ON STOCHASTIC OPTIMAL CONTROL IN FERROMAGNETISM

THOMAS DUNST, ANANTA K. MAJEE, ANDREAS PROHL, AND GUY VALLET

Abstract. A model is proposed to *e.g.* control the domain wall motion in ferromagnets in the presence of thermal fluctuations, and the existence of an optimal stochastic control process is proved. The convergence of a finite element approximation of the problem is shown in space-dimension one, which then allows to apply Pontryagin's maximum principle for this finite dimensional setting. The resulting coupled system of forward-backward stochastic differential equations is numerically solved by means of the stochastic gradient method to enable practical simulations.

1. Introduction

The ability to manipulate magnetic nanostructures is relevant to optimize data storage devices; typical examples are walls in a ferromagnetic nanowire ($d = 1$) separating domains of almost uniform magnetization m, whose control over their position, structure, and dynamic behavior is crucial to ensure a reliable transport of data which are represented by those magnetic structures. A central problem in this context is to design (controlling) field pulses u which enable prescribed precessional switching, in particular in the presence of thermal fluctuations. At zero temperature, the physical literature mainly discusses simplified settings, or 'trial-and-error approaches' to motivate certain field pulses which approximately serve this goal; see [5, p. 144]. The approaches in [1, 2] discuss controllability of finite spin ensembles through applied fields u, while a more practical approach which is based on the minimization of a functional $J(m, u)$ subject to solving the Landau-Lifshitz-Gilbert equation is studied in [13].

Let $D \subset \mathbb{R}^d$ $(1 \leq d \leq 3)$ be a bounded Lipschitz domain, and $T > 0$. The magnetization $m : D_T \times \Omega \to \mathbb{R}^3$ at elevated temperature $\mathtt{T} > 0$ is governed by the stochastic Landau-Lifshitz-Gilbert equation (SLLG)

$$\begin{aligned} dm &= \Big(m \times H_{\text{eff}} - \alpha\, m \times (m \times H_{\text{eff}})\Big)\, dt + \iota\, m \times \circ dW \quad && \text{in } D_T : (0,T) \times D\,, && (1.1)\\ \frac{\partial m}{\partial \nu} &= 0 && \text{on } \partial D_T := (0,T) \times \partial D\,, \\ m(0,\cdot) &= m_0 && \text{on } D, \end{aligned}$$

where $\iota = \iota(\mathtt{T})$ denotes the noise intensity. Moreover, $H_{\text{eff}} \equiv H_{\text{eff}}(m) = -D\mathcal{E}(m)$ denotes the effective field in this model of ferromagnetism, which is deduced from the Landau-Lifshitz energy $\mathcal{E}(m) \equiv \mathcal{E}_u(m)$ for some given external field $u : D_T \times \Omega \to \mathbb{R}^3$, and for simplicity in this work ensembles the exchange and external field energies,

$$\mathcal{E}(m) = \int_D \frac{A}{2} |\nabla m|^2 - \langle m, u \rangle \, dx\,, \tag{1.2}$$

where $A, \alpha > 0$; see *e.g.* [4, 27] for further details on this model. The system in (1.1) is driven by a Hilbert space-valued $\mathbf{Q}$-Wiener process W, where $\mathbf{Q} : \mathbb{K} \to \mathbb{K}$ is a symmetric, non-negative operator acting in a Hilbert space $\mathbb{K} \subset W^{1,\infty}(D;\mathbb{R}^3) \cap H^2(D;\mathbb{R}^3)$, and $\circ dW(t)$ in (1.1) denotes the Stratonovich differential. By [8], there exists a weak martingale solution of (1.1) with $u = 0$ for $1 \leq d \leq 3$, which is a 6-tuple $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P, W, m\big)$ such that (1.1) holds P-a.s. in analytically weak form; it may even be obtained as proper limit of iterates of an implementable numerical scheme as shown in [4]. These results may be sharpened to the existence of a unique strong solution for

2000 *Mathematics Subject Classification.* 45K05, 46S50, 49L20, 49L25, 91A23, 93E20.

Key words and phrases. Stochastic Landau-Lifshitz-Gilbert equation, Stratonovich noise, Young measures, Optimal control, Faedo-Galerkin approximation.

The fourth author would like to acknowledge the support of Institut Carnot ISIFoR .

$d=1$, where an $\mathcal{F}_t$-adapted process $m: D_T\times\Omega\mapsto\mathbb{S}^2$ satisfies (1.1) in analytically weak form for a given stochastic basis $\big(\Omega,\mathcal{F},\{\mathcal{F}_t\},P,W\big)$; cf. [9, 12].

Our goal is the stochastic optimal control in ferromagnetism: Let $\bar{m}\in H^1(D_T;\mathbb{S}^2)$[a] be given, $K>0$ and $q\geq 2$. Find a weak admissible solution $\pi^*:=\big(\Omega^*,\mathcal{F}^*,\{\mathcal{F}_t^*\},P^*,W^*,m^*,u^*\big)$ which minimizes

$$
\begin{aligned}
J(\pi)=E\Big[\int_0^T\Big(\|m-\bar{m}\|_{\mathbb{L}^2}^2+\|u\|_{\mathbb{H}^1}^{2q}\Big)\,dt+\psi(m(T))\Big]\ \text{ with }\ \pi=\big(\Omega,\mathcal{F},\{\mathcal{F}_t\},P,W,m,u\big)\\
\text{subject to (1.1) and } \|u(t)\|_{\mathbb{L}^2}\leq K \text{ for a.e. } t\in[0,T],\ P\text{-a.s.}
\end{aligned}
\tag{1.3}
$$

The existing literature (see *e.g.* [28]) on stochastic optimal control with SPDEs mainly considers those which have a mild solutions, which is not available for problem (1.1). For this reason, a minimizer π^* of (1.3) may be constructed by variational methods. For a minimizing sequence of weak admissible controls $\pi_n:=\big(\Omega_n,\mathcal{F}_n,\{\mathcal{F}_t^n\},P_n,W_n,m_n,u_n\big)$, we are looking for tightness of the laws of the process $\{m_n\}_{n\in\mathbb{N}}$ in the path space $C([0,T];\mathbb{L}^2)$ and therefore need to obtain uniform bounds in $W^{\gamma,p}(0,T;\mathbb{L}^2)$[b] for $\gamma\in(0,\frac{1}{2})$, $p\in[2,\infty)$ such that $p\gamma>1$. This constraint is in particular fulfilled for $p=4$, which is why we focus on $q=2$ in (1.3) in some parts below.

Once a minimizer π^* of problem (1.3) has been found we ask for its computation; in the deterministic setting (see e.g. [13]), a common numerical strategy to accomplish this goal may be based on Pontryagin's maximum principle. While the maximum principle has been obtained for several optimal control problems with prototypic SPDE constraints (see *e.g.* [19, 20]), a corresponding argumentation is not immediate in (1.3) where the nonlinear drift in (1.1) is not Lipschitz. To overcome this problem, we first show the convergence for a structure preserving finite element discretization (2.5) of (1.3) for $d=1$, where the corresponding drift is then Lipschitz because the (approximate) solutions of (1.1) are of unit length at the nodal points of the underlying mesh $\mathcal{T}_h$ covering D. To achieve this goal, we need a probabilistically strong solution of SLLG (1.1) with improved regularity properties to prove rates of local strong convergence for the approximate SPDE in (2.5) (see also (7.21)) towards (1.1), which is the key step to show the convergence of (2.5) to (1.3). We remark that our approach would not work for a general Galerkin discretization of (1.1) such as the one used in Subsection 7.1 where the preservation of the sphere property is not clear.

For the discretization (2.5) of (1.3), we then obtain Pontryagin's maximum principle resulting in a coupled forward-backward SDE system; cf. (2.7)-(2.8). For its solvability, we employ the control constraint in (1.3). Then, the coupled forward-backward SDE system is numerically solved by the

(i) least squares Monte-Carlo method to approximate conditional expectations which need to be computed at every time step to obtain approximate solutions of the adjoint equation,
(ii) and the stochastic gradient method from [14] to obtain updates of the feedback control u_h^* whose functional values monotonically decrease.

The rest of the paper is organized as follows. In Section 2, we give a proper definition of solvability for (the relaxed version of) problem (1.3) and state the main results. An optimal relaxed control for the relaxed version (3.3) of the problem (1.3) is constructed in Section 3 by using compactness properties of random Young measures on a suitable Polish space. A weak solution of problem (1.3) is then obtained in Section 4, which settles Theorem 2.3. Convergence of the value function and (suitable) minimizers of the finite element approximation (2.5) of (1.3) (for $d=1$) is shown in Section 5, and the corresponding maximum principle is obtained. In Section 6, the stochastic gradient method is detailed for discretization of (1.1), and simulations which approximately solve (1.3) are reported.

2. Reformulation of (1.3), Pontryagin's maximum principle, and main results

Throughout this paper, we use the letter $C>0$ to denote various generic constants. In the sequel, we denote by $\mathbb{L}^p$ the space $L^p(D;\mathbb{R}^3)$, and by $\mathbb{W}^{l,p}$ the space $W^{l,p}(D;\mathbb{R}^3)$ for any $l\geq 0$ and $p\geq 1$. We use $\langle,\rangle$ for an inner product in $\mathbb{R}^3$, and the euclidean norm is denoted by $|\cdot|$.

[a] $H^1(D_T;\mathbb{S}^2):=\big\{v(t)\in\mathbb{H}^1: v(t,x)\in\mathbb{S}^2 \text{ for almost every } x\in D \text{ and for all } t\in[0,T]\ \big\}$.

[b] $W^{\gamma,p}(0,T;\mathbb{X}):=\Big\{u\in L^p(0,T;\mathbb{X}): \displaystyle\int_0^T\int_0^T\frac{\|u(t)-u(s)\|_{\mathbb{X}}^p}{|t-s|^{1+\gamma p}}\,ds\,dt<+\infty\Big\}$.

2.1. **Stochastic optimal control problem (1.3): reformulation and main results.** Without loss of generality, we may assume in the following that the Wiener process W is of the form $a\beta(t)$, where $a \in \mathbb{W}^{1,\infty}$ and $\{\beta(t) : t \geq 0\}$ is a real-valued Brownian motion. In view of the effective field H_{eff} in (1.2) with $A = 1$, and avoiding the exchange energy in the phenomenological damping term of SLLG, equation (1.1) has the form

$$\begin{aligned} dm(t) &= \Big[m(t) \times \Delta m(t) + m(t) \times u(t) - \alpha\, m(t) \times (m(t) \times \Delta m(t))\Big]\, dt + \iota\, m(t) \times a \circ d\beta(t) \\ \frac{\partial m}{\partial \nu} &= 0 \qquad \text{on}\, \partial D_T \\ m(0,\cdot) &= m_0(\cdot) \qquad \text{on } D. \end{aligned} \tag{2.1}$$

Note that SLLG is a stochastic PDE with non-Lipschitz drift function and in general lacks a stochastically strong solution for $d \geq 2$. In [8], an (analytically) weak martingale solution is constructed for (2.1) with $u = 0$, which may easily be generalized to deterministic $u \in L^4(0,T;\mathbb{L}^2)$: there exists a 6-tuple $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P, \beta, m\big)$ such that

- $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ is a filtered probability space satisfying the usual hypotheses.
- β is an $\mathbb{R}$-valued $\mathcal{F}_t$-adapted Wiener process.
- m is an $\mathbb{H}^1$-valued $\mathcal{F}_t$-adapted stochastic process such that for P-a.e. $\omega \in \Omega$,
  a). $m(\omega,\cdot) \in C([0,T];\mathbb{L}^2)$ and $|m(\omega,t,x)| = 1$ for a.e. $x \in D$ and all $t \in [0,T]$.
  b). For all $t \in [0,T]$, and $\phi \in C^\infty(\overline{D};\mathbb{R}^3)$, the following equality holds: P-a.s.,

$$\begin{aligned} &\big(m(t), \phi\big)_{\mathbb{L}^2} - \big(m_0, \phi\big)_{\mathbb{L}^2} \\ &= \int_0^t \Big(\nabla m(s), m(s) \times \nabla\phi\Big)_{\mathbb{L}^2} ds + \alpha \int_0^t \Big(\nabla m(s), \nabla(\phi \times m(s)) \times m(s)\Big)_{\mathbb{L}^2} ds \\ &\qquad + \int_0^t \Big(m(s) \times u(s), \phi\Big)_{\mathbb{L}^2} ds + \iota \Big(\int_0^t m(s) \times a \circ d\beta(s), \phi\Big)_{\mathbb{L}^2}. \end{aligned} \tag{2.2}$$

Here, the first two terms of the right hand side of (2.2) are understood as

$$\sum_i \int_0^t \Big(\frac{\partial}{\partial x_i} m(s), m(s) \times \frac{\partial}{\partial x_i}\phi\Big)_{\mathbb{L}^2} ds \quad \text{and} \quad \alpha \sum_i \int_0^t \Big(\frac{\partial}{\partial x_i} m(s), \frac{\partial}{\partial x_i}(\phi \times m(s)) \times m(s)\Big)_{\mathbb{L}^2} ds$$

respectively. Next to the martingale representation theorem, its construction in particular uses the Skorokhod lemma which changes the underlying probability space to study sequences of approximate solutions with improved convergence properties. To solve (2.1) for an applied stochastic process $u \neq 0$ requires an extended argument in case only its law $\mu = \mathcal{L}(u)$ is prescribed. The following result is a generalization of [8, Theorem 2.11] whose proof is postponed to Subsection 7.1.

Theorem 2.1. *Let $D \subset \mathbb{R}^d$ $(1 \leq d \leq 3)$ be a bounded Lipschitz domain, $q \geq 2$, $T > 0$, $l \in \{0,1\}$, and $m_0 \in \mathbb{W}^{1,2}(D, \mathbb{S}^2)$*[c]*. Let $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P\big)$ be a given filtered probability space satisfying the usual hypotheses and β is a $\mathcal{F}_t$-adapted real-valued Wiener process on it. Let μ be a probability measure on $L^2(0,T;\mathbb{L}^2)$ such that $\int_{L^2(0,T;\mathbb{L}^2)} \Phi(v)\mu(dv) < +\infty$ where $\Phi(v) = \|v\|_{L^{2q}(0,T;\mathbb{W}^{l,2})}$ and $\mu\big\{v : \|v\|_{L^\infty(0,T;\mathbb{L}^2)} < K\big\} = 1$ for some given constant $K > 0$. Then for problem (2.1), there exist a weak martingale solution $\tilde{\pi} = \big(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}, \tilde{\beta}, \tilde{m}\big)$*[d] *and a $\tilde{\mathcal{F}}_t$-predictable stochastic process $\tilde{u} : \tilde{\Omega} \times D_T \mapsto \mathbb{R}^3$ in the sense given above such that (2.2) is valid, and* i) *$\mu = \mathcal{L}(\tilde{u})$ on $L^2(0,T;\mathbb{L}^2)$ and*

$$\tilde{P}\text{-a.s.}, \quad \|\tilde{u}(t)\|_{\mathbb{L}^2} \leq K \quad \text{for a.e. } t \in [0,T]; \qquad \tilde{E}\Big[\int_0^T \|\tilde{u}(t)\|_{\mathbb{W}^{l,2}}^{2q}\, dt\Big] < +\infty. \tag{2.3}$$

ii) *There exist positive constants C_1, C_2 such that*

$$\tilde{E}\Big[\sup_{0\leq t\leq T} \|\nabla\tilde{m}(t)\|_{\mathbb{L}^2}^{2q} + \Big(\int_0^T \|\tilde{m}(s) \times \Delta\tilde{m}(s)\|_{\mathbb{L}^2}^2\, ds\Big)^q\Big] \leq C_1 + C_2\tilde{E}\Big[\int_0^T \|\tilde{u}(t)\|_{\mathbb{L}^2}^{2q}\, dt\Big].$$

As a by-product of the above theorem, we have the following corollary.

[c] $\mathbb{W}^{1,2}(D, \mathbb{S}^2) = \big\{\varphi \in W^{1,2}(D;\mathbb{R}^3) : \ \varphi(x) \in \mathbb{S}^2 \text{ for almost every } x \in D\big\}$

[d] The associated expectation with P (resp. $\tilde{P}$) is denoted by E (resp. $\tilde{E}$).

Corollary 2.2. *Let $D \subset \mathbb{R}$ be bounded, $m_0 \in \mathbb{W}^{1,2}(D, \mathbb{S}^2)$, and $l \in \{0, 1\}$. Let $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P\big)$ be a given filtered probability space and u an $\mathbb{W}^{l,2}$-valued $\{\mathcal{F}_t\}$-predictable stochastic process on it whose law $\mathcal{L}(u)$ satisfies the properties given in Theorem 2.1 for $q \geq 4$. Then there exists a unique strong solution $\pi = \big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P, \beta, m, u\big)$ for the problem (2.1) satisfying the bounds as in Theorem 2.1. Moreover*

$$E\Big[\int_0^T \|\Delta m(t)\|^2_{\mathbb{L}^2}\, dt\Big] < +\infty.$$

Note that a weak martingale solution to the problem (2.1) is not unique for a given control process u satisfying (2.3) for $d = 2, 3$. We denote by $\mathcal{U}^w_{\rm ad}(m_0, T)$ the set of weak admissible solutions to the problem (2.1) in the sense of Theorem 2.1. From now onwards, we consider $l = 1$ in Theorem 2.1. The stochastic optimal control problem (1.3) may be rewritten as follows.

Definition 2.1. Let $0 < T < \infty$. Assuming the setup of data from Theorem 2.1, let ψ be a given Lipschitz continuous function on $\mathbb{L}^2$. A weak optimal solution for the control problem (1.3) is a 7-tuple $\pi^* = \big(\Omega^*, \mathcal{F}^*, \{\mathcal{F}^*_t\}, P^*, \beta^*, m^*, u^*\big) \in \mathcal{U}^w_{\rm ad}(m_0, T)$ such that

$$J(\pi^*) = \inf_{\pi \in \mathcal{U}^w_{\rm ad}(m_0,T)} J(\pi). \tag{2.4}$$

The associated control u^* in π^* is called weak optimal control of the underlying control problem. Our first main result is the following.

Theorem 2.3. *There exists a weak optimal solution π^* of* (1.3) *in the sense of Definition* 2.1.

The proof is detailed in Sections 3 and 4, and uses the variational solution concept for the SPDE (2.1). Note that there is no convenient compactness structure for the control sets, which is why we follow the strategy introduced by Fleming [17] to embed admissible controls into a larger space with proper compactness properties. We therefore consider a relaxed form of (1.3) first by considering the given control as a Young measure-valued control process taking values in the Polish space $\mathbb{L}^2$. Using compactness properties of Young measure-valued controls and Skorokhod's theorem, we establish the existence of a weak relaxed optimal control (cf. Theorem 3.1). Then, we exploit the convexity property of J with respect to the control variable, and linearity in (2.1) to settle Theorem 2.3. An alternative construction avoiding Young measures, which exploits Jakubowski's extension of Skorokhod's lemma to certain non-Polish spaces to verify Theorem 2.3 is discussed in Remark 4.1.

2.2. **Pontryagin's maximum principle for a finite element approximation of problem** (1.3) **for** $q \geq 4$ **and d=1.** Theorem 2.3 shows the existence of an optimal control in problem (2.4). In general cases, the optimal control may not be explicitly calculated and hence a numerical gradient descent method which is based on Pontryagin's maximum principle may be applied to approximately solve (1.3). Unfortunately, its validation is non-trivial due to the presence of the non-Lipschitz nonlinearities in (1.3), excluding its standard derivation *via* spike variations. For this reason, we proceed differently and discretize problem (1.3) first, to then validate a discrete maximum principle (and not *vice versa*). It turns out that convergence properties of a Galerkin method for (1.3) crucially depend on its respectation of the sphere property of states in (1.1). As a consequence, a *particular* Galerkin method is needed which preserves this property accordingly. For this purpose, we employ numerical concepts from [4, Chapter 2] to obtain the structure preserving finite element approximation (2.5) of the problem (1.3), and show its convergence when $h \to 0$, for h being the mesh size of the quasi-uniform triangulation $\mathcal{T}_h$ of $\overline{D}$; cf. [7]. Let $\mathbb{V}_h$ be the lowest order $\mathbb{H}^1$-conforming finite element space, $\mathcal{I}_h$ its (affine) nodal interpolation operator, and $\Delta_h : \mathbb{V}_h \to \mathbb{V}_h$ the discrete Laplacian; see Subsection 7.4 for further details. For every $h > 0$, the finite dimensional version of (1.3) then reads as follows: minimize

$$
\begin{aligned}
J_h(\pi_h) = E\Big[\int_0^T \Big(\big\|m(t) - \mathcal{I}_h[\bar{m}(t)]\big\|_h^2 + \|u(t)\|_{\mathbb{H}^1}^{2q}\Big)\, dt + \psi(m(T))\Big] \\
\text{with } \pi_h = \big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P, \beta, m, u\big), \text{ subject to the SDE} \\
dm(t) = \Big[\mathcal{I}_h\big[m(t)\times \Delta_h m(t)\big] + \mathcal{I}_h\big[m(t)\times u(t)\big] - \alpha\, \mathcal{I}_h\big[m(t)\times (m(t)\times \Delta_h m(t))\big]\Big] dt \\
+ \iota\, \mathcal{I}_h\big[m(t)\times a\big] \circ d\beta(t),\ t>0, \\
m(0) = \mathcal{I}_h[m_0], \quad \text{and} \quad u(t,\cdot)\in \mathcal{U}_h := \big\{v\in \mathbb{V}_h : \|v\|_{\mathbb{L}^2}\le K\big\} \text{ for a.e. } t\in[0,T],\ P\text{-a.s.}
\end{aligned}
\tag{2.5}
$$

Here, m is a $\mathbb{V}_h$-valued process on a given filtered probability space which takes values in $\mathbb{S}^2$ at nodal points $\{x_l\}_{l=1}^L$ of the mesh $\mathcal{T}_h$, and β is a given $\mathbb{R}$-valued Wiener process on it. There are several issues which need to be dealt with to validate the convergence of the value functions $J_h(\pi_h^*)$ to $J(\pi^*)$: our argumentation requires probabilistically strong solutions of SLLG, and higher spatial regularity properties of solutions of (2.1) for it to be valid. According to Corollary 2.2, a unique probabilistically strong solution of (2.1) exists for $d=1$, $q\ge 4$, and it exhibits improved spatial regularity properties. We denote the set of $\mathcal{U}_h$-valued controls to the problem (2.5) by $\mathcal{U}_{\mathrm{ad},h}^w(m_0,T)$. By following the proof of Theorem 2.3, one can show that there exists a 7-tuple $\pi_h^* = \big(\Omega_h, \mathcal{F}_h, \{\mathcal{F}_t^h\}, P^h, \beta_h^*, m_h^*, u_h^*\big)$ such that

$$
J_h(\pi_h^*) = \inf_{\pi\in\mathcal{U}_{\mathrm{ad},h}^w(m_0,T)} J_h(\pi), \qquad \text{while already} \qquad J(\pi^*) = \inf_{\pi\in\mathcal{U}_{\mathrm{ad}}^w(m_0,T)} J(\pi). \tag{2.6}
$$

Theorem 2.4. *Let $d=1$, $q\ge 4$, and $m_0\in \mathbb{W}^{1,2}(D,\mathbb{S}^2)$. Let π^* and π_h^* be 7-tuples as described above. Assume that ψ is Lipschitz continuous on $\mathbb{L}^2$. Then*

$$
J_h(\pi_h^*) \to J(\pi^*) \qquad (h\to 0).
$$

The proof uses the strong convergence of some related control processes in $\mathbb{H}^1$: we first consider the strong solution m_h of the finite dimensional equation in (2.5) for $u=\mathcal{R}_h u^*$ on the filtered probability space $\big(\Omega^*, \mathcal{F}^*, \{\mathcal{F}_t^*\}, P^*\big)$, where $\mathcal{R}_h$ is the Ritz projection defined in (7.11), and estimate the expected value of the $\mathbb{L}^2$-difference of the solutions m_h and m^* on a large subset of Ω^* (cf. Lemma 5.1). For this purpose, we use

- the reformulation of the PDE (2.1) as a semi-linear PDE by the identity $m^*\times(m^*\times\Delta m^*) = -\Delta m^* - |\nabla m^*|^2 m^*$ which exploits the unit-length property of its solution,
- a corresponding reformulation of the equation in (2.5) where we benefit from the numerical quadrature and the discrete sphere condition as relevant features of the equation in (2.5), and the (uniform) stability for the solution of it (resp. (7.21)) in Lemma 7.3 which is based on the discrete Gagliardo-Nirenberg inequality in Lemma 7.2.

As a result, we obtain the convergence of $J_h(\bar\pi_h)$ to $J(\pi^*)$ for $h\to 0$ where $\bar\pi_h = \big(\Omega^*, \mathcal{F}^*, \{\mathcal{F}_t^*\}, P^*, \beta^*, m_h, \mathcal{R}_h u^*\big)$; in a next step, we employ a standard variational argument to show $J(\pi^*) \le J_h(\pi_h^*) \le J_h(\bar\pi_h)$ which then settles Theorem 2.4.

A consequence of Theorem 2.4 is the strong convergence of related optimal controls on some probability space.

Corollary 2.5. *Let $d=1$, $q\ge 4$, and $m_0\in\mathbb{W}^{1,2}(D,\mathbb{S}^2)$. For every $h>0$, there exists a weak optimal solution $\tilde\pi_h^* = \big(\tilde\Omega, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t^h\}, \tilde P, \tilde\beta_h^*, \tilde m_h^*, \tilde u_h^*\big) \in \mathcal{U}_{\mathrm{ad,h}}^w(m_0,T)$ of (2.5). Moreover, there exists a weak optimal solution $\tilde\pi^* = \big(\tilde\Omega, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde P, \tilde\beta^*, \tilde m^*, \tilde u^*\big) \in \mathcal{U}_{\mathrm{ad}}^w(m_0,T)$ of (1.3) such that for $h\to 0$*

$$
\begin{aligned}
\tilde P\text{-a.s.,}\quad \tilde m_h^* \to \tilde m^* \text{ in } C([0,T];\mathbb{L}^2); \quad \tilde\beta_h^* \to \tilde\beta^* \text{ in } C([0,T];\mathbb{R}), \\
\text{and}\quad J_h(\tilde\pi_h^*) \to J(\tilde\pi^*); \quad \tilde u_h^* \to \tilde u^* \text{ in } L^{2q}(\tilde\Omega\times(0,T);\mathbb{H}^1).
\end{aligned}
$$

For every $h > 0$, the necessary first order optimality conditions for the 7-tuple π_h^* satisfying (2.6) involve a backward stochastic differential equation (BSDE in short) as adjoint equation:

$$\begin{aligned} dP_h(t) = \Big\{ & 2\big(m_h^*(t) - \mathcal{I}_h[\bar{m}(t)]\big) + \mathcal{I}_h[P_h(t) \times \Delta_h m_h^*(t)] - \Delta_h \mathcal{I}_h[P_h(t) \times m_h^*(t)] \\ & + \alpha \Delta_h \mathcal{I}_h[m_h^*(t)\langle m_h^*(t), P_h(t)\rangle - |m_h^*(t)|^2 P_h(t)] + \mathcal{I}_h[P_h(t) \times u_h^*(t)] \\ & - \alpha \mathcal{I}_h[\Delta_h m_h^*(t) \times (m_h^*(t) \times P_h(t))] - \alpha \mathcal{I}_h[P_h(t) \times (m_h^*(t) \times \Delta_h m_h^*(t))] \\ & + \iota \mathcal{I}_h[Q_h(t) \times a] - \frac{\iota^2}{2}\mathcal{I}_h[(P_h(t) \times a) \times a]\Big\} dt + Q_h(t) d\beta_h^*(t) + d\mathcal{N}_h(t), \; 0 \le t < T, \\ P_h(T) = & -\mathcal{I}_h[D\psi(m_h^*(T))]. \end{aligned} \tag{2.7}$$

We refer again to Subsection 7.4 for the used notation. In our case, the filtration $\{\mathcal{F}_t^h\}_{t\ge 0}$ is in general larger than the natural filtration generated by the Wiener process β_h^*, augmented by all the P^h- null sets in $\mathcal{F}_h$, leading to the additional $\mathbb{V}_h$-valued $\mathcal{F}_t^h$-martingale $\mathcal{N}_h$ in (2.7) (see *e.g.* [26] and the references therein). We may use the boundedness of m_h^* and of the control u_h^* to verify that the drift function in (2.7) is Lipschitz continuous from $\mathbb{V}_h$ onto $\mathbb{V}_h$ for every fixed $h > 0$. Following [26, 31], equation (2.7) then admits a unique predictable solution $(P_h, Q_h, \mathcal{N}_h) \in L^2_{\mathcal{F}_h}(\Omega_h; C([0,T]; \mathbb{V}_h)) \times L^2_{\mathcal{F}_h}(\Omega_h; L^2(0,T; \mathbb{V}_h)) \times L^2_{\mathcal{F}_h}(\Omega_h; L^2(0,T; \mathbb{V}_h))$ such that

$$E_h\Big[\Big(\mathcal{N}_h, \int_0^T Q_h(t) d\beta_h^*(t)\Big)_{\mathbb{L}^2}\Big] = 0.$$

Moreover, the following maximum principle holds: P^h-a.s. and for a.e. $t \in [0,T]$,

$$\big(P_h^*(t) \times m_h^*(t), \varphi - u_h^*(t)\big)_h - 4q\|u_h^*(t)\|_{\mathbb{H}^1}^{2q-2}\big(u_h^*(t), \varphi - u_h^*(t)\big)_{\mathbb{H}^1} \le 0 \tag{2.8}$$

for all $\varphi \in \mathcal{U}_h$, where the Hamiltonian $\mathcal{H}_h : \mathbb{V}_h \times \mathcal{U}_h \times \mathbb{V}_h \times \mathbb{V}_h \mapsto \mathbb{R}$ is defined as

$$\begin{aligned} \mathcal{H}_h(m, u, P_h, Q_h) = & - \big\|m - \mathcal{I}_h[\bar{m}]\big\|_h^2 - \|u\|_{\mathbb{H}^1}^{2q} + \Big(\mathcal{I}_h[m \times \Delta_h m], P_h\Big)_h + \Big(\mathcal{I}_h[m \times u], P_h\Big)_h \\ & - \alpha\Big(\mathcal{I}_h[m \times (m \times \Delta_h m)], P_h\Big)_h + \iota\Big(\mathcal{I}_h[m \times a], Q_h\Big)_h \\ & + \frac{\iota^2}{2}\Big(\mathcal{I}_h[(m \times a) \times a], P_h\Big)_h, \end{aligned}$$

and $(P_h, Q_h, \mathcal{N}_h)$ is the unique solution to the adjoint equation (2.7).

Note that the corresponding moment estimates for the solution of (2.7) depend on h which is essentially due to the quadratic and cubic cross-product terms in $\mathcal{H}_h$. Therefore, passing to the limit in (2.8) is not clear, but the coupled forward-backward SDE system (FBSDE for short) (2.5), (2.7)-(2.8) is now amenable to numerical techniques which are discussed next.

