - \boldsymbol{w}_h)\|_{\boldsymbol{L}^2(\Omega)} \leqslant \|\nabla \boldsymbol{u}_h\|_{\boldsymbol{L}^2(\Omega)} + \|\boldsymbol{w}_h\|_{\boldsymbol{L}^2(\Omega)} \lesssim \mathcal{J}_h(\boldsymbol{u}_h)^{1/2} + \mathcal{J}_h(\boldsymbol{w}_h)^{1/2} \leqslant C_0^{1/2} + \mathrm{tol}_h^{1/2}$$

is uniformly bounded. Therefore, we have that

$$\int_\Omega \big[\nabla(\boldsymbol{u}_h - \boldsymbol{w}_h) : \nabla(\overline{\boldsymbol{v}} - \overline{\boldsymbol{v}}_h) + \mathcal{I}_h(g_\gamma(\boldsymbol{u}_h - \boldsymbol{w}_h, \overline{\boldsymbol{v}} - \overline{\boldsymbol{v}}_h))\big] \,\mathrm{d}x \to 0 \quad \text{as } h \to 0.$$

With $\|\nabla \boldsymbol{w}_h\|^2_{\boldsymbol{L}^2(\Omega)} \leqslant \mathrm{tol}_h \to 0$ and the boundedness of $\overline{\boldsymbol{v}}$ in $\boldsymbol{H}^1(\Omega) \cap \boldsymbol{C}(\Omega)$, the Hölder inequality proves

$$\int_\Omega \big[\nabla \boldsymbol{w}_h : \nabla \overline{\boldsymbol{v}} + \mathcal{I}_h\big(g_\gamma(\boldsymbol{w}_h, \overline{\boldsymbol{v}})\big)\big] \,\mathrm{d}x \lesssim \mathcal{J}_h(\boldsymbol{w}_h) \lesssim \mathrm{tol}_h \to 0.$$

Step 7 (Proof of energy estimate (38)): Analogously to (50), it follows that

$$\int_\Omega \mathcal{I}_h\big(g_\gamma(\boldsymbol{u}_h, \boldsymbol{u}_h)\big) \,\mathrm{d}x \quad \xrightarrow{h \to 0} \quad \int_\Omega g_\gamma(\boldsymbol{u}, \boldsymbol{u}) \,\mathrm{d}x.$$

Therefore, it follows that

$$\begin{aligned} \liminf_{h \to 0} \mathcal{J}_h(\boldsymbol{u}_h) &= \liminf_{h \to 0} \int_\Omega \nabla \boldsymbol{u}_h : \nabla \boldsymbol{u}_h \,\mathrm{d}x + \lim_{h \to 0} \int_\Omega \mathcal{I}_h\big(g_\gamma(\boldsymbol{u}_h, \boldsymbol{u}_h)\big) \,\mathrm{d}x \\ &\geqslant \int_\Omega \big[\nabla \boldsymbol{u} : \nabla \boldsymbol{u} + g_\gamma(\boldsymbol{u}, \boldsymbol{u})\big] \,\mathrm{d}x = \mathcal{J}(\boldsymbol{u}). \end{aligned}$$

This concludes the proof. □

Remark 18. *We note that $\kappa = 0$ is only exploited in the crucial step 2 of the preceding proof of Theorem 10, where Lemma 15 is used. All other steps are valid for arbitrary $\kappa \in \mathbb{R}$.*

4.3. Necessity of nodal projection. Note that using mass lumping for the discrete energy is necessary in order to prove Lemma 15. Without the nodal projection, energy decrease for the g_γ-term (see (43)) does not hold true in general. In particular, for $\gamma \geqslant 0$, consider the unit triangle $T := \mathrm{conv}\{(0,0), (1,0), (0,1)\}$, where conv denotes the convex hull. An explicit computation with $\varepsilon > 0$ and $\delta := \sqrt{2 - \varepsilon^2/2}$ shows that the discrete function

$$\boldsymbol{v}_h := \begin{pmatrix} \delta \\ \delta \\ -\varepsilon \end{pmatrix} (1 - x - y) + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} x + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} y \ \in \mathcal{S}^1(T)$$

