CONVERGENCE OF MULTISTEP PROJECTION METHODS FOR HARMONIC MAP HEAT FLOWS INTO GENERAL SURFACES

GENMING BAI, XINPING GUI, AND BUYANG LI

Abstract. We propose a high-order multistep projection method for the harmonic map heat flow from a bounded domain $\Omega \subset \mathbb{R}^d$ into a given n-dimensional smooth surface $\Gamma \subset \mathbb{R}^{\mathcal{N}+1}$. At every time level, an auxiliary numerical solution is solved by a multistep backward difference formula with a mass-lumping finite element method in space, and then projected onto the surface Γ. The projected numerical solution is used in the backward difference formula and the extrapolation of nonlinearities in the following time levels. Such projection algorithms are convenient in computation while still preserving the pointwise geometric constraint of the solution to stay on the target surface Γ. The convergence of some low-order single-step projection algorithms based on the backward Euler and Crank–Nicolson schemes have been studied in many articles for harmonic map heat flow and related models into the unit sphere, while the convergence of high-order multistep projection methods still remains open. In this article, we propose a high-order multistep projection method for harmonic map heat flows into a general smooth surface (not necessarily the unit sphere) and prove its optimal-order convergence by combining four techniques, i.e., decomposition of the Nevanlinna–Odeh multiplier technique into approximately normal and tangential components separately, an almost orthogonal relation between the error functions associated to the auxiliary and projected numerical solutions, pointwise L^∞ error estimates, the use of orthogonal projection onto the target surface Γ. Numerical results are provided to support the theoretical anlaysis on the convergence of the high-order multistep projection methods.

Key words: harmonic map heat flow, target surface, backward difference formula, projection, finite element, mass lumping, error estimates, almost orthogonality

1. Introduction

We consider the harmonic map heat flow from a bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, into an $\mathcal{N}$-dimensional closed smooth surface $\Gamma \subset \mathbb{R}^{\mathcal{N}+1}$, defined as the solution of the following initial-boundary value problem:

$$\partial_t \mathbf{m} = \Delta \mathbf{m} - A(\mathbf{m})(\nabla \mathbf{m}, \nabla \mathbf{m}) \qquad \text{in } \Omega \times (0, T], \tag{1.1}$$

$$\partial_\nu \mathbf{m} = \mathbf{0} \qquad \text{on } \partial\Omega \times (0, T], \tag{1.2}$$

$$\mathbf{m} = \mathbf{m}^0 \qquad \text{in } \Omega \times \{0\}, \tag{1.3}$$

where $A(\mathbf{m})$ is the second fundamental of the target surface Γ at point $\mathbf{m} \in \Gamma$, ν denotes the outward unit normal vector on $\partial\Omega$, and $\mathbf{m}^0$ is a given initial value such that $\mathbf{m}^0(x) \in \Gamma$ for $x \in \Omega$. It can be shown that the solution automatically satisfies the pointwise geometric constraint $\mathbf{m}(x, t) \in \Gamma$ for all $(x, t) \in \Omega \times (0, T]$.

The harmonic map heat flows were first introduced by Eells and Sampson [26] to construct the harmonic map in a homotopy class of any given smooth map. Applications of the harmonic map heat flows can be found in color image denoising [51,53] and nematic liquid crystal theory [16,43]. In the latter application, the harmonic map heat flow is coupled with the Navier-Stokes equations to describe the molecular orientation. Closely related models to the harmonic map heat flows include the Landau–Lifshitz equation of ferromagnetism dynamics [7, 46] and the geometric evolution equation of the normal vector in the mean curvature flow [39, 40]. These equations are all nonlinear with respect to the gradient of the solution, and their solutions share the same structure of staying on the unit sphere. Many different numerical methods

have been considered for approximating the harmonic map heat flow and the related models, including the Landau–Lifshitz equation and the nematic liquid crystal equations.

Optimal-order convergence has been proved for many algorithms which do not preserve the pointwise geometric constraint $\mathbf{m}(x,t) \in \Gamma$. For example, Girault and Guillen-Gonzalez [30] proved first-order convergence of mixed finite element approximations to a penalized nematic liquid crystals model. Gao [29] proved optimal-order convergence of the linearly implicit backward Euler (FEMs) for the Landau–Lifshitz equation. Akrivis, Feischl, Kovács and Lubich [2] proved optimal H^1-norm error estimates for the Landau–Lifshitz equation with high-order backward difference formulae (BDF) in time and a linearly implicit FEM in space. More convergence results of numerical approximations to the Landau–Lifshitz equation and nematic liquid crystal equations can be found in [10, 23, 28].

Many nonlinearly implicit numerical algorithms were proposed to preserve the pointwise geometric constraint for the harmonic map heat flows. For example, Bartels and Prohl [14] proposed a nonlinearly implicit finite-element method based on a reformulation of equation (1.1) and a reduced spatial integration. Bartels, Lubich and Prohl [13] proposed a constrained FEM by introducing an approximate discrete Lagrange multiplier into the variational weak formulation. Baňas, Prohl and Schätzle [15] constructed a fully discrete nonlinear FEM for the harmonic map heat flows into spheres of nonconstant radii. Gutiérrez-Santacreu and Restelli [36] considered a unified saddle-point FEM for both equation (1.1) and the Landau-Lifshitz equation. These nonlinearly implicit constraint-preserving algorithms were all proved to be convergent, whereas no convergence rates have been proved yet. In [31–34, 38] the authors proposed a family of manifold-valued finite element spaces by which high-order convergence can be shown for a large class of discrete elliptic variational problems including the stationary harmonic map into general surfaces. In [12], Bartels proposed an iterative finite element algorithm for harmonic map into general two-dimensional surfaces where the global convergence can be shown under a mild mesh condition using compactness arguments.

The linearly implicit projection methods are computationally as cheap as the classical linearly implicit methods and capable of preserving the pointwise geometric constraint, and therefore have attracted considerable attention from the numerical analysts in recent years. In this class of methods, at every time level, the numerical solution is first solved by a linearly implicit scheme and then projected onto the surface Γ before proceeding to the next time level. Alouges and Jaissson [6] proposed a normalized tangent plane method for approximating the Landau–Lifshitz equation, which was later generalized to a θ-method by Alouges [5] (based on the classical θ-scheme for solving the heat equation), and extended to second-order discretizations by Alouges et al. [8]. These methods were proved convergent without explicit convergence rates. More recently, error estimates of the projection methods were obtained in [11] and [9] for the semi-implicit Euler method and the Crank–Nicolson method, respectively, under the stepsize restriciton $\tau \sim h$. Under this condition, the error estimates are $O(\tau + h^{r+1}) = O(h)$ and $O(\tau^2 + h^2) = O(h^2)$ for the semi-implicit Euler FEM (with finite elements of degree $r \geqslant 2$) and Crank–Nicolson finite difference method, respectively. Gui, Li and Wang [35] proved the optimal-order convergence $O(\tau + h^{r+1})$ of a projection method under a less restrictive condition $h^{r+1} \lesssim \tau$, with the semi-implicit Euler method in time and a mass-lumping FEM using tensor-product elements of degree $r \geqslant 1$. All these methods are single-step methods based on the backward Euler and Crank–Nicoson schemes.

The convergence of a semi-projection method with two-step BDF for the Landau–Lifshitz equation was proved in [19], where "semi-projection method" means that the numerical solution is projected onto the unit sphere in the extrapolation of the nonlinear terms, but not projected onto the unit sphere in the two-step BDF for approximating the time derivative. The construction of an unconditionally stable projection method for the Landau–Lifshitz equation by considering fast solvers and large damping parameters was given in [18] without error estimates. As far as we know, the convergence of projection methods with second- and higher-order multistep time discretizations still remain open. The convergence of projection methods with a general target surface (rather than the unit sphere) also remains open.

In this article, we propose a high-order multistep projection method for the harmonic map heat flow into a general smooth surface $\Gamma \subset \mathbb{R}^{\mathcal{N}+1}$, with k-step BDF in time ($k = 2, \ldots, 5$) and

a tensor-product mass-lumping FEM of degree $r \geqslant 1$ in space, and prove the optimal-order convergence $O(\tau^k + h^{r+1})$ of the k-step projection method under a mild condition $h^{r+1} \lesssim \tau^k$.

Let $\frac{1}{\tau}\delta_0 \widetilde{\mathbf{m}}_h^{n-j} + \frac{1}{\tau}\sum_{j=1}^{k} \delta_j \mathbf{m}_h^{n-j}$ be the k-step BDF for approximating the time derivative $\partial_t \mathbf{m}(t_n)$ in a multistep projection method, where $\mathbf{m}_h^n$ and $\widetilde{\mathbf{m}}_h^n$ denote the projected numerical solution and the auxiliary numerical solution (before being projected onto the surface Γ), respectively. Compared with the single-step projection methods, the main difficulty in the error analysis for multistep projection methods is from the following term:

$$\Big(\frac{\delta_0}{\tau}\widetilde{\mathbf{e}}_h^n + \frac{1}{\tau}\sum_{j=1}^{k} \delta_j \mathbf{e}_h^{n-j}\Big)(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}) \tag{1.4}$$

where $\mathbf{e}_h^{n-1} := I_h \mathbf{m}(t_{n-1}) - \mathbf{m}_h^{n-1}$ and $\widetilde{\mathbf{e}}_h^{n-1} := I_h \mathbf{m}(t_{n-1}) - \widetilde{\mathbf{m}}_h^{n-1}$ denote the errors of the projected numerical solution and auxiliary numerical solution, respectively, and $\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}$ is the Nevanlinna–Odeh multiplier for the k-step BDF with some parameter $\theta_k \in [0, 1)$; see [1, 3, 4, 44, 45]. The mismatch between $\mathbf{e}_h^{n-j}$ and $\widetilde{\mathbf{e}}_h^{n-j}$ in (1.4) must be shown to have very small (high-order) influence on the Nevanlinna–Odeh multiplier technique. Our approach is to decompose $\widetilde{\mathbf{e}}_h^i$ into its approximately normal and tangential components, i.e.,

$$\widetilde{\mathbf{e}}_h^i = \eta^i + \mathbf{e}_h^i \quad \text{with} \quad \eta^j = \widetilde{\mathbf{e}}_h^j - \mathbf{e}_h^j,$$

and then rewrite (1.4) into the following four parts:

$$\begin{aligned}
&\Big(\delta_0 \widetilde{\mathbf{e}}_h^n + \sum_{j=1}^{k} \delta_j \mathbf{e}_h^{n-j}\Big) \cdot (\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}) \\
&= \Big(\delta_0 \eta^n + \sum_{j=0}^{k} \delta_j \mathbf{e}_h^{n-j}, (\eta^n - \theta_k \eta^{n-1}) + (\mathbf{e}_h^n - \theta_k \mathbf{e}_h^{n-1})\Big)_h \\
&= \big(\delta_0 \eta^n\big) \cdot \big(\eta^n - \theta_k \eta^{n-1}\big) + \Big(\sum_{j=0}^{k} \delta_j \mathbf{e}_h^{n-j}\Big) \cdot \big(\eta^n - \theta_k \eta^{n-1}\big) \\
&\qquad + \big(\delta_0 \eta^n\big) \cdot \big(\mathbf{e}_h^n - \theta_k \mathbf{e}_h^{n-1}\big) + \Big(\sum_{j=0}^{k} \delta_j \mathbf{e}_h^{n-j}\Big) \cdot \big(\mathbf{e}_h^n - \theta_k \mathbf{e}_h^{n-1}\big).
\end{aligned}$$

The first and fourth parts are estimated directly by using the Nevanlinna–Odeh multiplier technique, while the second and third parts (the cross-product terms) are estimated by utilizing an almost orthogonality relation for a general target surface Γ:

$$|\mathbf{e}_h^{n-j} \cdot \eta^{n-i}| \leqslant C\tau(|\mathbf{e}_h^{n-j}|^2 + |\widetilde{\mathbf{e}}_h^{n-i}|^2 + |\mathbf{e}_h^{n-i}|^2), \tag{1.5}$$

with an additional factor $C\tau$ to eliminate the factor τ^{-1} in (1.4). This additional factor $C\tau$ is based on the proof of the pointwise error estimate

$$\|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^\infty(\Omega)} \leqslant \tau,$$

which guarantees that $\mathbf{e}_h^{n-j}$ is approximately tangential to the target surface Γ at point $\mathbf{m}(t_{n-j}) \in \Gamma$ up to a quantity of $O(\tau|\mathbf{e}_h^{n-j}|)$, provided that we define the projection onto the surface Γ through distance projection (orthogonal projection). The pointwise error estimate and the almost orthogonality relation in (1.5) allow us to apply the Nevanlinna–Odeh multiplier technique [1, 3, 4, 44, 45] in the presence of a projection stage and therefore is essential in proving optimal-order convergence of the multistep projection methods. The almost orthogonality and the Nevanlinna–Odeh multiplier estimates are first established at the finite element nodes and then utilized in the error analysis through the mass lumping technique.

Since the projection method and its analysis proposed in this article is quite general, it may also be applicable to other geometric partial differential equations with constrained target manifolds. For example, the similar projection method may be applied to the heat flow of Yang–Mills equation on $\Omega \subset \mathbb{R}^d$ with image constrained on the Lie algebra $\mathfrak{so}(d)$ [48, 52], as well as the wave map into general smooth surfaces [49].

The rest of this article is organized as follows. In Section 2, we present the notations, the numerical algorithm, and the main theorem on the convergence of the numerical solutions.

The proof of the main theorem is presented in Section 3. Numerical results are presented in Section 4 to support the theoretical analysis. Concluding remarks and extension to triangular meshes are presented in Section 5.

2. Main results

In this section, we introduce the basic notations and the geometry of the target surface $\Gamma \subset \mathbb{R}^{\mathcal{N}+1}$, including the second fundamental form on the surface and the distance projection onto the surface, which will be used in the definition of the multistep projection method for harmonic map heat flows into a general target surface. Then we present the main theorem of this article on the optimal-order convergence of the multistep projection method.

2.1. Notations

For $1 \leqslant p \leqslant +\infty$ and integer $j \geqslant 0$, we denote by $\mathbf{L}^p := (L^p)^{\mathcal{N}+1}$ and $\mathbf{W}^{j,p} := (W^{j,p})^{\mathcal{N}+1}$ the $(\mathcal{N}+1)$-dimensional vector-valued Lebesgue and Sobolev spaces defined on Ω; see [27, page 261]. The norm and semi-norm on $\mathbf{W}^{j,p}$ are denoted by $\|\cdot\|_{W^{j,p}}$ and $|\cdot|_{W^{j,p}}$, respectively. The abbreviations $H^j := W^{j,2}$ and $\mathbf{H}^j := \mathbf{W}^{j,2}$ are used as usual. The space of uniformly continuous functions on $\overline{\Omega}$ is denoted by $C(\overline{\Omega})$.

Throughout the article, we use C to denote a generic positive constant, and ε a small generic positive constant, which may assume different values at different occurrences but are always independent of τ, h, and N. For brevity, we denote by "$f \lesssim g$" the statement "$f \leqslant Cg$ for some constant C".

Let Ω be a rectangular domain and consider the H^1-conforming tensor-product finite element space on a rectangular mesh, i.e.,

$$\mathbf{S}_h^r := \{\mathbf{v} \in \mathbf{H}^1 : \mathbf{v}|_K \in (Q_r)^{\mathcal{N}+1} \,\forall\, K \in \mathcal{K}\}, \tag{2.1}$$

where $\mathcal{K}$ denotes the set of cuboids ($d = 3$) or rectangles ($d = 2$) in a quasi-uniform rectangular partition of the domain Ω with mesh size $h := \max_{K\in\mathcal{K}} \operatorname{diam}(K)$, and Q_r denotes the tensor-product polynomial space composed of polynomials of degree up to $r \geqslant 1$ in each variable. By using the piecewise Lagrange interpolation operator I_h, we can define the discrete inner product $(\cdot,\cdot)_h$ on $\mathbf{S}_h^r$ by

$$(\mathbf{f},\mathbf{g})_h := \int_\Omega I_h(\mathbf{f}\cdot\mathbf{g})\mathrm{d}x = \sum_{K\in\mathcal{K}} \int_K I_h(\mathbf{f}\cdot\mathbf{g})\mathrm{d}x \quad \forall\, \mathbf{f},\mathbf{g} \in \mathbf{S}_h^r,$$

where the interpolation nodes are chosen as Gauss–Lobatto points on each element K. Since these quadrature points coincide with the finite element nodes, the mass matrix is thus lumped; see [35, Section 2.1] and [42, Section 2] for more details. Similarly, the discrete L^p norm is defined as

$$\|\mathbf{u_h}\|_{L_h^p} := \begin{cases} \left(\displaystyle\int_\Omega I_h(|\mathbf{u}_h|^p)\mathrm{d}x\right)^{\frac{1}{p}} & \text{for } \mathbf{u}_h \in \mathbf{S}_h^r \text{ and } 1 \leqslant p < +\infty, \\ \displaystyle\max_{x\in\mathrm{Nodes}(\mathbf{S}_h^r)} |\mathbf{u}_h(x)| & \text{for } \mathbf{u}_h \in \mathbf{S}_h^r \text{ and } p = +\infty, \end{cases}$$

where $\mathrm{Nodes}(\mathbf{S}_h^r) \subseteq \bar{\Omega}$ denotes the set of all finite element nodes. which is equivalent to the conventional L^p norm for finite element functions, i.e.,

$$\|\mathbf{u_h}\|_{L^p} \lesssim \|\mathbf{u_h}\|_{L_h^p} \lesssim \|\mathbf{u_h}\|_{L^p} \quad \forall\, \mathbf{u}_h \in \mathbf{S}_h^r, \quad 1 \leqslant p \leqslant +\infty. \tag{2.2}$$

A proof of this equivalence relation can be found in [35, Lemma 2.1].

