Stochastic Control of Magnetization Dynamics

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Thomas Dunst
aus Nürtingen

Tübingen
2016

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Prüfung:	14.06.2016
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Andreas Prohl
2. Berichterstatter:	Prof. Dr. Ľubomír Baňas
3. Berichterstatter:	Prof. Dr. Michael Tretyakov

Abstract

This thesis is concerned with the approximation of various problems related to the stochastic Landau-Lifshitz-Gilbert equation (SLLG), which models the dynamics of a ferromagnetic body at elevated temperatures. The SLLG is a nonlinear stochastic partial differential equation which possesses an inherent non-convex side constraint. Firstly, the time discretization of the stochastic partial differential equation is addressed, where we study the convergence behavior of a structure-preserving discretization. Secondly, the approximation and simulation of the stochastic optimal control problem subject to the SLLG is studied by means of the necessary first order optimality conditions.

The thesis is split into three parts. In the first part we focus on the time discretization of the SLLG. We show convergence in probability with rate of order $1/2$ for a time discretized scheme which is based on the midpoint rule and preserves the sphere constraint. Main difficulties were the analytical and numerical treatment of the nonlinear and stochastic terms. Computational studies carried out in this part support this convergence rate.

The second and the third part are contributed to the stochastic optimal control problem. In the second part, we prove strong convergence with optimal rates for a spatial discretization of the forward-backward stochastic heat equation which describes the stochastic optimal control problem subject to the stochastic heat equation. As an intermediate step, we show optimal rates for a spatial discretization of the backward stochastic heat equation. A full discretization which is based on the implicit Euler method for a temporal discretization and a least squares Monte-Carlo method is then proposed. Next to an iterative solution strategy which is based on a well-known Picard-type algorithm, the new stochastic gradient method turns out to be much more flexible. Concluding computational experiments compare the efficiency of different discretization approaches.

The third part combines the methodology of the second part with the SLLG. Here, we control the dynamics of a fixed number of ferromagnetic spins at elevated temperatures by minimizing a quadratic functional subject to the SLLG. Existence of a minimum of the stochastic optimal control problem with control constraints is shown. The related first order optimality conditions consist of a coupled forward-backward SDE system, which is numerically solved by a structure-inheriting discretization, the least squares Monte-Carlo method to approximate related conditional expectations, and the stochastic gradient method. Computational experiments are reported which motivate optimal controls in the case of interacting anisotropy, stray field, exchange energies, and acting noise.

Zusammenfassung

Ziel dieser Arbeit ist es, numerische Approximationen verschiedener Problemstellungen, die sich mit der stochastischen Landau-Lifshitz-Gilbert-Gleichung (kurz SLLG) befassen, zu untersuchen. Die SLLG ist eine nichtlineare stochastische partielle Differentialgleichung mit einer nichtkonvexen Zwangsbedingung (Sphärenbedingung) und wird bei der Modellierung ferromagentischer Dynamiken eingesetzt.

Die Dissertation ist in drei wesentliche Teile untergliedert. Der erste Teil untersucht das Konvergenzverhalten einer auf dem Mittelpunktverfahren basierenden Zeitdiskretisierung der SLLG in einer Raumdimension, welche die Sphärenbedingung erhält. Für diese Zeitdiskretisierung wird Ratenkonvergenz in Wahrscheinlichkeit mit Ordnung $1/2$ bewiesen. Zu den Hauptschwierigkeiten, die zu überwinden sind, gehören sowohl die analytische, als auch die numerische Behandlung der Nichtlinearitäten, sowie des stochastischen Integralterms. Durchgeführte Simulationen unterstützen die nachgewiesene Konvergenzordnung.

Die beiden weiteren Teile der vorliegenden Dissertation befassen sich mit der Approximation stochastischer optimaler Steuerungsprobleme. Im zweiten Teil wird ein System stochastischer partieller Differentialgleichungen (kurz FBSHE) untersucht, welches aus einer vorwärtsgerichteten, sowie einer rückwärtsgerichteten stochastischen Wärmeleitungsgleichung (kurz BSHE) besteht, und aus den Bedingungen erster Ordnung eines stochastischen optimalen Steuerungsproblems motiviert ist. Ratenkonvergenz mit optimaler Ordnung für eine Raumdiskretisierung basierend auf $\mathbb{P}_1$-finiten Elementen wird für die BSHE sowie das System FBSHE nachgewiesen. Anschließend wird eine Volldiskretisierung der FBSHE vorgeschlagen, welche zusätzlich auf dem impliziten Eulerverfahren basiert. Dabei auftretende bedingte Erwartungen werden mit der *least squares Monte-Carlo method* approximiert. Für die Simulation der FBSHE wird neben einer Picard-Iteration ein stochastisches Gradientenverfahren vorgeschlagen, welches sich in durchgeführten Simulationen als flexibler erweist.

Im dritten Teil werden die Methoden des zweiten Teils mit der SLLG verbunden. Ziel ist es, die optimale Kontrolle der Dynamik einer fixierten Anzahl ferromagentischer Partikel bei erhöhten Temperaturen zu untersuchen. Das stochastische optimale Steuerungsproblem setzt sich dabei aus der Minimierung eines quadratischen Funktionals unter der Nebenbedingung, dass die SLLG erfüllt ist zusammen. Existenz eines solchen Minimums wird nachgewiesen. Damit verbundene Bedingungen erster Ordnung welche aus einem System von gekoppelten vorwarts-rückwärtsgerichteten stochastischen Differentialgleichungen besteht werden unter Verwendung einer strukturerhaltenden Zeitdiskretisierung, der *least squares Monte-Carlo method*, sowie des stochastischen Gradientenverfahrens diskretisiert sowie simuliert. In computergestützten Experimenten wird der Einfluss der optimalen stochastischen Kontrolle im Falle des Zusammenwirkens von Anisotropie, Streufeld, Austauschenergie und Rauschen untersucht.

Acknowledgment

I thank everyone who supported me during the preparation of this thesis.

There are some people deserving a special note of thanks.

I warmly thank my advisor, Prof. Dr. Andreas Prohl, for the opportunity to write my thesis in the interesting and important field of stochastic optimal control. I thank him for his scholastic guidance, his constant inspiration, and his valuable suggestions and patience. I also thank him, Prof. Dr. Ľubomír Baňas, and Prof. Dr. Michael Tretyakov for being the three evaluators of this thesis.

Moreover, I thank Prof. Dr. Christian Bender, Prof. Dr. Ulrich Nowak, and Dr. Elmar Teufl for comments and suggestions which improved this thesis.

I thank my recent and former colleagues of the numerical analysis group: Bernd Brumm, Erich Carelli, Jonas Hähnle, Jonas Konzelmann, Balázs Kovács, Heiko Kröner, Ananta Majee, Dhia Mansour, Mikhail Neklyudov, Christian Power, Anton Prochel, Ailyn Schäfer, Christian Schellnegger, Jonathan Seyrich, Hanna Wallach, and Daniel Weiß.

In particular, I am grateful to my colleagues and friends Erich Carelli, Markus Klein, and Elmar Teufl, who were involved in many helpful mathematical discussions, who gave me programming advice and a constant moral support over the last years.

Much gratitude to Leonard J. Konrad for reading parts of this thesis and giving me feedback.

Finally, I would like to thank my parents, my sister Daniela, and Nici, who have always been supporting and encouraging me.

The simulations carried out in Part II and III were performed on the computational resource bwUniCluster funded by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC.

Contents

List of Figures

List of Tables

Introduction

Magnetism is of fundamental importance for modern civilization: For example, compasses greatly simplified navigation on the sea, loudspeaker are part of every modern audio equipment and of every mobile phone, magnetic recording is the key technology for mass storage, and magnetic resonance imaging is used in medical diagnosis. Magnetization processes are described by ferromagnetic models. Their study leads to a deeper understanding of this process and provides information which is experimentally inaccessible or which is linked to unnecessary costs.

One well-accepted physical model describing the magnetization of a ferromagnetic body has been proposed in the pioneering works of Landau and Lifshitz [LL35] and of Gilbert [Gil55, Gil04]. The time evolution in $[0,T]$ of the magnetization of a ferromagnetic body $D \subset \mathbb{R}^d$, $d = 1, 2, 3$ is modeled by the function

$$\boldsymbol{m} : [0,T] \times D \to \{\boldsymbol{x} \in \mathbb{R}^3;\, |\boldsymbol{x}|_{\mathbb{R}^3} = M_s\}, \tag{SPHERE}$$

where $M_s > 0$ denotes the saturation magnetization. The vector $\boldsymbol{m}(t, \boldsymbol{x})$ is the direction of the magnetization at time $t \in [0,T]$ and at position $\boldsymbol{x} \in D$. Since we consider the equation without any physical dimensions, it is natural to set $M_s \equiv 1$. The Landau-Lifshitz-Gilbert equation (LLG)

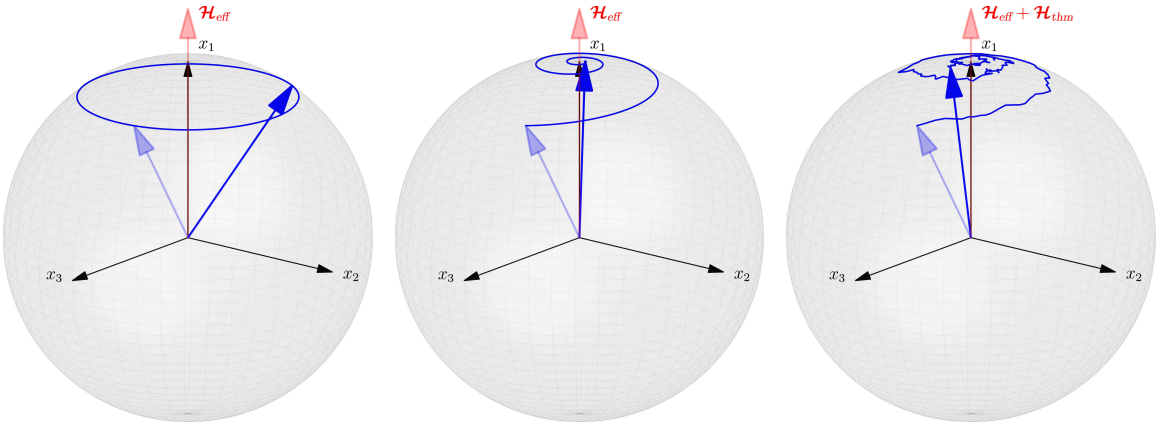
$$\boldsymbol{m}_t = -\alpha \boldsymbol{m} \times (\boldsymbol{m} \times \boldsymbol{\mathcal{H}}_{\mathit{eff}}) + \boldsymbol{m} \times \boldsymbol{\mathcal{H}}_{\mathit{eff}},$$

where $0 < \alpha \ll 1$, describes the dynamics of the magnetization in the presence of an effective field $\boldsymbol{\mathcal{H}}_{\mathit{eff}} = \boldsymbol{\mathcal{H}}_{\mathit{eff}}(\boldsymbol{m})$. The first term on the right-hand-side of the LLG is a phenomenological term which was introduced by Landau and Lifshitz in [LL35] and accounts for the damping of the magnetization $\boldsymbol{m}$ towards the effective field $\boldsymbol{\mathcal{H}}_{\mathit{eff}}$. The second term in the LLG is commonly referred to as precession term and describes the rotation of the magnetization $\boldsymbol{m}$ around the effective field $\boldsymbol{\mathcal{H}}_{\mathit{eff}}$. The dynamics of these terms are visualized in Figure I.

The effective field $\boldsymbol{\mathcal{H}}_{\mathit{eff}}$ in the LLG acting on ferromagnetic particle ensembles is the negative of the gradient of the total magnetic energy $\mathcal{E}$ which consists of several contributions:

$$\boldsymbol{\mathcal{H}}_{\mathit{eff}} = \boldsymbol{\mathcal{H}}_{\mathit{exch}} + \boldsymbol{\mathcal{H}}_{\mathit{ext}} + \boldsymbol{\mathcal{H}}_{\mathit{ani}} + \boldsymbol{\mathcal{H}}_{d}.$$

The first contribution is the exchange energy, which penalizes spatial changes in the magnetization and is usually modeled by $\boldsymbol{\mathcal{H}}_{\mathit{exch}} = \Delta \boldsymbol{m}$. The second part $\boldsymbol{\mathcal{H}}_{\mathit{ext}}$ is an external magnetic field (so called Zeeman contribution) and favors the alignment with an external field. It is modeled according to $\boldsymbol{\mathcal{H}}_{\mathit{ext}} = \boldsymbol{u}$, for some $\boldsymbol{u} : [0,T] \times D \to \mathbb{R}^3$. Crystallographic properties of the ferromagnetic material are taken into account by the anisotropy energy $\boldsymbol{\mathcal{H}}_{\mathit{ani}}$, where the magnetization $\boldsymbol{m}$ prefers to align with the (crystallographic) easy



(a) precession **(b)** precession/damping **(c)** precession/damping

Figure I. Dynamics of the precession and damping term in the LLG, respectively SLLG.

axis $\boldsymbol{e} \in \mathbb{R}^3$; its contribution is modeled by $\boldsymbol{\mathcal{H}}_{ani} = -\nabla\phi(\boldsymbol{m})$, where $\phi : \mathbb{S}^2 \to \mathbb{R}$ denotes the anisotropy density. The fourth contribution is the demagnetization field (also called stray-field), which takes the interaction with a surrounding magnetic field into account and is usually modeled using Maxwell's equations.

The partial differential equation LLG is strongly nonlinear and its solutions must satisfy the non-convex side constraint (SPHERE). Existence of a solution has been analyzed in the literature: for $d = 2, 3$ existence of weak solutions is shown for the prototype case $\boldsymbol{\mathcal{H}}_{eff} = \Delta\boldsymbol{m}$, see e.g. [AS92, GH93]. Numerical studies carried out in [BBP08, BP06] show that in this setup possible finite time blow-up from smooth initial data may be expected. For $d = 1$ however, existence of a weak solution with improved regularity properties is well known.

The solution of the LLG cannot be expressed by an explicit formula in general. Computer-based simulations are necessary to get insights concerning the behavior and properties of the solution of this equation, requiring fast and reliable numerical schemes, see [BP06, Cim08, Pro01] and references therein for an overview of numerical schemes for the LLG.

Thermal fluctuations $\boldsymbol{\mathcal{H}}_{thm}$ are included into the LLG by perturbing the effective field $\boldsymbol{\mathcal{H}}_{eff}$ to describe random changes which occur by the interaction of the ferromagnet with a surrounding heat bath. The study of random fluctuations in the dynamics of magnetism has been proposed by Néel [Née46] and later been studied by Brown Jr. [Bro63] where the focus is on a single nanomagnetic particle. After that, the stochastic equation has been considered by physicists; see e.g. [BMS09, BKM$^+$99, CL93, GPL98] among others. The random thermal fluctuations are usually modeled by $\boldsymbol{\mathcal{H}}_{thm} = \dot{\boldsymbol{W}}$ which has to be understood in the sense of Stratonovich in order to satisfy the saturation magnetization. Here $\boldsymbol{W}$ is a $\boldsymbol{Q}$-Wiener process; see Chapter 2 of Part I. The stochastic version of the Landau-Lifshitz-Gilbert equation (SLLG) takes the form

$$\boldsymbol{m}_t = -\alpha\boldsymbol{m} \times (\boldsymbol{m} \times \boldsymbol{\mathcal{H}}_{eff}) + \boldsymbol{m} \times (\boldsymbol{\mathcal{H}}_{eff} + \boldsymbol{\mathcal{H}}_{thm}),$$

where fluctuations in the damping part are neglected due to $\alpha \ll 1$, see [GPL98], and

the works [BBNP13, BMS09, GPL98] for further background of the physical model. The influence of the fluctuations in the equation is visualized in Figure I.

The modeling, analysis, and numerics of the stochastic Landau-Lifshitz-Gilbert equation is an active field of research: In recent works [AdBH14,BBNP13,BBNP14,BGJ12,BGJ13] the existence of a weak martingale solution of the SLLG is established. In [BGJ13], where a rigorous mathematical treatment of the SLLG is initiated, the existence of a weak martingale solution in the case of scalar-valued noise is provided using a standard Faedo-Galerkin approximation and compactness arguments. By similar arguments existence of a weak martingale solution is shown in [BGJ12] for one-dimensional domains, and moreover, by pathwise uniqueness which holds in one space dimension, existence of a strong solution is proven.

Numerical approximation schemes for the SLLG are studied in [AdBH14,BBNP13,BBNP14] in two and three space dimensions. In [BBNP13, BBNP14] a finite element method, a midpoint scheme, and a random walk approximation of Wiener increments are combined to obtain a fully practical discretization of the SLLG, whereas in [AdBH14] a semi-discrete scheme which extends the projection algorithm of [AJ06] to the stochastic case is used to construct a weak martingale solution.

The time discretization of the SLLG used in [BBNP13, BBNP14] takes the form

$$\begin{aligned} \boldsymbol{m}^{j+1} = \boldsymbol{m}^j &- \alpha k \boldsymbol{m}^{j+\frac{1}{2}} \times (\boldsymbol{m}^{j+\frac{1}{2}} \times \boldsymbol{\mathcal{H}}_{\mathit{eff}}^{j+1}) \\ &+ k \boldsymbol{m}^{j+\frac{1}{2}} \times \boldsymbol{\mathcal{H}}_{\mathit{eff}}^{j+1} + \boldsymbol{m}^{j+\frac{1}{2}} \times \Delta_j \boldsymbol{W}, \end{aligned} \tag{SLLG-MID}$$

where $I_k := \{t_j\}_{j=0}^J$ is a uniform partition of $[0,T]$ of equi-distant mesh size $k>0$, $\Delta_j \boldsymbol{W} := \boldsymbol{W}(t_{j+1}) - \boldsymbol{W}(t_j)$ denotes the stochastic increment, and $\boldsymbol{m}^{j+\frac{1}{2}} := \frac{1}{2}(\boldsymbol{m}^{j+1} + \boldsymbol{m}^j)$. The diffusive term is evaluated at $\boldsymbol{m}^{j+\frac{1}{2}}$ according to the definition of the Stratonovich integral, the first terms of the damping and precession term are evaluated at $\boldsymbol{m}^{j+\frac{1}{2}}$ in order to maintain the sphere constraint (in the $\mathbb{P}$-a.s. sense). Finally and the evaluation of the effective field $\boldsymbol{\mathcal{H}}_{\mathit{eff}}$ at time t_{j+1} allows for an estimate for the discrete energy ($q \in \mathbb{N}$)

$$\mathbb{E}\left[\sup_{r=0,\dots,J-1}\Big(\|\nabla \boldsymbol{m}^{r+1}\|_{\mathbf{L}^2}^{2q} + \sum_{j=0}^{r}\|\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\|_{\mathbf{L}^2}^2 + k\sum_{j=0}^{r}\|\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2\Big)\right] \le C_{T,q}$$

for $\boldsymbol{\mathcal{H}}_{\mathit{eff}} = \Delta\boldsymbol{m}$. This bound is essential in [BBNP13,BBP13] for the construction of a weak martingale solution and below to conduct the error analysis for the scheme (SLLG-MID). Schemes of the type (SLLG-MID) were already used for the LLG in [BP06], and, moreover, for the SLLG in [BBNP13, BBP13] to perform numerical studies.

In the case of a finite ensemble of nanomagnetic particles (SDE case) a midpoint type scheme is often used as a benchmark for accuracy, see e.g. [MTF+10]. In this particular case, the strong order of convergence 1/2 is shown in [MRT02, NP13]. However, for the SPDE no order of convergence is available so far. Part I of this thesis fills this gap.

In the first part an error analysis for the scheme (SLLG-MID) approximating the SLLG is carried out, which is limited to one-dimensional domains where strong solutions with improved regularity exist. The main tools for the error analysis are reformulations of the damping term in the SLLG as well as in the semi-discrete scheme (SLLG-MID). Local

strong rates of convergence with order 1/2 are shown on a subset $\widetilde{\Omega}_k \subset \Omega$, where $\mathbb{P}[\widetilde{\Omega}_k] \to 1$ for $k \to 0$. The restriction to this subset is needed, since terms of the form

$$Ck \sum_{j=0}^{J} \|\nabla \boldsymbol{m}^j\|_{\mathbf{L}^2}^4 \|\boldsymbol{m}(t_{j+1}) - \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2$$

occur amongst others in the error analysis. These terms emerge due to nonlinear drift terms which are only locally Lipschitz. The factor $\|\nabla \boldsymbol{m}^j\|_{\mathbf{L}^2}^4$ prevents a direct application of a Gronwall argument. The idea is to restrict the error analysis to subsets $\widetilde{\Omega}_k \subset \Omega$, where this factor is bounded such that a Gronwall argument is applicable, and then to show that $\mathbb{P}[\widetilde{\Omega}_k] \to 1$ for $k \to 0$ using the bound for the discrete energy above and further estimates. By Chebychev's inequality this implies convergence in probability with rate 1/2. Computational studies are conducted indicating a strong rate of order 1/2 and a weak rate of order 1. This part is based on the publication [Dun15]. In [CP12] a similar approach is used in a different context.

In many applications it is relevant to control the dynamics of the stochastic Landau-Lifshitz-Gilbert equation, using the external field $\boldsymbol{\mathcal{H}}_{ext}$, so that its solution is in some sense close to a given desirable target. For this purpose, we search for an external field $\boldsymbol{u}$, known as control, which minimizes the functional

$$\mathcal{J}_{SLLG} = \frac{1}{2}\mathbb{E}\left[\int_0^T \left(\delta\|\boldsymbol{m}(t) - \widetilde{\boldsymbol{m}}(t)\|^2 + \lambda\|\boldsymbol{u}(t)\|^2\right) \mathrm{d}t + \kappa\|\boldsymbol{m}(T) - \widetilde{\boldsymbol{m}}(T)\|^2\right]$$

subject to the stochastic Landau-Lifshitz-Gilbert equation, with $\boldsymbol{\mathcal{H}}_{ext} = \boldsymbol{u}$. This functional has three components which are scaled by $\delta \geq 0$, $\kappa \geq 0$, and $\lambda > 0$. The first and third measure the distance of the magnetization $\boldsymbol{m}$ to a given deterministic desirable target magnetization $\widetilde{\boldsymbol{m}}$. The second component measures the cost of the control $\boldsymbol{u}$.

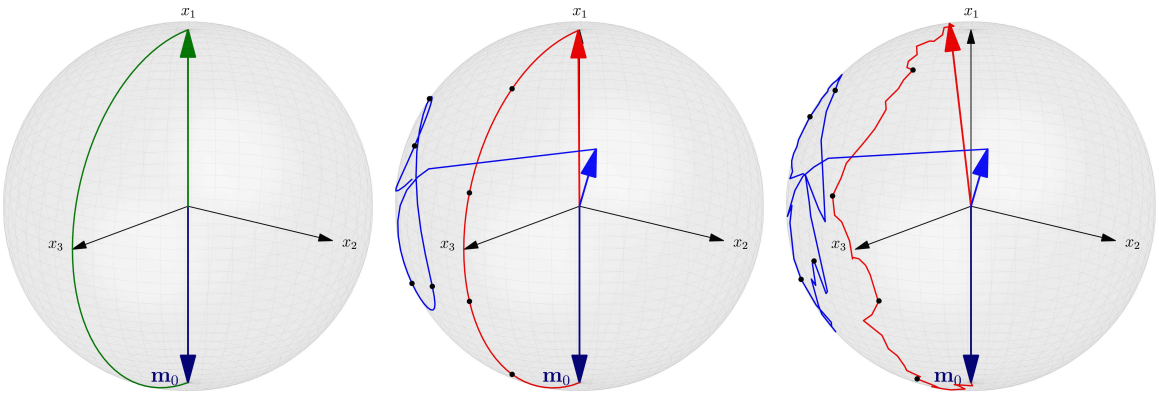
This stochastic control problem is of practical interest, e.g. in data recording, where fast and reliable ways to switch the magnetization are of great importance. Here one asks for an external field which switches the magnetization with minimal costs.

Controlling a ferromagnetic body has been addressed in the literature. However, the textbook [BMS09] admits on page 145:

> "One of the central problems in the research on precessional switching is the design of magnetic field pulses that will guarantee the magnetization reversal. In the literature, this problem has been mostly addressed experimentally or numerically by using a trial-and-error approach."

The problem of switching a single nanomagnetic particle optimally is studied mathematically in [AB09]. There an optimal control is derived for a special cost functional. Controllability of a finite ensemble of nanomagnetic particles is shown in [ACLP11]. In [DKPS15] existence of a minimum, and a characterization of this minimum by the first order necessary conditions is shown in the case of the one-dimensional LLG. Moreover, a convergent semi-discretization in time (semi-implicit Euler scheme) is proposed and computational studies using this scheme and a further modification by an additional projection step are carried out there.

In contrast to the previous work in this area, the third part of the thesis, which is based on the manuscript [DP16] submitted for publication, focuses on minimizing $\mathcal{J}_{SLLG}$ subject to a finite ensemble of nanomagnetic particles where (random) thermal fluctuations are taken into account. Our particular interest is to develop a fully practicable method which approximatively computes the explicit structure of the control. Figure II shows the behavior of one path of the optimal state and the direction of the optimal control in the case of a single nanomagnetic particle.



(a) deterministic target profile

(b) deterministic optimal control problem

(c) stochastic optimal control problem

Figure II. The target profile $\widetilde{\boldsymbol{m}}$ (green), one trajectory of the optimal state (red) and the direction of the control (blue) in the case of a stochastic control problem and the corresponding deterministic control problem for a single nanomagnetic particle.

First, we prove the existence of a minimum of the cost functional $\mathcal{J}_{SLLG}$ in terms of a weak control $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \boldsymbol{W}, \boldsymbol{u})$ of the stochastic optimal control problem. We propose a fully practicable method to approximatively solve the generalized Hamiltonian system which is formally linked to the optimal stochastic control problem. The generalized Hamiltonian system consists of the state equation SLLG, the adjoint equation, which is a backward stochastic differential equation, and an optimality condition. This forward-backward problem is discretized using adequate time discretizations. Emerging conditional expectations are approximated by a least squares Monte-Carlo method, which returns the (deterministic) regression functions $(\mathcal{P}_R^j(\boldsymbol{m}^j), \mathcal{Q}_R^j(\boldsymbol{m}^j))$ for the stochastic backward equation.

By a combination of the optimality condition, the (deterministic) regression functions, and a certain least squares Monte-Carlo method, a stochastic gradient method to resolve the forward-backward character of this problem is developed. This method is of crucial importance, since a Picard type algorithm, which is so far only reported in the literature to deal with this kind of problems (see e.g. [BZ08]), is only applicable for very limited time durations $T > 0$. We employ this stochastic gradient method in our computational experiments to study the stochastic optimal control of spin switching in the case of a single particle, and of a finite particle ensemble.

The second part is of independent relevancy, though it lays the fundaments to the numerical setup of the third part. This part is based on the manuscript [DP15] accepted for publication in SIAM Journal on Scientific Computing. Here the focus is on controlling the dynamics of

the stochastic heat equation (SHE)

$$\begin{aligned} \mathrm{d}X(t) &= \Big[\Delta X(t) + U(t)\Big]\,\mathrm{d}t + \sum_{i=1}^{n} \nu^i(t) X(t)\,\mathrm{d}W^i(t), \\ X(0) &= x_0, \end{aligned}$$

with homogeneous Dirichlet boundary conditions. For this purpose, we search for an $\mathbb{F}$-adapted stochastic control U which minimizes

$$\mathcal{J}_{SHE} = \mathbb{E}\left[\int_0^T \frac{1}{2}\Big(\|X(t) - \widetilde{X}(t)\|_{\mathbb{L}^2}^2 + \lambda \|U(t)\|_{\mathbb{L}^2}^2\Big)\,\mathrm{d}t + g\big(X(T)\big)\right],$$

subject to the SHE. Following the work [Ben83, Zho93] the first order necessary conditions consist of the state equation SHE, the adjoint equation which is the backward stochastic heat equation (BSHE)

$$\begin{aligned} \mathrm{d}Y(t) &= \Big[-\Delta Y(t) - \sum_{i=1}^{n} \nu^i(t) Z^i(t) - \Big(X(t) - \widetilde{X}(t)\Big)\Big]\,\mathrm{d}t + \sum_{i=1}^{n} Z^i(t)\,\mathrm{d}W^i(t), \\ Y(T) &= Dg\big(X(T)\big), \end{aligned}$$

with homogeneous Dirichlet boundary conditions, and the optimality condition $Y(t) + \lambda U(t) = 0$. By inserting the optimality condition in the SHE, we obtain the coupled system of forward and backward stochastic heat equations (FBSHE)

$$\begin{aligned} \mathrm{d}X(t) &= \Big[\Delta X(t) - \frac{1}{\lambda} Y(t)\Big]\,\mathrm{d}t + \sum_{i=1}^{n} \nu^i(t) X(t)\,\mathrm{d}W^i(t), \\ \mathrm{d}Y(t) &= \Big[-\Delta Y(t) - \sum_{i=1}^{n} \nu^i(t) Z^i(t) - \Big(X(t) - \widetilde{X}(t)\Big)\Big]\,\mathrm{d}t + \sum_{i=1}^{n} Z^i(t)\,\mathrm{d}W^i(t), \\ X(0) &= x_0, \qquad Y(T) = Dg\big(X(T)\big). \end{aligned}$$

These equations are studied in the second part, where the focus is on the approximation and simulation of the BSHE and the FBSHE.

Backward stochastic partial differential equations naturally extend backward stochastic differential equations (BSDE) and usually appear as adjoint equations in several stochastic optimal control problems. The field of backward stochastic differential equations is fastly growing; various numerical methods to solve and simulate backward stochastic differential equations have been developed. PDE based methods, e.g. [DMP96, MT06], yield very precise numerical results for low-dimensional BSDEs. For high-dimensional BSDEs a direct discretization in time, see e.g. [BT04, Zha04], combined with an approximation of the occurring conditional expectations, see e.g. [BD07, BT04, CM12, GLW05, GT14], seems to be more applicable.

To the best of our knowledge, the results in Part II firstly address the numerical approximation and simulation of a backward stochastic partial differential equation. We start with a spatial discretization using $\mathbb{P}_1$-finite elements of the BSHE and, in particular, of the FBSHE motivated above. We prove optimal strong rates of convergence for the BSHE using

a result on improved regularity of the strong solution from [DT12]. By a combination of this result and a Picard-iteration, we show optimal strong rates of convergence in the case of the FBSHE for short time durations $T > 0$.

Fully implementable algorithms to simulate the BSHE and the FBSHE are discussed, which are based on a direct discretization in time of the spatially discretized BSHE using the implicit Euler scheme, and a partition estimation method to approximate conditional expectations, which is a special case of the least squares Monte-Carlo method; see e.g. [GLW05]. The latter method for the approximation of $\mathbb{E}[\,\cdot\,|X_h^j]$ is based on a partition of the range of X_h^j, where X_h is the solution of the spatially discretized SHE. This is a high-dimensional task since the dimension of X_h^j depends on the spatial discretization parameter $h > 0$. To accurately resolve X_h^j, we propose two partitioning strategies, which depend on the law of X_h^j: the Binary Tree Cuboids **(BTC)**, which is built on the idea that the constructed regions should be equally likely; and the Voronoi Partition Method **(V)**, where a nearest neighbor clustering with respect to additional realizations of X_h^j is used to define the partition. By these approximations, we have an implementable algorithm to simulate the deterministic regression functions $(\mathcal{Y}_h^j(X_h^j), \mathcal{Z}_h^{i,j}(X_h^j))$ which approximatively solve the BSHE.

Computational experiments are carried out to study the behavior of both partition strategies in the case of the BSHE. The convergence behavior of the space-discretization of the BSHE is experimentally studied supporting the analytically obtained rates.

Three different schemes are considered to resolve the forward-backward character of the necessary optimality conditions: The first is a Picard-type iteration, where the optimality condition is inserted in the forward equation; see the FBSHE above. Similarly to [BZ08], the solution of the backward stochastic heat equation at the previous Picard iteration step is then used in the (forward) stochastic heat equation. However, this scheme converges only for restrictive choices of the data (small time durations $T > 0$). To overcome this restrictions we propose a new stochastic gradient method, which uses the optimality condition combined with the representation of the solution of the backward equation by (deterministic) regression functions. For stochastic LQ problems where $g(\cdot)$ in $\mathcal{J}_{SHE}$ is quadratic, we propose a third scheme which avoids the computation of conditional expectations by exploiting linearity of the problem.

These schemes are employed in computational experiments to study their convergence behavior: the stochastic gradient method returns the same approximation as the Picard-type iteration for those choices of the data, where the Picard-type iteration converges. In this case the Picard-type iteration is less time consuming than the stochastic gradient method. In the other case, the stochastic gradient method converges and returns an approximation of the control with a significant reduction in the cost functional $\mathcal{J}_{SHE}$. For stochastic LQ problems this approximation almost matches the approximation which is obtained by the scheme which avoids the computation of conditional expectations.

The simulation methodology of the second part is applied in the third part to the SLLG constrained stochastic control problem, where the corresponding optimality condition is slightly more difficult. The stochastic gradient method can be adapted to this framework by a further approximation exploiting the properties of the Voronoi Partition method.

Part I.

Convergence with rates for a time discretization of the stochastic Landau-Lifshitz-Gilbert equation

1. Introduction

This part[1] considers a system of equations describing the magnetization of a ferromagnetic wire $D := (\underline{x}, \overline{x}) \subset \mathbb{R}$ with periodic boundary conditions. Let $D_T := (0,T) \times D$ with $0 < T < \infty$, and $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space. The aim is to find a process $\boldsymbol{m} : D_T \times \Omega \to \mathbb{S}^2 := \{\boldsymbol{x} \in \mathbb{R}^3;\, |\boldsymbol{x}|_{\mathbb{R}^3} = 1\}$, which satisfies the stochastic Landau-Lifshitz-Gilbert equation (SLLG)

$$\begin{aligned} \boldsymbol{m}_t(t,x,\omega) = &-\alpha \boldsymbol{m}(t,x,\omega) \times \Big(\boldsymbol{m}(t,x,\omega) \times \Delta \boldsymbol{m}(t,x,\omega)\Big) + \boldsymbol{m}(t,x,\omega) \times \Delta \boldsymbol{m}(t,x,\omega) \\ &+ \boldsymbol{m}(t,x,\omega) \times \circ \dot{\boldsymbol{W}}(t,x,\omega) \qquad \forall (t,x,\omega) \in D_T \times \Omega, \\ \boldsymbol{m}(0,\cdot) &= \boldsymbol{m}_0. \end{aligned} \tag{1.1}$$

See the works [BMS09, KP06] and references therein for a detailed description of the deterministic Landau-Lifshitz-Gilbert equation and the recent works [BBNP13,BBNP14,BMS09, BGJ12, BGJ13] for a description of the stochastic version.
The right-hand-side of equation (1.1) consists of three terms: The first term is a phenomenological damping term (relaxation term) to model dissipation effects, which is scaled by $0 < \alpha \ll 1$. The second term is commonly referred to as precession term. The third term describes thermal fluctuations by multiplicative colored Gaussian noise $\boldsymbol{m}(t,x,\omega) \times \circ \dot{\boldsymbol{W}}(t,x,\omega)$, where the stochastic integral is understood in Stratonovich sense.

The stochastic Landau-Lifshitz-Gilbert equation is of interest in recent research: In [BGJ13] the existence of a weak martingale solution of problem (1.1) (for $D \subset \mathbb{R}^d$ with $d = 2, 3$ and Neumann boundary conditions) in the case of scalar noise is proven using a standard Faedo-Galerkin approximation and compactness methods. In [BGJ12], scalar noise in both, the precession and the damping term is considered for the special case $D \subset \mathbb{R}$. A weak martingale solution is constructed similarly as in [BGJ13], and by pathwise uniqueness which holds in 1d, the existence of a strong solution is proven. The works [BBNP13, BBNP14] use a fully practicable finite element scheme to construct a weak martingale solution of (1.1) (for $D \subset \mathbb{R}^d$ with $d = 2, 3$ and Neumann boundary conditions). The used scheme consists of $\mathbf{P}_1$-finite elements for the spatial discretization, and the midpoint scheme for the time discretization. In [BBNP13, BBP13], this scheme is used in order to computationally study possible blow-up behavior of the solution in 2d. Further results in this direction, including long-time dynamics, are contained in [BBNP13].

The focus of the present part lies on the numerical solution of problem (1.1) on the circle $D \subset \mathbb{R}$. We prove rates of convergence for the following time discretized scheme:

[1] Part I is based on the publication [Dun15] in IMA Journal of Numerical Analysis published by Oxford University Press.

Algorithm 1.1
Consider a uniform partition $I_k := \{t_j\}_{j=0}^J$ covering $[0,T]$ with equi-distant mesh-size $k = T/J > 0$, where $t_0 := 0$ and $t_J := T$. Let $\boldsymbol{m}^0 := \boldsymbol{m}_0$. For every $j \geq 0$ and $\Delta_j \boldsymbol{W} := \boldsymbol{W}(t_{j+1}) - \boldsymbol{W}(t_j) \sim \mathcal{N}(0, k\boldsymbol{Q})$ determine the $\mathbf{W}^{1,2}(D;\mathbb{R}^3)$-valued random variable $\boldsymbol{m}^{j+1}$, such that $\mathbb{P}$-a.s.

$$\boldsymbol{m}^{j+1} = \boldsymbol{m}^j - \alpha k \boldsymbol{m}^{j+\frac{1}{2}} \times (\boldsymbol{m}^{j+\frac{1}{2}} \times \Delta \boldsymbol{m}^{j+1}) + k\boldsymbol{m}^{j+\frac{1}{2}} \times \Delta \boldsymbol{m}^{j+1} + \boldsymbol{m}^{j+\frac{1}{2}} \times \Delta_j \boldsymbol{W}.$$

The scheme has been studied in [BP06] in the deterministic and in [BBNP14] in the stochastic context: iterates are $\mathbb{S}^2$-valued and satisfy a discrete energy bound (see Lemma 3.4). In [BBNP13] it is pointed out, that preserving the sphere constraint is essential to obtain a proper long-time behavior of the scheme.