Remark 2.6. From a mathematical view point, the inequality constraint on the controls in Theorem 2.1 is only used to show the existence of a solution of the adjoint equation, cf. (2.7), to validate Pontryagin's maximum principle.

2.3. **The least squares Monte-Carlo method and the stochastic gradient method to approximately solve** (2.6)**.** The starting point to simulate an approximate minimizer of (1.3) is the FBSDE system (2.5), (2.7)–(2.8), but further steps are needed to obtain an implementable scheme. We start with the time discretization of the equation in (2.5), where the midpoint rule is used on a mesh $I_k := \{t_j\}_{j=0}^J \subset [0,T]$ to obtain a scheme where $\mathbb{V}_h$-valued random variables $\{M^j\}_{j=0}^J$ take again values in $\mathbb{S}^2$ at nodal points $\{x_l\}_{l=1}^L \subset \overline{D}$ of $\mathcal{T}_h$. We denote $M^{j+\frac{1}{2}} := \frac{1}{2}(M^{j+1} + M^j)$.

Algorithm 2.7. Let M^0 be a $\mathbb{V}_h$-valued random variable, with $\{M^0(x_l)\}_{l=1}^L \subset \mathbb{S}^2$. Let $\{U^j; 0 \le j \le J-1\} \subset L^2(\Omega; \mathbb{V}_h)$ be given, as well as $\Delta_j\beta := \beta(t_{j+1}) - \beta(t_j) \sim \mathcal{N}(0,k)$. For every $0 \le j \le J-1$, find the $\mathbb{V}_h$-valued random variable M^{j+1}, such that

$$\begin{aligned} \big(M^{j+1} - M^j, \phi\big)_h = & k\big(M^{j+\frac{1}{2}} \times \big[\Delta_h M^{j+1} + U^j\big], \phi\big)_h - \alpha k\big(M^{j+\frac{1}{2}} \times \big[M^{j+\frac{1}{2}} \times \Delta_h M^{j+1}\big], \phi\big)_h \\ & + \iota\big(M^{j+\frac{1}{2}} \times a, \phi\big)_h \Delta_j\beta \quad \forall \phi \in \mathbb{V}_h. \end{aligned} \tag{2.9}$$

For every j, Equation (2.9) is a coupled nonlinear system of equations which is solved by Newton's method, where the Jacobian is approximated by finite differences. Again, for computational purposes, we approximate the Hamiltonian $\mathcal{H}_h$ by

$$\widetilde{\mathcal{H}}_h(m,u,P_h,Q_h) = -\|m-\bar{m}\|^2_{\mathbb{L}^2} - \|u\|^{2q}_{\mathbb{H}^1} + \big(\nabla m, m\times P_h\big)_{\mathbb{L}^2} + \big(m\times u, P_h\big)_{\mathbb{L}^2} - \alpha\big(\nabla m, \nabla P_h\big)_{\mathbb{L}^2}$$
$$+ \alpha\big(|\nabla m|^2 m, P_h\big)_{\mathbb{L}^2} + \iota\big(m\times a, Q_h\big)_{\mathbb{L}^2} + \frac{\iota^2}{2}\big((m\times a)\times a, P_h\big)_{\mathbb{L}^2},$$

which is suggested by

- using the discrete sphere property of the computed magnetization; *e.g.*, $m\times(m\times\Delta_h m)$ in $\mathcal{H}_h$ is replaced by $-\Delta_h m - |\nabla m|^2 m$, and
- replacing $\big(\,\cdot\,\big)_h$ by the scalar product in $\mathbb{L}^2$.

The associated adjoint equation and its maximum principle for π^*_h are then given by

$$\begin{cases} dP_h(t) = -\frac{\partial}{\partial m}\widetilde{\mathcal{H}}_h(m^*_h(t),u^*_h(t),P_h(t),Q_h(t))\,dt + Q_h(t)d\beta^*_h(t) + d\mathcal{N}_h(t),\ 0\le t<T,\\ P_h(T) = -D\psi(m^*_h(T)), \end{cases} \tag{2.10}$$

P^h-a.s., and for a.e. $t\in[0,T]$,

$$\big(P^*_h(t)\times m^*_h(t), \varphi - u^*_h(t)\big)_{\mathbb{L}^2} - 4q\|u^*_h(t)\|^{2q-2}_{\mathbb{H}^1}\big(u^*_h(t),\varphi-u^*_h(t)\big)_{\mathbb{H}^1} \le 0 \quad \forall\varphi\in\mathcal{U}_h. \tag{2.11}$$

For the time discretization of the backward equation (2.10) we use a semi-implicit scheme; see e.g. [6] for the time discretization of BSDEs.

Algorithm 2.8. Let $\{M^j;\,0\le j\le J\}\subset L^2(\Omega_h;\mathbb{V}_h)$ and $\{U^j;\,0\le j\le J-1\}\subset L^2(\Omega_h;\mathbb{V}_h)$ be given. Let $\big(P^J,\phi\big)_{\mathbb{L}^2} = -\big(D\Psi(M^J),\phi\big)_{\mathbb{L}^2}$ for every $\phi\in\mathbb{V}_h$. For every $0\le j\le J-1$, find the $\mathbb{V}_h$-valued random variables (P^j,Q^j), such that

$$\big(Q^j,\phi\big)_{\mathbb{L}^2} = E_h\Big[\frac{\Delta_j\beta}{k}\big(P^{j+1},\phi\big)_{\mathbb{L}^2}\big|\mathcal{F}^h_{t_j}\Big] \quad \forall\phi\in\mathbb{V}_h, \tag{2.12}$$

and

$$\begin{aligned}
&\big(P^j,\phi\big)_{\mathbb{L}^2} + \alpha k\big(\nabla P^j,\nabla\phi\big)_{\mathbb{L}^2} - \alpha k\big(|\nabla M^j|^2 P^j,\phi\big)_{\mathbb{L}^2} - k\big(M^j\times\nabla P^j,\nabla\phi\big)_{\mathbb{L}^2}\\
&\quad + k\big(\nabla M^j\times\nabla P^j,\phi\big)_{\mathbb{L}^2} - k\big(U^j\times P^j,\phi\big)_{\mathbb{L}^2}\\
&= E_h\Big[\Big\{\big(P^{j+1},\phi\big)_{\mathbb{L}^2} - 2k\big(M^{j+1}-\bar{M}(t_{j+1}),\phi\big)_{\mathbb{L}^2} + 2\alpha k\big(\langle P^{j+1},M^{j+1}\rangle\nabla M^{j+1},\nabla\phi\big)_{\mathbb{L}^2}\\
&\quad + \frac{\iota^2}{2}k\big((P^{j+1}\times a)\times a,\phi\big)_{\mathbb{L}^2}\Big\}\Big|\mathcal{F}^h_{t_j}\Big] - \iota k\big(Q^j\times a,\phi\big)_{\mathbb{L}^2} \quad \forall\phi\in\mathbb{V}_h.
\end{aligned} \tag{2.13}$$

For every $0\le j\le J$, we may represent the solution (P^j,Q^j) of (2.12)–(2.13) by two measurable, deterministic, but unknown functions $\mathcal{P}^j:\mathbb{V}_h\to\mathbb{V}_h$ and $\mathcal{Q}^j:\mathbb{V}_h\to\mathbb{V}_h$, such that $P^j=\mathcal{P}^j(M^j)$, and $Q^j=\mathcal{Q}^j(M^j)$. By the least squares Monte-Carlo method, see e.g. [21], these functions are approximated by $(\mathcal{P}^j_R(\cdot),\mathcal{Q}^j_R(\cdot))$ using the finite dimensional space span$\{\mathbb{1}_{C^j_r}(\cdot);\,r=1,\dots,R\}$ and coefficients $\{p^j_r\}^R_{r=1}$ resp. $\{q^j_r\}^R_{r=1}$. It is due to the high dimensionality of $\mathbb{V}_h$ that an adaptive partitioning of it is needed for this purpose. This computation is the bottleneck in the simulation of the BSDE (2.12)–(2.13) as part of the overall stochastic control problem, since it poses severe demands on computational times as well as memory storage. In order to weaken these demands and thus allow for the simulation of larger spin systems, we choose an approach similar to the stratified regression algorithm in [22], which allows for an efficient parallel implementation; see Section 6 for details. Algorithm 2.8 in combination with the least squares Monte-Carlo method then approximates (2.7) as part of the stochastic control problem (2.5), (2.7)–(2.8) which is now solved by the stochastic gradient method. This iterative scheme was proposed in [14] (in a related setting) to compute deterministic coefficient functions $\{u^{(v),j}_r\}^R_{r=1}$ on a given sequence of partitions $\{C^{(v),j}_r\}^R_{r=1}$ of $\mathbb{V}_h$ to generate a sequence of approximate feedback functions

$$\mathcal{U}^{(v),j}_R(\cdot) = \sum\nolimits_{r=1}^R u^{(v),j}_r\mathbb{1}_{C^{(v),j}_r}(\cdot), \tag{2.14}$$

such that the cost functional $v \mapsto J\big(M^{(v),\cdot}, \mathcal{U}_R^{(v),\cdot}(M^{(v),\cdot})\big)$ monotonically decreases. Hence, the overall algorithm to approximately solve (2.5), (2.7)–(2.8) consists in each iteration of the following steps:

(1) Simulate realizations of equation (2.5) using Algorithm 2.7 and the approximate feedback control function $\mathcal{U}_R^{(v-1),\cdot}(\cdot)$.
(2) Use these realizations to estimate the coefficients in Algorithm 2.8 where again the control $\mathcal{U}_R^{(v-1),\cdot}(\cdot)$ is used.
(3) Use these coefficients to proceed with an update step with Armijo step size rule to obtain $\mathcal{U}_R^{(v),\cdot}(\cdot)$ for the control, which is based on the maximum principle (2.11).

This procedure has been carried out in [15] for a simpler cost function in the case of one and three nanomagnetic particles. The present work deals with (1.3) which involves a SPDE and thus is far more complex to solve. Also, in the present context, the explicit time discretization of the state equation used in [15] is unfavorable, since multiple particles require a smaller time step size, which results in more computational effort in the simulation of the backward equation. To allow for larger time step sizes k, we use the midpoint rule in Algorithm 2.7, at the expense of solving a system of nonlinear equations per iteration step. Also, the present SPDE case requires more regions $\{C_r^{(v),j}\}$ to resolve the high dimensional state space. Finally, we modify the computation of the coefficients in the adjoint equation (BSDE) similar to [22], such that the computation can be done in parallel; see Section 6 for details.

3. Optimal relaxed controls in the weak formulation

Relaxation is a common strategy for optimal control problems to provide a necessary compactness structure for the control sets in (2.4). While taking primary motivation from [10], our construction of a relaxed control which minimizes the relaxed version of (2.4) bases on variational methods.

Let $\mathcal{P}(\mathbb{L}^2)$ denote the set of all probability measures on $\mathcal{B}(\mathbb{L}^2)$, the Borel σ-algebra on $\mathbb{L}^2$, and $\mathcal{Y}(0,T,\mathbb{L}^2)$ denotes [e] the set of all Young measures on $\mathbb{L}^2$. Let $\{q_t\}_{t\in[0,T]}$ be a $\mathcal{P}(\mathbb{L}^2)$-valued relaxed control process defined on the filtered probability space $\big(\Omega,\mathcal{F},\{\mathcal{F}_t\},P\big)$. We define the associated relaxed functional: for $q \ge 2$

$$\hat{J}(\hat{\pi}) = E\Big[\int_0^T \Big(\|m(t)-\bar{m}(t)\|_{\mathbb{L}^2}^2 + \|v\|_{\mathbb{H}^1}^{2q} q_t(dv)\Big)\,dt + \psi(m(T))\Big], \tag{3.1}$$

with $\hat{\pi} = \big(\Omega,\mathcal{F},\{\mathcal{F}_t\},P,\beta,m,\{q_t\}_{t\in[0,T]}\big)$, where m is a weak martingale solution of the relaxed version of SLLG; see (3.2) below.

Definition 3.1. Let $T>0$ be fixed and $m_0 \in \mathbb{W}^{1,2}(D;\mathbb{S}^2)$. A weak admissible relaxed solution of (2.1) is a 7-tuple $\hat{\pi} = \big(\Omega,\mathcal{F},\{\mathcal{F}_t\},P,\beta,m,\{q_t\}_{t\in[0,T]}\big)$ such that

- $(\Omega,\mathcal{F},P)$ is a complete probability space endowed with the filtration $\{\mathcal{F}_t\}$ satisfying the usual hypotheses, and $\beta(\cdot)$ is an $\mathbb{R}$-valued $\mathcal{F}_t$-adapted Wiener process.
- $\{q_t\}_{t\in[0,T]}$ is a $\mathcal{P}(\mathbb{L}^2)$-valued $\mathcal{F}_t$-predictable relaxed control stochastic process.
- m is a $\mathbb{H}^1$-valued $\mathcal{F}_t$-predictable stochastic process such that for P-a.e $\omega\in\Omega$,
  a). $m(\omega,\cdot)\in C([0,T];\mathbb{L}^2)$ and $|m(\omega,t,x)|=1$ for a.e. $x\in D$ and for all $t\in[0,T]$.
  b). For every $\phi\in C^\infty(\overline{D};\mathbb{R}^3)$ and every $t\in[0,T]$, there holds for P-a.e. $\omega\in\Omega$

$$\begin{aligned}
&(m(t),\phi)_{\mathbb{L}^2} - (m_0,\phi)_{\mathbb{L}^2} \\
&= \int_0^t \big(\nabla m(s), m(s)\times\nabla\phi\big)_{\mathbb{L}^2}\,ds + \alpha\int_0^t \Big(\nabla m(s), \nabla\big(\phi\times m(s)\big)\times m(s)\Big)_{\mathbb{L}^2}\,ds \\
&\quad + \int_0^t \Big(\int_{\mathbb{L}^2} m(s)\times v q_s(dv),\phi\Big)_{\mathbb{L}^2}\,ds + \iota\Big(\int_0^t m(s)\times a\circ d\beta(s),\phi\Big)_{\mathbb{L}^2}.
\end{aligned} \tag{3.2}$$

[e] We refer to [11, 18] and references therein for the definition of Young measures.

We denote the set of weak admissible relaxed solutions by $\hat{\mathcal{U}}_{\mathrm{ad},w}(m_0,T)$. The relaxed optimal control problem is then to find a weak admissible relaxed solution $\hat{\pi}^*$ such that

$$\hat{J}(\hat{\pi}^*) = \inf_{\hat{\pi}\in\hat{\mathcal{U}}_{\mathrm{ad},w}(m_0,T)} \hat{J}(\hat{\pi}) := \Lambda_2. \tag{3.3}$$

We say that $\hat{\pi}^*$ as in (3.3) (resp. q^*, the associated control in $\hat{\pi}^*$) is a weak optimal relaxed solution (resp. weak optimal relaxed control) for (3.1)-(3.2).

Theorem 3.1. *Let $T>0$ be fixed, and $m_0 \in \mathbb{W}^{1,2}(D;\mathbb{S}^2)$. Then the relaxed control problem* (3.3) *admits a weak optimal relaxed control.*

Proof. Thanks to Corollary 2.2, there exists a weak martingale solution of SLLG (2.1) for $u=0$, and hence Λ_2 is finite since ψ is Lipschitz-continuous and the solution m satisfies $|m(t)|=1$. Let $\hat{\pi}_n = \big(\Omega^n, \mathcal{F}^n, \{\mathcal{F}^n_t\}, P^n, \beta_n, m_n, \{q^n_t\}_{t\in[0,T]}\big), n\in\mathbb{N}$, be a minimizing sequence of weak admissible relaxed controls, *i.e.*, $\lim_{n\to\infty}\hat{J}(\hat{\pi}_n)=\Lambda_2$. We will prove the theorem in several steps.

Step 1: For each $n\in\mathbb{N}$, we define the random Young measure λ_n on $\big(\Omega^n,\mathcal{F}^n,P^n\big)$ by

$$\lambda_n(dv,dt) = q^n_t(dv)dt.$$

Since Λ_2 is finite, there exists $R>0$ such that

$$E^n\Big[\int_0^T\int_{\mathbb{L}^2}\|v\|_{\mathbb{H}^1}^{2q}\lambda_n(dv,dt)\Big] = E^n\Big[\int_0^T\int_{\mathbb{L}^2}\|v\|_{\mathbb{H}^1}^{2q}\,q^n_t(dv)\,dt\Big]\le R. \tag{3.4}$$

Define $\eta:[0,T]\times\mathbb{L}^2\to[0,\infty]$ as

$$\eta(t,v) = \begin{cases}\|v\|_{\mathbb{H}^1}^{2q}, & \text{if } v\in\mathbb{H}^1\\ +\infty & \text{otherwise}\end{cases}$$

In view of [10, Example 2.12], we see that η is measurable such that $\eta(t,\cdot)$ is an inf-compact function on $\mathbb{L}^2$ for all $t\in[0,T]$. Thus, by (3.4), we get $E^n\Big[\int_0^T\int_{\mathbb{L}^2}\eta(t,v)\lambda_n(dv,dt)\Big]\le R$, and hence thanks to [3, Definition 3.3] and [11, Theorem 4.3.5], we conclude that the family of laws of $\{\lambda_n\}_{n\in\mathbb{N}}$ is tight on $\mathcal{Y}(0,T;\mathbb{L}^2)$.

Step 2: Since $\{\hat{\pi}_n\}_n$ is a sequence of weak admissible relaxed controls, one has

$$\begin{aligned}&E^n\Big[\sup_{0\le t\le T}\|\nabla m_n(t)\|_{\mathbb{L}^2}^{2q}\Big] + E^n\Big[\Big(\int_0^T\|m_n(s)\times\Delta m_n(s)\|_{\mathbb{L}^2}^2 ds\Big)^q\Big]\\ &\le C_1 + C_2E^n\Big[\int_0^T\int_{\mathbb{L}^2}\|v\|_{\mathbb{H}^1}^{2q}q^n_t(dv)\,dt\Big]\le C,\end{aligned} \tag{3.5}$$

where the last inequality follows from (3.4). Again, P^n a.s., and for all $t\in[0,T]$, the following reformulation of (3.2) with Stratonovich correction term holds

$$\begin{aligned} m_n(t) &= m_0 + \int_0^t m_n(s)\times\Delta m_n(s)\,ds - \alpha\int_0^t m_n(s)\times\big(m_n(s)\times\Delta m_n(s)\big)\,ds\\ &\quad + \int_0^t\int_{\mathbb{L}^2} m_n(s)\times v q^n_s(dv)\,ds + \frac{\iota^2}{2}\int_0^t (m_n(s)\times a)\times a\,ds + \iota\int_0^t m_n(s)\times a\,d\beta_n(s)\\ &\equiv m_0 + \sum_{i=1}^5 \mathbf{B}_{n,i}(t). \end{aligned} \tag{3.6}$$

We want to establish tightness for the sequence of processes $\{m_n\}_{n\in\mathbb{N}}$. It is due to (3.5) that there remains to validate a uniform bound in $\|\cdot\|_{W^{\gamma,4}(0,T;\mathbb{L}^2)}$ for some suitable $0<\gamma<1$ to apply Prokhorov's lemma. By long but elementary estimates there follows for $0<\gamma<\frac{1}{2}$,

$$\sup_n E^n\Big[\|\mathbf{B}_{n,i}\|_{W^{\gamma,4}(0,T;\mathbb{L}^2)}^4\Big]\le C \qquad \forall\ 1\le i\le 5.$$

In view of the above estimates along with (3.6) and the moment estimate (3.5), we see that for $0 < \gamma < \frac{1}{2}$, there exists a constant $C > 0$, independent of n such that

$$\sup_n E^n\Big[\|m_n\|^4_{L^4(0,T;\mathbb{H}^1)\cap W^{\gamma,4}(0,T;\mathbb{L}^2)}\Big] \le C.$$

Let us choose $\gamma \in (0, \frac{1}{2})$ such that $4\gamma > 1$. Then, by [16, Theorems 2.2 and 2.1], $W^{\gamma,4}(0,T;\mathbb{L}^2)$ is compactly embedded into $C([0,T];(\mathbb{H}^1)^*)$, and the embedding of $L^4(0,T;\mathbb{H}^1)\cap W^{\gamma,4}(0,T;\mathbb{L}^2)$ into $L^4(0,T;\mathbb{L}^2)$ is compact. Moreover, for $\gamma > \frac{1}{4}$, we can apply [30, Lemma 5 and Theorem 3] to conclude that the family of the laws of the processes $\{m_n\}_{n\in\mathbb{N}}$ is tight on $C([0,T];\mathbb{L}^2)$. Also, for $i = 1, 2, \cdots, 5$, the family of the laws of the processes $\{\mathbf{B}_{n,i}\}_{n\in\mathbb{N}}$ is tight in $C([0,T];(\mathbb{H}^1)^*)$. Hence, by Prokhorov's theorem, there exist a probability measure μ on $\mathbb{X} := C([0,T];\mathbb{L}^2)\times C([0,T];(\mathbb{H}^1)^*)^5 \times \mathcal{Y}(0,T;\mathbb{L}^2)$ and a subsequence of $\{m_n, \mathbf{B}_{n,1}, \mathbf{B}_{n,2}, \mathbf{B}_{n,3}, \mathbf{B}_{n,4}, \mathbf{B}_{n,5}, \lambda_n\}_n$, still denoted by same index n, such that

$$\mathcal{L}\big(m_n, \mathbf{B}_{n,1}, \mathbf{B}_{n,2}, \mathbf{B}_{n,3}, \mathbf{B}_{n,4}, \mathbf{B}_{n,5}, \lambda_n\big) \to \mu \quad \text{weakly as } \; n \to \infty.$$

Step 3: The space $\mathbb{X}$ is separable and metrizable. Thus, by Dudley's generalization of the Skorokhod representation theorem, there exist a probability space $(\tilde\Omega, \tilde{\mathcal{F}}, \tilde P)$ and a sequence of random variables $\{\tilde m_n, \tilde{\mathbf{B}}_{n,1}, \tilde{\mathbf{B}}_{n,2}, \tilde{\mathbf{B}}_{n,3}, \tilde{\mathbf{B}}_{n,4}, \tilde{\mathbf{B}}_{n,5}, \tilde\lambda_n\}_n$ and $\big(\tilde m, \tilde{\mathbf{B}}_1, \tilde{\mathbf{B}}_2, \tilde{\mathbf{B}}_3, \tilde{\mathbf{B}}_4, \tilde{\mathbf{B}}_5, \tilde\lambda\big)$ with values in $\mathbb{X}$, which are defined on $(\tilde\Omega, \tilde{\mathcal{F}}, \tilde P)$ and satisfy $\tilde P$-a.s.,

$$\begin{cases} \tilde m_n \to \tilde m \; \text{ in } C([0,T];\mathbb{L}^2), \\ \tilde{\mathbf{B}}_{n,i} \to \tilde{\mathbf{B}}_i \; \text{ in } \; C([0,T];(\mathbb{H}^1)^*) \qquad (1 \le i \le 5), \\ \tilde\lambda_n \to \tilde\lambda \; \text{ stably in } \; \mathcal{Y}(0,T;\mathbb{L}^2), \end{cases} \tag{3.7}$$

and

$$\big(\tilde m_n, \tilde{\mathbf{B}}_{n,1}, \tilde{\mathbf{B}}_{n,2}, \tilde{\mathbf{B}}_{n,3}, \tilde{\mathbf{B}}_{n,4}, \tilde{\mathbf{B}}_{n,5}, \tilde\lambda_n\big) \overset{d}{=} \big(m_n, \mathbf{B}_{n,1}, \mathbf{B}_{n,2}, \mathbf{B}_{n,3}, \mathbf{B}_{n,4}, \mathbf{B}_{n,5}, \lambda_n\big) \quad \forall\, n \in \mathbb{N}. \tag{3.8}$$

The sequence $\{\tilde m_n\}_n$ satisfies the same estimates as the original sequence $\{m_n\}_n$. In particular,

$$\begin{cases} \tilde P\text{-a.s.}, \quad |\tilde m_n(t,x)| = 1, \text{ for a.e. } x \in D \text{ and for all } \; t \in [0,T], \\ \sup_n \tilde E\Big[\sup_{0\le t\le T} \|\tilde m_n(t)\|^{2q}_{\mathbb{H}^1} + \Big(\int_0^T \|\tilde m_n(t) \times \Delta\tilde m_n(t)\|^2_{\mathbb{L}^2}\, dt\Big)^q\Big] \le C. \end{cases} \tag{3.9}$$

Furthermore, in view of (3.6) and (3.8), one can conclude that $\tilde P$-a.s.,

$$\tilde m_n(t) = m_0 + \sum_{i=1}^5 \tilde{\mathbf{B}}_{n,i}(t), \quad \text{with} \quad \begin{cases} \tilde{\mathbf{B}}_{n,1}(t) = \int_0^t \tilde m_n(s) \times \Delta\tilde m_n(s)\, ds \\ \tilde{\mathbf{B}}_{n,2}(t) = -\alpha \int_0^t \tilde m_n(s) \times \big(\tilde m_n(s) \times \Delta\tilde m_n(s)\big)\, ds \\ \tilde{\mathbf{B}}_{n,3}(t) = \int_0^t \int_{\mathbb{L}^2} \tilde m_n(s) \times v\, \tilde\lambda_n(dv, ds) \\ \tilde{\mathbf{B}}_{n,4}(t) = \frac{\iota^2}{2} \int_0^t (\tilde m_n(s) \times a) \times a\, ds. \end{cases} \tag{3.10}$$

Step 4: In view of the disintegration theory of measures, since $\tilde\lambda_n : \tilde\Omega \to \mathcal{Y}(0,T;\mathbb{L}^2)$ is a random Young measure such that for every $A \in \mathcal{B}([0,T]\times\mathbb{L}^2)$, the mapping $\tilde\omega \mapsto \tilde\lambda_n(\tilde\omega)(A)$ is measurable and there exists a relaxed control process $\tilde q^n = \{\tilde q^n_t\}_{t\in[0,T]}$ defined on probability space $(\tilde\Omega, \tilde{\mathcal{F}}, \tilde P)$ such that $\tilde P$-a.s., $\tilde\lambda_n(dv, dt) = \tilde q^n_t(dv)\, dt$. Therefore, we can rewrite the term $\tilde{\mathbf{B}}_{n,3}(t)$ as

$$\tilde{\mathbf{B}}_{n,3}(t) = \int_0^t \int_{\mathbb{L}^2} \tilde m_n(s) \times v \tilde q^n_s(dv)\, ds, \quad \tilde P\text{-a.s., and for all } t \in [0,T].$$

Revisiting (3.6), for each $n \in \mathbb{N}$ and $t \in [0,T]$ we define the process $\mathcal{M}_n(t)$ on $(\Omega^n, \mathcal{F}^n, P^n)$ as

$$\mathcal{M}_n(t) = m_n(t) - m_0 - \sum_{i=1}^4 \mathbf{B}_{n,i}(t) = \iota \int_0^t m_n(s) \times a\, d\beta_n(s).$$

Note that $\mathcal{M}_n(\cdot)$ is adapted to the filtration generated by the processes $\{m_n, q^n\}$, and hence is adapted to the filtration $\{\mathcal{F}^n_t\}$. Moreover, it is a $\{\mathcal{F}^n_t\}_{t\in[0,T]}$-martingale with quadratic variation

$$Q_n(t) = \iota^2 \int_0^t |m_n(s) \times a|^2\, ds.$$

For $t \geq 0$, we now define the processes $\widetilde{\mathcal{M}}_n(t)$ and $\widetilde{\mathcal{M}}$ on the probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ via

$$\widetilde{\mathcal{M}}_n(t) = \tilde{m}_n(t) - m_0 - \sum_{i=1}^{4} \tilde{\mathbf{B}}_{n,i}(t) \quad \forall n \in \mathbb{N}, \tag{3.11}$$

$$\widetilde{\mathcal{M}}(t) = \tilde{m}(t) - m_0 - \sum_{i=1}^{4} \tilde{\mathbf{B}}_i(t). \tag{3.12}$$

Define a filtration $\{\tilde{\mathcal{F}}_t^n\}_{t\in[0,T]}$ as

$$\tilde{\mathcal{F}}_t^n = \sigma\big\{\tilde{m}_n(s), \tilde{q}_s^n : 0 \leq s \leq t\big\} \quad \forall t \in [0,T].$$

Observe that thanks to the expressions in (3.10), $\widetilde{\mathcal{M}}_n$ is an $\{\tilde{\mathcal{F}}_t^n\}_{t\in[0,T]}$-adapted process. Define

$$\tilde{Q}_n(t) = \iota^2 \int_0^t |\tilde{m}_n(s) \times a|^2 \, ds.$$

Since $(m_n, q^n) \stackrel{d}{=} (\tilde{m}_n, \tilde{q}^n)$, we see that $\forall t \in [0,T]$, $\mathcal{M}_n(t) \stackrel{d}{=} \widetilde{\mathcal{M}}_n(t)$ and $Q_n(t) \stackrel{d}{=} \tilde{Q}_n(t)$. Thus, we infer that $\widetilde{\mathcal{M}}_n$ is an $\{\tilde{\mathcal{F}}_t^n\}_{t\in[0,T]}$-adapted martingale with quadratic variation $\tilde{Q}_n$.