2.2. Geometry of the target surface

Given a closed oritentable surface $\Gamma \subset \mathbb{R}^{\mathcal{N}+1}$, there exists a smooth unit outward normal vector field $\boldsymbol{\nu} : \Gamma \to \mathbb{S}^{\mathcal{N}}$. The second fundamental form at point $p \in \Gamma$ is a bilinear form

defined by

$$A(p)(X,Y) := -\langle X, \mathrm{d}\boldsymbol{\nu}(p)(Y)\rangle\, \boldsymbol{\nu}(p) = -\langle Y, \mathrm{d}\boldsymbol{\nu}(p)(X)\rangle\, \boldsymbol{\nu}(p) \quad \forall\, X, Y \in T\Gamma_p, \tag{2.3}$$

where $T\Gamma_p$ denotes the space of vectors in $\mathbb{R}^{\mathcal{N}+1}$ tangential to the surface at $p \in \Gamma$, and $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product of $\mathbb{R}^{\mathcal{N}+1}$. The last equality in (2.3) is due to the self-adjointness of $\mathrm{d}\boldsymbol{\nu}(p)$ with respect to the inner product $\langle\cdot,\cdot\rangle$.

We can extend the second fundamental form to vectors $X, Y \in \mathbb{R}^{\mathcal{N}+1}$ which are not necessarily in the tangent space $T\Gamma_p$, by first projecting X, Y onto the tangent space $T\Gamma_p$ and then substituting them into the second fundamental form, i.e.,

$$A(p)(X,Y) := A(p)\big(X - (X\cdot\boldsymbol{\nu}(p))\boldsymbol{\nu}(p), Y - (Y\cdot\boldsymbol{\nu}(p))\boldsymbol{\nu}(p)\big) \quad \forall\, X, Y \in \mathbb{R}^{\mathcal{N}+1}. \tag{2.4}$$

Correspondingly, we can define a matrix $\mathbf{H}(p) \in \mathbb{R}^{(\mathcal{N}+1)\times(\mathcal{N}+1)}$ associated to the second fundamental form by

$$\mathbf{H}(p)X\cdot Y = -\langle X, \mathrm{d}\boldsymbol{\nu}(p)(Y)\,\rangle \quad \forall\, X, Y \in T\Gamma_p.$$

If X and Y are matrices with rows X_i and Y_i, $i = 1,\dots,d$, respectively, then we define

$$A(p)(X,Y) := \sum_{i=1}^{d} A(p)(X_i, Y_i). \tag{2.5}$$

For any column vector field $\mathbf{u} = (u_1, \cdots, u_{\mathcal{N}+1})^\top$, we define $\nabla\mathbf{u} = (\nabla u_1, \cdots, \nabla u_{\mathcal{N}+1})$ with the gradient of the components as column vectors. According to the definition in (2.5), we have

$$A(p)(\nabla\mathbf{u}, \nabla\mathbf{v}) = \sum_{i=1}^{d} A(p)(\partial_i\mathbf{u}, \partial_i\mathbf{v}). \tag{2.6}$$

For the simplicity of notation, we define the interpolated second fundamental form

$$A_h(\mathbf{u})(X,Y) := -\langle X, I_h(\mathbf{H}\circ\mathbf{u})Y\rangle I_h(\boldsymbol{\nu}\circ\mathbf{u}) \tag{2.7}$$

for $\mathbf{u} \in C(\overline{\Omega};\Gamma)$ and $X, Y \in C(\overline{\Omega};\mathbb{R}^{\mathcal{N}+1})$. If X, Y in (2.7) are two $\mathbb{R}^{d\times(\mathcal{N}+1)}$-valued functions, we define

$$A_h(\mathbf{u})(X,Y) := \sum_{i=1}^{d} A_h(\mathbf{u})(X_i, Y_i). \tag{2.8}$$

In the δ neighbourhood U_δ of Γ, with a sufficiently small δ, the distance function $d : U_\delta \to \mathbb{R}$ defined by $d(p) = \inf_{q\in\Gamma}|p-q|$ is well-defined and smooth. Correspondingly, the distance projection $\mathbf{a} : U_\delta \to \Gamma$ defined by

$$\mathbf{a}(p) = p - d(p)\boldsymbol{\nu}(p) \tag{2.9}$$

is also smooth. The gradient of $\mathbf{a}$ at point $p \in U_\delta$ has the following expression:

$$\nabla\mathbf{a}(p) = \mathbf{I} - d(p)\mathbf{H}(p) - \boldsymbol{\nu}(p)\otimes\boldsymbol{\nu}(p). \tag{2.10}$$

Hence, if we are given a function $\mathbf{u} \in W^{1,\infty}(\Omega; U_\delta)$, then $\mathbf{a}(\mathbf{u}) \in W^{1,\infty}(\Omega)^{\mathcal{N}+1}$ and therefore

$$\|\nabla[\mathbf{a}(\mathbf{u})]\|_{L^\infty} = \|(\nabla\mathbf{a})(\mathbf{u})\cdot\nabla\mathbf{u}\|_{L^\infty} \lesssim \|\nabla\mathbf{u}\|_{L^\infty}. \tag{2.11}$$

Moreover, for two measurable functions $\mathbf{u}, \mathbf{v} : \Omega \to U_\delta \subset \mathbb{R}^{\mathcal{N}+1}$, we can use the local Lipschitz continuity to obtain the following pointwise estimate:

$$|\mathbf{a}(\mathbf{u}(x)) - \mathbf{a}(\mathbf{v}(x))| \lesssim |\mathbf{u}(x) - \mathbf{v}(x)| \quad \forall\, x \in \Omega. \tag{2.12}$$

The stability results in (2.11) and (2.12) also hold for $\boldsymbol{\nu}$ and $\mathbf{H}$.

2.3. The numerical scheme and its error estimates

Let $t_n = n\tau\,(n = 0,\cdots,N)$ be a uniform partition of the interval $[0,T]$ with stepsize $\tau = T/N$. For $k = 2,\dots,5$, we denote by δ_j and γ_j the coefficients of the following polynomials:

$$\delta(\zeta) = \sum_{j=1}^{k}\frac{1}{j}(1-\zeta)^j = \sum_{j=0}^{k}\delta_j\zeta^j \quad \text{and} \quad \gamma(\zeta) = \frac{1}{\zeta}[1-(1-\zeta)^k] = \sum_{j=0}^{k-1}\gamma_j\zeta^j,$$

where $\delta(\zeta)$ and $\gamma(\zeta)$ are the generating polynomials of the linearly implicit k-step BDF; see [37]. For simplicity, we use the following notations to denote the k-step extrapolations of the exact and numerical solutions, respectively:

$$\widehat{u}(t_n) = \sum_{j=0}^{k-1} \gamma_j u(t_{n-j-1}) \quad \text{and} \quad \widehat{u}_h^n = \sum_{j=0}^{k-1} \gamma_j u_h^{n-j-1}.$$

For the k-step BDF, the following consistency estimates are well known.

Lemma 2.1. *Let X be any given Banach space. Then the following two estimates hold:*

$$\Big\| \sum_{j=0}^{k-1} \gamma_j u(t_{n-j-1}) - u(t_n) \Big\|_X \lesssim \tau^k \|u\|_{C^k([0,T];X)} \qquad \forall\, u \in C^k([0,T];X), \tag{2.13}$$

$$\Big\| \sum_{j=0}^{k} \delta_j u(t_{n-j}) - \partial_t u \Big\|_X \lesssim \tau^k \|u\|_{C^{k+1}([0,T];X)} \qquad \forall\, u \in C^{k+1}([0,T];X). \tag{2.14}$$

We assume that the numerical solutions $\mathbf{m}_h^j = \widetilde{\mathbf{m}}_h^j$ at the starting time levels $j = 0, \ldots, k-1$ are given. Then we determine $\mathbf{m}_h^n$, $k \leqslant n \leqslant N$, by first solving an auxiliary numerical solution $\widetilde{\mathbf{m}}_h^n \in \mathbf{S}_h^r$ from the weak formulation

$$\Big(\frac{\delta_0}{\tau} \widetilde{\mathbf{m}}_h^n + \frac{1}{\tau} \sum_{j=1}^{k} \delta_j \mathbf{m}_h^{n-j}, \mathbf{v}_h \Big)_h - (\Delta_h \widetilde{\mathbf{m}}_h^n, \mathbf{v}_h) = -\Big(A_h(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)(\nabla \widehat{\mathbf{m}}_h^n, \nabla \widehat{\mathbf{m}}_h^n), \mathbf{v}_h \Big) \tag{2.15}$$

where the discrete Laplacian operator $\Delta_h : \mathbf{S}_h^r \to \mathbf{S}_h^r$ is defined by

$$(\Delta_h \mathbf{v}_h, \mathbf{w}_h) := -(\nabla \mathbf{v}_h, \nabla \mathbf{w}_h) \quad \forall\, \mathbf{v}_h, \mathbf{w}_h \in \mathbf{S}_h^r,$$

and then projecting $\widetilde{\mathbf{m}}_h^n$ onto Γ at the finite element nodes, i.e.,

$$\mathbf{m}_h^n = I_h \left(\mathbf{a} \circ \widetilde{\mathbf{m}}_h^n \right) \in \mathbf{S}_h^r, \tag{2.16}$$

where $\mathbf{a}$ is the distance projection defined in (2.9).

Remark 2.1. Since the nonlinear terms are treated fully explicitly, the unique solvability of (2.15) is clear. Then (2.16) is well defined provided the projection $\mathbf{a}$ is well defined. This is the case when Γ encloses a convex domain. For a general smooth surface which does not enclose a convex domain, (2.16) is well defined if $\widetilde{\mathbf{m}}_h^n$ is in a small neighbourhood of the surface Γ. This can be guaranteed by the error estimates (3.43), since at each finite element node it holds that

$$\|\widetilde{\mathbf{m}}_h^n - \mathbf{a} \circ \widetilde{\mathbf{m}}_h^n\|_{L^\infty} \leqslant \|\widetilde{\mathbf{m}}_h^n - I_h \mathbf{m}(t_n)\|_{L^\infty} = \|\widetilde{\mathbf{e}}_h^n\|_{L^\infty} \leqslant \tau,$$

where $\widetilde{\mathbf{e}}_h^n := I_h \mathbf{m}(t_n) - \widetilde{\mathbf{m}}_h^n$. Hence we can choose sufficiently small τ such that all of the finite element nodal values lie in U_δ ensuring the nodal projection is well-defined.

In order to obtain error estimates for the k-step projection method defined in (2.15)–(2.16), we shall work with the following conditions:

(C1) The solution of problem (1.1)–(1.3) is sufficiently smooth. More specifically, we require

$$\mathbf{m}_0 \in \mathbf{H}^{2r} \cap \mathbf{W}^{2,4} \text{ and } \mathbf{m} \in L^\infty(0,T;\mathbf{H}^{2r} \cap \mathbf{W}^{2,4}) \cap C^1(0,T;\mathbf{H}^{2r}) \cap C^{k+1}(0,T;L^2).$$

(C2) The starting values $\widetilde{\mathbf{m}}_h^j = \mathbf{m}_h^j \in \mathbf{S}_h^r$, $j = 0, \ldots, k-1$, are sufficiently accurate approximations to the exact solution, satisfying the following estimates:

$$\max_{0 \leqslant \leqslant k-1} \|\mathbf{m}_h^j - I_h m(t_j)\|_{L^\infty} \leqslant \tau, \quad \max_{0 \leqslant \leqslant k-1} \|\Delta_h(\mathbf{m}_h^j - I_h m(t_j))\|_{L^2} \leqslant \tau^{\frac{2}{3}}, \tag{2.17}$$

$$\max_{0 \leqslant \leqslant k-1} \left(\|\mathbf{m}_h^j - I_h m(t_j)\|_{L^2}^2 + \tau \|\nabla(\mathbf{m}_h^j - I_h m(t_j))\|_{L^2}^2 \right) \lesssim \tau^{2k} + h^{2(r+1)}. \tag{2.18}$$

Remark 2.2. The local existence and uniqueness of the smooth solutions to problem (1.1)–(1.3) have been proved in [20] with some finite blow-up time $T_0 > 0$. For $T < T_0$ (before the solution blows up), the solution can be assumed to be sufficiently smooth. Moreover, the global existence and uniqueness of smooth solutions of harmonic map heat flow are available for a large class of general target surfaces including surfaces with non-positive sectional curvature (including hyperbolic spaces $\mathbb{H}^n$ and hyperplanes, see [50, Theorem 6.4]) and surfaces with

vanishing 2nd homotopy group (including n-sphere, i.e. $\pi_2(\mathbb{S}^n)=0$, for $n\in\mathbb{N}\backslash\{2\}$, see [50, Theorem 6.5]).

Remark 2.3. If the solution is sufficiently smooth, then the starting values $\widetilde{\mathbf{m}}_h^j=\mathbf{m}_h^j\in\mathbf{S}_h^r$, $j=0,\dots,k-1$, can be computed by a single-step high-order Runge–Kutta method or simply by temporal Taylor expansion using the partial differential equation. The computed numerical solution would have the desired accuracy satisfying (2.17)–(2.18).

The main theoretical result of this article is the following theorem.

Theorem 2.2. *Let Ω be a rectangular domain in $\mathbb{R}^d$ with $d\in\{1,2,3\}$. Under conditions (C1)–(C2), there exists a constant $\tau_0>0$ such that, under the stepsize condition $\kappa h^{r+1}\leqslant\tau^k\leqslant\tau_0$ (where $\kappa>0$ can be any constant independent of τ and h), the numerical solutions $\widetilde{\mathbf{m}}_h^n,\mathbf{m}_h^n\in\mathbf{S}_h^r$ given by the k-step projection method (2.15)–(2.16) satisfy the following estimate:*

$$\max_{k\leqslant n\leqslant N}(\|\mathbf{m}_h^n-\mathbf{m}(t_n)\|_{L^2}+\|\widetilde{\mathbf{m}}_h^n-\mathbf{m}(t_n)\|_{L^2})\lesssim\tau^k+h^{r+1}. \tag{2.19}$$

The results in this article can also be extended to a more general bounded polyhedral domain $\Omega\subset\mathbb{R}^d$ with triangular meshes, and adapted to a curved domain by combining the mass-lumping techniques [21,24] with the iso-parametric finite element method [22], and using the approximation results for a fixed surface [25]. More detailed comments are presented in the conclusion section. Moreover, our proofs apply to Dirichlet boundary condition and periodic boundary condition as well. Instead of being closed and compact, the surface Γ can also be open and unbounded (without boundary), since the boundedness of the exact solution is sufficient to conclude the same results.

3. Proof of Theorem 2.2

The following three lemmas will be frequently used in the proof of Theorem 2.2. They are concerned with the error of Lagrange interpolation, the inverse inequality of finite element functions, and the multiplier technique for the k-step BDF.

Lemma 3.1 (Error of the interpolation operator [17, Theorem 4.4.20])**.** *Let $0\leqslant s\leqslant r$ and $p>d/(s+1)$. Then the following results hold*

$$\|\mathbf{v}-I_h\mathbf{v}\|_{L^p}+h\|\mathbf{v}-I_h\mathbf{v}\|_{W^{1,p}}\lesssim h^{s+1}|v|_{W^{s+1,p}}, \tag{3.1}$$

$$\|\mathbf{v}-I_h\mathbf{v}\|_{L^\infty}\lesssim h^{2-\frac{d}{2}}|v|_{H^2}, \tag{3.2}$$

where $I_h:C(\overline{\Omega})\to\mathbf{S}_h^r$ is the Lagrange interpolation operator.