satisfies $|\boldsymbol{v}_h(0,0)| = 2$ and $\|\Pi_h \boldsymbol{v}_h \cdot \mathbf{e}_3\|^2_{L^2(T)} > \|\boldsymbol{v}_h \cdot \mathbf{e}_3\|^2_{L^2(T)}$ for $\varepsilon < 4/3$, since

$$\|\boldsymbol{v}_h \cdot \mathbf{e}_3\|^2_{L^2(T)} = \|(1+\varepsilon)x + (1+\varepsilon)y - \varepsilon\|^2_{L^2(T)} = \frac{1}{48}\big(12 - 8\varepsilon + 4\varepsilon^2\big),$$

$$\|\Pi_h \boldsymbol{v}_h \cdot \mathbf{e}_3\|^2_{L^2(T)} = \|(1+\varepsilon/2)x + (1+\varepsilon/2)y - \varepsilon/2\|^2_{L^2(T)} = \frac{1}{48}\big(12 - 4\varepsilon + \varepsilon^2\big).$$

For $\gamma < 0$, a similar counterexample can be constructed.

4.4. Increase in helical energy for $\kappa \neq 0$. Since, analytically, it is not straightforward to either prove or to disprove energy decrease in every step of Algorithm B in the case $\kappa \neq 0$ (see (40)), we seek to decide on that matter numerically. To this end, we look at the mesh $\mathcal{T}_h = \{T\}$ consisting only of the unit triangle $T = \operatorname{conv}\{(0,0),(1,0),(0,1)\}$. We stress that $\mathcal{T}_h$ satisfies the angle condition (37).

On this mesh, random vectors $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ with $|\mathbf{a}|, |\mathbf{b}|, |\mathbf{c}| > 1$ are generated to define

$$\boldsymbol{v}_h(x,y) := (1-x-y)\,\mathbf{a} + x\,\mathbf{b} + y\,\mathbf{c} \in \mathcal{S}^1(\mathcal{T}_h).$$

Subsequently, the energy difference $\mathcal{J}_h(\boldsymbol{v}_h) - \mathcal{J}_h(\Pi_h \boldsymbol{v}_h)$ is computed with Mathematica [Wol]. We found that, e.g., the values

$$\mathbf{a} := \begin{pmatrix} 0.44353334 \\ 0.86741656 \\ 0.22558999 \end{pmatrix}, \quad \mathbf{b} := \begin{pmatrix} 0.46138525 \\ 0.63580881 \\ 0.61893662 \end{pmatrix}, \quad \mathbf{c} := \begin{pmatrix} 0.5304891 \\ 0.66534908 \\ -0.52526736 \end{pmatrix},$$

with $|\mathbf{a}|, |\mathbf{b}|, |\mathbf{c}| > 1 + 5 \cdot 10^{-6}$ lead to an energy increase $\mathcal{J}_h(\Pi_h \boldsymbol{v}_h) \approx \mathcal{J}_h(\boldsymbol{v}_h) + 9.27 \cdot 10^{-6}$.

We stress that similar results can be found for different shapes and scalings of the triangle T, although these examples are scarce and the energy increase is small. This suggests that a generalization of the energy decrease (40) to $\kappa \neq 0$ cannot be true in general. However, since Algorithm B appears to converge also for $\kappa \neq 0$, we expect that this is a shortcoming of our analysis which potentially can be overcome.

4.5. Numerical experiment on reduced 3D model. To test our algorithm in the case of $\kappa \neq 0$ (which is thus not covered by Theorem 10), we reproduce an experiment from [HPP+19, Section 4.2]. This work considers the three dimensional micromagnetic model

$$\mathcal{E}_{3\mathrm{D}}(\boldsymbol{u}) = \int_\Omega A\Big[|\nabla \boldsymbol{u}|^2 + D\,\boldsymbol{\pi}[\boldsymbol{u}] \cdot \boldsymbol{u} + K|\boldsymbol{u} \times \mathbf{e}_3|^2 + \frac{\mu_0 M_s^2}{2} |\boldsymbol{h}_{\mathrm{d}}[\boldsymbol{u}]|^2\Big]\,\mathrm{d}x \tag{51}$$

with the interfacial Dzyaloshinskii–Moriya interaction (DMI)

$$\boldsymbol{\pi}[\boldsymbol{u}] = \mathbf{e}_1 \times \partial_2 \boldsymbol{u} - \mathbf{e}_2 \times \partial_1 \boldsymbol{u} = \begin{pmatrix} -\partial_1 u_3 \\ -\partial_2 u_3 \\ \partial_1 u_1 + \partial_2 u_2 \end{pmatrix}. \tag{52}$$