Step 5: We know that $\tilde{P}$-a.s., $\tilde{m}_n \to \tilde{m}$ in $C([0,T];\mathbb{L}^2)$ as $n \to \infty$. Now we claim that

$$\tilde{E}\Big[\sup_{t\in[0,T]} \|\tilde{m}_n(t) - \tilde{m}(t)\|_{\mathbb{L}^2}^2\Big] \to 0 \quad \big(n \to \infty\big). \tag{3.13}$$

To prove this claim, define $\mathcal{A}_n := \sup_{t\in[0,T]} \|\tilde{m}_n(t) - \tilde{m}(t)\|_{\mathbb{L}^2}^2$. Then $\mathcal{A}_n \to 0$, $\tilde{P}$-a.s. as $n \to \infty$. Invoking Fatou's lemma and the uniform moment estimate (3.9), we have

$$\begin{aligned}
\tilde{E}\big[|\mathcal{A}_n|^2\big] &\leq C\tilde{E}\Big[\sup_{t\in[0,T]} \|\tilde{m}_n(t)\|_{\mathbb{L}^2}^4 + \sup_{t\in[0,T]} \|\tilde{m}(t)\|_{\mathbb{L}^2}^4\Big] \\
&\leq C\bigg\{\tilde{E}\Big[\sup_{t\in[0,T]} \|\tilde{m}_n(t)\|_{\mathbb{L}^2}^4\Big] + \liminf_n \tilde{E}\Big[\sup_{t\in[0,T]} \|\tilde{m}_n(t)\|_{\mathbb{L}^2}^4\Big]\bigg\} \\
&\leq C \sup_n \tilde{E}\Big[\sup_{t\in[0,T]} \|\tilde{m}_n(t)\|_{\mathbb{L}^2}^4\Big] \leq C.
\end{aligned}$$

This implies that the family $\{\mathcal{A}_n\}_{n\in\mathbb{N}}$ is uniformly integrable. Hence, we may use Vitali's convergence theorem to validate (3.13). Again, in view of (3.7) and the definition of $\widetilde{\mathcal{M}}_n(t)$, we have that $\tilde{P}$-a.s., $\widetilde{\mathcal{M}}_n \to \widetilde{\mathcal{M}}$ in $C([0,T];(\mathbb{H}^1)^*)$ as $n \to \infty$. Now we claim that

$$\tilde{E}\Big[\sup_{t\in[0,T]} \|\widetilde{\mathcal{M}}_n(t) - \widetilde{\mathcal{M}}(t)\|_{(\mathbb{H}^1)^*}^2\Big] \to 0 \quad \big(n \to \infty\big). \tag{3.14}$$

To show that (3.14) holds, we define $\mathcal{B}_n := \sup_{t\in[0,T]} \|\widetilde{\mathcal{M}}_n(t) - \widetilde{\mathcal{M}}(t)\|_{(\mathbb{H}^1)^*}^2$. Then $\mathcal{B}_n \to 0$, $\tilde{P}$-a.s. for $n \to \infty$. By using the uniform moment estimate (3.9) along with (3.11), we see that

$$\sup_n \tilde{E}\Big[\sup_{t\in[0,T]} \|\widetilde{\mathcal{M}}_n(t)\|_{(\mathbb{H}^1)^*}^4\Big] \leq C. \tag{3.15}$$

Again, we may invoke Fatou's lemma and (3.15) to establish (3.14) by Vitali's convergence theorem.

Step 6: In this step, we will identify the limiting processes $\tilde{\mathbf{B}}_1, \tilde{\mathbf{B}}_2, \tilde{\mathbf{B}}_3$ and $\tilde{\mathbf{B}}_4$ in (3.12). First we want to identify $\tilde{\mathbf{B}}_3$. Since $\tilde{\lambda}(\tilde{\omega}) \in \mathcal{Y}(0,T;\mathbb{L}^2)$, like in Step 4, there exists a relaxed control process $\{\tilde{q}_t\}_{t\in[0,T]}$ defined on $\big(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}\big)$ such that $\tilde{P}$-a.s., $\tilde{\lambda}(dv, dt) = \tilde{q}_t(dv)\, dt$. Let

$$\tilde{f}_n(t) = \int_{\mathbb{L}^2} \tilde{m}_n(t) \times v \, \tilde{q}_t^n(dv), \quad \hat{f}_n(t) = \int_{\mathbb{L}^2} \tilde{m}(t) \times v \, \tilde{q}_t^n(dv), \quad \text{and} \quad \tilde{f}(t) = \int_{\mathbb{L}^2} \tilde{m}(t) \times v \, \tilde{q}_t(dv).$$

By using (3.13) and

$$|\tilde{m}(t,x)| \leq |\tilde{m}_n(t,x)| = 1, \text{ for a.e. } x \in D, \text{ and all } t \in [0,T], \tilde{P}\text{-a.s.}, \tag{3.16}$$

we easily find

$$\tilde{f}_n - \hat{f}_n \to 0 \ \text{ in } L^2(\tilde{\Omega} \times [0,T]; \mathbb{L}^2).$$

Next we infer that

$$\hat{f}_n \rightharpoonup \tilde{f} \text{ in } L^2(\tilde{\Omega} \times [0,T]; (\mathbb{H}^1)^*) \quad \big(n \to \infty\big). \tag{3.17}$$

To prove this, let $\psi \in (L^2(\tilde{\Omega} \times [0,T]; (\mathbb{H}^1)^*))^* = L^2(\tilde{\Omega} \times [0,T]; \mathbb{H}^1)$ be fixed. We need to show that

$$\tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} \Big\langle \tilde{m}(t) \times v, \psi(t) \Big\rangle \tilde{\lambda}_n(dv, dt)\Big] \overset{n\to\infty}{\longrightarrow} \tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} \Big\langle \tilde{m}(t) \times v, \psi(t) \Big\rangle \tilde{\lambda}(dv, dt)\Big],$$

where $\big\langle \cdot \big\rangle$ denotes the duality pairing between $(\mathbb{H}^1)^*$ and $\mathbb{H}^1$. For this purpose, denote by $T_k \in C^\infty(\mathbb{R})$ a truncation such that

$$1_{[-k,k]} \le T_k \le 1_{[-k-1,k+1]}.$$

Thus, one gets that

$$\begin{aligned}
&\Big|\tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} \Big\langle \tilde{m}(t) \times v, \psi(t) \Big\rangle \tilde{\lambda}_n(dv, dt)\Big] - \tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} \Big\langle \tilde{m}(t) \times v, \psi(t) \Big\rangle \tilde{\lambda}(dv, dt)\Big]\Big| \\
&\le \Bigg|\tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} T_k(\|v\|_{\mathbb{L}^2}) \Big\langle \tilde{m}(t) \times v, \psi(t) \Big\rangle \tilde{\lambda}_n(dv, dt)\Big] \\
&\qquad\qquad - \tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} T_k(\|v\|_{\mathbb{L}^2}) \Big\langle \tilde{m}(t) \times v, \psi(t) \Big\rangle \tilde{\lambda}(dv, dt)\Big]\Bigg| \\
&\quad + \Big|\tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} [1 - T_k(\|v\|_{\mathbb{L}^2})] \|v\|_{\mathbb{L}^2} \|\psi(t)\|_{\mathbb{H}^1} \tilde{\lambda}_n(dv, dt)\Big]\Big| \\
&\qquad + \Big|\tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} [1 - T_k(\|v\|_{\mathbb{L}^2})] \|v\|_{\mathbb{L}^2} \|\psi(t)\|_{\mathbb{H}^1} \tilde{\lambda}(dv, dt)\Big]\Big|.
\end{aligned}$$

Note that

$$\begin{aligned}
&\Big|\tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} [1 - T_k(\|v\|_{\mathbb{L}^2})] \|v\|_{\mathbb{L}^2} \|\psi(t)\|_{\mathbb{H}^1} \tilde{\lambda}_n(dv, dt)\Big]\Big| \\
&\le \Big|\tilde{E}\Big[\int_0^T \|\psi(t)\|_{\mathbb{H}^1} \int_{\|v\|_{\mathbb{L}^2} > k} \|v\|_{\mathbb{L}^2} \tilde{\lambda}_n(dv, dt)\Big]\Big| \\
&\le C \|\psi\|_{L^2(\tilde{\Omega} \times [0,T]; \mathbb{H}^1)} \left(\int_{\tilde{\Omega}} \int_0^T \int_{\{v \in \mathbb{L}^2 : \|v\|_{\mathbb{L}^2} > k\}} \|v\|^2_{\mathbb{L}^2} \, \tilde{\lambda}_n(\tilde{\omega})(dv, dt) \tilde{P}(d\tilde{\omega}) \right)^{\frac{1}{2}} \\
&\le C \|\psi\|_{L^2(\tilde{\Omega} \times [0,T]; \mathbb{H}^1)} \frac{1}{k} \left(\tilde{E}\Big[\int_0^T \int_{\{v \in \mathbb{L}^2 : \|v\|_{\mathbb{L}^2} > k\}} \|v\|^4_{\mathbb{L}^2} \, \tilde{\lambda}_n(dv, dt)\Big] \right)^{\frac{1}{2}} \\
&\le \frac{C}{k} \|\psi\|_{L^2(\tilde{\Omega} \times [0,T]; \mathbb{H}^1)}
\end{aligned} \tag{3.18}$$

which holds uniformly with respect to $n \in \mathbb{N}$. Note that $\mathbb{L}^2 \ni v \mapsto \|v\|^4_{\mathbb{L}^2}$ is lower semi-continuous. Thus, as $\tilde{\lambda}_n \to \tilde{\lambda}$ stably in $\mathcal{Y}(0,T;\mathbb{L}^2)$, we obtain $\tilde{P}-$a.s.,

$$\int_0^T \int_{\mathbb{L}^2} \|v\|^4_{\mathbb{L}^2} \tilde{\lambda}(dv, dt) \le \liminf \int_0^T \int_{\mathbb{L}^2} \|v\|^4_{\mathbb{L}^2} \tilde{\lambda}_n(dv, dt),$$

and hence, thanks to Fatou's lemma,

$$\tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} \|v\|^4_{\mathbb{L}^2} \tilde{\lambda}(dv, dt)\Big] \le \liminf \tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} \|v\|^4_{\mathbb{L}^2} \tilde{\lambda}_n(dv, dt)\Big] \le C.$$

Therefore, (3.18) holds if we replace $\tilde{\lambda}_n$ by $\tilde{\lambda}$ on the left-hand side. In other words, one has

$$\tilde{E}\Big[\int_0^T \int_{\{v \in \mathbb{L}^2 : \|v\|_{\mathbb{L}^2} > k\}} \Big\langle \tilde{m}(t) \times v, \psi(t) \Big\rangle \, \tilde{\lambda}_n(dv, dt)\Big] \le \frac{C}{k} \|\psi\|_{L^2(\tilde{\Omega} \times [0,T]; \mathbb{H}^1)}.$$

Finally, in view of [10, Theorem 2.16], one gets

$$\limsup_n \Big|\tilde{E}\Big[\int_0^T\int_{\mathbb{L}^2}\Big\langle \tilde{m}(t)\times v,\psi(t)\Big\rangle\tilde{\lambda}_n(dv,dt)\Big]-\tilde{E}\Big[\int_0^T\int_{\mathbb{L}^2}\Big\langle \tilde{m}(t)\times v,\psi(t)\Big\rangle\tilde{\lambda}(dv,dt)\Big]\Big| \\ \le \frac{C}{k}\|\psi\|_{L^2(\tilde{\Omega}\times[0,T];\mathbb{H}^1)}$$

for any k, and then (3.17) holds by passing to the limit with respect to k.

We are now in a position to identify the limiting process $\tilde{\mathbf{B}}_3(t)$. As, $\tilde{P}$-a.s., $\tilde{\mathbf{B}}_{n,3}\to\tilde{\mathbf{B}}_3$ in $C([0,T];(\mathbb{H}^1)^*)$, in view of the above discussion, we get that for all $t\in[0,T]$

$$\tilde{\mathbf{B}}_3(t)=\int_0^t\int_{\mathbb{L}^2}\tilde{m}(s)\times v\,\tilde{q}_s(dv)\,ds,\quad \tilde{P}\text{-a.s.}$$

Again, one can show by arguments similar to those in [8, Lemmas 4.5 and 4.9] that for all $t\in[0,T]$ and $\tilde{P}$-a.s.,

$$\begin{cases}\tilde{\mathbf{B}}_{n,1}(t)\rightharpoonup\tilde{\mathbf{B}}_1(t)=\int_0^t\tilde{m}(s)\times\Delta\tilde{m}(s)\,ds\\ \tilde{\mathbf{B}}_{n,2}(t)\rightharpoonup\tilde{\mathbf{B}}_2(t)=-\alpha\int_0^t\tilde{m}(s)\times\big(\tilde{m}(s)\times\Delta\tilde{m}(s)\big)\,ds \ \text{ in } L^2(\tilde{\Omega}\times[0,T];(\mathbb{H}^1)^*)\quad\big(n\to\infty\big)\\ \tilde{\mathbf{B}}_{n,4}(t)\rightharpoonup\tilde{\mathbf{B}}_4(t)=\frac{\iota^2}{2}\int_0^t(\tilde{m}(s)\times a)\times a\,ds.\end{cases}$$

Step 7: Define a right continuous filtration $\{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}$ by

$$\tilde{\mathcal{F}}_t=\sigma\big\{(\tilde{m}(s),\tilde{q}_s):0\le s\le t\big\},\quad t\in[0,T].$$

In view of the definition of $\widetilde{\mathcal{M}}$ along with the identification of the terms $\tilde{\mathbf{B}}_1,\tilde{\mathbf{B}}_2,\tilde{\mathbf{B}}_3$ and $\tilde{\mathbf{B}}_4$, it is obvious that $\widetilde{\mathcal{M}}$ is an $\{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}$ -adapted stochastic process with values in $(\mathbb{H}^1)^*$. Next we claim that $\widetilde{\mathcal{M}}(\cdot)$ *is an $\tilde{\mathcal{F}}_t$-martingale with quadratic variation*

$$\tilde{Q}(t)=\iota^2\int_0^t|\tilde{m}(s)\times a|^2\,ds. \tag{3.19}$$

Let $0<s<t\le T$ and $\phi\in C_b\big(C([0,s];(\mathbb{H}^1)^*)\times\mathcal{Y}(0,s;\mathbb{L}^2)\big)$. Note that, since $\widetilde{\mathcal{M}}_n$ is an $\tilde{\mathcal{F}}_t^n$-martingale, we have

$$\tilde{E}\Big[\big(\widetilde{\mathcal{M}}_n(t)-\widetilde{\mathcal{M}}_n(s)\big)\phi(\tilde{m}_n,\tilde{\lambda}_n)\Big]=0.$$

Moreover, we use (3.13) and (3.14) to have

$$0=\tilde{E}\Big[\big(\widetilde{\mathcal{M}}_n(t)-\widetilde{\mathcal{M}}_n(s)\big)\phi(\tilde{m}_n,\tilde{\lambda}_n)\Big]\overset{n\to\infty}{\longrightarrow}\tilde{E}\Big[\big(\widetilde{\mathcal{M}}(t)-\widetilde{\mathcal{M}}(s)\big)\phi(\tilde{m},\tilde{\lambda})\Big].$$

This gives the martingale property of $\widetilde{\mathcal{M}}$ with respect to the filtration $\tilde{\mathcal{F}}_t$. Again, since $a\in\mathbb{W}^{1,\infty}$, thanks to (3.16) and (3.13), we verify

$$\tilde{E}\Big[\sup_{t\in[0,T]}\|\tilde{Q}_n(t)-\tilde{Q}(t)\|^2_{L^2(D)}\Big]\le C(\iota,a)\tilde{E}\Big[\sup_{0\le t\le T}\|\tilde{m}_n(s)-\tilde{m}(s)\|^2_{\mathbb{L}^2}\Big]\to 0\ \big(n\to\infty\big).$$

Combining these results, as $\widetilde{\mathcal{M}}_n$ is an $\tilde{\mathcal{F}}_t^n$-martingale with quadratic variation $\tilde{Q}_n$, we have for any $\phi\in C_b\big(C([0,s];(\mathbb{H}^1)^*)\times\mathcal{Y}(0,s;\mathbb{L}^2)\big)$ with $0<s<t\le T$,

$$0=\tilde{E}\Big[\big(\widetilde{\mathcal{M}}_n^2(t)-\widetilde{\mathcal{M}}_n^2(s)-(\tilde{Q}_n(t)-\tilde{Q}_n(s))\big)\phi(\tilde{m}_n,\tilde{\lambda}_n)\Big] \\ \overset{n\to\infty}{\longrightarrow}\tilde{E}\Big[\big(\widetilde{\mathcal{M}}^2(t)-\widetilde{\mathcal{M}}^2(s)-(\tilde{Q}(t)-\tilde{Q}(s))\big)\phi(\tilde{m},\tilde{\lambda})\Big].$$

Thus, noting the martingale property of $\widetilde{\mathcal{M}}$, we conclude that $\tilde{Q}$ is the quadratic variation of $\widetilde{\mathcal{M}}$ where $\tilde{Q}$ is defined by (3.19).

Step 8: We have shown that $\tilde{P}$-a.s.,

$$\widetilde{\mathcal{M}}(t)=\tilde{m}(t)-m_0-\int_0^t\tilde{m}(s)\times\Delta\tilde{m}(s)\,ds+\alpha\int_0^t\tilde{m}(s)\times\big(\tilde{m}(s)\times\Delta\tilde{m}(s)\big)\,ds \\ -\int_0^t\int_{\mathbb{L}^2}\tilde{m}(s)\times v\,\tilde{q}_s(dv)\,ds-\frac{\iota^2}{2}\int_0^t(\tilde{m}(s)\times a)\times a\,ds \tag{3.20}$$

is a $\tilde{\mathcal{F}}_t$-adapted martingale with quadratic variation $\tilde{Q}(t)$. Thus, by the martingale representation theorem, there exist an extension of probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, still denoted by $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, and a $\mathbb{R}$-valued Wiener process $\tilde{\beta}$ defined on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ such that

$$\widetilde{\mathcal{M}}(t) = \iota \int_0^t \tilde{m}(s) \times a \, d\tilde{\beta}(s). \tag{3.21}$$

Combining (3.20) and (3.21), we conclude that $\tilde{P}$-a.s.,

$$\begin{aligned} \tilde{m}(t) = m_0 &+ \int_0^t \tilde{m}(s) \times \Delta \tilde{m}(s) \, ds - \alpha \int_0^t \tilde{m}(s) \times \big(\tilde{m}(s) \times \Delta \tilde{m}(s)\big) \, ds \\ &+ \int_0^t \int_{\mathbb{L}^2} \tilde{m}(s) \times v \, \tilde{q}_s(dv) \, ds + \iota \int_0^t \tilde{m}(s) \times a \circ d\tilde{\beta}(s) \quad \text{in } (\mathbb{H}^1)^*. \end{aligned} \tag{3.22}$$

i.e., $\hat{\pi} = \big(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}, \tilde{\beta}, \tilde{m}, \{\tilde{q}_t\}_{t \in [0,T]}\big) \in \hat{\mathcal{U}}_{\mathrm{ad},w}(m_0, T)$ if we show that $\tilde{P}$-a.s.,

$$|\tilde{m}(t,x)| = 1 \quad \forall\, t \in [0,T], \text{ and a.e. } x \in D. \tag{3.23}$$

We can achieve (3.23) in the following way: Let $\phi \in C_0^\infty(D)$. We apply Itô's formula to the function $\mathbb{L}^2 \ni u \mapsto \big(u, \phi u\big)_{\mathbb{L}^2}$ and arrive at (cf. [8, Proof of (2.11)] under small changes)

$$\big(\tilde{m}(t), \phi \tilde{m}(t)\big)_{\mathbb{L}^2} = \big(m_0, \phi m_0\big)_{\mathbb{L}^2}.$$

for all $t \in [0,T]$. Since ϕ is arbitrary and $|m_0(x)| = 1$ for a.e. $x \in D$, we infer that $\tilde{P}$-a.s., $|\tilde{m}(t,x)| = 1$, for a.e. $x \in D$ and for all $t \in [0,T]$.

Step 9: Observe that

$$\begin{aligned} S : [0,T] \times \mathbb{L}^2 &\longrightarrow [0, \infty] \\ (t, v) &\longmapsto \|v\|_{\mathbb{H}^1}^{2q} \end{aligned}$$

is a measurable, non-negative, and lower semi-continuous convex function. Thus, since $\tilde{\lambda}_n \to \tilde{\lambda}$ stably in $\mathcal{Y}(0,T;\mathbb{L}^2)$, invoking [11, Proposition 2.1.12], (3.7), and the property (3.8) along with Fatou's lemma, we get

$$\begin{aligned} \hat{J}(\hat{\pi}) &= \tilde{E}\Big[\int_0^T \|\tilde{m}(t) - \bar{m}(t)\|_{\mathbb{L}^2}^2 \, dt + \int_0^T \int_{\mathbb{L}^2} \|v\|_{\mathbb{H}^1}^{2q} \, \tilde{\lambda}(dv, dt)\Big] + \tilde{E}\big[\psi(\tilde{m}(T))\big] \\ &\le \liminf_{n \to \infty} \bigg\{ \tilde{E}\Big[\int_0^T \|\tilde{m}_n(t) - \bar{m}(t)\|_{\mathbb{L}^2}^2 \, dt + \int_0^T \int_{\mathbb{L}^2} \|v\|_{\mathbb{H}^1}^{2q} \, \tilde{\lambda}_n(dv, dt)\Big] + \tilde{E}\big[\psi(\tilde{m}_n(T))\big] \bigg\} \\ &= \liminf_{n \to \infty} \bigg\{ E^n\Big[\int_0^T \int_{\mathbb{L}^2} \big(\|m_n(t) - \bar{m}(t)\|_{\mathbb{L}^2}^2 + \|v\|_{\mathbb{H}^1}^{2q}\big) \lambda_n(dv, dt)\Big] + E^n\big[\psi(m_n(T))\big] \bigg\} \\ &= \liminf_{n \to \infty} \hat{J}(\hat{\pi}_n) = \Lambda_2. \end{aligned}$$

This implies that $\hat{\pi}$ is a weak optimal relaxed solution for the control problem (3.3) and this finishes the proof. □

4. Proof of Theorem 2.3

With the help of Theorem 3.1, we may now prove Theorem 2.3.
For $\pi_n^1 = \big(\Omega^n, \mathcal{F}^n, \{\mathcal{F}_t^n\}, P^n, \beta_n, m_n, u_n\big)$, let $\{\pi_n^1; n \in \mathbb{N}\}$ be a minimizing sequence of weak admissible controls *i.e.*,

$$\lim_{n \to \infty} J(\pi_n^1) = \Lambda_1. \tag{4.1}$$

Define $q_t^n(dv) = \delta_{u_n(t)}(dv)$ in the weak relaxed control problem. Then $\{\hat{\pi}_n^2; n \in \mathbb{N}\}$ where $\hat{\pi}_n^2 = \big(\Omega^n, \mathcal{F}^n, \{\mathcal{F}_t^n\}, P^n, \beta_n, m_n, \{q_t^n\}_{t \in [0,T]}\big)$ is a decreasing sequence of weak admissible relaxed controls for the same underlying problem. By evidence, $J(\pi_n^1) = \hat{J}(\hat{\pi}_n^2)$. By the proof of Theorem

3.1, there exists a weak solution $\hat{\pi} = \big(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}, \tilde{\beta}, \tilde{m}, \{\tilde{q}_t\}_{t\in[0,T]}\big)$ such that (3.22) holds. We now define the $\{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}$-predictable stochastic process $\tilde{u}$ *via*

$$\tilde{u}(t) = \int_{\mathbb{L}^2} v\,\tilde{q}_t(dv). \tag{4.2}$$

Evidently $\tilde{E}\Big[\int_0^T \|\tilde{u}(s)\|_{\mathbb{H}^1}^{2q}\,ds\Big] \le \tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} \|v\|_{\mathbb{H}^1}^{2q} \tilde{q}_s(dv)\,ds\Big] < +\infty$. Moreover, we have $\tilde{P}$-a.s.,

$$\|\tilde{u}(t)\|_{\mathbb{L}^2} \le K \quad \text{for a.e. } t \in [0,T]. \tag{4.3}$$

To prove this, let $\Psi : x \mapsto (|x| - K)^+$. For $\tilde{\lambda} = \tilde{q}_s(dv)\,ds$, and $\tilde{\lambda}_n = \tilde{q}^n_s(dv)\,ds$, we have $\tilde{\lambda}_n \to \tilde{\lambda}$ stably in $\mathcal{Y}(0,T;\mathbb{L}^2)$ and $\tilde{\lambda}_n \overset{d}{=} \lambda_n\big(= \delta_{u_n(t,\cdot)}(dv)dt\big)$ (see (3.7)). Thus by Jensen inequality, stable convergence of Young measures, and then equality of laws, one has that

$$\begin{aligned}
\tilde{E}\Big[\int_0^T \Psi\Big(\|\tilde{u}(t)\|_{\mathbb{L}^2}\Big)\,ds\Big] &\le \tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} \Psi\Big(\|v\|_{\mathbb{L}^2}\Big)\,\tilde{q}_s(dv)\,ds\Big] \\
&\le \liminf_n \tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} \Psi\Big(\|v\|_{\mathbb{L}^2}\Big)\,\tilde{q}^n_s(dv)\,ds\Big] \\
&\le \liminf_n E^n\Big[\int_0^T \int_{\mathbb{L}^2} \Psi\Big(\|v\|_{\mathbb{L}^2}\Big)\,q^n_s(dv)\,ds\Big] \\
&= \liminf_n E^n\Big[\int_0^T \Psi\Big(\|u_n(s)\|_{\mathbb{L}^2}\Big)\,ds\Big] = 0,
\end{aligned}$$

since $\|u_n(s)\|_{\mathbb{L}^2} \le K$ for a.e. $s \in [0,T]$ and P^n- a.s. This validates (4.3). Now, since the control acts linearly, we have

$$\int_0^t \int_{\mathbb{L}^2} \tilde{m}(s) \times v\,\tilde{q}_s(dv)\,ds = \int_0^t \big(\tilde{m}(s) \times \int_{\mathbb{L}^2} v\,\tilde{q}_s(dv)\big)\,ds = \int_0^t \tilde{m}(s) \times \tilde{u}(s)\,ds.$$

Thus, we see that the following stochastic PDE holds: $\tilde{P}$-a.s. and for all $t \in [0,T]$

$$\begin{aligned}
\tilde{m}(t) = m_0 + \int_0^t \tilde{m}(s) \times \Delta\tilde{m}(s)\,ds - \alpha \int_0^t \tilde{m}(s) \times \big(\tilde{m}(s) \times \Delta\tilde{m}(s)\big)\,ds \\
+ \int_0^t \tilde{m}(s) \times \tilde{u}(s)\,ds + \iota \int_0^t \tilde{m}(s) \times a \circ d\tilde{\beta}(s).
\end{aligned}$$

Hence $\pi^* = \big(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}, \tilde{\beta}, \tilde{m}, \tilde{u}\big) \in \mathcal{U}^w_{\text{ad}}(m_0,T)$. Now, in view of Jensen's inequality, we obtain

$$\begin{aligned}
J(\pi^*) &= \tilde{E}\Big[\int_0^T \big(\|\tilde{m}(t) - \bar{m}(t)\|_{\mathbb{L}^2}^2 + \|\tilde{u}(t)\|_{\mathbb{H}^1}^{2q}\,dt\Big] + \tilde{E}\big[\psi(\tilde{m}(T))\big] \\
&= \tilde{E}\Big[\int_0^T \big(\|\tilde{m}(t) - \bar{m}(t)\|_{\mathbb{L}^2}^2 + \|\int_{\mathbb{L}^2} v\tilde{q}_t(dv)\|_{\mathbb{H}^1}^{2q}\,dt\Big] + \tilde{E}\big[\psi(\tilde{m}(T))\big] \\
&\le \tilde{E}\Big[\int_0^T \int_{\mathbb{L}^2} \Big(\|\tilde{m}(t) - \bar{m}(t)\|_{\mathbb{L}^2}^2 + \|v\|_{\mathbb{H}^1}^{2q}\Big)\tilde{\lambda}(dv,dt)\Big] + \tilde{E}\big[\psi(\tilde{m}(T))\big] \\
&\le \liminf_{n\to\infty} \hat{J}(\hat{\pi}^2_n) = \lim_{n\to\infty} J(\pi^1_n) = \Lambda_1.
\end{aligned}$$

In other words, $\pi^* = \big(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}, \tilde{\beta}, \tilde{m}, \tilde{u}\big)$ is a weak optimal solution for the control problem (2.1)-(2.4) and this finishes the proof of Theorem 2.3.

Remark 4.1. We proved Theorem 2.3 by constructing an optimal relaxed solution $\hat{\pi}^*$ of (3.3) first via the compactness of Young measures on $\mathcal{Y}(0,T;\mathbb{L}^2)$, and then the process defined in (4.2) was shown to be part of a weak optimal solution π^* of the problem (1.3) where we exploited the fact that the control acts linearly in the PDE (2.1). A different strategy to verify Theorem 2.3, which is also based on the linearity of control in (2.1), is to use the Jakubowski-Skorokhod representation theorem (cf. [25]) instead of the classical Skorokhod representation theorem. Let $\{\pi_n\}_n$ be such that (4.1) holds. Then $\sup_n E_n\big[\int_0^T \|u_n(t)\|_{\mathbb{H}^1}^{2q}\,dt\big] < R$ for some $R > 0$. Following Step 2 in the proof of Theorem 3.1, we see that $\big\{\mathcal{L}(m_n, \mathbf{B}_{n,1}, \mathbf{B}_{n,2}, \mathbf{B}_{n,3}, \mathbf{B}_{n,4}, \mathbf{B}_{n,5}, u_n)\big\}_{n\in\mathbb{N}}$ is tight on the space $\mathbb{X} := C([0,T];\mathbb{L}^2)\times C([0,T];(\mathbb{H}^1)^*)^5 \times (L^{2q}(0,T;\mathbb{H}^1),w)^{\mathrm{f}}$; see Step 2 in the proof of Theorem 3.1 for notation, where $\mathbf{B}_{n,3}$ is replaced by $\int_0^t m_n(s)\times u_n(s)\,ds$. Note that $(\mathbb{X},w)$ is not Polish but a separable space endowed with the weak topology. We may apply the Jakubowski-Skorokhod representation theorem on the space $(\mathbb{X},w)$ to ensure the existence of a probability space $(\tilde\Omega,\tilde{\mathcal{F}},\tilde P)$, and $\mathbb{X}$-valued Borel-measurable random variables $\{\tilde m_n,\tilde{\mathbf{B}}_{n,1},\tilde{\mathbf{B}}_{n,2},\tilde{\mathbf{B}}_{n,3},\tilde{\mathbf{B}}_{n,4},\tilde{\mathbf{B}}_{n,5},\tilde u_n\}_n$ and $\big(\tilde m,\tilde{\mathbf{B}}_1,\tilde{\mathbf{B}}_2,\tilde{\mathbf{B}}_3,\tilde{\mathbf{B}}_4,\tilde{\mathbf{B}}_5,\tilde u\big)$ such that

$$\big(\tilde m_n,\tilde{\mathbf{B}}_{n,1},\tilde{\mathbf{B}}_{n,2},\tilde{\mathbf{B}}_{n,3},\tilde{\mathbf{B}}_{n,4},\tilde{\mathbf{B}}_{n,5},\tilde u_n\big)\overset{d}{=}\big(m_n,\mathbf{B}_{n,1},\mathbf{B}_{n,2},\mathbf{B}_{n,3},\mathbf{B}_{n,4},\mathbf{B}_{n,5},u_n\big)\quad\forall\, n\in\mathbb{N}$$

and $\tilde P$-a.s., $\big(\tilde m_n,\tilde{\mathbf{B}}_{n,1},\tilde{\mathbf{B}}_{n,2},\tilde{\mathbf{B}}_{n,3},\tilde{\mathbf{B}}_{n,4},\tilde{\mathbf{B}}_{n,5},\tilde u_n\big)$ converges to $\big(\tilde m,\tilde{\mathbf{B}}_1,\tilde{\mathbf{B}}_2,\tilde{\mathbf{B}}_3,\tilde{\mathbf{B}}_4,\tilde{\mathbf{B}}_5,\tilde u\big)$ in the topology of $\mathbb{X}$. One may now go through the following Steps 4-5 and identity the limiting processes $\tilde{\mathbf{B}}_i : i=1,2,4$ as before. Since the control acts linearly in the equation (2.1), by using the weak convergence of $\tilde u_n$ and the strong convergence of $\tilde m_n$, it is easy to see that $\tilde{\mathbf{B}}_3(t)=\int_0^t \tilde m(s)\times\tilde u(s)\,ds$. The argumentation in Steps 7-9 then establishes that $\tilde\pi := \big(\tilde\Omega,\tilde{\mathcal{F}},\{\tilde{\mathcal{F}}_t\},\tilde P,\tilde\beta,\tilde m,\tilde u\big)$ is a weak optimal solution for the control problem (2.1)-(2.4).