Lemma 3.2 (Inverse inequalities [17, Lemma 4.5.3, Theorem 4.5.11])**.** *Let $\mathbf{v}_h\in\mathbf{S}_h^r$, and let $1\leqslant p\leqslant\infty$, $1\leqslant q\leqslant\infty$, $0\leqslant m\leqslant l$. Then*

$$\|\mathbf{v}_h\|_{W^{l,p}(K)}\lesssim h^{m-l+d/p-d/q}\|\mathbf{v}_h\|_{W^{m,q}(K)}, \tag{3.3}$$

$$\Big(\sum_{K\in\mathcal{K}}\|\mathbf{v}_h\|_{W^{l,p}(K)}^p\Big)^{1/p}\lesssim h^{m-l+\min\{0,d/p-d/q\}}\Big(\sum_{K\in\mathcal{K}}\|\mathbf{v}_h\|_{W^{m,q}(K)}^q\Big)^{1/q}. \tag{3.4}$$

Lemma 3.3 (cf. [4, Section 2.4])**.** *Suppose that $1\leqslant k\leqslant 5$, $n\geqslant k$, and $\mathbf{u}_h^j\in\mathbf{S}_h^r$ for $0\leqslant j\leqslant k$. Then there exists a symmetric positive definite matrix $G=(g_{ij})\in\mathbb{R}^{k\times k}$ such that*

$$\Big(\sum_{j=0}^k\delta_j\mathbf{u}_h^{k-j},\mathbf{u}_h^k-\theta_k\mathbf{u}_h^{k-1}\Big)_h\geqslant\sum_{i,j=1}^k g_{ij}(\mathbf{u}_h^i,\mathbf{u}_h^j)_h-\sum_{i,j=1}^k g_{ij}(\mathbf{u}_h^{i-1},\mathbf{u}_h^{j-1})_h, \tag{3.5}$$

where $\theta_1=\theta_2=0$, $\theta_3=0.0836$, $\theta_4=0.2878$, $\theta_5=0.8160$.

3.1. The consistency error

To start with, it can be easily seen that the exact solution of problem (1.1)–(1.3) satisfies the following equation:

$$(\partial_t\mathbf{m},\mathbf{v}_h)+(\nabla\mathbf{m},\nabla\mathbf{v}_h)=-(A(\mathbf{m})(\nabla\mathbf{m},\nabla\mathbf{m}),\mathbf{v}_h)\quad\forall\,\mathbf{v}_h\in\mathbf{S}_h^r,$$

from which we further derive that

$$\begin{aligned}&\Big(\frac{1}{\tau}\sum_{j=0}^{k}\delta_j I_h\mathbf{m}(t_{n-j}),\mathbf{v}_h\Big)_h+(\nabla I_h\mathbf{m}(t_n),\nabla\mathbf{v}_h)\\&=-\Big(A_h(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))(\nabla I_h\widehat{\mathbf{m}}(t_n),\nabla I_h\widehat{\mathbf{m}}(t_n)),\mathbf{v}_h\Big)+\mathcal{E}_1(\mathbf{v}_h),\end{aligned}\tag{3.6}$$

where $\mathcal{E}_1(\mathbf{v}_h)$ is the consistency error due to the temporal discretization by the k-step BDF and the spatial discretization by the mass-lumping FEM, i.e.,

$$\begin{aligned}\mathcal{E}_1(\mathbf{v}_h):=&\Big(\frac{1}{\tau}I_h\sum_{j=0}^{k}\delta_j\mathbf{m}(t_{n-j}),\mathbf{v}_h\Big)_h-\Big(\frac{1}{\tau}I_h\sum_{j=0}^{k}\delta_j\mathbf{m}(t_{n-j}),\mathbf{v}_h\Big)\\&+\Big(\frac{1}{\tau}I_h\sum_{j=0}^{k}\delta_j\mathbf{m}(t_{n-j}),\mathbf{v}_h\Big)-\Big(\partial_t\mathbf{m}(t_n),\mathbf{v}_h\Big)\\&+\Big(\nabla(I_h\mathbf{m}(t_n)-\mathbf{m}(t_n)),\nabla\mathbf{v}_h\Big)\\&-\Big(A(\mathbf{m}(t_n))(\nabla\mathbf{m}(t_n),\nabla\mathbf{m}(t_n))-A_h(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))(\nabla I_h\widehat{\mathbf{m}}(t_n),\nabla I_h\widehat{\mathbf{m}}(t_n)),\mathbf{v}_h\Big)\\=:&\,\mathcal{E}_{1,1}(\mathbf{v}_h)+\mathcal{E}_{1,2}(\mathbf{v}_h)+\mathcal{E}_{1,3}(\mathbf{v}_h)+\mathcal{E}_{1,4}(\mathbf{v}_h).\end{aligned}$$

The term $\mathcal{E}_{1,3}(\mathbf{v}_h)$ is exactly the same as $\mathcal{E}_3(\mathbf{v}_h)$ in [35, Section 3.2]. The following result has been proved in [35, Lemma 3.7]:

$$|\mathcal{E}_{1,3}(\mathbf{v}_h)|\lesssim h^{r+1}\|\mathbf{v}_h\|_{H^1}.\tag{3.7}$$

Furthermore, with the help of Lemma 2.1, the remaining three terms $\mathcal{E}_{1,1}(\mathbf{v}_h)$, $\mathcal{E}_{1,2}(\mathbf{v}_h)$ and $\mathcal{E}_{1,4}(\mathbf{v}_h)$ can be estimated similarly as the estimates of $\mathcal{E}_1(\mathbf{v}_h)$, $\mathcal{E}_2(\mathbf{v}_h)$ and $\mathcal{E}_4(\mathbf{v}_h)$ in [35, Section 3.2]. Specifically, the following results hold:

$$|\mathcal{E}_{1,1}(\mathbf{v}_h)|+|\mathcal{E}_{1,2}(\mathbf{v}_h)|+|\mathcal{E}_{1,4}(\mathbf{v}_h)|]\lesssim h^{r+1}\|\mathbf{v}_h\|_{H^1}+\tau^k\|\mathbf{v}_h\|_{L^2}.\tag{3.8}$$

Then, by collecting the results in (3.7)–(3.8), we obtain the following estimate of the consistency error:

$$|\mathcal{E}_1(\mathbf{v}_h)|\lesssim(\tau^k+h^{r+1})\|\mathbf{v}_h\|_{L^2}+h^{r+1}\|\nabla\mathbf{v}_h\|_{L^2}.\tag{3.9}$$

3.2. Mathematical induction and error equation

We consider two types of error functions corresponding to the projected numerical solution and the auxiliary numerical solution, respectively, i.e.,

$$\mathbf{e}_h^n:=I_h\mathbf{m}(t_n)-\mathbf{m}_h^n\quad\text{and}\quad\widetilde{\mathbf{e}}_h^n:=I_h\mathbf{m}(t_n)-\widetilde{\mathbf{m}}_h^n.\tag{3.10}$$

We shall prove the smallness of these error functions by mathematical induction: For $1\leqslant n\leqslant l$, we assume that the numerical solution $\mathbf{m}_h^{n-1}$ is uniquely determined and

$$\|\widetilde{\mathbf{e}}_h^{n-1}\|_{L^\infty}\leqslant\tau,\tag{3.11}$$

$$\|\Delta_h\widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}\leqslant\tau^{\frac{2}{3}},\tag{3.12}$$

$$\|\widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}^2+\tau\|\nabla\widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}^2\leqslant\tau^{2k-\frac{2}{5}}+h^{2(r+\frac{23}{25})},\tag{3.13}$$

which are all naturally true for $1\leqslant n\leqslant k$ since the functions $\mathbf{m}_h^j=\widetilde{\mathbf{m}}_h^j$, $0\leqslant j\leqslant k-1$, are given and satisfying (2.17)–(2.18). Under the induction assumption we shall prove that (3.11)–(3.13) also hold for $n=l+1$.

Recalling that the mesh size restriction $h^{r+1}\lesssim\tau^k$, the constant $\frac{23}{25}$ in (3.13) is chosen so that the term

$$h^{(r+\frac{23}{25})}\lesssim\tau^{\frac{k(r+\frac{23}{25})}{r+1}}=\tau^{k-\frac{2k}{25(r+1)}}\tag{3.14}$$

can always be bounded by $\tau^{k-\frac{1}{5}}$ for any $r\geqslant 1$ and $k\leqslant 5$. From (3.10) and (3.11) we immediately get the following L^∞-estimates for $1\leqslant n\leqslant l$:

$$\|I_h\mathbf{m}(t_{n-1})\|_{L^\infty}\lesssim 1\quad\text{and}\quad\|\widetilde{\mathbf{m}}_h^{n-1}\|_{L^\infty}\lesssim 1.\tag{3.15}$$

These L^∞-estimates will be frequently used.

The error equation can be obtained by substracting equation (2.15) from (3.6), i.e.,

$$\begin{aligned}
&\Big(\frac{\delta_0}{\tau}\widetilde{\mathbf{e}}_h^n+\frac{1}{\tau}\sum_{j=1}^{k}\delta_j\mathbf{e}_h^{n-j},\mathbf{v}_h\Big)_h-(\Delta_h\widetilde{\mathbf{e}}_h^n,\mathbf{v}_h)\\
&=-\Big(A_h(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))(\nabla I_h\widehat{\mathbf{m}}(t_n),\nabla I_h\widehat{\mathbf{m}}(t_n))-A_h(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)(\nabla\widehat{\mathbf{m}}_h^n,\nabla\widehat{\mathbf{m}}_h^n),\mathbf{v}_h\Big)+\mathcal{E}_1(\mathbf{v}_h) \qquad (3.16)\\
&=\mathcal{E}_1(\mathbf{v}_h)-\Big(\nabla I_h\widehat{\mathbf{m}}(t_n)\cdot\big(I_h\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)\nabla I_h\widehat{\mathbf{m}}(t_n)\big)-\nabla\widehat{\mathbf{m}}_h^n\cdot I_h\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)\nabla\widehat{\mathbf{m}}_h^n,I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n)\cdot\mathbf{v}_h\Big)\\
&\quad-\Big(\nabla I_h\widehat{\mathbf{m}}(t_n)\cdot\big(I_h\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))-I_h\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)\big)\nabla I_h\widehat{\mathbf{m}}(t_n),I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n)\cdot\mathbf{v}_h\Big)\\
&\quad-\Big(\nabla I_h\widehat{\mathbf{m}}(t_n)\cdot I_h\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))\nabla I_h\widehat{\mathbf{m}}(t_n),\big(I_h\boldsymbol{\nu}(I_h\widehat{\mathbf{m}}(t_n))-I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n)\big)\cdot\mathbf{v}_h\Big)\\
&=:\mathcal{E}_1(\mathbf{v}_h)+\mathcal{E}_2(\mathbf{v}_h)+\mathcal{E}_3(\mathbf{v}_h)+\mathcal{E}_4(\mathbf{v}_h), \qquad (3.17)
\end{aligned}$$

which holds for all $k\leqslant n\leqslant l$ and $\mathbf{v}_h\in\mathbf{S}_h^r$. In view of the multiplier technique for the k-step BDF in Lemma 3.3, we choose $\mathbf{v}_h=\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1}$ in (3.17) and estimate $\mathcal{E}_j(\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1})$ below for $j=1,2,3,4$.

The first term on the right-hand side of (3.17) can be estimated by using (3.9), i.e.,

$$|\mathcal{E}_1(\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1})|\lesssim(\tau^k+h^{r+1})(\|\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}+\|\nabla(\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1})\|_{L^2}). \qquad (3.18)$$

The third and fourth terms on the right-hand side of (3.17) depend on the estimation of $\|I_h\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))-I_h\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)\|_{L^2}$ and $\|I_h\boldsymbol{\nu}(I_h\widehat{\mathbf{m}}(t_n))-I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n)\|_{L^2}$, which can be estimated by using the interpolation error estimates in Lemma 3.1 (with $p=4$ and $s=0$ therein), i.e.,

$$\begin{aligned}
&\|I_h\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))-I_h\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)\|_{L^2}\\
&\leqslant\|\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))-\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)\|_{L^2}+\|(I_h-1)\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))-(I_h-1)\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)\|_{L^2}\\
&\qquad\qquad\text{(triangle inequality is used)}\\
&\lesssim\|\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))-\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)\|_{L^2}+h|\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))-\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)|_{W^{1,4}}\\
&\qquad\qquad\text{(Lemma 3.1 and the inclusion } H^1\subseteq W^{1,4}\text{ are used)}\\
&\leqslant\|\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_n))-\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^n)\|_{L^2}+h\|\nabla(\mathbf{H}\circ\mathbf{a})(\widehat{\mathbf{m}}_h^n)\cdot\nabla(I_h\widehat{\mathbf{m}}(t_n)-\widehat{\mathbf{m}}_h^n)\|_{L^4}\\
&\quad+h\|(\nabla(\mathbf{H}\circ\mathbf{a})(I_h\widehat{\mathbf{m}}(t_n))-\nabla(\mathbf{H}\circ\mathbf{a})(I_h\widehat{\mathbf{m}}_h^n))\cdot\nabla I_h\widehat{\mathbf{m}}(t_n)\|_{L^4}\\
&\qquad\qquad\text{(chain rule and triangle inequality are used)}\\
&\lesssim\|\widehat{\mathbf{m}}(t_n)-\widehat{\mathbf{m}}_h^n\|_{L^2}+h\|\nabla(I_h\widehat{\mathbf{m}}(t_n)-\widehat{\mathbf{m}}_h^n)\|_{L^4}+h\|I_h\widehat{\mathbf{m}}(t_n)-I_h\widehat{\mathbf{m}}_h^n\|_{L^4}\\
&\qquad\qquad\text{(Lipschitz continuity of }\mathbf{H}\circ\mathbf{a}\text{ and }\nabla(\mathbf{H}\circ\mathbf{a})\text{ is used)}\\
&\lesssim\sum_{j=0}^{k-1}\|\mathbf{e}_h^{n-j-1}\|_{L^2}+\sum_{j=0}^{k-1}\|\mathbf{e}_h^{n-j-1}\|_{L^4}\qquad\text{(Lemma 3.2 is used)}\\
&\leqslant C_\varepsilon\sum_{j=0}^{k-1}\|\mathbf{e}_h^{n-j-1}\|_{L^2}+\varepsilon\sum_{j=0}^{k-1}\|\mathbf{e}_h^{n-j-1}\|_{L^6}, \qquad (3.19)
\end{aligned}$$

which holds for any $\varepsilon>0$ (with a constant C_ε depending on ε^{-1}). In the second to last inequality we have used the inverse inequality (Lemma 3.2), the local Lipschitz continuity of $\nabla(\mathbf{H}\circ\mathbf{a})$, and the boundedness of $I_h\widehat{\mathbf{m}}(t_{n-j-1})$ and $\widehat{\mathbf{m}}_h^{n-j-1}$ as shown in (3.15). In the last inequality we have used the Sobolev interpolation inequality for $d\leqslant 3$ and Young's inequality. Estimate (3.19) holds as well if $I_h\mathbf{H}(\mathbf{a}\circ\cdot)$ is replaced by $I_h\boldsymbol{\nu}(\cdot)$. From (3.19) and the local Lipschitz continuity of $\mathbf{H}$ and $\boldsymbol{\nu}$ we immediately obtain the following estimate:

$$\begin{aligned}
&|\mathcal{E}_3(\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1})|+|\mathcal{E}_4(\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1})|\\
&\leqslant\sum_{j=0}^{k-1}(C_\varepsilon\|\mathbf{e}_h^{n-j-1}\|_{L^2}+\varepsilon\|\mathbf{e}_h^{n-j-1}\|_{L^6})\|\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1}\|_{L^2} \qquad (3.20)
\end{aligned}$$

for $k \leqslant n \leqslant l$ and any $\varepsilon > 0$.

The estimation of $|\mathcal{E}_2(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})|$ is presented in the next subsection.

3.3. Consequences of the induction assumption

Under the induction assumption (3.11)–(3.13), the two types of the errors defined in (3.10) are connected through the following lemmas.