The main result of the present part is strong convergence with rate $1/2$ (cf. Theorem 4.1)

$$\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_k} \sup_{j=0,\dots,J-1} \|\boldsymbol{m}(t_{j+1}) - \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2\Big] \leq Ck^{1-\epsilon} \qquad (\epsilon > 0) \tag{1.2}$$

on a subset $\widetilde{\Omega}_k \subset \Omega$, such that $\mathbb{P}[\Omega \setminus \widetilde{\Omega}_k] \to 0$ for $k \to 0$. It is due to the non-Lipschitz nonlinear terms, that this local estimate is needed to single out those solution paths where a Gronwall argument can be accomplished. A simple consequence of (1.2) is the convergence in probability with rate $1/2$ for iterates of Algorithm 1.1. Here, the work [CP12] is followed, where a similar strategy was first proposed for the stochastic incompressible Navier-Stokes equation to account for nonlinear effects in the problem.

The computational studies detailed in Chapter 5 suggest a strong rate of order $1/2$, and a weak rate of order 1 in both cases: for infinite dimensional colored noise, and also in the case of space-time white noise where no theoretical result even with respect to solvability exists. Those results are consistent with the work [NP13], where the strong rate of order $1/2$ and the weak rate of order 1 is proven for a similar time discretization in the case of a finite ensemble of spins.

The main tool which we use is to reformulate the damping term for both, equation (1.1), and Algorithm 1.1: For the analytical problem, formula $\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = \langle \boldsymbol{a}, \boldsymbol{c}\rangle_{\mathbb{R}^3}\boldsymbol{b} - \langle \boldsymbol{a}, \boldsymbol{b}\rangle_{\mathbb{R}^3}\boldsymbol{c}$ gives

$$\boldsymbol{m} \times (\boldsymbol{m} \times \Delta \boldsymbol{m}) = \langle \boldsymbol{m}, \Delta \boldsymbol{m}\rangle_{\mathbb{R}^3}\boldsymbol{m} - |\boldsymbol{m}|_{\mathbb{R}^3}^2 \Delta \boldsymbol{m},$$

and since the constraint $|\boldsymbol{m}(t,\boldsymbol{x},\cdot)|_{\mathbb{R}^3}^2 = 1$ *Leb*-a.e. in D for all $t \in [0,T]$ $\mathbb{P}$-a.s. holds, it can be further reformulated as

$$\boldsymbol{m} \times (\boldsymbol{m} \times \Delta \boldsymbol{m}) = -|\nabla \boldsymbol{m}|_{\mathbb{R}^3}^2 \boldsymbol{m} - \Delta \boldsymbol{m} \tag{1.3}$$

to achieve a semilinear problem. This reformulation of the cubic nonlinearity allows us to show improved regularity results in space for the strong solution, by assuming more regularity on the initial value $\boldsymbol{m}_0$ and the $\boldsymbol{Q}$-Wiener process $\boldsymbol{W}$ (see Propositions 3.2 and 3.3).
On the other hand, if the above formula for the corresponding term in Algorithm 1.1 is employed, we obtain

$$\boldsymbol{m}^{j+\frac{1}{2}} \times (\boldsymbol{m}^{j+\frac{1}{2}} \times \Delta \boldsymbol{m}^{j+1}) \tag{1.4}$$

$$
\begin{aligned}
&= \langle \boldsymbol{m}^{j+\frac{1}{2}}, \Delta \boldsymbol{m}^{j+1} \rangle_{\mathbb{R}^3} \boldsymbol{m}^{j+\frac{1}{2}} - |\boldsymbol{m}^{j+\frac{1}{2}}|^2_{\mathbb{R}^3} \Delta \boldsymbol{m}^{j+1} \\
&= - |\nabla \boldsymbol{m}^{j+1}|^2_{\mathbb{R}^3} \boldsymbol{m}^{j+\frac{1}{2}} - \frac{1}{2} \langle \boldsymbol{m}^{j+1} - \boldsymbol{m}^j, \Delta \boldsymbol{m}^{j+1} \rangle_{\mathbb{R}^3} \boldsymbol{m}^{j+\frac{1}{2}} - |\boldsymbol{m}^{j+\frac{1}{2}}|^2_{\mathbb{R}^3} \Delta \boldsymbol{m}^{j+1}.
\end{aligned}
$$

However, this reformulation does not allow for better regularity results, as for the limiting problem the prefactor may degenerate, since $|\boldsymbol{m}^{j+\frac{1}{2}}|_{\mathbb{R}^3} = 0$ may be valid at places where $\boldsymbol{m}^{j+1} = \boldsymbol{m}^j$, which prevents the property of a strongly elliptic operator, which justifies a regularity shift.

The reformulation in equation (1.4) will be the starting point for the error analysis below. It is the lack of control of $\{\Delta \boldsymbol{m}^j;\, j \geq 0\}$ uniformly in $k > 0$ which is a main impediment to derive rates of convergence for the corresponding fully discrete schemes.

We consider equation (1.1) with periodic boundary conditions, in order to avoid technical problems due to boundary effects in the analysis below. The one-dimensionality is needed at several places: First, the existence of a strong solution of the SLLG in two or three space dimensions is an open problem. Second, to verify the improved regularity results in Section 3.1, and to derive the error estimates in Chapter 4, we frequently use the Gagliardo-Nirenberg inequality, where the dimension one is relevant.

This part of the thesis is organized as follows: We first collect some preliminaries in Chapter 2, including some notation, the definition of the stochastic integral and the needed assumptions on equation (1.1). In Chapter 3, we show a-priori bounds on both, the strong solution $\{\boldsymbol{m}(t);\, t \in [0,T]\}$ and the iterates of the semi-discrete scheme $\{\boldsymbol{m}^j;\, j = 0, \ldots, J\}$. We then use these results in Chapter 4 to prove the main result given in Theorem 4.1, the rate of convergence in probability for Algorithm 1.1. Computational studies are finally detailed in Chapter 5.

2. Preliminaries

First, we introduce some notation. The norm and the scalar product in $\mathbb{R}^3$ are denoted by $|\cdot|_{\mathbb{R}^3}$ and $\langle\cdot,\cdot\rangle_{\mathbb{R}^3}$, while $\|\cdot\|_{\mathbf{L}^2}$ and $(\cdot,\cdot)$ denote the norm and the scalar product in $\mathbf{L}^2(D;\mathbb{R}^3)$ respectively, which is the standard Lebesgue space of (equivalence classes of) square integrable functions $\boldsymbol{u}: D \to \mathbb{R}^3$. The norm in $\mathbf{W}^{k,2}(D;\mathbb{R}^3)$ for $k = 1, 2, 3$, the space of functions $\boldsymbol{u} \in \mathbf{L}^2(D;\mathbb{R}^3)$ where the weak derivative up to the k-th order also belongs to $\mathbf{L}^2(D;\mathbb{R}^3)$, is denoted by $\|\cdot\|_{\mathbf{W}^{k,2}}$.
A standard reference which is used in this section is [DPZ92]. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete probability space with a filtration $\mathbb{F} := \{\mathcal{F}_t;\, t \in [0,T]\}$, and let us consider a sequence $\{\beta^l(t);\, t \in [0,T]\}_{l\in\mathbb{N}}$ of i.i.d. $\mathbb{R}$-valued Brownian motions on the given filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. Let $\boldsymbol{\mathcal{K}}$ and $\boldsymbol{\mathcal{H}}$ be Hilbert spaces. By $\mathcal{S}_1(\boldsymbol{\mathcal{K}})$ we denote the space of trace class operators on $\boldsymbol{\mathcal{K}}$. Let $\boldsymbol{Q} \in \mathcal{S}_1(\boldsymbol{\mathcal{K}})$ be symmetric and non-negative, and let $\{\mathbf{e}_l\}_{l\in\mathbb{N}}$ be an orthonormal basis of $\boldsymbol{\mathcal{K}}$ consisting of eigenfunctions of $\boldsymbol{Q}$ with $\mathbb{R}^+$-valued eigenvalues $\{q_l\}_{l\in\mathbb{N}}$. The $\boldsymbol{\mathcal{K}}$-valued $\boldsymbol{Q}$-Wiener process $\boldsymbol{W} = \{\boldsymbol{W}(t);\, t \in [0,T]\}$ is defined by

$$\boldsymbol{W}(t) = \sum_{l=1}^{\infty} \sqrt{q_l}\mathbf{e}_l\beta^l(t) \quad \forall t \in [0,T].$$

For $p \geq 1$ we denote the space of equivalence classes of $\mathbb{F}$-progressively measurable processes $X : [0,T]\times\Omega \to \boldsymbol{\mathcal{H}}$ such that $\mathbb{E}[\int_0^T \|X(t)\|_{\boldsymbol{\mathcal{H}}}^p \,\mathrm{d}t] < \infty$ by $M^p([0,T],\mathbb{F};\boldsymbol{\mathcal{H}})$.
The space $\mathcal{L}(\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{H}})$ contains all linear bounded operators from $\boldsymbol{\mathcal{K}}$ to $\boldsymbol{\mathcal{H}}$, and the space $\mathcal{S}_2(\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{H}})$ consists of linear Hilbert-Schmidt operators from $\boldsymbol{\mathcal{K}}$ to $\boldsymbol{\mathcal{H}}$. The stochastic integral $\{\int_0^t \phi(s)\,\mathrm{d}\boldsymbol{W}(s);\, t \in [0,T]\}$ for any $\phi \in M^2([0,T],\mathbb{F};\mathcal{L}(\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{H}}))$ is defined as the continuous $\boldsymbol{\mathcal{H}}$-valued $\mathbb{F}$-martingale, such that

$$\int_0^t \phi(s)\,\mathrm{d}\boldsymbol{W}(s) = \sum_{m=1}^{M} \phi(t_{m-1})\Big(\boldsymbol{W}(t\wedge t_m) - \boldsymbol{W}(t\wedge t_{m-1})\Big) \quad \forall t \in [0,T],$$

for all step processes ϕ. This stochastic integral satisfies the Itô-isometry, i.e., for each $\phi \in M^2([0,T],\mathbb{F};\mathcal{L}(\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{H}}))$ there holds

$$\mathbb{E}\Big[\|\int_0^t \phi(s)\,\mathrm{d}\boldsymbol{W}(s)\|_{\boldsymbol{\mathcal{H}}}^2\Big] = \mathbb{E}\Big[\int_0^t \|\phi(s)\boldsymbol{Q}^{\frac{1}{2}}\|_{\mathcal{S}_2(\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{H}})}^2\,\mathrm{d}s\Big] \quad \forall t \in [0,T].$$

The following assumptions on data $\boldsymbol{W}$, $\boldsymbol{Q}$ and $\boldsymbol{m}_0$ are needed below:

Assumption $\mathbf{A_1}$
Let $\boldsymbol{\mathcal{K}} \subset \mathbf{W}^{3,2}(D;\mathbb{R}^3)$ and the embedding is continuous.

Assumption $\mathbf{A_2}$
Let $\boldsymbol{W} = \{\boldsymbol{W}(t,\cdot);\, t \in [0,T]\}$ be a $\boldsymbol{Q}$-Wiener process with values in $\boldsymbol{\mathcal{K}}$ and $\boldsymbol{Q}^{\frac{1}{2}} \in \mathcal{S}_1(\boldsymbol{\mathcal{K}})$ is a symmetric, non-negative operator.

Assumption $\mathbf{A_3}$
Let $\boldsymbol{m}_0 \in \mathbf{W}^{3,2}(D;\mathbb{S}^2)$.

Remark 2.1
Assumptions A_1 and A_3 are necessary to show additional regularity properties for the strong solution (see Chapter 3). We need assumption $\boldsymbol{Q}^{\frac{1}{2}} \in \mathcal{S}_1(\boldsymbol{\mathcal{K}})$ in the error estimate (see Chapter 4), as well as for most of the a-priori estimates in Chapter 3.

The following remark will be frequently used later, see [Ich82, Corollary 1.1].

Remark 2.2
The $\boldsymbol{Q}$-Wiener process $\boldsymbol{W}$ satisfies for all $r \in \mathbb{N}$ and all $0 \leq s < t$

$$\mathbb{E}\Big[\|\boldsymbol{W}(t) - \boldsymbol{W}(s)\|_{\boldsymbol{\mathcal{K}}}^{2r}\Big] \leq C(r)(\operatorname{Tr}\boldsymbol{Q})^r|t-s|^r.$$

By Assumption A_1, there exists $C > 0$, such that $\|\boldsymbol{u}\|_{\mathbf{W}^{3,2}} \leq C\|\boldsymbol{u}\|_{\boldsymbol{\mathcal{K}}}$ for all $\boldsymbol{u} \in \boldsymbol{\mathcal{K}}$, and thus

$$\mathbb{E}\Big[\|\boldsymbol{W}(t) - \boldsymbol{W}(s)\|_{\mathbf{W}^{3,2}}^{2r}\Big] \leq C(r)(\operatorname{Tr}\boldsymbol{Q})^r|t-s|^r.$$

3. A-priori estimates

In this chapter, we motivate important a-priori estimates for the strong solution of equation (1.1) and the iterates of Algorithm 1.1. These estimates will be useful for the error analysis in Chapter 4.

3.1. Strong solution

The following definition is deduced from [BGJ12].

Definition 3.1 (Strong solution to equation (1.1))
Let $T > 0$ and $\mathfrak{P} = (\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be given and let Assumptions A_1–A_3 hold. An $\mathbf{W}^{1,2}(D;\mathbb{R}^3)$-valued $\mathbb{F}$-adapted stochastic process $\{\boldsymbol{m}(t);\, t \in [0,T]\}$ on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is a strong solution of equation (1.1), if

1. $\mathbb{P}$-a.s. there holds $\boldsymbol{m}(\cdot,\cdot,\omega) \in C\big([0,T];\mathbf{L}^2(D;\mathbb{R}^3)\big) \cap L^8\big(0,T;\mathbf{W}^{1,2}(D;\mathbb{R}^3)\big)$;
2. $\mathbb{P}$-a.s. there holds $|\boldsymbol{m}(t,\cdot,\omega)|^2_{\mathbb{R}^3} = 1$ a.e. in D for all $t \in [0,T]$;
3. $\mathbb{P}$-a.s. there holds for every $t \in [0,T]$ and all $\phi \in \mathbf{W}^{1,2}(D;\mathbb{R}^3)$

$$\begin{aligned}(\boldsymbol{m}(t),\phi) - (\boldsymbol{m}_0,\phi) = \alpha \int_0^t \big(\nabla[\phi \times \boldsymbol{m}(s)] \times \boldsymbol{m}(s), \nabla \boldsymbol{m}(s)\big)\,\mathrm{d}s \\ - \int_0^t \big(\nabla\phi \times \boldsymbol{m}(s), \nabla \boldsymbol{m}(s)\big)\,\mathrm{d}s + \int_0^t \big(\phi, \boldsymbol{m}(s) \times \circ\, \mathrm{d}\boldsymbol{W}(s)\big).\end{aligned}$$

The existence of a strong solution of problem (1.1) can be shown similarly as in [BGJ12], where the existence is shown in the scalar noise case. The following proofs in this subsection are formal and can be done rigorously by using an appropriate finite dimensional approximation of problem (1.1), and tending the related approximation to the limit. For the sake of readability, in this subsection, the notation $\partial_x^i \cdot$ for the i-th derivative in space is used. In the proofs of Propositions 3.2 and 3.3 small constants $\delta, \widetilde{\delta} > 0$ are used for terms which will be absorbed to the left-hand-side, and inequalities are understood in the $\mathbb{P}$-a.s. sense.
Below, we reformulate the Stratonovich integral as Itô integral together with a Stratonovich correction term,

$$\sum_{l=1}^{\infty} \sqrt{q_l} \int_0^t \boldsymbol{m}(s) \times \mathbf{e}_l \circ \mathrm{d}\beta^l(s) = \sum_{l=1}^{\infty} \Big(\sqrt{q_l} \int_0^t \boldsymbol{m}(s) \times \mathbf{e}_l \,\mathrm{d}\beta^l(s) + \frac{1}{2} q_l \int_0^t \big(\boldsymbol{m}(s) \times \mathbf{e}_l\big) \times \mathbf{e}_l \,\mathrm{d}s \Big).$$

Moreover, by using $|\boldsymbol{m}(t,\cdot)|^2_{\mathbb{R}^3} = 1$ $\mathbb{P}$-a.s., the argumentation around equation (1.3) allows us to restate equation (1.1) as follows

$$\boldsymbol{m}_t - \alpha\partial_x^2\boldsymbol{m} = \alpha|\partial_x\boldsymbol{m}|^2_{\mathbb{R}^3}\boldsymbol{m} + \boldsymbol{m}\times\partial_x^2\boldsymbol{m} + \boldsymbol{m}\times\circ\dot{\boldsymbol{W}} \quad \text{on } D_T\times\Omega. \tag{3.1}$$

Proposition 3.2
Let $\{\boldsymbol{m}(t);\, t\in[0,T]\}$ be a strong solution of equation (1.1). There holds for all $0<t\le T$ and $p\ge 2$:

a) $\mathbb{E}\Big[\sup\limits_{0\le s\le t}\|\partial_x\boldsymbol{m}(s)\|^{2p}_{\mathbf{L}^2} + \alpha\int_0^t\|\boldsymbol{m}(s)\times\partial_x^2\boldsymbol{m}(s)\|^2_{\mathbf{L}^2}\,\mathrm{d}s\Big] \le C(T,p,\operatorname{Tr}\boldsymbol{Q}^{\frac12},\boldsymbol{m}_0,D)$;

b) $\mathbb{E}\Big[\sup\limits_{0\le s\le t}\|\partial_x^2\boldsymbol{m}(s)\|^{2p}_{\mathbf{L}^2} + \alpha\int_0^t\|\partial_x^3\boldsymbol{m}(s)\|^2_{\mathbf{L}^2}\,\mathrm{d}s\Big] \le C(T,p,\operatorname{Tr}\boldsymbol{Q}^{\frac12},\boldsymbol{m}_0,D)$;

c) $\mathbb{E}\Big[\sup\limits_{0\le s\le t}\|\partial_x^3\boldsymbol{m}(s)\|^{2p}_{\mathbf{L}^2} + \alpha\int_0^t\|\partial_x^4\boldsymbol{m}(s)\|^2_{\mathbf{L}^2}\,\mathrm{d}s\Big] \le C(T,p,\operatorname{Tr}\boldsymbol{Q}^{\frac12},\boldsymbol{m}_0,D)$.

Proof
a) The first assertion follows by the application of Itô's formula for the functional $\boldsymbol{m}\mapsto\|\partial_x\boldsymbol{m}\|^2_{\mathbf{L}^2}$ applied to equation (1.1) and Burkholder-Davis-Gundy's inequality, see for instance [BGJ13, Theorem 3.5].

b) Fix $0\le r\le t\le T$. We apply Itô's formula for the functional $\boldsymbol{m}\mapsto\|\partial_x^2\boldsymbol{m}\|^2_{\mathbf{L}^2}$ to equation (3.1):

$$\begin{aligned}
&\|\partial_x^2\boldsymbol{m}(r)\|^2_{\mathbf{L}^2} - \|\partial_x^2\boldsymbol{m}(0)\|^2_{\mathbf{L}^2}\\
&= 2\alpha\int_0^r\Big(\partial_x^2[\partial_x^2\boldsymbol{m}(s)],\partial_x^2\boldsymbol{m}(s)\Big)\,\mathrm{d}s + 2\alpha\int_0^r\Big(\partial_x^2[|\partial_x\boldsymbol{m}(s)|^2_{\mathbb{R}^3}\boldsymbol{m}(s)],\partial_x^2\boldsymbol{m}(s)\Big)\,\mathrm{d}s\\
&\quad + 2\int_0^r\Big(\partial_x^2[\boldsymbol{m}(s)\times\partial_x^2\boldsymbol{m}(s)],\partial_x^2\boldsymbol{m}(s)\Big)\,\mathrm{d}s + \sum_{l=1}^\infty q_l\int_0^r\Big(\partial_x^2[(\boldsymbol{m}(s)\times\mathbf{e}_l)\times\mathbf{e}_l],\partial_x^2\boldsymbol{m}(s)\Big)\,\mathrm{d}s\\
&\quad + 2\sum_{l=1}^\infty\sqrt{q_l}\int_0^r\Big(\partial_x^2[\boldsymbol{m}(s)\times\mathbf{e}_l],\partial_x^2\boldsymbol{m}(s)\Big)\,\mathrm{d}\beta^l(s)\\
&\quad + 2\frac12\sum_{l=1}^\infty q_l\int_0^r\Big(\partial_x^2[\boldsymbol{m}(s)\times\mathbf{e}_l],\partial_x^2[\boldsymbol{m}(s)\times\mathbf{e}_l]\Big)\,\mathrm{d}s\\
&=: I + II + III + IV + 2\sum_{l=1}^\infty\sqrt{q_l}\int_0^r\Big(\partial_x^2[\boldsymbol{m}(s)\times\mathbf{e}_l],\partial_x^2\boldsymbol{m}(s)\Big)\,\mathrm{d}\beta^l(s) + V.
\end{aligned}$$

First, the terms I–V are simplified. For the first term I, by integration by parts (where we benefit from $\partial D=\emptyset$), we obtain

$$I = -2\alpha\int_0^r\|\partial_x^3\boldsymbol{m}(s)\|^2_{\mathbf{L}^2}\,\mathrm{d}s.$$

The second term II can be simplified using integration by parts and the triangle inequality:

$$2\alpha\Big|\Big(\partial_x^2[|\partial_x\boldsymbol{m}(s)|^2_{\mathbb{R}^3}\boldsymbol{m}(s)],\partial_x^2\boldsymbol{m}(s)\Big)\Big|$$

$$
\begin{aligned}
&= 2\alpha \Big| \Big(\partial_x [|\partial_x \boldsymbol{m}(s)|^2_{\mathbb{R}^3} \boldsymbol{m}(s)], \partial_x^3 \boldsymbol{m}(s) \Big) \Big| \\
&\leq 2\alpha \Big| \Big(|\partial_x \boldsymbol{m}(s)|^2_{\mathbb{R}^3} \partial_x \boldsymbol{m}(s) + 2\langle \partial_x \boldsymbol{m}(s), \partial_x^2 \boldsymbol{m}(s) \rangle_{\mathbb{R}^3} \boldsymbol{m}(s), \partial_x^3 \boldsymbol{m}(s) \Big) \Big| \\
&\leq 2\alpha \Big| \Big(|\partial_x \boldsymbol{m}(s)|^2_{\mathbb{R}^3} \partial_x \boldsymbol{m}(s), \partial_x^3 \boldsymbol{m}(s) \Big) \Big| + 4\alpha \Big| \Big(\langle \partial_x \boldsymbol{m}(s), \partial_x^2 \boldsymbol{m}(s) \rangle_{\mathbb{R}^3} \boldsymbol{m}(s), \partial_x^3 \boldsymbol{m}(s) \Big) \Big| \\
&=: II_A + II_B.
\end{aligned}
$$

For the first part II_A, we obtain using Young's and Gagliardo-Nirenberg's inequalities, i.e., $\|\partial_x \boldsymbol{m}(s)\|_{\mathbf{L}^\infty} \leq C \|\partial_x \boldsymbol{m}(s)\|^{\frac{1}{2}}_{\mathbf{L}^2} \|\partial_x \boldsymbol{m}(s)\|^{\frac{1}{2}}_{\mathbf{W}^{1,2}}$, which holds in 1d. Let $\delta > 0$, then

$$
\begin{aligned}
II_A &\leq \delta \|\partial_x^3 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^4_{\mathbf{L}^\infty} \|\partial_x \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} \\
&\leq \delta \|\partial_x^3 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^2_{\mathbf{W}^{1,2}} \|\partial_x \boldsymbol{m}(s)\|^4_{\mathbf{L}^2} \\
&\leq \delta \|\partial_x^3 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C \Big(\|\partial_x^2 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + \|\partial_x \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} \Big) \|\partial_x \boldsymbol{m}(s)\|^4_{\mathbf{L}^2}.
\end{aligned}
$$

We deduce $\|\partial_x^2 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} = |(\partial_x^2 \boldsymbol{m}(s), \partial_x^2 \boldsymbol{m}(s))| \leq \|\partial_x \boldsymbol{m}(s)\|_{\mathbf{L}^2} \|\partial_x^3 \boldsymbol{m}(s)\|_{\mathbf{L}^2}$ using integration by parts, such that

$$
\begin{aligned}
II_A &\leq \delta \|\partial_x^3 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C \|\partial_x^3 \boldsymbol{m}(s)\|_{\mathbf{L}^2} \|\partial_x \boldsymbol{m}(s)\|^5_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^6_{\mathbf{L}^2} \\
&\leq 2\delta \|\partial_x^3 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^{10}_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^6_{\mathbf{L}^2}.
\end{aligned}
$$

Similarly, we obtain for the second term II_B the estimate

$$
II_B \leq 3\delta \|\partial_x^3 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^{10}_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^6_{\mathbf{L}^2}.
$$

The third term III can be estimated using integration by parts, the identity $\langle \boldsymbol{a} \times \boldsymbol{b}, \boldsymbol{b} \rangle_{\mathbb{R}^3} = 0$, Young's inequality and $\|\partial_x \boldsymbol{m}(s) \times \partial_x^2 \boldsymbol{m}(s)\|_{\mathbf{L}^2} \leq C \|\partial_x \boldsymbol{m}(s)\|_{\mathbf{L}^\infty} \|\partial_x^2 \boldsymbol{m}(s)\|_{\mathbf{L}^2}$ by

$$
\Big| \Big(\partial_x^2 [\boldsymbol{m}(s) \times \partial_x^2 \boldsymbol{m}(s)], \partial_x^2 \boldsymbol{m}(s) \Big) \Big| \leq \delta \|\partial_x^3 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^2_{\mathbf{L}^\infty} \|\partial_x^2 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2}.
$$

Again by the estimate $\|\partial_x^2 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} \leq \|\partial_x \boldsymbol{m}(s)\|_{\mathbf{L}^2} \|\partial_x^3 \boldsymbol{m}(s)\|_{\mathbf{L}^2}$ as well as Young's and Gagliardo-Nirenberg's inequalities, we obtain

$$
\Big| \Big(\partial_x^2 [\boldsymbol{m}(s) \times \partial_x^2 \boldsymbol{m}(s)], \partial_x^2 \boldsymbol{m}(s) \Big) \Big| \leq 3\delta \|\partial_x^3 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^{10}_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^6_{\mathbf{L}^2}.
$$

For the fourth term IV, we arrive using $\mathbf{e}_l \in \mathbf{W}^{1,\infty}(D; \mathbb{R}^3)$ and similar arguments as in the steps before at

$$
\begin{aligned}
\Big| \Big(\partial_x^2 [(\boldsymbol{m}(s) \times \mathbf{e}_l) \times \mathbf{e}_l], \partial_x^2 \boldsymbol{m}(s) \Big) \Big| &= \Big| \Big(\partial_x [(\boldsymbol{m}(s) \times \mathbf{e}_l) \times \mathbf{e}_l], \partial_x^3 \boldsymbol{m}(s) \Big) \Big| \\
&\leq \delta \|\partial_x^3 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C.
\end{aligned}
$$

Similarly, we obtain

$$
\begin{aligned}
\|\partial_x^2 [\boldsymbol{m}(s) \times \mathbf{e}_l]\|^2_{\mathbf{L}^2} &= \|\partial_x^2 \boldsymbol{m}(s) \times \mathbf{e}_l + 2\partial_x \boldsymbol{m}(s) \times \partial_x \mathbf{e}_l + \boldsymbol{m}(s) \times \partial_x^2 \mathbf{e}_l\|^2_{\mathbf{L}^2} \\
&\leq C \|\partial_x^2 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C \|\partial_x \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} + C
\end{aligned}
$$

for the last term V using $\|\boldsymbol{u} \times \boldsymbol{v}\|_{\mathbf{L}^2} \leq C \|\boldsymbol{u}\|_{\mathbf{L}^2} \|\boldsymbol{v}\|_{\mathbf{L}^\infty}$ and $\mathbf{e}_l \in \mathbf{W}^{2,2}(D; \mathbb{R}^3)$. Thus, we arrive at

$$
\|\partial_x^2 \boldsymbol{m}(r)\|^2_{\mathbf{L}^2} - \|\partial_x^2 \boldsymbol{m}(0)\|^2_{\mathbf{L}^2} + 2\alpha \int_0^r \|\partial_x^3 \boldsymbol{m}(s)\|^2_{\mathbf{L}^2} \, \mathrm{d}s
$$

$$\leq C\int_0^r \|\partial_x \boldsymbol{m}(s)\|_{\mathbf{L}^2}^{10}\,\mathrm{d}s + C + C\int_0^r \|\partial_x^2 \boldsymbol{m}(s)\|_{\mathbf{L}^2}^2\,\mathrm{d}s + 2\sum_{l=1}^{\infty}\sqrt{q_l}\int_0^r \Big(\partial_x^2[\boldsymbol{m}(s)\times \mathbf{e}_l], \partial_x^2 \boldsymbol{m}(s)\Big)\,\mathrm{d}\beta^l(s).$$

Neglecting the third term on the left-hand-side, taking everything to the power $p \geq 2$, and then taking the supremum over $0 \leq r \leq t$ leads us to

$$\sup_{0\leq r\leq t}\|\partial_x^2 \boldsymbol{m}(r)\|_{\mathbf{L}^2}^{2p} \leq C\|\partial_x^2 \boldsymbol{m}(0)\|_{\mathbf{L}^2}^{2p} + C\int_0^t \|\partial_x \boldsymbol{m}(s)\|_{\mathbf{L}^2}^{10p}\,\mathrm{d}s + C\int_0^t \|\partial_x^2 \boldsymbol{m}(s)\|_{\mathbf{L}^2}^{2p}\,\mathrm{d}s + C + C(\operatorname{Tr}\boldsymbol{Q}^{\frac{1}{2}})^{p-1}\sum_{l=1}^{\infty}\sqrt{q_l}\sup_{0\leq r\leq t}\Big|\int_0^r \Big(\partial_x^2[\boldsymbol{m}(s)\times \mathbf{e}_l], \partial_x^2 \boldsymbol{m}(s)\Big)\,\mathrm{d}\beta^l(s)\Big|^p,$$

where we have used Young's and Hölder's inequalities. Applying expectations, we obtain

$$\mathbb{E}\Big[\sup_{0\leq r\leq t}\Big|\int_0^r \Big(\partial_x^2[\boldsymbol{m}(s)\times \mathbf{e}_l], \partial_x^2 \boldsymbol{m}(s)\Big)\,\mathrm{d}\beta^l(s)\Big|^p\Big] \leq C\mathbb{E}\Big[\Big|\int_0^t \Big|\Big(\partial_x^2[\boldsymbol{m}(s)\times \mathbf{e}_l], \partial_x^2 \boldsymbol{m}(s)\Big)\Big|^2\,\mathrm{d}s\Big|^{\frac{p}{2}}\Big]$$

for the last term using Burkholder-Davis-Gundy's inequality. Similarly as for term V above, by applying $\langle \boldsymbol{a}\times\boldsymbol{b}, \boldsymbol{b}\rangle_{\mathbb{R}^3} = 0$, we obtain the estimate

$$\Big|\Big(\partial_x^2[\boldsymbol{m}(s)\times \mathbf{e}_l], \partial_x^2 \boldsymbol{m}(s)\Big)\Big|^2 \leq C\|\partial_x^2 \boldsymbol{m}(s)\|_{\mathbf{L}^2}^4 + C\|\partial_x \boldsymbol{m}(s)\|_{\mathbf{L}^2}^4 + C.$$

Putting things together leads us to

$$\mathbb{E}\Big[\sup_{0\leq r\leq t}\|\partial_x^2 \boldsymbol{m}(r)\|_{\mathbf{L}^2}^{2p}\Big] \leq C\mathbb{E}\Big[\|\partial_x^2 \boldsymbol{m}(0)\|_{\mathbf{L}^2}^{2p}\Big] + C\int_0^t \mathbb{E}\Big[\sup_{0\leq r\leq s}\|\partial_x^2 \boldsymbol{m}(r)\|_{\mathbf{L}^2}^{2p}\Big]\,\mathrm{d}s + C \leq C,$$

after the use of Gronwall's inequality and Assumption A_3.

c) This assertion follows by Itô's formula for the functional $\boldsymbol{m} \mapsto \|\partial_x^3 \boldsymbol{m}\|_{\mathbf{L}^2}^2$ applied to equation (3.1). Here, the assumptions $\mathbf{e}_l \in \mathbf{W}^{3,2}(D;\mathbb{R}^3)$ and $\boldsymbol{m}_0 \in \mathbf{W}^{3,2}(D;\mathbb{R}^3)$ are needed. □

The following result on Hölder regularity in time of the strong solution is relevant to later conclude strong rates of convergence for iterates of the corresponding discretization.

Proposition 3.3
Let $\{\boldsymbol{m}(t);\, t\in[0,T]\}$ be a strong solution of equation (1.1). There holds for all $0 \leq s \leq t$ and $p \in \mathbb{N}$:

a) $\mathbb{E}\Big[\sup_{s\leq r\leq t}\|\boldsymbol{m}(r) - \boldsymbol{m}(s)\|_{\mathbf{L}^2}^{2p}\Big] \leq C(T, p, \operatorname{Tr}\boldsymbol{Q}^{\frac{1}{2}}, \boldsymbol{m}_0, D)|t-s|^p$;

b) $\mathbb{E}\Big[\sup_{s\leq r\leq t}\|\partial_x[\boldsymbol{m}(r) - \boldsymbol{m}(s)]\|_{\mathbf{L}^2}^{2p}\Big] \leq C(T, p, \operatorname{Tr}\boldsymbol{Q}^{\frac{1}{2}}, \boldsymbol{m}_0, D)|t-s|^p$;

c) $\mathbb{E}\Big[\sup_{s\leq r\leq t}\|\partial_x^2[\boldsymbol{m}(r) - \boldsymbol{m}(s)]\|_{\mathbf{L}^2}^{2}\Big] \leq C(T, p, \operatorname{Tr}\boldsymbol{Q}^{\frac{1}{2}}, \boldsymbol{m}_0, D)|t-s|$.