While [HPP+19] employs the so-called Landau–Lifshitz–Gilbert equation, which models the time-dependent evolution of the magnetization (as a generalized gradient flow for $\mathcal{E}_{3\mathrm{D}}$), we compute the steady state of the system by energy minimization. Furthermore, we consider the anti-symmetric exchange term $\operatorname{curl}(\boldsymbol{u}) \cdot \boldsymbol{u}$ instead of interfacial DMI $\boldsymbol{\pi}[\boldsymbol{u}] \cdot \boldsymbol{u}$; see Remark 19 below. The material parameters are those of cobalt, i.e., $A = 1.5 \cdot 10^{-11}\,\mathrm{J\,m^{-1}}$, $K = 8 \cdot 10^5\,\mathrm{J\,m^{-3}}$, and $M_s = 5.8 \cdot 10^5\,\mathrm{A\,m^{-1}}$. The strength of the anti-symmetric exchange term is varied in the range $D = 0, 1, \ldots, 8 \cdot 10^{-3}\,\mathrm{J\,m^{-2}}$. The computational domain Ω is a disk of diameter 80 nm within the plane spanned by $\mathbf{e}_1$ and $\mathbf{e}_2$.

To obtain a two-dimensional model that fits in the framework of the present paper, we first non-dimensionalize the model in the sense of, e.g., [Rug16]: With the exchange length $\ell_{\mathrm{ex}} = \sqrt{2A/(\mu_0 M_s^2)}$ and the vacuum permeability $\mu_0 = 4\pi \cdot 10^{-7}\,\mathrm{H\,m^{-1}}$, we rescale $x \to$

x/ℓ_{ex}. Together with $\kappa = D/(\mu_0 M_s^2 \ell_{\text{ex}})$, passing to the thin-film limit as in [DDPR22] yields the two-dimensional energy

$$\mathcal{E}_{\text{2D}}(\boldsymbol{u}) = \int_\Omega \left[\frac{1}{2}|\nabla \boldsymbol{u}|^2 + \kappa \operatorname{curl} \boldsymbol{u} \cdot \boldsymbol{u} + \frac{K}{\mu_0 M_s^2}|\boldsymbol{u} \times \mathbf{e}_3|^2 + \frac{1}{2}|\boldsymbol{u} \cdot \mathbf{e}_3|^2\right] \mathrm{d}x. \tag{53}$$

Using (8) to summarize the anisotropy terms and rewriting the first two terms as helical derivative, this energy reads as (5) with $\kappa = D/(\mu_0 M_s^2 \ell_{\text{ex}})$ and $\gamma = 1 - 2K/(\mu_0 M_s^2) - \kappa^2$.

We perform two different experiments that differ only by the initial values

$$\boldsymbol{u}^0 = \mathbf{e}_3, \tag{54}$$

$$\boldsymbol{u}^0 = \begin{cases} -\mathbf{e}_3 & r < 15\,\text{nm} - \varepsilon, \\ \mathbf{e}_3 & r > 15\,\text{nm} + \varepsilon, \end{cases} \tag{55}$$

with $\varepsilon > 0$ small and a smooth transition layer such that $\boldsymbol{u}^0 \in \mathcal{M}$ in both cases, which represent a constant and a skyrmion-like initial magnetization, respectively. The results for (54) are shown in Figure 2 and those for (55) in Figure 3. For the constant initial magnetization (54), the magnetization stays essentially uniform up to $D = 6$; for $D = 7, 8$, horseshoe-like magnetic domains are visible. For the skyrmion-like initial magnetization (55), the magnetization relaxes to a uniform magnetization for $D = 0, 1, 2$, but forms a skyrmion for $D = 3, \dots, 6$ and a so-called target skyrmion for $D = 8$. This is in total agreement with [HPP+19]. However, for the skyrmion-like initial magnetization and $D = 7$ we get a horseshoe-like state, whereas [HPP+19] gets a target skyrmion. This may be explained by the heavy dependence of the bifurcation point between simple skyrmion and target skyrmion on the relaxation method as suggested by numerical evidence in [FPPR22], as well as the effects explained in Remark 19 below. Since our algorithm is designed around physical energy minimization without the need for user-supplied parameters, we cannot control the relaxation process in our algorithm.