Lemma 3.4. *Under the induction assumption in (3.11), the following inequality holds for sufficient small τ:*

$$|\mathbf{e}_h^{n-1}| \lesssim |\widetilde{\mathbf{e}}_h^{n-1}| \quad \text{at all finite element nodes},$$

and

$$\|\mathbf{e}_h^{n-1}\|_{L_h^p} \lesssim \|\widetilde{\mathbf{e}}_h^{n-1}\|_{L_h^p} \quad \text{and} \quad \|\mathbf{e}_h^{n-1}\|_{L^p} \lesssim \|\widetilde{\mathbf{e}}_h^{n-1}\|_{L^p} \quad \forall\, 1 \leqslant p \leqslant +\infty. \tag{3.21}$$

Proof. The proof of these results in the case $\Gamma = \mathbb{S}^n$ can be found in [35, Lemma 3.9] (the case $p = \infty$ was not explicitly stated in [35, Lemma 3.9] but the proof is the same as $1 \leqslant p < \infty$). The proof of these results for a general target surface Γ follows from a similar argument with the help of smallness condition (3.11), which guarantees that $\widetilde{\mathbf{m}}_h^{n-1}$ is in a small neighborhood of the surface Γ so that the distance projection onto the surface Γ is well defined. □

This lemma connects the two types of errors in L^p-norm. Hence using this lemma and result (3.20), we can directly deduce that

$$\begin{aligned}
&|\mathcal{E}_3(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})| + |\mathcal{E}_4(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})| \\
&\leqslant \sum_{j=0}^{k-1} (C_\varepsilon \|\widetilde{\mathbf{e}}_h^{n-j-1}\|_{L^2} + \varepsilon \|\widetilde{\mathbf{e}}_h^{n-j-1}\|_{L^6}) \|\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}\|_{L^2} \\
&\leqslant \sum_{j=0}^{k-1} (C_\varepsilon \|\widetilde{\mathbf{e}}_h^{n-j-1}\|_{L^2} + \varepsilon \|\nabla \widetilde{\mathbf{e}}_h^{n-j-1}\|_{L^2}) \|\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}.
\end{aligned} \tag{3.22}$$

Next we provide a lemma which connects the two types of errors in $W^{1,p}$-norm.

Lemma 3.5. *Under the induction assumption in (3.11), the following estimate holds for sufficiently small τ:*

$$\|\mathbf{e}_h^{n-1}\|_{W^{1,p}} \lesssim \|\widetilde{\mathbf{e}}_h^{n-1}\|_{W^{1,p}} + h^r \quad \forall\, 2 \leqslant p \leqslant +\infty. \tag{3.23}$$

Proof. The L^p-norm estimate is contained in Lemma 3.4. Hence, we focus on the estimation of the $W^{1,p}$ semi-norm, i.e.,

$$\begin{aligned}
\|\nabla \mathbf{e}_h^{n-1}\|_{L^p} &= \|\nabla \big(I_h(\mathbf{a} \circ \widetilde{\mathbf{m}}_h^{n-1}) - I_h \mathbf{m}(t_{n-1})\big)\|_{L^p} \\
&= \|\nabla \big(I_h(\mathbf{a} \circ \widetilde{\mathbf{m}}_h^{n-1}) - \mathbf{a} \circ \widetilde{\mathbf{m}}_h^{n-1}\big)\|_{L^p} + \|\nabla \big(\mathbf{a} \circ \widetilde{\mathbf{m}}_h^{n-1} - \mathbf{a} \circ (I_h \mathbf{m}(t_{n-1}))\big)\|_{L^p} \\
&\quad + \|\nabla \big(\mathbf{a} \circ (I_h \mathbf{m}(t_{n-1})) - \mathbf{a} \circ \mathbf{m}(t_{n-1})\big)\|_{L^p} + \|\nabla \big(\mathbf{m}(t_{n-1}) - I_h \mathbf{m}(t_{n-1})\big)\|_{L^p} \\
&=: \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4.
\end{aligned} \tag{3.24}$$

Since $\mathbf{a}$ is smooth in a neighborhood of Γ, it follows that $|(\nabla^j \mathbf{a})(\widetilde{\mathbf{m}}_h^{n-1})| \lesssim 1$ for $j = 1, \cdots, m$. By using this result and the Hölder inequality, the inverse inequality, the Leibniz rule, and the boundedness condition (3.11), we derive

$$\begin{aligned}
&\|\nabla \big(I_h(\mathbf{a} \circ \widetilde{\mathbf{m}}_h^{n-1}) - \mathbf{a} \circ \widetilde{\mathbf{m}}_h^{n-1}\big)\|_{L^p(K)} \\
&\lesssim h^{r+d/p} |\mathbf{a} \circ \widetilde{\mathbf{m}}_h^{n-1}|_{W^{r+1,\infty}(K)} \\
&\lesssim h^{r+d/p} \big(\|(\nabla^{r+1}\mathbf{a})(\widetilde{\mathbf{m}}_h^{n-1}) : (\nabla \widetilde{\mathbf{m}}_h^{n-1})^{r+1}\|_{L^\infty(K)} + \cdots + \|(\nabla \mathbf{a})(\widetilde{\mathbf{m}}_h^{n-1}) \cdot \nabla^{r+1} \widetilde{\mathbf{m}}_h^{n-1}\|_{L^\infty(K)}\big) \\
&\lesssim h^{r+d/p} \big(\|\nabla \widetilde{\mathbf{m}}_h^{n-1}\|_{L^\infty(K)}^{r+1} + \cdots + \|\nabla^{r+1} \widetilde{\mathbf{m}}_h^{n-1}\|_{L^\infty(K)}\big) \\
&\lesssim h^{r+d/p} \big(\|\nabla \widetilde{\mathbf{e}}_h^{n-1}\|_{L^\infty(K)}^{r+1} + \cdots + \|\nabla^{r+1} \widetilde{\mathbf{e}}_h^{n-1}\|_{L^\infty(K)}\big) \\
&\quad + h^{r+d/p} \big(\|\nabla I_h \mathbf{m}(t_{n-1})\|_{L^\infty(K)}^{r+1} + \cdots + \|\nabla^{r+1} I_h \mathbf{m}(t_{n-1})\|_{L^\infty(K)}\big)
\end{aligned}$$

$$\lesssim h^{d/p}\|\nabla\widetilde{\mathbf{e}}_h^{n-1}\|_{L^\infty(K)}+h^{r+d/p}\big(\|\nabla I_h\mathbf{m}(t_{n-1})\|_{L^\infty(K)}+\cdots+\|\nabla^{r+1}I_h\mathbf{m}(t_{n-1})\|_{L^\infty(K)}\big)$$
$$\lesssim \|\nabla\widetilde{\mathbf{e}}_h^{n-1}\|_{L^p(K)}+h^r\|\mathbf{m}(t_{n-1})\|_{W^{r+1,p}(K)}. \tag{3.25}$$

The condition $p\geqslant 2$ is required to ensure that I_h is stable in $W^{2,p}$ for $d\leqslant 3$. Then after taking l^p norm of the above estimate over $K\in\mathcal{K}$, we get

$$\mathcal{A}_1\lesssim \|\nabla\widetilde{\mathbf{e}}_h^{n-1}\|_{L^p}+h^r\|\mathbf{m}(t_{n-1})\|_{W^{r+1,p}}. \tag{3.26}$$

By using the Leibniz rule and the triangle inequality, we can estimate $\mathcal{A}_2$ as follows:

$$\begin{aligned}
\mathcal{A}_2&=\|\nabla\mathbf{a}(\widetilde{\mathbf{m}}_h^{n-1})\cdot\nabla\widetilde{\mathbf{m}}_h^{n-1}-\nabla\mathbf{a}(I_h\mathbf{m}(t_{n-1}))\cdot\nabla(I_h\mathbf{m}(t_{n-1}))\|_{L^p}\\
&\lesssim \|\nabla\big(\widetilde{\mathbf{m}}_h^{n-1}-I_h\mathbf{m}(t_{n-1})\big)\|_{L^p}+\|\nabla\mathbf{a}(\widetilde{\mathbf{m}}_h^{n-1})-\nabla\mathbf{a}(I_h\mathbf{m}(t_{n-1}))\|_{L^p}\\
&\lesssim \|\nabla\big(\widetilde{\mathbf{m}}_h^{n-1}-I_h\mathbf{m}(t_{n-1})\big)\|_{L^p}+\Big(\sum_{K\in\mathcal{K}}h^d\|\nabla\mathbf{a}(\widetilde{\mathbf{m}}_h^{n-1})-\nabla\mathbf{a}(I_h\mathbf{m}(t_{n-1}))\|_{L^\infty(K)}^p\Big)^{1/p}\\
&\lesssim \|\nabla\big(\widetilde{\mathbf{m}}_h^{n-1}-I_h\mathbf{m}(t_{n-1})\big)\|_{L^p}+\Big(\sum_{K\in\mathcal{K}}h^d\|\widetilde{\mathbf{m}}_h^{n-1}-I_h\mathbf{m}(t_{n-1})\|_{L^\infty(K)}^p\Big)^{1/p}\\
&\lesssim \|\widetilde{\mathbf{m}}_h^{n-1}-I_h\mathbf{m}(t_{n-1})\|_{W^{1,p}}\qquad\text{(the inverse inequality is used)}\\
&=\|\widetilde{\mathbf{e}}_h^{n-1}\|_{W^{1,p}},
\end{aligned}\tag{3.27}$$

where we have used the local Lipschitz continuity of $\nabla\mathbf{a}$ and the boundedness of $\widetilde{\mathbf{m}}_h^{n-1}$ in the third inequality. The term $\mathcal{A}_3$ can be estimated analogously, i.e.,

$$\begin{aligned}
\mathcal{A}_3&=\|\nabla\big(\mathbf{a}\circ(I_h\mathbf{m}(t_{n-1}))-\mathbf{a}\circ\mathbf{m}(t_{n-1})\big)\|_{L^p}\\
&\lesssim \|\nabla\big(\mathbf{m}(t_{n-1})-I_h\mathbf{m}(t_{n-1})\big)\|_{L^p}+\|\nabla\mathbf{a}(\mathbf{m}(t_{n-1}))-\nabla\mathbf{a}(I_h\mathbf{m}(t_{n-1}))\|_{L^p}\\
&\lesssim \|\mathbf{m}(t_{n-1})-I_h\mathbf{m}(t_{n-1})\|_{W^{1,\infty}}\\
&\lesssim h^r\|\mathbf{m}(t_{n-1})\|_{W^{r+1,\infty}}.
\end{aligned}\tag{3.28}$$

Substituting the estimates of $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ into (3.24), and using the standard estimates for the Lagrange interpolation error, we obtain

$$\begin{aligned}
\|\nabla\mathbf{e}_h^{n-1}\|_{L^p}&\lesssim \|\nabla\widetilde{\mathbf{e}}_h^{n-1}\|_{L^p}+h^r\|\mathbf{m}(t_{n-1})\|_{W^{r+1,p}}\\
&\quad+\|\widetilde{\mathbf{e}}_h^{n-1}\|_{W^{1,p}}+h^r\|\mathbf{m}(t_{n-1})\|_{W^{r+1,\infty}}+h^r\|\mathbf{m}(t_{n-1})\|_{W^{r+1,p}}\\
&\lesssim \|\widetilde{\mathbf{e}}_h^{n-1}\|_{W^{1,p}}+h^r.
\end{aligned}\tag{3.29}$$

This completes the proof of Lemma 3.5. □

Lemma 3.6. *Under the induction assumptions in (3.11)–(3.12), the following result holds:*

$$\|\nabla\mathbf{e}_h^{n-1}\|_{L^6}\lesssim\tau^{\frac{2}{3}}+h^r. \tag{3.30}$$

Proof. By using Lemma 3.5 and the following discrete Sobolev embedding inequality for $d\leqslant 3$ (see [35, Lemma 3.5]):

$$\|\nabla\mathbf{v}_h\|_{L^6}\lesssim\|\Delta_h\mathbf{v}_h\|_{L^2}\quad\forall\,\mathbf{v}_h\in\mathbf{S}_h^r,$$

we have

$$\|\nabla\mathbf{e}_h^{n-1}\|_{L^6}\lesssim\|\widetilde{\mathbf{e}}_h^{n-1}\|_{W^{1,6}}+h^r\lesssim\|\widetilde{\mathbf{e}}_h^{n-1}\|_{L^\infty}+\|\Delta_h\widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}+h^r\lesssim\tau^{\frac{2}{3}}+h^r \tag{3.31}$$

for $1\leqslant n\leqslant l$, where (3.11)–(3.12) and $\tau\lesssim 1$ are used. □

By using this lemma and the triangle inequality, we obtain the $W^{1,6}$-boundedness of the numerical solutions, i.e.,

$$\|\nabla\mathbf{m}_h^{l-j-1}\|_{L^6}\leqslant\|\nabla\mathbf{e}_h^{l-j-1}\|_{L^6}+\|\nabla I_h\mathbf{m}(t_{l-j-1})\|_{L^6}\lesssim 1\quad\text{for }\ 0\leqslant j\leqslant k-1.$$

Now we turn to the estimation of $\mathcal{E}_2(\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1})$ for $1\leqslant n\leqslant l$ by utilizing the estimates in Lemma 3.4, Lemma 3.5, and Lemma 3.6.

Lemma 3.7. *Under the induction assumptions in (3.11)–(3.12), the following results hold for* $1 \leqslant n \leqslant l$*:*

$$|\mathcal{E}_2(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})| \leqslant C_\varepsilon \sum_{j=0}^{k} \|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2 + \varepsilon \sum_{j=0}^{k} \|\nabla \widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2, \tag{3.32}$$

$$|\mathcal{E}_2(\widetilde{\mathbf{e}}_h^n)| \leqslant C_\varepsilon \sum_{j=0}^{k} \|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2 + \varepsilon \sum_{j=0}^{k} \|\nabla \widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2. \tag{3.33}$$

Proof. Recall the definiton of $\mathcal{E}_2$. Since $\mathbf{H}$ is symmetric, $I_h\mathbf{H}$ induces a symmetric bilinear form. Hence we have

$$\begin{aligned}
\mathcal{E}_2(\mathbf{v}_h) &= \Big(\nabla I_h \widehat{\mathbf{m}}(t_n) \cdot \big(I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)\nabla I_h \widehat{\mathbf{m}}(t_n)\big) - \nabla \widehat{\mathbf{m}}_h^n \cdot I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)\nabla \widehat{\mathbf{m}}_h^n, I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n) \cdot \mathbf{v}_h\Big) \\
&= -\Big(\nabla(I_h \widehat{\mathbf{m}}(t_n) - \widehat{\mathbf{e}}_h^n) \cdot I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)\nabla(I_h \widehat{\mathbf{m}}(t_n) - \widehat{\mathbf{e}}_h^n) \\
&\qquad - \nabla I_h \widehat{\mathbf{m}}(t_n) \cdot \big(I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)\nabla I_h \widehat{\mathbf{m}}(t_n)\big), I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n) \cdot \mathbf{v}_h\Big) \\
&= -\Big(\nabla I_h \widehat{\mathbf{e}}_h^n \cdot \big(I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)\nabla I_h \widehat{\mathbf{e}}_h^n\big) - 2\nabla \widehat{\mathbf{e}}_h^n \cdot I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)\nabla I_h \widehat{\mathbf{m}}(t_n), I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n) \cdot \mathbf{v}_h\Big).
\end{aligned}$$