Proof
c) Fix $0 \le s \le t \le T$. By taking the norm $\|\partial_x^2 \cdot\|_{\mathbf{L}^2}^2$ on both sides of equation (1.1), taking supremum and expectation, we obtain

$$\begin{aligned}
\mathbb{E}\Big[\sup_{s\le r\le t}\|\partial_x^2[\boldsymbol{m}(t)-\boldsymbol{m}(s)]\|_{\mathbf{L}^2}^2\Big] &\le C\mathbb{E}\Big[\sup_{s\le r\le t}\Big\|\int_s^r \partial_x^2[\boldsymbol{m}(z)\times(\boldsymbol{m}(z)\times\Delta\boldsymbol{m}(z))]\,\mathrm{d}z\Big\|_{\mathbf{L}^2}^2\Big]\\
&\quad + C\mathbb{E}\Big[\sup_{s\le r\le t}\Big\|\int_s^r \partial_x^2[\boldsymbol{m}(z)\times\Delta\boldsymbol{m}(z)]\,\mathrm{d}z\Big\|_{\mathbf{L}^2}^2\Big]\\
&\quad + C\mathbb{E}\Big[\sup_{s\le r\le t}\Big\|\sum_{l=1}^\infty q_l\int_s^r \partial_x^2[(\boldsymbol{m}(z)\times\mathbf{e}_l)\times\mathbf{e}_l]\,\mathrm{d}z\Big\|_{\mathbf{L}^2}^2\Big]\\
&\quad + C\mathbb{E}\Big[\sup_{s\le r\le t}\Big\|\sum_{l=1}^\infty \sqrt{q_l}\int_s^r (\boldsymbol{m}(z)\times\mathbf{e}_l)\,\mathrm{d}\beta^l(z)\Big\|_{\mathbf{W}^{2,2}}^2\Big]\\
&=: I + II + III + IV. \qquad (3.2)
\end{aligned}$$

The first and the second term can be estimated in a similar way. For the sake of simplicity, we focus only on the first one, where we arrive at

$$\begin{aligned}
I &\le C|t-s|\mathbb{E}\Big[\int_s^t \|\partial_x^2[\boldsymbol{m}(z)\times(\boldsymbol{m}(z)\times\Delta\boldsymbol{m}(z))]\|_{\mathbf{L}^2}^2\,\mathrm{d}z\Big]\\
&\le C|t-s|\bigg(\int_s^t \mathbb{E}\Big[\|\partial_x^2\boldsymbol{m}(z)\|_{\mathbf{L}^\infty}^2\|\partial_x^2\boldsymbol{m}(z)\|_{\mathbf{L}^2}^2\|\boldsymbol{m}(z)\|_{\mathbf{L}^\infty}^2\Big]\,\mathrm{d}z\\
&\qquad + \int_s^t \mathbb{E}\Big[\|\partial_x^4\boldsymbol{m}(z)\|_{\mathbf{L}^2}^2\|\boldsymbol{m}(z)\|_{\mathbf{L}^\infty}^4\Big]\,\mathrm{d}z\\
&\qquad + \int_s^t \mathbb{E}\Big[\|\partial_x^3\boldsymbol{m}(z)\|_{\mathbf{L}^2}^2\|\partial_x^2\boldsymbol{m}(z)\|_{\mathbf{L}^\infty}^2\|\boldsymbol{m}(z)\|_{\mathbf{L}^\infty}^2\Big]\,\mathrm{d}z\bigg)\\
&\le C|t-s|\Big(C|t-s|+C\Big), \qquad (3.3)
\end{aligned}$$

by using Hölder's and Gagliardo-Nirenberg's inequality, the property $\|\boldsymbol{m}(z)\|_{\mathbf{L}^\infty}=1$ $\mathbb{P}$-a.s., and Proposition 3.2. Similarly, we obtain for the third term III the estimate

$$\begin{aligned}
III &\le C(\operatorname{Tr}\boldsymbol{Q})\mathbb{E}\Big[\sum_{l=1}^\infty q_l\sup_{s\le r\le t}\Big\|\int_s^r \partial_x^2[(\boldsymbol{m}(z)\times\mathbf{e}_l)\times\mathbf{e}_l]\,\mathrm{d}z\Big\|_{\mathbf{L}^2}^2\Big]\\
&\le C(\operatorname{Tr}\boldsymbol{Q})\sum_{l=1}^\infty q_l|t-s|\int_s^t \mathbb{E}\Big[\|\partial_x^2[(\boldsymbol{m}(z)\times\mathbf{e}_l)\times\mathbf{e}_l]\|_{\mathbf{L}^2}^2\Big]\,\mathrm{d}z\\
&\le C\big(\operatorname{Tr}\boldsymbol{Q}\big)^2|t-s|^2. \qquad (3.4)
\end{aligned}$$

By applying Burkholder-Davis-Gundy's inequality and Proposition 3.2, we arrive at

$$\begin{aligned}
IV &\le C\operatorname{Tr}\boldsymbol{Q}^{\frac12}\sum_{l=1}^\infty \sqrt{q_l}\mathbb{E}\Big[\sup_{s\le r\le t}\Big\|\int_s^r (\boldsymbol{m}(z)\times\mathbf{e}_l)\,\mathrm{d}\beta^l(z)\Big\|_{\mathbf{W}^{2,2}}^2\Big]\\
&\le C\operatorname{Tr}\boldsymbol{Q}^{\frac12}\sum_{l=1}^\infty \sqrt{q_l}\int_s^t \mathbb{E}\Big[\|\boldsymbol{m}(z)\times\mathbf{e}_l\|_{\mathbf{W}^{2,2}}^2\Big]\,\mathrm{d}z\\
&\le C\big(\operatorname{Tr}\boldsymbol{Q}^{\frac12}\big)^2|t-s|. \qquad (3.5)
\end{aligned}$$

Thus, we obtain the assertion by combining estimates (3.2)–(3.5).

a) This assertion can be shown similarly as in c) by using the norm $\|\cdot\|_{\mathbf{L}^2}^{2p}$ on both sides of equation (1.1).

b) This assertion can be shown similarly as in c) by using the norm $\|\partial_x \cdot\|_{\mathbf{L}^2}^{2p}$ on both sides of equation (1.1). □

3.2. Semi-discretization in time

In this section, we derive a-priori bounds for the iterates of the semi-discretized scheme given in Algorithm 1.1.

Lemma 3.4
Let Assumptions A_1–A_2 be fulfilled and the timestep size $k > 0$ be sufficiently small.

a) There exists an adapted sequence of $\mathbf{W}^{1,2}(D;\mathbb{R}^3)$-valued random variables, which satisfy $\mathbb{P}$-a.s. the scheme given in Algorithm 1.1.

b) The iterates of Algorithm 1.1 take values in $\mathbb{S}^2$, i.e., there holds for each $j \in \mathbb{N}$ $\mathbb{P}$-a.s. $\|\boldsymbol{m}^j\|_{\mathbf{L}^\infty} = 1$.

c) Let $\boldsymbol{m}^0 \in \mathbf{W}^{1,2}(D;\mathbb{S}^2)$. There holds

$$\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\Big(\|\nabla \boldsymbol{m}^{r+1}\|_{\mathbf{L}^2}^2 + \sum_{j=0}^{r}\|\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\|_{\mathbf{L}^2}^2 + k\sum_{j=0}^{r}\|\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2\Big)\Big] \leq C_T.$$

d) Let $\boldsymbol{m}^0 \in \mathbf{W}^{1,2}(D;\mathbb{S}^2)$. There holds for $q \in \mathbb{N}$

$$\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\Big(\|\nabla \boldsymbol{m}^{r+1}\|_{\mathbf{L}^2}^{2q} + \sum_{j=0}^{r}\Big|\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^{2q-1} - \|\nabla \boldsymbol{m}^{j}\|_{\mathbf{L}^2}^{2q-1}\Big|^2\Big)\Big] \leq C_{T,q}.$$

e) There holds $\mathbb{P}$-a.s.

$$\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}^4 \leq Ck^2(\alpha^2+1)\|\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 + C\|\Delta_j \boldsymbol{W}\|_{\mathbf{L}^2}^4.$$

Proof
a) This assertion can be shown similarly as in [DBD04] with the tools obtained in [BBNP14].

b) Multiplying the equation in Algorithm 1.1 for one fixed $\omega \in \Omega$ with $\boldsymbol{m}^{j+\frac{1}{2}}(\omega)$ in $\mathbb{R}^3$-sense, and using $\langle \boldsymbol{a}, \boldsymbol{a}\times\boldsymbol{b}\rangle_{\mathbb{R}^3} = 0$ leads to the equality

$$\langle \boldsymbol{m}^{j+1}-\boldsymbol{m}^j, \frac{1}{2}(\boldsymbol{m}^{j+1}+\boldsymbol{m}^j)\rangle_{\mathbb{R}^3} = 0.$$

Since $\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{a}+\boldsymbol{b}\rangle_{\mathbb{R}^3} = |\boldsymbol{a}|_{\mathbb{R}^3}^2 - |\boldsymbol{b}|_{\mathbb{R}^3}^2$ is valid, we obtain $|\boldsymbol{m}^{j+1}|_{\mathbb{R}^3}^2 = |\boldsymbol{m}^j|_{\mathbb{R}^3}^2$ $\mathbb{P}$-a.s., which concludes the proof.

c) The first assertion follows analogously to the proof of Lemma 4.1 in [BBNP14] by multiplying the equation in Algorithm 1.1 for one $\omega \in \Omega$ with $-\Delta \boldsymbol{m}^{j+1}(\omega)$ and afterwards integrating in space. There, we obtain $\mathbb{P}$-a.s.

$$\begin{aligned}
&\frac{1}{2}\Big(\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 - \|\nabla \boldsymbol{m}^{j}\|_{\mathbf{L}^2}^2\Big) \\
&\leq \Big(\nabla[\boldsymbol{m}^j \times \Delta_j \boldsymbol{W}], \nabla \boldsymbol{m}^j\Big) \\
&\quad + \Big(C\|\nabla \boldsymbol{m}^j\|_{\mathbf{L}^2}^2 + C\Big)\Big(\|\Delta_j \boldsymbol{W}\|_{\mathbf{L}^2}^2 + \|\Delta_j \boldsymbol{W}\|_{\mathbf{L}^\infty}^2 + \|\nabla[\Delta_j \boldsymbol{W}]\|_{\mathbf{L}^2}^2\Big),
\end{aligned} \tag{3.6}$$

which is the starting point for assertion d). After summation, taking supremum over iterates $j = 0, \dots, J-1$ and taking expectations, the first term on the right-hand-side can be estimated using a Burkholder-Davis-Gundy argument, for the second term we use Remark 2.2, and finally assertion c) follows by the discrete version of Gronwall's inequality.

d) In order to show the second assertion, we use an inductive argument: To obtain the result for $q = 2$, we multiply inequality (3.6) by $\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2$. By using the identity $(a-b)a = \frac{1}{2}(a^2 - b^2) + \frac{1}{2}(a-b)^2$, Young's inequality, estimate $\|\boldsymbol{u} \times \boldsymbol{v}\|_{\mathbf{L}^2} \leq C\|\boldsymbol{u}\|_{\mathbf{L}^2}\|\boldsymbol{v}\|_{\mathbf{L}^\infty}$, and $\|\boldsymbol{m}^j\|_{\mathbf{L}^\infty} = 1$ $\mathbb{P}$-a.s., we arrive at

$$\begin{aligned}
&\frac{1}{4}\Big(\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4 - \|\nabla \boldsymbol{m}^{j}\|_{\mathbf{L}^2}^4 + \Big|\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 - \|\nabla \boldsymbol{m}^{j}\|_{\mathbf{L}^2}^2\Big|^2\Big) \\
&\leq \Big(\nabla[\boldsymbol{m}^j \times \Delta_j \boldsymbol{W}], \nabla \boldsymbol{m}^j\Big)\|\nabla \boldsymbol{m}^j\|_{\mathbf{L}^2}^2 + C\Big(\|\nabla \boldsymbol{m}^j\|_{\mathbf{L}^2}^4 + C\Big) \\
&\quad \times \Big(\|\Delta_j \boldsymbol{W}\|_{\mathbf{L}^2}^4 + \|\Delta_j \boldsymbol{W}\|_{\mathbf{L}^\infty}^4 + \|\nabla[\Delta_j \boldsymbol{W}]\|_{\mathbf{L}^2}^4 + \|\Delta_j \boldsymbol{W}\|_{\mathbf{L}^2}^2 + \|\Delta_j \boldsymbol{W}\|_{\mathbf{L}^\infty}^2 + \|\nabla[\Delta_j \boldsymbol{W}]\|_{\mathbf{L}^2}^2\Big).
\end{aligned} \tag{3.7}$$

After summation, taking supremum over iterates $j = 0, \dots, J-1$ and taking expectations, the first term on the left-hand-side of inequality (3.7) can be estimated using Burkholder-Davis-Gundy's inequality and the other terms by using independence and Remark 2.2.
To obtain the result for $q = 3$, we multiply inequality (3.7) by $\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4$ and obtain it as above. By repeating this procedure, we obtain the result for each $q \in \mathbb{N}$.

e) Multiplication of the equation in Algorithm 1.1 for one $\omega \in \Omega$ with $\boldsymbol{m}^{j+1}(\omega) - \boldsymbol{m}^j(\omega)$ then yields that $\mathbb{P}$-a.s.

$$\|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{L}^2}^2 \leq Ck^2(\alpha^2+1)\|\boldsymbol{m}^{j+\frac{1}{2}} \times \Delta \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 + C\|\Delta_j \boldsymbol{W}\|_{\mathbf{L}^2}^2. \tag{3.8}$$

Multiplication of both sides of inequality (3.8) by $\|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{L}^2}^2$ and using the $\mathbb{P}$-a.s. estimate $\|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{L}^2}^2 \leq C$ yields the assertion. □

4. Error estimate for the semi-discretization in time

In this chapter we state the main theorem and prove rates of convergence in probability for the iterates $\{\boldsymbol{m}^j;\, j=0,\dots,J\}$ from Algorithm 1.1 and the strong solution $\{\boldsymbol{m}(t);\, t\in[0,T]\}$ of equation (1.1).

Theorem 4.1
Suppose that Assumptions A_1–A_3 hold. Let $\{\boldsymbol{m}(t);\, t\in[0,T]\}$ be the strong solution of (1.1) and $\{\boldsymbol{m}^j;\, j=0,\dots,J\}$ be the iterates of Algorithm 1.1. For every $\epsilon>0$, the set

$$\widetilde{\Omega}_k := \Big\{\omega\in\Omega : \sup_{j=0,\dots,J-1}\|\nabla \boldsymbol{m}^j\|_{\mathbf{L}^2}^4 + \sup_{s\in[0,t_{J-1}]}\|\nabla \boldsymbol{m}(s)\|_{\mathbf{L}^2}^4 + \sup_{s\in[0,t_{J-1}]}\|\nabla \boldsymbol{m}(s)\|_{\mathbf{W}^{1,2}}^2 \le \log(k^{-\frac{\epsilon}{2}})\Big\}$$

satisfies

$$\mathbb{P}[\widetilde{\Omega}_k] \ge 1 + \frac{C}{\epsilon\log(k)},$$

and there holds the local error estimate

$$\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_k}\sup_{j=0,\dots,J-1}\|\boldsymbol{m}(t_{j+1})-\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2\Big] \le Ck^{1-\epsilon}.$$

By Theorem 4.1 we obtain strong convergence for the semi-discretized scheme with rate $0<\nu<\frac{1}{2}(1-\epsilon)$ for every $\epsilon>0$ on arbitrary large subsets of the probability space. This implies convergence in probability [Pri01, Definition 2.7] with rate ν.

Corollary 4.2
Let the assumptions of Theorem 4.1 be fulfilled. The iterates $\{\boldsymbol{m}^j;\, j=0,\dots,J\}$ of the semi-discretized scheme given in Algorithm 1.1 converge in probability with order $0<\nu<\frac{1}{2}(1-\epsilon)$ for every $\epsilon>0$, i.e., there exists a constant $\widetilde{C}>0$, such that

$$\lim_{k\to 0}\mathbb{P}\Big[\sup_{j=0,\dots,J-1}\|\boldsymbol{m}(t_{j+1})-\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}\ge \widetilde{C}k^{\nu}\Big] = 0.$$

Proof
Due to Theorem 4.1, there exists a subset $\widetilde{\Omega}_k\subset\Omega$, such that, by using Chebychev's inequality, the following holds:

$$\mathbb{P}\Big[\sup_{j=0,\dots,J-1}\|\boldsymbol{m}(t_{j+1})-\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}\ge \widetilde{C}k^{\nu}\Big]$$

$$
\begin{aligned}
&\leq \mathbb{P}\Big[\Big\{\sup_{j=0,\dots,J-1}\|\boldsymbol{m}(t_{j+1})-\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}\geq \widetilde{C}k^{\nu}\Big\}\cap\widetilde{\Omega}_k\Big]+\mathbb{P}[\Omega\setminus\widetilde{\Omega}_k]\\
&\leq \frac{\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_k}\sup\limits_{j=0,\dots,J-1}\|\boldsymbol{m}(t_{j+1})-\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2\Big]}{\widetilde{C}^2k^{2\nu}}+\Big(1-\mathbb{P}[\widetilde{\Omega}_k]\Big)\\
&\leq \frac{Ck^{1-\epsilon}}{\widetilde{C}^2k^{2\nu}}-\frac{C}{\epsilon\log(k)}.
\end{aligned}
$$

Since $2\nu<1-\epsilon$, we obtain

$$
\lim_{k\to 0}\mathbb{P}\Big[\sup_{j=0,\dots,J-1}\|\boldsymbol{m}(t_{j+1})-\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}\geq\widetilde{C}k^{\nu}\Big]=0.
$$

Proof (Proof of Theorem 4.1)
The proof is based on the reformulation (1.4) of the equation in Algorithm 1.1. Hölder regularity statements (Proposition 3.3) for the strong solution are exploited to pathwisely control temporal discretization errors. Nonlinear effects are accounted for by regularity results (Proposition 3.2 and Lemma 3.4). Due to the occurrence of mixed terms a subset argument is needed to apply Gronwall's inequality.

The error below is defined by $\boldsymbol{z}^j:=\boldsymbol{m}(t_j)-\boldsymbol{m}^j\ \forall j=0,\dots,J$. We subtract equation (3.1) and the equation in Algorithm 1.1 where the damping term is reformulated according to (1.4), and test with $\boldsymbol{z}^{j+1}$.

Since a usual application of Gronwall's inequality is not possible on the whole probability space Ω (see therefore inequality (4.25) below), we restrict the error analysis to the subset $\widetilde{\Omega}_{\kappa,j}\subset\Omega$, defined for a fixed $\kappa>0$ by

$$
\widetilde{\Omega}_{\kappa,j}:=\Big\{\omega\in\Omega:\sup_{r=0,\dots,j}\|\nabla\boldsymbol{m}^r\|_{\mathbf{L}^2}^4+\sup_{t_0\leq s\leq t_j}\|\nabla\boldsymbol{m}(s)\|_{\mathbf{L}^2}^4+\sup_{t_0\leq s\leq t_j}\|\nabla\boldsymbol{m}(s)\|_{\mathbf{W}^{1,2}}^2\leq\kappa\Big\}
$$

for all $j=0,\dots,J-1$. There holds $\widetilde{\Omega}_{\kappa,j+1}\subset\widetilde{\Omega}_{\kappa,j}$. We arrive at

$$
\begin{aligned}
&\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(\boldsymbol{z}^{j+1}-\boldsymbol{z}^j,\boldsymbol{z}^{j+1}\Big)-\alpha\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big(\Delta\boldsymbol{m}(s)-\Delta\boldsymbol{m}^{j+1},\boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s\\
&\leq\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big|\alpha k\Big([|\boldsymbol{m}^{j+\frac12}|_{\mathbb{R}^3}^2-1]\Delta\boldsymbol{m}^{j+1},\boldsymbol{z}^{j+1}\Big)\Big|\\
&\quad+\alpha\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big(|\nabla\boldsymbol{m}(s)|_{\mathbb{R}^3}^2\boldsymbol{m}(s)-|\nabla\boldsymbol{m}^{j+1}|_{\mathbb{R}^3}^2\boldsymbol{m}^{j+1},\boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s\\
&\quad+\frac{\alpha k}{2}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(|\nabla\boldsymbol{m}^{j+1}|_{\mathbb{R}^3}^2(\boldsymbol{m}^{j+1}-\boldsymbol{m}^j),\boldsymbol{z}^{j+1}\Big)\\
&\quad+\frac{\alpha k}{2}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big|\Big(\langle\boldsymbol{m}^{j+1}-\boldsymbol{m}^j,\Delta\boldsymbol{m}^{j+1}\rangle_{\mathbb{R}^3}\boldsymbol{m}^{j+\frac12},\boldsymbol{z}^{j+1}\Big)\Big|\\
&\quad+\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big(\boldsymbol{m}(s)\times\Delta\boldsymbol{m}(s)-\boldsymbol{m}^{j+1}\times\Delta\boldsymbol{m}^{j+1},\boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s\\
&\quad+\frac{k}{2}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big((\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\times\Delta\boldsymbol{m}^{j+1},\boldsymbol{z}^{j+1}\Big)\\
&\quad+\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(\int_{t_j}^{t_{j+1}}\boldsymbol{m}(s)\times\circ\,\mathrm{d}\boldsymbol{W}(s)-\boldsymbol{m}^{j+\frac12}\times\Delta_j\boldsymbol{W},\boldsymbol{z}^{j+1}\Big)
\end{aligned}
$$

$$=: I + II + III + IV + V + VI + VII. \tag{4.1}$$

The left-hand-side of equation (4.1) consists of seven terms, which are considered separately in the following: The first four terms I–IV correspond to the damping term, the terms V–VI to the precessional term, and the last one VII attributes error effects of the stochastic integral term.

During the whole proof, small constants $\delta, \widetilde{\delta} > 0$ are used for terms which are in the end absorbed to the left-hand-side and inequalities are understood in the $\mathbb{P}$-a.s. sense.
First Step: *Estimate for the left-hand-side and terms I–VI.*
LHS (first term): Using the identity $2\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{a}\rangle_{\mathbb{R}^3} = |\boldsymbol{a}|^2_{\mathbb{R}^3} - |\boldsymbol{b}|^2_{\mathbb{R}^3} + |\boldsymbol{a}-\boldsymbol{b}|^2_{\mathbb{R}^3}$, we obtain

$$\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(\boldsymbol{z}^{j+1}-\boldsymbol{z}^j, \boldsymbol{z}^{j+1}\Big) = \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\frac{1}{2}\Big(\|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2} - \|\boldsymbol{z}^j\|^2_{\mathbf{L}^2} + \|\boldsymbol{z}^{j+1}-\boldsymbol{z}^j\|^2_{\mathbf{L}^2}\Big).$$

LHS (second term): This term can be rewritten using integration by parts as

$$\begin{aligned}
&- \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big(\Delta[\boldsymbol{m}(s)-\boldsymbol{m}^{j+1}], \boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s\\
&= \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big(\nabla[\boldsymbol{m}(s)-\boldsymbol{m}^{j+1}], \nabla\boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s\\
&= \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\|\nabla\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2} + \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big(\nabla[\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})], \nabla\boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s.
\end{aligned}$$

The second part of this equation is bounded by Young's inequality, i.e.,

$$\begin{aligned}
&\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big|-\Big(\nabla[\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})], \nabla\boldsymbol{z}^{j+1}\Big)\Big|\,\mathrm{d}s\\
&\le C\int_{t_j}^{t_{j+1}}\|\nabla[\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})]\|^2_{\mathbf{L}^2}\,\mathrm{d}s + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\|\nabla\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2}.
\end{aligned} \tag{4.2}$$

RHS (first term I): We decompose the first term I using integration by parts into

$$\begin{aligned}
I &\le \alpha\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\Big|\Big(\nabla\boldsymbol{m}^{j+1}, [|\boldsymbol{m}^{j+\frac{1}{2}}|^2_{\mathbb{R}^3}-1]\nabla\boldsymbol{z}^{j+1}\Big)\Big| + \alpha\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\Big|\Big(\nabla\boldsymbol{m}^{j+1}, \boldsymbol{z}^{j+1}\nabla[|\boldsymbol{m}^{j+\frac{1}{2}}|^2_{\mathbb{R}^3}-1]\Big)\Big|\\
&=: I_A + I_B.
\end{aligned}$$

For the first term I_A, we use the parallelogram identity together with the fact, that iterates $\boldsymbol{m}^j$ are $\mathbb{S}^2$-valued, in order to conclude

$$\begin{aligned}
|\boldsymbol{m}^{j+\frac{1}{2}}|^2_{\mathbb{R}^3} - 1 &= |\frac{1}{2}\boldsymbol{m}^{j+1} + \frac{1}{2}\boldsymbol{m}^j|^2_{\mathbb{R}^3} - 2\Big(|\frac{1}{2}\boldsymbol{m}^{j+1}|^2_{\mathbb{R}^3} + |\frac{1}{2}\boldsymbol{m}^j|^2_{\mathbb{R}^3}\Big)\\
&= -|\frac{1}{2}\boldsymbol{m}^{j+1} - \frac{1}{2}\boldsymbol{m}^j|^2_{\mathbb{R}^3}.
\end{aligned} \tag{4.3}$$

This yields $\||\boldsymbol{m}^{j+\frac{1}{2}}|^2_{\mathbb{R}^3} - 1\|_{\mathbf{L}^\infty} \le C\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|^2_{\mathbf{L}^\infty} \le C\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^\infty}(\|\boldsymbol{m}^{j+1}\|_{\mathbf{L}^\infty} + \|\boldsymbol{m}^j\|_{\mathbf{L}^\infty})$, and thus by using $\|\boldsymbol{m}^j\|_{\mathbf{L}^\infty} \le 1$ $\mathbb{P}$-a.s., we obtain

$$\begin{aligned}
I_A &= C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\Big|\Big(\nabla\boldsymbol{m}^{j+1}, [|\boldsymbol{m}^{j+\frac{1}{2}}|^2_{\mathbb{R}^3}-1]\nabla\boldsymbol{z}^{j+1}\Big)\Big|\\
&\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^\infty}\|\nabla\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}.
\end{aligned}$$

By Young's and Gagliardo-Nirenberg's inequalities, i.e., by $\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^\infty} \le C\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}^{\frac{1}{2}}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}^{\frac{1}{2}}$ in 1d, we obtain

$$I_A \le Ck\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2} + \delta \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2. \tag{4.4}$$

Using again identity (4.3), and Gagliardo-Nirenberg's and Young's inequalities, we obtain

$$\begin{aligned}
I_B &= \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\frac{1}{4}\Big|\Big(\nabla \boldsymbol{m}^{j+1}, \boldsymbol{z}^{j+1}\nabla[|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j|_{\mathbb{R}^3}^2]\Big)\Big| \\
&= \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\frac{1}{2}\Big|\Big(\nabla \boldsymbol{m}^{j+1}, \boldsymbol{z}^{j+1}\langle\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j], \boldsymbol{m}^{j+1}-\boldsymbol{m}^j\rangle_{\mathbb{R}^3}\Big)\Big| \\
&\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^\infty}\|\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\|_{\mathbf{L}^2}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^\infty} \\
&\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^{\frac{1}{2}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^{\frac{1}{2}}\|\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\|_{\mathbf{L}^2}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}^{\frac{1}{2}} \\
&\quad \times \|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}^{\frac{1}{2}} \\
&\le \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2 + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\
&\quad + Ck\|\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\|_{\mathbf{L}^2}^2\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}.
\end{aligned} \tag{4.5}$$

RHS (second term II): We obtain

$$\begin{aligned}
II &\le \alpha\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big(|\nabla \boldsymbol{m}(s)|_{\mathbb{R}^3}^2(\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})), \boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s \\
&\quad + \alpha\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big(|\nabla \boldsymbol{m}(t_{j+1})|_{\mathbb{R}^3}^2(\boldsymbol{m}(t_{j+1})-\boldsymbol{m}^{j+1}), \boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s \\
&\quad + \alpha\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big((|\nabla \boldsymbol{m}(s)|_{\mathbb{R}^3}^2-|\nabla \boldsymbol{m}(t_{j+1})|_{\mathbb{R}^3}^2)\boldsymbol{m}(t_{j+1}), \boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s \\
&\quad + \alpha\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big((|\nabla \boldsymbol{m}(t_{j+1})|_{\mathbb{R}^3}^2-|\nabla \boldsymbol{m}^{j+1}|_{\mathbb{R}^3}^2)\boldsymbol{m}^{j+1}, \boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s \\
&=: II_A + II_B + II_C + II_D.
\end{aligned}$$

The first term II_A can be simplified to

$$\begin{aligned}
II_A &\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\|\nabla \boldsymbol{m}(s)\|_{\mathbf{L}^\infty}^2\|\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}\,\mathrm{d}s \\
&\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\|\nabla \boldsymbol{m}(s)\|_{\mathbf{L}^2}\|\nabla \boldsymbol{m}(s)\|_{\mathbf{W}^{1,2}}\|\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}\,\mathrm{d}s,
\end{aligned}$$

using Gagliardo-Nirenberg's inequality, i.e., $\|\boldsymbol{u}\|_{\mathbf{L}^\infty} \le C\|\boldsymbol{u}\|_{\mathbf{L}^2}^{\frac{1}{2}}\|\boldsymbol{u}\|_{\mathbf{W}^{1,2}}^{\frac{1}{2}}$ in 1d. By Young's and triangle inequalities, we further conclude

$$\begin{aligned}
II_A &\le Ck\sup_{t_j\le s\le t_{j+1}}\|\nabla \boldsymbol{m}(s)\|_{\mathbf{L}^2}^2\|\nabla \boldsymbol{m}(s)\|_{\mathbf{W}^{1,2}}^2\sup_{t_j\le s\le t_{j+1}}\|\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2 \\
&\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2.
\end{aligned} \tag{4.6}$$

For the second term II_B, there holds

$$II_B \le \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^\infty}^2\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\,\mathrm{d}s$$

$$\leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}\|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{W}^{1,2}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2$$
$$\leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{W}^{1,2}}^2\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \tag{4.7}$$

using Gagliardo-Nirenberg's and Young's inequalities. For the terms II_C and II_D the identity

$$|\boldsymbol{a}|_{\mathbb{R}^3}^2 - |\boldsymbol{b}|_{\mathbb{R}^3}^2 = (|\boldsymbol{a}|_{\mathbb{R}^3} + |\boldsymbol{b}|_{\mathbb{R}^3})(|\boldsymbol{a}|_{\mathbb{R}^3} - |\boldsymbol{b}|_{\mathbb{R}^3}) \leq (|\boldsymbol{a}|_{\mathbb{R}^3} + |\boldsymbol{b}|_{\mathbb{R}^3})|\boldsymbol{a} - \boldsymbol{b}|_{\mathbb{R}^3} \tag{4.8}$$

is used. For the first term $II_{C,1}$ corresponding to the $|\boldsymbol{a}|_{\mathbb{R}^3}(|\boldsymbol{a}|_{\mathbb{R}^3} - |\boldsymbol{b}|_{\mathbb{R}^3})$ part in equation (4.8), we obtain using $\|\boldsymbol{m}^{j+1}\|_{\mathbf{L}^\infty} = 1$ $\mathbb{P}$-a.s., Gagliardo-Nirenberg's, and Young's inequalities

$$\begin{aligned} II_{C,1} &\leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \int_{t_j}^{t_{j+1}} \|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}\|\nabla[\boldsymbol{m}(t_{j+1}) - \boldsymbol{m}(s)]\|_{\mathbf{L}^2}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^{\frac{1}{2}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^{\frac{1}{2}}\,\mathrm{d}s \\ &\leq Ck \sup_{t_j\leq s\leq t_{j+1}} \|\nabla[\boldsymbol{m}(t_{j+1}) - \boldsymbol{m}(s)]\|_{\mathbf{L}^2}^2 + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^4\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\quad + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2. \end{aligned} \tag{4.9}$$

Using the same arguments for the second term $II_{C,2}$, we get

$$\begin{aligned} II_{C,2} &\leq Ck \sup_{t_j\leq s\leq t_{j+1}} \|\nabla[\boldsymbol{m}(t_{j+1}) - \boldsymbol{m}(s)]\|_{\mathbf{L}^2}^2 + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\quad + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2. \end{aligned} \tag{4.10}$$

Again, by using identity (4.8), Gagliardo-Nirenberg's and Young's inequalities as well as $\|\boldsymbol{m}^{j+1}\|_{\mathbf{L}^\infty} = 1$ $\mathbb{P}$-a.s., we obtain for the first part of II_D

$$\begin{aligned} II_{D,1} &\leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^4\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\quad + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2, \end{aligned} \tag{4.11}$$

and, similarly

$$II_{D,2} \leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2. \tag{4.12}$$

The term II requires to later restrict to the subset $\widetilde{\Omega}_{\kappa,j}$, see estimates (4.7), (4.11) and (4.12) since we obtain mixed terms, e.g. $\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2$, which prevents a direct use of a discrete version of Gronwall's inequality.

RHS (third term III): We use again Gagliardo-Nirenberg's and Young's inequalities to conclude

$$\begin{aligned} III &\leq k\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2\|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{L}^\infty}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^\infty} \\ &\leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2\|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{L}^2}^{\frac{1}{2}}\|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}^{\frac{1}{2}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^{\frac{1}{2}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^{\frac{1}{2}} \\ &\leq Ck\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4\|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}\|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{L}^2} \\ &\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2. \end{aligned} \tag{4.13}$$

RHS (fourth term IV): For the fourth term IV, integration by parts, the product formula, and the identities $\langle\langle \boldsymbol{a},\boldsymbol{b}\rangle_{\mathbb{R}^3}\boldsymbol{c},\boldsymbol{d}\rangle_{\mathbb{R}^3} = \langle\langle \boldsymbol{c},\boldsymbol{d}\rangle_{\mathbb{R}^3}\boldsymbol{a},\boldsymbol{b}\rangle_{\mathbb{R}^3}$ and $\boldsymbol{m}^{j+\frac{1}{2}} = \boldsymbol{m}^{j+1} - \frac{1}{2}(\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)$ are used to obtain

$$\begin{aligned}
IV &\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\Big|\Big(\nabla[(\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\langle \boldsymbol{m}^{j+1}-\boldsymbol{m}^j,\boldsymbol{z}^{j+1}\rangle_{\mathbb{R}^3}],\nabla\boldsymbol{m}^{j+1}\Big)\Big| \\
&\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\Big|\Big(\nabla[(\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\langle \boldsymbol{m}^{j+1},\boldsymbol{z}^{j+1}\rangle_{\mathbb{R}^3}],\nabla\boldsymbol{m}^{j+1}\Big)\Big| \\
&\le \Big(C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\Big|\Big(\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\langle \boldsymbol{m}^{j+1}-\boldsymbol{m}^j,\boldsymbol{z}^{j+1}\rangle_{\mathbb{R}^3},\nabla\boldsymbol{m}^{j+1}\Big)\Big| \\
&\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\Big|\Big((\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\langle \nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j],\boldsymbol{z}^{j+1}\rangle_{\mathbb{R}^3},\nabla\boldsymbol{m}^{j+1}\Big)\Big|\Big) \\
&\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\Big|\Big((\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\langle \boldsymbol{m}^{j+1}-\boldsymbol{m}^j,\nabla\boldsymbol{z}^{j+1}\rangle_{\mathbb{R}^3},\nabla\boldsymbol{m}^{j+1}\Big)\Big| \\
&\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\Big|\Big(\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\langle \boldsymbol{m}^{j+1},\boldsymbol{z}^{j+1}\rangle_{\mathbb{R}^3},\nabla\boldsymbol{m}^{j+1}\Big)\Big| \\
&\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\Big|\Big((\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\langle \nabla\boldsymbol{m}^{j+1},\boldsymbol{z}^{j+1}\rangle_{\mathbb{R}^3},\nabla\boldsymbol{m}^{j+1}\Big)\Big| \\
&\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\Big|\Big((\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\langle \boldsymbol{m}^{j+1},\nabla\boldsymbol{z}^{j+1}\rangle_{\mathbb{R}^3},\nabla\boldsymbol{m}^{j+1}\Big)\Big| \\
&=: IV_A + IV_B + IV_C + IV_D + IV_E,
\end{aligned}$$

where IV_A represents the first two summands. Those first two summands contained in IV_A can be simplified using Gagliardo-Nirenberg's inequality to

$$\begin{aligned}
IV_A &\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\|_{\mathbf{L}^2}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^\infty}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^\infty}\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2} \\
&\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\|_{\mathbf{L}^2}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}^{\frac{1}{2}}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}^{\frac{1}{2}} \\
&\quad \times \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^{\frac{1}{2}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^{\frac{1}{2}}\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}.
\end{aligned}$$

By Young's inequality, we arrive at

$$\begin{aligned}
IV_A &\le \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2 \\
&\quad + Ck\sup_{r=0,\dots,j+1}\|\nabla\boldsymbol{m}^r\|_{\mathbf{L}^2}^2\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{W}^{1,2}} \\
&\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2.
\end{aligned} \tag{4.14}$$

For the term IV_B, we obtain using Gagliardo-Nirenberg's and Young's inequalities

$$\begin{aligned}
IV_B &\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}^2\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}^2\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 \\
&\quad + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2.
\end{aligned} \tag{4.15}$$

For the term IV_C, Gagliardo-Nirenberg's and Young's inequalities again yield

$$\begin{aligned}
IV_C &\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\|_{\mathbf{L}^2}\|\boldsymbol{m}^{j+1}\|_{\mathbf{L}^\infty}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^\infty}\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2} \\
&\le Ck\|\nabla[\boldsymbol{m}^{j+1}-\boldsymbol{m}^j]\|_{\mathbf{L}^2}^2 + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\
&\quad + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2.
\end{aligned} \tag{4.16}$$

Similarly, we obtain for the term IV_D

$$\begin{aligned} IV_D &\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^\infty}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^\infty}\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\le Ck\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2, \end{aligned} \tag{4.17}$$

and for IV_E,

$$\begin{aligned} IV_E &\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^\infty}\|\boldsymbol{m}^{j+1}\|_{\mathbf{L}^\infty}\|\nabla\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2} \\ &\le Ck\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}\|\nabla\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2. \end{aligned} \tag{4.18}$$

RHS (fifth term V): In order to simplify the fifth term V, we consider

$$\begin{aligned} &\boldsymbol{m}(s)\times\Delta\boldsymbol{m}(s)-\boldsymbol{m}^{j+1}\times\Delta\boldsymbol{m}^{j+1} \\ &= (\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1}))\times\Delta\boldsymbol{m}(s)+\boldsymbol{m}(t_{j+1})\times\Delta[\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})] \\ &\quad + (\boldsymbol{m}(t_{j+1})-\boldsymbol{m}^{j+1})\times\Delta\boldsymbol{m}(t_{j+1})+\boldsymbol{m}^{j+1}\times\Delta[\boldsymbol{m}(t_{j+1})-\boldsymbol{m}^{j+1}], \end{aligned}$$

and, due to $\langle\boldsymbol{a}\times\boldsymbol{b},\boldsymbol{a}\rangle_{\mathbb{R}^3}=0$, we are able to rewrite the term V in the form

$$\begin{aligned} V &= \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big((\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1}))\times\Delta\boldsymbol{m}(s),\boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s \\ &\quad + \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big(\boldsymbol{m}(t_{j+1})\times\Delta[\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})],\boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s \\ &\quad + \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big(\boldsymbol{m}^{j+1}\times\Delta\boldsymbol{z}^{j+1},\boldsymbol{z}^{j+1}\Big)\,\mathrm{d}s \\ &=: V_A+V_B+V_C. \end{aligned}$$

Each term is now considered separately. Using Gagliardo-Nirenberg's and Young's inequalities, we obtain for the first term V_A

$$\begin{aligned} V_A &\le \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\|\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}\|\Delta\boldsymbol{m}(s)\|_{\mathbf{L}^2}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^\infty}\,\mathrm{d}s \\ &\le Ck\sup_{t_j\le s\le t_{j+1}}\|\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2\sup_{t_j\le s\le t_{j+1}}\|\Delta\boldsymbol{m}(s)\|_{\mathbf{L}^2}^2 + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\quad + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2. \end{aligned} \tag{4.19}$$

The second term V_B can be simplified using $\langle\boldsymbol{b},\boldsymbol{c}\times\boldsymbol{a}\rangle_{\mathbb{R}^3}=\langle\boldsymbol{a}\times\boldsymbol{b},\boldsymbol{c}\rangle_{\mathbb{R}^3}$ and integration by parts to

$$\begin{aligned} V_B &\le \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}\Big|\Big(\nabla[\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})],\nabla\boldsymbol{z}^{j+1}\times\boldsymbol{m}(t_{j+1})+\boldsymbol{z}^{j+1}\times\nabla\boldsymbol{m}(t_{j+1})\Big)\Big|\,\mathrm{d}s \\ &=: V_{B,1}+V_{B,2}. \end{aligned}$$

First, Gagliardo-Nirenberg's and Young's inequalities and $\|\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^\infty}=1$ $\mathbb{P}$-a.s. yield

$$V_{B,1} \le Ck\sup_{t_j\le s\le t_{j+1}}\|\nabla[\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})]\|_{\mathbf{L}^2}^2 + \delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k\|\nabla\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2. \tag{4.20}$$

Similarly, we obtain

$$V_{B,2} \leq Ck \sup_{t_j \leq s \leq t_{j+1}} \|\nabla[\boldsymbol{m}(s) - \boldsymbol{m}(t_{j+1})]\|_{\mathbf{L}^2}^2 + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^4 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + \delta \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2. \tag{4.21}$$

The third term V_C simplifies using integration by parts, Gagliardo-Nirenberg's and Young's inequality to

$$\begin{aligned} V_C &\leq \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2} \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^\infty} \\ &\leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + \delta \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + \delta \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2. \end{aligned} \tag{4.22}$$

The part V_C of term V also makes again a restriction of the error analysis on the subset $\widetilde{\Omega}_{\kappa,j}$ necessary.