5. Finite element approximation of the optimal control problem

In this section, we prove Theorem 2.4 for the finite element approximation of problem (1.3) and $d=1$, $q\geq 4$. We already pointed out that probabilistically strong solutions of SLLG (2.1), and their higher spatial regularity properties are needed for this purpose, which exist thanks to Corollary 2.2. Let $\pi^*=\big(\Omega^*,\mathcal{F}^*,\{\mathcal{F}_t^*\},P^*,\beta^*,m^*,u^*\big)$ be a 7-tuple from Theorem 2.4. By Corollary 2.2, (m^*,u^*) satisfies the following estimates:

$$\begin{cases} P^*\text{-a.s.},\ \ |m^*(t,x)|=1,\ \text{for all } t\in[0,T]\ \text{ and every } x\in D,\\ E^*\Big[\sup\limits_{0\le t\le T}\|m^*(t)\|_{\mathbb{H}^1}^{2q}+\int_0^T\|\Delta m^*(t)\|_{\mathbb{L}^2}^2\,dt\Big]<+\infty,\\ P^*\text{-a.s.}, \|u^*(t)\|_{\mathbb{L}^2}\le K\ \text{ for a.e. } t\in[0,T] \text{ and } \ E^*\Big[\int_0^T\|u^*(t)\|_{\mathbb{H}^1}^{2q}\,dt\Big]<+\infty.\end{cases}\tag{5.1}$$

Consider also (7.21), where u is replaced by $\mathcal{R}_h u^*$ on the stochastic basis $\big(\Omega^*,\mathcal{F}^*,\{\mathcal{F}_t^*\},P^*\big)$ and by β^*. By Lemma 7.3, the SDE (7.21) with $u=\mathcal{R}_h u^*$ has a unique continuous, $\{\mathcal{F}_t^*\}$-adapted strong solution m_h. Below, we use the $\mathbb{L}^2$-projection $\mathcal{P}_h$ defined in (7.9).

Lemma 5.1. *Let $\big(\Omega^*,\mathcal{F}^*,\{\mathcal{F}_t^*\},P^*,\beta^*,m^*,u^*\big)$ be a 7-tuple as stated in Theorem 2.4. Let m_h be a strong solution of (7.21) with $u=\mathcal{R}_h u^*$ on the same filtered probability space $\big(\Omega^*,\mathcal{F}^*,\{\mathcal{F}_t^*\},P^*\big)$. Set*

$$\Omega_{R,t}^{*,h}=\Big\{\omega\in\Omega^*:\sup_{s\in[0,t]}\|m^*(s)\|_{\mathbb{H}^1}^4\le R\Big\}\cap\Big\{\omega\in\Omega^*:\sup_{s\in[0,t]}\|m_h(s)\|_{\mathbb{H}^1}^4\le R\Big\}$$

where $R>0$ is fixed. Then, there exists a constant $C>0$, independent of h and R such that

$$\sup_{0\le t\le T}E^*\Big[\mathbf{1}_{\Omega_{R,t}^{*,h}}\|\mathcal{P}_h(m_h-m^*)(t)\|_{\mathbb{L}^2}^2\Big]\le ChR\exp(CTR).$$

Proof. **Step 1:** *Error equation and its estimate.* Define $\xi_h=m_h-m^*$. Then $\mathcal{P}_h\xi_h$ satisfies the SDE

$$\begin{aligned} d\mathcal{P}_h\xi_h(t)=\Big\{&-\alpha\big(\mathcal{I}_h[m_h(t)\times(m_h(t)\times\Delta_h m_h(t))]-\mathcal{P}_h[m^*(t)\times(m^*(t)\times\Delta m^*(t))]\big)\\ &+\big(\mathcal{I}_h[m_h(t)\times\Delta_h m_h(t)]-\mathcal{P}_h[m^*(t)\times\Delta m^*(t)]\big)\\ &+\big(\mathcal{I}_h[m_h(t)\times\mathcal{R}_h u^*(t)]-\mathcal{P}_h[m^*(t)\times u^*(t)]\big)\\ &+\frac{\iota^2}{2}\big(\mathcal{I}_h[(m_h(t)\times a)\times a]-\mathcal{P}_h[(m^*(t)\times a)\times a]\big)\Big\}\,dt\\ &+\iota\big(\mathcal{I}_h[m_h(t)\times a]-\mathcal{P}_h[m^*(t)\times a]\big)\,d\beta^*(t),\end{aligned}\tag{5.2}$$

and $\mathcal{P}_h\xi_h(0)=\mathcal{I}_h[m_0]-\mathcal{P}_h[m_0]$. Apply Itô's formula to the function $x\to\|x\|_{\mathbb{L}^2}^2$ to get

$\|\mathcal{P}_h\xi_h(t)\|_{\mathbb{L}^2}^2$

[f] We denote by $(\mathbb{X},w)$ the topological space $\mathbb{X}$ equipped with the weak topology.

$$
\begin{aligned}
&= \|\mathcal{P}_h\xi_h(0)\|_{\mathbb{L}^2}^2 + 2\int_0^t \Big(\mathcal{I}_h[m_h(s)\times\Delta_h m_h(s)] - \mathcal{P}_h[m^*(s)\times\Delta m^*(s)], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\,ds \\
&+ 2\int_0^t \Big(\mathcal{I}_h[m_h(s)\times\mathcal{R}_h u^*(s)] - \mathcal{P}_h[m^*(s)\times u^*(s)], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\,ds \\
&- 2\alpha\int_0^t \Big(\mathcal{I}_h[m_h(s)\times(m_h(s)\times\Delta_h m_h(s))] - \mathcal{P}_h[m^*(s)\times(m^*(s)\times\Delta m^*(s))], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\,ds \\
&+ \iota^2\int_0^t \Big(\mathcal{I}_h[(m_h(s)\times a)\times a] - \mathcal{P}_h[(m^*(s)\times a)\times a], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\,ds \\
&+ 2\iota\int_0^t \Big(\mathcal{I}_h[m_h(s)\times a] - \mathcal{P}_h[m^*(s)\times a], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\,d\beta^*(s) \\
&+ \iota^2\int_0^t \big\|\mathcal{I}_h[m_h(s)\times a] - \mathcal{P}_h[m^*(s)\times a]\big\|_{\mathbb{L}^2}^2\,ds \\
&\equiv \|\mathcal{P}_h\xi_h(0)\|_{\mathbb{L}^2}^2 + \sum_{i=1}^6 \mathbf{B}_i(t).
\end{aligned}
\tag{5.3}
$$

Observe that, thanks to (7.5), the boundedness of m_h in $\mathbb{L}^\infty$, the $\mathbb{H}^1$- stability of $\mathcal{I}_h$(cf. [7, Theorem 4.4.4]), the interpolation error estimate (7.6), and the inverse estimate (7.7)

$$
\begin{aligned}
&- 2\alpha\Big(\mathcal{I}_h[m_h(s)\times(m_h(s)\times\Delta_h m_h(s))], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2} \\
&= -2\alpha\Big\{\Big(\mathcal{I}_h[m_h(s)\times(m_h(s)\times\Delta_h m_h(s))], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2} \\
&\qquad - \Big(\mathcal{I}_h[m_h(s)\times(m_h(s)\times\Delta_h m_h(s))], \mathcal{P}_h\xi_h(s)\Big)_h\Big\} \\
&\qquad\quad - 2\alpha\Big(\mathcal{I}_h[m_h(s)\times(m_h(s)\times\Delta_h m_h(s))], \mathcal{P}_h\xi_h(s)\Big)_h \\
&\le C(\alpha)h\Big(\|(\mathcal{I}_h - \mathrm{Id})[m_h(s)\times(m_h(s)\times\Delta_h m_h(s))]\|_{\mathbb{L}^2} + \|\Delta_h m_h(s)\|_{\mathbb{L}^2}\Big)\|\nabla\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2} \\
&\qquad\quad - 2\alpha\Big(\mathcal{I}_h[m_h(s)\times(m_h(s)\times\Delta_h m_h(s))], \mathcal{P}_h\xi_h(s)\Big)_h \\
&\le C(\alpha)h\|\Delta_h m_h(s))\|_{\mathbb{L}^2}\|\nabla\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2} - 2\alpha\Big(\mathcal{I}_h[m_h(s)\times(m_h(s)\times\Delta_h m_h(s))], \mathcal{P}_h\xi_h(s)\Big)_h \\
&\equiv \overline{\mathbf{B}}_{3,1}(s) + \overline{\mathbf{B}}_{3,2}(s).
\end{aligned}
$$

By using the vector identity $b\times(c\times e) = c\langle e,b\rangle - e\langle b,c\rangle \quad \forall\, b,c,e\in\mathbb{R}^3$, the discrete sphere property and the definition of the discrete Laplacian Δ_h, we rewrite $\overline{\mathbf{B}}_{3,2}(s)$ as

$$
\begin{aligned}
\overline{\mathbf{B}}_{3,2}(s) &= -2\alpha\Big(\Delta_h m_h(s), m_h(s)\times(m_h(s)\times\mathcal{P}_h\xi_h(s))\Big)_h \\
&= -2\alpha\Big(\Delta_h m_h(s), m_h(s)\big\langle m_h(s), \mathcal{P}_h\xi_h(s)\big\rangle - \mathcal{P}_h\xi_h(s)|m_h(s)|^2\Big)_h \\
&= 2\alpha\big(\Delta_h m_h(s), \mathcal{P}_h\xi_h(s)\big)_h - 2\alpha\Big(\Delta_h m_h(s), \mathcal{I}_h[m_h(s)\big\langle m_h(s), \mathcal{P}_h\xi_h(s)\big\rangle]\Big)_h \\
&= -2\alpha\big(\nabla m_h(s), \nabla\mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2} + 2\alpha\Big(\nabla m_h(s), \nabla\mathcal{I}_h[m_h(s)\big\langle m_h(s), \mathcal{P}_h\xi_h(s)\big\rangle]\Big)_{\mathbb{L}^2} \\
&= -2\alpha\big(\nabla m_h(s), \nabla\mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2} + 2\alpha\Big(\nabla m_h(s), \nabla[m_h(s)\big\langle m_h(s), \mathcal{P}_h\xi_h(s)\big\rangle]\Big)_{\mathbb{L}^2} \\
&\qquad + 2\alpha\Big(\nabla m_h(s), \nabla(\mathcal{I}_h - \mathrm{Id})[m_h(s)\big\langle m_h(s), \mathcal{P}_h\xi_h(s)\big\rangle]\Big)_{\mathbb{L}^2} \\
&= -2\alpha\big(\nabla m_h(s), \nabla\mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2} + 2\alpha\big(|\nabla m_h(s)|^2 m_h(s), \mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2} \\
&\qquad + 2\alpha\Big(\big\langle m_h(s), \nabla m_h(s)\big\rangle, \nabla\big\langle m_h(s), \mathcal{P}_h\xi_h(s)\big\rangle\Big)_{\mathbb{L}^2} \\
&\qquad\quad + 2\alpha\Big(\nabla m_h(s), \nabla(\mathcal{I}_h - \mathrm{Id})[m_h(s)\big\langle m_h(s), \mathcal{P}_h\xi_h(s)\big\rangle]\Big)_{\mathbb{L}^2} \\
&\equiv -2\alpha\big(\nabla m_h(s), \nabla\mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2} + 2\alpha\big(|\nabla m_h(s)|^2 m_h(s), \mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2} + \mathbf{B}_{3,1}^1(s) + \mathbf{B}_{3,1}^2(s).
\end{aligned}
$$

On the other hand, thanks to (7.9), and unit length property of m^*, we infer for the other term in $\mathbf{B}_3$ that

$$\Big(\mathcal{P}_h[m^*(s)\times(m^*(s)\times\Delta m^*(s))],\mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}$$
$$=\big(\nabla m^*(s),\nabla\mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2}-\big(|\nabla m^*(s)|^2 m^*(s),\mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2}.$$

A combination of these considerations yields

$$\begin{aligned}\mathbf{B}_3(t)&\le -2\alpha\int_0^t\Big(\nabla\xi_h(s),\nabla\mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}ds+2\alpha\int_0^t\Big(|\nabla m_h(s)|^2\xi_h(s),\mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}ds\\&\quad+2\alpha\int_0^t\Big(\big(|\nabla m_h(s)|^2-|\nabla m^*(s)|^2\big)m^*(s),\mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}ds+\sum_{i=1}^2\int_0^t\mathbf{B}_{3,1}^i(s)\,ds\\&\quad+C(\alpha)h\int_0^t\|\Delta_h m_h(s)\|_{\mathbb{L}^2}\|\nabla\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\,ds\\&\equiv\sum_{i=1}^3\mathbf{B}_{3,i}(t)+\sum_{i=1}^2\int_0^t\mathbf{B}_{3,1}^i(s)\,ds+\int_0^t\overline{\mathbf{B}}_{3,1}(s)\,ds.\end{aligned}$$

In view of (7.10), and Young's inequality, we have for $\varepsilon>0$

$$\begin{aligned}\mathbf{B}_{3,1}(t)&=-2\alpha\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+2\alpha\int_0^t\big(\nabla\xi_h(s),\nabla[m^*(s)-\mathcal{P}_h m^*(s)]\big)_{\mathbb{L}^2}ds\\&\le-2\alpha\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+C(\alpha)h\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}\|\Delta m^*(s)\|_{\mathbb{L}^2}\,ds\\&\le(-2\alpha+\varepsilon)\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+C(\alpha,\varepsilon)h\int_0^t\|\Delta m^*(s)\|_{\mathbb{L}^2}^2\,ds.\end{aligned}$$

By elementary estimates, we easily obtain ($\varepsilon_1,\varepsilon_2>0$)

$$\begin{aligned}\mathbf{B}_{3,2}(t)&\le\varepsilon_1\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+C(\alpha,\varepsilon_1)\Big(1+\sup_{0\le s\le t}\|\nabla m_h(s)\|_{\mathbb{L}^2}^4\Big)\int_0^t\|\xi_h(s)\|_{\mathbb{L}^2}^2\,ds,\\\mathbf{B}_{3,3}(t)&\le\varepsilon_2\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+C\sup_{0\le s\le t}\Big(1+\|\nabla m_h(s)\|_{\mathbb{L}^2}^4+\|\nabla m^*(s)\|_{\mathbb{L}^2}^4\Big)\int_0^t\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds.\end{aligned}$$

An argumentation similar to $\mathbf{A}_{3,1}$ (cf. (7.26)), reveals that

$$\begin{aligned}\int_0^t\mathbf{B}_{3,1}^1(s)\,ds&\le C(\alpha)h\int_0^t\Big(\|\nabla\xi_h(s)\|_{\mathbb{L}^2}\|\nabla m_h(s)\|_{\mathbb{L}^4}^2+\|\nabla m_h(s)\|_{\mathbb{L}^4}^2\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^\infty}\|\nabla m_h(s)\|_{\mathbb{L}^2}\Big)ds\\&\le\varepsilon_3\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+C(\varepsilon_3,\alpha)h^2\int_0^t\|\nabla m_h(s)\|_{\mathbb{L}^4}^4\,ds\\&\qquad+C\int_0^t h\|\nabla m_h(s)\|_{\mathbb{L}^4}^2\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^\infty}\|\nabla m_h(s)\|_{\mathbb{L}^2}\,ds\\&\le\varepsilon_3\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+C(\varepsilon_3,\alpha)h\int_0^t\|\nabla m_h(s)\|_{\mathbb{L}^2}^4\,ds+\mathbf{B}_{3,1}^{1,1}(t),\end{aligned}$$

where in the last line we invoked the inverse estimate (7.7). In view of the inverse estimate (7.7), Young's inequality, the Gagliardo-Nirenberg inequality for $d=1$, and the $\mathbb{H}^1$-stability of the projection operator $\mathcal{P}_h$, we obtain for $\varepsilon_4>0$

$$\begin{aligned}\mathbf{B}_{3,1}^{1,1}(t)&\le C\int_0^t h^{\frac12}\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^\infty}\|\nabla m_h(s)\|_{\mathbb{L}^2}^3\,ds\\&\le Ch\int_0^t\|\nabla m_h(s)\|_{\mathbb{L}^2}^6\,ds+C\int_0^t\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\|\xi_h(s)\|_{\mathbb{H}^1}\,ds\\&\le Ch\int_0^t\|\nabla m_h(s)\|_{\mathbb{L}^2}^6\,ds+\varepsilon_4\int_0^t\|\xi_h(s)\|_{\mathbb{H}^1}^2\,ds+C(\varepsilon_4)\int_0^t\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds.\end{aligned}$$

Next we estimate $\mathbf{B}_{3,1}^2(s)$. Let $\psi = m_h(s)\langle m_h(s), \mathcal{P}_h\xi_h(s)\rangle$ and $\phi = [\mathcal{I}_h - \mathrm{Id}]\psi$. By the definition of $\mathcal{R}_h$ in (7.11), and of discrete Laplacian Δ_h, we may rewrite $\mathbf{B}_{3,1}^2(s)$ in the form

$$\begin{aligned}
\mathbf{B}_{3,1}^2(s) &= 2\alpha\big(\nabla m_h(s), \nabla\mathcal{R}_h\phi\big)_{\mathbb{L}^2} - 2\alpha\Big(m_h(s), (\mathrm{Id} - \mathcal{R}_h)\phi\Big)_{\mathbb{L}^2} \\
&= -2\alpha\big(\Delta_h m_h(s), \mathcal{R}_h\phi\big)_h - 2\alpha\Big(m_h(s), [\mathrm{Id} - \mathcal{R}_h]\phi\Big)_{\mathbb{L}^2} \\
&= \mathbf{B}_{3,1}^{2,1}(s) + \mathbf{B}_{3,1}^{2,2}(s).
\end{aligned}$$

We first consider $\mathbf{B}_{3,1}^{2,2}(s)$. By the estimate (7.13), the $\mathbb{H}^1$-stability of $\mathcal{R}_h$, interpolation estimate (7.6), the $\mathbb{H}^1$- stability of $\mathcal{P}_h$, and the inverse estimate (7.7), Young's inequality, and the fact that $\nabla^2\phi_h\big|_K = 0$ for all $\phi_h \in \mathbb{V}_h$, we have for $\varepsilon_5 > 0$

$$\begin{aligned}
\mathbf{B}_{3,1}^{2,2}(s) &\leq C(\alpha)h\|\nabla[\mathcal{R}_h - \mathrm{Id}]\phi\|_{\mathbb{L}^2} \leq C(\alpha)h\|[\mathcal{I}_h - \mathrm{Id}]\psi\|_{\mathbb{H}^1} \\
&\leq C(\alpha)h^2\|\nabla^2[m_h(s)\langle m_h(s), \mathcal{P}_h\xi_h(s)\rangle]\|_{\mathbb{L}^2} \\
&\leq C(\alpha)h^2\Big(\|\nabla m_h(s)\|_{\mathbb{L}^\infty}\|\nabla\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2} + \||\nabla m_h(s)|^2\|_{\mathbb{L}^2}\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^\infty}\Big) \\
&\leq C(\alpha)h^2\|\nabla m_h(s)\|_{\mathbb{L}^\infty}\|\nabla\xi_h(s)\|_{\mathbb{L}^2} + C(\alpha)h^2\Big(\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\|\nabla\xi_h(s)\|_{\mathbb{L}^2} + \|\nabla m_h(s)\|_{\mathbb{L}^4}^4\Big) \\
&\leq \varepsilon_5\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2 + C(\varepsilon_5)h\Big(\|\nabla m_h(s)\|_{\mathbb{L}^2}^2 + \|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\Big) + C(\alpha)h\|\nabla m_h(s)\|_{\mathbb{L}^2}^4.
\end{aligned}$$

We use (7.13), (7.5), (7.8), the $\mathbb{H}^1$- stabilities of $\mathcal{R}_h$ and $\mathcal{P}_h$, the interpolation error estimate (7.6), and the inverse estimate (7.7) in combination with discrete sphere property to conclude

$$\begin{aligned}
\mathbf{B}_{3,1}^{2,1}(s) &\leq C(\alpha)\|\Delta_h m_h(s)\|_h\Big(\|[\mathrm{Id} - \mathcal{R}_h]\phi\|_{\mathbb{L}^2} + \|\phi\|_{\mathbb{L}^2}\Big) \\
&\leq C(\alpha)\|\Delta_h m_h(s)\|_{\mathbb{L}^2}\Big(h\|[\mathcal{I}_h - \mathrm{Id}]\psi\|_{\mathbb{H}^1} + \|[\mathcal{I}_h - \mathrm{Id}]\psi\|_{\mathbb{L}^2}\Big) \\
&\leq C(\alpha)h^2\|\Delta_h m_h(s)\|_{\mathbb{L}^2}\|\nabla^2[m_h(s)\langle m_h(s), \mathcal{P}_h\xi_h(s)\rangle]\|_{\mathbb{L}^2} \\
&\leq C(\alpha)h^2\|\Delta_h m_h(s)\|_{\mathbb{L}^2}\Big(\|\nabla m_h(s)\|_{\mathbb{L}^\infty}\|\nabla\xi_h(s)\|_{\mathbb{L}^2} + \|\nabla m_h(s)\|_{\mathbb{L}^4}^2\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^\infty}\Big) \\
&\leq C(\alpha)h^{\frac{1}{2}}\|\nabla m_h(s)\|_{\mathbb{L}^2}^2\|\nabla\xi_h(s)\|_{\mathbb{L}^2} + C(\alpha)h\|\nabla m_h(s)\|_{\mathbb{L}^2}\|\nabla m_h(s)\|_{\mathbb{L}^4}^2\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^\infty} \\
&\leq C(\alpha)h^{\frac{1}{2}}\|\nabla m_h(s)\|_{\mathbb{L}^2}^2\|\nabla\xi_h(s)\|_{\mathbb{L}^2} + C(\alpha)h^{\frac{1}{2}}\|\nabla m_h(s)\|_{\mathbb{L}^2}^3\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^\infty} \\
&\leq C(\alpha)h^{\frac{1}{2}}\|\nabla m_h(s)\|_{\mathbb{L}^2}^2\|\nabla\xi_h(s)\|_{\mathbb{L}^2} + C\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\|\nabla\xi_h(s)\|_{\mathbb{L}^2} + C(\alpha)h\|\nabla m_h(s)\|_{\mathbb{L}^2}^6 \\
&\leq \varepsilon_6\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2 + C(\varepsilon_6, \alpha)\Big(\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2 + h\|\nabla m_h(s)\|_{\mathbb{L}^2}^4\Big) + C(\alpha)h\|\nabla m_h(s)\|_{\mathbb{L}^2}^6.
\end{aligned}$$

Combining all the above estimates, we obtain

$$\begin{aligned}
\mathbf{B}_3(t) &\leq \big(-2\alpha + \varepsilon + \sum_{i=1}^{6}\varepsilon_i\big)\int_0^t \|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds + \big(C + C(\varepsilon_4, \varepsilon_5, \varepsilon_6, \alpha)\big)\int_0^t \|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds \\
&\quad + C\sup_{0\leq s\leq t}\Big(\|\nabla m_h(s)\|_{\mathbb{L}^2}^4 + \|\nabla m^*(s)\|_{\mathbb{L}^2}^4\Big)\int_0^t \|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds \\
&\quad + Ch\int_0^t \Big\{\big(1 + C(\varepsilon_3, \varepsilon_4, \varepsilon_6)\big)\|\nabla m_h(s)\|_{\mathbb{L}^2}^4 + \|\nabla m_h(s)\|_{\mathbb{L}^2}^6 + C(\varepsilon_5)\|\nabla m_h(s)\|_{\mathbb{L}^2}^2 \\
&\quad + \|\Delta m^*(s)\|_{\mathbb{L}^2}^2\Big\}\,ds + C(\varepsilon_1)h\sup_{0\leq s\leq t}\|\nabla m_h(s)\|_{\mathbb{L}^2}^4\int_0^t \|\nabla m^*(s)\|_{\mathbb{L}^2}^2\,ds + \int_0^t \overline{\mathbf{B}}_{3,1}(s)\,ds. \qquad (5.4)
\end{aligned}$$

Next we consider $\mathbf{B}_1(t)$. Notice that in view of (7.5), and the definitions of the discrete Laplacian and projection operator $\mathcal{P}_h$ (cf. (7.9)), one has

$$\begin{aligned}
&\big(\mathcal{I}_h[m_h(s) \times \Delta_h m_h(s)], \mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2} \\
&= \Big\{\big(\mathcal{I}_h[m_h(s) \times \Delta_h m_h(s)], \mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2} - \big(\mathcal{I}_h[m_h(s) \times \Delta_h m_h(s)], \mathcal{P}_h\xi_h(s)\big)_h\Big\} \\
&\qquad + \big(\mathcal{I}_h[m_h(s) \times \Delta_h m_h(s)], \mathcal{P}_h\xi_h(s)\big)_h
\end{aligned}$$

$$
\begin{aligned}
&\leq Ch\|\nabla\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\Big(\|(\mathcal{I}_h-\mathrm{Id})[m_h(s)\times\Delta_h m_h(s)]\|_{\mathbb{L}^2}+\|\Delta_h m_h(s)\|_{\mathbb{L}^2}\Big)\\
&\qquad+\big(\mathcal{I}_h[m_h(s)\times\Delta_h m_h(s)],\mathcal{P}_h\xi_h(s)\big)_h\\
&\leq\big(\nabla m_h(s),\nabla[m_h(s)\times\mathcal{P}_h\xi_h(s)]\big)_{\mathbb{L}^2}+Ch\|\Delta_h m_h(s)\|_{\mathbb{L}^2}\ \|\nabla\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\\
&=\big(\nabla m_h(s),m_h(s)\times\nabla\mathcal{P}_h\xi_h(s)]\big)_{\mathbb{L}^2}+\overline{\mathbf{B}}_{3,1}(s),
\end{aligned}
$$

$$
\text{and }\big(\mathcal{P}_h[m^*(s)\times\Delta m^*(s)],\mathcal{P}_h\xi_h(s)\big)_{\mathbb{L}^2}=\big(\nabla m^*(s),m^*(s)\times\nabla\mathcal{P}_h\xi_h(s)]\big)_{\mathbb{L}^2}.
$$

As a consequence, we obtain the bound

$$
\begin{aligned}
\mathbf{B}_1(t)&\leq 2\int_0^t\big(\nabla\xi_h(s),m_h(s)\times\nabla\mathcal{P}_h\xi_h(s)]\big)_{\mathbb{L}^2}\,ds+\int_0^t\overline{\mathbf{B}}_{3,1}(s)\,ds\\
&\qquad+2\int_0^t\big(\nabla m^*(s),\xi_h(s)\times\nabla\mathcal{P}_h\xi_h(s)]\big)_{\mathbb{L}^2}\,ds\\
&\equiv\mathbf{B}_{1,1}(t)+\int_0^t\overline{\mathbf{B}}_{3,1}(s)\,ds+\mathbf{B}_{1,2}(t).
\end{aligned}
$$

Since $\big(\nabla\xi_h(s),m_h(s)\times\nabla\xi_h(s)\big)_{\mathbb{L}^2}=0$, in view of a Gagliardo-Nirenberg inequality for $d=1$, (7.10), the $\mathbb{H}^1$- stability of $\mathcal{P}_h$ and Young's inequality, we obtain the following bounds for $\mathbf{B}_{1,i}(t)\quad(i=1,2)$

$$
\begin{aligned}
\mathbf{B}_{1,1}(t)&=2\int_0^t\big(\nabla\xi_h(s),m_h(s)\times\nabla[m^*(s)-\mathcal{P}_h m^*(s)]\big)_{\mathbb{L}^2}\,ds\\
&\leq Ch\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}\|\Delta m^*(s)\|_{\mathbb{L}^2}\,ds\leq\varepsilon_7\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+C(\varepsilon_7)h^2\int_0^t\|\Delta m^*(s)\|_{\mathbb{L}^2}^2\,ds,\\
\mathbf{B}_{1,2}(t)&\leq C\int_0^t\|\nabla m^*(s)\|_{\mathbb{L}^2}\|\xi_h(s)\|_{\mathbb{L}^\infty}\|\nabla\xi_h(s)\|_{\mathbb{L}^2}\,ds\\
&\leq\varepsilon_8\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+C(\varepsilon_8)\Big(1+\sup_{0\leq s\leq t}\|\nabla m^*(s)\|_{\mathbb{L}^2}^4\Big)\int_0^t\|\xi_h(s)\|_{\mathbb{L}^2}^2\,ds\\
&\leq\varepsilon_8\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+C(\varepsilon_8)\Big(1+\sup_{0\leq s\leq t}\|\nabla m^*(s)\|_{\mathbb{L}^2}^4\Big)\int_0^t\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds\\
&\qquad+C(\varepsilon_8)h\Big(1+\sup_{0\leq s\leq t}\|\nabla m^*(s)\|_{\mathbb{L}^2}^4\Big)\int_0^t\|m^*(s)\|_{\mathbb{H}^1}^2\,ds,
\end{aligned}
$$

where $\varepsilon_7,\varepsilon_8>0$. Thus, we have

$$
\begin{aligned}
\mathbf{B}_1(t)&\leq(\varepsilon_7+\varepsilon_8)\int_0^t\|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds+C(\varepsilon_8)\Big(1+\sup_{0\leq s\leq t}\|\nabla m^*(s)\|_{\mathbb{L}^2}^4\Big)\int_0^t\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds\\
&\quad+C(\varepsilon_8)h\Big(1+\sup_{0\leq s\leq t}\|\nabla m^*(s)\|_{\mathbb{L}^2}^4\Big)\int_0^t\|m^*(s)\|_{\mathbb{H}^1}^2\,ds+C(\varepsilon_7)h^2\int_0^t\|\Delta m^*(s)\|_{\mathbb{L}^2}^2\,ds. \quad (5.5)
\end{aligned}
$$

Now we estimate $\mathbf{B}_2(t)$. Thanks to (7.9)

$$
\begin{aligned}
&\Big(\mathcal{I}_h[m_h(s)\times\mathcal{R}_h u^*(s)]-\mathcal{P}_h[m^*(s)\times u^*(s)],\mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\\
&=\Big(\xi_h(s)\times\mathcal{R}_h u^*(s),\mathcal{P}_h\xi(s)\Big)_{\mathbb{L}^2}+\Big(m^*(s)\times(\mathcal{R}_h u^*(s)-u^*(s)),\mathcal{P}_h\xi(s)\Big)_{\mathbb{L}^2}\\
&\qquad+\Big((\mathcal{I}_h-\mathrm{Id})[m_h(s)\times\mathcal{R}_h u^*(s)],\mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\\
&\equiv\mathbf{B}_{2,1}(s)+\mathbf{B}_{2,2}(s)+\mathbf{B}_{2,3}(s).
\end{aligned}
$$