Then we estimate $\mathcal{E}_2(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})$ as follows, by utilizing Lemma 3.4, Lemma 3.5, and Lemma 3.6:

$$\begin{aligned}
\mathcal{E}_2(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}) &= -\Big(\nabla \widehat{\mathbf{e}}_h^n \cdot I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)(2\nabla I_h \widehat{\mathbf{m}}(t_n) - \nabla \widehat{\mathbf{e}}_h^n), I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n) \cdot (\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})\Big) \\
&\qquad \text{(the symmetry of } \mathbf{H} \text{ is used)} \\
&= -\Big(\nabla \widehat{\mathbf{e}}_h^n \cdot I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)(2\nabla I_h \widehat{\mathbf{m}}(t_n) - 2\nabla \widehat{\mathbf{m}}(t_n)), I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n) \cdot (\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})\Big) \\
&\quad - \Big(\nabla \widehat{\mathbf{e}}_h^n \cdot I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)(2\nabla \widehat{\mathbf{m}}(t_n) - \nabla \widehat{\mathbf{e}}_h^n), I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n) \cdot (\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})\Big) \\
&\lesssim \sum_{j=0}^{k-1} h\|\nabla \mathbf{e}_h^{n-j-1}\|_{L^2}\|\widehat{\mathbf{m}}(t_n)\|_{W^{2,\infty}}\|\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}\|_{L^2} \\
&\quad + \Big|\Big(\nabla \widehat{\mathbf{e}}_h^n \cdot I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)\nabla \widehat{\mathbf{e}}_h^n, I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n) \cdot (\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})\Big)\Big| \\
&\quad + \Big|\Big(\widehat{\mathbf{e}}_h^n \nabla \cdot \big(I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)2\nabla \widehat{\mathbf{m}}(t_n)\big), I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n) \cdot (\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})\Big)\Big| \\
&\quad + \Big|\Big(\widehat{\mathbf{e}}_h^n \cdot I_h\mathbf{H}(\mathbf{a} \circ \widehat{\mathbf{m}}_h^n)2\nabla \widehat{\mathbf{m}}(t_n), \nabla\big(I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^n) \cdot (\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})\big)\Big)\Big| \\
&\qquad \text{(integration by parts)} \\
&\lesssim \sum_{j=0}^{k-1} \|\mathbf{e}_h^{n-j-1}\|_{L^2}\|\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}\|_{H^1} \qquad \text{(inverse inequality)} \\
&\quad + \sum_{j=0}^{k-1}(\|\widetilde{\mathbf{e}}_h^{n-j-1}\|_{H^1} + h^r) \sum_{j=0}^{k-1} \|\nabla \mathbf{e}_h^{n-j-1}\|_{L^4}\|\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}\|_{L^4} \\
&\qquad \text{(Lemma 3.5 is used)} \\
&\quad + \sum_{j=0}^{k-1} \|\mathbf{e}_h^{n-j-1}\|_{L^2}(\|\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}\|_{L^2} + \|\nabla(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})\|_{L^2}) \\
&\qquad \text{(Lemma 3.6 is used)} \\
&\leqslant C \sum_{j=0}^{k-1} \|\widetilde{\mathbf{e}}_h^{n-j-1}\|_{L^2}^2 + C\|\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}^2 + \varepsilon\|\nabla(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})\|_{L^2}^2
\end{aligned}$$

$$+ C\sum_{j=0}^{k-1}\|\widetilde{\mathbf{e}}_h^{n-j-1}\|_{H^1}\|\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1}\|_{L^4}+C\sum_{j=0}^{k-1}\|\widetilde{\mathbf{e}}_h^{n-j-1}\|_{L^4}\|\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1}\|_{L^4}$$

(Lemma 3.4 and Lemma 3.6 are used)

$$\leqslant C_\varepsilon\sum_{j=0}^{k}\|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2+\varepsilon\sum_{j=0}^{k}\|\nabla\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2,$$

where we have used the Sobolev interpolation inequality and Young's inequality in the last inequality, i.e.,

$$\|u\|_{L^4}\lesssim\|u\|_{L^2}^{\frac{1}{4}}\|u\|_{L^6}^{\frac{3}{4}}\leqslant C_\varepsilon\|u\|_{L^2}+\varepsilon\|\nabla u\|_{L^2}.$$

By setting $\theta_k=0$ we also obtain the following result:

$$|\mathcal{E}_2(\widetilde{\mathbf{e}}_h^n)|\leqslant C_\varepsilon\sum_{j=0}^{k}\|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2+\varepsilon\sum_{j=0}^{k}\|\nabla\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2. \tag{3.34}$$

This completes the proof of Lemma 3.7. □

3.4. A sub-optimal error estimate

As a preliminary result, we prove a sub-optimal error estimate by choosing the test function $\mathbf{v}_h=\widetilde{\mathbf{e}}_h^n$ in equation (3.17). This sub-optimal error estimate will be used in the next subsection to prove an L^∞-norm error bound of $O(\tau)$, which is needed in proving the optimal-order convergence.

For $k\leqslant n\leqslant l$, we choose $\mathbf{v}_h=\widetilde{\mathbf{e}}_h^n$ in (3.17) and rewrite its left-hand side as follows:

$$\Big(\frac{\delta_0}{\tau}\widetilde{\mathbf{e}}_h^n+\frac{1}{\tau}\sum_{j=1}^{k}\delta_j\mathbf{e}_h^{n-j},\widetilde{\mathbf{e}}_h^n\Big)_h-(\Delta_h\widetilde{\mathbf{e}}_h^n,\widetilde{\mathbf{e}}_h^n)=\frac{\delta_0}{\tau}\|\widetilde{\mathbf{e}}_h^n\|_{L_h^2}^2+\frac{1}{\tau}\Big(\sum_{j=1}^{k}\delta_j\mathbf{e}_h^{n-j},\widetilde{\mathbf{e}}_h^n\Big)_h+\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2.$$

The second term on the right-hand side can be estimated by using (3.21), (2.2), induction assumption (3.13), and $h^{r+1}\lesssim\tau^k$, which imply that

$$\begin{aligned}\Big|\frac{1}{\tau}\Big(\sum_{j=1}^{k}\delta_j\mathbf{e}_h^{n-j},\widetilde{\mathbf{e}}_h^n\Big)_h\Big|&\leqslant\frac{\varepsilon_1}{\tau}\|\widetilde{\mathbf{e}}_h^n\|_{L_h^2}^2+\frac{C}{\tau}\sum_{j=1}^{k}\|\mathbf{e}_h^{n-j}\|_{L_h^2}^2\leqslant\frac{\varepsilon_1}{\tau}\|\widetilde{\mathbf{e}}_h^n\|_{L_h^2}^2+\frac{C}{\tau}(\tau^{2k-\frac{2}{5}}+h^{2(r+\frac{23}{25})})\\&\leqslant\frac{\varepsilon_1}{\tau}\|\widetilde{\mathbf{e}}_h^n\|_{L_h^2}^2+C\tau^{2k-\frac{7}{5}}.\end{aligned}$$

The right-hand side of (3.17) can be estimated by using the estimates of $|\mathcal{E}_j(\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1})|$, $j=1,2,3,4$, in (3.18), (3.22) and (3.34). Then we obtain

$$\frac{\delta_0}{\tau}\|\widetilde{\mathbf{e}}_h^n\|_{L_h^2}^2+\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2\leqslant\frac{\varepsilon_1}{\tau}\|\widetilde{\mathbf{e}}_h^n\|_{L_h^2}^2+C\tau^{2k-\frac{7}{5}}+Ch^{2(r+1)}+C\sum_{j=0}^{k}\|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2+\varepsilon_2\sum_{j=0}^{k}\|\nabla\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2$$

for $k\leqslant n\leqslant l$. By choosing sufficiently small constants ε_1 and ε_2, and multiplying the inequality by τ, we have

$$\delta_0\|\widetilde{\mathbf{e}}_h^n\|_{L^2}^2+\tau\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2\lesssim\tau^{2k-\frac{2}{5}}+\tau h^{2(r+1)}+\tau\sum_{j=0}^{k}\|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2+\tau\sum_{j=1}^{k}\|\nabla\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2$$

for $k\leqslant n\leqslant l$. Then, by using (2.18) and (3.13), we obtain

$$\begin{aligned}\|\widetilde{\mathbf{e}}_h^n\|_{L^2}^2+\tau\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2&\lesssim\tau^{2k-\frac{2}{5}}+\tau h^{2(r+1)}+\tau\sum_{j=1}^{k}\|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2+\tau\sum_{j=1}^{k}\|\nabla\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2\\&\lesssim\tau^{2k-\frac{2}{5}}+h^{2(r+\frac{23}{25})}\lesssim\tau^{2k-\frac{2}{5}}\quad\text{for } k\leqslant n\leqslant l.\end{aligned}\tag{3.35}$$

where the last inequality uses (3.14). Since $k\geqslant 2$, the inequality above also implies that

$$\|\widetilde{\mathbf{e}}_h^n\|_{L^2}\lesssim\tau^{\frac{9}{5}}\quad\text{and}\quad\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}\lesssim\tau^{\frac{13}{10}}\quad\text{for } k\leqslant n\leqslant l. \tag{3.36}$$

3.5. Proof of (3.11)–(3.12) for $n = l + 1$

When $n = l$ the error equation (3.16) becomes

$$\Big(\frac{\delta_0}{\tau}\widetilde{\mathbf{e}}_h^l + \frac{1}{\tau}\sum_{j=1}^{k}\delta_j\mathbf{e}_h^{l-j}, \mathbf{v}_h\Big)_h - (\Delta_h\widetilde{\mathbf{e}}_h^l, \mathbf{v}_h) \tag{3.37}$$

$$= -\Big(A_h(I_h\widehat{\mathbf{m}}(t_l))(\nabla I_h\widehat{\mathbf{m}}(t_l), \nabla I_h\widehat{\mathbf{m}}(t_l)) - A_h(\widehat{\mathbf{m}}_h^l)(\nabla\widehat{\mathbf{m}}_h^l, \nabla\widehat{\mathbf{m}}_h^l), \mathbf{v}_h\Big) + \mathcal{E}_1(\mathbf{v}_h), \tag{3.38}$$

which holds for all $\mathbf{v}_h \in \mathbf{S}_h^r$. By using (3.9), the inverse inequality in Lemma 3.2, and $h^{r+1} \lesssim \tau^k$, we have

$$|\mathcal{E}_1(\mathbf{v}_h)| \lesssim (\tau^k + h^r)\|\mathbf{v}_h\|_{L^2} \lesssim \tau^{\frac{rk}{r+1}}\|\mathbf{v}_h\|_{L^2} \lesssim \tau\|\mathbf{v}_h\|_{L^2}, \tag{3.39}$$

where the last inequality uses $2 \leqslant k \leqslant 5$ and $r \geqslant 1$. From (3.36) we also conclude that

$$\Big|\Big(\frac{\delta_0}{\tau}\widetilde{\mathbf{e}}_h^l + \frac{1}{\tau}\sum_{j=1}^{k}\delta_j\mathbf{e}_h^{l-j}, \mathbf{v}_h\Big)_h\Big| \lesssim \frac{1}{\tau}\|\widetilde{\mathbf{e}}_h^l\|_{L^2}\|\mathbf{v}_h\|_{L^2} + \frac{1}{\tau}\sum_{j=1}^{k}\|\mathbf{e}_h^{l-j}\|_{L^2}\|\mathbf{v}_h\|_{L^2} \lesssim \tau^{\frac{4}{5}}\|\mathbf{v}_h\|_{L^2}. \tag{3.40}$$

As a result, we have

$$\begin{aligned}
&\Big|\Big(A_h(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_l))(\nabla I_h\widehat{\mathbf{m}}(t_l), \nabla I_h\widehat{\mathbf{m}}(t_l)) - A_h(\mathbf{a}\circ\widehat{\mathbf{m}}_h^l)(\nabla\widehat{\mathbf{m}}_h^l, \nabla\widehat{\mathbf{m}}_h^l), \mathbf{v}_h\Big)\Big| \\
&\leqslant \Big|\Big(\nabla I_h\widehat{\mathbf{m}}(t_l)\cdot\big(I_h\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_l)) - I_h\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^l)\big)\nabla I_h\widehat{\mathbf{m}}(t_l), I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^l)\cdot\mathbf{v}_h\Big)\Big| \\
&\quad + \Big|\Big(\nabla I_h\widehat{\mathbf{m}}(t_l)\cdot I_h\mathbf{H}(\mathbf{a}\circ I_h\widehat{\mathbf{m}}(t_l))\nabla I_h\widehat{\mathbf{m}}(t_l), \big(I_h\boldsymbol{\nu}(I_h\widehat{\mathbf{m}}(t_l)) - I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^l)\big)\cdot\mathbf{v}_h\Big)\Big| \\
&\quad + \Big|\Big(\nabla I_h\widehat{\mathbf{m}}(t_l)\cdot I_h\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^l)\nabla I_h\widehat{\mathbf{m}}(t_l) - \nabla\widehat{\mathbf{m}}_h^l\cdot I_h\mathbf{H}(\mathbf{a}\circ\widehat{\mathbf{m}}_h^l)\nabla\widehat{\mathbf{m}}_h^l, I_h\boldsymbol{\nu}(\widehat{\mathbf{m}}_h^l)\cdot\mathbf{v}_h\Big)\Big| \\
&\lesssim \sum_{j=0}^{k-1}\|\mathbf{e}_h^{l-j-1}\|_{L^\infty}\|\mathbf{v}_h\|_{L^2} + \sum_{j=0}^{k-1}(\|\nabla\widetilde{\mathbf{e}}_h^{l-j-1}\|_{L^2} + \|\nabla\widetilde{\mathbf{e}}_h^{l-j-1}\|_{L^4}^2)\|\mathbf{v}_h\|_{L^2} \\
&\lesssim \tau\|\mathbf{v}_h\|_{L^2} + (h^r + \tau^{\frac{13}{10}} + \tau^{\frac{4}{3}})\|\mathbf{v}_h\|_{L^2} \\
&\lesssim \tau\|\mathbf{v}_h\|_{L^2}, \quad \text{(the condition } h^{r+1} \lesssim \tau^k \text{ implies } h^r \lesssim \tau \text{ for } k \geqslant 2 \text{ and } r \geqslant 1\text{).}
\end{aligned} \tag{3.41}$$

In the second to last inequality we have used (3.13) to estimate $\|\mathbf{e}_h^{l-j-1}\|_{L^\infty}$ and (3.36) to estimate $\|\nabla\widetilde{\mathbf{e}}_h^{l-j-1}\|_{L^2}$, as well as Lemma 3.6 to estimate $\|\nabla\widetilde{\mathbf{e}}_h^{l-j-1}\|_{L^4}^2$.

By substituting (3.39)–(3.41) into (3.37) and using the duality argument, we obtain the following result:

$$\|\Delta_h\widetilde{\mathbf{e}}_h^l\|_{L^2} = \sup_{\mathbf{v}_h\in\mathbf{S}_h^r}\frac{|(\Delta_h\widetilde{\mathbf{e}}_h^l, \mathbf{v}_h)|}{\|\mathbf{v}_h\|_{L^2}} \lesssim \tau^{\frac{4}{5}} + \tau\|\widetilde{\mathbf{e}}_h^l\|_{L^\infty} \lesssim \tau^{\frac{4}{5}} + \tau\|\Delta_h\widetilde{\mathbf{e}}_h^l\|_{L^2}, \tag{3.42}$$

where the last step is due to the discrete Sobolev embedding inequality ([35, Lemma 3.5]). For sufficiently small stepsize τ, we obtain

$$\|\Delta_h\widetilde{\mathbf{e}}_h^l\|_{L^2} \lesssim \tau^{\frac{4}{5}} \leqslant \tau^{\frac{2}{3}}.$$

Then, by using the discrete Sobolev interpolation inequality ([35, Lemma 3.5]), together with (3.36), we have

$$\|\widetilde{\mathbf{e}}_h^l\|_{L^\infty} \lesssim (\|\Delta_h\widetilde{\mathbf{e}}_h^l\|_{L^2} + \|\widetilde{\mathbf{e}}_h^l\|_{L^2})^{\frac{d}{4}}\|\widetilde{\mathbf{e}}_h^l\|_{L^2}^{1-\frac{d}{4}} \lesssim (\tau^{\frac{4}{5}} + \tau^{\frac{9}{5}})^{\frac{d}{4}}\tau^{\frac{9}{5}(1-\frac{d}{4})} \lesssim \tau^{\frac{9}{5}-\frac{d}{4}} \leqslant \tau \tag{3.43}$$

for $d = 1, 2, 3$. This proves (3.11) and (3.12) for $n = l + 1$.

3.6. Error estimates and proof of (3.13) for $n = l + 1$

Before proceeding further, we turn to prove the almost orthogonality relation between the two types of error functions. This lemma plays a crucial role later in proving the optimal-order convergence.

Lemma 3.8 (Almost orthogonality)**.** *If (3.11) holds for $k \leqslant n \leqslant l+1$, then the following inequality holds:*

$$|(\mathbf{e}_h^{n-j}, \widetilde{\mathbf{e}}_h^{n-i} - \mathbf{e}_h^{n-i})_h| \lesssim \tau(\|\mathbf{e}_h^{n-j}\|_{L_h^2}^2 + \|\widetilde{\mathbf{e}}_h^{n-i}\|_{L_h^2}^2 + \|\mathbf{e}_h^{n-i}\|_{L_h^2}^2) \tag{3.44}$$

for all $k \leqslant n \leqslant l$ and $0 \leqslant i, j \leqslant k$.

Proof. In the proof, we consider a more general setting where the target surface $\Gamma \subset \mathbb{R}^{\mathcal{N}+1}$ is of arbitrary co-dimension. We denote the dimension of Γ as $\dim(\Gamma) = \mathcal{N}'$ with $1 \leqslant \mathcal{N}' \leqslant \mathcal{N}$.