RHS (sixth term VI): Using $\langle \boldsymbol{a} \times \boldsymbol{b}, \boldsymbol{c} \rangle_{\mathbb{R}^3} = \langle \boldsymbol{b}, \boldsymbol{c} \times \boldsymbol{a} \rangle_{\mathbb{R}^3}$ and integration by parts, we obtain

$$\begin{aligned} VI &\leq \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \frac{k}{2} \Big| \Big(\Delta \boldsymbol{m}^{j+1}, \boldsymbol{z}^{j+1} \times (\boldsymbol{m}^{j+1} - \boldsymbol{m}^j) \Big) \Big| \\ &\leq \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \frac{k}{2} \Big| \Big(\nabla \boldsymbol{m}^{j+1}, \nabla \boldsymbol{z}^{j+1} \times (\boldsymbol{m}^{j+1} - \boldsymbol{m}^j) \Big) \Big| \\ &\quad + \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \frac{k}{2} \Big| \Big(\nabla \boldsymbol{m}^{j+1}, \boldsymbol{z}^{j+1} \times \nabla[\boldsymbol{m}^{j+1} - \boldsymbol{m}^j] \Big) \Big| \\ &=: VI_A + VI_B. \end{aligned}$$

For the first term VI_A, by Gagliardo-Nirenberg's and Young's inequalities, we arrive at

$$\begin{aligned} VI_A &\leq \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2} \|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2} \|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{L}^\infty} \\ &\leq Ck \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 \|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{W}^{1,2}} \|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{L}^2} + \delta \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2, \end{aligned} \tag{4.23}$$

and the second term VI_B can be similarly simplified to

$$\begin{aligned} VI_B &\leq \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2} \|\nabla[\boldsymbol{m}^{j+1} - \boldsymbol{m}^j]\|_{\mathbf{L}^2} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^\infty} \\ &\leq Ck \|\nabla[\boldsymbol{m}^{j+1} - \boldsymbol{m}^j]\|_{\mathbf{L}^2}^2 + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\quad + \delta \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\boldsymbol{z}^{j+1}\|_{\mathbf{W}^{1,2}}^2. \end{aligned} \tag{4.24}$$

So far, estimates for the terms I–VI in inequality (4.1) are derived. If we combine the obtained estimates (4.4)–(4.24) of the first step, we arrive at

$$\begin{aligned} &\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \frac{1}{2} \Big(\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 - \|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2 + \|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|_{\mathbf{L}^2}^2 \Big) + \alpha \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\quad + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} k \Big(\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4 + \|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^4 + \|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{W}^{1,2}}^2 \Big) \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\quad + Ck \|\nabla[\boldsymbol{m}^{j+1} - \boldsymbol{m}^j]\|_{\mathbf{L}^2}^2 \\ &\quad + kA_{j+1} \sup_{t_j \leq s \leq t_{j+1}} \|\nabla[\boldsymbol{m}(s) - \boldsymbol{m}(t_{j+1})]\|_{\mathbf{L}^2}^2 + kB_{j+1} \|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{W}^{1,2}} \|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{L}^2} \end{aligned}$$

$$+ kD_{j+1} \sup_{t_j \le s \le t_{j+1}} \|\boldsymbol{m}(s) - \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2 + kE_{j+1}\|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{L}^2}^2$$
$$+ \Big(\int_{t_j}^{t_{j+1}} \boldsymbol{m}(s) \times \circ \, \mathrm{d}\boldsymbol{W}(s) - \boldsymbol{m}^{j+\frac{1}{2}} \times \Delta_j \boldsymbol{W}, \boldsymbol{z}^{j+1}\Big). \tag{4.25}$$

In inequality (4.25) above, the term A_{j+1} is constructed by (4.2), (4.9), (4.10), (4.19), (4.20), and (4.21), such that

$$A_{j+1} := C + C \sup_{t_j \le s \le t_{j+1}} \|\Delta \boldsymbol{m}(s)\|_{\mathbf{L}^2}^2.$$

The definition of term B_{j+1} is motivated from (4.4), (4.5), (4.13), (4.14), (4.17), (4.18), and (4.23)

$$B_{j+1} := C \sup_{r=0,\dots,j+1} \|\nabla \boldsymbol{m}^r\|_{\mathbf{L}^2}^2 + C\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4.$$

The following term D_{j+1} comes from (4.6) and (4.19)

$$D_{j+1} := C \sup_{t_j \le s \le t_{j+1}} \|\Delta \boldsymbol{m}(s)\|_{\mathbf{L}^2}^2 + \sup_{t_j \le s \le t_{j+1}} \|\nabla \boldsymbol{m}(s)\|_{\mathbf{L}^2}^2 \|\nabla \boldsymbol{m}(s)\|_{\mathbf{W}^{1,2}}^2;$$

term E_{j+1} is motivated by estimate (4.15)

$$E_{j+1} := C\|\boldsymbol{m}^{j+1} - \boldsymbol{m}^j\|_{\mathbf{W}^{1,2}}^2 \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2.$$

The terms A_{j+1} to E_{j+1} share a common property. It is due to Proposition 3.2 and Lemma 3.4, that $\mathbb{E}[(A_{j+1})^p] \le C$, $\mathbb{E}[(B_{j+1})^p] \le C$, $\mathbb{E}[(D_{j+1})^p] \le C$ and $\mathbb{E}[(E_{j+1})^p] \le C$ for $p \ge 1$ hold.
We first sum over $j = 0, \dots, r$ and then take the supremum over $r = 0, \dots, J-1$ and finally take expectations in inequality (4.25). The lines 2 to 4 on the right-hand-side of inequality (4.25) can now be estimated by using Hölder's inequality and the results in Proposition 3.3 and Lemma 3.4, and we obtain at least rate 1 for those terms.
On the left-hand-side of (4.25), we use $\widetilde{\Omega}_{\kappa,j+1} \subset \widetilde{\Omega}_{\kappa,j}$ for all $j = 0, \dots, J-2$ to obtain

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{r=0,\dots,J-1} \sum_{j=0}^{r} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \big(\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 - \|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\big)\Big] \\
&= \mathbb{E}\Big[\sup_{r=0,\dots,J-1} \Big(\mathbb{1}_{\widetilde{\Omega}_{\kappa,r}} \|\boldsymbol{z}^{r+1}\|_{\mathbf{L}^2}^2 + \sum_{j=0}^{r} (\mathbb{1}_{\widetilde{\Omega}_{\kappa,j-1}} - \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}) \|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big)\Big] \\
&\ge \mathbb{E}\Big[\sup_{r=0,\dots,J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,r}} \|\boldsymbol{z}^{r+1}\|_{\mathbf{L}^2}^2\Big].
\end{aligned}$$

Thus, we arrive at

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{r=0,\dots,J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,r}} \|\boldsymbol{z}^{r+1}\|_{\mathbf{L}^2}^2 + \sum_{j=0}^{J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|_{\mathbf{L}^2}^2 + \alpha k \sum_{j=0}^{J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big] \qquad (4.26) \\
&\le Ck + Ck \sum_{j=0}^{J-1} \mathbb{E}[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2] + Ck \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big]
\end{aligned}$$

$$+ Ck \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^4 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \Big] + Ck \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{W}^{1,2}}^2 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \Big]$$
$$+ \mathbb{E}\Big[\sup_{r=0,\dots,J-1} \sum_{j=0}^{r} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \Big(\int_{t_j}^{t_{j+1}} \boldsymbol{m}(s) \times \circ \, \mathrm{d}\boldsymbol{W}(s) - \boldsymbol{m}^{j+\frac{1}{2}} \times \Delta_j \boldsymbol{W}, \boldsymbol{z}^{j+1} \Big) \Big].$$

Before starting to estimate the stochastic integral term, the three mixed terms (first and second line on the right-hand-side of inequality (4.26)) are considered: For the third term on the right-hand-side of (4.26), there holds

$$\begin{aligned} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 &\le \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \Big(\|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4 - \|\nabla \boldsymbol{m}^{j}\|_{\mathbf{L}^2}^4 \Big) \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\ &\quad + \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{m}^{j}\|_{\mathbf{L}^2}^4 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \\ &=: \widehat{I} + \widehat{II}. \end{aligned}$$

The first part $\widehat{I}$ yields

$$\begin{aligned} k\mathbb{E}\Big[\sum_{j=0}^{J-1} |\widehat{I}|\Big] &\le k\mathbb{E}\Bigg[\Big(\sum_{j=0}^{J-1} \Big| \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4 - \|\nabla \boldsymbol{m}^{j}\|_{\mathbf{L}^2}^4 \Big|^2 \Big)^{\frac{1}{2}} \\ &\qquad \times \Big(\sum_{j=0}^{J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \Big(\|\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 + \|\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2 \Big) \Big)^{\frac{1}{2}} \Bigg] \\ &\le Ck\mathbb{E}\Big[\sum_{j=0}^{J-1} \Big| \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4 - \|\nabla \boldsymbol{m}^{j}\|_{\mathbf{L}^2}^4 \Big|^2 \Big] \\ &\quad + Ck\mathbb{E}\Big[\sum_{j=0}^{J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \Big(\|\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 + \|\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2 \Big) \Big] \\ &\le Ck + Ck \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \Big], \end{aligned}$$

by using Lemma 3.4 d), the $\mathbb{P}$-a.s. estimate $\sup_j (\|\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2 + \|\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2) \le C$, and Hölder's inequality. For the second part, we obtain $\widehat{II} \le C\kappa \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2$. Thus, we get for the first mixed term in inequality (4.26)

$$k \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^4 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \Big] \le Ck + Ck(\kappa+1) \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \Big].$$

The second mixed term in inequality (4.26) can be simplified in a similar way, just with the use of Proposition 3.2 b) instead of Lemma 3.4 d). Here, we finally obtain

$$k \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^4 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \Big] \le Ck^2 + Ck\kappa \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 \Big].$$

In order to simplify the term $\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{W}^{1,2}}^2 \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2$, we rewrite $\|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{W}^{1,2}}^2$ to $\|\nabla \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2 + \|\Delta \boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2$, where the first part can be handled as before. For the

second part $\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\Delta \boldsymbol{m}(t_{j+1})\|^2_{\mathbf{L}^2}\|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2}$, we arrive by similar arguments as before using $\sup_j \|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2} \leq C$ $\mathbb{P}$-a.s. at

$$\begin{aligned}
&\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\Delta \boldsymbol{m}(t_{j+1})\|^2_{\mathbf{L}^2}\|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2}\\
&\leq \Big(\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} C\|\Delta[\boldsymbol{m}(t_{j+1})-\boldsymbol{m}(t_j)]\|^2_{\mathbf{L}^2} + \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} C\|\Delta \boldsymbol{m}(t_j)\|^2_{\mathbf{L}^2}\Big)\|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2}\\
&\leq \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} C\|\Delta[\boldsymbol{m}(t_{j+1})-\boldsymbol{m}(t_j)]\|^2_{\mathbf{L}^2}\|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2} + \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} C\kappa\|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2}\\
&\leq C\|\Delta[\boldsymbol{m}(t_{j+1})-\boldsymbol{m}(t_j)]\|^2_{\mathbf{L}^2} + \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} C\kappa\|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2}.
\end{aligned}$$

After summation and taking expectations, we obtain

$$k\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\nabla \boldsymbol{m}(t_{j+1})\|^2_{\mathbf{W}^{1,2}}\|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2}\Big] \leq Ck + Ck(\kappa+1)\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2}\Big]$$

using Proposition 3.3 c). As a consequence, inequality (4.26) can be rewritten as

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\mathbb{1}_{\widetilde{\Omega}_{\kappa,r-1}}\|\boldsymbol{z}^r\|^2_{\mathbf{L}^2} + \sum_{j=0}^{J-1}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^j\|^2_{\mathbf{L}^2} + \alpha k\sum_{j=0}^{J-1}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\nabla \boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2}\Big]\\
&\leq Ck + Ck(1+\kappa)\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}\|^2_{\mathbf{L}^2}\Big]\\
&\quad + \mathbb{E}\Big[\sup_{r=0,\dots,J-1}\sum_{j=0}^{r}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(\int_{t_j}^{t_{j+1}}\boldsymbol{m}(s)\times\circ\,\mathrm{d}\boldsymbol{W}(s) - \boldsymbol{m}^{j+\frac{1}{2}}\times\Delta_j\boldsymbol{W},\boldsymbol{z}^{j+1}\Big)\Big], \qquad (4.27)
\end{aligned}$$

where the last term will be estimated in the second step below. From inequality (4.27), we observe, that the terms I–VI in equation (4.1) yield a convergence rate of order close to one.

Second Step: *Stochastic term VII.*
An estimate for the last expression in inequality (4.27) is derived. Therefore, we first rewrite the Stratonovich integral term and consider the Itô correction:

$$\begin{aligned}
\int_{t_j}^{t_{j+1}}\boldsymbol{m}(s)\times\circ\,\mathrm{d}\boldsymbol{W}(s) &= \sum_{l=1}^{\infty}\sqrt{q_l}\int_{t_j}^{t_{j+1}}\boldsymbol{m}(s)\times\mathbf{e}_l\circ\,\mathrm{d}\beta^l(s)\\
&= \sum_{l=1}^{\infty}\sqrt{q_l}\int_{t_j}^{t_{j+1}}\boldsymbol{m}(s)\times\mathbf{e}_l\,\mathrm{d}\beta^l(s) + \frac{1}{2}\sum_{l=1}^{\infty}q_l\int_{t_j}^{t_{j+1}}(\boldsymbol{m}(s)\times\mathbf{e}_l)\times\mathbf{e}_l\,\mathrm{d}s\\
&= \int_{t_j}^{t_{j+1}}\boldsymbol{m}(s)\times\,\mathrm{d}\boldsymbol{W}(s) + \frac{1}{2}\sum_{l=1}^{\infty}q_l\int_{t_j}^{t_{j+1}}(\boldsymbol{m}(s)\times\mathbf{e}_l)\times\mathbf{e}_l\,\mathrm{d}s.
\end{aligned}$$

On the other hand, by using $\boldsymbol{m}^{j+\frac{1}{2}} = \boldsymbol{m}^j + \frac{1}{2}(\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)$, we obtain

$$\begin{aligned}
\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta_j\boldsymbol{W} &= \sum_{l=1}^{\infty}\sqrt{q_l}\Delta_j\beta^l\boldsymbol{m}^j\times\mathbf{e}_l + \frac{1}{2}\sum_{l=1}^{\infty}\sqrt{q_l}\Delta_j\beta^l(\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\times\mathbf{e}_l\\
&= \sum_{l=1}^{\infty}\sqrt{q_l}\Delta_j\beta^l\boldsymbol{m}^j\times\mathbf{e}_l + \frac{1}{2}k\sum_{l=1}^{\infty}q_l(\boldsymbol{m}^{j+\frac{1}{2}}\times\mathbf{e}_l)\times\mathbf{e}_l + \mathcal{A}_j,
\end{aligned}$$

where the term $\mathcal{A}_j$ is defined by

$$\mathcal{A}_j := \frac{1}{2}\sum_{l=1}^{\infty}\sqrt{q_l}\Delta_j\beta^l(\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\times\mathbf{e}_l - \frac{1}{2}k\sum_{l=1}^{\infty}q_l(\boldsymbol{m}^{j+\frac{1}{2}}\times\mathbf{e}_l)\times\mathbf{e}_l.$$

The equation in Algorithm 1.1 is used in order to restate the increment $\boldsymbol{m}^{j+1}-\boldsymbol{m}^j$, such that

$$\begin{aligned}\mathcal{A}_j &= \frac{1}{2}\sum_{l=1}^{\infty}q_l(|\Delta_j\beta^l|^2-k)(\boldsymbol{m}^{j+\frac{1}{2}}\times\mathbf{e}_l)\times\mathbf{e}_l\\ &\quad+\frac{1}{2}k\sum_{l=1}^{\infty}\sqrt{q_l}\Delta_j\beta^l\Big[\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}-\alpha\boldsymbol{m}^{j+\frac{1}{2}}\times(\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1})\Big]\times\mathbf{e}_l\\ &\quad+\frac{1}{2}\sum_{l_1=1}^{\infty}\sum_{l_2\neq l_1}^{\infty}\sqrt{q_{l_1}}\sqrt{q_{l_2}}\Delta_j\beta^{l_1}\Delta_j\beta^{l_2}(\boldsymbol{m}^{j+\frac{1}{2}}\times\mathbf{e}_{l_2})\times\mathbf{e}_{l_1}.\end{aligned}$$

This argument, together with the identity

$$\boldsymbol{m}(s)\times\mathbf{e}_l-\boldsymbol{m}^j\times\mathbf{e}_l=\big[\boldsymbol{m}(s)-\boldsymbol{m}(t_j)\big]\times\mathbf{e}_l+\boldsymbol{z}^j\times\mathbf{e}_l$$

is used to decompose

$$\begin{aligned}&\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(\int_{t_j}^{t_{j+1}}\boldsymbol{m}(s)\times\circ\,\mathrm{d}\boldsymbol{W}(s)-\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta_j\boldsymbol{W},\boldsymbol{z}^{j+1}\Big)\\ &=\Big(\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times\,\mathrm{d}\boldsymbol{W}(s),\boldsymbol{z}^{j+1}\Big)+\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(\boldsymbol{z}^j\times\Delta_j\boldsymbol{W},\boldsymbol{z}^{j+1}\Big)\\ &\quad+\frac{1}{2}\sum_{l=1}^{\infty}q_l\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(\Big((\boldsymbol{m}(s)-\boldsymbol{m}^{j+\frac{1}{2}})\times\mathbf{e}_l\Big)\times\mathbf{e}_l\,\mathrm{d}s,\boldsymbol{z}^{j+1}\Big)\\ &\quad+\frac{1}{2}\sum_{l=1}^{\infty}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}q_l(|\Delta_j\beta^l|^2-k)\Big((\boldsymbol{m}^{j+\frac{1}{2}}\times\mathbf{e}_l)\times\mathbf{e}_l,\boldsymbol{z}^{j+1}\Big)\\ &\quad+\frac{1}{2}k\sum_{l=1}^{\infty}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\sqrt{q_l}\Delta_j\beta^l\Big(\Big[\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}-\alpha\boldsymbol{m}^{j+\frac{1}{2}}\times(\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1})\Big]\times\mathbf{e}_l,\boldsymbol{z}^{j+1}\Big)\\ &\quad+\frac{1}{2}\sum_{l_1=1}^{\infty}\sum_{l_2\neq l_1}^{\infty}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\sqrt{q_{l_1}}\sqrt{q_{l_2}}\Delta_j\beta^{l_1}\Delta_j\beta^{l_2}\Big((\boldsymbol{m}^{j+\frac{1}{2}}\times\mathbf{e}_{l_2})\times\mathbf{e}_{l_1},\boldsymbol{z}^{j+1}\Big)\\ &=:\widetilde{I}+\widetilde{II}+\widetilde{III}+\widetilde{IV}+\widetilde{V}+\widetilde{VI}.\end{aligned}$$

In the following, each term is treated separately, starting with the first term $\widetilde{I}$.

Term $\widetilde{I}$: Here, by using $\boldsymbol{z}^{j+1}=(\boldsymbol{z}^{j+1}-\boldsymbol{z}^j)+\boldsymbol{z}^j$, we obtain:

$$\begin{aligned}\widetilde{I}&\le\Big(\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times\,\mathrm{d}\boldsymbol{W}(s),\boldsymbol{z}^{j+1}-\boldsymbol{z}^j\Big)\\ &\quad+\Big(\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times\,\mathrm{d}\boldsymbol{W}(s),\boldsymbol{z}^{j}\Big)\\ &=:\widetilde{I}_A+\widetilde{I}_B.\end{aligned}$$

For the first term corresponding to term $\widetilde{I}_A$, we conclude using Young's inequality, Itô isometry, and finally Hölder's inequality

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\sum_{j=0}^{r}\Big(\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times \mathrm{d}\boldsymbol{W}(s),\boldsymbol{z}^{j+1}-\boldsymbol{z}^{j}\Big)\Big]\\
&\le \mathbb{E}\Big[\sum_{j=0}^{J-1}\Big(C\|\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times \mathrm{d}\boldsymbol{W}(s)\|_{\mathbf{L}^2}^2+\delta\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^{j}\|_{\mathbf{L}^2}^2\Big)\Big]\\
&= \delta\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^{j}\|_{\mathbf{L}^2}^2\Big]\\
&\quad + C\sum_{j=0}^{J-1}\mathbb{E}\Big[\|\sum_{l=1}^{\infty}\sqrt{q_l}\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times\mathbf{e}_l\,\mathrm{d}\beta^l(s)\|_{\mathbf{L}^2}^2\Big]\\
&\le \delta\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^{j}\|_{\mathbf{L}^2}^2\Big]\\
&\quad + C\,\mathrm{Tr}\,\boldsymbol{Q}^{\frac{1}{2}}\sum_{j=0}^{J-1}\sum_{l=1}^{\infty}\sqrt{q_l}\mathbb{E}\Big[\|\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times\mathbf{e}_l\,\mathrm{d}\beta^l(s)\|_{\mathbf{L}^2}^2\Big]\\
&\le \delta\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^{j}\|_{\mathbf{L}^2}^2\Big]\\
&\quad + C\,\mathrm{Tr}\,\boldsymbol{Q}^{\frac{1}{2}}\sum_{j=0}^{J-1}\sum_{l=1}^{\infty}\sqrt{q_l}\int_{t_j}^{t_{j+1}}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times\mathbf{e}_l\|_{\mathbf{L}^2}^2\Big]\,\mathrm{d}s.
\end{aligned}
$$

By using $\|\boldsymbol{u}\times\boldsymbol{v}\|_{\mathbf{L}^2}\le C\|\boldsymbol{u}\|_{\mathbf{L}^2}\|\boldsymbol{v}\|_{\mathbf{L}^\infty}$ together with Assumption A_1 and Proposition 3.3, we arrive at

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\sum_{j=0}^{r}\Big(\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times \mathrm{d}\boldsymbol{W}(s),\boldsymbol{z}^{j+1}-\boldsymbol{z}^{j}\Big)\Big]\\
&\le \delta\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^{j}\|_{\mathbf{L}^2}^2\Big]+C(\mathrm{Tr}\,\boldsymbol{Q}^{\frac{1}{2}})^2\sum_{j=0}^{J-1}\int_{t_j}^{t_{j+1}}\mathbb{E}\Big[\|\boldsymbol{m}(s)-\boldsymbol{m}(t_j)\|_{\mathbf{L}^2}^2\Big]\,\mathrm{d}s\\
&\le \delta\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^{j}\|_{\mathbf{L}^2}^2\Big]+C(\mathrm{Tr}\,\boldsymbol{Q}^{\frac{1}{2}})^2k.
\end{aligned}
$$

The second term $\widetilde{I}_B$ can be handled using Burkholder-Davis-Gundy's inequality:

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\sum_{j=0}^{r}\Big(\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times \mathrm{d}\boldsymbol{W}(s),\boldsymbol{z}^{j}\Big)\Big]\\
&= \mathbb{E}\Big[\sup_{r=0,\dots,J-1}\sum_{j=0}^{r}\sum_{l=1}^{\infty}\sqrt{q_l}\Big(\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times\mathbf{e}_l\,\mathrm{d}\beta^l(s),\boldsymbol{z}^{j}\Big)\Big]\\
&\le \sum_{l=1}^{\infty}\sqrt{q_l}\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\Big|\sum_{j=0}^{r}\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big((\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times\mathbf{e}_l,\boldsymbol{z}^{j}\Big)\,\mathrm{d}\beta^l(s)\Big|\Big]
\end{aligned}
$$

$$\leq C\sum_{l=1}^{\infty}\sqrt{q_l}\mathbb{E}\Big[\Big|\sum_{j=0}^{J-1}\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big|\Big((\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times\mathbf{e}_l,\boldsymbol{z}^j\Big)\Big|^2\,\mathrm{d}s\Big|^{\frac{1}{2}}\Big].$$

By using $\|\boldsymbol{u}\times\boldsymbol{v}\|_{\mathbf{L}^2}\leq C\|\boldsymbol{u}\|_{\mathbf{L}^2}\|\boldsymbol{v}\|_{\mathbf{L}^\infty}$ together with Assumption A_1, Young's inequality, making the supremum, and Proposition 3.3 a), we continue with

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\sum_{j=0}^{r}\Big(\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\int_{t_j}^{t_{j+1}}(\boldsymbol{m}(s)-\boldsymbol{m}(t_j))\times\,\mathrm{d}\boldsymbol{W}(s),\boldsymbol{z}^j\Big)\Big]\\
&\leq\sum_{l=1}^{\infty}\sqrt{q_l}\mathbb{E}\Big[\Big|\sum_{j=0}^{J-1}\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{m}(s)-\boldsymbol{m}(t_j)\|_{\mathbf{L}^2}^2\|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\,\mathrm{d}s\Big|^{\frac{1}{2}}\Big]\\
&\leq\sum_{l=1}^{\infty}\sqrt{q_l}\mathbb{E}\Big[\sup_{r=0,\dots,J-1}(\mathbb{1}_{\widetilde{\Omega}_{\kappa,r}}\|\boldsymbol{z}^r\|_{\mathbf{L}^2})\Big|\sum_{j=0}^{J-1}\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{m}(s)-\boldsymbol{m}(t_j)\|_{\mathbf{L}^2}^2\,\mathrm{d}s\Big|^{\frac{1}{2}}\Big]\\
&\leq\widetilde{\delta}\sum_{l=1}^{\infty}\sqrt{q_l}\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\mathbb{1}_{\widetilde{\Omega}_{\kappa,r}}\|\boldsymbol{z}^r\|_{\mathbf{L}^2}^2\Big]+C\sum_{l=1}^{\infty}\sqrt{q_l}\mathbb{E}\Big[\sum_{j=0}^{J-1}\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{m}(s)-\boldsymbol{m}(t_j)\|_{\mathbf{L}^2}^2\,\mathrm{d}s\Big]\\
&\leq\widetilde{\delta}\,\mathrm{Tr}\,\boldsymbol{Q}^{\frac{1}{2}}\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\mathbb{1}_{\widetilde{\Omega}_{\kappa,r}}\|\boldsymbol{z}^r\|_{\mathbf{L}^2}^2\Big]+C\,\mathrm{Tr}\,\boldsymbol{Q}^{\frac{1}{2}}k.
\end{aligned}$$

Here the property $\widetilde{\Omega}_{\kappa,r}\subset\widetilde{\Omega}_{\kappa,r-1}$ is used, which yields $\mathbb{1}_{\widetilde{\Omega}_{\kappa,r}}\leq\mathbb{1}_{\widetilde{\Omega}_{\kappa,r-1}}$. The first part can be absorbed later in inequality (4.27).

Term $\widetilde{II}$: Using $\langle\boldsymbol{a}\times\boldsymbol{b},\boldsymbol{a}\rangle_{\mathbb{R}^3}=0$, we obtain

$$\widetilde{II}=\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(\boldsymbol{z}^j\times\Delta_j\boldsymbol{W},\boldsymbol{z}^{j+1}\Big)=\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(\boldsymbol{z}^j\times\Delta_j\boldsymbol{W},\boldsymbol{z}^{j+1}-\boldsymbol{z}^j\Big),$$

thus, using Young's inequality together with the independence of all, $\|\Delta_j\boldsymbol{W}\|_{\mathbf{L}^\infty}^2$, $\|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2$, and $\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}$, we conclude

$$\begin{aligned}
\mathbb{E}\Big[\sum_{j=0}^{J-1}|\widetilde{II}|\Big]&\leq\mathbb{E}\Big[\sum_{j=0}^{J-1}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big|\Big(\boldsymbol{z}^j\times\Delta_j\boldsymbol{W},\boldsymbol{z}^{j+1}-\boldsymbol{z}^j\Big)\Big|\Big]\\
&\leq\mathbb{E}\Big[\sum_{j=0}^{J-1}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\|\Delta_j\boldsymbol{W}\|_{\mathbf{L}^\infty}^2\Big]+\delta\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big]\\
&\leq Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big]+\delta\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}-\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big],
\end{aligned}$$

where we have used again $\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\leq\mathbb{1}_{\widetilde{\Omega}_{\kappa,j-1}}$.

Term $\widetilde{III}$: The third term $\widetilde{III}$ can be decomposed using $\boldsymbol{m}(s)-\boldsymbol{m}^{j+\frac{1}{2}}=\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})+\boldsymbol{z}^{j+1}+\frac{1}{2}(\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)$ into three parts:

$$\begin{aligned}
\widetilde{III}=\frac{1}{2}\sum_{l=1}^{\infty}q_l\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(((\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1}))\times\mathbf{e}_l)\times\mathbf{e}_l\,\mathrm{d}s,\boldsymbol{z}^{j+1}\Big)\\
+\frac{1}{2}\sum_{l=1}^{\infty}q_l\int_{t_j}^{t_{j+1}}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big((\boldsymbol{z}^{j+1}\times\mathbf{e}_l)\times\mathbf{e}_l\,\mathrm{d}s,\boldsymbol{z}^{j+1}\Big)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{4}\sum_{l=1}^{\infty} q_l \int_{t_j}^{t_{j+1}} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big(((\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\times\mathbf{e}_l)\times\mathbf{e}_l\,\mathrm{d}s, \boldsymbol{z}^{j+1}\Big)\\
&=: \widetilde{III}_A + \widetilde{III}_B + \widetilde{III}_C.
\end{aligned}
$$

Starting with the first part, we obtain by using $\|\boldsymbol{u}\times\boldsymbol{v}\|_{\mathbf{L}^2}\le C\|\boldsymbol{u}\|_{\mathbf{L}^2}\|\boldsymbol{v}\|_{\mathbf{L}^\infty}$ together with Assumption A_1, and Hölder's and Young's inequalities,

$$
\begin{aligned}
|\widetilde{III}_A| &\le \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\| \int_{t_j}^{t_{j+1}} \sum_{l=1}^{\infty} q_l((\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1}))\times\mathbf{e}_l)\times\mathbf{e}_l\,\mathrm{d}s\|_{\mathbf{L}^2}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}\\
&\le C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\sum_{l=1}^{\infty} q_l k^{\frac{1}{2}}\Big(\int_{t_j}^{t_{j+1}}\|\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2\,\mathrm{d}s\Big)^{\frac{1}{2}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}\\
&\le C\,\mathrm{Tr}\,\boldsymbol{Q}\int_{t_j}^{t_{j+1}}\|\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2\,\mathrm{d}s + C\,\mathrm{Tr}\,\boldsymbol{Q}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2.
\end{aligned}
$$

Summation, taking expectations, and using Proposition 3.3 a) yield

$$
\begin{aligned}
\mathbb{E}\Big[\sum_{j=0}^{J-1}|\widetilde{III}_A|\Big] &\le C\sum_{j=0}^{J-1}\int_{t_j}^{t_{j+1}}\mathbb{E}\Big[\|\boldsymbol{m}(s)-\boldsymbol{m}(t_{j+1})\|_{\mathbf{L}^2}^2\Big]\,\mathrm{d}s + Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big]\\
&\le Ck + Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big].
\end{aligned}
$$

For the second term $\widetilde{III}_B$, by $\|\boldsymbol{u}\times\boldsymbol{v}\|_{\mathbf{L}^2}\le C\|\boldsymbol{u}\|_{\mathbf{L}^2}\|\boldsymbol{v}\|_{\mathbf{L}^\infty}$ together with Assumption A_1, we obtain that

$$
|\widetilde{III}_B| \le \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big|\frac{1}{2}\sum_{l=1}^{\infty} q_l k\Big((\boldsymbol{z}^{j+1}\times\mathbf{e}_l)\times\mathbf{e}_l, \boldsymbol{z}^{j+1}\Big)\Big| \le C\,\mathrm{Tr}\,\boldsymbol{Q}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2.
$$

After summation and taking expectation, we obtain

$$
\mathbb{E}\Big[\sum_{j=0}^{J-1}|\widetilde{III}_B|\Big] \le Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big].
$$

Finally, for the last term $\widetilde{III}_C$, we conclude

$$
\begin{aligned}
|\widetilde{III}_C| &\le \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\Big|\sum_{l=1}^{\infty} q_l\Big(((\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\times\mathbf{e}_l)\times\mathbf{e}_l, \boldsymbol{z}^{j+1}\Big)\Big|\\
&\le C(\mathrm{Tr}\,\boldsymbol{Q})^2 k\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}^2 + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2
\end{aligned}
$$

as before. This yields

$$
\mathbb{E}\Big[\sum_{j=0}^{J-1}|\widetilde{III}_C|\Big] \le Ck + Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big]
$$

using inequality (3.8) and Lemma 3.4.