Remark 19. *We note that considering the two-dimensional curl*

$$\operatorname{curl}(\boldsymbol{u}) = \mathbf{e}_1 \times \partial_1 \boldsymbol{u} + \mathbf{e}_2 \times \partial_2 \boldsymbol{u} = \begin{pmatrix} \partial_2 u_3 \\ -\partial_1 u_3 \\ \partial_1 u_2 - \partial_2 u_1 \end{pmatrix}$$

instead of the DMI term $\boldsymbol{\pi}[\boldsymbol{u}]$ of the three dimensional model does not have any effects on the topological structure of the magnetization. However, it changes the type of the domain walls: In [HPP+19], only Neel-type domain walls occur, whereas our model produces only Bloch-type walls.

Remark 20. *Numerically, the initial configuration (55) is meta-stable for $D = 0$ in the thin-film case. In the thin-film 2D model of [DDPR22], the stray-field contribution is approximated by $|\boldsymbol{u} \cdot \mathbf{e}_3|^2$, which causes the stray field of $\boldsymbol{u}_h^0$ to be parallel to $\mathbf{e}_3$. Together with $\Delta \boldsymbol{u}_h^0$ and anisotropy also being parallel to $\mathbf{e}_3$, and $\mathbf{e}_3 \perp \mathcal{K}_h[\boldsymbol{u}_h^0]$, the initial tangential update $\boldsymbol{w}_h$ vanishes. In 3D, however, the stray field would not be pointing purely in the $\mathbf{e}_3$ direction because of boundary effects, thus exerting a small force perpendicular to $\mathbf{e}_3$ and perturbing that initial state. For this reason, we introduced a small divergent vector field for the disc part $r < 15\,\text{nm}$ to artificially perturb the initial state: $\boldsymbol{u}^0(x, y) = \Pi_h\big(\varepsilon \cdot (x, y, 0)^\intercal - \mathbf{e}_3\big)$ for some small $\varepsilon > 0$. For $D = 0$, this proved sufficient to prevent the system from being trapped in the initial meta-stable state. For $D \neq 0$, this modification is not necessary.*

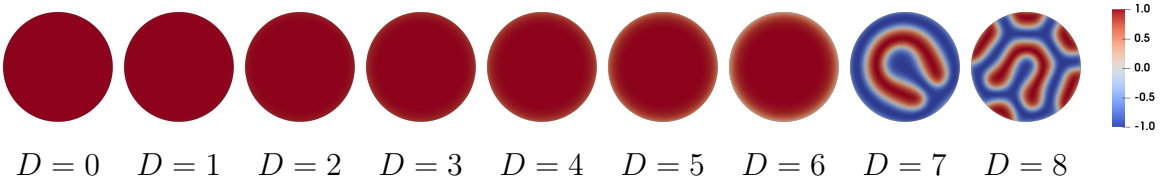

FIGURE 2. Simulation results for the model from Section 4.5 with different values of D and initial condition (54). The color represents the contribution of the magnetization in the $\mathbf{e}_3$-direction.

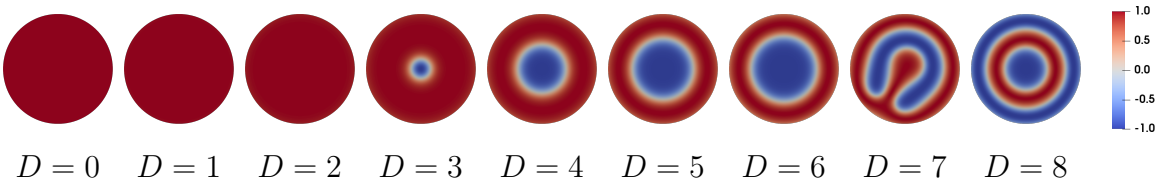

FIGURE 3. Simulation results for the model from Section 4.5 with different values of D and initial condition (55). The color represents the contribution of the magnetization in the $\mathbf{e}_3$-direction.

REFERENCES

[Alo97] F. Alouges. A new algorithm for computing liquid crystal stable configurations: the harmonic mapping case. *SIAM J. Numer. Anal.*, 34(5):1708–1726, 1997. DOI: 10.1137/S0036142994264249.