In the first step we show that Γ can be locally identified as a vector-valued graph which takes values in $\mathbb{R}^{\mathcal{N}+1-\mathcal{N}'}$. For any point $p \in \Gamma$, we denote the chart in which p lives as (U_p, φ_p), with $\varphi_p^{-1} : \mathbb{R}^{\mathcal{N}'} \supset \varphi_p(U_p) \to \Gamma \subset \mathbb{R}^{\mathcal{N}+1}$ being a parametrization of U_p. Since φ_p is a C^∞-diffeomorphism, it follows that $\mathrm{d}\varphi_p^{-1} : \mathbb{R}^{\mathcal{N}'} \to \mathbb{R}^{\mathcal{N}+1}$ is injective. In other words, if we denote the coordinate systems on $\varphi_p(U_p)$ and U_p as $\{x_i\}_{i=1}^{\mathcal{N}'}$ and $\{y_i\}_{i=1}^{\mathcal{N}+1}$ respectively, the injectivity implies that the Jacobian matrix

$$J(p) := \begin{pmatrix} \partial_{x_1}(\varphi_p^{-1})^1 & \partial_{x_2}(\varphi_p^{-1})^1 & \dots & \partial_{x_{\mathcal{N}'}}(\varphi_p^{-1})^1 \\ \partial_{x_1}(\varphi_p^{-1})^2 & \partial_{x_2}(\varphi_p^{-1})^2 & \dots & \partial_{x_{\mathcal{N}'}}(\varphi_p^{-1})^2 \\ \vdots & \vdots & \ddots & \vdots \\ \partial_{x_1}(\varphi_p^{-1})^{\mathcal{N}+1} & \partial_{x_2}(\varphi_p^{-1})^{\mathcal{N}+1} & \dots & \partial_{x_{\mathcal{N}'}}(\varphi_p^{-1})^{\mathcal{N}+1} \end{pmatrix} \in \mathbb{R}^{(\mathcal{N}+1)\times\mathcal{N}'} \tag{3.45}$$

is of full rank $\mathcal{N}'$. Equivalently, there exists an $(\mathcal{N}'\times\mathcal{N}')$-minor which is invertible. We denote the row labels of this minor as $k_1, ..., k_{\mathcal{N}'}$, and denote the minor as

$$J_{\mathcal{N}'}(p) := \left(\partial_j(\varphi_p^{-1})^{k_i}\right)_{i,j=1,...,\mathcal{N}'}.$$

Then we define the coordinate projection associated to this set of labels as

$$\Pi_p : \mathbb{R}^{\mathcal{N}+1} \to \mathbb{R}^{\mathcal{N}'}, (y_1, ..., y_{\mathcal{N}+1}) \mapsto (y_{k_1}, ..., y_{k_{\mathcal{N}'}}).$$

Therefore, $J_{\mathcal{N}'}(p)$ is the Jacobian of $\Pi_p \circ \varphi_p^{-1} : \mathbb{R}^{\mathcal{N}'} \to \mathbb{R}^{\mathcal{N}'}$ at point p. By the invertibility and inverse function theorem we know that $\Pi_p \circ \varphi_p^{-1}$ is a diffeomorphism on some neighbourhood $W_p \subset \varphi_p(U_p)$ of p. This means that $\Pi_p^{-1} : \Pi_p \circ \varphi_p^{-1}(W_p) \to \varphi_p^{-1}(W_p) \subset \mathbb{R}^{\mathcal{N}+1}$ is a diffeomorphism and thus Π_p^{-1} is the local vector-valued graph around point p taking values in $\mathbb{R}^{\mathcal{N}+1-\mathcal{N}'}$. Therefore, we have shown that for any $p \in \Gamma$, there exists a neighborhood V_p of p in Γ such that V_p is the graph of a smooth function which has one of the following $\binom{\mathcal{N}+1}{\mathcal{N}'}$ forms: $(y_{k_1}, ..., y_{k_{\mathcal{N}'}})^\wedge = f(y_{k_1}, ..., y_{k_{\mathcal{N}'}})$ where $\{k_1, ..., k_{\mathcal{N}'}\} \subseteq \{1, ..., \mathcal{N}+1\}$ are $\mathcal{N}'$ different labels determined by the process above.

In the second step, we are going to show that there exists $r > 0$ such that for any $p \in \Gamma$, $\Gamma \cap B_r(p)$ can be identified as a vector-valued graph on $T\Gamma_p$.

Up to linear transformation, we can assume $T\Gamma_p = \mathbb{R}^{\mathcal{N}'} \times \{0\}^{\mathcal{N}+1-\mathcal{N}'}$ without loss of generality. From the previous step, the only choice now is $(y_{\mathcal{N}'+1}, ..., y_{\mathcal{N}+1}) = f(y_1, ..., y_{\mathcal{N}'})$ otherwise the gradient of the graph will blow up at p contradicting with the smoothness. Then we denote $B_{r_p}(p)$ as the ball such that $\Gamma \cap B_{r_p}(p) \subset V_p$ according to the previous step. If $\inf_{p\in\Gamma} r_p = 0$, we can pick out a sequence such that $p_i \to p_0$ with $r_{p_i} \to 0$ which contradicts with the fact that $r_{p_0} > 0$. So we may set $r = \inf_{p\in\Gamma} r_p > 0$. Hence this step is complete.

In the last step we are ready to conclude this lemma. Let r be that in step 2. For any $p \in \Gamma$ we define the corresponding bijective projection in step 1 as $\Pi_p : \Gamma \cap B_r(p) \to T\Gamma_p$. Using this projection, we can define the smooth graph function associated to point p as $g_p := \Pi_p^{-1} : \Pi_p(\Gamma \cap B_r(p)) \to \Gamma \cap B_r(p)$ and moreover define the following two domains

$$\mathcal{D}_p := \Pi_p(\Gamma \cap B_{r/2}(p)) \quad \text{and} \quad \mathcal{D}_p^* := \Pi_p(\Gamma \cap B_{2r/3}(p)).$$

If τ is sufficiently small, then by the induction assumption in (3.11) and the estiamte in (3.43) we know that, at each finite element node $x \in \mathrm{Nodes}(\mathbf{S}_h^r) \subseteq \bar{\Omega}$, $\mathbf{m}_h^{n-j}(x)$ and $I_h\mathbf{m}(x, t_{n-j})$ are both on the surface Γ and contained in $\Gamma \cap B_{r/4}(\mathbf{m}_h^n(x))$ and therefore, by the triangle inequality, contained in $\Gamma \cap B_{r/2}(\mathbf{m}_h^{n-i}(x))$ for any $0 \leqslant i, j \leqslant k$. We now work locally on the

graph of $\Gamma \cap B_r(\mathbf{m}_h^{n-i}(x))$. It holds that

$$\mathbf{e}_h^{n-j} \cdot (\widetilde{\mathbf{e}}_h^{n-i} - \mathbf{e}_h^{n-i}) = |\mathbf{e}_h^{n-j}| \cdot |\widetilde{\mathbf{e}}_h^{n-i} - \mathbf{e}_h^{n-i}| \sin\gamma, \tag{3.46}$$

where γ is the angle between $\mathbf{e}_h^{n-j}$ and $T\Gamma_{\mathbf{m}_h^{n-i}(x)}$. Then by applying Taylor's theorem around $\Pi_{\mathbf{m}_h^{n-i}} \circ I_h\mathbf{m}(t_{n-j})$ and $\Pi_{\mathbf{m}_h^{n-i}} \circ \mathbf{m}_h^{n-i}$ successively, at each element node $x \in \mathrm{Nodes}(\mathbf{S}_h^r)$, we have

$$\begin{aligned}
|\sin\gamma| &= |g_{\mathbf{m}_h^{n-i}(x)}(\Pi_{\mathbf{m}_h^{n-i}(x)} \circ \mathbf{m}_h^{n-j}(x)) - g_{\mathbf{m}_h^{n-i}(x)}(\Pi_{\mathbf{m}_h^{n-i}(x)} \circ I_h\mathbf{m}(x, t_{n-j}))|/|\mathbf{e}_h^{n-i}| \\
&\lesssim |\nabla g_{\mathbf{m}_h^{n-i}(x)}(\Pi_{\mathbf{m}_h^{n-i}(x)} \circ I_h\mathbf{m}(x, t_{n-j}))| + \max_{z \in \bar{\mathcal{D}}_{\mathbf{m}_h^{n-i}(x)}} |\nabla^2 g_{\mathbf{m}_h^{n-i}(x)}(z)| \cdot |\mathbf{e}_h^{n-i}| \\
&\lesssim \max_{z \in \bar{\mathcal{D}}_{\mathbf{m}_h^{n-i}(x)}} |\nabla^2 g_{\mathbf{m}_h^{n-i}(x)}(z)| \cdot (|\mathbf{m}(x, t_{n-j}) - \mathbf{m}(x, t_{n-i})| + |\mathbf{e}_h^{n-i}|) \\
&\lesssim \max_{z \in \bar{\mathcal{D}}_{\mathbf{m}_h^{n-i}(x)}} |\nabla^2 g_{\mathbf{m}_h^{n-i}(x)}(z)| \cdot \tau.
\end{aligned} \tag{3.47}$$

The second to last inequality of (3.47) follows from the fact that we are considering the graph with domain $T\Gamma_{\mathbf{m}_h^{n-i}(x)}$, so the gradient of this graph at $\mathbf{m}_h^{n-i}(x)$ vanishes by the tangency. The last inequality follows from $\|\mathbf{e}_h^{n-i}\|_{L^\infty} \lesssim \tau$, as shown in (3.43).

The last thing we need to show is that $\max_{w\in\Gamma} \max_{z\in\bar{\mathcal{D}}_w} |\nabla^2 g_w(z)| < +\infty$. Note that this is a pure geometric quantity, so it doesn't depend on τ. Suppose that it is not true. Then, by the compactness of Γ, there exist sequences $w_i \in \Gamma$ and $z_i \in \bar{\mathcal{D}}_{w_i}$ such that $w_i \to w_0 \in \Gamma$ for some w_0 and $|\nabla^2 g_{w_i}(z_i)| \to +\infty$. We take sufficiently large i such that $|w_i - w_0| < \frac{1}{6}r$. Then $\Gamma \cap B_{r/2}(w_i) \subset \Gamma \cap B_{2r/3}(w_0)$ is also a graph locally on $T\Gamma_{w_0}$. This means the quantity $|\nabla^2 g_{w_i}(z_i)|$ can not blow up because of the smoothness of the local graph of $\Gamma \cap B_{2r/3}(w_0)$, leading to contradiction.

Substituting the boundedness of $\max_{w\in\Gamma} \max_{z\in\bar{\mathcal{D}}_w} |\nabla^2 g_w(z)|$ into (3.46)–(3.47), we obtain

$$|\mathbf{e}_h^{n-j} \cdot (\widetilde{\mathbf{e}}_h^{n-i} - \mathbf{e}_h^{n-i})| \lesssim \tau |\mathbf{e}_h^{n-j}| \cdot |\widetilde{\mathbf{e}}_h^{n-i} - \mathbf{e}_h^{n-i}| \quad \text{at the nodes.}$$

Then the result of Lemma 3.8 follows from the application of Young's inequality. □

Remark 3.1. Since this technical lemma holds for the surface Γ with arbitrary co-dimension, it will allow us to extend the results in this paper to the surface of arbitrary co-dimension without any difficulty. Many equations of interests fall into this category. For example, the constraint of Yang-Mills equation on $\Omega \subset \mathbb{R}^d$, i.e., $\mathfrak{so}(d) \hookrightarrow \mathbb{R}^{d\times d}$, has co-dimension $\frac{(d-1)d}{2}$.

In the following, we choose $\mathbf{v}_h = \widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^n$ in the error equation (3.17), i.e.,

$$\begin{aligned}
&\Big(\frac{\delta_0}{\tau}\widetilde{\mathbf{e}}_h^n + \frac{1}{\tau}\sum_{j=1}^{k} \delta_j \mathbf{e}_h^{n-j}, \widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}\Big)_h - (\Delta_h \widetilde{\mathbf{e}}_h^n, \widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}) \\
&= \mathcal{E}_1(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}) + \mathcal{E}_2(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}) + \mathcal{E}_3(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}) + \mathcal{E}_4(\widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1})
\end{aligned} \tag{3.48}$$

for $k \leqslant n \leqslant l$. If we denote

$$\eta^j = \widetilde{\mathbf{e}}_h^j - \mathbf{e}_h^j$$

for $0 \leqslant j \leqslant l$, then replacing $\widetilde{\mathbf{e}}_h^i$ with $\eta^i + \mathbf{e}_h^i$ for $i \in \{n-1, n\}$ in the first term of the left-hand side of equation (3.48) provides

$$\begin{aligned}
\Big(\delta_0\widetilde{\mathbf{e}}_h^n + \sum_{j=1}^{k} \delta_j \mathbf{e}_h^{n-j}, \widetilde{\mathbf{e}}_h^n - \theta_k \widetilde{\mathbf{e}}_h^{n-1}\Big)_h &= \Big(\delta_0\eta^n + \sum_{j=0}^{k} \delta_j \mathbf{e}_h^{n-j}, (\eta^n - \theta_k\eta^{n-1}) + (\mathbf{e}_h^n - \theta_k \mathbf{e}_h^{n-1})\Big)_h \\
&= (\delta_0\eta^n, \eta^n - \theta_k\eta^{n-1})_h + \Big(\sum_{j=0}^{k} \delta_j \mathbf{e}_h^{n-j}, \eta^n - \theta_k\eta^{n-1}\Big)_h \\
&\quad + (\delta_0\eta^n, \mathbf{e}_h^n - \theta_k \mathbf{e}_h^{n-1})_h + \Big(\sum_{j=0}^{k} \delta_j \mathbf{e}_h^{n-j}, \mathbf{e}_h^n - \theta_k \mathbf{e}_h^{n-1}\Big)_h \\
&=: \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4.
\end{aligned}$$

Since $0<\delta_0$ and $0\leqslant\theta_k<1$ for $2\leqslant k\leqslant 5$, it follows that

$$\begin{aligned}\mathcal{I}_1&\geqslant\delta_0\|\eta^n\|_{L_h^2}^2-\frac{\delta_0\theta_k}{2}(\|\eta^n\|_{L_h^2}^2+\|\eta^{n-1}\|_{L_h^2}^2)\\&=\delta_0(1-\theta_k)\|\eta^n\|_{L_h^2}^2+\frac{\delta_0\theta_k}{2}(\|\eta^n\|_{L_h^2}^2-\|\eta^{n-1}\|_{L_h^2}^2)\end{aligned}\tag{3.49}$$

for $k\leqslant n\leqslant l$. And from Lemma 3.3 we know that

$$\mathcal{I}_4\geqslant\sum_{i,j=1}^{k}g_{ij}(\mathbf{e}_h^{n-k+i},\mathbf{e}_h^{n-k+j})_h-\sum_{i,j=1}^{k}g_{ij}(\mathbf{e}_h^{n-k+i-1},\mathbf{e}_h^{n-k+j-1})_h\tag{3.50}$$

for all $k\leqslant n\leqslant l$, where $(g_{ij})_{i,j=1}^k\in\mathbb{R}^{k\times k}$ is a given positive definite matrix.