We use the interpolation error estimate (7.6), i) of Lemma 7.3, the $\mathbb{H}^1$- stability of $\mathcal{R}_h$, and the inverse estimate (7.7) to estimate

$$
\begin{aligned}
\mathbf{B}_{2,3}(s)&\leq Ch\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\|\nabla(m_h(s)\times\mathcal{R}_h u^*(s))\|_{\mathbb{L}^2}\\
&\leq Ch\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\Big(\|u^*(s)\|_{\mathbb{H}^1}+\|\nabla m_h(s)\|_{\mathbb{L}^\infty}\|\mathcal{R}_h u^*(s)\|_{\mathbb{L}^2}\Big)\\
&\leq C\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\Big(h\|u^*(s)\|_{\mathbb{H}^1}+h^{\frac{1}{2}}\|\nabla m_h(s)\|_{\mathbb{L}^2}\|u^*(s)\|_{\mathbb{H}^1}\Big)
\end{aligned}
$$

$$\le C\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2 + Ch\Big(\|u^*(s)\|_{\mathbb{H}^1}^4 + \|\nabla m_h(s)\|_{\mathbb{L}^2}^4 + \|u^*(s)\|_{\mathbb{H}^1}^2\Big).$$

Similarly,

$$\mathbf{B}_{2,1}(s) = \big(\xi_h(s)\times\mathcal{R}_h u^*(s), \mathcal{P}_h\xi_h(s) - \xi_h(s)\big)_{\mathbb{L}^2} \le Ch\Big(\|u^*(s)\|_{\mathbb{H}^1}^2 + \|m^*(s)\|_{\mathbb{H}^1}^2\Big),$$
$$\mathbf{B}_{2,2}(s) \le C\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\|(\mathrm{Id}-\mathcal{R}_h)u^*(s)\|_{\mathbb{L}^2} \le Ch^2\|u^*(s)\|_{\mathbb{H}^1}^2 + C\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2,$$

where in the last inequality, we invoked (7.13) and the $\mathbb{H}^1$- stability of $\mathcal{R}_h$. Therefore, we obtain

$$\mathbf{B}_2(t) \le Ch\int_0^t \Big(\|\nabla m_h(s)\|_{\mathbb{L}^2}^4 + \|u^*(s)\|_{\mathbb{H}^1}^4 + \|u^*(s)\|_{\mathbb{H}^1}^2 + \|m^*(s)\|_{\mathbb{H}^1}^2\Big)\,ds + C\int_0^t \|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds. \tag{5.6}$$

Since $a \in \mathbb{W}^{1,\infty}$, by the interpolation error estimate (7.6), i) of Lemma 7.3, and (7.10), we obtain

$$\begin{aligned}
&\Big(\mathcal{I}_h[(m_h(s)\times a)\times a] - \mathcal{P}_h[(m^*(s)\times a)\times a], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\\
&= \Big((\mathcal{I}_h - \mathrm{Id})[(m_h(s)\times a)\times a], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2} + \Big((\xi_h(s)\times a)\times a, \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\\
&\le C\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2 + Ch\Big(\|\nabla m_h(s)\|_{\mathbb{L}^2}^2 + \|\nabla a\|_{\mathbb{L}^2}^2 + \|m^*(s)\|_{\mathbb{H}^1}^2\Big),
\end{aligned}$$

and therefore

$$\mathbf{B}_4(t) \le C\int_0^t \|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds + Ch\int_0^t \Big(\|\nabla m_h(s)\|_{\mathbb{L}^2}^2 + \|\nabla a\|_{\mathbb{L}^2}^2 + \|m^*(s)\|_{\mathbb{H}^1}^2\Big)\,ds. \tag{5.7}$$

We continue with $\mathbf{B}_6(t)$. Thanks to the boundedness of the solutions in $\mathbb{L}^\infty$, (7.6), (7.10), and keeping in mind that $a \in \mathbb{W}^{1,\infty}$ we get

$$\begin{aligned}
\mathbf{B}_6(t) &\le \iota^2\int_0^t \Big(\big\|(\mathcal{I}_h - \mathrm{Id})[m_h(s)\times a]\big\|_{\mathbb{L}^2}^2 + \big\|(\mathrm{Id}-\mathcal{P}_h)[m^*(s)\times a]\big\|_{\mathbb{L}^2}^2 + \big\|\xi_h(s)\times a\big\|_{\mathbb{L}^2}^2\Big)\,ds\\
&\le Ch\int_0^t \Big(\|\nabla[m_h(s)\times a]\|_{\mathbb{L}^2}^2 + \|\nabla[m^*(s)\times a]\|_{\mathbb{L}^2}^2\Big)\,ds + C\int_0^t \|\xi_h(s)\|_{\mathbb{L}^2}^2\,ds\\
&\le Ch\int_0^t \Big(\|\nabla m_h(s)\|_{\mathbb{L}^2}^2 + \|m^*(s)\|_{\mathbb{H}^1}^2 + \|\nabla a\|_{\mathbb{L}^2}^2\Big)\,ds + C\int_0^t \|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds.
\end{aligned} \tag{5.8}$$

Thanks to Young's inequality, for $\varepsilon_9 > 0$

$$\int_0^t \bar{\mathbf{B}}_{3,1}(s)\,ds \le \varepsilon_9\int_0^t \|\nabla\xi_h(s)\|_{\mathbb{L}^2}^2\,ds + C(\varepsilon_9)h\int_0^t \|\Delta_h m_h(s)\|_{\mathbb{L}^2}^2\,ds. \tag{5.9}$$

Step 2: *Estimates on $\Omega_{R,t}^{*,h}$ and Gronwall's lemma.* We combine (5.4),(5.5), (5.6), (5.7), and (5.8) along with (5.9) in (5.3) and choose $\varepsilon, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6, \varepsilon_7, \varepsilon_8$ and $\varepsilon_9 > 0$ such that $2\alpha > \varepsilon + \sum_{i=1}^9 \varepsilon_i$. Then

$$\begin{aligned}
\|\mathcal{P}_h\xi_h(t)\|_{\mathbb{L}^2}^2 &\le Ch\|m_0\|_{\mathbb{H}^1}^2 + C\sup_{0\le s\le t}\Big(\|\nabla m_h(s)\|_{\mathbb{L}^2}^4 + \|\nabla m^*(s)\|_{\mathbb{L}^2}^4\Big)\int_0^t \|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds\\
&\quad + C\int_0^t \|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}^2\,ds + Ch\sup_{0\le s\le t}\Big(1 + \|\nabla m_h(s)\|_{\mathbb{L}^2}^4 + \|\nabla m^*(s)\|_{\mathbb{L}^2}^4\Big)\int_0^t \|m^*(s)\|_{\mathbb{H}^1}^2\,ds\\
&\quad + Ch\int_0^t \Big\{\|\nabla a\|_{\mathbb{L}^2}^2 + \|u^*(s)\|_{\mathbb{H}^1}^2 + \|u^*(s)\|_{\mathbb{H}^1}^4 + \|m^*(s)\|_{\mathbb{H}^1}^2 + \|\Delta m^*(s)\|_{\mathbb{L}^2}^2\\
&\qquad + \|\Delta_h m_h(s)\|_{\mathbb{L}^2}^2 + \|m_h(s)\|_{\mathbb{H}^1}^2 + \|\nabla m_h(s)\|_{\mathbb{L}^2}^4 + \|\nabla m_h(s)\|_{\mathbb{L}^2}^6\Big\}\,ds\\
&\quad + 2\iota\int_0^t \Big(\mathcal{I}_h[m_h(s)\times a] - \mathcal{P}_h[m^*(s)\times a], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\,d\beta^*(s).
\end{aligned} \tag{5.10}$$

We now restrict the estimate on $\Omega^{*,h}_{R,t}$. Note that

$$\Omega^{*,h}_{R,t} \subset \Omega^{*,h}_{R,s} \quad (0 \le s < t). \tag{5.11}$$

Thanks to (5.1), i)-iii) of Lemma 7.3, and (5.11), we obtain from (5.10)

$$\begin{aligned} E^*\Big[\mathbf{1}_{\Omega^{*,h}_{R,t}} \|\mathcal{P}_h\xi_h(t)\|^2_{\mathbb{L}^2}\Big] &\le Ch + CRh + CR\, E^*\Big[\int_0^t \mathbf{1}_{\Omega^{*,h}_{R,s}} \|\mathcal{P}_h\xi_h(s)\|^2_{\mathbb{L}^2}\, ds\Big] \\ &\quad + 2\iota\, E^*\Big[\mathbf{1}_{\Omega^{*,h}_{R,t}} \int_0^t \Big(\mathcal{I}_h[m_h(s)\times a] - \mathcal{P}_h[m^*(s)\times a], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2} d\beta^*(s)\Big]. \end{aligned}$$

By using (7.6), (7.9), (7.10), the boundedness of the solutions in $\mathbb{L}^\infty$, Itô's-isometry, Cauchy-Schwarz inequality, the $\mathbb{L}^2$-stability of $\mathcal{P}_h$, and the fact that $\big(\xi_h(s)\times a, \xi_h(s)\big)_{\mathbb{L}^2} = 0$, we get

$$\begin{aligned} &\Big|E^*\Big[\mathbf{1}_{\Omega^{*,h}_{R,t}} \int_0^t \Big(\mathcal{I}_h[m_h(s)\times a] - \mathcal{P}_h[m^*(s)\times a], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2} d\beta^*(s)\Big]\Big| \\ &\le \Big\{E^*\Big[\int_0^t \Big|\Big(\mathcal{I}_h[m_h(s)\times a] - \mathcal{P}_h[m^*(s)\times a], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2}\Big|^2 ds\Big]\Big\}^{\frac12} \\ &\le C\Big\{E^*\Big[\int_0^t \Big|\Big((\mathcal{I}_h - \mathrm{Id})[m_h(s)\times a], \mathcal{P}_h\xi_h(s)\Big)_{\mathbb{L}^2} + \Big(\xi_h(s)\times a, \mathcal{P}_h\xi_h(s) - \xi_h(s)\Big)_{\mathbb{L}^2}\Big|^2 ds\Big]\Big\}^{\frac12} \\ &\le Ch\Big\{E^*\Big[\int_0^t \Big(\|\nabla m_h(s)\|^4_{\mathbb{L}^2} + \|\nabla a\|^4_{\mathbb{L}^2} + \|m^*(s)\|^4_{\mathbb{H}^1} + \|\xi_h(s)\|^4_{\mathbb{L}^2}\Big) ds\Big]\Big\}^{\frac12} \le Ch. \end{aligned}$$

Thus,

$$E^*\Big[\mathbf{1}_{\Omega^{*,h}_{R,t}} \|\mathcal{P}_h\xi_h(t)\|^2_{\mathbb{L}^2}\Big] \le CRh + CR\, E^*\Big[\int_0^t \mathbf{1}_{\Omega^{*,h}_{R,s}} \|\mathcal{P}_h\xi_h(s)\|^2_{\mathbb{L}^2}\, ds\Big].$$

Hence, thanks to Gronwall's inequality, we obtain

$$\sup_{0\le t\le T} E^*\Big[\mathbf{1}_{\Omega^{*,h}_{R,t}} \|\mathcal{P}_h\xi_h(t)\|^2_{\mathbb{L}^2}\Big] \le ChR\exp(CRT),$$

for some constant $C > 0$, independent of h and R. This completes the proof. □

5.1. **Proof of Theorem 2.4.** With the help of Lemma 5.1, we will prove this theorem in three steps.

Step 1: Let $\bar\pi_h = \big(\Omega^*, \mathcal{F}^*, \{\mathcal{F}^*_t\}, P^*, \beta^*, m_h, \mathcal{R}_h u^*\big) \in \mathcal{U}^w_{\mathrm{ad},h}(m_0,T)$ and $\pi^* = \big(\Omega^*, \mathcal{F}^*, \{\mathcal{F}^*_t\}, P^*, \beta^*, m^*, u^*\big) \in \mathcal{U}^w_{\mathrm{ad}}(m_0,T)$ such that $J(\pi^*) = \inf_{\pi\in\mathcal{U}^w_{\mathrm{ad}}(m_0,T)} J(\pi)$. In this step, we show that $J_h(\bar\pi_h) \to J(\pi^*)$ as $h \to 0$. Let $C > 0$ be the constant in Lemma 5.1. We set $R = \frac{1}{CT}\log(h^{-\delta})$ for some constant $\delta > 0$ which will be chosen later. Then,

$$\sup_{0\le t\le T} E^*\Big[\mathbf{1}_{\Omega^{*,h}_{R,t}} \|\mathcal{P}_h\xi_h(t)\|^2_{\mathbb{L}^2}\Big] \le \frac{h^{1-\delta}}{T}\log(h^{-\delta}). \tag{5.12}$$

For $0 \le t \le T$, let us calculate $P^*((\Omega^{*,h}_{R,t})^c)$, where $(\Omega^{*,h}_{R,t})^c$ denotes the complement of $\Omega^{*,h}_{R,t}$ in Ω^* . In view of Chebyshev's inequality, and the special choice of R,

$$\begin{aligned} P^*((\Omega^{*,h}_{R,t})^c) &\le P^*\Big(\big\{\omega\in\Omega^* : \sup_{s\in[0,t]} \|m^*(s)\|^4_{\mathbb{H}^1} > R\big\}\Big) + P^*\Big(\big\{\omega\in\Omega^* : \sup_{s\in[0,t]} \|m_h(s)\|^4_{\mathbb{H}^1} > R\big\}\Big) \\ &\le \frac{CT}{\log(h^{-\delta})} E^*\Big[\sup_{0\le s\le T} \big(\|m_h(s)\|^4_{\mathbb{H}^1} + \|m^*(s)\|^4_{\mathbb{H}^1}\big)\Big] \le \frac{\widetilde{C}}{\log(h^{-\delta})}. \end{aligned} \tag{5.13}$$

Next we compute $P^*\Big(\|\mathcal{P}_h\xi_h(t)\|_{\mathbb{L}^2} > Ch^\tau\Big)$ for some $\tau > 0$, which we will use to estimate the functional in (5.15). By using Chebyshev's inequality, estimates (5.12), and (5.13), we get

$$\begin{aligned} P^*\Big(\|\mathcal{P}_h\xi_h(t)\|_{\mathbb{L}^2} > Ch^\tau\Big) &\le P^*\Big(\{\|\mathcal{P}_h\xi_h(t)\|_{\mathbb{L}^2} > Ch^\tau\} \cap \Omega^{*,h}_{R,t}\Big) + P^*\big((\Omega^{*,h}_{R,t})^c\big) \\ &\le \frac{1}{C^2h^{2\tau}} E^*\Big[\mathbf{1}_{\Omega^{*,h}_{R,t}} \|\mathcal{P}_h\xi_h(t)\|^2_{\mathbb{L}^2}\Big] + P^*\big((\Omega^{*,h}_{R,t})^c\big) \end{aligned}$$

$$\leq \frac{h^{1-\delta-2\tau}}{TC^2}\log(h^{-\delta}) + \frac{\widetilde{C}}{\log(h^{-\delta})},$$

which yields that

$$\sup_{0\leq t\leq T} P^*\Big(\|\mathcal{P}_h\xi_h(t)\|_{\mathbb{L}^2} > Ch^\tau\Big) \leq \frac{h^{1-\delta-2\tau}}{TC^2}\log(h^{-\delta}) + \frac{\widetilde{C}}{\log(h^{-\delta})}. \tag{5.14}$$

Note that P^*-a.s., $|m_h(t,x)| \leq 1$ and $|m^*(t,x)| = 1$ for all $t \in [0,T]$ and every $x \in D$. Thus, since ψ is Lipschitz continuous on $\mathbb{L}^2$ by using the triangle inequality, and the binomial formula, we have

$$\begin{aligned}
\big|J_h(\bar{\pi}_h) - J(\pi^*)\big| &\leq E^*\Big[\int_0^T \Big|\big\|m_h(s) - \bar{m}(s)\big\|_{\mathbb{L}^2}^2 - \big\|m^*(s) - \bar{m}(s)\big\|_{\mathbb{L}^2}^2\Big|\,ds \\
&\quad + E^*\Big[\int_0^T \Big|\big\|m_h(s) - \mathcal{I}_h[\bar{m}(s)]\big\|_{\mathbb{L}^2}^2 - \big\|m_h(s) - \bar{m}(s)\big\|_{\mathbb{L}^2}^2\Big|\,ds\Big] \\
&\quad + E^*\Big[\int_0^T \Big|\big\|m_h(s) - \mathcal{I}_h[\bar{m}(s)]\big\|_h^2 - \big\|m_h(s) - \mathcal{I}_h[\bar{m}(s)]\big\|_{\mathbb{L}^2}^2\Big|\,ds\Big] \\
&\quad + E^*\Big[\int_0^T \Big|\|\mathcal{R}_h u^*(s)\|_{\mathbb{H}^1}^{2q} - \|u^*(s)\|_{\mathbb{H}^1}^{2q}\Big|\,ds\Big] + \big|\psi(m_h(T)) - \psi(m^*(T))\big|\Big] \\
&\equiv \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3 + \mathcal{J}_4 + \mathcal{J}_5.
\end{aligned} \tag{5.15}$$

Let us first consider the term $\mathcal{J}_1$. Thanks to the estimate (5.14), the $\mathbb{L}^2$-stability of the projection operator $\mathcal{P}_h$, and the boundedness of m^* and m_h in $\mathbb{L}^\infty$, we have

$$\begin{aligned}
\mathcal{J}_1 &\leq CE^*\Big[\int_0^T \|m_h(s) - m^*(s)\|_{\mathbb{L}^2}\,ds\Big] \leq CE^*\Big[\int_0^T \|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\,ds\Big] + Ch \\
&\leq C\int_0^T \Big\{\int_{\{\omega:\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2} > Ch^\tau\}} + \int_{\{\omega:\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2} \leq Ch^\tau\}}\Big\}\|\mathcal{P}_h\xi_h(s)\|_{\mathbb{L}^2}\,dP^*(\omega)\,ds + Ch \\
&\leq \frac{h^{1-\delta-2\tau}}{TC}\log(h^{-\delta}) + \frac{\widetilde{C}}{\log(h^{-\delta})} + Ch^\tau + Ch.
\end{aligned}$$

We proceed similarly with $\mathcal{J}_5$ and find the same bound. Since $\bar{m} \in H^1(D_T;\mathbb{S}^2)$, by using the interpolation error estimate (7.6), and the fact that $\|\mathcal{I}_h\phi\|_{\mathbb{L}^\infty} \leq \|\phi\|_{\mathbb{L}^\infty}$ for all $\phi \in C(\overline{D};\mathbb{R}^3)$, we find that $\mathcal{J}_2 \leq Ch$. Again in view of (7.5) and (7.6), one can find the same bound for $\mathcal{J}_3$. Next we estimate $\mathcal{J}_4$. Note that for a.s. $\omega \in \Omega^*$, $\mathcal{R}_h u^*(s)$ converges to $u^*(s)$ in $\mathbb{H}^1$ for all $s \in [0,T]$, and therefore $\|\mathcal{R}_h u^*(s)\|_{\mathbb{H}^1}^{2q}$ converges to $\|u^*(s)\|_{\mathbb{H}^1}^{2q}$. Thanks to the $\mathbb{H}^1$- stability of $\mathcal{R}_h$ and (5.1), one can use the dominated convergence theorem to conclude that $\mathcal{J}_5 \to 0$ as $h \to 0$.

Let us choose $\delta, \tau > 0$ such that $2\tau + \delta < 1$. Now, we combine these results in (5.15), and choose δ and τ as in the above discussion. The result is

$$\big|J_h(\bar{\pi}_h) - J(\pi^*)\big| \leq \frac{h^{1-\delta-2\tau}}{TC}\log(h^{-\delta}) + \frac{\widetilde{C}}{\log(h^{-\delta})} + Ch^\tau + Ch + \mathcal{J}_5 \longrightarrow 0 \ \ (h \to 0).$$

i.e.,

$$J_h(\bar{\pi}_h) \to J(\pi^*) \quad (h \to 0). \tag{5.16}$$

Step 2: Fix $h > 0$. Let $\pi_h^* = \big(\Omega_h, \mathcal{F}_h, \{\mathcal{F}_t^h\}, P^h, \beta_h^*, m_h^*, u_h^*\big) \in \mathcal{U}_{\mathrm{ad},h}^w(m_0,T)$ be such that

$$J_h(\pi_h^*) = \inf_{\pi\in\mathcal{U}_{\mathrm{ad},h}^w(m_0,T)} J_h(\pi).$$

In this step, we show that $J(\pi^*) \leq J_h(\pi_h^*)$. Note that problem (7.21) has a probabilistically strong solution on a given filtered probability space for $u = 0$. Thus, there exists $R > 0$ such that

$$E^h\Big[\int_0^T \|u_h^*(t)\|_{\mathbb{H}^1}^{2q}\,dt\Big] \leq R. \tag{5.17}$$

Moreover, m_h^* satisfies the bounds from Lemma 7.3. Furthermore, P^h- a.s., and for all $t \in [0,T]$, we have

$$\begin{aligned} m_h^*(t) &= \mathcal{I}_h[m_0] + \int_0^t \mathcal{I}_h[m_h^*(s) \times \Delta_h m_h^*(s)]\,ds + \int_0^t \mathcal{I}_h[m_h^*(s) \times \int_{\mathbb{L}^2} v\delta_{u_h^*(s,\cdot)}(dv)]\,ds \\ &\quad - \alpha \int_0^t \mathcal{I}_h[m_h^*(s) \times \big(m_h^*(s) \times \Delta_h m_h^*(s)\big)]\,ds + \frac{\iota^2}{2} \int_0^t \mathcal{I}_h[(m_h^*(s) \times a) \times a]\,ds \\ &\quad + \iota \int_0^t \mathcal{I}_h[m_h^*(s) \times a]\,d\beta_h^*(s). \end{aligned}$$

Define the associated Young measure $\lambda_h(dv, dt) = \delta_{u_h^*(t,\cdot)}(dv)\,dt$. Then, repeating the same arguments (cf. Step 1 of the proof of Theorem 3.1) we infer that the family of laws of $\{\lambda_h\}_{h>0}$ is tight on $\mathcal{Y}(0,T;\mathbb{L}^2)$. Moreover, by proceeding similar to Steps 2-3 of the proof of Theorem 3.1, there exist a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, a sequence of random variables $\big\{\big(\tilde{m}_h^*, \tilde{\lambda}_h, \tilde{\beta}_h^*\big)\big\}_{h>0}$ and $\big(\tilde{m}^*, \tilde{\lambda}, \tilde{\beta}^*\big)$ defined on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ with values in $C([0,T];\mathbb{L}^2) \times \mathcal{Y}(0,T;\mathbb{L}^2) \times C([0,T];\mathbb{R})$, such that for all $h > 0$

$$\big(\tilde{m}_h^*, \tilde{\lambda}_h, \tilde{\beta}_h^*\big) \overset{d}{=} \big(m_h^*, \lambda_h, \beta_h^*\big), \quad \text{and} \quad \tilde{P}\text{-a.s.}, \quad \begin{cases} \tilde{m}_h^* \to \tilde{m}^* \ \text{ in } C([0,T];\mathbb{L}^2), \\ \tilde{\lambda}_h \to \tilde{\lambda} \ \text{ stably in } \mathcal{Y}(0,T;\mathbb{L}^2), \\ \tilde{\beta}_h^* \to \tilde{\beta}^* \ \text{ in } C([0,T];\mathbb{R}). \end{cases} \tag{5.18}$$

Since $\tilde{P}$-a.s., $\tilde{\lambda}_h, \tilde{\lambda} \in \mathcal{Y}(0,T;\mathbb{L}^2)$, there exist relaxed control processes $\{\tilde{q}_t^h\}_{t\in[0,T]}$ and $\{\tilde{q}_t\}_{t\in[0,T]}$ defined on the probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ such that $\tilde{P}$-a.s.,

$$\tilde{\lambda}_h(dv, dt) = \tilde{q}_t^h(dv)\,dt, \qquad \text{and} \qquad \tilde{\lambda}(dv, dt) = \tilde{q}_t(dv)\,dt.$$

We define the filtrations: for $t \in [0,T]$

$$\widetilde{\mathcal{F}}_t^h = \sigma\big\{\big(\tilde{m}_h^*(s), \tilde{q}_s^h, \tilde{\beta}_h^*(s)\big) : 0 \le s \le t\big\}, \quad \text{and} \quad \widetilde{\mathcal{F}}_t = \sigma\big\{\big(\tilde{m}^*(s), \tilde{q}_s, \tilde{\beta}^*(s)\big) : 0 \le s \le t\big\}.$$

Thanks to (5.18), we have $\tilde{\pi}_h = \big(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\widetilde{\mathcal{F}}_t^h\}, \tilde{P}, \tilde{\beta}_h^*, \tilde{m}_h^*, \tilde{u}_h^*\big) \in \mathcal{U}_{\text{ad},h}^w(m_0, T)$ and

$$J_h(\tilde{\pi}_h) = J_h(\pi_h^*), \tag{5.19}$$

where $\tilde{u}_h^*(t) = \int_{\mathbb{L}^2} v\,\tilde{q}_t^h(dv)$ satisfies (2.3). One can use (5.18), i)-iii) of Lemma 7.3, and adapt a similar argument as in Steps 5, 6 and 8 of the proof of Theorem 3.1 along with [8, Lemma 5.2] to obtain a 7-tuple $\tilde{\pi}^* = \big(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\widetilde{\mathcal{F}}_t\}, \tilde{P}, \tilde{\beta}^*, \tilde{m}^*, \tilde{u}^*\big) \in \mathcal{U}_{\text{ad}}^w(m_0, T)$, where $\tilde{u}^*$ is defined by $\tilde{u}^*(t) = \int_{\mathbb{L}^2} v\,\tilde{q}_t(dv)$ satisfying (2.3), and $\tilde{\beta}^*$ is a $\{\widetilde{\mathcal{F}}_t\}$-adapted real-valued Brownian motion.

Next we calculate $J(\tilde{\pi}_h)$. We see that

$$\begin{aligned} J(\tilde{\pi}_h) &= \tilde{E}\Big[\int_0^T \Big(\big\|\tilde{m}_h^*(t) - \bar{m}(t)\big\|_{\mathbb{L}^2}^2 + \|\tilde{u}_h^*(t)\|_{\mathbb{H}^1}^{2q}\Big)\,dt\Big] + \tilde{E}\big[\psi(\tilde{m}_h^*(T))\big] \\ &= \tilde{E}\Big[\int_0^T \Big(\big\|\tilde{m}_h^*(t) - \mathcal{I}_h[\bar{m}(t)]\big\|_h^2 + \|\tilde{u}_h^*(t)\|_{\mathbb{H}^1}^{2q}\Big)\,dt\Big] + \tilde{E}\big[\psi(\tilde{m}_h(T))\big] \\ &\qquad + \tilde{E}\Big[\int_0^T \Big(\big\|\tilde{m}_h^*(t) - \bar{m}(t)\big\|_{\mathbb{L}^2}^2 - \big\|\tilde{m}_h^*(t) - \mathcal{I}_h[\bar{m}(t)]\big\|_h^2\Big)\,dt\Big] \\ &= J_h(\tilde{\pi}_h) + \tilde{E}\Big[\int_0^T \Big(\big\|\tilde{m}_h^*(t) - \bar{m}(t)\big\|_{\mathbb{L}^2}^2 - \big\|\tilde{m}_h^*(t) - \mathcal{I}_h[\bar{m}(t)]\big\|_h^2\Big)\,dt\Big] \\ &:= J_h(\tilde{\pi}_h) + J_{\text{error}}(\tilde{\pi}_h). \end{aligned}$$

We proceed similarly (cf. $\mathcal{J}_3$ and $\mathcal{J}_4$) to have $J_{\text{error}}(\tilde{\pi}_h) \le Ch$ yielding $J(\tilde{\pi}_h) = J_h(\tilde{\pi}_h) + \mathcal{O}(h)$. We may then adapt the arguments in Step 9 of the proof of Theorem 3.1 to obtain

$$\begin{aligned} J(\pi^*) \le J(\tilde{\pi}^*) \le \liminf_h J(\tilde{\pi}_h) &= \liminf_h \Big(J_h(\tilde{\pi}_h) + \mathcal{O}(h)\Big) \\ &= \liminf_h J_h(\tilde{\pi}_h) = \liminf_h J_h(\pi_h^*) \le J_h(\pi_h^*), \end{aligned}$$

where the first inequality follows from the fact that $J(\pi^*) = \inf\limits_{\pi \in \mathcal{U}_{\text{ad}}^w(m_0,T)} J(\pi)$.

Step 3: Since $J_h(\pi_h^*) = \inf_{\pi \in \mathcal{U}^w_{\text{ad},h}(m_0,T)} J_h(\pi)$, by Steps 1 and 2 we have

$$J(\pi^*) \le J_h(\pi_h^*) \le J_h(\bar{\pi}_h) \to J(\pi^*) \text{ as } h \to 0,$$

where we have invoked (5.16) to have the last expression. In other words, $J_h(\pi_h^*) \to J(\pi^*)$ as $h \to 0$. This completes the proof.

5.2. **Proof of Corollary 2.5.** Following the proof of Theorem 2.4, we see that there exist a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$, and related weak optimal solutions $\tilde{\pi}_h^* = (\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t^h\}, \tilde{P}, \tilde{\beta}_h^*, \tilde{m}_h^*, \tilde{u}_h^*) \in \mathcal{U}^w_{\text{ad},h}(m_0,T)$ and $\tilde{\pi}^* = (\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}, \tilde{\beta}^*, \tilde{m}^*, \tilde{u}^*) \in \mathcal{U}^w_{\text{ad}}(m_0,T)$ such that $\tilde{P}$-a.s.,

$$\tilde{m}_h^* \to \tilde{m}^* \quad \text{in } C([0,T];\mathbb{L}^2); \qquad \tilde{\beta}_h^* \to \tilde{\beta}^* \quad \text{in } C([0,T];\mathbb{R}).$$

for $h \to 0$. Moreover

$$\tilde{u}_h^* \rightharpoonup \tilde{u}^* \text{ in } L^{2q}(\tilde{\Omega} \times (0,T);\mathbb{H}^1), \text{ and } J_h(\tilde{\pi}_h^*) \to J(\tilde{\pi}^*) = J(\pi^*) \qquad (h \to 0).$$

Therefore, it remains to show the strong convergence of $\tilde{u}_h^*$ to $\tilde{u}^*$. Define

$$\Theta(\tilde{u}_h^*) := \tilde{E}\Big[\int_0^T \|\tilde{u}_h^*(s)\|_{\mathbb{H}^1}^{2q}\, ds\Big], \text{ and } \Theta(\tilde{u}^*) := \tilde{E}\Big[\int_0^T \|\tilde{u}^*(s)\|_{\mathbb{H}^1}^{2q}\, ds\Big].$$

Since $J_h(\tilde{\pi}_h^*) \to J(\tilde{\pi}^*)$ $(h \to 0)$, we have $\Theta(\tilde{u}_h^*) \to \Theta(\tilde{u}^*)$ $(h \to 0)$. The weak convergence, and uniform convexity of the space $L^{2q}(\tilde{\Omega} \times (0,T);\mathbb{H}^1)$ then leads to the conclusion.