Term $\widetilde{IV}$: The fourth term has to be rewritten again using $\boldsymbol{z}^{j+1} = (\boldsymbol{z}^{j+1} - \boldsymbol{z}^j) + \boldsymbol{z}^j$ into the parts $\widetilde{IV}_A$ and $\widetilde{IV}_B$. For the first part $\widetilde{IV}_A$ corresponding to $(\boldsymbol{z}^{j+1} - \boldsymbol{z}^j)$, we obtain

$$\begin{aligned}|\widetilde{IV}_A| &= \frac{1}{2}\Big|\sum_{l=1}^{\infty} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} q_l(|\Delta_j\beta^l|^2 - k)\Big((\boldsymbol{m}^{j+\frac{1}{2}} \times \mathbf{e}_l) \times \mathbf{e}_l, \boldsymbol{z}^{j+1} - \boldsymbol{z}^j\Big)\Big| \\ &\leq C\sum_{l=1}^{\infty} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} q_l\Big||\Delta_j\beta^l|^2 - k\Big|\|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|_{\mathbf{L}^2}\end{aligned}$$

using the estimate $\|\boldsymbol{u} \times \boldsymbol{v}\|_{\mathbf{L}^2} \leq C\|\boldsymbol{u}\|_{\mathbf{L}^2}\|\boldsymbol{v}\|_{\mathbf{L}^\infty}$ together with Assumption A_1. Taking expectations, summing up over $j = 0, \dots, J-1$, using Hölder's inequality and the equality $\mathbb{E}\Big[\Big||\Delta_j\beta^l|^2 - k\Big|^2\Big] = 2k^2$, which holds due to $\Delta_j\beta^l \sim \mathcal{N}(0,k)$, we arrive at

$$\begin{aligned}\mathbb{E}\Big[\sum_{j=0}^{J-1}|\widetilde{IV}_A|\Big] &\leq C\sum_{j=0}^{J-1}\sum_{l=1}^{\infty} q_l\Big(\mathbb{E}\Big[\Big||\Delta_j\beta^l|^2 - k\Big|^2\Big]\Big)^{\frac{1}{2}}\Big(\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big]\Big)^{\frac{1}{2}} \\ &\leq C(\operatorname{Tr}\boldsymbol{Q})^2 k + \delta\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big].\end{aligned}$$

For the second part $\widetilde{IV}_B$, we simplify using $\boldsymbol{m}^{j+\frac{1}{2}} = \boldsymbol{m}^j + \frac{1}{2}(\boldsymbol{m}^{j+1} - \boldsymbol{m}^j)$, thus

$$\begin{aligned}\widetilde{IV}_B &= \frac{1}{2}\sum_{l=1}^{\infty}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} q_l(|\Delta_j\beta^l|^2 - k)\Big((\boldsymbol{m}^j \times \mathbf{e}_l) \times \mathbf{e}_l, \boldsymbol{z}^j\Big) \\ &\quad + \frac{1}{4}\sum_{l=1}^{\infty}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} q_l(|\Delta_j\beta^l|^2 - k)\Big(((\boldsymbol{m}^{j+1} - \boldsymbol{m}^j) \times \mathbf{e}_l) \times \mathbf{e}_l, \boldsymbol{z}^j\Big) \\ &=: \widetilde{IV}_{B,1} + \widetilde{IV}_{B,2}.\end{aligned}$$

To get an estimate for $\widetilde{IV}_{B,1}$, we define

$$\begin{aligned}X_j &:= \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\sum_{l=1}^{\infty} q_l(|\Delta_j\beta^l|^2 - k)\Big((\boldsymbol{m}^j \times \mathbf{e}_l) \times \mathbf{e}_l, \boldsymbol{z}^j\Big) \qquad \forall j = 0, \dots, J-1; \\ Y_r &:= \sum_{j=0}^{r-1} X_j \qquad \forall r = 1, \dots, J.\end{aligned}$$

Since X_j is $\mathcal{F}_{t_w}$-measurable ($j < w$), we obtain that Y_r is $\mathcal{F}_{t_r}$-measurable. By $\mathbb{E}[|Y_r|] < \infty$ and $\mathbb{P}$-a.s.

$$\begin{aligned}\mathbb{E}[Y_r|\mathcal{F}_{t_{r-1}}] &= Y_{r-1} + \mathbb{1}_{\widetilde{\Omega}_{\kappa,r-1}}\sum_{l=1}^{\infty} q_l((\boldsymbol{m}^{r-1} \times \mathbf{e}_l) \times \mathbf{e}_l, \boldsymbol{z}^{r-1})\mathbb{E}[(|\Delta_{r-1}\beta^l|^2 - k)|\mathcal{F}_{t_{r-1}}] \\ &= Y_{r-1}\end{aligned}$$

for every $r = 1, \dots, J$, we obtain that the process $\{Y_j\}_j$ is a $\{\mathcal{F}_{t_j}\}_j$-martingale.
Thus, we are able to apply Doob's inequality

$$\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\sum_{j=0}^{r}\widetilde{IV}_{B,1}\Big] \leq \mathbb{E}\Big[\sup_{r=1,\dots,J}|\sum_{j=0}^{r-1} X_j|\Big]$$

$$\leq \Big(\mathbb{E}\Big[\Big(\sup_{r=1,\dots,J}|\sum_{j=0}^{r-1}X_j|\Big)^2\Big]\Big)^{\frac{1}{2}}$$
$$\leq C\Big(\mathbb{E}\Big[\Big(|\sum_{j=0}^{J-1}X_j|\Big)^2\Big]\Big)^{\frac{1}{2}}$$
$$\leq C\Big(\sum_{j=0}^{J-1}\mathbb{E}[|X_j|^2]+2\sum_{j<w}\mathbb{E}[X_jX_w]\Big)^{\frac{1}{2}}.$$

Since $j<w$, we obtain

$$\begin{aligned}\mathbb{E}[X_jX_w]&=\mathbb{E}[X_j\mathbb{E}[X_w|\mathcal{F}_{t_w}]]\\&=\sum_{l=1}^{\infty}q_l\mathbb{E}\Big[X_j\mathbb{1}_{\widetilde{\Omega}_{\kappa,w}}((\boldsymbol{m}^w\times\mathbf{e}_l)\times\mathbf{e}_l,\boldsymbol{z}^w)\mathbb{E}[(|\Delta_w\beta^l|^2-k)|\mathcal{F}_{t_w}]\Big]\\&=0,\end{aligned}$$

and thus, using Young's inequality and independence of $|\Delta_j\beta^l|^2$ and $\|\boldsymbol{z}^j\|^2_{\mathbf{L}^2}$, again estimate $\mathbb{E}[\big||\Delta_j\beta^l|^2-k\big|^2]=2k^2$, and finally $\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\leq\mathbb{1}_{\widetilde{\Omega}_{\kappa,j-1}}$, we arrive at

$$\begin{aligned}\mathbb{E}\Big[\sup_{r=0,\dots,J-1}\sum_{j=0}^{r}\widetilde{IV}_{B,1}\Big]&\leq C\Big(\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big|\sum_{l=1}^{\infty}q_l(|\Delta_j\beta^l|^2-k)\Big((\boldsymbol{m}^j\times\mathbf{e}_l)\times\mathbf{e}_l,\boldsymbol{z}^j\Big)\Big|^2\Big]\Big)^{\frac{1}{2}}\\&\leq C\Big(\sum_{j=0}^{J-1}\operatorname{Tr}\boldsymbol{Q}\sum_{l=1}^{\infty}q_l\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}(|\Delta_j\beta^l|^2-k)^2\|\boldsymbol{z}^j\|^2_{\mathbf{L}^2}\Big]\Big)^{\frac{1}{2}}\\&\leq Ck\Big(\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^j\|^2_{\mathbf{L}^2}\Big]\Big)^{\frac{1}{2}}\\&\leq Ck+Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j-1}}\|\boldsymbol{z}^j\|^2_{\mathbf{L}^2}\Big].\end{aligned}$$

For the terms corresponding to $\widetilde{IV}_{B,2}$, we use $\|\boldsymbol{u}\times\boldsymbol{v}\|_{\mathbf{L}^2}\leq C\|\boldsymbol{u}\|_{\mathbf{L}^2}\|\boldsymbol{v}\|_{\mathbf{L}^\infty}$ together with Assumption A_1 to obtain

$$\begin{aligned}|\widetilde{IV}_{B,2}|&\leq\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\frac{1}{2}\Big|\Big(\sum_{l=1}^{\infty}q_l(|\Delta_j\beta^l|^2-k)\Big(((\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\times\mathbf{e}_l)\times\mathbf{e}_l,\boldsymbol{z}^j\Big)\Big|\\&\leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\sum_{l=1}^{\infty}q_l\Big||\Delta_j\beta^l|^2-k\Big|\|((\boldsymbol{m}^{j+1}-\boldsymbol{m}^j)\times\mathbf{e}_l)\times\mathbf{e}_l\|_{\mathbf{L}^2}\|\boldsymbol{z}^j\|_{\mathbf{L}^2}\\&\leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\sum_{l=1}^{\infty}q_l\Big||\Delta_j\beta^l|^2-k\Big|\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}\|\boldsymbol{z}^j\|_{\mathbf{L}^2}.\end{aligned}$$

Thus summation over $j=0,\dots,J-1$, taking expectations, and using Hölder's inequality, as well as $\mathbb{E}[\big||\Delta_j\beta^l|^2-k\big|^4]\leq Ck^4$, since $\Delta_j\beta^l\sim\mathcal{N}(0,k)$, yield

$$\mathbb{E}\Big[\sum_{j=0}^{J-1}|\widetilde{IV}_{B,2}|\Big]\leq C\sum_{j=0}^{J-1}\sum_{l=1}^{\infty}q_l\Big(\mathbb{E}\Big[\big||\Delta_j\beta^l|^2-k\big|^4\Big]\mathbb{E}\Big[\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|^4_{\mathbf{L}^2}\Big]\Big)^{\frac{1}{4}}\Big(\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^j\|^2_{\mathbf{L}^2}\Big]\Big)^{\frac{1}{2}}$$

$$\leq C\sum_{j=0}^{J-1}\sum_{l=1}^{\infty} q_l k\Big(\mathbb{E}\Big[\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}^4\Big]\Big)^{\frac{1}{2}} + C\sum_{j=0}^{J-1}\sum_{l=1}^{\infty} q_l k\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big].$$

By Hölder's inequality, Lemma 3.4 c), e) and $\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \leq \mathbb{1}_{\widetilde{\Omega}_{\kappa,j-1}}$, we may conclude that

$$\begin{aligned}
\mathbb{E}\Big[\sum_{j=0}^{J-1}|\widetilde{IV}_{B,2}|\Big] &\leq C\,\mathrm{Tr}\,\boldsymbol{Q}k^{\frac{1}{2}}\Big(\sum_{j=0}^{J-1}\mathbb{E}\Big[\|\boldsymbol{m}^{j+1}-\boldsymbol{m}^j\|_{\mathbf{L}^2}^4\Big]\Big)^{\frac{1}{2}} + C\,\mathrm{Tr}\,\boldsymbol{Q}k\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big]\\
&\leq Ck^{\frac{1}{2}}\Big(\sum_{j=0}^{J-1}C\big(k^2\mathbb{E}\Big[\|\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2\Big]+\mathbb{E}\Big[\|\Delta_j\boldsymbol{W}\|_{\mathbf{L}^2}^4\Big]\big)\Big)^{\frac{1}{2}}\\
&\quad + Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big]\\
&\leq Ck + Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j-1}}\|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big].
\end{aligned}$$

Term $\widetilde{V}$: Let $r>0$ be fixed. For term $\widetilde{V}$, we arrive by using $\|\boldsymbol{u}\times\boldsymbol{v}\|_{\mathbf{L}^2}\leq C\|\boldsymbol{u}\|_{\mathbf{L}^2}\|\boldsymbol{v}\|_{\mathbf{L}^\infty}$ together with $\|\boldsymbol{m}^{j+\frac{1}{2}}\|_{\mathbf{L}^\infty}\leq 2$ $\mathbb{P}$-a.s. and Young's inequality at

$$\begin{aligned}
|\widetilde{V}| &= \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\Big|k\Big(\Big[\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}-\alpha\boldsymbol{m}^{j+\frac{1}{2}}\times(\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1})\Big]\times\Delta_j\boldsymbol{W},\boldsymbol{z}^{j+1}\Big)\Big|\\
&\leq C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\|\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}\|\Delta_j\boldsymbol{W}\|_{\mathbf{L}^\infty}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}\\
&\leq Ck\|\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^{\frac{2+r}{1+r}}\|\Delta_j\boldsymbol{W}\|_{\mathbf{L}^\infty}^{\frac{2+r}{1+r}} + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^{2+r}\\
&\leq Ck\|\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^{\frac{2+r}{1+r}}\|\Delta_j\boldsymbol{W}\|_{\mathbf{L}^\infty}^{\frac{2+r}{1+r}} + C\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}k\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^r.
\end{aligned}$$

Since $\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^r\leq C$ $\mathbb{P}$-a.s. holds, we obtain

$$\begin{aligned}
\mathbb{E}\Big[\sum_{j=0}^{J-1}|\widetilde{V}|\Big] &\leq Ck\Big(\sum_{j=0}^{J-1}\mathbb{E}\Big[\|\boldsymbol{m}^{j+\frac{1}{2}}\times\Delta\boldsymbol{m}^{j+1}\|_{\mathbf{L}^2}^2\Big]\Big)^{\frac{2+r}{2+2r}}\Big(\sum_{j=0}^{J-1}\mathbb{E}\Big[\|\Delta_j\boldsymbol{W}\|_{\mathbf{L}^\infty}^{2+\frac{4}{r}}\Big]\Big)^{\frac{r}{2+2r}}\\
&\quad + Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big]\\
&\leq Ck^{\frac{2+r}{2+2r}} + Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big]\\
&\leq Ck^{1-\tilde{\epsilon}} + Ck\sum_{j=0}^{J-1}\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big],
\end{aligned}$$

by using Hölder's inequality, Lemma 3.4 c), Remark 2.2, and setting $\tilde{\epsilon}:=\frac{r}{2+2r}>0$. For small choices of $r>0$, $\tilde{\epsilon}$ gets small, and we obtain a rate close to one.

Term $\widetilde{VI}$: Finally, the estimates obtained for the term $\widetilde{VI}$ are similar to those for term $\widetilde{IV}$. Analogously, using $\boldsymbol{z}^{j+1}=(\boldsymbol{z}^{j+1}-\boldsymbol{z}^j)+\boldsymbol{z}^j$, we decompose term $\widetilde{VI}$ into two parts $\widetilde{VI}_A$ and $\widetilde{VI}_B$. For part $\widetilde{VI}_A$, we obtain using Young's inequality and Assumption A_1

$$|\widetilde{VI}_A|\leq C(\mathrm{Tr}\,\boldsymbol{Q}^{\frac{1}{2}})^2\sum_{l_1=1}^{\infty}\sum_{l_2\neq l_1}\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}}\sqrt{q_{l_1}}\sqrt{q_{l_2}}|\Delta_j\beta^{l_1}|^2|\Delta_j\beta^{l_2}|^2\|(\boldsymbol{m}^{j+\frac{1}{2}}\times\mathbf{e}_{l_2})\times\mathbf{e}_{l_1}\|_{\mathbf{L}^2}^2$$

$$+ \delta \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|_{\mathbf{L}^2}^2.$$

By using Assumption A_1, the estimate $\|\boldsymbol{u} \times \boldsymbol{v}\|_{\mathbf{L}^2} \leq C\|\boldsymbol{u}\|_{\mathbf{L}^2}\|\boldsymbol{v}\|_{\mathbf{L}^\infty}$ in addition with $\|\boldsymbol{m}^{j+\frac{1}{2}}\|_{\mathbf{L}^\infty}^2 \leq 2$ $\mathbb{P}$-a.s., independence of $\Delta_j\beta^{l_1}$ and $\Delta_j\beta^{l_2}$ for $l_1 \neq l_2$, and taking expectations and summing up over $j = 0, \ldots, J-1$ yield

$$\mathbb{E}\Big[\sum_{j=0}^{J-1} |\widetilde{VI}_A|\Big] \leq C(\operatorname{Tr} \boldsymbol{Q}^{\frac{1}{2}})^2 k + \delta \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big].$$

For the second part $\widetilde{VI}_B$, we use again $\boldsymbol{m}^{j+\frac{1}{2}} = \boldsymbol{m}^j + \frac{1}{2}(\boldsymbol{m}^{j+1} - \boldsymbol{m}^j)$ to conclude

$$\begin{aligned}
\widetilde{VI}_B &= \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \frac{1}{2} \sum_{l_1=1}^{\infty} \sum_{l_2 \neq l_1} \sqrt{q_{l_1}}\sqrt{q_{l_2}} \Delta_j\beta^{l_1} \Delta_j\beta^{l_2} \Big((\boldsymbol{m}^j \times \mathbf{e}_{l_2}) \times \mathbf{e}_{l_1}, \boldsymbol{z}^j\Big) \\
&\quad + \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \frac{1}{4} \sum_{l_1=1}^{\infty} \sum_{l_2 \neq l_1} \sqrt{q_{l_1}}\sqrt{q_{l_2}} \Delta_j\beta^{l_1} \Delta_j\beta^{l_2} \Big(((\boldsymbol{m}^{j+1} - \boldsymbol{m}^j) \times \mathbf{e}_{l_2}) \times \mathbf{e}_{l_1}, \boldsymbol{z}^j\Big) \\
&=: \widetilde{VI}_{B,1} + \widetilde{VI}_{B,2}.
\end{aligned}$$

Similar arguments as in $\widetilde{IV}_{B,1}$ and $\widetilde{IV}_{B,2}$ yield

$$\mathbb{E}\Big[\sup_{r=0,\ldots,J-1} \sum_{j=0}^{r} \widetilde{VI}_{B,1}\Big] \leq C \operatorname{Tr} \boldsymbol{Q}^{\frac{1}{2}} k + C \operatorname{Tr} \boldsymbol{Q}^{\frac{1}{2}} k \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j-1}} \|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big]$$

for term $\widetilde{VI}_{B,1}$, and for $\widetilde{VI}_{B,2}$

$$\mathbb{E}\Big[\sum_{j=0}^{J-1} |\widetilde{VI}_{B,2}|\Big] \leq Ck + Ck \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j-1}} \|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big].$$

Combining the results obtained in the second step, we arrive at

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{r=0,\ldots,J-1} \sum_{j=0}^{r} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \Big(\int_{t_j}^{t_{j+1}} \boldsymbol{m}(s) \times \circ \,\mathrm{d}\boldsymbol{W}(s) - \boldsymbol{m}^{j+\frac{1}{2}} \times \Delta_j \boldsymbol{W}, \boldsymbol{z}^{j+1}\Big)\Big] \\
&\leq \delta \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big] + \widetilde{\delta}\mathbb{E}\Big[\sup_{j=0,\ldots,J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j-1}} \|\boldsymbol{z}^j\|_{\mathbf{L}^2}^2\Big] \\
&\quad + Ck \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big] + Ck + Ck^{1-\tilde{\epsilon}}.
\end{aligned} \tag{4.28}$$

Third Step: *Gronwall argument*
Finally combining the results (4.27) and (4.28) from the previous steps, we arrive at

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{j=0,\ldots,J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + \sum_{j=0}^{J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1} - \boldsymbol{z}^j\|_{\mathbf{L}^2}^2 + \alpha k \sum_{j=0}^{J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big] \\
&\leq Ck + Ck^{1-\tilde{\epsilon}} + Ck(1+\kappa) \sum_{j=0}^{J-1} \mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big],
\end{aligned}$$

where we have absorbed the first two terms of the right-hand-side of inequality (4.28). The discrete Gronwall inequality finally leads us to

$$\mathbb{E}\Big[\sup_{j=0,\dots,J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2 + \sum_{j=0}^{J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}-\boldsymbol{z}^j\|_{\mathbf{L}^2}^2 + \alpha k \sum_{j=0}^{J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\nabla \boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big] \leq C\exp(Ct_J\kappa)k^{1-\tilde{\epsilon}}.$$

Set $\kappa := \log(k^{-\frac{\epsilon}{2}})$ for a $\epsilon > 0$ and $\tilde{\epsilon} := \frac{\epsilon}{2}$. Thus using $\widetilde{\Omega}_{\kappa,J-1} \subset \widetilde{\Omega}_{\kappa,j}\ \forall j = 0,\dots,J-1$, we obtain

$$\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_{\kappa,J-1}} \sup_{j=0,\dots,J-1} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big] \leq \mathbb{E}\Big[\sup_{j=0,\dots,J-1} \mathbb{1}_{\widetilde{\Omega}_{\kappa,j}} \|\boldsymbol{z}^{j+1}\|_{\mathbf{L}^2}^2\Big] \leq Ck^{1-\epsilon}.$$

Define $\widetilde{\Omega}_k := \widetilde{\Omega}_{\kappa,J-1}$. Thus

$$\widetilde{\Omega}_k = \Big\{\omega\in\Omega;\ \sup_{j=0,\dots,J-1} \|\nabla \boldsymbol{m}^j\|_{\mathbf{L}^2}^4 + \sup_{s\in[0,t_{J-1}]} \|\nabla \boldsymbol{m}(s)\|_{\mathbf{L}^2}^4 + \sup_{s\in[0,t_{J-1}]} \|\nabla \boldsymbol{m}(s)\|_{\mathbf{W}^{1,2}}^2 \leq \log(k^{-\frac{\epsilon}{2}})\Big\},$$

and there holds

$$\begin{aligned}\mathbb{P}\Big[\widetilde{\Omega}_{\kappa,J-1}\Big] &\geq 1 - \frac{\mathbb{E}\Big[\sup\|\nabla \boldsymbol{m}^j\|_{\mathbf{L}^2}^4\Big] + \mathbb{E}\Big[\sup\|\nabla \boldsymbol{m}(s)\|_{\mathbf{L}^2}^4\Big] + \mathbb{E}\Big[\sup\|\nabla \boldsymbol{m}(s)\|_{\mathbf{W}^{1,2}}^2\Big]}{\log(k^{-\frac{\epsilon}{2}})} \\ &\geq 1 + \frac{C}{\epsilon\log(k)}\end{aligned}$$

due to Proposition 3.2 and Lemma 3.4. □

5. Computational studies

In this chapter we focus on numerical studies concerning the strong and weak convergence order of the time discretized scheme given in Algorithm 1.1 as well as the performance of different nonlinear solvers for different types of noise. Computational studies, using this scheme, are already presented in [BBNP13, BBP13], where possible blow-up and switching behavior in two space dimensions is studied. However, detailed numerical studies about convergence behavior are not reported in [BBNP13, BBP13].

For spatial discretization, the finite element space $\mathbf{V}_h \subset \mathbf{W}^{1,2}(D;\mathbb{R}^3)$ is defined by

$$\mathbf{V}_h := \Big\{ \mathbf{\Phi} \in \mathbf{C}(\bar{D};\mathbb{R}^3);\ \mathbf{\Phi}_{|K} \in \mathbf{P}_1(K;\mathbb{R}^3)\ \forall K \in \mathcal{T}_h \Big\},$$

where $\mathcal{T}_h$ denotes a regular triangulation of D into intervals K with a maximum mesh-size $h := \max\{\mathrm{diam}(K);\ K \in \mathcal{T}_h\} > 0$. For each element $K \in \mathcal{T}_h$, let $\mathbf{P}_1(K;\mathbb{R}^3)$ denote the space of all $\mathbb{R}^3$-valued polynomials of degree one.
The nodal interpolation operator $\mathcal{I}_h : \mathbf{C}(\bar{D};\mathbb{R}^3) \to \mathbf{V}_h$ is defined by

$$\mathcal{I}_h[\mathbf{\Phi}](x_r) := \mathbf{\Phi}(x_r) \qquad \forall \mathbf{\Phi} \in \mathbf{C}(\bar{D};\mathbb{R}^3),$$

for all nodes $\{x_r;\ r = 1,\dots,R\}$. The bilinear form $(\cdot,\cdot)_h : \mathbf{C}(\bar{D};\mathbb{R}^3) \times \mathbf{C}(\bar{D};\mathbb{R}^3) \to \mathbb{R}$ is defined via

$$(\mathbf{\Phi},\mathbf{\Xi})_h := \int_D \mathcal{I}_h[\langle \mathbf{\Phi}(x), \mathbf{\Xi}(x)\rangle_{\mathbb{R}^3}]\,\mathrm{d}x = \sum_{r=1}^{R} \beta_r \langle \mathbf{\Phi}(x_r), \mathbf{\Xi}(x_r)\rangle_{\mathbb{R}^3} \qquad \forall \mathbf{\Phi},\mathbf{\Xi} \in \mathbf{C}(\bar{D};\mathbb{R}^3),$$
$$\|\mathbf{\Phi}\|_h^2 := (\mathbf{\Phi},\mathbf{\Phi})_h \qquad \forall \mathbf{\Phi} \in \mathbf{C}(\bar{D};\mathbb{R}^3)$$

for certain weights $\beta_r > 0$. The mapping $\|\cdot\|_h$ is a norm on $\mathbf{V}_h$. The discrete Laplace operator $\widetilde{\Delta}_h : \mathbf{V}_h \to \mathbf{V}_h$ is defined by

$$-(\widetilde{\Delta}_h \mathbf{\Phi}_h, \mathbf{\Xi}_h)_h = (\nabla \mathbf{\Phi}_h, \nabla \mathbf{\Xi}_h)_{\mathbf{L}^2} \qquad \forall \mathbf{\Phi}_h, \mathbf{\Xi}_h \in \mathbf{V}_h.$$

The finite element formulation of Algorithm 1.1 reads as follows.

Algorithm 5.1
Let $\boldsymbol{M}^0 = \mathcal{I}_h[\boldsymbol{m}_0]$. For every $j = 0,\dots,J-1$ and $\Delta_j \boldsymbol{W} := \boldsymbol{W}(t_{j+1}) - \boldsymbol{W}(t_j) \sim \mathcal{N}(0, k\boldsymbol{Q})$ determine the $\mathbf{V}_h$-valued random variable $\boldsymbol{M}^{j+1}$, such that $\mathbb{P}$-a.s.

$$\Big(\boldsymbol{M}^{j+1} - \boldsymbol{M}^j, \mathbf{\Phi}\Big)_h + \alpha k \Big(\boldsymbol{M}^{j+\frac{1}{2}} \times [\boldsymbol{M}^{j+\frac{1}{2}} \times \widetilde{\Delta}_h \boldsymbol{M}^{j+1}], \mathbf{\Phi}\Big)_h$$
$$- k\Big(\boldsymbol{M}^{j+\frac{1}{2}} \times \widetilde{\Delta}_h \boldsymbol{M}^{j+1}, \mathbf{\Phi}\Big)_h = \Big(\boldsymbol{M}^{j+\frac{1}{2}} \times \Delta_j \boldsymbol{W}, \mathbf{\Phi}\Big)_h \qquad \forall \mathbf{\Phi} \in \mathbf{V}_h.$$

We consider a simple fixed point method (SFM), Newtons method (DNM), where the Jacobian is approximated by finite differences, and a modification of Powell's hybridmethod (PHM) in order to solve the nonlinear equation in Algorithm 5.1. For a more detailed description of these methods, we refer to [BBNP13] for SFM and DNM, and to [Bre73] for PHM.
First, different effects according to the different solvers are studied by simulating one path of Algorithm 5.1 on $D = (0, 1)$ with periodic boundary conditions up to final time $T = 5$. The initial value $\boldsymbol{m}_0(x) := (0, \sin(2\pi x), \cos(2\pi x))^T \in \mathbb{S}^2$, the damping parameter $\alpha = 0.25$, the noise intensity $\nu = 1.0$, and the spatial discretization parameter $h = 0.05$ are fixed. For all methods, the thresholding condition is set to be $\texttt{TOL} = 10^{-10}$, and the maximal amount of iterations is set to be 10.000.

Table 5.1. Average, minimum and maximum amount of iterations to perform one time iteration. Different types of noise, colored noise (cn) and approximate white noise (st), are compared with the deterministic case (det). The elapsed time (in seconds) to simulate one path with timestep size $k = 2^{-14}$ is compared.

	k	2^{-14}		2^{-13}		2^{-12}		2^{-11}		2^{-10}		2^{-9}		time
		av	iter	av	iter	av	iter	av	iter	av	iter	av	iter	
	det	1.3	1-2	1.3	1-3	1.4	1-3	1.5	1-3					9.7
SFM	cn	3.0	3-4	4.0	3-5	6.0	4-8	20.2	5-28					26.5
	st	5.0	5-5	7.0	6-7	11.2	10-12	39.2	32-44					35.8
	det	1.5	1-2	1.5	1-3	1.6	1-4	1.6	1-4	1.7	1-33			46.5
DNM	cn	2.0	2-3	2.0	2-3	2.0	2-3	2.1	2-3	2.7	2-3	3.0	2-3	73.8
	st	3.0	2-3	3.0	3-3	3.0	3-4	3.0	3-4	3.3	3-4	4.0	3-4	94.9
	det	1.5	1-4	1.5	1-6	1.6	1-7	1.7	1-6	1.7	1-10	1.8	1-16	42.4
PHM	cn	2.7	2-5	3.0	2-4	3.2	3-4	3.9	3-5	4.5	4-7	5.3	4-9	54.7
	st	4.0	3-5	4.7	4-5	5.5	5-6	6.9	6-8	8.3	7-10	10.0	8-12	54.7

Table 5.1 shows the average, the minimum and the maximum amount of iterations to compute one timestep of the path for the three different methods for different timestep sizes k. Here one can observe, that the SFM needs more iteration steps, but has the advantage to be faster compared to both, DNM and PHM. However, SFM has the disadvantage, that it does not converge for larger timestep sizes, e.g. $k = 2^{-10}$ or $k = 2^{-9}$, while DNM and PHM does. The DNM and PHM even work for larger timestep sizes $k = 2^{-8}, 2^{-7}, 2^{-6}, 2^{-5}, 2^{-4}$. Another observation is, that due to noise, the amount of iteration steps as well as the simulation time increase. This increase is for SFM significantly more pronounced compared to DNM and PHM.

In the following convergence studies, PHM is used, since it allows larger timestep size and it is almost twice as fast as DNM. Here, one considers equation (1.1) on $D = (0, 1)$ with periodic boundary conditions up to final time $T = 2$. Again, the initial value $\boldsymbol{m}_0(x) := (0, \sin(2\pi x), \cos(2\pi x))^T \in \mathbb{S}^2$ for all $x \in D$, the damping parameter is set to be $\alpha = 0.25$, and the maximal mesh-size $h = 0.05$ is fixed.

To compute the convergence order for the time discretization of Algorithm 1.1, different approximations $\{\boldsymbol{M}_i^j;\, j = 0, \dots, J\}$ for different timestep sizes $k_i = 2^{-(8+i)}$ for $i \in \{1, 2, 3, 4\}$

are simulated and compared with a finer simulated approximation $\{\boldsymbol{M}^j_{\text{ex}};\, j = 0, \dots, J\}$ using timestep size $k_{\text{ex}} = 2^{-14}$. The strong and weak errors are defined by

$$\boldsymbol{se}(i) := \Big(\mathbb{E}\Big[\max_{j=0,\dots,J} \|\boldsymbol{M}^j_{\text{ex}} - \boldsymbol{M}^j_i\|^2_{\mathbf{L}^2}\Big]\Big)^{\frac{1}{2}},$$

$$\boldsymbol{we}(i) := \Big(\max_{j=0,\dots,J} \|\mathbb{E}[\boldsymbol{M}^j_{\text{ex}}] - \mathbb{E}[\boldsymbol{M}^j_i]\|^2_{\mathbf{L}^2}\Big)^{\frac{1}{2}}.$$

In all simulations, the amount of 3000 paths is used to approximate the expectation values. One path of the solution using colored noise described below is illustrated in Figure 5.1.

5.1. Colored noise

The colored noise term is simulated by

$$\Delta_j \boldsymbol{W}(x) = \nu \sum_{l=1}^{L} \sum_{r=1}^{3} \sqrt{q_{3(l-1)+r}}\sqrt{2}\sin(\pi l x)\mathbf{e}_r \Delta_j \beta^{3(l-1)+r} \qquad \forall x \in D, \tag{5.1}$$

where $\Delta_j \beta^z$ are independent increments of a $\mathbb{R}$-valued Wiener process for $z = 1, \dots, 3L$, and $\mathbf{e}_r$ for $r = 1, 2, 3$ is a basis of $\mathbb{R}^3$. The parameter $\nu = 1.0$ is fixed. Figure 5.2 shows the convergence behavior using the sequence $q_r = r^{-2.5}$ and fix $L = 20$. This simulation fits to the setting considered in this work. The experimental strong order is 0.44, and the experimental weak order is 1.07, which is close to the results obtained in Theorem 4.1. Increasing parameter L does not change the convergence behavior.

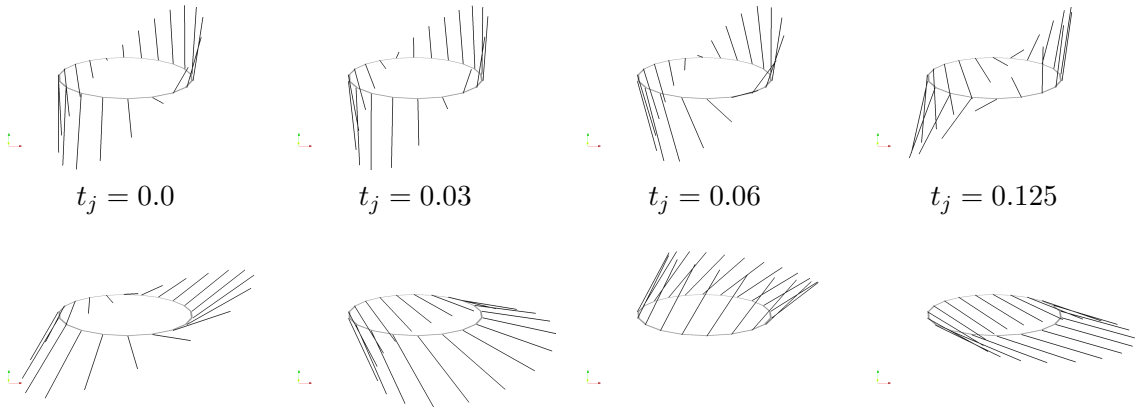
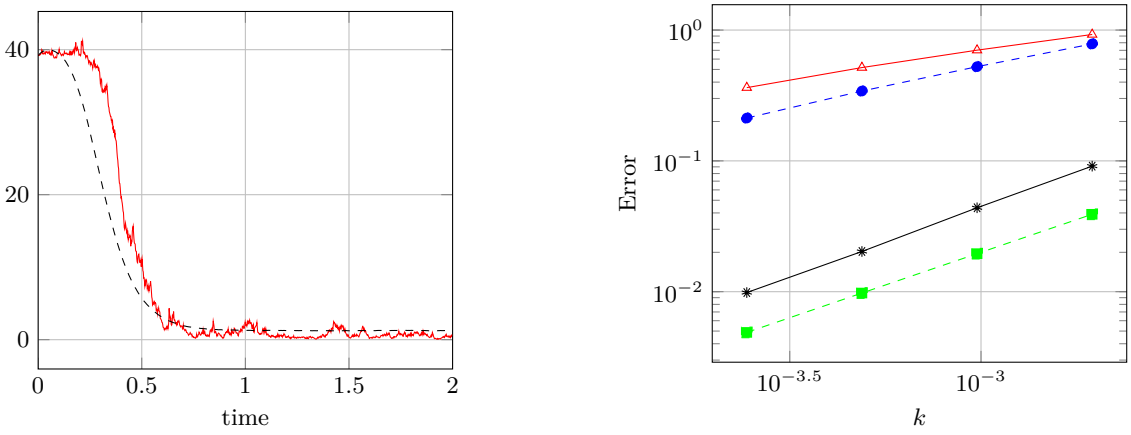


Figure 5.1. One trajectory $t_j \mapsto \boldsymbol{M}^j(\omega)$ of the solution of equation (1.1) with colored noise.



(a) $t_j \mapsto \|\nabla \boldsymbol{M}^j(\omega)\|_{\mathbf{L}^2}$ **(b)** rates of convergence

Figure 5.2. Colored noise where $q_r = r^{-2.5}$ and $L = 20$: The time evolution of $t_j \mapsto \|\nabla \boldsymbol{M}^j(\omega)\|_{\mathbf{L}^2}$ (——) and corresponding expectation value (- - -) is shown in **(a)**. Strong (—△—) and weak (—✳—) rates of convergence and the reference slopes 0.5 (- ● -) and 1.0 (- ■ -) are illustrated in **(b)**.

5.2. Approximate white noise

Experiments corresponding to space-time white noise, which is motivated physically are considered. However, mathematical results for equation (1.1) with space-time white noise are currently not available, even in the case of one-dimensional domains.
A similar approach as in [BBNP13, BBP13] (without adaptive mesh refinement) is considered, where the space-time white noise is approximated using

$$\Delta_j \boldsymbol{W}(x) := \nu \sum_{r=1}^{R} \frac{\boldsymbol{\Phi}_r(x)}{\sqrt{\frac{1}{2}|\mathrm{supp}\boldsymbol{\Phi}_r|}} \Delta_j \beta^r \qquad \forall x \in D, \tag{5.2}$$

where $\boldsymbol{\Phi}_r(x)$ are the basis functions of $\mathbf{V}_h$, and $\Delta_j \beta^r$ are independent increments of the $\mathbb{R}$-valued Wiener process β^r for $r = 1, \ldots, R$. The parameter $\nu = 0.5$ is fixed. Note that, as it is depicted in [BSDDM05], approximation (5.2) has a finite numerical correlation in space. An additional adaptive mesh refinement would be necessary to capture effects of space-time white noise more correctly. Figure 5.4 illustrates the obtained experimental strong convergence order of 0.34 and an experimental weak convergence order of 0.98.

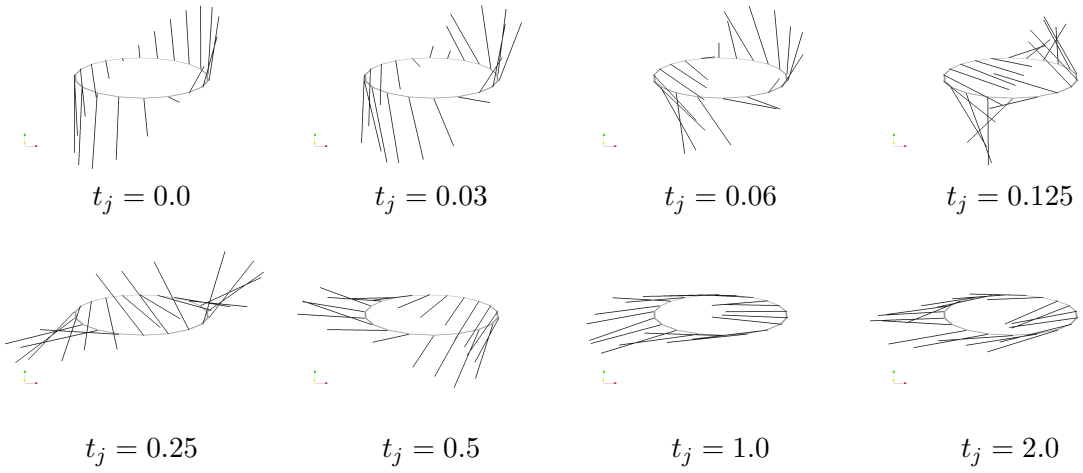
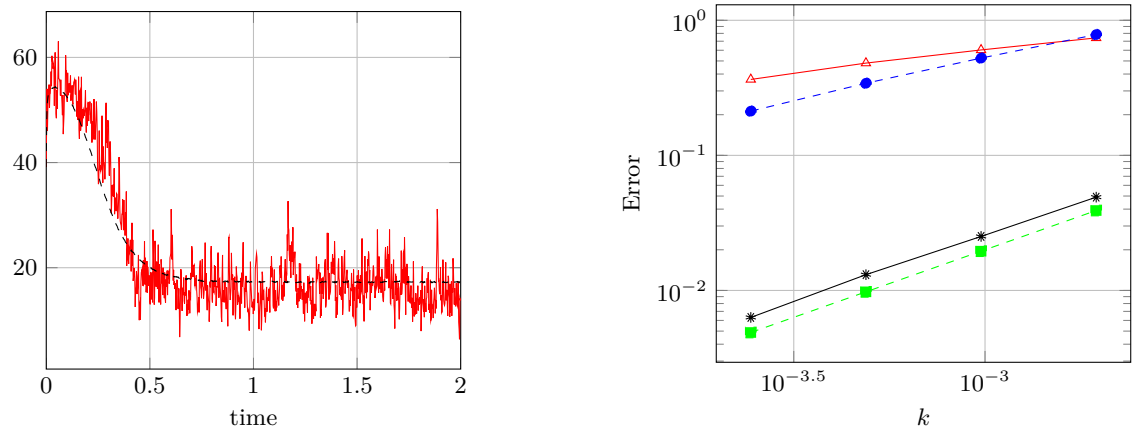


Figure 5.3. One trajectory $t_j \mapsto \boldsymbol{M}^j(\omega)$ of the solution of equation (1.1) with approximate white noise.