On the other hand, by using $\eta^i=\widetilde{\mathbf{e}}_h^i-\mathbf{e}_h^i$ for $i\in\{n-1,n\}$, we have

$$\begin{aligned}|\mathcal{I}_2|&\leqslant\sum_{j=0}^{k}|\delta_j|\,|(\mathbf{e}_h^{n-j},\eta^n)_h|+\sum_{j=0}^{k}|\delta_j\theta_k|\,|(\mathbf{e}_h^{n-j},\eta^{n-1})_h|\\&=\sum_{j=0}^{k}|\delta_j|\,|(\mathbf{e}_h^{n-j},\widetilde{\mathbf{e}}_h^n-\mathbf{e}_h^n)_h|+\sum_{j=0}^{k}|\delta_j\theta_k|\,|(\mathbf{e}_h^{n-j},\widetilde{\mathbf{e}}_h^{n-1}-\mathbf{e}_h^{n-1})_h|\end{aligned}$$

for all $k\leqslant n\leqslant l$. Recall that $\|\widetilde{\mathbf{e}}_h^i\|_{L^\infty}\leqslant\tau$ holds for all $0\leqslant i\leqslant l$, which is proved by the mathematical induction method, namely it is firstly assumed by the induction hypothesis (3.11) for $0\leqslant i\leqslant l-1$ and then proved by Subsection 3.5 for $i=l$. Hence it follows from Lemma 3.8 that

$$\begin{aligned}|\mathcal{I}_2|&\lesssim\sum_{j=0}^{k}\tau(\|\mathbf{e}_h^{n-j}\|_{L_h^2}^2+\|\widetilde{\mathbf{e}}_h^n\|_{L_h^2}^2+\|\mathbf{e}_h^n\|_{L_h^2}^2)+\sum_{j=0}^{k}\tau(\|\mathbf{e}_h^{n-j}\|_{L_h^2}^2+\|\widetilde{\mathbf{e}}_h^{n-1}\|_{L_h^2}^2+\|\mathbf{e}_h^{n-1}\|_{L_h^2}^2)\\&\lesssim\tau\sum_{j=0}^{k}\|\mathbf{e}_h^{n-j}\|_{L_h^2}^2+\tau\|\widetilde{\mathbf{e}}_h^n\|_{L_h^2}^2+\tau\|\widetilde{\mathbf{e}}_h^{n-1}\|_{L_h^2}^2\end{aligned}\tag{3.51}$$

for $k\leqslant n\leqslant l$. Similarly,

$$\begin{aligned}|\mathcal{I}_3|\leqslant\delta_0|\big(\mathbf{e}_h^n,\eta^n\big)_h|+\delta_0\theta_k|\big(\mathbf{e}_h^{n-1},\eta^n\big)_h|&=\delta_0|\big(\mathbf{e}_h^n,\widetilde{\mathbf{e}}_h^n-\mathbf{e}_h^n\big)_h|+\delta_0\theta_k|\big(\mathbf{e}_h^{n-1},\widetilde{\mathbf{e}}_h^n-\mathbf{e}_h^n\big)_h|\\&\lesssim\tau(\|\mathbf{e}_h^n\|_{L_h^2}^2+\|\widetilde{\mathbf{e}}_h^n\|_{L_h^2}^2+\|\mathbf{e}_h^{n-1}\|_{L_h^2}^2)\end{aligned}\tag{3.52}$$

holds for all $k\leqslant n\leqslant l$. By (3.51) and (3.52) we have

$$|\mathcal{I}_2|+|\mathcal{I}_3|\lesssim\tau\sum_{j=0}^{k}\|\mathbf{e}_h^{n-j}\|_{L^2}^2+\tau\|\widetilde{\mathbf{e}}_h^n\|_{L^2}^2+\tau\|\widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}^2,\tag{3.53}$$

where the norm equivalence relation (2.2) has been used.

Combining the error equation (3.48) and the estimates of $|\mathcal{E}_j(\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1})|$, $j=1,2,3,4$, in (3.18), (3.22) and (3.32), and noting that

$$\begin{aligned}-(\Delta_h\widetilde{\mathbf{e}}_h^n,\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1})&=\big(\nabla\widetilde{\mathbf{e}}_h^n,\nabla(\widetilde{\mathbf{e}}_h^n-\theta_k\widetilde{\mathbf{e}}_h^{n-1})\big)\\&\geqslant(1-\theta_k)\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2+\frac{\theta_k}{2}(\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2-\|\nabla\widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}^2),\end{aligned}$$

we obtain the following result for $k\leqslant n\leqslant l$:

$$\begin{aligned}&\frac{\mathcal{I}_1}{\tau}+\frac{\mathcal{I}_4}{\tau}+(1-\theta_k)\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2+\frac{\theta_k}{2}(\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2-\|\nabla\widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}^2)\\&\leqslant C\frac{|\mathcal{I}_2|+|\mathcal{I}_3|}{\tau}+C\tau^{2k}+Ch^{2(r+1)}+C\sum_{j=0}^{k}\|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2+\varepsilon\sum_{j=0}^{k}\|\nabla\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2.\end{aligned}$$

Hence, by applying (3.49)–(3.50) and (3.53) we get

$$\frac{\delta_0(1-\theta_k)}{\tau}\|\eta^n\|_{L_h^2}^2+\frac{\delta_0\theta_k}{2\tau}(\|\eta^n\|_{L_h^2}^2-\|\eta^{n-1}\|_{L_h^2}^2)$$

$$
\begin{aligned}
&+\frac{1}{\tau}\sum_{i,j=1}^{k} g_{ij}(\mathbf{e}_h^{n-k+i},\mathbf{e}_h^{n-k+j})_h-\frac{1}{\tau}\sum_{i,j=1}^{k} g_{ij}(\mathbf{e}_h^{n-k+i-1},\mathbf{e}_h^{n-k+j-1})_h\\
&+(1-\theta_k)\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2+\frac{\theta_k}{2}(\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2-\|\nabla\widetilde{\mathbf{e}}_h^{n-1}\|_{L^2}^2)\\
&\leqslant C\tau^{2k}+Ch^{2(r+1)}+C\sum_{j=0}^{k}\|\mathbf{e}_h^{n-j}\|_{L^2}^2+C\sum_{j=0}^{k}\|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2+\varepsilon\sum_{j=0}^{k}\|\nabla\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2
\end{aligned}
$$

for $k\leqslant n\leqslant l$. Then multiplying both sides by τ and summing up this inequality over $n=k,\cdots,q$ with $k\leqslant q\leqslant l$ lead to

$$
\begin{aligned}
&\delta_0(1-\theta_k)\sum_{n=k}^{q}\|\eta^n\|_{L_h^2}^2+\frac{\delta_0\theta_k}{2}(\|\eta^q\|_{L_h^2}^2-\|\eta^{k-1}\|_{L_h^2}^2) \qquad (3.54)\\
&+\sum_{i,j=1}^{k} g_{ij}(\mathbf{e}_h^{q-k+i},\mathbf{e}_h^{q-k+j})_h-\sum_{i,j=1}^{k} g_{ij}(\mathbf{e}_h^{i-1},\mathbf{e}_h^{j-1})_h\\
&+(1-\theta_k)\tau\sum_{n=k}^{q}\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2+\frac{\theta_k\tau}{2}(\|\nabla\widetilde{\mathbf{e}}_h^q\|_{L^2}^2-\|\nabla\widetilde{\mathbf{e}}_h^{k-1}\|_{L^2}^2)\\
&\leqslant C\tau^{2k}+Ch^{2(r+1)}+C\tau\sum_{n=k}^{q}\sum_{j=0}^{k}(\|\mathbf{e}_h^{n-j}\|_{L^2}^2+\|\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2)+\varepsilon\tau\sum_{n=k}^{q}\sum_{j=0}^{k}\|\nabla\widetilde{\mathbf{e}}_h^{n-j}\|_{L^2}^2. \qquad (3.55)
\end{aligned}
$$

Since the matrix $(g_{ij})_{i,j=1}^k$ is positive definite, it follows that

$$
\sum_{i,j=1}^{k} g_{ij}(\mathbf{e}_h^{q-k+i},\mathbf{e}_h^{q-k+j})_h\geqslant\lambda_1\|\mathbf{e}_h^q\|_{L_h^2}^2\quad\text{and}\quad\sum_{i,j=1}^{k} g_{ij}(\mathbf{e}_h^{i-1},\mathbf{e}_h^{j-1})_h\lesssim\sum_{j=1}^{k}\|\mathbf{e}_h^{j-1}\|_{L_h^2}^2,
$$

where λ_1 is the smallest eigenvalue of $(g_{ij})_{i,j=1}^k$. Using these results and rearranging (3.55) yield

$$
\begin{aligned}
&\delta_0(1-\theta_k)\sum_{n=k}^{q}\|\eta^n\|_{L_h^2}^2+\frac{\delta_0\theta_k}{2}\|\eta^q\|_{L_h^2}^2+\lambda_1\|\mathbf{e}_h^q\|_{L_h^2}^2+(1-\theta_k)\tau\sum_{n=k}^{q}\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2+\frac{\theta_k\tau}{2}\|\nabla\widetilde{\mathbf{e}}_h^q\|_{L^2}^2\\
&\leqslant C\tau^{2k}+Ch^{2(r+1)}+C\tau\sum_{n=0}^{q}(\|\mathbf{e}_h^n\|_{L^2}^2+\|\widetilde{\mathbf{e}}_h^n\|_{L^2}^2)+\varepsilon\tau\sum_{n=0}^{q}\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2\\
&+\frac{\delta_0\theta_k}{2}\|\eta^{k-1}\|_{L_h^2}^2+C\sum_{j=1}^{k}\|\mathbf{e}_h^{j-1}\|_{L_h^2}^2+\frac{\theta_k\tau}{2}\|\nabla\widetilde{\mathbf{e}}_h^{k-1}\|_{L^2}^2 \qquad (3.56)
\end{aligned}
$$

for $k\leqslant q\leqslant l$. Moreover, we can duduce from

$$
\|\widetilde{\mathbf{e}}_h^q\|_{L_h^2}^2\leqslant 2(\|\mathbf{e}_h^q\|_{L_h^2}^2+\|\widetilde{\mathbf{e}}_h^q-\mathbf{e}_h^q\|_{L_h^2}^2)=2(\|\mathbf{e}_h^q\|_{L_h^2}^2+\|\eta^q\|_{L_h^2}^2),
$$

that

$$
\begin{aligned}
&\delta_0(1-\theta_k)\sum_{n=k}^{q}\|\eta^n\|_{L_h^2}^2+\frac{\delta_0\theta_k}{2}\|\eta^q\|_{L_h^2}^2+\lambda_1\|\mathbf{e}_h^q\|_{L_h^2}^2\\
&\geqslant\delta_0(1-\theta_k/2)\|\eta^q\|_{L_h^2}^2+\lambda_1\|\mathbf{e}_h^q\|_{L_h^2}^2\\
&\geqslant\begin{cases}\frac{\lambda_1}{4}\|\widetilde{\mathbf{e}}_h^q\|_{L_h^2}+\frac{\lambda_1}{2}\|\mathbf{e}_h^q\|_{L_h^2}, & \text{if } \delta_0(1-\theta_k/2)\geqslant\frac{\lambda_1}{2},\\ \delta_0(1/2-\theta_k/4)\|\widetilde{\mathbf{e}}_h^q\|_{L_h^2}+\frac{\lambda_1}{2}\|\mathbf{e}_h^q\|_{L_h^2}, & \text{if } \delta_0(1-\theta_k/2)<\frac{\lambda_1}{2}.\end{cases}
\end{aligned}
$$

Due to $\lambda_1,\delta_0>0$ and $0\leqslant\theta_k<1$ for all $k=2,3,4,5$, we substitute this into (3.56) and get

$$
\begin{aligned}
&\|\widetilde{\mathbf{e}}_h^q\|_{L^2}^2+\|\mathbf{e}_h^q\|_{L^2}^2+\tau\sum_{n=k}^{q}\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2\\
&\leqslant C\tau^{2k}+Ch^{2(r+1)}+C\tau\sum_{n=0}^{q}(\|\mathbf{e}_h^n\|_{L^2}^2+\|\widetilde{\mathbf{e}}_h^n\|_{L^2}^2)+C\varepsilon\tau\sum_{n=0}^{q}\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2
\end{aligned}
$$

$$+ C\|\eta^{k-1}\|_{L^2}^2 + C\sum_{j=1}^{k}\|\mathbf{e}_h^{j-1}\|_{L^2}^2 + C\tau\|\nabla\widetilde{\mathbf{e}}_h^{k-1}\|_{L^2}^2$$

for $k \leqslant q \leqslant l$, where we have also used the norm equivalence (2.2). Therefore a sufficiently small ε and the discrete Gronwall inequality give

$$\begin{aligned}
&\|\widetilde{\mathbf{e}}_h^q\|_{L^2}^2 + \|\mathbf{e}_h^q\|_{L^2}^2 + \tau\sum_{n=k}^{q}\|\nabla\widetilde{\mathbf{e}}_h^n\|_{L^2}^2 \\
&\leqslant C\tau^{2k} + Ch^{2(r+1)} + C\|\eta^{k-1}\|_{L^2}^2 + C\sum_{j=1}^{k}(\|\mathbf{e}_h^{j-1}\|_{L^2}^2 + \|\widetilde{\mathbf{e}}_h^{j-1}\|_{L^2}^2 + \tau\|\nabla\widetilde{\mathbf{e}}_h^{j-1}\|_{L^2}^2) \\
&\leqslant C\tau^{2k} + Ch^{2(r+1)}
\end{aligned}$$

for all $k \leqslant q \leqslant l$, where the last inequality has employed the condition (C2). This not only proves (3.13) for $n = l+1$ (when τ and h are sufficiently small) but also completes the proof of Theorem 2.2 in view of the interpolation error estimates in Lemma 3.1.

4. Numerical Results

In this section we present numerical results to support the theoretical analysis for the optimal-order convergence of the k-step projection method. All the computations are performed by Firedrake [47] with double precision.

We consider the harmonic map heat flow from a rectangular domain $\Omega = [0,1]\times[0,1]$ to the two-dimensional unit sphere $\Gamma = \mathbb{S}^2 \subset \mathbb{R}^3$ and the three-dimensional unit sphere $\Gamma = \mathbb{S}^3 \subset \mathbb{R}^4$, respectively. The corresponding initial functions are chosen as $\mathbf{m}^0 = \frac{1}{S}[m_1^0, m_2^0, m_3^0]^\top$ and $\overline{\mathbf{m}}^0 = \frac{1}{M}[m_1^0, m_2^0, m_3^0, m_4^0]^\top$ where

$$\begin{aligned}
m_1^0(x,y) &= e^{-x} + \frac{(1-e)}{2e}x^2 + x, \\
m_2^0(x,y) &= e^{-\frac{1}{2}y}(y^2+3y+6), \\
m_3^0(x,y) &= e^{-\frac{1}{2}y}\big(e^{-x} + \frac{1-e}{2e}x^2 + x\big)(y^2+3y+6), \\
m_4^0(x,y) &= e^{-\frac{1}{2}x}(x^2+3x+6)\big(-\ln(y+1) - \frac{y^2}{4} + y + 4\big),
\end{aligned}$$

and

$$\begin{aligned}
S(x,y) &= \sqrt{m_1^0(x,y)^2 + m_2^0(x,y)^2 + m_3^0(x,y)^2}, \\
M(x,y) &= \sqrt{m_1^0(x,y)^2 + m_2^0(x,y)^2 + m_3^0(x,y)^2 + m_4^0(x,y)^2}.
\end{aligned}$$

Clearly, $\mathbf{m}^0 \in \mathbb{S}^2$ and $\overline{\mathbf{m}}^0 \in \mathbb{S}^3$ both satisfy the boundary condition (1.2). With the two initial functions, we solve problem (1.1)–(1.3) by the proposed k-step projection method (2.15)–(2.16) up to time $T = 0.5$.

And the starting values are calculated by l-step ($1 \leqslant l \leqslant k-1$) projection methods (2.15)–(2.16) with sufficiently small time stepsizes so that the temporal k-th convergence order can be maintained.

We compare the numerical solution with the reference solution to calculate both temporal and spatial discretization errors. The corresponding L^2 errors of the numerical solutions at time $T = 0.5$ are respectively depicted in Figures 1–4.

To verify the temporal convergence rates, we set the reference time step $\tau_{\text{ref}} = \frac{1}{1000}$, and calculate the errors for $\tau = \frac{1}{100}, \frac{1}{120}, \frac{1}{140}, \frac{1}{160}$, with sufficiently small mesh sizes used to ensure that spatial discretization errors are negligible. To access the spatial convergence orders, we set the reference mesh size $h_{\text{ref}} = \frac{1}{150}$, and calculate the errors for $h = \frac{1}{15}, \frac{1}{20}, \frac{1}{25}, \frac{1}{30}$, with high-order BDF temporal schemes and adequately small time steps empolyed so that the temporal discretization errors do not impact the observation of optimal spatial convergence.