[Bar05] S. Bartels. Stability and convergence of finite-element approximation schemes for harmonic maps. *SIAM J. Numer. Anal.*, 43(1):220–238, 2005. DOI: 10.1137/040606594.

[BBF13] D. Boffi, F. Brezzi, and M. Fortin. *Mixed Finite Element Methods and Applications*. Springer Berlin Heidelberg, 2013. DOI: 10.1007/978-3-642-36519-5.

[BS08] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*. Springer, New York, third edition, 2008. DOI: 10.1007/978-0-387-75934-0. URL: https://doi.org/10.1007/978-0-387-75934-0.

[DDPR22] E. Davoli, G. Di Fratta, D. Praetorius, and M. Ruggeri. Micromagnetics of thin films in the presence of Dzyaloshinskii–Moriya interaction. *Math. Models Methods Appl. Sci.*, 32(05):911–939, 2022. DOI: 10.1142/s0218202522500208.

[Di 20] G. Di Fratta. Micromagnetics of curved thin films. *Zeitschrift für angewandte Mathematik und Physik*, 71(4), June 2020. DOI: 10.1007/s00033-020-01336-2.

[DIP20] G. Di Fratta, M. Innerberger, and D. Praetorius. Weak-strong uniqueness for the Landau–Lifshitz–Gilbert equation in micromagnetics. *Nonlinear Anal. Real World Appl.*, 55:103122, 2020. DOI: 10.1016/j.nonrwa.2020.103122.

[DMRS20] G. Di Fratta, C. B. Muratov, F. N. Rybakov, and V. V. Slastikov. Variational Principles of Micromagnetics Revisited. *SIAM Journal on Mathematical Analysis*, 52(4):3580–3599, January 2020. DOI: 10.1137/19m1261365.

[FPPR22] G. D. Fratta, C.-M. Pfeiler, D. Praetorius, and M. Ruggeri. The Mass-Lumped Midpoint Scheme for Computational Micromagnetics: Newton Linearization and Application to Magnetic Skyrmion Dynamics. *Comput. Meth. Appl. Math.*, published online, 2022. DOI: 10.1515/cmam-2022-0060.

[HPP+19] G. Hrkac, C.-M. Pfeiler, D. Praetorius, M. Ruggeri, A. Segatti, and B. Stiftner. Convergent tangent plane integrators for the simulation of chiral magnetic skyrmion dynamics. *Adv. Comput. Math.*, 45(3):1329–1368, 2019. DOI: 10.1007/s10444-019-09667-z.

[KPP+19] J. Kraus, C.-M. Pfeiler, D. Praetorius, M. Ruggeri, and B. Stiftner. Iterative solution and preconditioning for the tangent plane scheme in computational micromagnetics. *J. Comput. Phys.*, 398:108866, 27, 2019. DOI: 10.1016/j.jcp.2019.108866.

[Mel14] C. Melcher. Chiral skyrmions in the plane. *Proc. R. Soc. A*, 470:20140394, 2014. DOI: 10.1098/rspa.2014.0394.

[Rug16] M. Ruggeri. Coupling and numerical integration of the Landau–Lifshitz–Gilbert equation. PhD thesis, TU Wien, Institut of Analysis and Scientific Computing, 2016.

[SS86] Y. Saad and M. H. Schultz. GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems. *SIAM J. Sci. Statist. Comput.*, 7(3):856–869, 1986. DOI: 10.1137/0907058.

[Wol] Wolfram Research, Inc. Mathematica, Version 12.1. URL: https://www.wolfram.com/mathematica.

Dipartimento di Matematica e Applicazioni "R. Caccioppoli", Università degli Studi di Napoli "Federico II", Via Cintia, Complesso Monte S. Angelo, 80126 Napoli, Italy
Email address: Giovanni.DiFratta@unina.it

Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA
Email address: innerbergerm@hhmi.org (corresponding author)

TU Wien, Institute for Analysis and Scientific Computing, Wiedner Hauptstr. 8–10/E101/4, 1040 Wien, Austria
Email address: Dirk.Praetorius@asc.tuwien.ac.at

University of Bristol, School of Mathematics, Bristol BS8 1TW, United Kingdom
Email address: Valeriy.Slastikov@bristol.ac.uk