6. Discretization of Pontryagin's maximum principle

In this section, we detail steps to obtain an implementable algorithm to approximately solve equations (2.5), (2.7)–(2.8). Key tools next to the structure preserving discretization in time (Algorithm 2.7) of (2.5) and the semi-implicit time discretization for the approximation of the BSDE (2.7) are the least-squares Monte Carlo method to approximate conditional expectations in Algorithm 2.8, and the stochastic gradient method to generate a convergent sequence of control processes. We are interested in moving a initial profile along a ferromagnetic wire:

Example 6.1. Fix $T > 0$, $\alpha > 0$, and let $\delta, \iota, \kappa, \lambda_1, \lambda_2 \ge 0$. Find a tuple (m,u), which minimizes

$$J(\pi) = E\Big[\int_0^T \Big(\delta\|m - \bar{m}\|_{\mathbb{L}^2}^2 + \big(1 + \lambda_1\|u\|_{\mathbb{L}^2}^2 + \lambda_2\|\nabla u\|_{\mathbb{L}^2}^2\big)^4\Big) dt + \kappa\psi(m(T))\Big] \tag{6.1}$$

subject to (1.1) with periodic boundary conditions, and the control constraint P-a.s. $\|u(t)\|_{\mathbb{L}^2} \le K$ for a.e. $t \in [0,T]$.

6.1. **Approximation of the adjoint equation.** If compared to Section 2.3, we use 'algebraic versions' of Algorithms 2.7, and 2.8, since the finite element space $\mathbb{V}_h$ may be identified with $(\mathbb{R}^3)^L$. Since nodal values of iterates of (2.9) are in $\mathbb{S}^2$, each $M^j(\omega) \in \mathbb{V}_h$ can be represented by a vector $\overrightarrow{M}^j(\omega) \in (\mathbb{S}^2)^L$. We denote by 'Stiff' the stiffness matrix consisting of entries $(\nabla\phi_l, \nabla\phi_k)_{\mathbb{L}^2}$, while 'Mass' denotes the mass matrix with entries $(\phi_l, \phi_k)_{\mathbb{L}^2}$. The P_1-finite element projection of the deterministic target profile at time t_j is denoted by $\underline{\overrightarrow{M}}^j$.

The stochastic gradient method in Subsection 6.2 generates a sequence of approximating feedback functions $\{\mathcal{U}_R^{(v),j}\,;\, v \in \mathbb{N}\}$ which then enter the adjoint equation in Algorithm 2.8 in the next iteration. Hence, suppose for the following that the approximation of the control $\overrightarrow{U}^j$ may be written in terms of $\overrightarrow{M}^j$, i.e. $\overrightarrow{U}^j = \mathcal{U}^j(\overrightarrow{M}^j)$, where $\mathcal{U}^j : (\mathbb{S}^2)^L \to (\mathbb{R}^3)^L$ is a given deterministic function. Then $\overrightarrow{M}^{j+1}$ can be expressed by $\overrightarrow{M}^{j+1} = F^{j+1}(\overrightarrow{M}^j, \Delta_j\beta)$, where $F^{j+1} : (\mathbb{S}^2)^L \times \mathbb{R} \to (\mathbb{R}^3)^L$ is a deterministic function.

Algorithm 6.2. Let $\overrightarrow{P}^J := -\kappa D\Psi(\overrightarrow{M}^J)$. For every $j = J-1, \dots, 0$ determine

$$\text{Mass}\,\overrightarrow{Q}^j = E\Big[\overrightarrow{\Psi}^j_Q(\Delta_j\beta, \overrightarrow{P}^{j+1})\big|\mathcal{F}_{t_j}\Big], \tag{6.2}$$

and

$$\text{LS}(\overrightarrow{M}^j, \overrightarrow{U}^j) \cdot \overrightarrow{P}^j = E\Big[\overrightarrow{\Psi}^j_P(\overrightarrow{P}^{j+1}, \overrightarrow{M}^{j+1}, a)\big|\mathcal{F}_{t_j}\Big] - \iota k\,\text{B}_3(\overrightarrow{Q}^j, a), \tag{6.3}$$

where

$$\begin{aligned}\overrightarrow{\Psi}^j_Q(\Delta_j\beta,\overrightarrow{P}^{j+1}) &:= \frac{\Delta_j\beta}{k}\,\mathrm{Mass}\,\overrightarrow{P}^{j+1},\\ \overrightarrow{\Psi}^j_P(\overrightarrow{P}^{j+1},\overrightarrow{M}^{j+1},a) &:= \mathrm{Mass}\,\overrightarrow{P}^{j+1} - 2\delta k\,\mathrm{Mass}\big(\overrightarrow{M}^{j+1}-\overrightarrow{\underline{M}}^{j+1}\big) + 2\alpha k\,\mathrm{B}_1(\overrightarrow{P}^{j+1},\overrightarrow{M}^{j+1})\\ &\quad + \frac{\iota^2}{2}k\,\mathrm{B}_2(\overrightarrow{P}^{j+1},a).\end{aligned}$$

Moreover, the matrix LS is given by

$$\mathrm{LS}(\overrightarrow{M}^j,\overrightarrow{U}^j) := \mathrm{Mass} + \alpha k\,\mathrm{Stiff} - \alpha k\,\mathrm{A}_1(\overrightarrow{M}^j) - k\,\mathrm{A}_2(\overrightarrow{M}^j) + k\,\mathrm{A}_3(\overrightarrow{M}^j) - k\,\mathrm{A}_4(\overrightarrow{U}^j).$$

The objects $\mathrm{A}_1(\cdot)$ to $\mathrm{A}_4(\cdot)$ are matrices, whose (l,k)-th entry is

$$\begin{aligned}\mathrm{A}_1(M) &:= \big(|\nabla M|^2\phi_l,\phi_k\big)_{\mathbb{L}^2}, &\qquad \mathrm{A}_2(M) &:= \big(M\times\nabla\phi_l,\nabla\phi_k\big)_{\mathbb{L}^2},\\ \mathrm{A}_3(M) &:= \big(\nabla M\times\nabla\phi_l,\phi_k\big)_{\mathbb{L}^2}, & \mathrm{A}_4(U) &:= \big(U\times\phi_l,\phi_k\big)_{\mathbb{L}^2},\end{aligned}$$

and $\mathrm{B}_1(\cdot)$ to $\mathrm{B}_3(\cdot)$ are vectors, whose k-th entry consists of

$$\begin{aligned}\mathrm{B}_1(P,M) &:= \big(\langle P,M\rangle_{\mathbb{R}^3}\nabla M,\nabla\phi_k\big)_{\mathbb{L}^2}, &\qquad \mathrm{B}_2(M,a) &:= \big((M\times a)\times a,\phi_k\big)_{\mathbb{L}^2},\\ \mathrm{B}_3(Q,a) &:= \big(Q\times a,\phi_k\big)_{\mathbb{L}^2}.\end{aligned}$$

Thanks to $\overrightarrow{M}^{j+1} = F^{j+1}(\overrightarrow{M}^j,\Delta_j\beta)$, we may represent the solution of (6.2)–(6.3) by two measurable, deterministic, but unknown functions $\overrightarrow{\mathcal{P}}^j : (\mathbb{S}^2)^L \to (\mathbb{R}^3)^L$ and $\overrightarrow{\mathcal{Q}}^j : (\mathbb{S}^2)^L \to (\mathbb{R}^3)^L$, such that

$$\overrightarrow{P}^j = \overrightarrow{\mathcal{P}}^j(\overrightarrow{M}^j) \qquad \text{and} \qquad \overrightarrow{Q}^j = \overrightarrow{\mathcal{Q}}^j(\overrightarrow{M}^j)\,, \tag{6.4}$$

where

$$\overrightarrow{\mathcal{Q}}^j(\overrightarrow{X}) = E\Big[\frac{\Delta_j\beta}{k}\overrightarrow{\mathcal{P}}^{j+1}(\overrightarrow{M}^{j+1})\Big|\overrightarrow{M}^j = \overrightarrow{X}\Big], \tag{6.5}$$

$$\begin{aligned}\overrightarrow{\mathcal{P}}^j(\overrightarrow{X}) = E\Big[&\Big(\mathrm{LS}(\overrightarrow{M}^j,\overrightarrow{\mathcal{U}}^j(\overrightarrow{X}))\Big)^{-1}\cdot\Big(\mathrm{Mass}\,\overrightarrow{\mathcal{P}}^{j+1}(\overrightarrow{M}^{j+1}) - 2k\,\mathrm{Mass}\big(\overrightarrow{M}^{j+1}-\overrightarrow{\underline{M}}^{j+1}\big)\\ &+ 2\alpha k\,\mathrm{B}_1(\overrightarrow{\mathcal{P}}^{j+1}(\overrightarrow{M}^{j+1}),\overrightarrow{M}^{j+1}) + \frac{\iota^2}{2}k\,\mathrm{B}_2(\overrightarrow{\mathcal{P}}^{j+1}(\overrightarrow{M}^{j+1}),a)\Big)\Big|\overrightarrow{M}^j = \overrightarrow{X}\Big]\\ &- \iota k\Big(\mathrm{LS}(\overrightarrow{M}^j,\overrightarrow{\mathcal{U}}^j(\overrightarrow{X}))\Big)^{-1}\cdot\mathrm{B}_3(\overrightarrow{\mathcal{Q}}^j(\overrightarrow{X}),a).\end{aligned} \tag{6.6}$$

Our aim is to approximate and simulate the deterministic functions $\overrightarrow{\mathcal{Q}}^j(\cdot)$ and $\overrightarrow{\mathcal{P}}^j(\cdot)$ of the feedback representation given in (6.4). This is carried out using the partition estimation method, which is a special case of the least squares Monte-Carlo method; cf. [21]. This method approximates $\overrightarrow{\mathcal{Q}}^j(\cdot)$ and $\overrightarrow{\mathcal{P}}^j(\cdot)$ by $\overrightarrow{\mathcal{Q}}^j_R(\cdot)$ and $\overrightarrow{\mathcal{P}}^j_R(\cdot)$ using the finite dimensional space $\mathrm{span}\big\{\mathbb{1}_{C^j_r}(\cdot);\, r=1,\ldots,R\big\}$, where the regions $\{C^j_r\}_{r=1}^R$ form a partition of $(\mathbb{S}^2)^L$, i.e. $\bigcup_{r=1}^R C^j_r = \overrightarrow{M}^j[\Omega] = (\mathbb{S}^2)^L$. Let $\Theta^{\mathcal{Q}^j}$ resp. $\Theta^{\mathcal{P}^j}$ denote a component of the argument in the conditional expectation in (6.5) resp. (6.6). The r-th coefficient q^j_r of $\overrightarrow{\mathcal{Q}}^j_R(\cdot)$ may then be computed using

$$q^j_r = \frac{1}{\#\{\overrightarrow{M}^j_m \in C^j_r\}}\sum_{m=1}^M \mathbb{1}_{C^j_r}(\overrightarrow{M}^j_m)\Theta^{\mathcal{Q}^j}_m\,, \tag{6.7}$$

where $M \gg R$ many independent samples $(\overrightarrow{M}^j_m,\Theta^{\mathcal{Q}^j}_m)$ of $(\overrightarrow{M}^j,\Theta^{\mathcal{Q}^j})$ are used. The computation of the coefficients in (6.7) at a fixed time iteration point $j \in \{T/k-1,\ldots,0\}$ consists of the following steps:

(A) Simulate M many paths of the forward SPDE via Algorithm 2.7 starting at t_0 up to t_{j+1}.
(B) Evaluate in which region C^j_r each path at time t_j has been fallen into.
(C) Compute for each region C^j_r the mean value of those $\Theta^{\mathcal{Q}^j}$ resp. $\Theta^{\mathcal{P}^j}$, whose path at time t_j is contained in it.

Simulating a nonlinear system of equations in Algorithm 2.7 from t_0 up to t_{j+1} in step (A) and locating the region C_r^j which contains the specific realization at time t_j in step (B) is time consuming; in addition, all R coefficients $\{q_r^j\}_{r=1}^R$ resp. $\{p_r^j\}_{r=1}^R$ are computed at once, requiring the storage of M realizations of $(\overrightarrow{M}_m^j, \Theta_m^{\mathcal{Q}^j})$ resp. $(\overrightarrow{M}_m^j, \Theta_m^{\mathcal{P}^j})$. To weaken these limitations, we use a further approximation: For each $1 \le r \le R$, choose a representative element $\overrightarrow{M}_{\text{repr},r}^j \in C_r^j$, and then proceed for each region C_r^j as follows:

(A') Simulate M/R many paths of the forward SPDE using Algorithm 2.7 starting at t_j to t_{j+1} using the (local) start value $\overrightarrow{M}_{r,repr}^j$.

(B') Compute the mean value of $\Theta^{\mathcal{Q}^j}$ resp. $\Theta^{\mathcal{P}^j}$ for this region C_r^j.

Both steps (A')–(B') correspond only to one region C_r^j. This allows to compute different regions in parallel, and moreover, reduces the huge computational memory demands otherwise needed, since only M/R many realizations $(\overrightarrow{M}_m^j, \Theta_m^{\mathcal{Q}^j})$ resp. $(\overrightarrow{M}_m^j, \Theta_m^{\mathcal{P}^j})$ of the r-th region C_r^j have to be stored. This approach was suggested in [22], using a hypercube partition in combination with drawing the (local) start values $\overrightarrow{M}_{\text{repr},r}^j$ from a logistic distribution. In our context, a partition by hypercubes is not suitable, since the discretization of a SPDE is a high dimensional problem ($3L$ dimensions). Instead we proceed as in [14] adaptively to partition $\overrightarrow{M}^j[\Omega]$:

(1) Simulate R additional realizations $\{\overrightarrow{M}_{\text{add},r}^j\}_{r=1}^R$ of the $(\mathbb{S}^2)^L$-valued random variable $\overrightarrow{M}^j$.
(2) Define the region C_r^j by

$$C_r^j := \big\{\overrightarrow{X} \in (\mathbb{S}^2)^L;\, |\overrightarrow{X} - \overrightarrow{M}_{\text{add},r}^j|_{(\mathbb{R}^3)^L} < \inf_{r \neq s} |\overrightarrow{X} - \overrightarrow{M}_{\text{add},s}^j|_{(\mathbb{R}^3)^L}\big\}.$$

(3) Define the local basis function $\eta_r^j(\cdot) := \mathbb{1}_{C_r^j}(\cdot)$.

This strategy decomposes $(\mathbb{S}^2)^L$ according to the distribution of $\overrightarrow{M}^j$: it creates more regions in areas where $\overrightarrow{M}^j$ is more likely to take values, and may be quickly realized in actual simulations. We use the center $\overrightarrow{M}_{\text{add},r}^j$ of the region C_r^j as the (local) starting value $\overrightarrow{M}_{\text{repr},r}^j$ in the local computation instead of drawing the starting value by some distribution. This choice has the advantage that we have only to build the matrix $\text{LS}(\overrightarrow{M}_{\text{add},r}^j, \overrightarrow{U}_r^j)$ and a corresponding LR-decomposition of it at one time for each region, instead of for each realization, and thus saves computation time.

6.2. **The stochastic gradient method.** The stochastic gradient method which was introduced in [14, 15] is an iterative scheme which generates a sequence of approximate feedback control functions $\mathcal{U}_R^{(v),j}(\cdot)$ on a sequence of partitions $\{C_r^{(v),j}\}_{r=1}^R$ of $\mathbb{V}_h$, which decrease monotonically the cost functional $v \mapsto J(M^{(v),\cdot}, \mathcal{U}_R^{(v),\cdot}(M^{(v),\cdot}))$. The stochastic gradient method updates the coefficients of the feedback function $\mathcal{U}_R^{(v),j}(\cdot)$ (2.14) according to an approximation of the maximum principle (2.11), where (local) approximations of the state, the adjoint, and the control considered in each region $C_r^{(v),j}$ are involved. By a similar calculation as in [15], the (local) gradient step $\overrightarrow{G}_r^{(v-1),j}$ for region $r = 1, \ldots, R$ is given by

$$\begin{aligned}\overrightarrow{G}_r^{(v-1),j} := -8\Big(1 + \lambda_1 \overrightarrow{U}_r^{(v-1),j}\,\text{Mass}\,\overrightarrow{U}_r^{(v-1),j} + \lambda_2 \overrightarrow{U}_r^{(v-1),j}\,\text{Stiff}\,\overrightarrow{U}_r^{(v-1),j}\Big)^7 \\ \times \Big(2\lambda_1\,\text{Mass}\,\overrightarrow{U}_r^{(v-1),j} + 2\lambda_2\,\text{Stiff}\,\overrightarrow{U}_r^{(v-1),j}\Big) + \text{A}_4(\overrightarrow{M}_{\text{add},r}^{(v),j})\overrightarrow{P}_r^{(v),j}.\end{aligned}$$

Scheme 6.3 (Stochastic gradient method for the stochastic control problem)**.**

(1) Set $\overrightarrow{U}^{(0),j} \equiv \overrightarrow{U}_{init}^j$ for each $j = 0, \ldots, J$, and for each basis region indexed by $r = 1, \ldots, R$ in the first gradient iteration step.
(2) Iterate $v = 1, 2, \ldots$ until a stopping criterion is met:
    (i) **Forward SPDE:** For each $j = 0, \ldots, J-1$ simulate the $(\mathbb{S}^2)^L$-valued random variable $\overrightarrow{M}^{(v),j+1}$ by Algorithm 2.7 using $\overrightarrow{\mathcal{U}}_R^{(v-1),j}(\cdot)$ and $\overrightarrow{M}^{(v),j}$.

(ii) **Backward SPDE:** Set $\overrightarrow{\mathcal{U}}_R^{(v),J}(\overrightarrow{X}) := -2\kappa\big(\overrightarrow{X} - \overrightarrow{\underline{M}}^J\big)$. For each $j = J-1,\ldots,0$, approximate $\overrightarrow{\mathcal{Q}}_R^{(v),j}(\cdot)$ and $\overrightarrow{\mathcal{P}}_R^{(v),j}(\cdot)$ from (6.5)–(6.6) using the least squares Monte-Carlo method, as well as $\overrightarrow{\mathcal{U}}_R^{(v-1),j}(\cdot)$, $\overrightarrow{\mathcal{P}}_R^{(v),j+1}(\cdot)$, $\overrightarrow{M}^{(v),j+1}$, and $\overrightarrow{M}^{(v),j}$.
Obtain $\overrightarrow{\mathcal{P}}_R^{(v),j}(\cdot)$, resp. $\overrightarrow{\mathcal{Q}}_R^{(v),j}(\cdot)$ with coefficients $\{\overrightarrow{P}_r^{(v),j}\}_{r=1}^R$, resp. $\{\overrightarrow{Q}_r^{(v),j}\}_{r=1}^R$.

(iii) **Gradient step:** Compute the coefficients $\{\overrightarrow{U}_r^{(v),\cdot}\}_{r=1}^R$ according to

$$\overrightarrow{U}_\cdot^{(v),\cdot} = P_{\bar{U}}\big[\overrightarrow{U}_\cdot^{(v-1),\cdot} + \sigma^{(v)}\cdot\overrightarrow{G}_\cdot^{(v-1),\cdot}\big], \tag{6.8}$$

using a suitable step size $\sigma^{(v)}$.

(iv) Evaluate the cost function $J(\cdot)$ or the gradient $\overrightarrow{G}_\cdot^{(v-1),\cdot}$ to decide if a stopping criterion is met.

The projection $P_{\bar{U}}$ in (6.8) is understood as a projection in each time step j, each region r, and each position in space x_l to the ball $B_{\bar{U}}(x_l)$. For the computation of the step size $\sigma^{(v)}$ in equation (6.8) we use a modification of Armijo's rule:

- Approximate the current cost function $J^{(v-1)}$ using the coefficients $\overrightarrow{U}_\cdot^{(v-1),\cdot}$.
- Iterate $s = 0,1,2,\ldots$ until a stopping criterion is met:
  - Set $\overrightarrow{U}_\cdot^{(v),\cdot,s} = P_{\bar{U}}\big[\overrightarrow{U}_\cdot^{(v-1),\cdot} + \sigma^\star\beta^s\overrightarrow{G}_\cdot^{(v-1),\cdot}\big]$.
  - Approximate the cost function $J^{(v),s}$ using the coefficients $\overrightarrow{U}_\cdot^{(v),\cdot,s}$.
  - Stop, if $J^{(v),s} - J^{(v-1)} \leq -\underline{\sigma}\sigma^\star\beta^s kR^{-1}\sum_{j=0}^{J-1}\sum_{r=1}^R\|\overrightarrow{G}_r^{(v-1),j}\|^2_{(\mathbb{R}^3)^L}$.

6.3. **Computational studies.** In Example (6.1), we fix $T = 0.5$, $\alpha = 0.02$, and $\psi(x) := \|x - \bar{m}(T)\|^2_{\mathbb{L}^2}$. The initial value m_0 is different on the two disjoint regions, which supports a plateau (orange in Figure 1) where m_0 attains the value $(0,0,1)^T$, while the other supports a rotation on the sphere using sine and cosine functions. This plateau is moved in space at a constant speed; see Figure 2, where the deterministic target profile $\bar{m}$ is shown. In the cost functional (6.1), we use the parameters $\delta = 1.0$, $(\lambda_1,\lambda_2) = (10^{-5},10^{-8})$ and $\kappa = 0.1$.

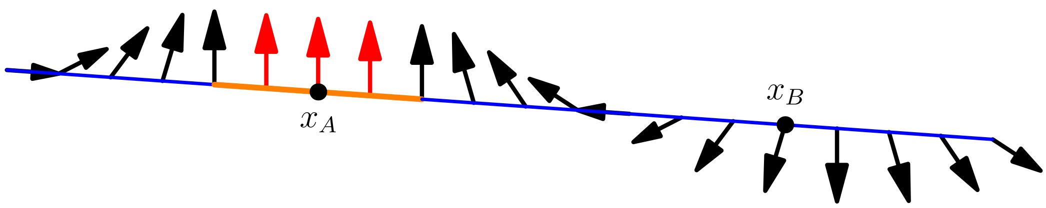


FIGURE 1. Initial value m_0: at each of the three red spins an independent three-dimensional Wiener noise is applied, which is scaled by ι. Positions x_A, resp. x_B, which are considered in Figure 4, resp. Figure 5, below.

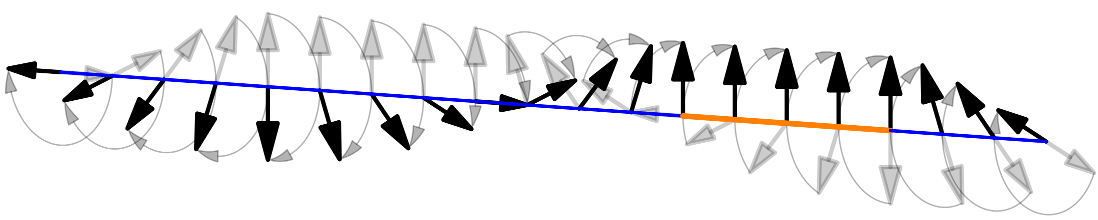
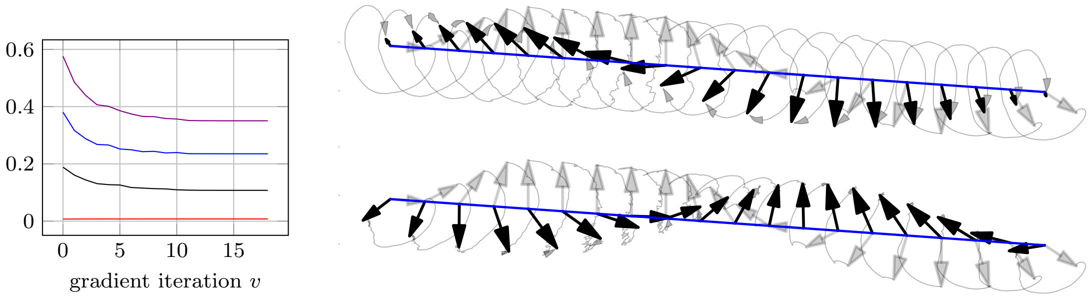


FIGURE 2. Deterministic target profile $\bar{m}$ at final time T (dark): the orange plateau region is moved at constant speed from the initial state m_0 (shaded) to the illustrated state.



(A) Cost functional (B) A single realization of $M^{(0)}$ (top) and $M^\star$ (bottom) at final time T

FIGURE 3. Iterates of the stochastic gradient method: the cost functional $J^{(v)}$ (——) and its parts $\mathbb{E}\big[\delta k \sum_{j=0}^{J} \|M^{(v),j} - \bar{M}^{(v)}(t_j)\|_{\mathbb{L}^2}^2\big]$ (——), $\mathbb{E}\big[k \sum_{j=0}^{J-1}\big(1 + \lambda_1 \|U^{(v),j}\|_{\mathbb{L}^2}^2 + \lambda_2 \|\nabla U^{(v),j}\|_{\mathbb{L}^2}^2\big)^4\big]$ (——), and $\mathbb{E}\big[\kappa \|M^{(v),J} - \bar{M}^{(v)}(T)\|_{\mathbb{L}^2}^2\big]$ (——).

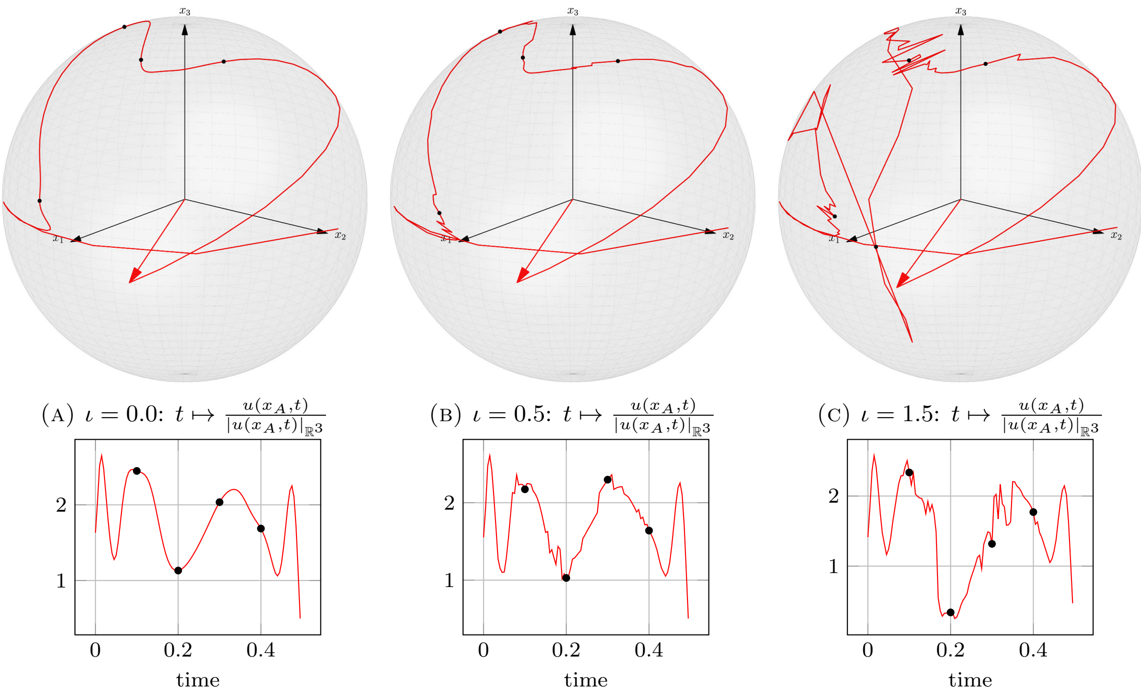
We use Scheme 6.3 ($k = 0.005$, $h = 0.05$, $M = 250000$, $R \in \{500, 1000\}$) and the partition estimation method to simulate an optimal control of Example 6.1. $\widetilde{M} = 1000$ paths are simulated to approximate the expectation value of the cost functional (6.1), which are independent of those before. The initial value of the stochastic gradient method is the optimal control of the corresponding deterministic optimal control problem, i.e., Example 6.1 with $\iota = 0$. We stop the stochastic gradient method when the difference of two successive values of the cost function is less than a given tolerance $tol = 10^{-8}$. In our simulations we use $\bar{U} = 10^3$ which was not met in all realizations. The simulations require huge storage capabilities and computation times and are carried out in parallel on a cluster.

The decay of the cost functional in the procedure of the stochastic gradient method is illustrated in Figure 3(A). In the case without any control, i.e., $u \equiv 0$, the cost functional attains the value $J = 0.9385$. By using the stochastic gradient method we are able to find a stochastic control which yields the value $J = 0.2323$ for the cost functional in the case of $R = 500$ regions. This value can be further reduced by increasing the amount of regions R. We use the deterministic optimal control as starting value for the stochastic gradient method; its value is $J = 0.4899$, see Figure 3(A) at $v = 0$. This large value of the cost functional can be explained by Figure 3(B), where the illustration at the top shows a single realization of the state $M^{(0)}$ at final time T, which was controlled by the deterministic optimal control. Here, the deterministic optimal control is not sufficient to enforce the shape of the target profile in the presence of noise; in contrast, the stochastic gradient method yields a stochastic control which forces realizations for the magnetization to approximate the target profile $\bar{m}$; see e.g. Figure 3(B) (bottom).

The evolution of one path at the certain positions $x_A = 0.3$, resp. $x_B = 0.75$, in the case of $\iota \in \{0.0, 1.0, 1.5\}$ are shown in Figure 4, resp. Figure 5. The position x_A is within the range where the Wiener process acts directly; see Figure 1. We observe in Figure 4 abrupt changes in the direction of the optimal control at position x_A to compensate for noise effects; the magnitude of the control varies slightly. A less pronounced dependence of controls on growing noise intensity ι is observed at the distant point x_B due to exchange effects in the SPDE; see Figure 5.