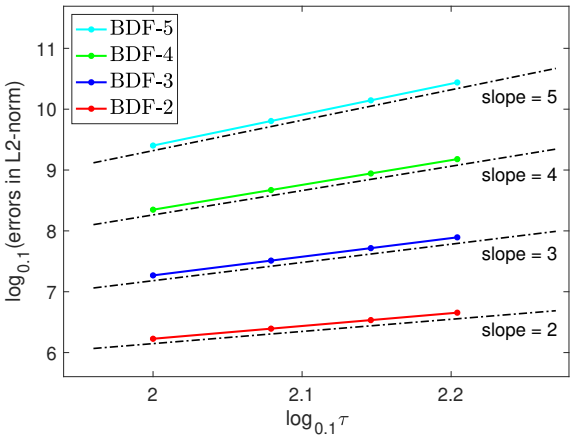

FIGURE 1. Time discretization errors for $\Gamma = \mathbb{S}^2 \subset \mathbb{R}^3$

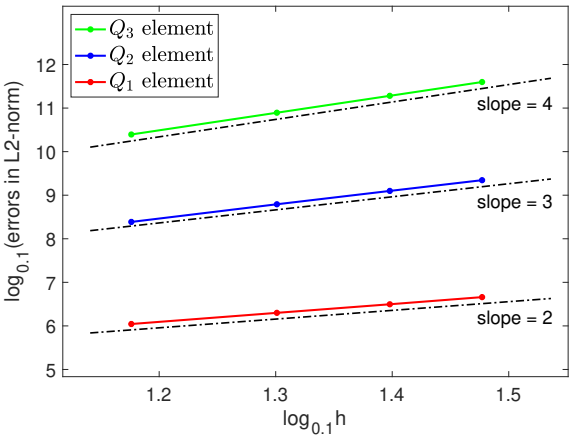

FIGURE 2. Spatial discretization errors for $\Gamma = \mathbb{S}^2 \subset \mathbb{R}^3$

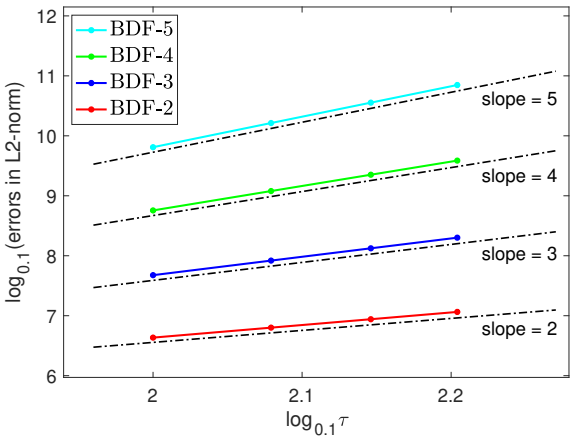

FIGURE 3. Time discretization errors for $\Gamma = \mathbb{S}^3 \subset \mathbb{R}^4$

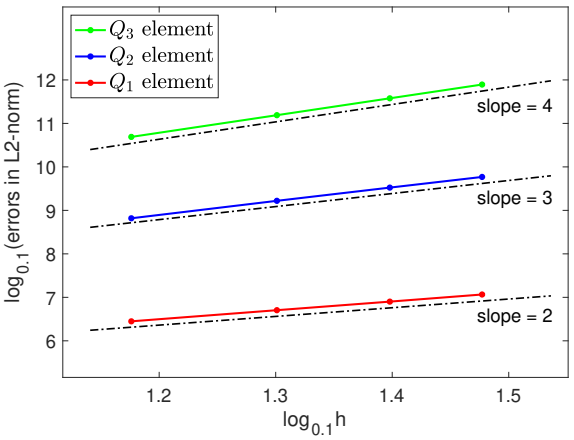

FIGURE 4. Spatial discretization errors for $\Gamma = \mathbb{S}^3 \subset \mathbb{R}^4$

From Figures 1–4 we see that the temporal and spatial discretization errors are $O(\tau^k)$ and $O(h^{r+1})$, respectively. These results are consistent with the theoretical analysis in Theorem 2.2. Moreover, the numerical results in Figure 4 also indicate that the stepsize restriction $\tau^k \gtrsim h^{r+1}$ may not be necessary, though it is required in the error analysis.

5. Conclusions

We have proposed a k-step projection method for solving the harmonic map heat flow from a rectangular domain into a general smooth surface $\Gamma \subset \mathbb{R}^{\mathcal{N}+1}$, and proved the optimal-order convergence of the proposed method under the mesh size condition $h^{r+1} \lesssim \tau^k$, where $r \geqslant 1$ denotes the degree of finite elements in a tensor-product mass-lumping FEM. The proof utilizes a new geometric relation between the error functions associated to the auxiliary and projected numerical solutions. The geometric relation allows us to apply the Nevanlinna–Odeh multiplier technique [4, 45] in the presence of a projection stage. We have only considered an explicit treatment of the nonlinearity for the simplicity of implementation (with time-independent matrix assembling at every time level, thus the matrix only needs to be assembled once and for all time levels), while the convergence analysis actually could be extended to schemes with semi-implicit treatment of the nonlinearity.

We have proved the optimal-order convergence based on mass-lumping techniques on rectangular meshes. The proof can be extended to triangular meshes by using the mass-lumping

techniques in [21, 24] based on the $\widetilde{P}_r$ finite element space for $1 \leqslant r \leqslant 5$. In this case, the quadrature error bound and the superconvergence result should be replaced by

$$|\mathcal{E}_{1,2}(\mathbf{v}_h)| \lesssim h^r \|\mathbf{v}_h\|_{H^1}, \tag{5.1}$$

$$|\mathcal{E}_{1,3}(\mathbf{v}_h)| \lesssim h^r \|\mathbf{v}_h\|_{H^1}, \tag{5.2}$$

respectively, where (5.1) holds due to [24, Lemma 5.2], and (5.2) is the standard approximation property of the Lagrange interpolation. These changes will lead to the following sub-optimal error estimate under the mesh size condition $h^r \lesssim \tau^k$:

$$\max_{k \leqslant n \leqslant N} (\|\mathbf{m}_h^n - \mathbf{m}(t_n)\|_{L^2} + \|\widetilde{\mathbf{m}}_h^n - \mathbf{m}(t_n)\|_{L^2}) \lesssim \tau^k + h^r. \tag{5.3}$$

The extension to curved domain using triangular mesh and iso-parametric finite element method is also possible.

The analysis in this article also applies to unbounded surfaces without boundary, provided that the exact solution of the problem is sufficiently smooth. For example, if Γ is a hyperbolic surface, i.e., $\mathbb{H}^{\mathcal{N}} \subset (\mathbb{R}^{\mathcal{N}+1}, \langle\cdot,\cdot\rangle_L)$ with L being the Lorentz product [41], then the theorems in [26] imply the existence of a unique smooth global solution given smooth initial data. In this case, error estimates can also be obtained for the proposed multistep projection method.

Since the projection method in this article is quite general, it may also be applicable to other geometric partial differential equations with constrained target manifolds. For example, the similar projection method may be applied to the heat flow of Yang–Mills equation on $\Omega \subset \mathbb{R}^d$ whose image is constrained on the Lie algebra $\mathfrak{so}(d)$ [48, 52], as well as the wave map into general smooth surfaces [49].

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China (project no. 12071020). The work of G. Bai and X. Gui was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region, China (GRF project no. PolyU15300920 and PolyU15301321). The work of B. Li was supported in part by a fellowship award from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU/RFS2324-5S03) and an internal grant at The Hong Kong Polytechnic University (Project ID: P0038843).

References

[1] Georgios Akrivis. Stability of implicit–explicit backward difference formulas for nonlinear parabolic equations. *SIAM J. Numer. Anal.*, 53:464–484, 2015.

[2] Georgios Akrivis, Michael Feischl, Balázs Kovács, and Christian Lubich. Higher-order linearly implicit full discretization of the Landau-Lifshitz-Gilbert equation. *Math. Comp.*, 90(329):995–1038, 2021.

[3] Georgios Akrivis and Emmanouil Katsoprinakis. Backward difference formulae: New multipliers and stability properties for parabolic equations. *Math. Comp.*, 85:2195–2216, 2016.

[4] Georgios Akrivis and Christian Lubich. Fully implicit, linearly implicit and implicit-explicit backward difference formulae for quasi-linear parabolic equations. *Numer. Math.*, 131(4):713–735, 2015.

[5] François Alouges. A new finite element scheme for Landau-Lifchitz equations. *Discrete Contin. Dyn. Syst. Ser. S*, 1(2):187–196, 2008.

[6] François Alouges and Pascal Jaisson. Convergence of a finite element discretization for the Landau-Lifshitz equations in micromagnetism. *Math. Models Methods Appl. Sci.*, 16(2):299–316, 2006.

[7] François Alouges and Alain Soyeur. On global weak solutions for Landau–Lifshitz equations: existence and nonuniqueness. *Nonlinear Anal.*, 18:1071–1084, 1992.

[8] François Alouges, Evaggelos Kritsikis, Jutta Steiner, and Jean-Christophe Toussaint. A convergent and precise finite element scheme for Landau–Lifshitz–Gilbert equation. *Numer. Math.*, 128:407–430, 2014.

[9] Rong An, Huadong Gao, and Weiwei Sun. Optimal error analysis of Euler and Crank–Nicolson projection finite difference schemes for Landau–Lifshitz equation. *SIAM J. Numer. Anal.*, 59(3):1639–1662, 2021.

[10] Rong An and Jian Su. Optimal error estimates of semi-implicit Galerkin method for time-dependent nematic liquid crystal flows. *J. Sci. Comput.*, 74:979–1008, 2018.

[11] Rong An and Weiwei Sun. Analysis of backward Euler projection FEM for the Landau–Lifshitz equation. *IMA Journal of Numerical Analysis*, 2021.

[12] Sören Bartels. Numerical analysis of a finite element scheme for the approximation of harmonic maps into surfaces. *Mathematics of computation*, 79(271):1263–1301, 2010.

[13] Sören Bartels, Christian Lubich, and Andreas Prohl. Convergent discretization of heat and wave map flows to spheres using approximate discrete Lagrange multipliers. *Math. Comp.*, 78:1269–1292, 2009.

[14] Sören Bartels and Andreas Prohl. Constraint preserving implicit finite element discretization of harmonic map flow into spheres. *Math. Comp.*, 76:1847–1859, 2007.

[15] L'ubomír Baňas, Andreas Prohl, and Reiner Schätzle. Finite element approximations of harmonic map heat flows and wave maps into spheres of nonconstant radii. *Numer. Math.*, 115:395–432, 2010.

[16] Roland Becker, Xiaobing Feng, and Andreas Prohl. Finite element approximations of the Ericksen–Leslie model for nematic liquid crystal flow. *SIAM J. Numer. Anal.*, 46:1704–1731, 2008.

[17] Susanne C. Brenner and L. Ridgway Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[18] Yongyong Cai, Jingrun Chen, Cheng Wang, and Changjian Xie. A second-order numerical method for Landau-Lifshitz-Gilbert equation with large damping parameters. *J. Comput. Phys.*, 451:Paper No. 110831, 12, 2022.

[19] Jingrun Chen, Cheng Wang, and Changjian Xie. Convergence analysis of a second-order semi-implicit projection method for Landau-Lifshitz equation. *Appl. Numer. Math.*, 168:55–74, 2021.

[20] Yunmei Chen and Fang Hua Lin. Evolution equations with a free boundary condition. *J. Geom. Anal.*, 8(2):179–197, 1998.

[21] M. J. S. Chin-Joe-Kong, W. A. Mulder, and M. Van Veldhuizen. Higher-order triangular and tetrahedral finite elements with mass lumping for solving the wave equation. *J. Engrg. Math.*, 35(4):405–426, 1999.

[22] Philippe G. Ciarlet and P. A. Raviart. Interpolation theory over curved elements, with applications to finite element methods. *Comput. Methods Appl. Mech. Eng.*, 1(2):217–249, 1972.

[23] Ivan Cimrák. Error estimates for a semi-implicit numerical scheme solving the Landau–Lifshitz equation with an exchange field. *IMA. J. Numer. Anal.*, 25:611–634, 2005.

[24] Gary Cohen, Patrick Joly, Jean Elizabeth Roberts, and Nathalie Tordjman. Higher order triangular finite elements with mass lumping for the wave equation. *SIAM J. Numer. Anal.*, 38(6):2047–2078, 2001.

[25] Alan Demlow. Higher-order finite element methods and pointwise error estimates for elliptic problems on surfaces. *SIAM J. Numer. Anal.*, 47(2):805–827, 2009.

[26] James Eells, Jr. and Joseph H. Sampson. Harmonic mappings of Riemannian manifolds. *Amer. J. Math.*, 86:109–160, 1964.

[27] Lawrence Craig Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[28] Michael Feischl and Thanh Tran. The eddy current–LLG equations: FEM-BEM coupling and a priori error estimates. *SIAM J. Numer. Anal.*, 55:1786–1819, 2017.

[29] Huadong Gao. Optimal error estimates of a linearized backward Euler FEM for the Landau–Lifshitz equation. *SIAM J. Numer. Anal.*, 52:2574–2593, 2014.

[30] Vivette Girault and Francisco Guillen-Gonzalez. Mixed formulation, approximation and decoupling algorithm for a penalized nematic liquid crystals model. *Math. Comp.*, 80:781–819, 2011.

[31] Philipp Grohs. Quasi-interpolation in riemannian manifolds. *IMA Journal of Numerical Analysis*, 33(3):849–874, 2013.

[32] Philipp Grohs, Hanne Hardering, and Oliver Sander. Optimal a priori discretization error bounds for geodesic finite elements. *Foundations of Computational Mathematics*, 15(6):1357–1411, 2015.

[33] Philipp Grohs, Hanne Hardering, Oliver Sander, and Markus Sprecher. Projection-based finite elements for nonlinear function spaces. *SIAM Journal on Numerical Analysis*, 57(1):404–428, 2019.

[34] Philipp Grohs and Markus Sprecher. Projection-based quasiinterpolation in manifolds. *SAM Report*, 23, 2013.

[35] Xinping Gui, Buyang Li, and Jilu Wang. Convergence of renormalized finite element methods for heat flow of harmonic maps. *SIAM J. Numer. Anal.*, 60(1):312–338, 2022.

[36] Juan Vicente Gutiérrez-Santacreu and Marco Restelli. Inf-sup stable finite element methods for the Landau–Lifshitz–Gilbert and harmonic map heat flow equations. *SIAM J. Numer. Anal.*, 55:2565–2591, 2017.

[37] E. Hairer and G. Wanner. *Solving ordinary differential equations. II*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2010. Stiff and differential-algebraic problems, Second revised edition, paperback.

[38] Hanne Hardering. L^2-discretization error bounds for maps into Riemannian manifolds. *Numerische Mathematik*, 139:381–410, 2018.

[39] Balázs Kovács, Buyang Li, and Christian Lubich. A convergent evolving finite element algorithm for mean curvature flow of closed surfaces. *Numer. Math.*, 143:797–853, 2019.

[40] Balázs Kovács, Buyang Li, and Christian Lubich. A convergent algorithm for forced mean curvature flow driven by diffusion on the surface. *Interfaces Free Bound.*, 22(4):443–464, 2020.

[41] Urs Lang. *Lecture notes in Riemannian and metric geometry*. ETH Zürich, 2022.

[42] Buyang Li, Jiang Yang, and Zhi Zhou. Arbitrarily high-order exponential cut-off methods for preserving maximum principle of parabolic equations. *SIAM J. Sci. Comput.*, 42(6):A3957–A3978, 2020.

[43] Fang-Hua Lin. Nonlinear theory of defects in nematic liquid crystals; phase transition and flow phenomena. *Comm. Pure Appl. Math.*, 42(6):789–814, 1989.
[44] Christian Lubich, Dhia Mansour, and Chandrasekhar Venkataraman. Backward difference time discretization of parabolic differential equations on evolving surfaces. *IMA J. Numer. Anal.*, 33:1365–1385, 2013.
[45] Olavi Nevanlinna and Farouk M. Odeh. Multiplier techniques for linear multistep methods. *Numer. Funct. Anal. Optim.*, 3(4):377–423, 1981.
[46] Andreas Prohl. *Computational micromagnetism*. Advances in Numerical Mathematics. B. G. Teubner, Stuttgart, 2001.
[47] Florian Rathgeber, David A. Ham, Lawrence Mitchell, Michael Lange, Fabio Luporini, Andrew T. T. McRae, Gheorghe-Teodor Bercea, Graham R. Markall, and Paul H. J. Kelly. Firedrake: automating the finite element method by composing abstractions. *ACM Trans. Math. Software*, 43(3):Art. 24, 27, 2017.
[48] Tristan Riviere and Michael Struwe. Partial regularity for harmonic maps and related problems. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 61(4):451–463, 2008.
[49] Jalal M Ihsan Shatah and Michael Struwe. *Geometric wave equations*, volume 2. American Mathematical Soc., 2000.
[50] Michael Struwe. *Variational Methods: Applications to Nonlinear Partial Differential Equations and Hamiltonian Systems*, volume 34. Springer Science & Business Media, 2008.
[51] Bei Tang, Guillermo Sapiro, and Vicent Caselles. Color image enhancement via chromaticity diffusion. *IEEE Trans. Image Process.*, 10(5):701–707, 2001.
[52] Karen Keskulla Uhlenbeck. Removable singularities in Yang-Mills fields. *Comm. Math. Phys.*, 83(1):11–29, 1982.
[53] Luminita A. Vese and Stanley J. Osher. Numerical methods for p-harmonic flows and applications to image processing. *SIAM J. Numer. Anal.*, 40(6):2085–2104 (2003), 2002.

Genming Bai: Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong. *E-mail address*: genming.bai@connect.polyu.hk

Xinping Gui: Beijing Computational Science Research Center, Beijing 100193, China. *E-mail address*: gui@csrc.ac.cn

Buyang Li: Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong. *E-mail address*: buyang.li@polyu.edu.hk