(a) $t_j \mapsto \|\nabla \boldsymbol{M}^j(\omega)\|_{\mathbf{L}^2}$ **(b)** rates of convergence

Figure 5.4. Approximate white noise: The time evolution of $t_j \mapsto \|\nabla \boldsymbol{M}^j(\omega)\|_{\mathbf{L}^2}$ (——) and corresponding expectation value (- - -) is shown in **(a)**. Strong (—△—) and weak (—✷—) rates of convergence and the reference slopes 0.5 (-●-) and 1.0 (-■-) are illustrated in **(b)**.

Part II.

The forward-backward stochastic heat equation: numerical analysis and simulation

6. Introduction

Let $D \subset \mathbb{R}^d$ be a bounded domain with C^2 boundary, and $0 < T < \infty$ be the terminal time. The aim of this part[1] of the thesis is to formulate and study a fully implementable scheme to simulate the backward stochastic heat equation ($0 \le t \le T$)

$$\mathrm{d}Y(t) = \Big[-\Delta Y(t) - \sum_{i=1}^{n} \nu^i(t) Z^i(t) - F(t)\Big]\,\mathrm{d}t + \sum_{i=1}^{n} Z^i(t)\,\mathrm{d}W^i(t), \qquad Y(T) = \Psi, \tag{6.1}$$

with homogeneous Dirichlet boundary conditions. Here, $\{W(t);\, t \in [0,T]\}$ denotes an $\mathbb{R}^n$-valued Wiener process $W(t) = (W^1(t), \dots, W^n(t))$, and F an external free forcing term, both of it supported on the stochastic basis $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$.
This backward stochastic partial differential equation (BSPDE) serves as a prototype example for more general cases where the drift involves a linear second order elliptic operator (see [DT12]) to which most of the results below apply as well. While of independent interest, the BSDPE (6.1) also appears as first order adjoint equation in the first order optimality conditions of the following prototype linear-convex stochastic optimal control problem (see also [Ben83]): given a deterministic profile $\{\widetilde{X}(t,\cdot);\, t \in [0,T]\}$, find an $\mathbb{F}$-adapted process $\{U(t,\cdot);\, t \in [0,T]\}$ which minimizes the cost functional

$$\mathcal{J}(X,U) = \mathbb{E}\Big[\frac{1}{2}\int_0^T \Big(\|X(t) - \widetilde{X}(t)\|_{\mathbb{L}^2}^2 + \alpha\|U(t)\|_{\mathbb{L}^2}^2\Big)\,\mathrm{d}t + g(X(T))\Big] \qquad (\alpha > 0), \tag{6.2}$$

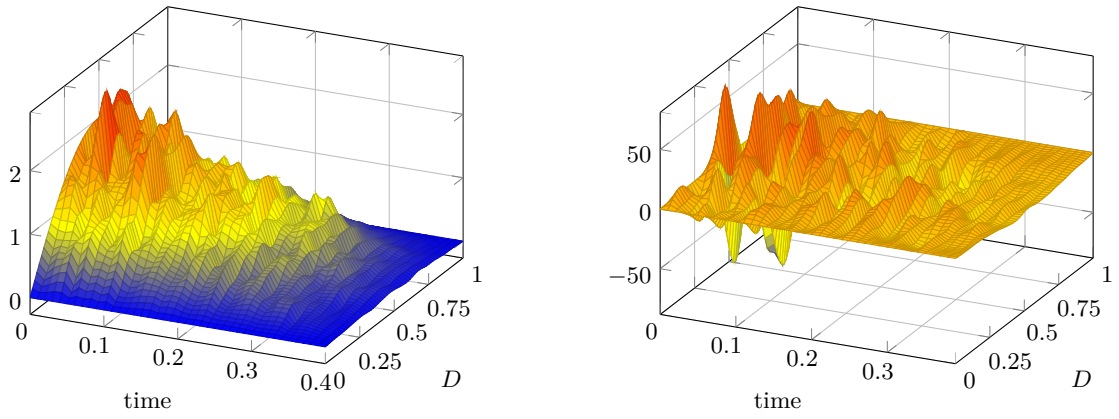
subject to the controlled (forward) stochastic heat equation ($0 \le t \le T$)

$$\mathrm{d}X(t) = \Big[\Delta X(t) + U(t)\Big]\,\mathrm{d}t + \sum_{i=1}^{n} \nu^i(t) X(t)\,\mathrm{d}W^i(t), \qquad X(0) = x_0, \tag{6.3}$$

with homogeneous Dirichlet boundary conditions. This part of the thesis is a first step to access this problem numerically by dealing with the adjoint equation (6.1) as part of the corresponding optimality system (6.3), (7.3)–(7.4). Figure 6.1 displays one approximate path of the optimal control $U^\star$ as well as the related optimal state $X^\star$; see Section 10.2 for further details.

Backward stochastic partial differential equations (BSPDEs for short) naturally extend backward stochastic differential equations (BSDEs for short), which is a subject of increasing interest in the last two decades. Various numerical methods to approximate BSDEs have been developed: references for procedures which exploit the connection of BSDEs and corresponding deterministic (parabolic) PDEs include [DMP96, MT06], which are efficient for low-dimensional problems. A different approach is realized in [BT04, Zha04], where

[1] Part II is based on the manuscript [DP15] accepted for publication in SIAM Journal on Scientific Computing. Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.



(a) $t_j \mapsto X_h^{\star,j}(\omega)$ **(b)** $t_j \mapsto U_h^{\star,j}(\omega)$

Figure 6.1. One path of the optimal state $X_h^{\star,j}$ and the optimal control $U_h^{\star,j}$ for (6.2)–(6.3), see Setup **D** in Table 10.3 for further details.

the backward equation is discretized directly by using an explicit/implicit time stepping scheme which involves conditional expectations. In order to compute the occurring conditional expectations, several techniques have been developed: methods based on Malliavin calculus [BT04], cubature method [CM12] or the least squares Monte-Carlo method; see e.g. [BD07, GLW05, GT14]. A combination of the implicit Euler method and the least squares Monte-Carlo method is used in this work to simulate the BSPDE (6.1), and the first order optimality conditions of the stochastic optimal control problem (6.2)–(6.3), which is a forward-backward stochastic partial differential equation (FBSPDE); see (6.3), (7.3)–(7.4).

From a numerical viewpoint, the computational setup and resources required to approximate the solution (Y, Z) of a BSPDE such as (6.1) is substantially more demanding than simulating a SPDE. Even further, computational demands increase to approximate a triple (X, Y, Z) of ($\mathbb{F}$-adapted) processes which solve a FBSPDE such as (6.3), (7.3)–(7.4). In Section 8 we consider the spatial discretization based on $\mathbb{P}_1$-finite elements for both problems; see equations (8.1) and (8.4)–(8.5). The main theoretical results are on optimal strong rates of convergence with respect to the (spatial) mesh parameter $h > 0$: see Theorem 8.2 for the BSPDE (6.1), whose proof uses (simple) variational arguments, resting on improved regularity properties of the solution of (6.1), Itô's formula, and approximation results for the finite element method; and Theorem 8.4 for the FBSPDE, where we use Theorem 8.2 in combination with a contraction argument to show strong convergence with optimal order for Y as well as Z for short time durations $T > 0$.

In the second part, we discuss fully implementable algorithms to simulate the BSPDE (6.1) resp. the FBSDPE (6.3), (7.3)–(7.4): an algorithm is proposed for problem (6.1), which is based on the implicit Euler method for the BSDE (8.1), and the one-step forward dynamic programming scheme ((ODP) for short); see Scheme 9.1. This scheme requires to compute conditional expectations $\mathbb{E}[\,\cdot\,|\mathcal{F}_{t_j}]$ in each time step, which is the reason for the significant computational complexity of the problem. For its approximation we use the representation $(Y_h^j, Z_h^{i,j}) = (\mathcal{Y}_h^j(X_h^j), \mathcal{Z}_h^{i,j}(X_h^j))$ of iterates of Scheme 9.1 with the help of deterministic

functions $\mathcal{Y}_h^j(\cdot)$ and $\mathcal{Z}_h^{i,j}(\cdot)$, which is evaluated on the iterates of the (forward) explanatory process $\{X_h^j;\, j = 0, \dots, J\}$. Those functions will be computed via conditional expectations of the form $\mathbb{E}[\,\cdot\,|X_h^j]$, which in turn will be approximated with the help of the partitioning estimation method as a special case of the least squares Monte-Carlo method. This method partitions the image $X_h^j[\Omega]$ into R regions, which here is a set of finite element functions, and then approximates $\mathbb{E}[\,\cdot\,|X_h^j]$ by a function which for any given realization $X_h^j(\omega)$ returns the sample mean for the corresponding region in which $X_h^j(\omega)$ is contained. The choice of how to construct such a partitioning in crucial: for small systems of BSDEs, usually a uniform mesh (so-called hypercubes **(HC)**) is used, see e.g. [GLW05, GT14]. However, this strategy is not practicable for the approximation of a BSPDE as (6.1) where a high dimensional state space is involved. To resolve $X_h^j[\Omega]$ more accurately, two further partitioning methods are discussed where the partitioning is chosen according to the sampling for X_h^j:

- the Voronoi partition method **(V)**, which uses R (additional) realizations of X_h^j in combination with nearest neighbor clustering;
- the Binary Tree Cuboids **(BTC)**, where every region is equally visited by X_h^j.

To realize both (a-posteriori) approaches only requires to be able to simulate X_h^j, while the explicit knowledge of the distribution of X_h^j is not needed; see Chapter 9 for a detailed description. Computational studies which quantify different approximation effects throughout the simulation of BSPDEs are carried out in Chapter 10. The studies support the convergence rates of Theorem 8.2, and show that the implicit Euler method is exempted from a restrictive CFL condition, which is in turn needed for the explicit Euler method.

A series of computational studies motivate uniform dependence of the spatial mesh-size of the time discretization error of order $\mathcal{O}(\sqrt{k})$, which complements the theoretical result on BSDEs in [Zha04], where involved stability constants in related error estimates depend on the dimension of the state space. The given algorithmic setup is general and may be adopted to more general (nonlinear) BSPDEs. However, for restricted terminal data in (6.1), we propose a simpler scheme which avoids the computation of conditional expectations by exploiting linearity of the problem and thus leads to a significant speed-up in computation time; cf. Remark 9.3.

These results on how to simulate problem (6.1) provide the basis for a proper approximation of the FBSPDE (6.3), (7.3)–(7.4), and here we consider two schemes. The first combines Scheme 9.1 with a Picard type iteration as suggested in [BZ08]; see Scheme 9.4. We combine this scheme with the methods **(BTC)** and **(V)**, and compute related coefficients as well as the basis functions in $\mathcal{Y}_h^{(v-1),j}(\cdot)$ for the computation of $X_h^{(v),j}$ according to the law of $X_h^{(v-1),j}$, rather than $X_h^{(v),j}$ in (9.12). Computational studies confirm that Picard iterates only converge for short time durations $T > 0$. A significant reduction of iterations may be obtained when the related deterministic control problem is solved in a precursory step to provide good initial data; further computational savings may be achieved when the forward equation yielding $X_{h_F}^{(v),j}$ is discretized using a comparably rougher spatial discretization than for the adjoint equation yielding $(Y_{h_B}^{(v),j}, Z_{h_B}^{(v),i,j})$, i.e. $h_B < h_F$. In contrast to the Picard type iteration, the newly proposed stochastic gradient method exploits the structure of the approximation of the deterministic functions, and computational studies here suggest

convergence for general terminal times $T > 0$; see Scheme 9.5. The stochastic gradient method provides a general strategy to solve stochastic optimal control problems. For the restricted setting of stochastic LQ problems (where $g(\cdot)$ is quadratic in (6.2)) we may exploit linearity of (6.3) and use successive substitution of iterates to arrive at a semi-explicit discretization of the stochastic Riccati equation, which enables an explicit formula for optimal feedback control; cf. Remark 9.6. While of independent interest, we focus in the computational studies in Section 10.2 on the study of the stochastic gradient method because of its broader applicability.

This part of the thesis is organized as follows: We first collect preliminary results in Chapter 7, including a discussion to ensure existence of a strong solution for the BSPDE (6.1), and for the FBSPDE (6.3), (7.3)–(7.4). In Section 8.1, we focus on the spatial discretization of (6.1) using $\mathbb{P}_1$-finite elements, and a corresponding error analysis; see Theorem 8.2, while Section 8.2 is devoted to a space discretization and a corresponding error analysis for the FBSPDE (6.3), (7.3)–(7.4); see Theorem 8.4. Fully implementable algorithms to simulate problems (6.1) and (6.3), (7.3)–(7.4) as well as different partitioning approaches are detailed in Chapter 9. Computational studies are reported in Chapter 10.

7. Preliminaries

7.1. Notation

Let $\mathbb{K}$ be a separable Hilbert space. By $\|\cdot\|_{\mathbb{L}^2}$ resp. $(\cdot,\cdot)$ we denote the norm resp. the scalar product in $\mathbb{L}^2 := \mathbb{L}^2(D;\mathbb{R})$. The norm in $\mathbb{W}^{k,p} := \mathbb{W}^{k,p}(D;\mathbb{R})$ for $k = 1, 2$ is denoted by $\|\cdot\|_{\mathbb{W}^{k,2}}$. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ be a complete filtered probability space, where $\{\mathcal{F}_t\}_{t\in[0,T]}$ is the filtration generated by the n-dimensional Wiener process $W = \{W(t);\, t \in [0,T]\}$, augmented by all the $\mathbb{P}$-null sets. The σ-algebra of predictable sets on $\Omega \times [0,T]$ is denoted by $\mathcal{P}$. The space of all predictable $\mathbb{K}$-valued processes $X : \Omega \times [0,T] \to \mathbb{K}$ satisfying $\mathbb{E}\big[\int_0^T \|X(t)\|_{\mathbb{K}}^2 \,\mathrm{d}t\big] < \infty$ is denoted by $L^2_{\mathcal{P}}(\Omega; L^2(0,T;\mathbb{K}))$. The space of all predictable $\mathbb{K}$-valued continuous processes $X : \Omega \times [0,T] \to \mathbb{K}$ satisfying $\mathbb{E}\big[\sup_{t\in[0,T]} \|X(t)\|_{\mathbb{K}}^2\big] < \infty$ is denoted by $L^2_{\mathcal{P}}(\Omega; C([0,T];\mathbb{K}))$.

7.2. Strong solution of the backward stochastic heat equation

Let $\Psi \in L^2(\Omega;\mathcal{F}_T,\mathbb{W}_0^{1,2})$, $F \in L^2_{\mathcal{P}}(\Omega; L^2(0,T;\mathbb{L}^2))$, and $\nu^i \in L^\infty(0,T;\mathbb{W}^{1,\infty})$ for $i = 1,\dots,n$ be given. The strong solution of (6.1) is a pair of square integrable adapted processes (Y,Z) such that the analytically weak form of (6.1) is satisfied. The following theorem states its existence, uniqueness, and regularity, and is adapted from [DT12, Theorem 3.1].

Theorem 7.1
Let $0 < T < \infty$. There exists a unique strong solution

$$(Y,Z) \in L^2_{\mathcal{P}}(\Omega; C([0,T];\mathbb{W}_0^{1,2}) \cap L^2(0,T;\mathbb{W}_0^{1,2} \cap \mathbb{W}^{2,2})) \times L^2_{\mathcal{P}}(\Omega; L^2(0,T;\mathbb{W}_0^{1,2}(D;\mathbb{R}^n)))$$

of (6.1), i.e., for a.e. $(\omega,t) \in \Omega \times [0,T]$ there holds

$$\begin{aligned}\int_D Y(t)\phi \,\mathrm{d}x = \int_D \Psi\phi \,\mathrm{d}x - \int_t^T \int_D \nabla Y(s)\nabla\phi \,\mathrm{d}x\,\mathrm{d}s + \sum_{i=1}^n \int_t^T \int_D \nu^i(s) Z^i(s)\phi \,\mathrm{d}x\,\mathrm{d}s \\ - \int_t^T \int_D F(s)\phi \,\mathrm{d}x\,\mathrm{d}s - \sum_{i=1}^n \int_t^T \int_D Z^i(s)\phi \,\mathrm{d}x\,\mathrm{d}W^i(s) \qquad \forall \phi \in \mathbb{W}_0^{1,2},\end{aligned}$$

and there exists a constant $C \equiv C(D,T,\nu^i) > 0$ such that

$$\begin{aligned}&\mathbb{E}\Big[\sup_{t\in[0,T]} \|Y(t)\|_{\mathbb{W}^{1,2}}^2 + \int_0^T \Big(\|Y(t)\|_{\mathbb{W}^{2,2}}^2 + \sum_{i=1}^n \|Z^i(t)\|_{\mathbb{W}^{1,2}}^2\Big)\,\mathrm{d}t\Big] \\ &\le C\mathbb{E}\Big[\|\Psi\|_{\mathbb{W}^{1,2}}^2 + \int_0^T \|F(t)\|_{\mathbb{L}^2}^2 \,\mathrm{d}t\Big]. \qquad (7.1)\end{aligned}$$

7.3. Stochastic maximum principle for (6.2)–(6.3)

Let $x_0 \in L^2(\Omega;\mathcal{F}_0,\mathbb{W}_0^{1,2})$ be given, $T>0$, and $g \in C^2(\mathbb{L}^2,\mathbb{R})$ be bounded from below, convex, and there exists $C>0$ such that for all $x \in \mathbb{L}^2$

$$\|Dg(x)\|_{\mathbb{W}^{\alpha,2}} \le C(1+\|x\|_{\mathbb{W}^{\alpha,2}}) \quad (\alpha \in \{0,1\}) \quad \text{and} \quad \|D^2 g(x)\|_{\mathcal{L}(\mathbb{L}^2,\mathbb{L}^2)} \le C. \tag{7.2}$$

By [Ben83], the optimum of (6.2)–(6.3) $(X^\star,U^\star)$ with (unique) solution $(Y^\star,Z^\star)$ solves the state equation (6.3), the adjoint equation $(0 \le t \le T)$

$$\begin{aligned} \mathrm{d}Y^\star(t) &= \Big[-\Delta Y^\star(t) - \sum_{i=1}^n \nu^i(t) Z^{\star,i}(t) - \Big(X^\star(t) - \widetilde{X}(t)\Big)\Big]\,\mathrm{d}t + \sum_{i=1}^n Z^{\star,i}(t)\,\mathrm{d}W^i(t), \\ Y^\star(T) &= Dg\big(X^\star(T)\big), \end{aligned} \tag{7.3}$$

with homogeneous Dirichlet boundary conditions, and the maximum principle

$$0 = \alpha U^\star(t) + Y^\star(t). \tag{7.4}$$

Inserting (7.4) into the state equation (6.3) yields the FBSPDE which is denoted below by (FBSHE) and consists of (6.3), (7.3)–(7.4).

7.4. Strong solution of the forward-backward stochastic heat equation

Throughout this part of the thesis, let $\widetilde{X} \in L^2(0,T;\mathbb{L}^2)$ in (7.3) be given. Consider the Banach space

$$\mathcal{M}_{\mathbb{L}^2}[0,T] := \Big[L^2_{\mathcal{P}}(\Omega;C([0,T];\mathbb{L}^2))\Big]^2 \times L^2_{\mathcal{P}}(\Omega;L^2(0,T;\mathbb{L}^2(D;\mathbb{R}^n))),$$

which is endowed with the norm

$$\|(X,Y,Z)\|^2_{\mathcal{M}_{\mathbb{L}^2}[0,T]} := \mathbb{E}\Big[\sup_{t\in[0,T]}\|X(t)\|^2_{\mathbb{L}^2} + \sup_{t\in[0,T]}\|Y(t)\|^2_{\mathbb{L}^2}\Big] + \mathbb{E}\Big[\sum_{i=1}^n \int_0^T \|Z^i(t)\|^2_{\mathbb{L}^2}\,\mathrm{d}t\Big].$$

A triple $(X,Y,Z) \in \mathcal{M}_{\mathbb{L}^2}[0,T]$ is called strong solution of (FBSHE), if it satisfies $\mathbb{P}$-a.s. the analytically weak form of the system (FBSHE).

Our goal for the numerical treatment of the problem in Section 8.2 is to approximate its strong solution via (space-time approximations of) the following Picard iteration method: Let $Y^{(0)} \equiv 0$. Iterate $v = 1,2,\ldots$

$$\mathrm{d}X^{(v)}(t) = \Big[\Delta X^{(v)}(t) - \frac{1}{\alpha}Y^{(v-1)}(t)\Big]\,\mathrm{d}t + \sum_{i=1}^n \nu^i(t) X^{(v)}(t)\,\mathrm{d}W^i(t), \tag{7.5}$$

$$\mathrm{d}Y^{(v)}(t) = \Big[-\Delta Y^{(v)}(t) - \sum_{i=1}^n \nu^i(t) Z^{(v),i}(t) - \Big(X^{(v)}(t) - \widetilde{X}(t)\Big)\Big]\,\mathrm{d}t + \sum_{i=1}^n Z^{(v),i}(t)\,\mathrm{d}W^i(t), \tag{7.6}$$

$$X^{(v)}(0) = x_0, \qquad Y^{(v)}(T) = Dg(X^{(v)}(T)),$$

with homogeneous Dirichlet boundary conditions for $X^{(v)}$ and $Y^{(v)}$. For every $v \in \mathbb{N}$, the existence of a strong solution $X^{(v)} \in L^2_{\mathcal{P}}(\Omega; C([0,T];\mathbb{L}^2))$ of (7.5), resp. $(Y^{(v)}, Z^{(v)}) \in L^2_{\mathcal{P}}(\Omega; C([0,T];\mathbb{L}^2)) \times L^2_{\mathcal{P}}(\Omega; L^2(0,T;\mathbb{L}^2(D;\mathbb{R}^n)))$ e.g. follows from [Cho07, Chapter 6.7], resp. Theorem 7.1.

Theorem 7.2
Let $0 < T \leq \widetilde{T}_0$ be sufficiently small. Then $\{\Theta^{(v)} \equiv (X^{(v)}, Y^{(v)}, Z^{(v)});\, v \in \mathbb{N}\}$ of (7.5)–(7.6) converge to the unique strong solution $\Theta^\star \equiv (X^\star, Y^\star, Z^\star) \in \mathcal{M}_{\mathbb{L}^2}[0,T]$ of (FBSHE).

Proof
The proof is based on a fixed point argument for the scheme (7.5)–(7.6). We follow the arguments detailed in the proof of Theorem 8.3, by there replacing $\mathcal{M}_{\mathbb{V}_h}[0,T]$ with $\mathcal{M}_{\mathbb{L}^2}[0,T]$ and by considering instead the fixed point map $\mathcal{T} : \mathcal{M}_{\mathbb{L}^2}[0,T] \to \mathcal{M}_{\mathbb{L}^2}[0,T]$ for (7.5)–(7.6). Accordingly, we can establish the existence of a unique strong solution for small time durations $0 < T \leq \widetilde{T}_0$ by contraction property. Moreover, we obtain the following estimate with $0 < \overline{q} < 1$ similarly to (8.10),

$$\|\Theta^{(v)} - \Theta^\star\|^2_{\mathcal{M}_{\mathbb{L}^2}[0,T]} \leq \frac{\overline{q}^{v-1}}{1-\overline{q}} \|\Theta^{(2)} - \Theta^{(1)}\|^2_{\mathcal{M}_{\mathbb{L}^2}[0,T]}. \tag{7.7}$$

8. Spatial discretization and rates of convergence

Let $\mathcal{T}_h$ be a regular triangulation of $D \subset \mathbb{R}^d$ into intervals K with a maximum mesh size $h := \max\{\mathrm{diam}(K);\, K \in \mathcal{T}_h\}$. For each element $K \in \mathcal{T}_h$, let $\mathbb{P}_1(K)$ denote the set of all polynomials of degree less or equal to one. We define the finite element space $\mathbb{V}_h \subset \mathbb{W}_0^{1,2}$ by

$$\mathbb{V}_h := \{\phi \in C_0(\overline{D});\, \phi_{|K} \in \mathbb{P}_1(K)\ \forall K \in \mathcal{T}_h\}.$$

The $\mathbb{L}^2$-projection $\Pi_h : \mathbb{L}^2 \to \mathbb{V}_h$ is defined by

$$(\Pi_h \xi - \xi, \phi_h) = 0 \qquad \forall \phi_h \in \mathbb{V}_h,$$

and satisfies

$$\|\Pi_h \xi - \xi\|_{\mathbb{L}^2} + h\|\nabla[\Pi_h \xi - \xi]\|_{\mathbb{L}^2} \leq \begin{cases} Ch\|\nabla \xi\|_{\mathbb{L}^2} & \forall \xi \in \mathbb{W}_0^{1,2}, \\ Ch^2\|\nabla^2 \xi\|_{\mathbb{L}^2} & \forall \xi \in \mathbb{W}_0^{1,2} \cap \mathbb{W}^{2,2}. \end{cases}$$

The Ritz-projection $\mathcal{R}_h : \mathbb{W}_0^{1,2} \to \mathbb{V}_h$ is defined by

$$(\nabla[\mathcal{R}_h \xi - \xi], \nabla \phi_h) = 0 \qquad \forall \phi_h \in \mathbb{V}_h.$$

We define the discrete Laplace operator $\Delta_h : \mathbb{V}_h \to \mathbb{V}_h$ by

$$-(\Delta_h \xi_h, \phi_h) = (\nabla \xi_h, \nabla \phi_h) \qquad \forall \phi_h, \xi_h \in \mathbb{V}_h.$$

8.1. Backward stochastic heat equation

Suppose $F \equiv 0$ throughout this section. The spatially discrete version of (6.1) reads as follows: For all $t \in [0,T]$, there holds $\mathbb{P}$-a.s.

$$\begin{aligned} (Y_h(t), \phi_h) = (\Psi, \phi_h) - \int_t^T (\nabla Y_h(s), \nabla \phi_h)\, \mathrm{d}s + \sum_{i=1}^n \int_t^T (\nu^i(s) Z_h^i(s), \phi_h)\, \mathrm{d}s \\ - \sum_{i=1}^n \int_t^T (Z_h^i(s), \phi_h)\, \mathrm{d}W^i(s) \qquad \forall \phi_h \in \mathbb{V}_h. \end{aligned} \tag{8.1}$$

For every fixed $h > 0$, equation (8.1) is a linear backward stochastic differential equation, and existence and uniqueness of a strong solution $(Y_h, Z_h) \in L^2_{\mathcal{P}}(\Omega; C([0,T]; \mathbb{V}_h)) \times L^2_{\mathcal{P}}(\Omega; L^2(0,T; \mathbb{V}_h^n))$ e.g. follow from [MY07, Chapter 1, Theorem 4.2]. In particular, the martingale representation theorem states that the processes Z_h^i have to be $\mathbb{V}_h$-valued.

Lemma 8.1
The strong solution (Y_h, Z_h) of (8.1) satisfies

$$\text{(i)} \sup_{t\in[0,T]} \mathbb{E}\Big[\|Y_h(t)\|_{\mathbb{L}^2}^2\Big] + \mathbb{E}\Big[\int_0^T \Big(\|\nabla Y_h(t)\|_{\mathbb{L}^2}^2 + \sum_{i=1}^n \|Z_h^i(t)\|_{\mathbb{L}^2}^2\Big)\,\mathrm{d}t\Big] \le C\mathbb{E}\Big[\|\Psi\|_{\mathbb{L}^2}^2\Big];$$

$$\text{(ii)} \sup_{t\in[0,T]} \mathbb{E}\Big[\|\nabla Y_h(t)\|_{\mathbb{L}^2}^2\Big] + \mathbb{E}\Big[\int_0^T \Big(\|\Delta_h Y_h(t)\|_{\mathbb{L}^2}^2 + \sum_{i=1}^n \|\nabla Z_h^i(t)\|_{\mathbb{L}^2}^2\Big)\,\mathrm{d}t\Big] \le C\mathbb{E}\Big[\|\nabla\Psi\|_{\mathbb{L}^2}^2\Big],$$

where $C \equiv C(D,T) > 0$ does not depend on $h > 0$.

Proof
Statement (i): Fix $t \in [0,T]$. Using Itô's formula with $x \mapsto \|x\|_{\mathbb{L}^2}^2$ for equation (8.1) yields $\mathbb{P}$-a.s.

$$\begin{aligned}\|Y_h(T)\|_{\mathbb{L}^2}^2 - \|Y_h(t)\|_{\mathbb{L}^2}^2 &= 2\int_t^T \|\nabla Y_h(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s - 2\sum_{i=1}^n \int_t^T \big(\Pi_h[\nu^i(s)Z_h^i(s)], Y_h(s)\big)\,\mathrm{d}s \\ &\quad + 2\sum_{i=1}^n \int_t^T \big(-Z_h^i(s), Y_h(s)\big)\,\mathrm{d}W^i(s) + \sum_{i=1}^n \int_t^T \|Z_h^i(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s.\end{aligned}$$

By taking expectations, using Gronwall's inequality and stability of Π_h, we arrive at

$$\mathbb{E}\Big[\|Y_h(t)\|_{\mathbb{L}^2}^2\Big] + \mathbb{E}\Big[\sum_{i=1}^n \int_t^T \|Z_h^i(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big] + \mathbb{E}\Big[\int_t^T \|\nabla Y_h(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big] \le C\mathbb{E}\Big[\|\Pi_h\Psi\|_{\mathbb{L}^2}^2\Big].$$

Statement (ii) may be obtained using Itô's formula with $x \mapsto \|\nabla x\|_{\mathbb{L}^2}^2$ for equation (8.1). □

Theorem 8.2
Let (Y, Z) be the strong solution of (6.1) and (Y_h, Z_h) be the strong solution of (8.1). Then

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|Y(t) - Y_h(t)\|_{\mathbb{L}^2}^2 + \int_0^T \Big(\|\nabla[Y(t) - Y_h(t)]\|_{\mathbb{L}^2}^2 + \sum_{i=1}^n \|Z^i(t) - Z_h^i(t)\|_{\mathbb{L}^2}^2\Big)\,\mathrm{d}t\Big] \le Ch^2.$$

Moreover, there holds

$$\mathbb{E}\Big[\int_0^T \|Y(t) - Y_h(t)\|_{\mathbb{L}^2}^2\,\mathrm{d}t\Big] \le Ch^4.$$

Proof
Define $\mathcal{Y}(t) := Y(t) - Y_h(t)$ and $\mathcal{Z}^i(t) := Z^i(t) - Z_h^i(t)$. Subtracting equation (8.1) from (6.1) leads to $(0 \le t \le T)$

$$\begin{aligned}\mathrm{d}\Pi_h\mathcal{Y}(t) &= \Big[-\Delta_h\mathcal{R}_h\mathcal{Y}(t) - \sum_{i=1}^n \Pi_h[\nu^i(t)\mathcal{Z}^i(t)]\Big]\,\mathrm{d}t + \sum_{i=1}^n \Pi_h\mathcal{Z}^i(t)\,\mathrm{d}W^i(t), \\ \Pi_h\mathcal{Y}(T) &= 0.\end{aligned} \tag{8.2}$$

Step 1: By using Itô's formula with $x \mapsto \|x\|_{\mathbb{L}^2}^2$ for equation (8.2) and taking expectations we obtain

$$\mathbb{E}\Big[\|\Pi_h\mathcal{Y}(t)\|_{\mathbb{L}^2}^2\Big] + \mathbb{E}\Big[\sum_{i=1}^n \int_t^T \|\Pi_h\mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big]$$

$$\leq -2\mathbb{E}\Big[\int_t^T \Big(\nabla\mathcal{Y}(s), \nabla\Pi_h\mathcal{Y}(s)\Big)\,\mathrm{d}s\Big] + 2\mathbb{E}\Big[\sum_{i=1}^n \int_t^T \|\Pi_h\mathcal{Y}(s)\|_{\mathbb{L}^2}\|\Pi_h[\nu^i(s)\mathcal{Z}^i(s)]\|_{\mathbb{L}^2}\,\mathrm{d}s\Big].$$

In order to simplify the last term, we get

$$\|\Pi_h[\nu^i(s)\mathcal{Z}^i(s)]\|_{\mathbb{L}^2} \leq \|\nu^i(s)\|_{\mathbb{L}^\infty}\Big(\|\Pi_h\mathcal{Z}^i(s)\|_{\mathbb{L}^2} + \|[\mathrm{Id}-\Pi_h]\mathcal{Z}^i(s)\|_{\mathbb{L}^2}\Big),$$

such that

$$\begin{aligned}
&\mathbb{E}\Big[\|\Pi_h\mathcal{Y}(t)\|_{\mathbb{L}^2}^2\Big] + \mathbb{E}\Big[\sum_{i=1}^n\int_t^T \|\Pi_h\mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big]\\
&\leq -2\mathbb{E}\Big[\int_t^T \|\nabla\mathcal{Y}(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big] + 2\mathbb{E}\Big[\int_t^T (\nabla\mathcal{Y}(s), \nabla[\mathrm{Id}-\Pi_h]\mathcal{Y}(s))\,\mathrm{d}s\Big]\\
&\quad + C\mathbb{E}\Big[\int_t^T \|\Pi_h\mathcal{Y}(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big] + \frac{1}{2}\mathbb{E}\Big[\sum_{i=1}^n\int_t^T \|\Pi_h\mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big]\\
&\quad + Ch^2\mathbb{E}\Big[\sum_{i=1}^n\int_t^T \|\nabla Z^i(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big].
\end{aligned}$$

By interpolation estimates, Theorem 7.1, and Gronwall's inequality we arrive at

$$\mathbb{E}\Big[\|\Pi_h\mathcal{Y}(t)\|_{\mathbb{L}^2}^2\Big] + \mathbb{E}\Big[\int_t^T \|\nabla\mathcal{Y}(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big] + \frac{1}{2}\mathbb{E}\Big[\sum_{i=1}^n\int_t^T \|\Pi_h\mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big] \leq Ch^2. \tag{8.3}$$

Step 2: Taking the supremum over $t\in[0,T]$ in the beginning of Step 1 before applying expectations leads to

$$\begin{aligned}
&\sup_{t\in[0,T]}\|\Pi_h\mathcal{Y}(t)\|_{\mathbb{L}^2}^2\\
&\leq C\sum_{i=1}^n\int_0^T \Big|\big(\Pi_h[\nu^i(s)\mathcal{Z}^i(s)], \Pi_h\mathcal{Y}(s)\big)\Big|\,\mathrm{d}s + \int_0^T \Big|\big(\Delta_h\mathcal{R}_h\mathcal{Y}(s), \Pi_h\mathcal{Y}(s)\big)\Big|\,\mathrm{d}s\\
&\quad + \sum_{i=1}^n\Big|\int_0^T \big(\Pi_h\mathcal{Z}^i(s), \Pi_h\mathcal{Y}(s)\big)\,\mathrm{d}W^i(s)\Big| + \sum_{i=1}^n \sup_{t\in[0,T]}\Big|\int_0^t \big(\Pi_h\mathcal{Z}^i(s), \Pi_h\mathcal{Y}(s)\big)\,\mathrm{d}W^i(s)\Big|,
\end{aligned}$$

where the stochastic integral term is decomposed into the last two terms in order to apply the Burkholder-Davis-Gundy inequality below. Now we take expectations and apply the results in (8.3), interpolation estimates, and Theorem 7.1 for the first three terms on the right-hand-side to arrive at

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[0,T]}\|\Pi_h\mathcal{Y}(t)\|_{\mathbb{L}^2}^2\Big]\\
&\leq C\mathbb{E}\Big[\sum_{i=1}^n\int_0^T \|\Pi_h\mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big] + C\mathbb{E}\Big[\int_0^T \|\Pi_h\mathcal{Y}(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big] + \mathbb{E}\Big[\int_0^T \big|(\nabla\mathcal{R}_h\mathcal{Y}(s), \nabla\mathcal{Y}(s))\big|\,\mathrm{d}s\Big]\\
&\quad + C\mathbb{E}\Big[\sum_{i=1}^n\int_0^T \|[\mathrm{Id}-\Pi_h]Z^i(s)\|_{\mathbb{L}^2}^2\,\mathrm{d}s\Big] + 2\sum_{i=1}^n\mathbb{E}\Big[\sup_{t\in[0,T]}\Big|\int_0^t \big(\Pi_h\mathcal{Z}^i(s), \Pi_h\mathcal{Y}(s)\big)\,\mathrm{d}W^i(s)\Big|\Big]\\
&\leq Ch^2 + I + II.
\end{aligned}$$

By Theorem 7.1 we obtain $I \leq Ch^2$. The last term II may be estimated according to

$$\begin{aligned} II &\leq \sum_{i=1}^{n} \mathbb{E}\Big[\Big|\int_0^T \|\Pi_h \mathcal{Y}(s)\|_{\mathbb{L}^2}^2 \|\Pi_h \mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2 \,\mathrm{d}s\Big|^{\frac{1}{2}}\Big] \\ &\leq \delta \mathbb{E}\Big[\sup_{t\in[0,T]} \|\Pi_h \mathcal{Y}(s)\|_{\mathbb{L}^2}^2\Big] + C_\delta \mathbb{E}\Big[\sum_{i=1}^{n} \int_0^T \|\Pi_h \mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2 \,\mathrm{d}s\Big] \end{aligned}$$

using Burkholder-Davis-Gundy's, Hölder's, and Young's inequalities ($\delta > 0$). Thus, by using estimate (8.3), we arrive at $\mathbb{E}[\sup_{t\in[0,T]} \|\Pi_h \mathcal{Y}(s)\|_{\mathbb{L}^2}^2] \leq Ch^2$.