7. APPENDIX

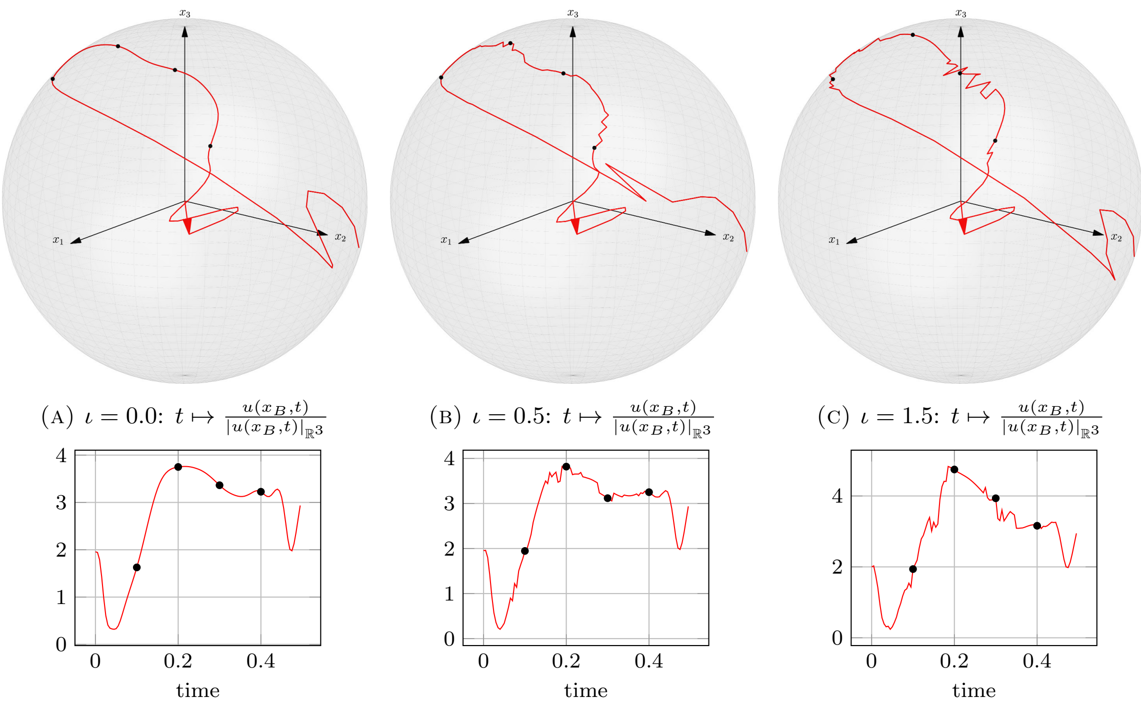
7.1. **Proof of Theorem 2.1.** Let $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P\big)$ be a given filtered probability space satisfying the usual hypotheses and β is a $\mathcal{F}_t$-adapted real-valued Wiener process. Thanks to the assumption on μ, by Skorokhod's theorem, there exist a probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$, and random variables $(\bar{u}, \bar{\beta})$ defined on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$ with values in $L^{2q}(0,T; \mathbb{W}^{l,2}) \times C([0,T]; \mathbb{R})$, $l \in \{0,1\}$ and $q \geq 2$ such



(A) $\iota = 0.0$: $t \mapsto \frac{u(x_A,t)}{|u(x_A,t)|_{\mathbb{R}^3}}$ (B) $\iota = 0.5$: $t \mapsto \frac{u(x_A,t)}{|u(x_A,t)|_{\mathbb{R}^3}}$ (C) $\iota = 1.5$: $t \mapsto \frac{u(x_A,t)}{|u(x_A,t)|_{\mathbb{R}^3}}$

(D) $\iota = 0.0$: $t \mapsto |u(x_A,t)|_{\mathbb{R}^3}$ (E) $\iota = 0.5$: $t \mapsto |u(x_A,t)|_{\mathbb{R}^3}$ (F) $\iota = 1.5$: $t \mapsto |u(x_A,t)|_{\mathbb{R}^3}$

FIGURE 4. Time evolution of the direction of the optimal control at position x_A and its magnitude in the case of different intensities of the noise.



(A) $\iota = 0.0$: $t \mapsto \frac{u(x_B,t)}{|u(x_B,t)|_{\mathbb{R}^3}}$ (B) $\iota = 0.5$: $t \mapsto \frac{u(x_B,t)}{|u(x_B,t)|_{\mathbb{R}^3}}$ (C) $\iota = 1.5$: $t \mapsto \frac{u(x_B,t)}{|u(x_B,t)|_{\mathbb{R}^3}}$

(D) $\iota = 0.0$: $t \mapsto |u(x_B,t)|_{\mathbb{R}^3}$ (E) $\iota = 0.5$: $t \mapsto |u(x_B,t)|_{\mathbb{R}^3}$ (F) $\iota = 1.5$: $t \mapsto |u(x_B,t)|_{\mathbb{R}^3}$

FIGURE 5. Time evolution of the direction of the optimal control at position x_B and its magnitude in the case of different intensities of the noise.

that $\mathcal{L}(\bar{u}) = \mu$ on $L^2(0,T;\mathbb{L}^2)$. Now define the filtration

$$\bar{\mathcal{F}}_t := \sigma\big\{\big(\bar{u}(s), \bar{\beta}(s)\big) : 0 \le s \le t\big\}$$

Then $\bar{u}$ is a $\mathbb{W}^{l,2}$-valued $\bar{\mathcal{F}}_t$-predictable stochastic process satisfying

$$\bar{P}\text{-a.s., } \|\bar{u}(t)\|_{\mathbb{L}^2} \le K \text{ for a.e. } t \in [0,T] \text{ and } \bar{E}\Big[\int_0^T \|\bar{u}(t)\|^{2q}_{\mathbb{W}^{l,2}}\,dt\Big] < +\infty, \tag{7.1}$$

and $\bar{\beta}$ is an $\bar{\mathcal{F}}_t$-adapted Wiener process. For each fixed $n \in \mathbb{N}$, define

$$T_n(\bar{u}) = \begin{cases} \bar{u}, \text{ if } |\bar{u}| < n \\ 0, \text{ otherwise} \end{cases}$$

For each $n \in \mathbb{N}$, let $\mathbb{H}_n$ be a finite-dimensional subspace of $\mathbb{L}^2$, and $P_n : \mathbb{L}^2 \to \mathbb{H}_n$ is an orthonormal projection. Following [8], we consider the Faedo-Galerkin equation in $\mathbb{H}_n$:

$$\begin{aligned} dm(t) &= \Big[P_n\big(m(t)\times\Delta m(t)\big) + P_n\big(m(t)\times T_n(\bar{u}(t))\big) - \alpha\, P_n\big(m(t)\times(m(t)\times\Delta m(t))\big) \\ &\quad + \frac{\iota^2}{2}P_n\big(P_n(m(t)\times a)\times a\big)\Big]\,dt + \iota\, P_n\big(m(t)\times a\big)\beta(t),\ t>0, \\ m(0) &= P_n(m_0). \end{aligned} \tag{7.2}$$

Note that $\bar{P}$-a.s., $\|u_n(t)\|_{\mathbb{L}^\infty} \le C(n)$, and hence for each fixed $n \in \mathbb{N}$ the drift function of the Galerkin equation (7.2) is locally Lipschitz from $\mathbb{H}_n$ into $\mathbb{H}_n$. Consequently, there exists a unique continuous, $\bar{\mathcal{F}}_t$-adapted strong solution m_n on $\big(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}, \bar{P}\big)$ under $u_n := T_n(\bar{u})$. Itô's formula for the functions $x \mapsto \|x\|^2_{\mathbb{L}^2}$, $x \mapsto \|\nabla x\|^2_{\mathbb{L}^2}$ then yields the following bounds (see *e.g.* [8, Theorem 3.5]):

$$\begin{cases} \bar{P}\text{-a.s., and for all } t \in [0,T], \quad \|m_n(t)\|_{\mathbb{L}^2} \le \|m_0\|_{\mathbb{L}^2} \\ \bar{E}\Big[\sup\limits_{0\le t\le T}\|\nabla m_n(t)\|^{2q}_{\mathbb{L}^2}\Big] + \bar{E}\Big[\Big(\int_0^T \|m_n(s)\times\Delta m_n(s)\|^2_{\mathbb{L}^2}\,ds\Big)^q\Big] \le C, \end{cases} \tag{7.3}$$

where $C > 0$ is a constant, independent of n. We remark here that m_n does not preserve the sphere condition, in contrast to i) of Lemma 7.3.

Note that $u_n \to \bar{u}$ in $L^2(\bar{\Omega}\times D_T)$ and u_n satisfies (7.1). We may then adapt the arguments in [8, Sections 4 and 5] along with (7.3) to have the existence of a 7-tuple $\tilde{\pi} := \big(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}, \tilde{\beta}, \tilde{m}, \tilde{u}\big) \in \mathcal{U}_{\mathrm{ad},w}(m_0,T)$ such that $\mathcal{L}(\tilde{u}) = \mu$ on $L^2(0,T;\mathbb{L}^2)$, and $\tilde{u}$ satisfies (7.1). Moreover, the bounds stated in ii) of Theorem 2.1 is satisfied by $(\tilde{m}, \tilde{u})$. In other words, $\tilde{\pi}$ is a weak martinagle solution to the problem (2.1). This completes the proof.

7.2. **On pathwise uniqueness of weak martingale solutions for** (2.1)**.** We show pathwise uniqueness of weak martingale solutions in some appropriate path space for $d = 1$ and a given control $\tilde{u}$. Let $\big(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}, \tilde{\beta}, \tilde{m}_1, \tilde{u}\big)$ and $\big(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}, \tilde{P}, \tilde{\beta}, \tilde{m}_2, \tilde{u}\big)$ be two weak solutions for the problem (2.1) such that paths of $\tilde{m}_i : i = 1, 2$ lie in $C([0,T];\mathbb{L}^2)\cap L^8(0,T;\mathbb{H}^1)$. Note that for $q \ge 4$ in Theorem 2.1, paths of weak solutions for the problem (2.1) lie in $C([0,T];\mathbb{L}^2)\cap L^8(0,T;\mathbb{H}^1)$. Let $m = \tilde{m}_1 - \tilde{m}_2$. Since $\tilde{P}$-a.s., $|\tilde{m}_i(t,x)| = 1$ for all $x \in D$ and $t \in [0,T]$, we see that m satisfies the following equation: $\tilde{P}$-a.s.

$$\begin{aligned} m(t) &= \alpha\int_0^t \Delta m(s)\,ds + \alpha\int_0^t |\nabla\tilde{m}_1(s)|^2_{\mathbb{R}^3} m(s)\,ds + \int_0^t m(s)\times\tilde{u}(s)\,ds + \int_0^t m(s)\times\Delta\tilde{m}_1(s)\,ds \\ &\quad + \alpha\int_0^t \big(|\nabla\tilde{m}_1(s)|_{\mathbb{R}^3} - |\nabla\tilde{m}_2(s)|_{\mathbb{R}^3}\big)\big(|\nabla\tilde{m}_1(s)|_{\mathbb{R}^3} + |\nabla\tilde{m}_2(s)|_{\mathbb{R}^3}\big)\tilde{m}_2(s)\,ds \\ &\quad + \int_0^t \tilde{m}_2(s)\times\Delta m(s)\,ds + \frac{\iota^2}{2}\int_0^t \big(m(s)\times a\big)\times a\,ds + \iota\int_0^t m(s)\times a\,d\tilde{\beta}(s). \end{aligned}$$

Application of Itô's formula for $\|m(t)\|^2_{\mathbb{L}^2}$ and the use of Gagliardo-Nirenberg, Cauchy-Schwarz and Young inequalities, the boundedness property $|\tilde{m}_i(t,x)| = 1$ for all $x \in D$ and $t \in [0,T]$, and the fact that $\langle b\times c, b\rangle = 0$ for any $b, c \in \mathbb{R}^3$ then yield

$$\|m(t)\|^4_{\mathbb{L}^2} \le C\int_0^t \|m(s)\|^4_{\mathbb{L}^2}\Big(1 + \int_0^t \|\tilde{m}_1(s)\|^8_{\mathbb{H}^1}\,ds + \int_0^t \|\tilde{m}_2(s)\|^8_{\mathbb{H}^1}\,ds\Big).$$

Now, for each $n \in \mathbb{N}$, we define the $\tilde{\mathcal{F}}_t$-stopping time:

$$\tau_n := \inf\Big\{r \in [0,T] : \int_0^r \|\tilde{m}_1(s)\|_{\mathbb{H}^1}^8\,ds + \int_0^r \|\tilde{m}_2(s)\|_{\mathbb{H}^1}^8\,ds \ge n\Big\} \wedge T.$$

Then, $\tilde{P}$-a.s., and for all $t \in [0,T]$, we have

$$\sup_{r\in[0,t]} \|m(r\wedge\tau_n)\|_{\mathbb{L}^2}^4$$
$$\le C\int_0^t \mathbf{1}_{[0,\tau_n]}(s) \sup_{r\in[0,s]} \|m(r\wedge\tau_n)\|_{\mathbb{L}^2}^4 \Big(1 + \int_0^{t\wedge\tau_n} \|\tilde{m}_1(s)\|_{\mathbb{H}^1}^8\,ds + \int_0^{t\wedge\tau_n} \|\tilde{m}_2(s)\|_{\mathbb{H}^1}^8\,ds\Big).$$

Taking expectation, and using Gronwall's lemma, we arrive at $\tilde{E}\big[\sup_{r\in[0,T]} \|m(r\wedge\tau_n)\|_{\mathbb{L}^2}^4\big] = 0$. Since $\tilde{P}$-a.s., τ_n increases to T, by monotone convergence theorem we get $\tilde{E}\big[\sup_{r\in[0,T]} \|m(r)\|_{\mathbb{L}^2}^4\big] = 0$. This implies that $\tilde{P}$-a.s., $\tilde{m}_1 = \tilde{m}_2$ on $C([0,T];\mathbb{L}^2)\cap L^8(0,T;\mathbb{H}^1)$.

7.3. **Proof of Corollary 2.2.** We use Gyöngy-Krylov's characterization of convergence in probability introduced in [23] along with pathwise uniqueness of weak martingale solutions from Subsection 7.2 to prove Corollary 2.2. The following result is the Gyöngy-Krylov characterization of convergence in probability.

Lemma 7.1. *Let $(\mathbb{G},\rho)$ be a Polish space equipped with Borel σ-algebra. A sequence of $\mathbb{G}$-valued random variables $\{X_n : n\in\mathbb{N}\}$ converges in probability to a $\mathbb{G}$-valued random element if and only if for every subsequence of joint laws $\{\mu_{p_k,n_k} : k\in\mathbb{N}\}$ of the pairs of sequences X_{p_k} and X_{n_k}, there exists a further subsequence which converges weakly to a probability measure μ such that*

$$\mu\big((x,y)\in\mathbb{G}\times\mathbb{G} : x = y\big) = 1.$$

To apply Lemma 7.1, we follow the setup as in [24]. Let $\mathbb{G} = C([0,T];\mathbb{L}^2)\times L^2(0,T;\mathbb{L}^2)$, and $\mu_{n,p}$ be the joint law of (m_n,u_n,m_p,u_p) on $\mathbb{G}\times\mathbb{G}$, and $\nu_{n,p}$ the joint law of (m_n,u_n,m_p,u_p,β).

A similar argument as in Subsection 7.1 yields tightness of the family $\big\{\nu_{n,p} : n,p\in\mathbb{N}\big\}$ on $\mathbb{G}\times\mathbb{G}\times C([0,T];\mathbb{R})$. Let us consider any subsequence $\big\{\nu_{n_k,p_k} : k\in\mathbb{N}\big\}$ of the family $\big\{\nu_{n,p} : n,p\in\mathbb{N}\big\}$. Then, by the Prokhorov theorem, it has a weakly convergent subsequence. Without loss of generality we may assume that the original sequence $\big\{\nu_{n_k,p_k} : k\in\mathbb{N}\big\}$ itself converges to a measure ν. Thus, by the Skorokhod representation theorem, there exist a probability space $(\bar{\Omega},\bar{\mathcal{F}},\bar{P})$ and a sequence of random variables $\big\{\big(\tilde{m}_{n_k},\tilde{u}_{n_k},\hat{m}_{p_k},\hat{u}_{p_k},\bar{\beta}_k\big)\big\}_{k\in\mathbb{N}}$ converging almost surely in $\mathbb{G}\times\mathbb{G}\times C([0,T];\mathbb{R})$ to a random element $\big(\tilde{m},\tilde{u},\hat{m},\hat{u},\bar{\beta}\big)$ such that

$$\bar{P}\Big(\big(\tilde{m}_{n_k},\tilde{u}_{n_k},\hat{m}_{p_k},\hat{u}_{p_k},\bar{\beta}_k\big)\in\cdot\Big) = \nu_{n_k,p_k}(\cdot), \quad \bar{P}\Big(\big(\tilde{m},\tilde{u},\hat{m},\hat{u},\bar{\beta}\big)\in\cdot\Big) = \nu(\cdot).$$

Moreover, there exists a sequence of perfect functions $\phi_k : \bar{\Omega}\to\Omega$ such that $\tilde{u}_{n_k} = u_{n_k}\circ\phi_k$ and $\hat{u}_{p_k} = u_{p_k}\circ\phi_k$. Since $u_n(t) = T_n(u)$, we have $u_n\to u$ in $L^2(\Omega\times D_T)$. As a consequence,

$$\bar{E}\Big[\int_0^T \|\tilde{u}_{n_k}(t)-\hat{u}_{p_k}(t)\|_{\mathbb{L}^2}^2\,dt\Big] = E\Big[\int_0^T \|u_{n_k}(t)-u_{p_k}(t)\|_{\mathbb{L}^2}^2\,dt\Big] \longrightarrow 0 \quad (k\to\infty)$$

which yields $\tilde{u} = \hat{u}$. Notice that μ_{n_k,p_k} converges weakly to a measure μ where μ is defined as

$$\mu(\cdot) = \bar{P}\Big(\big(\tilde{m},\tilde{u},\hat{m},\hat{u}\big)\in\cdot\Big).$$

As before, we can show that $(\tilde{m},\tilde{u},\bar{\beta})$ and $(\hat{m},\hat{u},\bar{\beta})$ are martingale solutions defined on the filtered probability space $(\bar{\Omega},\bar{\mathcal{F}},\{\bar{\mathcal{F}}_t\},\bar{P})$ where $\{\bar{\mathcal{F}}_t\}_{t\in[0,T]}$ is the filtration given by

$$\bar{\mathcal{F}}_t = \sigma\Big\{\big(\tilde{m}(s),\tilde{u}(s),\hat{m}(s),\hat{u}(s),\bar{\beta}(s)\big) : 0\le s\le t\Big\}.$$

Now, in view of *a priori* estimates, regularity of solution, and pathwise uniqueness of martingale solutions (cf. Subsection 7.2), we have

$$\mu\big((x,y)\in\mathbb{G}\times\mathbb{G} : x = y\big) = \bar{P}\big(\tilde{m} = \hat{m} \text{ in } C([0,T];\mathbb{L}^2)\big) = 1.$$

Thus, by using Lemma 7.1, we conclude that the original sequence $\{m_n\}$ defined on the given probability space $\big(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\ge0},P\big)$ converges in probability on $C([0,T];\mathbb{L}^2)$ to a random variable, say m. Without loss of generality we may assume P- almost sure convergence. Hence one can repeat the arguments as we have done for the sequence $\{\tilde{m}_n\}$ in Subsection 7.1 and conclude that (m,u) is a pathwise strong solution to the problem (2.1) on the given stochastic basis $\big(\Omega,\mathcal{F},\{\mathcal{F}_t\},P\big)$. Since $m\in\mathbb{S}^2$, we may reformulate the damping term $m\times(m\times\Delta m)$ as $-\Delta m - m|\nabla m|^2$ and argue similarly as in the proof of [9, Theorem 5.3] to conclude that $\Delta m\in L^2(\Omega\times[0,T];\mathbb{L}^2)$.

7.4. **Finite element discretization and its stability.** Let $\{\mathcal{T}_h\}_{h>0}$ be a family of quasi-uniform triangulations of $D\subset\mathbb{R}$ a bounded Lipschitz domain, cf. [7]. We denote by $\mathcal{E}_h:=\{x_l; l\in L\}$ the set of nodes of the triangulation $\mathcal{T}_h$. Consider the finite element space $\mathbb{V}_h\subset\mathbb{H}^1$,

$$\mathbb{V}_h=\big\{\phi_h\in C(\overline{D};\mathbb{R}^3):\phi_h\big|_K\in P_1(K;\mathbb{R}^3)\quad\forall K\in\mathcal{T}_h\big\},$$

where $P_1(K;\mathbb{R}^3)$ is the space of $\mathbb{R}^3$-valued functions on K which are polynomials of degree less or equal to one in each component. Define the nodal interpolation operator $\mathcal{I}_h: C(\overline{D};\mathbb{R}^3)\to\mathbb{V}_h$ as a bounded linear operator such that for every $\phi\in C(\overline{D};\mathbb{R}^3)$,

$$\mathcal{I}_h\phi(x_l)=\phi(x_l)\qquad\forall\ l\in L.$$

Let $\widetilde{\mathbb{V}}_h$ be the counterpart of $\mathbb{V}_h$ for real-valued mappings. For each $l\in L$, let $\varphi_l\in\widetilde{\mathbb{V}}_h$ be the nodal basis function *i.e.,* $\varphi_l(x_l)=1$, and $\varphi_l(x_m)=0$ for all $m\in L\setminus\{l\}$. Define the bilinear form $\big(\cdot,\cdot\big)_h: C(\overline{D};\mathbb{R}^3)\times C(\overline{D};\mathbb{R}^3)\to\mathbb{R}$ by

$$\big(\phi,\psi\big)_h=\sum_{l\in L}\zeta_l\langle\phi(x_l),\psi(x_l)\rangle\quad\forall\ \phi,\psi\in C(\overline{D};\mathbb{R}^3),$$

where $\zeta_l=\int_D\varphi_l\,dx$. Note that the induced mapping

$$\|\phi\|_h=\sqrt{\big(\phi,\phi\big)_h}\qquad\forall\ \phi\in C(\overline{D};\mathbb{R}^3)$$

is a norm on $\mathbb{V}_h$. One can show that (see [4]) for all $\phi_h,\psi_h\in\mathbb{V}_h$

$$\|\phi_h\|_{\mathbb{L}^2}\le\|\phi_h\|_h\le C\|\phi_h\|_{\mathbb{L}^2}\tag{7.4}$$

$$\Big|\big(\phi_h,\psi_h\big)_h-\big(\phi_h,\psi_h\big)_{\mathbb{L}^2}\Big|\le Ch\|\phi_h\|_{\mathbb{L}^2}\|\nabla\psi_h\|_{\mathbb{L}^2}.\tag{7.5}$$

We define the discrete Laplacian $\Delta_h:\mathbb{V}_h\to\mathbb{V}_h$ by the variational identity

$$-\big(\Delta_h\phi_h,\psi_h\big)_h=\big(\nabla\phi_h,\nabla\psi_h\big)_{\mathbb{L}^2}\qquad\forall\ \phi_h,\psi_h\in\mathbb{V}_h.$$

The following interpolation error estimates are well-known, see e.g. [7, Chapter 4]: for all $p\ge1$ and $m\in\{0,1,2\}$

$$\|\phi-\mathcal{I}_h\phi\|_{\mathbb{W}^{m,p}}\le Ch^{2-m}\big|\phi\big|_{\mathbb{W}^{2,p}}\quad\forall\phi\in\mathbb{W}^{2,p},\tag{7.6}$$

with the semi-norm $\big|\phi\big|_{\mathbb{W}^{m,p}}:=\Big(\sum_{|\gamma|=m}\|\nabla^\gamma\phi\|_{\mathbb{L}^p}^p\Big)^{\frac{1}{p}}$. We frequently use inverse estimates [7, Chapter 4]: for any $1\le r,p\le\infty$, there exists a constant $C=C(p,r)>0$ independent of h such that $\forall\phi_h\in\mathbb{V}_h$

$$\|\nabla\phi_h\|_{\mathbb{L}^r}\le Ch^{l-1+\min\{0,\frac{1}{r}-\frac{1}{p}\}}\big|\phi_h\big|_{\mathbb{W}^{l,p}}\quad(l=0,1).\tag{7.7}$$

In view of (7.7) with $l=0$, one also has the following inverse estimate

$$\|\Delta_h\phi_h\|_{\mathbb{L}^2}\le Ch^{-1}\|\nabla\phi_h\|_{\mathbb{L}^2}\quad\forall\phi_h\in\mathbb{V}_h.\tag{7.8}$$

We consider the $\mathbb{L}^2$-orthonormal projection $\mathcal{P}_h:\mathbb{L}^2\to\mathbb{V}_h$ i.e., for all $f\in\mathbb{L}^2$

$$\big(f-\mathcal{P}_hf,\phi_h\big)_{\mathbb{L}^2}=0\quad\forall\phi_h\in\mathbb{L}^2\tag{7.9}$$

with the following well-known properties:

$$\begin{cases}\|f-\mathcal{P}_hf\|_{\mathbb{L}^2}\le Ch\|f\|_{\mathbb{H}^1}\qquad\forall f\in\mathbb{H}^1\\ \|f-\mathcal{P}_hf\|_{\mathbb{L}^2}+h\|\nabla[f-\mathcal{P}_hf]\|_{\mathbb{L}^2}\le Ch^2\|\Delta f\|_{\mathbb{L}^2}\quad\forall f\in\mathbb{H}^2.\end{cases}\tag{7.10}$$

Now consider the Ritz projection $\mathcal{R}_h:\mathbb{H}^1\to\mathbb{V}_h$ via

$$(\psi-\mathcal{R}_h\psi,\varphi_h)_{\mathbb{L}^2}+(\nabla[\psi-\mathcal{R}_h\psi],\nabla\varphi_h)_{\mathbb{L}^2}=0\quad\forall\varphi_h\in\mathbb{V}_h.\tag{7.11}$$

A similar argument as in [29] reveals that there exists a constant $0 < C = C(p) < \infty$, independent of h, such that

$$\|\mathcal{R}_h\phi\|_{\mathbb{W}^{1,p}} \le C\|\phi\|_{\mathbb{W}^{1,p}} \quad \forall \phi \in \mathbb{W}^{1,p} \quad (2 \le p \le \infty). \tag{7.12}$$

Moreover, a relevant property of the Ritz projection $\mathcal{R}_h$, whose proof uses a duality argument similar to [7, Theorem 5.4.4] is as follows: for sufficiently small $h > 0$, there holds

$$\big\|(\mathrm{Id} - \mathcal{R}_h)\phi\big\|_{\mathbb{L}^2} \le Ch\big\|\nabla[\mathrm{Id} - \mathcal{R}_h]\phi\big\|_{\mathbb{L}^2} \quad \forall \phi \in \mathbb{H}^1. \tag{7.13}$$

Let us prove the following discrete Gagliardo-Nirenberg inequality.

Lemma 7.2. *Let $d = 1$. For any $2 \le p \le \infty$, there exists a constant $C = C(p) > 0$, independent of h such that*

$$\|\nabla\phi_h\|_{\mathbb{L}^p} \le C\|\phi_h\|_{\mathbb{H}^1}^{\frac{p+2}{2p}}\Big(\|\phi_h\|_{\mathbb{H}^1} + \|\Delta_h\phi_h\|_{\mathbb{L}^2}\Big)^{\frac{p-2}{2p}} \qquad \forall \phi_h \in \mathbb{V}_h. \tag{7.14}$$

Proof. To prove (7.14), we use an auxiliary problem: Fix $\phi_h \in \mathbb{V}_h$. Let $\psi \in \mathbb{H}^1$ be the unique solution of

$$\psi - \Delta\psi = \phi_h - \Delta_h\phi_h \quad \text{in } D\,, \qquad \partial_\nu\psi = 0 \quad \text{on } \partial D\,. \tag{7.15}$$

Fix $2 \le p < \infty$. We use the following decomposition, and the inverse estimate (7.7) along with the $\mathbb{W}^{1,p}$-stability of the Ritz projection (7.12) to have

$$\begin{aligned} \|\nabla\phi_h\|_{\mathbb{L}^p} &\le \|\nabla[\phi_h - \mathcal{R}_h\psi]\|_{\mathbb{L}^p} + \|\nabla\mathcal{R}_h\psi\|_{\mathbb{L}^p} \\ &\le Ch^{-\frac{p-2}{2p}}\|\nabla[\phi_h - \mathcal{R}_h\psi]\|_{\mathbb{L}^2} + C\|\psi\|_{\mathbb{W}^{1,p}} =: I + II. \end{aligned} \tag{7.16}$$

We bound the term II in (7.16) with the help of Gagliardo-Nirenberg's estimate ($d = 1$)

$$\|\nabla\psi\|_{\mathbb{L}^p} \le C_1\|\psi\|_{\mathbb{H}^1}^{\frac{p+2}{p}}\|\psi\|_{\mathbb{H}^2}^{\frac{p-2}{2p}} + C_2\|\psi\|_{\mathbb{H}^1} \tag{7.17}$$

To estimate $\|\psi\|_{\mathbb{H}^l}$ $(l = 1, 2)$, we consider (7.15) in weak form: for every $\phi \in \mathbb{H}^1$, let

$$\begin{aligned} &\Big|\big(\psi - \phi_h, \phi\big)_{\mathbb{L}^2} + \big(\nabla[\psi - \phi_h], \nabla\phi\big)_{\mathbb{L}^2}\Big| \\ &\le \Big|\big(-\Delta_h\phi_h, [\mathcal{I}_h - \mathrm{Id}]\phi\big)_{\mathbb{L}^2}\Big| + \Big|\big(\Delta_h\phi_h, \mathcal{I}_h\phi\big)_h - \big(\Delta_h\phi_h, \mathcal{I}_h\phi\big)_{\mathbb{L}^2}\Big| + \Big|\big(\nabla\phi_h, \nabla[\mathcal{I}_h - \mathrm{Id}]\phi\big)_{\mathbb{L}^2}\Big| \\ &\le III + IV + C\|\nabla\phi_h\|_{\mathbb{L}^2}\|\phi\|_{\mathbb{H}^1} \end{aligned} \tag{7.18}$$

by definition of the discrete Laplacian. Thanks to the interpolation estimate (7.6), and the inverse estimate (7.8) we see that $III \le C\|\nabla\phi_h\|_{\mathbb{L}^2}\|\phi\|_{\mathbb{H}^1}$. Again, by the $\mathbb{H}^1$- stability of the Lagrange interpolation operator $\mathcal{I}_h$ and (7.5), along with the inverse estimate (7.8) we obtain $IV \le C\|\nabla\phi_h\|_{\mathbb{L}^2}\|\phi\|_{\mathbb{H}^1}$. Taking $\phi = \psi$ in (7.18) and using Young's inequality lead to the estimate

$$\|\psi\|_{\mathbb{H}^1}^2 \le C\big(\|\phi_h\|_{\mathbb{L}^2}^2 + \|\phi_h\|_{\mathbb{H}^1}^2\big)\,. \tag{7.19}$$

Next, consider the strong form (7.15), multiply with $-\Delta\psi$, and ψ and then integrate. Due to the Neumann boundary condition, by addition, we have

$$\|\psi\|_{\mathbb{H}^2}^2 \le C\Big(\|\Delta_h\phi_h\|_{\mathbb{L}^2}^2 + \|\phi_h\|_{\mathbb{L}^2}^2\Big). \tag{7.20}$$