Step 3: Standard interpolation estimates and Theorem 7.1 justify

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|\mathcal{Y}(t)\|_{\mathbb{L}^2}^2\Big] \leq C\mathbb{E}\Big[\sup_{t\in[0,T]} \|\Pi_h \mathcal{Y}(t)\|_{\mathbb{L}^2}^2\Big] + C\mathbb{E}\Big[\sup_{t\in[0,T]} \|[\Pi_h - \mathrm{Id}]\mathcal{Y}(t)\|_{\mathbb{L}^2}^2\Big] \leq Ch^2,$$

and $\mathbb{E}[\sum_{i=1}^n \int_0^T \|\mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2 \,\mathrm{d}s] \leq Ch^2$ accordingly.

Step 4: Use Itô's formula with $x \mapsto \|\nabla \Delta_h^{-1} x\|_{\mathbb{L}^2}^2$ for equation (8.2). Then, $\mathbb{P}$-a.s.

$$\begin{aligned} &- \|\nabla \Delta_h^{-1} \Pi_h \mathcal{Y}(t)\|_{\mathbb{L}^2}^2 \\ &= \int_t^T \Big[2\Big(-\nabla \Delta_h^{-1} \Delta_h \mathcal{R}_h \mathcal{Y}(s), \nabla \Delta_h^{-1} \Pi_h \mathcal{Y}(s)\Big) - 2\sum_{i=1}^{n} \Big(\nabla \Delta_h^{-1} \Pi_h [\nu^i(s) \mathcal{Z}^i(s)], \nabla \Delta_h^{-1} \Pi_h \mathcal{Y}(s)\Big) \\ &\quad + \sum_{i=1}^{n} \|\nabla \Delta_h^{-1} \Pi_h \mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2 \Big] \,\mathrm{d}s + 2\sum_{i=1}^{n} \int_t^T \Big(\nabla \Delta_h^{-1} \Pi_h \mathcal{Z}^i(s), \nabla \Delta_h^{-1} \Pi_h \mathcal{Y}(s)\Big) \,\mathrm{d}W^i(s). \end{aligned}$$

Taking expectations, using Young's inequality and the definition of Δ_h yields

$$\begin{aligned} &\mathbb{E}\Big[\|\nabla \Delta_h^{-1} \Pi_h \mathcal{Y}(t)\|_{\mathbb{L}^2}^2\Big] + \mathbb{E}\Big[\sum_{i=1}^{n} \int_t^T \|\nabla \Delta_h^{-1} \Pi_h \mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2 \,\mathrm{d}s\Big] + 2\mathbb{E}\Big[\int_t^T \Big(\mathcal{R}_h \mathcal{Y}(s), \Pi_h \mathcal{Y}(s)\Big) \,\mathrm{d}s\Big] \\ &\leq \mathbb{E}\Big[\sum_{i=1}^{n} \int_t^T \epsilon_1 \|\nabla \Delta_h^{-1} \Pi_h [\nu^i(s) \mathcal{Z}^i(s)]\|_{\mathbb{L}^2}^2 \,\mathrm{d}s\Big] + C_{\epsilon_1} \mathbb{E}\Big[\int_t^T \|\nabla \Delta_h^{-1} \Pi_h \mathcal{Y}(s)\|_{\mathbb{L}^2} \,\mathrm{d}s\Big] \end{aligned}$$

for a $\epsilon_1 > 0$. In order to estimate the last term on the right-hand-side, write $V(s) := \Delta_h^{-1} \Pi_h [\nu^i(s) \mathcal{Z}^i(s)]$ and consider the problem $(\Delta_h V, \phi_h) = (\nu^i \mathcal{Z}^i, \phi_h)$ $\forall \phi_h \in \mathbb{V}_h$. By Galerkin orthogonality, the definition of Δ_h, interpolation estimates, stability of Π_h, and Poincare's inequality, there holds

$$\begin{aligned} &\epsilon_1 \|\nabla V\|_{\mathbb{L}^2}^2 \\ &= \epsilon_1 (\Pi_h \mathcal{Z}^i, \nu^i V) + \epsilon_1 ([\mathrm{Id} - \Pi_h] Z^i, \nu^i V) \\ &= \epsilon_1 (\Delta_h \Delta_h^{-1} \Pi_h \mathcal{Z}^i, \nu^i V) + \epsilon_1 ([\mathrm{Id} - \Pi_h] Z^i, [\mathrm{Id} - \Pi_h] \nu^i V) \\ &\leq \epsilon_1 \Big|(\nabla \Delta_h^{-1} \Pi_h \mathcal{Z}^i, \nabla \Pi_h [\nu^i V])\Big| + \epsilon_1 \|[\mathrm{Id} - \Pi_h] Z^i\|_{\mathbb{L}^2} \|[\mathrm{Id} - \Pi_h] \nu^i V\|_{\mathbb{L}^2} \\ &\leq \frac{1}{2} \|\nabla \Delta_h^{-1} \Pi_h \mathcal{Z}^i\|_{\mathbb{L}^2}^2 + C_{\epsilon_2} h^4 \|\nabla Z^i\|_{\mathbb{L}^2}^2 + \epsilon_1^2 (C + \epsilon_2) \Big(\|\nu^i \nabla V\|_{\mathbb{L}^2}^2 + \|V \nabla \nu^i\|_{\mathbb{L}^2}^2\Big) \\ &\leq \frac{1}{2} \|\nabla \Delta_h^{-1} \Pi_h \mathcal{Z}^i\|_{\mathbb{L}^2}^2 + C_{\epsilon_2} h^4 \|\nabla Z^i\|_{\mathbb{L}^2}^2 + \epsilon_1^2 (C + \epsilon_2) \Big(\|\nu^i\|_{\mathbb{L}^\infty}^2 + C\|\nabla \nu^i\|_{\mathbb{L}^\infty}^2\Big) \|\nabla V\|_{\mathbb{L}^2}^2 \end{aligned}$$

for any $\epsilon_2 > 0$. For sufficiently small $\epsilon_1 > 0$, we absorb the last term of the right-hand-side, and hence, there exists a constant $0 < \widetilde{\delta} < 1$, such that

$$\epsilon_1 \|\nabla V\|_{\mathbb{L}^2}^2 \le \widetilde{\delta} \|\nabla \Delta_h^{-1} \Pi_h \mathcal{Z}^i\|_{\mathbb{L}^2}^2 + C_{\epsilon_2} h^4 \|\nabla Z^i\|_{\mathbb{L}^2}^2.$$

We obtain using inequality (7.1) and the equality

$$\Big(\mathcal{R}_h \mathcal{Y}(s), \Pi_h \mathcal{Y}(s)\Big) = \|\mathcal{Y}(s)\|_{\mathbb{L}^2}^2 + \Big([\mathcal{R}_h - \mathrm{Id}]\mathcal{Y}(s), \Pi_h \mathcal{Y}(s)\Big) + \Big([\Pi_h - \mathrm{Id}]\mathcal{Y}(s), \mathcal{Y}(s)\Big)$$

such that

$$\begin{aligned}
&\mathbb{E}\Big[\|\nabla \Delta_h^{-1} \Pi_h \mathcal{Y}(t)\|_{\mathbb{L}^2}^2\Big] + (1 - \widetilde{\delta}) \sum_{i=1}^n \mathbb{E}\Big[\int_t^T \|\nabla \Delta_h^{-1} \Pi_h \mathcal{Z}^i(s)\|_{\mathbb{L}^2}^2 \,\mathrm{d}s\Big] + 2\mathbb{E}\Big[\int_t^T \|\mathcal{Y}(s)\|_{\mathbb{L}^2}^2 \,\mathrm{d}s\Big] \\
&\le C\mathbb{E}\Big[\int_t^T \|\nabla \Delta_h^{-1} \Pi_h \mathcal{Y}(s)\|_{\mathbb{L}^2}^2 \,\mathrm{d}s\Big] + Ch^4 \mathbb{E}\Big[\sum_{i=1}^n \int_t^T \|\nabla Z^i(s)\|_{\mathbb{L}^2}^2 \,\mathrm{d}s\Big] \\
&\quad + C\mathbb{E}\Big[\int_t^T \|[\mathcal{R}_h - \mathrm{Id}]\mathcal{Y}(s)\|_{\mathbb{L}^2}^2 \,\mathrm{d}s\Big] + C\mathbb{E}\Big[\int_t^T \|[\Pi_h - \mathrm{Id}]\mathcal{Y}(s)\|_{\mathbb{L}^2}^2 \,\mathrm{d}s\Big].
\end{aligned}$$

Standard approximation results, Theorem 7.1, and Gronwall's inequality validates the assertion. □

8.2. Forward-backward stochastic heat equation

For the sake of simplicity, suppose $\nu^i \equiv 1$ throughout this section. However, the results can be shown for ν^i specified in Section 7.4. We use Theorem 8.2 to derive strong rates of convergence for the following spatial discretization of (FBSHE): Find $\Theta_h^\star \equiv (X_h^\star, Y_h^\star, Z_h^\star)$ such that $(0 \le t \le T)$

$$\mathrm{d}X_h^\star(t) = \Big[\Delta_h X_h^\star(t) - \frac{1}{\alpha} Y_h^\star(t)\Big] \,\mathrm{d}t + \sum_{i=1}^n X_h^\star(t)] \,\mathrm{d}W^i(t), \tag{8.4}$$

$$\mathrm{d}Y_h^\star(t) = \Big[-\Delta_h Y_h^\star(t) - \sum_{i=1}^n Z_h^{\star,i}(t) - \Big(X_h^\star(t) - \Pi_h \widetilde{X}(t)\Big)\Big] \,\mathrm{d}t + \sum_{i=1}^n Z_h^{\star,i}(t) \,\mathrm{d}W^i(t), \tag{8.5}$$

$$X_h^\star(0) = \Pi_h x_0, \qquad Y_h^\star(T) = \Pi_h [Dg(X_h^\star(T))].$$

Below we use the Banach space

$$\mathcal{M}_{\mathbb{V}_h}[0,T] := \Big[L^2_{\mathcal{P}}(\Omega; C([0,T]; \mathbb{V}_h))\Big]^2 \times L^2_{\mathcal{P}}(\Omega; L^2(0,T; \mathbb{V}_h^n)),$$

which is endowed with the norm $\|(X_h, Y_h, Z_h)\|^2_{\mathcal{M}_{\mathbb{V}_h}[0,T]} := \|(X_h, Y_h, Z_h)\|^2_{\mathcal{M}_{\mathbb{L}^2}[0,T]}$.

System (8.4)–(8.5) may be interpreted as first order optimality system of the finite element discetization of the stochastic optimal control problem (6.2)–(6.3), for which [YZ99, Chapter 3, Theorem 3.2] ensures solvability. However, we are here interested in using the practical solution strategy via (discrete) Picard iterates, which solve the following system of equations: Let $Y_h^{(0)} \equiv 0$. Iterate $v = 1, 2, \ldots$

$$\mathrm{d}X_h^{(v)}(t) = \Big[\Delta_h X_h^{(v)}(t) - \frac{1}{\alpha} Y_h^{(v-1)}(t)\Big] \,\mathrm{d}t + \sum_{i=1}^n X_h^{(v)}(t) \,\mathrm{d}W^i(t), \tag{8.6}$$

$$\begin{aligned} \mathrm{d}Y_h^{(v)}(t) &= \Big[-\Delta_h Y_h^{(v)}(t) - \sum_{i=1}^{n} Z_h^{(v),i}(t) - \Big(X_h^{(v)}(t) - \Pi_h \widetilde{X}(t)\Big)\Big]\,\mathrm{d}t \\ &\quad + \sum_{i=1}^{n} Z_h^{(v),i}(t)\,\mathrm{d}W^i(t), \\ X_h^{(v)}(0) &= \Pi_h x_0, \qquad Y_h^{(v)}(T) = \Pi_h\big[Dg(X_h^{(v)}(T))\big]. \end{aligned} \tag{8.7}$$

Let $Y_h^{(v-1)} \in L^2_{\mathcal{P}}(\Omega; L^2(0,T;\mathbb{V}_h))$ be given. There exists a unique strong solution $X_h^{(v)} \in L^2_{\mathcal{P}}(\Omega; C([0,T];\mathbb{V}_h))$ of the stochastic differential equation (8.6), while solvability of (8.7) follows from [MY07, Chapter 1, Theorem 4.2].

Theorem 8.3
Let $0 < T \le \widehat{T}_0$ be sufficiently small and independent of h. Then iterates $\{\Theta_h^{(v)} \equiv (X_h^{(v)}, Y_h^{(v)}, Z_h^{(v)});\ v \in \mathbb{N}\}$ of equations (8.6)–(8.7) converges to the unique strong solution $\Theta_h^\star \equiv (X_h^\star, Y_h^\star, Z_h^\star) \in \mathcal{M}_{\mathbb{V}_h}[0,T]$ of (8.4)–(8.5).

Proof
We show a contraction property for the (fixed point) mapping $\mathcal{T} : \mathcal{M}_{\mathbb{V}_h}[0,T] \to \mathcal{M}_{\mathbb{V}_h}[0,T]$ which maps $\Theta_h^{(v-1)}$ to $\Theta_h^{(v)}$ according to equations (8.6)–(8.7). Suppose that $\Theta_h^{(v-2)}, \Theta_h^{(v-1)} \in \mathcal{M}_{\mathbb{V}_h}[0,T]$ are given, and both satisfy system (8.6)–(8.7). For $w \in \{v-1, v\}$ we define the differences

$$(\overline{\mathcal{X}}_h^{(w)}, \overline{\mathcal{Y}}_h^{(w)}, \overline{\mathcal{Z}}_h^{(w)}) := (X_h^{(w)} - X_h^{(w-1)}, Y_h^{(w)} - Y_h^{(w-1)}, Z_h^{(w)} - Z_h^{(w-1)}),$$

and consider the forward difference process $(0 \le t \le T)$

$$\begin{aligned} \mathrm{d}\overline{\mathcal{X}}_h^{(v)}(t) &= \Big[\Delta_h \overline{\mathcal{X}}_h^{(v)}(t) - \frac{1}{\alpha}\overline{\mathcal{Y}}_h^{(v-1)}(t)\Big]\,\mathrm{d}t + \sum_{i=1}^{n} \overline{\mathcal{X}}_h^{(v)}(t)\,\mathrm{d}W^i(t), \\ \overline{\mathcal{X}}_h^{(v)}(0) &= 0. \end{aligned}$$

By applying Itô's formula with $x \mapsto \|x\|^2_{\mathbb{L}^2}$, using the Burkholder-Davis-Gundy inequality and Gronwall's lemma, there exists a constant $\overline{C} > 0$ independent of h and v, such that

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|\overline{\mathcal{X}}_h^{(v)}(t)\|^2_{\mathbb{L}^2}\Big] \le \exp(\overline{C}T)\frac{2}{\alpha^2}\mathbb{E}\Big[\int_0^T \|\overline{\mathcal{Y}}_h^{(v-1)}(t)\|^2_{\mathbb{L}^2}\,\mathrm{d}t\Big]. \tag{8.8}$$

The equation for the backward difference process is $(0 \le t \le T)$

$$\begin{aligned} \mathrm{d}\overline{\mathcal{Y}}_h^{(v)}(t) &= \Big[-\Delta_h \overline{\mathcal{Y}}_h^{(v)}(t) - \sum_{i=1}^{n} \overline{\mathcal{Z}}_h^{(v),i}(t) - \overline{\mathcal{X}}_h^{(v)}(t)\Big]\,\mathrm{d}t + \sum_{i=1}^{n} \overline{\mathcal{Z}}_h^{(v),i}(t)\,\mathrm{d}W^i(t), \\ \overline{\mathcal{Y}}_h^{(v)}(T) &= \Pi_h\big[Dg(X_h^{(v)}(T)) - Dg(X_h^{(v-1)}(T))\big]. \end{aligned}$$

Applying again Itô's formula with $x \mapsto \|x\|^2_{\mathbb{L}^2}$, and using (7.2) leads to

$$\begin{aligned} &\sup_{t\in[0,T]} \mathbb{E}\big[\|\overline{\mathcal{Y}}_h^{(v)}(t)\|^2_{\mathbb{L}^2}\big] + \mathbb{E}\Big[\sum_{i=1}^{n}\int_0^T \|\overline{\mathcal{Z}}_h^{(v),i}(t)\|^2_{\mathbb{L}^2}\,\mathrm{d}t\Big] \\ &\le \exp(\overline{C}T)\Big(\mathbb{E}\Big[\int_0^T \|\overline{\mathcal{X}}_h^{(v)}(t)\|^2_{\mathbb{L}^2}\,\mathrm{d}t\Big] + C\mathbb{E}\big[\|\overline{\mathcal{X}}_h^{(v)}(T)\|^2_{\mathbb{L}^2}\big]\Big), \end{aligned} \tag{8.9}$$

and with the Burkholder-Davis-Gundy inequality (see Step 2 in the proof of Theorem 8.2)

$$\begin{aligned}&\mathbb{E}\Big[\sup_{t\in[0,T]}\|\overline{\mathcal{Y}}_h^{(v)}(t)\|_{\mathbb{L}^2}^2+\sum_{i=1}^n\int_0^T\|\overline{\mathcal{Z}}_h^{(v),i}(t)\|_{\mathbb{L}^2}^2\,\mathrm{d}t\Big]\\&\leq\Big(\overline{C}\exp(\overline{C}T)+\overline{C}\Big)\Big(\mathbb{E}\Big[\int_0^T\|\overline{\mathcal{X}}_h^{(v)}(t)\|_{\mathbb{L}^2}^2\,\mathrm{d}t\Big]+\mathbb{E}[\|\overline{\mathcal{X}}_h^{(v)}(T)\|_{\mathbb{L}^2}^2]\Big).\end{aligned}$$

Combining this estimate with (8.8) leads to

$$\|(\overline{\mathcal{X}}_h^{(v)},\overline{\mathcal{Y}}_h^{(v)},\overline{\mathcal{Z}}_h^{(v)})\|_{\mathcal{M}_{\mathbb{V}_h}[0,T]}^2\leq\overline{q}\|(\overline{\mathcal{X}}_h^{(v-1)},\overline{\mathcal{Y}}_h^{(v-1)},\overline{\mathcal{Z}}_h^{(v-1)})\|_{\mathcal{M}_{\mathbb{V}_h}[0,T]}^2,$$

where $\overline{q}<1$ (independent of h) for $0<T\leq\widehat{T}_0$ small enough. Thus we obtain that the mapping $\mathcal{T}$ is contractive. As a consequence, there exists a fixed point $(X_h^\star,Y_h^\star,Z_h^\star)\in\mathcal{M}_{\mathbb{V}_h}[0,T]$ for $\mathcal{T}$. Moreover,

$$\|\Theta_h^{(v)}-\Theta_h^\star\|_{\mathcal{M}_{\mathbb{V}_h}[0,T]}^2\leq\frac{\overline{q}^{v-1}}{1-\overline{q}}\|\Theta_h^{(2)}-\Theta_h^{(1)}\|_{\mathcal{M}_{\mathbb{V}_h}[0,T]}^2.\tag{8.10}$$

The following theorem asserts rates of convergence for the solution $(X_h^\star,Y_h^\star,Z_h^\star)$ of (8.4)–(8.5) towards $(X^\star,Y^\star,Z^\star)$ from (FBSHE) with respect to $h>0$ for terminal times $T\leq\min\{\widetilde{T}_0,\widehat{T}_0\}$. Its proof combines an error analysis for the approximation (8.6)–(8.7) of (7.5)–(7.6) for every $v\in\mathbb{N}$ with a fixed point argument, and a verification of improved regularity properties for solutions of (7.5)–(7.6). For this purpose, we consider the Banach space

$$\begin{aligned}\mathcal{N}[0,T]:=&\Big[L_{\mathcal{P}}^2\Big(\Omega;C([0,T];\mathbb{W}_0^{1,2})\cap L^2(0,T;\mathbb{W}_0^{1,2}\cap\mathbb{W}^{2,2})\Big)\Big]^2\\&\times L_{\mathcal{P}}^2\Big(\Omega;L^2(0,T;\mathbb{W}_0^{1,2}(D;\mathbb{R}^n))\Big)\subset\mathcal{M}_{\mathbb{L}^2}[0,T],\end{aligned}$$

which is endowed with the norm

$$\begin{aligned}\|(X,Y,Z)\|_{\mathcal{N}[0,T]}^2:=&\,\mathbb{E}\Big[\sup_{t\in[0,T]}\|X(t)\|_{\mathbb{W}^{1,2}}^2+\sup_{t\in[0,T]}\|Y(t)\|_{\mathbb{W}^{1,2}}^2\Big]\\&+\mathbb{E}\Big[\int_0^T\Big(\|X(t)\|_{\mathbb{W}^{2,2}}^2+\|Y(t)\|_{\mathbb{W}^{2,2}}^2+\sum_{i=1}^n\|Z^i(t)\|_{\mathbb{W}^{1,2}}^2\Big)\,\mathrm{d}t\Big].\end{aligned}$$

We recall the numbers $\widetilde{T}_0,\widehat{T}_0>0$ from Theorems 7.2 and 8.3.

Theorem 8.4
Let $\Theta^\star\equiv(X^\star,Y^\star,Z^\star)$ solve (FBSHE), and $\Theta_h^\star\equiv(X_h^\star,Y_h^\star,Z_h^\star)$ be the solution of (8.4)–(8.5). There exists a $0<T_0\leq\min\{\widetilde{T}_0,\widehat{T}_0\}$, such that for every $0<T\leq T_0$ holds

$$\big\|(X^\star,Y^\star,Z^\star)-(X_h^\star,Y_h^\star,Z_h^\star)\big\|_{\mathcal{M}_{\mathbb{L}^2}[0,T]}^2\leq Ch^2.$$

For $g(\cdot)$ quadratic we additionally have

$$\mathbb{E}\Big[\int_0^T\Big(\|X^\star(t)-X_h^\star(t)\|_{\mathbb{L}^2}^2+\|Y^\star(t)-Y_h^\star(t)\|_{\mathbb{L}^2}^2\Big)\,\mathrm{d}t\Big]\leq Ch^4.$$

Proof

Step 1: We show improved regularity properties of the solution of (7.5)–(7.6) for every $v = 1, 2, \ldots$. There exist $T_1 > 0$ such that $\{\Theta^{(v)} \equiv (X^{(v)}, Y^{(v)}, Z^{(v)});\, v \in \mathbb{N}\} \subset \mathcal{N}[0,T]$ for all $T \le T_1$, and a constant $C > 0$ such that

$$\max_{v=1,2,\ldots} \big\|(X^{(v)}, Y^{(v)}, Z^{(v)})\big\|_{\mathcal{N}[0,T]} \le C. \tag{8.11}$$

It is sufficient to verify the bound (8.11). For equation (7.5) we obtain

$$\begin{aligned}&\mathbb{E}\Big[\sup_{t\in[0,T]} \|X^{(v)}(t)\|^2_{\mathbb{W}^{1,2}}\Big] + \mathbb{E}\Big[\int_0^T \|X^{(v)}(t)\|^2_{\mathbb{W}^{2,2}}\,\mathrm{d}t\Big]\\ &\le \exp(\widetilde{C}T)\Big(\widetilde{C} + \frac{\widetilde{C}}{\alpha^2}\mathbb{E}\Big[\int_0^T \|Y^{(v-1)}(t)\|^2_{\mathbb{L}^2}\,\mathrm{d}t\Big]\Big),\end{aligned}$$

for some $\widetilde{C} > 0$ depending only on the data of the limiting problem (FBSHE). For equation (7.6) we use Theorem 7.1 to establish

$$\begin{aligned}&\mathbb{E}\Big[\sup_{t\in[0,T]} \|Y^{(v)}(t)\|^2_{\mathbb{W}^{1,2}}\Big] + \mathbb{E}\Big[\int_0^T \Big(\|Y^{(v)}(t)\|^2_{\mathbb{W}^{2,2}} + \sum_{i=1}^n \|Z^{(v),i}(t)\|^2_{\mathbb{W}^{1,2}}\Big)\,\mathrm{d}t\Big]\\ &\le \widetilde{C}\exp(\widetilde{C}T)(T^2+T)\Big(\widetilde{C} + \frac{1}{\alpha^2}\mathbb{E}\Big[\int_0^T \|X^{(v)}(t)\|^2_{\mathbb{L}^2}\,\mathrm{d}t\Big]\Big) + +\widetilde{C}\mathbb{E}[\|X^{(v)}(T)\|^2_{\mathbb{W}^{1,2}}].\end{aligned}$$

By Poincare's inequality and a contraction argument for $0 < T \le T_1 \equiv T_1(\widetilde{C})$ small enough, we obtain estimate (8.11).

We now start with the error analysis. Each component of the error is split into three parts, for example

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\in[0,T]} \|X^\star(t) - X^\star_h(t)\|^2_{\mathbb{L}^2}\Big] \le{}& \mathbb{E}\Big[\sup_{t\in[0,T]} \|X^\star(t) - X^{(v)}(t)\|^2_{\mathbb{L}^2}\Big]\\ &+ \mathbb{E}\Big[\sup_{t\in[0,T]} \|X^{(v)}(t) - X^{(v)}_h(t)\|^2_{\mathbb{L}^2}\Big]\\ &+ \mathbb{E}\Big[\sup_{t\in[0,T]} \|X^{(v)}_h(t) - X^\star_h(t)\|^2_{\mathbb{L}^2}\Big],\end{aligned} \tag{8.12}$$

and accordingly for $(Y^\star, Z^\star)$ and $(Y^\star_h, Z^\star_h)$. In Step 2 below, the second term on the right-hand-side of (8.12) will be estimated, while the remaining errors are dealt with in Step 3.

Step 2: Let $0 < T \le T_1$. Fix $v \in \mathbb{N}$ and define

$$(\mathcal{X}^{(v)}, \mathcal{Y}^{(v)}, \mathcal{Z}^{(v)}) := (X^{(v)} - X^{(v)}_h, Y^{(v)} - Y^{(v)}_h, Z^{(v)} - Z^{(v)}_h),$$

subtract system (7.5)–(7.6) from (8.6)–(8.7) and apply Itô's formula with $x \mapsto \|x\|^2_{\mathbb{L}^2}$. Standard estimates together with the uniform estimate (8.11), the Burkholder-Davis-Gundy and Gronwall's inequality then lead to the following estimate,

$$\mathbb{E}\big[\sup_{t\in[0,T]} \|\Pi_h \mathcal{X}^{(v)}(t)\|^2_{\mathbb{L}^2}\big] \le \exp(\widehat{C}T)\frac{1}{\alpha^2}\mathbb{E}\Big[\int_0^T \|\Pi_h \mathcal{Y}^{(v-1)}(t)\|^2_{\mathbb{L}^2}\,\mathrm{d}t\Big] + Ch^2, \tag{8.13}$$

with $\widehat{C}, C > 0$ depending only on the data, but not on h and v. By Theorem 8.2 and adding additional errors at the terminal time T using (7.2) we obtain

$$\begin{aligned}&\mathbb{E}\big[\sup_{t\in[0,T]}\|\Pi_h\mathcal{Y}^{(v)}(t)\|_{\mathbb{L}^2}^2\big]\\&\leq \widehat{C}\exp(\widehat{C}T)(T+T^2)\Big(\mathbb{E}\Big[\int_0^T\|\Pi_h\mathcal{X}^{(v)}(t)\|_{\mathbb{L}^2}^2\,\mathrm{d}t\Big]+\mathbb{E}\big[\|\Pi_h\mathcal{X}^{(v)}(T)\|_{\mathbb{L}^2}^2\big]\Big)+Ch^2.\end{aligned}\tag{8.14}$$

Thanks to (8.13) we then arrive at

$$\begin{aligned}&\mathbb{E}\big[\sup_{t\in[0,T]}\|\Pi_h\mathcal{Y}^{(v)}(t)\|_{\mathbb{L}^2}^2\big]\\&\leq \widehat{C}\exp(\widehat{C}T)\frac{T+T^2}{\alpha^2}\Big(T\mathbb{E}\big[\sup_{t\in[0,T]}\|\Pi_h\mathcal{Y}^{(v-1)}(t)\|_{\mathbb{L}^2}^2\big]+\mathbb{E}\big[\|\Pi_h\mathcal{X}^{(v)}(T)\|_{\mathbb{L}^2}^2\big]\Big)+Ch^2.\end{aligned}$$

Again, there exists $0 < T_2 \leq T_1$, such that $\widehat{q} := \widehat{C}\exp(\widehat{C}T_2)\frac{T_2^2+T_2^3}{\alpha^2} < 1$, and hence

$$\mathbb{E}\big[\sup_{t\in[0,T_2]}\|\Pi_h\mathcal{Y}^{(v)}(t)\|_{\mathbb{L}^2}^2\big]\leq \widehat{q}^v\sup_{t\in[0,T_2]}\mathbb{E}\big[\|\Pi_h\mathcal{Y}^{(0)}(t)\|_{\mathbb{L}^2}^2\big]+Ch^2\frac{1-\widehat{q}^v}{1-\widehat{q}}\leq Ch^2\frac{1-\widehat{q}^v}{1-\widehat{q}},\tag{8.15}$$

since $\mathcal{Y}^{(0)}\equiv 0$. By (8.13), (8.15), approximation estimates, and (8.11), we arrive at

$$\max_{v=1,2,\dots}\mathbb{E}\Big[\sup_{t\in[0,T_2]}\|\mathcal{X}^{(v)}(t)\|_{\mathbb{L}^2}^2+\sup_{t\in[0,T_2]}\|\mathcal{Y}^{(v)}(t)\|_{\mathbb{L}^2}^2+\sum_{i=1}^n\int_0^{T_2}\|\mathcal{Z}^{(v),i}(t)\|_{\mathbb{L}^2}^2\,\mathrm{d}t\Big]\leq Ch^2.\tag{8.16}$$

Step 3: Set $0 < T \leq \min\{T_2, \widehat{T}_0, \widetilde{T}_0\}$. According to the proofs of Theorems 7.2 and 8.3, the first and third error contributions in (8.12) can be estimated by (7.7) and (8.10), such that both terms vanish for $v \to \infty$, while the second is bounded according to (8.16) (uniformly in v). This implies the first assertion in Theorem 8.4.

Step 4: The second assertion of Theorem 8.4 can be obtained by using Itô's formula with $x \mapsto \|\nabla\Delta_h^{-1}x\|_{\mathbb{L}^2}^2$ in Step 2, and exploiting $Dg(\cdot)$ being affine in (8.14). □

9. Simulation

We present fully implementable algorithms to simulate the semi-discrete backward stochastic heat equation (8.1), and the semi-discrete forward-backward stochastic heat equation considered in (8.4)–(8.5).

9.1. Backward stochastic heat equation

Consider (8.1). Assume that $\Psi = g(X_h(T))$, where the $\mathbb{V}_h$-valued process $\{X_h(t);\, t \in [0,T]\}$ is the solution of a spatially discretized (forward) stochastic partial differential equation driven by W. Its time discretization is denoted by $\{X_h^j;\, j = 0, \dots, J\}$. We use an implicit version of a time discretization for the forward and backward equation to avoid the restrictive mesh constraint $k \le Ch^2$ otherwise; see Figure 10.1.

Scheme 9.1 (One-step forward dynamic programming (ODP), Implicit Euler)
Let $k = t_{j+1} - t_j$ be the uniform time step for a net $\{t_j\}_{j=0}^J$ which covers $[0,T]$.

(i) Simulate $Y_h^J = \Pi_h[g(X_h^J)]$.

(ii) For each $j = J-1, \dots, 0$, simulate the $\mathbb{V}_h$-valued random variables $Z_h^{i,j}$ and Y_h^j such that $\forall \phi_h \in \mathbb{V}_h$

$$(Z_h^{i,j}, \phi_h) = \frac{1}{k}\mathbb{E}\Big[\Delta_j W^i (Y_h^{j+1}, \phi_h)\Big|\mathcal{F}_{t_j}\Big] \qquad \forall i = 1, \dots, n, \tag{9.1}$$

$$(Y_h^j, \phi_h) + k(\nabla Y_h^j, \nabla \phi_h) = \mathbb{E}\Big[(Y_h^{j+1}, \phi_h)\Big|\mathcal{F}_{t_j}\Big] + k\sum_{i=1}^n (\nu^i(t_j) Z_h^{i,j}, \phi_h). \tag{9.2}$$

Equations (9.1)–(9.2) may be interpreted as a projection of the solution onto the available information in each step while going backward in time; see also [BT04].

Scheme 9.1 is restated as an algebraic problem: Let $\phi_h^\ell \in \mathbb{V}_h$ for $\ell = 1, \dots, L$ be basis functions of $\mathbb{V}_h$. Consider $X_h^j(x) = \sum_{\ell=1}^L [\overrightarrow{\mathbf{X}}_h^j]_\ell \phi_h^\ell(x)$, $Y_h^j(x) = \sum_{\ell=1}^L [\overrightarrow{\mathbf{Y}}_h^j]_\ell \phi_h^\ell(x)$ and $Z_h^{i,j}(x) = \sum_{\ell=1}^L [\overrightarrow{\mathbf{Z}}_h^{i,j}]_\ell \phi_h^\ell(x)$ with coefficient vectors $\overrightarrow{\mathbf{X}}_h^j$, $\overrightarrow{\mathbf{Y}}_h^j$, $\overrightarrow{\mathbf{Z}}_h^{i,j} \in \mathbb{R}^L$, where $[\cdot]_\ell$ denotes the ℓ-th coordinate of the vector. Let $\overrightarrow{\mathbf{\Upsilon}}_{\mathbb{V}_h} : \mathbb{V}_h \to \mathbb{R}^L$ be the mapping which returns for each element in $\mathbb{V}_h$ the unique vector of coefficients according to the basis $\{\phi_h^\ell;\, \ell = 1, \dots, L\}$.

We denote by **Stiff** the stiffness matrix consisting of entries $(\nabla \phi_h^\ell, \nabla \phi_h^w)$, where $\phi_h^\ell, \phi_h^w \in \mathbb{V}_h$ are basis functions of $\mathbb{V}_h$, while **Mass** resp. $\mathbf{Mass}_{\nu^i}^j$ denote the mass matrices consisting of entries (ϕ_h^ℓ, ϕ_h^w) resp. $(\nu^i(t_j)\phi_h^\ell, \phi_h^w)$. The equations (9.1)–(9.2) in Scheme 9.1 may then be reformulated as follows:

Scheme 9.2

(i) Compute $\mathbf{Mass}\,\overrightarrow{\mathbf{Y}}_h^J = \overrightarrow{\mathbf{g}}_h$, where $\big[\overrightarrow{\mathbf{g}}_h\big]_\ell = \big(g(X_h^J), \phi_h^\ell\big)$ for each basis function ϕ_h^ℓ of $\mathbb{V}_h$.

(ii) For $j = J-1, \dots, 0$, find the $\mathbb{R}^L$-valued random variables $\overrightarrow{\mathbf{Z}}_h^{i,j}$ and $\overrightarrow{\mathbf{Y}}_h^j$ such that

$$\mathbf{Mass}\,\overrightarrow{\mathbf{Z}}_h^{i,j} = \frac{1}{k}\mathbb{E}\big[\Delta_j W^i\,\mathbf{Mass}\,\overrightarrow{\mathbf{Y}}_h^{j+1}\big|\mathcal{F}_{t_j}\big] \qquad \forall i = 1, \dots, n, \tag{9.3}$$

and

$$(\mathbf{Mass} + k\,\mathbf{Stiff})\overrightarrow{\mathbf{Y}}_h^j = \mathbb{E}\big[\mathbf{Mass}\,\overrightarrow{\mathbf{Y}}_h^{j+1}\big|\mathcal{F}_{t_j}\big] + k\sum_{i=1}^n \mathbf{Mass}_{\nu^i}^j\,\overrightarrow{\mathbf{Z}}_h^{i,j}. \tag{9.4}$$

Due to the linearity of the problem and the tower property of conditional expectations, we may reformulate equation (9.3) to

$$\overrightarrow{\mathbf{Z}}_h^{i,j} = \mathbb{E}\bigg[\frac{\Delta_j W^i}{k}\prod_{r=j+1}^{J-1}\Big((\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}(\mathbf{Mass} + \sum_{i'=1}^n \Delta_r W^{i'}\,\mathbf{Mass}_{\nu^{i'}}^r)\Big)\overrightarrow{\mathbf{Y}}_h^J\bigg|\mathcal{F}_{t_j}\bigg], \tag{9.5}$$

and equation (9.4) to

$$\overrightarrow{\mathbf{Y}}_h^j = \mathbb{E}\bigg[\prod_{r=j}^{J-1}\Big((\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}(\mathbf{Mass} + \sum_{i'=1}^n \Delta_r W^{i'}\,\mathbf{Mass}_{\nu^{i'}}^r)\Big)\overrightarrow{\mathbf{Y}}_h^J\bigg|\mathcal{F}_{t_j}\bigg]. \tag{9.6}$$

This reformulation avoids nested conditional expectations thus a related error propagation.