We combine (7.19), (7.20) in (7.17), and use Gagliardo-Nirenberg inequality to obtain

$$II \le C\|\psi\|_{\mathbb{H}^1} + C_1\|\psi\|_{\mathbb{H}^1}^{\frac{2+p}{2p}}\|\psi\|_{\mathbb{H}^2}^{\frac{p-2}{2p}} \le C\|\phi_h\|_{\mathbb{H}^1} + \|\phi_h\|_{\mathbb{H}^1}^{\frac{p+2}{2p}}\Big(\|\Delta_h\phi_h\|_{\mathbb{L}^2} + \|\phi_h\|_{\mathbb{L}^2}\Big)^{\frac{p-2}{2p}}.$$

To bound I in (7.16), we decompose the error and use a standard error estimate for $\mathcal{R}_h\psi$,

$$\|\nabla[\phi_h - \mathcal{R}_h\psi]\|_{\mathbb{L}^2} \le \|\nabla[\phi_h - \psi]\|_{\mathbb{L}^2} + \|\nabla[\psi - \mathcal{R}_h\psi]\|_{\mathbb{L}^2} \le \|\nabla[\phi_h - \psi]\|_{\mathbb{L}^2} + Ch\|\Delta\psi\|_{\mathbb{L}^2}.$$

Note that, setting $\phi = \psi - \phi_h$ in (7.18) we get $\|\psi - \phi_h\|_{\mathbb{H}^1}^2 \le Ch^2\|\Delta_h\phi_h\|_{\mathbb{L}^2}^2$, and thanks to (7.20), we obtain

$$\|\nabla[\phi_h - \mathcal{R}_h\psi]\|_{\mathbb{L}^2} \le Ch\big(\|\Delta_h\phi_h\|_{\mathbb{L}^2} + \|\phi_h\|_{\mathbb{L}^2}\big).$$

From (7.16), we may now conclude that

$$\|\nabla\phi_h\|_{\mathbb{L}^p}^p \le Ch^{\frac{p+2}{2}}\Big(\|\Delta_h\phi_h\|_{\mathbb{L}^2}^p + \|\phi_h\|_{\mathbb{L}^2}^p\Big) + C\|\phi_h\|_{\mathbb{H}^1}^{\frac{p+2}{2}}\Big(\|\Delta_h\phi_h\|_{\mathbb{L}^2} + \|\phi_h\|_{\mathbb{L}^2}\Big)^{\frac{p-2}{2}} + C\|\phi_h\|_{\mathbb{H}^1}^p$$

$$:= V + C\|\phi_h\|_{\mathbb{H}^1}^{\frac{p+2}{2}}\Big(\|\Delta_h\phi_h\|_{\mathbb{L}^2} + \|\phi_h\|_{\mathbb{L}^2}\Big)^{\frac{p-2}{2}} + C\|\phi_h\|_{\mathbb{H}^1}^p.$$

We use the inverse estimate (7.8) to get $V \le C\big(\|\phi_h\|_{\mathbb{H}^1}^{\frac{p+2}{2}}\|\Delta_h\phi_h\|_{\mathbb{L}^2}^{\frac{p-2}{2}} + \|\phi_h\|_{\mathbb{L}^2}^p\big)$. Finally, we obtain

$$\|\nabla\phi_h\|_{\mathbb{L}^p} \le C\|\phi_h\|_{\mathbb{H}^1}^{\frac{p+2}{2p}}\Big(\|\Delta_h\phi_h\|_{\mathbb{L}^2} + \|\phi_h\|_{\mathbb{H}^1}\Big)^{\frac{p-2}{2p}}.$$

Estimate (7.14) is also easily seen to hold for $p = \infty$. This completes the proof. □

The following SDE is a part of the optimal control problem (2.5);

$$\begin{aligned} dm_h(t) = \Big\{&\mathcal{I}_h[m_h(t)\times\Delta_h m_h(t)] + \mathcal{I}_h[m_h(t)\times u(t)] - \alpha\mathcal{I}_h[m_h(t)\times(m_h(t)\times\Delta_h m_h(t))] \\ &+ \frac{\iota^2}{2}\mathcal{I}_h[(m_h(t)\times a)\times a]\Big\}\,dt + \mathcal{I}_h[m_h\times a]\,d\beta(t), \\ m_h(0) = \mathcal{I}_h&[m_0]. \end{aligned} \tag{7.21}$$

Lemma 7.3. *Let $D\subset\mathbb{R}$ be a bounded interval, $q\ge 1$, $T>0$, and $m_0\in\mathbb{W}^{1,2}(D,\mathbb{S}^2)$. Let $\big(\Omega,\mathcal{F},\{\mathcal{F}_t\},P\big)$ be a given filtered probability space, and u a $\mathbb{H}^1$-valued $\{\mathcal{F}_t\}$-predictable stochastic process on it such that $E\Big[\int_0^T\|u(t)\|_{\mathbb{H}^1}^{2q}\,dt\Big] < +\infty$, and β is a $\{\mathcal{F}_t\}$-adapted real-valued Wiener process on $\big(\Omega,\mathcal{F},\{\mathcal{F}_t\},P\big)$. Then the SDE (7.21) has a unique strong global solution $m_h = \{m_h(t): t\ge 0\}$. Moreover, the following estimates hold:*

i) *P-a.s., and for all $t\in[0,T]$, $|m_h(t,x_l)| = 1$ for all $x_l\in\mathcal{E}_h$.*

ii) *$E\Big[\sup_{t\in[0,T]}\|\nabla m_h(s)\|_{\mathbb{L}^2}^{2q} + \Big(\int_0^T\big\|\mathcal{I}_h[m_h(s)\times\Delta_h m_h(s)]\big\|_{\mathbb{L}^2}^2\,ds\Big)^q\Big] \le C.$*

iii) *For $q\ge 3$, $E\Big[\int_0^T\big\|\Delta_h m_h(s)\big\|_{\mathbb{L}^2}^2\,ds\Big]\le C.$*

Proof. Since the drift of the SDE (7.21) is locally Lipschitz, by a continuation argument using the Lyapunov structure of the problem (7.21) we easily conclude the existence and uniqueness of a global strong solution $m_h = \{m_h(t): t\ge 0\}$.

i). Fix $1\le l\le L$. Apply Itô's formula to the functional $\big(m_h,\varphi_l m_h\big)_h$ which involves the nodal basis function $\varphi_l\in\widetilde{\mathbb{V}}_h$ and use the identity $\langle b\times c, b\rangle = 0\quad\forall\, b,c\in\mathbb{R}^3$ to have

$$\big(m_h(t),\varphi_l m_h(t)\big)_h = \big(\mathcal{I}_h[m_0],\varphi_l\mathcal{I}_h[m_0]\big)_h\quad\forall t\in[0,T].$$

Since $m_0(x_l)\in\mathbb{S}^2$, we infer that P-a.s., $|m_h(t,x_l)| = 1$ for all $x_l\in\mathcal{E}_h$ and for all $t\in[0,T]$. Since m_h is $\mathbb{V}_h$-valued, we see that P-a.s., $\|m_h(t,\cdot)\|_{\mathbb{L}^\infty}\le 1\quad\forall\, t\in[0,T]$.

ii). Apply Itô's formula to the functional $x\mapsto\|\nabla x\|_{\mathbb{L}^2}^2$ to get

$$\begin{aligned} \|\nabla m_h(t)\|_{\mathbb{L}^2}^2 &= \|\nabla\mathcal{I}_h[m_0]\|_{\mathbb{L}^2}^2 - 2\int_0^t\Big\{\big(\mathcal{I}_h[m_h(s)\times\Delta_h m_h(s)] + \mathcal{I}_h[m_h(s)\times u(s)],\Delta_h m_h(s)\big)_h\Big\}ds \\ &+ \int_0^t\Big(2\alpha\mathcal{I}_h[m_h(s)\times(m_h(s)\times\Delta_h m_h(s))] - \iota^2\mathcal{I}_h[(m_h(s)\times a)\times a],\Delta_h m_h(s)\Big)_h ds \\ &+ \iota^2\int_0^t\big\|\mathcal{I}_h[m_h(s)\times a]\big\|_{\mathbb{L}^2}^2\,ds - 2\int_0^t\big(\mathcal{I}_h[m_h(s)\times a],\Delta_h m_h(s)\big)_h\,d\beta(s). \end{aligned}$$

Note that

$$\begin{cases} \Big(\mathcal{I}_h[m_h(s)\times\Delta_h m_h(s)],\Delta_h m_h(s)\Big)_h = 0, \\ \mathbf{A} := \Big(\mathcal{I}_h[m_h(s)\times(m_h(s)\times\Delta_h m_h(s))],\Delta_h m_h(s)\Big)_h = -\big\|\mathcal{I}_h[m_h(s)\times\Delta_h m_h(s)]\big\|_h^2, \\ \mathbf{B} := \Big(\mathcal{I}_h[m_h(s)\times u(s)],\Delta_h m_h(s)\Big)_h = -\Big(\mathcal{I}_h[u(s)],\mathcal{I}_h[m_h(s)\times\Delta_h m_h(s)]\Big)_h \\ \quad\le\theta\big\|\mathcal{I}_h[m_h(s)\times\Delta_h m_h(s)]\big\|_h^2 + C(\theta)\big\|\mathcal{I}_h[u(s)]\big\|_h^2,\quad\text{for }\theta>0. \end{cases} \tag{7.22}$$

Since $a\in\mathbb{W}^{1,\infty}$, by using the boundedness of m_h in $\mathbb{L}^\infty$, and the $\mathbb{H}^1$-stability of $\mathcal{I}_h$, we have

$$\Big(\mathcal{I}_h[(m_h(s)\times a)\times a],\Delta_h m_h(s)\Big)_h = \big(\nabla\mathcal{I}_h[(m_h(s)\times a)\times a],\nabla m_h(s)\big)_{\mathbb{L}^2}$$

$$
\begin{aligned}
&\le C\big(\|\nabla m_h(s)\|_{\mathbb{L}^2} + \|\nabla a\|_{\mathbb{L}^2}\big)\|\nabla m_h(s)\|_{\mathbb{L}^2}\\
&\le C\big(\|\nabla m_h(s)\|_{\mathbb{L}^2}^2 + \|\nabla a\|_{\mathbb{L}^2}^2\big).
\end{aligned}
$$

Choosing $\theta > 0$ such that $2\alpha > \theta$, and (7.4) along with the $\mathbb{H}^1$- stability of $\mathcal{I}_h$, we obtain

$$
\begin{aligned}
&\sup_{r\in[0,t]} \|\nabla m_h(s)\|_{\mathbb{L}^2}^2 + (2\alpha-\theta)\int_0^t \|\mathcal{I}_h[m_h(s)\times\Delta_h m_h(s)]\|_{\mathbb{L}^2}^2\,ds\\
&\le \|\nabla m_0\|_{\mathbb{L}^2}^2 + 2\sup_{r\in[0,t]}\Big|\int_0^r \big(\nabla\mathcal{I}_h[m_h(s)\times a],\nabla m_h(s)\big)_{\mathbb{L}^2}\,d\beta(s)\Big|\\
&\quad + C\int_0^t \Big\{\|\nabla m_h(s)\|_{\mathbb{L}^2}^2 + \|u(s)\|_{\mathbb{H}^1}^2 + 1\Big\}\,ds. \qquad (7.23)
\end{aligned}
$$

Thus invoking BDG and Jensen inequalities, and the $\mathbb{H}^1$- stability of $\mathcal{I}_h$, we obtain for any $t\in[0,T]$

$$
\begin{aligned}
E\Big[\sup_{r\in[0,t]}\|\nabla m_h(s)\|_{\mathbb{L}^2}^{2q}\Big] \le\,& CE\Big[\int_0^T \big(1+\|u(s)\|_{\mathbb{H}^1}^{2q}\big)\,ds\Big] + C\int_0^t E\Big[\sup_{r\in[0,s]}\|\nabla m_h(r)\|_{\mathbb{L}^2}^{2q}\Big]\,ds\\
&+ CE\Big[\Big(\int_0^t \big(\nabla\mathcal{I}_h[m_h(s)\times a],\nabla m_h(s)\big)_{\mathbb{L}^2}^2\,ds\Big)^{\frac{q}{2}}\Big]\\
\le\,& CE\Big[\int_0^T \big(1+\|u(s)\|_{\mathbb{H}^1}^{2q}\big)\,ds\Big] + C\int_0^t E\Big[\sup_{r\in[0,s]}\|\nabla m_h(r)\|_{\mathbb{L}^2}^{2q}\Big]\,ds.
\end{aligned}
$$

Finally, we use Gronwall's inequality to conclude

$$
E\Big[\sup_{t\in[0,T]}\|\nabla m_h(s)\|_{\mathbb{L}^2}^{2q}\Big] \le C. \qquad (7.24)
$$

Furthermore, using (7.24) in (7.23) we obtain ii).
iii). We rewrite the term $\mathbf{A}$ in (7.22) and use i) to have

$$
\begin{aligned}
\mathbf{A} &= -\big(\mathcal{I}_h[\Delta_h m_h(s)|m_h(s)|^2],\Delta_h m_h(s)\big)_h + \big(\mathcal{I}_h[m_h(s)\langle m_h(s),\Delta_h m_h(s)\rangle],\Delta_h m_h(s)\big)_h\\
&= -\|\Delta_h m_h(s)\|_h^2 - \big(\nabla(\mathcal{I}_h-\mathrm{Id})[m_h(s)\langle m_h(s),\Delta_h m_h(s)\rangle],\nabla m_h(s)\big)_{\mathbb{L}^2}\\
&\qquad\qquad - \big(\nabla[m_h(s)\langle m_h(s),\Delta_h m_h(s)\rangle],\nabla m_h(s)\big)_{\mathbb{L}^2}\\
&\equiv -\|\Delta_h m_h(s)\|_h^2 + \mathbf{A}_2 + \mathbf{A}_3.
\end{aligned}
$$

We first consider $\mathbf{A}_3$. By using the product rule, i), and Hölder inequality, we have

$$
\begin{aligned}
\mathbf{A}_3 &\le C\|\Delta_h m_h(s)\|_{\mathbb{L}^2}\|\nabla m_h(s)\|_{\mathbb{L}^4}^2 - \Big(\nabla\langle m_h(s),\Delta_h m_h(s)\rangle,\langle m_h(s),\nabla m_h(s)\rangle\Big)_{\mathbb{L}^2}\\
&= C\|\Delta_h m_h(s)\|_{\mathbb{L}^2}\|\nabla m_h(s)\|_{\mathbb{L}^4}^2 - \sum_{K\in\mathcal{T}_h}\Big(\nabla\big\langle m_h(s),\Delta_h m_h(s)\big\rangle,\frac{1}{2}\nabla\big(|m_h(s)|_{\mathbb{R}^3}^2-1\big)\Big)_{\mathbb{L}^2(K)}\\
&\equiv C\|\Delta_h m_h(s)\|_{\mathbb{L}^2}\|\nabla m_h(s)\|_{\mathbb{L}^4}^2 + \mathbf{A}_{3,1}.
\end{aligned}
$$

We want to estimate $\mathbf{A}_{3,1}$. Note that the function $|m_h(s)|^2-1$ is continuous and zero at the nodal points x_l, $l\in L$. Thus, for every $K\in\mathcal{T}_h$, there exists a point $\xi_K\in K$ such that

$$
\nabla(|m_h(s)|_{\mathbb{R}^3}^2-1)(\xi_K) = 2\big\langle m_h(s),\nabla m_h(s)\big\rangle(\xi_K) = 0.
$$

Thus, since $m_h\in\mathbb{V}_h$, and $\nabla m_h\big|_K$ is constant, we have for any $x\in K$

$$
\begin{aligned}
\big\langle m_h(s,x),\nabla m_h(s,x)\big\rangle &= \Big\{\big\langle m_h(s,x),\nabla m_h(s,x)\big\rangle - \big\langle m_h(s,x),\nabla m_h(s,\xi_K)\big\rangle\Big\}\\
&\quad + \Big\{\big\langle m_h(s,x),\nabla m_h(s,\xi_K)\big\rangle - \big\langle m_h(s,\xi_K),\nabla m_h(s,\xi_K)\big\rangle\Big\}\\
&\le |\nabla m_h(s,x)-\nabla m_h(s,\xi_K)| + |m_h(s,x)-m_h(s,\xi_K)||\nabla m_h(s,\xi_K)|\\
&\le h|\nabla m_h(s,\zeta_K)||\nabla m_h(s,\xi_K)| \le h|\nabla m_h(s,x)|^2, \qquad (7.25)
\end{aligned}
$$

for some point $\zeta_K \in K$. Therefore, by using (7.25), and the product rule along with the inverse estimate (7.7), we obtain

$$\begin{aligned}
\mathbf{A}_{3,1} &\le C\Big(\sum_{K\in\mathcal{T}_h} \|\nabla m_h(s)\|_{\mathbb{L}^4(K)}^4\Big)^{\frac12}\Big(\sum_{K\in\mathcal{T}_h} h^2\|\nabla\Delta_h m_h(s)\|_{\mathbb{L}^2(K)}^2\Big)^{\frac12} \\
&\quad + C\Big(\sum_{K\in\mathcal{T}_h} h^2\|\nabla m_h(s)\|_{\mathbb{L}^6(K)}^6\Big)^{\frac12}\Big(\sum_{K\in\mathcal{T}_h} \|\Delta_h m_h(s)\|_{\mathbb{L}^2(K)}^2\Big)^{\frac12} \\
&\le C\|\nabla m_h(s)\|_{\mathbb{L}^4}^2\Big(\sum_{K\in\mathcal{T}_h} h^2\|\nabla\Delta_h m_h(s)\|_{\mathbb{L}^2(K)}^2\Big)^{\frac12} \\
&\quad + C\Big(\sum_{K\in\mathcal{T}_h} h^2\Big(h^{-\frac13}\|\nabla m_h(s)\|_{\mathbb{L}^2(K)}\Big)^6\Big)^{\frac12}\Big(\sum_{K\in\mathcal{T}_h} \|\Delta_h m_h(s)\|_{\mathbb{L}^2(K)}^2\Big)^{\frac12} \\
&\le C\|\Delta_h m_h(s)\|_{\mathbb{L}^2}\Big(\|\nabla m_h(s)\|_{\mathbb{L}^4}^2 + \|\nabla m_h(s)\|_{\mathbb{L}^2}^3\Big).
\end{aligned} \tag{7.26}$$

Next we consider $\mathbf{A}_2$. The interpolation error estimate (7.6), the inverse estimate (7.7), discrete Gagliardo-Nirenberg inequality (7.14), the boundedness of m_h in $\mathbb{L}^\infty$, and Cauchy-Schwarz inequality along with the product rule (keeping in mind that $\nabla^2\phi_h\big|_K = 0 \quad \forall \phi_h \in \mathbb{V}_h$) yields

$$\begin{aligned}
\mathbf{A}_2 &\le Ch\|\nabla m_h(s)\|_{\mathbb{L}^2}\Big(\sum_{K\in\mathcal{T}_h} \|\nabla^2[m_h(s)\langle m_h(s), \Delta_h m_h(s)\rangle]\|_{\mathbb{L}^2(K)}^2\Big)^{\frac12} \\
&\le Ch\|\nabla m_h(s)\|_{\mathbb{L}^2}\Big(\sum_{K\in\mathcal{T}_h} \|\nabla m_h(s)\|_{\mathbb{L}^\infty(K)}^4\|\Delta_h m_h(s)\|_{\mathbb{L}^2(K)}^2 \\
&\qquad\qquad + \|\nabla m_h(s)\|_{\mathbb{L}^\infty(K)}^2\|\nabla\Delta_h m_h(s)\|_{\mathbb{L}^2(K)}^2\Big)^{\frac12} \\
&\le C\|\nabla m_h(s)\|_{\mathbb{L}^2}^3\|\Delta_h m_h(s)\|_{\mathbb{L}^2} + C\|\nabla m_h(s)\|_{\mathbb{L}^2}\|\nabla m_h(s)\|_{\mathbb{L}^\infty}\|\Delta_h m_h(s)\|_{\mathbb{L}^2} \\
&\le \theta\|\Delta_h m_h(s)\|_{\mathbb{L}^2}^2 + C(\theta)\|\nabla m_h(s)\|_{\mathbb{L}^2}^2\|\nabla m_h(s)\|_{\mathbb{L}^\infty}^2 + C\|\nabla m_h(s)\|_{\mathbb{L}^2}^3\|\Delta_h m_h(s)\|_{\mathbb{L}^2} \\
&\le \theta\|\Delta_h m_h(s)\|_{\mathbb{L}^2}^2 + C(\theta)\|\nabla m_h(s)\|_{\mathbb{L}^2}^2\|m_h(s)\|_{\mathbb{H}^1}\Big(\|m_h(s)\|_{\mathbb{H}^1} + \|\Delta_h m_h(s)\|_{\mathbb{L}^2}\Big) \\
&\qquad + C\|\nabla m_h(s)\|_{\mathbb{L}^2}^3\|\Delta_h m_h(s)\|_{\mathbb{L}^2},
\end{aligned}$$

for $\theta > 0$. Again, for $\theta_0 > 0$, by (7.4) we may write the term $\mathbf{B}$ in (7.22) as

$$\mathbf{B} \le \theta_0\|\Delta_h m_h(s)\|_{\mathbb{L}^2}^2 + C(\theta_0)\|u(s)\|_{\mathbb{H}^1}^2.$$

Itô's formula to the functional $x \mapsto \|\nabla x\|_{\mathbb{L}^2}^2$ along with the above estimates then yields

$$\begin{aligned}
&\big(2\alpha - \theta_0 - \theta\big)E\Big[\int_0^T \|\Delta_h m_h(s)\|_{\mathbb{L}^2}^2\,ds\Big] \\
&\le C\|m_0\|_{\mathbb{H}^1}^2 + C_T + CE\Big[\int_0^T \|\nabla m_h(s)\|_{\mathbb{L}^2}^2\,ds\Big] + C(\theta_0)E\Big[\int_0^T \|u(s)\|_{\mathbb{H}^1}^2\,ds\Big] \\
&\quad + C(\theta)E\Big[\int_0^T \|m_h(s)\|_{\mathbb{H}^1}^4\,ds\Big] + CE\Big[\int_0^T \|\Delta_h m_h(s)\|_{\mathbb{L}^2}\Big(\|\nabla m_h(s)\|_{\mathbb{L}^4}^2 + \|\nabla m_h(s)\|_{\mathbb{L}^2}^3\Big)\,ds\Big] \\
&\qquad + C(\theta)E\Big[\int_0^T \|\Delta_h m_h(s)\|_{\mathbb{L}^2}\|m_h(s)\|_{\mathbb{H}^1}^2\,ds\Big] \\
&\le C_T + CE\Big[\int_0^T \|\nabla m_h(s)\|_{\mathbb{L}^2}^2\,ds\Big] + C(\theta_0)E\Big[\int_0^T \|u(s)\|_{\mathbb{H}^1}^2\,ds\Big] + C(\theta,\theta_2)E\Big[\int_0^T \|m_h(s)\|_{\mathbb{H}^1}^4\,ds\Big] \\
&\quad + \big(\theta_1 + \theta_2\big)E\Big[\int_0^t \|\Delta_h m_h(s)\|_{\mathbb{L}^2}^2\,ds\Big] + C(\theta_1)E\Big[\int_0^T \Big(\|\nabla m_h(s)\|_{\mathbb{L}^4}^4 + \|\nabla m_h(s)\|_{\mathbb{L}^2}^6\Big)\,ds\Big].
\end{aligned} \tag{7.27}$$

Thanks to the discrete Gagliardo-Nirenberg inequality (7.14) and Young's inequality, we see that

$$E\Big[\int_0^T \|\nabla m_h(s)\|_{\mathbb{L}^4}^4\,ds\Big] \le \theta_3 E\Big[\int_0^T \|\Delta_h m_h(s)\|_{\mathbb{L}^2}^2\,ds\Big] + C(\theta_3)E\Big[\int_0^T \|m_h(s)\|_{\mathbb{H}^1}^6\,ds\Big]$$

$$+ CE\Big[\int_0^T \Big(\|m_h(s)\|_{\mathbb{L}^2}^4 + \|m_h(s)\|_{\mathbb{L}^2}^2\Big)\, ds\Big], \tag{7.28}$$

for $\theta_3 > 0$. Using (7.28) in (7.27) and choosing $\theta_0, \theta, \theta_1, \theta_2$ and θ_3 such that $2\alpha - \theta - \sum_{i=0}^3 \theta_i > 0$ along with the $\mathbb{H}^1$- stability of interpolation operator $\mathcal{I}_h$, and ii), we conclude the estimate iii). □

Conflict of Interest: The authors declare that they have no conflict of interest.

References

[1] Agarwal, S; Carbou, G; Labbé, S; Prieur, C. Control of a network of magnetic ellipsoidal samples. *Math. Control Relat. F.* **1**, no. 2, pp. 129-147 (2011).

[2] Alouges, F; Beauchard, K. Magnetization switching on small ferromagnetic ellipsoidal samples. *ESAIM Control Optim. Calc. Var.* **15**, no. 3, pp. 676-711 (2009).

[3] Balder, E. J. Lectures on Young measure theory and its applications in economics. *Rend. Istit. Mat.Univ. Trieste*, 31 Suppl. 1:1-69, 2000.

[4] Baňas, L; Brzeźniak, Z; Neklyudov, M; Prohl, A. Stochastic ferromagnetism - Analysis and Numerics. *Studies in Mathematics* **58** (2014).

[5] Bertotti, G; Mayergoyz, I. D; Serpico, C. Nonlinear magnetization dynamics in nanosystems. *Elsevier series in electromagnetism, Amsterdam* (2009).

[6] Bouchard, B; Touzi, N. Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations. *Stochastic Proc. Appl.* 111 2004.

[7] Brenner, S. C; Scott, L. R. The mathematical theory of finite element methods. *Springer* (1994).

[8] Brzeźniak, Z; Goldys, B; Jegaraj, T. Weak solutions of a stochastic Landau-Lifshitz-Gilbert equation. *Appl. Math. Res. Express.* AMRX 2013, no. 1, 1-33.

[9] Brzeźniak, Z; Goldys, B; Jegaraj, T. Large deviations and transitions between equilibria for stochastic Landau-Lifshitz equation. *arXiv:1202.0370v2.*

[10] Brzeźniak, Z; Serrano, R. Optimal relaxed control of dissipative stochastic partial differential equations in Banach spaces *SIAM J. Control, Optim.* Vol. 51, No. 3, pp. 2664-2703.

[11] Castaing, C; Raynaud de Fitte, P; Valadier, M. Young Measures on Topological Spaces with Applications in Control Theory and Probability Theory. Mathematics and its Applications, 571 (2004).

[12] Dunst, T. Convergence with rates for a time-discretization of the Stochastic Landau-Lifschitz-Gilbert equation. *IMA J. Numer. Anal.* 35 (2015), no. 3, 1342-1380.

[13] Dunst, T; Klein, M; Prohl, A; Schäfer, A. Optimal control in evolutionary micromagnetism. *IMA J. Numer. Anal.* **35**, no. 3, pp. 1342-1380 (2015).

[14] Dunst, T; Prohl, A. The forward-backward stochastic heat equation: numerical analysis and simulation. *SIAM J. Sci. Comput.* 38 (2016), no. 5.

[15] Dunst, T; Prohl, A. Stochastic optimal control of finite ensembles of nanomagnets. *Preprint* (2016).

[16] Flandoli, F; Gatarek, D. Martingale and stationary solutions for stochastic Navier-Stokes equations. *Probability Theory and Related Fields.* 102 no. 3 (1995): 367-391.

[17] Fleming, W. H. Measure-valued processes in the control of partially-observable stochastic systems. *Appl. Math. Optim.* 6 (1980), pp. 271-285.

[18] Florescu, L. C.; Godet-Thobie, C. Young Measures and Compactness in Measure Spaces. De Gruyter (2012).

[19] Fuhrman, M; Hu, Y; Tessitore, G. Stochastic Maximum Principle for Optimal Control of SPDEs. *Appl. Math. Optim.* 68 (2013), no. 2, 181-217.

[20] Fuhrman, M; Orrier, C. Stochastic maximum principle for optimal control of a class of nonlinear SPDEs with dissipative drift. *SIAM J. Control Optim.* 54 (2016), no. 1, 341-371.

[21] Gobet, E; Lemor, J; Warin, X. A regression-based Monte Carlo Method to solve Backward Stochastic Differential equations. *Ann. Appl. Probab.* 15 no. 3 (2005): 2172–2202.

[22] Gobet, E; Lopez-Salas, J.G.; Turkedjiew, P.; Vazquez, C. Stratified regression Monte-Carlo scheme for semilinear PDEs and BSDEs with large scale parallelization on GPUs. ¡hal-01186000¿ (2015).

[23] Gyöngy, I; Krylov, N. Existence of strong solutions for Itô's stochastic equations *via* approximations. *Probab. Theory Relat. Fields* 105, 143-158 (1996).

[24] Hofmanová, M. Degenerate parabolic stochastic partial differential equations. *Stochastic Process. Appl.*, 123(12):4294–4336, 2013.

[25] Jakubowski, A. The almost sure Skorokhod representation for subsequences in nonmetric spaces. *Theory Probab. Appl.* Vol. 42, No. 1 (1998) 164-174.

[26] Karoui N. El and Huang S.-J. A general result of existence and uniqueness of backward stochastic differential equations. *In: Backward Stochastic Differential Equations, N. El Karoui et al ed., Pitman Research Notes in Math. Series*, No. 364., Chapman and Hall/CRC, Boca Raton, 1997, 27-36.

[27] Kružík, M; Prohl, A. Recent developments in the modeling, analysis, and numerics of ferromagnetism. *SIAM Rev.* 48 (2006), no. 3, 439-483.

[28] Nagase, N; Nisio, M. Optimal controls for stochastic partial differential equations. *SIAM. Control Optim.* 28 (1990), no. 1, 186-213.

[29] Rannacher, R; Scott, R. Some optimal error estimates for piecewise linear finite element approximations. *Math. Comp.* 38 (1982), no. 158, 437-445.
[30] Simon, J. Compact sets in the space $L^p(0,T;B)$. *Ann. Mat. Pura Appl.* (4) 146 (1987), 65-96.
[31] Yong, J; Zhou, X. Y. Stochastic controls. Hamiltonian systems and HJB equations. *Applications of Mathematics* (New York), 43. Springer-Verlag, New York, 1999.

(Thomas Dunst)
Mathematisches Institut Universität Tübingen Auf der Morgenstelle 10 D-72076 Tübingen, Germany
E-mail address: dunst@na.uni-tuebingen.de

(Ananta K. Majee)
Mathematisches Institut Universität Tübingen Auf der Morgenstelle 10 D-72076 Tübingen, Germany
E-mail address: majee@na.uni-tuebingen.de

(Andreas Prohl)
Mathematisches Institut Universität Tübingen Auf der Morgenstelle 10 D-72076 Tübingen, Germany
E-mail address: prohl@na.uni-tuebingen.de

(Guy Vallet)
CNRS / Univ Pau & Pays Adour, Laboratoire de Mathématiques et de leurs Applications de Pau IPRA, UMR5142 64000, Pau, France
E-mail address: guy.vallet@univ-pau.fr