The key ingredient for the simulation of the conditional expectations is the representation of the coefficient vectors $(\overrightarrow{\mathbf{Y}}_h^j, \overrightarrow{\mathbf{Z}}_h^{i,j})$ of equations (9.5)–(9.6) at time t_j by deterministic functions $(\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^j(\overrightarrow{\mathbf{X}}_h^j), \overrightarrow{\boldsymbol{\mathcal{Z}}}_h^{i,j}(\overrightarrow{\mathbf{X}}_h^j))$ evaluated at the approximation of the (forward) state equation $\overrightarrow{\mathbf{X}}_h^j$, via $\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^J(\overrightarrow{\boldsymbol{x}}_h) = \mathbf{Mass}^{-1}\,\overrightarrow{\mathbf{g}}_h(\overrightarrow{\boldsymbol{x}}_h)$, and

$$\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^j(\overrightarrow{\boldsymbol{x}}_h) = \mathbb{E}\bigg[\prod_{r=j}^{J-1}\Big((\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}(\mathbf{Mass} + \sum_{i'=1}^n \Delta_r W^{i'}\,\mathbf{Mass}_{\nu^{i'}}^r)\Big)\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^J(\overrightarrow{\mathbf{X}}_h^J)\Big|\overrightarrow{\mathbf{X}}_h^j = \overrightarrow{\boldsymbol{x}}_h\bigg] \tag{9.7}$$

$$\begin{aligned}\overrightarrow{\boldsymbol{\mathcal{Z}}}_h^{i,j}(\overrightarrow{\boldsymbol{x}}_h) = \mathbb{E}\bigg[&\frac{\Delta_j W^i}{k} \\ &\times \prod_{r=j+1}^{J-1}\Big((\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}(\mathbf{Mass} + \sum_{i'=1}^n \Delta_r W^{i'}\,\mathbf{Mass}_{\nu^{i'}}^r)\Big)\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^J(\overrightarrow{\mathbf{X}}_h^J)\Big|\overrightarrow{\mathbf{X}}_h^j = \overrightarrow{\boldsymbol{x}}_h\bigg]\end{aligned} \tag{9.8}$$

for all $j = 0, \dots, J-1$ and $i = 1, \dots, n$. These relations can be shown by induction, the Markov chain property of $\{X_h^j;\, j = 0, \dots, J\}$, and a corollary of the monotone class theorem; see also [GT14, Lemma 4.1] and [BM10, Theorem 2.1]. Hence it remains to approximate or compute the deterministic functions $\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^j(\cdot)$ and $\overrightarrow{\boldsymbol{\mathcal{Z}}}_h^{i,j}(\cdot)$ in (9.7)–(9.8). An explicit formula for $\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^j(\cdot)$ and $\overrightarrow{\boldsymbol{\mathcal{Z}}}_h^{i,j}(\cdot)$ may be obtained in the special case where the terminal datum uses a linear map $g(\cdot)$.

Remark 9.3
If the forward SPDE is the stochastic heat equation (6.3) (without control) which is discretized in time by the implicit Euler method, and moreover the function used for the terminal condition is $g(x) = cx$ with $c \in \mathbb{R} \setminus \{0\}$, we can use

$$\overrightarrow{\mathbf{X}}_h^{j+1} = \mathbf{Q}_h^{j,j+1}(\overrightarrow{\mathbf{X}}_h^j) = (\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}\Big(\mathbf{Mass}\,\overrightarrow{\mathbf{X}}_h^j + \sum_{i'=1}^{n} \mathbf{Mass}_{\sigma^{i'}}^j\,\overrightarrow{\mathbf{X}}_h^j \Delta_j W^{i'}\Big)$$

to express the coefficient vector $(\overrightarrow{\mathbf{Y}}_h^j, \overrightarrow{\mathbf{Z}}_h^{i,j})$ of equations (9.3)–(9.4) at time t_j by

$$\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^j(\overrightarrow{\mathbf{X}}_h^j) = \boldsymbol{A}_{\boldsymbol{Y}^j}\overrightarrow{\mathbf{X}}_h^j, \qquad \text{and} \qquad \overrightarrow{\boldsymbol{\mathcal{Z}}}_h^{i,j}(\overrightarrow{\mathbf{X}}_h^j) = \boldsymbol{A}_{\boldsymbol{Z}^{i,j}}\overrightarrow{\mathbf{X}}_h^j,$$

with (deterministic) matrices $\boldsymbol{A}_{\boldsymbol{Y}^j}, \boldsymbol{A}_{\boldsymbol{Z}^{i,j}} \in \mathbb{R}^{L\times L}$, which can be determined by the recursion:

1. Set $\boldsymbol{A}_{\boldsymbol{Y}^J} := c\boldsymbol{I}$.
2. For $j = J-1$ to 0 compute

$$\begin{aligned}
\boldsymbol{A}_{\boldsymbol{Z}^{i,j}} &:= \boldsymbol{A}_{\boldsymbol{Y}^{j+1}}(\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}\,\mathbf{Mass}_{\sigma^i}^j,\\
\boldsymbol{A}_{\boldsymbol{Y}^j} &:= (\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}\,\mathbf{Mass}\,\boldsymbol{A}_{\boldsymbol{Y}^{j+1}}(\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}\,\mathbf{Mass}\\
&\quad + k\sum_{i'=1}^{n}(\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}\,\mathbf{Mass}_{\nu^{i'}}^j\,\boldsymbol{A}_{\boldsymbol{Y}^{j+1}}(\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}\,\mathbf{Mass}_{\sigma^{i'}}^j\,.
\end{aligned}$$

Thus no conditional expectations need to be computed in this case to determine the deterministic functions $\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^j(\cdot)$ and $\overrightarrow{\boldsymbol{\mathcal{Z}}}_h^{i,j}(\cdot)$.

General terminal conditions $g(\cdot)$ however require to compute the conditional expectations in (9.7)–(9.8). Several techniques exist for general BSDEs to estimate the elements of the vectors in (9.7)–(9.8). In this thesis, the focus is on partitioning estimation, which is a special case of the least squares Monte-Carlo method. Let the $\mathbb{R}$-valued Θ^j denote an entry of the vectors in (9.7)–(9.8). Then the least squares Monte-Carlo method (approximately) evaluates $v(\overrightarrow{\boldsymbol{x}}_h) = \mathbb{E}[\Theta^j|\overrightarrow{\mathbf{X}}_h^j = \overrightarrow{\boldsymbol{x}}_h]$ based on the representation

$$v = \operatorname{argmin}_{\phi(\cdot)}\mathbb{E}\big[|\phi(\overrightarrow{\boldsymbol{x}}_h) - \Theta^j|^2\big], \tag{9.9}$$

among all $\mathcal{F}_{t_j}$-measurable functions $\phi : \mathbb{R}^L \to \mathbb{R}$ such that $\mathbb{E}\big[|\phi(\overrightarrow{\boldsymbol{x}}_h)|^2\big] < \infty$. In order to allow for the computation of $v(\cdot)$, problem (9.9) is replaced by a finite-dimensional minimization problem where the measurable function $v : \mathbb{R}^L \to \mathbb{R}$ is replaced by a function $v_R : \mathbb{R}^L \to \mathbb{R}$ in a finite dimensional linear subspace $\operatorname{span}\{\eta_r^j(\cdot);\, r = 1,\dots,R\}$ of $L^2(\Omega;\mathbb{P})$. As a consequence, coefficients $\{\mathbf{a}_r^j;\, r = 1,\dots,R\}$ in the representation $v_R(\cdot) = \sum_{r=1}^R \mathbf{a}_r^j\eta_r^j(\cdot)$ are then determined by minimizing the following least-squares problem

$$\frac{1}{M}\sum_{m=1}^{M}\Big|\sum_{r=1}^{R}\mathbf{a}_r^j\eta_r^j(\overrightarrow{\mathbf{X}}_h^{j,m}) - \Theta^{j,m}\Big|^2, \tag{9.10}$$

using M (where $M \gg R$) independent samples $\{(\Theta^{j,m}, \overrightarrow{\mathbf{X}}_h^{j,m});\, m = 1,\dots,M\}$ of (Θ^j, X_h^j). In the case of partitioning estimates, the basis functions η_r^j are given as indicator functions

$\eta_r^j(\cdot) := \mathbb{1}_{C_r^j}(\cdot)$ for disjoint regions C_r^j which partition $X_h^j[\Omega]$. The minimizing coefficients α_r^j in (9.10) are of the form

$$\mathbf{a}_r^j = \frac{1}{\#\{\overrightarrow{\mathbf{X}}_h^{j,m} \in C_r^j\}} \sum_{m=1}^{M} \mathbb{1}_{C_r^j}(X_h^{j,m})\Theta^{j,m} \approx \mathbb{E}\Big[\Theta^j \Big| \overrightarrow{\mathbf{X}}_h^{j,m} \in C_r^j\Big],$$

where the convention $\frac{0}{0} = 0$ is used. Thus, the conditional expectation $\mathbb{E}[\Theta^j|\overrightarrow{\mathbf{X}}_h^{j,m} = \overrightarrow{\boldsymbol{x}}_h]$ is approximated by

$$\mathbb{E}\Big[\Theta^j \Big| \overrightarrow{\mathbf{X}}_h^{j,m} = \overrightarrow{\boldsymbol{x}}_h\Big] \approx \sum_{r=1}^{R}\Big(\frac{1}{\#\{\overrightarrow{\mathbf{X}}_h^{j,m} \in C_r^j\}} \sum_{m=1}^{M} \mathbb{1}_{C_r^j}(\overrightarrow{\mathbf{X}}_h^{j,m})\Theta^{j,m}\Big) \cdot \mathbb{1}_{C_r^j}(\overrightarrow{\boldsymbol{x}}_h). \tag{9.11}$$

To sum up, for every $\overrightarrow{\boldsymbol{x}}_h \in \overrightarrow{\mathbf{X}}_h^{j,m}[\Omega]$, the partitioning estimator returns the local average of those $\Theta^{j,m}$ whose $\overrightarrow{\mathbf{X}}_h^{j,m}$ has been in the same region C_r^j as $\overrightarrow{\boldsymbol{x}}_h$. Note that this approach is less time and memory consuming in comparison with a usual least squares Monte-Carlo method, where a singular value decomposition of a $M \times R$ matrix has to be computed and stored at each time-step. Here, we need only two vectors, one indicating which region C_r^j contains the realization of $\overrightarrow{\mathbf{X}}_h^{j,m}$, and a second that stores the amount of visits $\#\{\overrightarrow{\mathbf{X}}_h^{j,m} \in C_r^j\}$ for each region C_r^j.

In order to verify convergence (see [GKKW02]) of the partitioning estimator (9.11) in the computation of the vectors in (9.7) resp. (9.8), truncation criteria were specified there which were never met in the simulations below. In addition, the condition $\mathbb{E}[|\Theta^j|^2] < \infty$ is needed and can be verified for each entry in the argument of the conditional expectations in (9.7) resp. (9.8). For a detailed summary and error analysis of the least squares Monte-Carlo method for BSDEs see the works [BD07, GLW05, GT14], the work [GKKW02] for a error analysis and a summary of partitioning estimates.

Partitioning estimates for BSDEs are discussed in [GLW05], including local hypercube basis functions or local Voronoi basis functions. Since the (forward) SPDE is a high dimensional problem (L dimensions with $L \sim h^{-d}$ for a discretization of a d-dimensional domain), a usual hypercube basis approach **(HC)** is not practicable, where each entry is discretized using e.g. a uniform mesh.

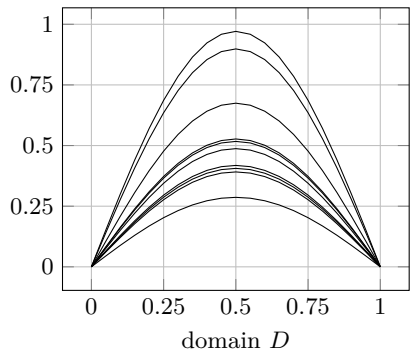
In the simulations which are carried out below, the (forward) SPDE (6.3) is driven by a discretization of colored (in space) noise, which acts on all L entries of the coefficient vector $\overrightarrow{\mathbf{X}}_h^j$, see e.g. Figure 9.1. The values at (neighboring) nodes show dependencies in those cases. We use numerical strategies which take advantage of the related spatial regularity of the solution and discretizes $\overrightarrow{\mathbf{X}}_h^j[\Omega]$ according to a partition of the space of functions $\mathbb{V}_h$, such as the Voronoi partition basis approach **(V)**.

Voronoi Partition Basis **(V)**:

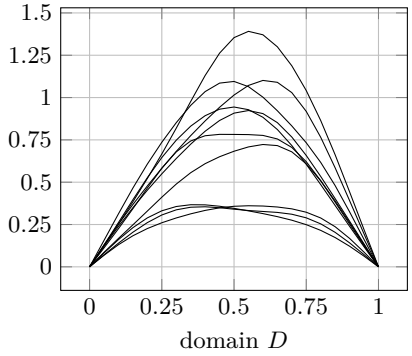
(I) Simulate R additional paths $\{\widehat{X}_{h,r}^j;\, j = 0, \dots, J\}$ of the (forward) SPDE.

(II) Define for each time-step $j = 1, \dots, J-1$:
$C_r^j := \{\Phi \in \mathbb{V}_h;\, \|\Phi - \widehat{X}_{h,r}^j\|_{\mathbb{L}^2} < \inf_{r \neq v} \|\Phi - \widehat{X}_{h,v}^j\|_{\mathbb{L}^2}\}$.

(III) Define the local basis function $\eta_r^j(\overrightarrow{\Phi}_h) := \mathbb{1}_{C_r^j}(\overrightarrow{\boldsymbol{\Upsilon}}_{\mathbb{V}_h}^{-1}(\overrightarrow{\Phi}_h))$ for $\overrightarrow{\Phi}_h \in \overrightarrow{\mathbf{X}}_h^j[\Omega]$.



(a) Setup A ($n = 1$, $\nu^1(x) \equiv 1$)



(b) Setup B ($n = 5$, $\nu^i(x) = \frac{1}{i}\sin(\pi x i)$)

Figure 9.1. Realizations of the approximation of the (forward) SPDE X_h^j at time $t_j = 0.2$ for Example 10.1 using Setups **A** and **B**.

As a second strategy we partition the set of functions $\overrightarrow{\mathbf{X}}_h^j[\Omega]$ into regions which are equally likely. This strategy is similar to the "adaptive local basis approach" in [BW12], but different in the construction. Suppose we want to construct $R = 2^{\texttt{amount}}$ many regions in $\mathbb{R}^L$. Suppose we construct $R = 2^{\texttt{amount}}$ many regions in $\mathbb{R}^L$.

Binary Tree Cuboids **(BTC)**:

(I) Simulate $V = P \cdot 2^{\texttt{paths}}$, $P \in \mathbb{N}$ many additional paths $\{\widehat{X}_{h,v}^j;\, j = 0, \dots, J\}$ of the (forward) SPDE ($V \gg R$ should hold).

(II) For each time-step $j = 1, \dots, J-1$ do:

(1) Define $\mathcal{S}_{0,1}^j$ consisting of $\{\widehat{X}_{h,v}^j;\, v = 1, \dots, V\}$.

(2) For $p = 1, \dots, \texttt{amount} + 1$ do:
For $q = 1, \dots, 2^{p-1}$ do:

- Define $\mathcal{S}^j := \mathcal{S}_{p-1,q}^j$ consisting of $\{\widehat{X}_{h,v'}^j;\, v' = 1, \dots, 2^{1-p} \cdot V\}$.
- Find the entry $\ell = 1, \dots, L$ in the coefficient vector $\overrightarrow{\boldsymbol{\Upsilon}}_{\mathbb{V}_h}(\widehat{X}_{h,v'}^j)$ for $\widehat{X}_{h,v'}^j \in \mathcal{S}^j$ which possesses the largest standard deviation; denote it by $\ell_{p,q}$.
- Compute the median $\texttt{med}_{p,q}$ of $\big[\overrightarrow{\boldsymbol{\Upsilon}}_{\mathbb{V}_h}(\widehat{X}_{h,v'}^j)\big]_{\ell_{p,q}}$.
- Divide the (sub-) sample $\mathcal{S}^j$ into two parts $\mathcal{S}^j = \mathcal{S}_{p,q}^j \cup \mathcal{S}_{p,2^{p-1}+q}^j$ according to whether $\big[\overrightarrow{\boldsymbol{\Upsilon}}_{\mathbb{V}_h}(\widehat{X}_{h,v'}^j)\big]_{\ell_{p,q}} > \texttt{med}_{p,q}$ holds or not. Note that both subsamples $\mathcal{S}_{p,q}^j$ and $\mathcal{S}_{p,2^{p-1}+q}^j$ contain the same amount of realizations $\widehat{X}_{h,v'}^j \in \mathcal{S}_{p-1,q}^j$.

(3) Set $R := 2^{\texttt{amount}}$. Construct the region C_r^j which contains all realizations of $\mathcal{S}_{\texttt{amount}+1,r}^j$ using the information $(\ell_{p,q}, \texttt{med}_{p,q})$.

(III) Define the local basis function $\eta_r^j(\overrightarrow{\Phi}_h) := \mathbb{1}_{C_r^j}(\overrightarrow{\Phi}_h)$ for $\overrightarrow{\Phi}_h \in \overrightarrow{\mathbf{X}}_h^j[\Omega]$.

The first two steps in the construction of the **(BTC)** mesh are illustrated in Figures 9.2 and 9.3.

This algorithm divides the initial sample $\mathcal{S}_{0,1}^j$ into R many subsamples each containing the same amount of realizations $\widehat{X}_{h,v}^j$. In order to identify the region C_r^j where a new realization of X_h^j has taken its value, we have to analyze the current binary tree. Therefore, only `amount` many checks are necessary. The crucial point of the approach **(BTC)** is to decide which entry $\ell = 1, \ldots, L$ of the coefficient vector should be divided into two parts. In the proposed procedure above and in all performed simulations, the entry which possesses the maximum standard deviation is taken, but other criteria are also possible.

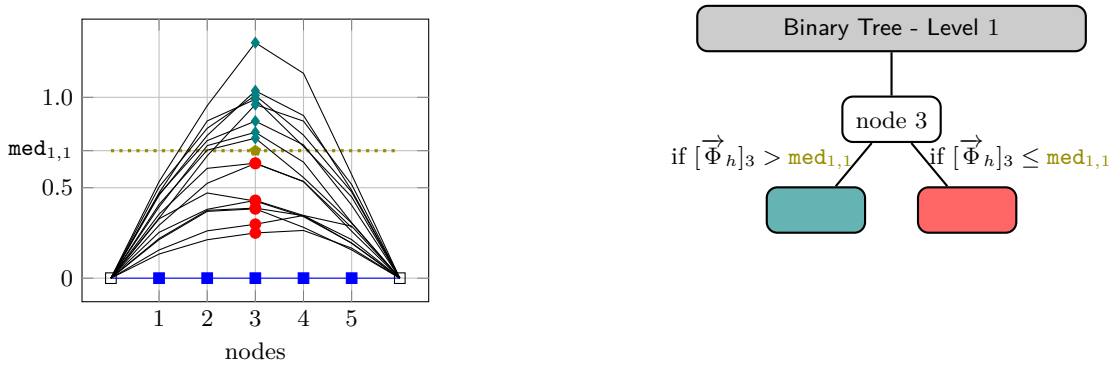


Figure 9.2. First step in the construction of the **(BTC)** mesh: The additional realizations of X_h^j (——) are divided according to the value at node 3 and the median $\mathtt{med}_{1,1}$ (·····) into two subsets $\mathcal{S}_{1,1}^j$ (♦) and $\mathcal{S}_{1,2}^j$ (●).

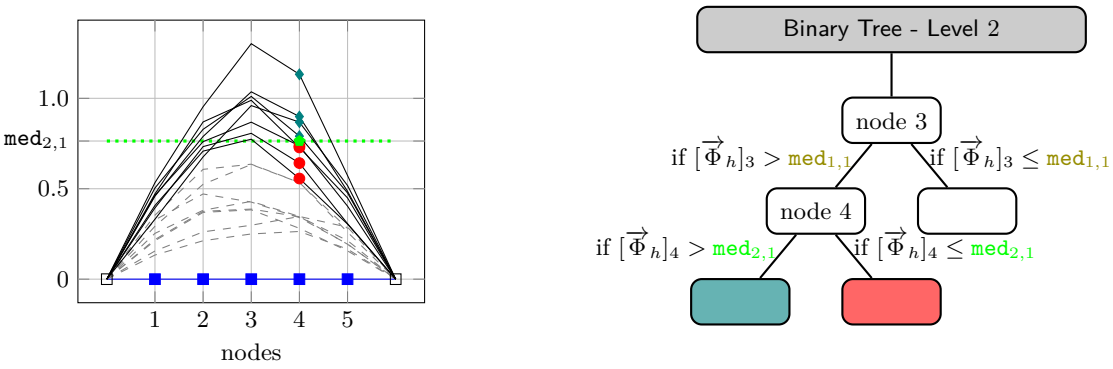


Figure 9.3. Second step in the construction of the **(BTC)** mesh: The realizations in $\mathcal{S}_{1,1}^j$ (——) are divided according to the value at node 4 and the median $\mathtt{med}_{2,1}$ (·····) into two subsets $\mathcal{S}_{2,1}^j$ (♦) and $\mathcal{S}_{2,2}^j$ (●).

The Voronoi partition basis **(V)**, as well as the Binary Tree Cuboids **(BTC)** construct the regions according to the distribution of X_h^j by evaluating additional independent simulations of X_h^j. The main difference is that the method **(V)** constructs one region based on a single additional realization, while **(BTC)** concentrates several realizations and constructs then one region for an ensemble of realizations: this makes the partition more robust. Despite of their success in simulations, a disadvantage of both "adaptive strategies" is the lack of

theoretical support, which exists for **(HC)**; see e.g. [GLW05]. For a general convergence analysis of partitioning estimation procedures similar to **(V)**, **(BTC)** and **(HC)** to estimate a single conditional expectation we refer to [GKKW02, Chapter 13].

A computational study of Scheme 9.1 in combination with **(BTC)** or **(V)** is performed in Section 10.1, where the scheme in Remark 9.3 is used to provide reference solutions in the special case of terminal data $g(x) = cx$.

9.2. Forward-backward stochastic heat equation

We consider two algorithms to simulate the FBSPDE (8.4)–(8.5) by approximating the deterministic functions via least squares Monte-Carlo. The first algorithm combines the Picard iteration (8.6)–(8.7) with Scheme 9.1. In practical studies the algorithm only terminates for moderate terminal times $T > 0$, which reflects the use of a contraction property in the proof of Theorem 7.2. To overcome this limitation, the new stochastic gradient method is proposed as a second algorithm in Subsection 9.2.2. The implicit Euler method is applied to both, the forward equation (8.4) as well as the backward equation (8.5).

A third strategy exploits the linearity of the problem and uses iterative substitution to arrive at representations of involved deterministic functions. This new scheme (see Remark 9.6) is restricted to stochastic linear-quadratic problems, and hence serves here as a source for reference data for the two algorithms above in this special case.

9.2.1. Picard type algorithm

We consider the following Picard type algorithm to resolve the forward-backward character of the system (8.4)–(8.5):

Scheme 9.4 (Picard type (ODP) scheme for stochastic control)

(1) Set $\mathcal{Y}_h^{(0),j}(\cdot) \equiv -\alpha U_h^{j,\text{init}}$ for each $j = 0, \dots, J-1$.

(2) Iterate $v = 1, 2, \dots$ until a stopping criterion is met:

(i) **FSPDE**: Compute $X_h^{(v),0} = \Pi_h x_0$. For each $j = 0, \dots, J-1$, simulate the $\mathbb{V}_h$-valued random variables $X_h^{(v),j+1}$ such that $\forall \phi_h \in \mathbb{V}_h$

$$\begin{aligned}&(X_h^{(v),j+1}, \phi_h) + k(\nabla X_h^{(v),j+1}, \nabla \phi_h)\\ &= (X_h^{(v),j}, \phi_h) - k(\frac{1}{\alpha}\mathcal{Y}_h^{(v-1),j}(X_h^{(v),j}), \phi_h) + \sum_{i=1}^{n}(\nu^i(t_j)X_h^{(v),j}, \phi_h)\Delta_j W^i.\end{aligned} \tag{9.12}$$

(ii) **BSPDE**: Set $Y_h^{(v),J} = \Pi_h\big[Dg(X_h^{(v),J})\big]$. For each $j = J-1, \dots, 0$, simulate the $\mathbb{V}_h$-valued random variables $Z_h^{(v),i,j}$ and $Y_h^{(v),j}$ such that $\forall \phi_h \in \mathbb{V}_h$

$$(Z_h^{(v),i,j}, \phi_h) = \frac{1}{k}\mathbb{E}\Big[\Delta_j W^i (Y_h^{(v),j+1}, \phi_h)\Big|\mathcal{F}_{t_j}\Big] \qquad \forall i = 1, \dots, n, \tag{9.13}$$

and

$$
\begin{aligned}
&(Y_h^{(v),j}, \phi_h) + k(\nabla Y_h^{(v),j}, \nabla \phi_h) \qquad (9.14)\\
&= \mathbb{E}\Big[(Y_h^{(v),j+1}, \phi_h) + k(X_h^{(v),j} - \widetilde{X}(t_j), \phi_h)\Big|\mathcal{F}_{t_j}\Big] + k\sum_{i=1}^{n}(\nu^i(t_j) Z_h^{(v),i,j}, \phi_h).
\end{aligned}
$$

Set $\mathcal{Y}_h^{(v),j}(X_h^{(v),j}) = Y_h^{(v),j}$.

This scheme is proposed in [BZ08] for general FBSDEs. If compared to the Picard iteration used in the proof of Theorem 8.4, the term $Y_h^{(v-1),j}$ in the state equation is replaced by $\mathcal{Y}_h^{(v-1),j}(X_h^{(v),j})$ to guarantee that the dimension of the underlying Markovian process $X_h^{(v)}$ does not increase with the amount of Picard iterations. Computational experiments show that constructing the basis of $\mathcal{Y}_h^{(v-1),j}(\cdot)$ according to the law of $X_h^{(v-1),j}$ while being evaluated by $X_h^{(v),j}$ in (9.12) does not seriously affect the simulation results. Similarly to Remark 9.3, a scheme may also be constructed from Scheme 9.4 for stochastic LQ problems, where the linearity in X_h of the drift in the adjoint equation (7.3) can be exploited to express the solution of (9.13)–(9.14) explicitly in terms of a deterministic function of the forward SPDE avoiding the approximation of conditional expectations (see also Remark 9.6). For general $g(\cdot)$ however the deterministic functions $\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^{(v),j}(\cdot)$ and $\overrightarrow{\boldsymbol{\mathcal{Z}}}_h^{(v),i,j}(\cdot)$ have to be approximated by $\overrightarrow{\boldsymbol{\mathcal{Y}}}_{h,R}^{(v),j}(\cdot)$ and $\overrightarrow{\boldsymbol{\mathcal{Z}}}_{h,R}^{(v),i,j}(\cdot)$ using the partition estimation together with the approaches **(V)** and **(BTC)**. Here we may again exploit linearity of the backward equation to reformulate equation (9.13)–(9.14) similarly to the reformulation (9.5)–(9.6) in order to avoid nested conditional expectations.

9.2.2. Stochastic gradient algorithm

This algorithm constructs iterates which successively decrease the functional (6.2), by using the maximum principle (7.4), where control iterates (after discretization in space and time) are approximated by the deterministic function $\mathcal{U}_{h,R}^j(\cdot) = \sum_{r=1}^R \mathbf{u}_{h,r}^j \mathbb{1}_{C_r^j}(\cdot)$ with the Voronoi partition method **(V)**. The formulation of the stochastic gradient method then uses the coefficients $\mathbf{u}_{h,r}^j$, and can be formulated as follows:

Scheme 9.5 (Stochastic gradient method)

(1) Set $\mathbf{u}_{h,r}^{(0),j} \equiv U_h^{j,\mathrm{init}}$ for each basis region $r = 1, \dots, R$ and each $j = 0, \dots, J-1$.

(2) Iterate $v = 1, 2, \dots$ until a stopping criterion is met:

(i) **FSPDE**: Compute $X_h^{(v),0} = \Pi_h x_0$. For each $j = 0, \dots, J-1$, simulate the $\mathbb{V}_h$-valued random variables $X_h^{(v),j+1}$ such that $\forall \phi_h \in \mathbb{V}_h$

$$
\begin{aligned}
&(X_h^{(v),j+1}, \phi_h) + k(\nabla X_h^{(v),j+1}, \nabla \phi_h)\\
&= (X_h^{(v),j}, \phi_h) + k(\mathcal{U}_{h,R}^{(v-1),j}(X_h^{(v),j}), \phi_h) + \sum_{i=1}^{n}(\nu^i(t_j) X_h^{(v),j}, \phi_h)\Delta_j W^i.
\end{aligned} \qquad (9.15)
$$

(ii) **BSPDE**: Set $Y_h^{(v),J} = \Pi_h[Dg(X_h^{(v),J})]$. For each $j = J-1,\dots,0$ simulate the $\mathbb{V}_h$-valued random variables $Z_h^{(v),i,j}$ and $Y_h^{(v),j}$ according to (9.13)–(9.14) and the partition estimation method. Obtain the approximation of the regression function $\mathcal{Y}_{h,R}^{(v),j}(\cdot) = \sum_{r=1}^R \mathbf{y}_{h,r}^{(v),j} \mathbb{1}_{C_r^{(v),j}}(\cdot)$.

(iii) **Gradient step**: Compute the coefficients $\mathbf{u}_{h,r}^{(v),j}$ of the function $\mathcal{U}_{h,R}^{(v),j}(\cdot) = \sum_{r=1}^R \mathbf{u}_{h,r}^{(v),j} \mathbb{1}_{C_r^{(v),j}}(\cdot)$ through

$$\mathbf{u}_h^{(v)} = \mathbf{u}_h^{(v-1)} - \sigma^{(v)} \mathbf{g}_h^{(v-1)}, \tag{9.16}$$

where $\mathbf{g}_h^{(v-1)} := -(\alpha \mathbf{u}_h^{(v-1)} + \mathbf{y}_h^{(v)})$.

Note that the objects $\mathbf{u}_h^{(v)}$, $\mathbf{g}_h^{(v-1)}$, and $\mathbf{y}_h^{(v)}$ in equation (9.16) involve all regions and relevant time steps $\{C_r^{(\cdot),j};\, r = 1,\dots,R,\, j = 0,\dots,J-1\}$. The basis regions $C_r^{(v),j}$ change in each iteration, since they depend on the law of $X_h^{(v),j}$. However, the same increments $\{\Delta_j W_r^i\}$ are used in the construction of the additional paths and thus in the construction of the basis regions $C_r^{(v),j}$. In (9.16) we use the coefficient $\mathbf{u}_{h,r}^{(v-1)}$ corresponding to the region $C_r^{(v-1),j}$ as precursory coefficient to update the coefficient $\mathbf{u}_{h,r}^{(v)}$ corresponding to the region $C_r^{(v),j}$. Computational experiments show that the regions $C_r^{(v),j}$ only slightly change and that this approximation does not seriously influence the simulation results. If compared to Scheme 9.4, the Scheme 9.5 terminates for arbitrary times $T > 0$ in all studies. For the computation of the step size $\sigma^{(v)}$ in equation (9.16) we consider an adaption of the Armijo method **(AR)**:

1. Approximate the current functional $\mathcal{J}^{(v-1)}$ using $\mathbf{u}_{h,r}^{(v-1),j}$.
2. Iterate $s = 0, 1, 2, \dots$ until a stopping criterion is met:
   - Compute $\overline{\mathbf{u}}_h^{(v),s} = \mathbf{u}_h^{(v-1)} - \sigma^\star \beta^s \mathbf{g}_h^{(v-1)}$.
   - Approximate the functional $\mathcal{J}^{(v-1),s}$ using $\overline{\mathbf{u}}_h^{(v),s}$.
   - Stop if $\mathcal{J}^{(v-1),s} - \mathcal{J}^{(v-1)} \leq -\underline{\sigma}\sigma^\star \beta^s \sum_{j=0}^{J-1} \sum_{r=1}^{R} \|\mathbf{g}_h^{(v-1),j,r}\|_{\mathbb{L}^2}^2$.
3. Set $\mathbf{u}_h^{(v)} := \overline{\mathbf{u}}_h^{(v),s}$.

This scheme offers a general strategy to solve stochastic optimal control problems. For the stochastic LQ problem (6.2)–(6.3) we may substitute iterates successively to obtain a formula for an approximation of the optimal feedback control without computing conditional expectations. This leads to restricted huge computational savings, and improved resolution, thus providing a convenient platform for comparative computational studies for the more general schemes above.

Remark 9.6

1. Suppose $g(x) := \frac{1}{2}\|x - \widetilde{X}(T)\|_{\mathbb{L}^2}^2$ in the cost functional (6.2). Consider equations (9.12)–(9.14) in the limit (i.e., without Picard iteration). The coefficient vectors $(\overrightarrow{\mathbf{Y}}_h^{\star,j}, \overrightarrow{\mathbf{Z}}_h^{\star,i,j})$

at time t_j of its corresponding reformulation as algebraic problems can be expressed using

$$\begin{aligned}\overrightarrow{\mathbf{X}}_h^{\star,j+1} = (\mathbf{Mass} + k\,\mathbf{Stiff})^{-1}\Big(\mathbf{Mass}\,\overrightarrow{\mathbf{X}}_h^{\star,j} - \frac{k}{\alpha}\,\mathbf{Mass}(\boldsymbol{A}_{\boldsymbol{Y}^{\star,j}}\overrightarrow{\mathbf{X}}_h^{\star,j} + \overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,j}}) \\ + \sum_{i'=1}^{n} \mathbf{Mass}^j_{\sigma^{i'}}\,\overrightarrow{\mathbf{X}}_h^{\star,j}\Delta_j W^{i'}\Big)\end{aligned} \tag{9.17}$$

by

$$\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^{\star,j}(\overrightarrow{\mathbf{X}}_h^{\star,j}) = \boldsymbol{A}_{\boldsymbol{Y}^{\star,j}}\overrightarrow{\mathbf{X}}_h^{\star,j} + \overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,j}}, \tag{9.18}$$

with (deterministic) $\boldsymbol{A}_{\boldsymbol{Y}^{\star,j}} \in \mathbb{R}^{L\times L}$ and $\overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,j}} \in \mathbb{R}^L$, which can be determined by the recursion:

a) Set $\boldsymbol{A}_{\boldsymbol{Y}^{\star,J}} := \boldsymbol{I}$ and $\overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,J}} := -\overrightarrow{\widetilde{\boldsymbol{X}}}_h^J$.

b) For $j = J-1$ to 0 compute

$$\begin{aligned}\boldsymbol{A}_{\boldsymbol{Y}^{\star,j}} := \Bigg(&\boldsymbol{I} + \frac{k}{\alpha}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\,\boldsymbol{A}_{\boldsymbol{Y}^{\star,j+1}}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass} \\ &+ \frac{k^2}{\alpha}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\Bigg)^{-1} \\ \cdot \Big(&\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\,\boldsymbol{A}_{\boldsymbol{Y}^{\star,j+1}}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass} \\ &+ k\sum_{i'=1}^{n}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}^j_{\nu^{i'}}\,\boldsymbol{A}_{\boldsymbol{Y}^{\star,j+1}}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}^j_{\sigma^{i'}} \\ &+ k\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\Big),\end{aligned}$$

and

$$\begin{aligned}\overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,j}} := \Bigg(&\boldsymbol{I} + \frac{k}{\alpha}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\,\boldsymbol{A}_{\boldsymbol{Y}^{\star,j+1}}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass} \\ &+ \frac{k^2}{\alpha}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\Bigg)^{-1} \\ \cdot \Big(&\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\,\overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,j+1}} - k\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}\,\overrightarrow{\widetilde{\boldsymbol{X}}}_h^{j+1}\Big).\end{aligned}$$

We motivate the first and the second step of this recursion: By $\overrightarrow{\mathbf{Y}}_h^{\star,J} = (\overrightarrow{\mathbf{X}}_h^{\star,J} - \overrightarrow{\widetilde{\boldsymbol{X}}}_h^J)$ and step a), we obtain (9.18) for $j = J$.
Now, use (9.18) and (9.17) to compute $\overrightarrow{\mathbf{Z}}_h^{\star,i,J-1}$

$$\begin{aligned}\overrightarrow{\mathbf{Z}}_h^{\star,i,J-1} &= \frac{1}{k}\mathbb{E}\big[\overrightarrow{\mathbf{Y}}_h^{\star,J}\Delta_{J-1}W^i|\mathcal{F}_{t_{J-1}}\big] = \frac{1}{k}\mathbb{E}\big[(\boldsymbol{A}_{\boldsymbol{Y}^{\star,J}}\overrightarrow{\mathbf{X}}_h^{\star,J} + \overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,J}})\Delta_{J-1}W^i|\mathcal{F}_{t_{J-1}}\big] \\ &= \boldsymbol{A}_{\boldsymbol{Y}^{\star,J}}\big(\mathbf{Mass} + k\,\mathbf{Stiff}\big)^{-1}\mathbf{Mass}^{J-1}_{\sigma^i}\,\overrightarrow{\mathbf{X}}_h^{\star,J-1}.\end{aligned}$$

Using this in $\overrightarrow{\mathbf{Y}}_h^{\star,J-1}$ yields

$$\begin{aligned}
\overrightarrow{\mathbf{Y}}_h^{\star,J-1} &= \mathbb{E}\Big[(\mathbf{Mass}+k\,\mathbf{Stiff})^{-1}\,\mathbf{Mass}(\overrightarrow{\mathbf{Y}}_h^{\star,J}+k(\overrightarrow{\mathbf{X}}_h^{\star,J}-\overrightarrow{\widetilde{\boldsymbol{X}}}_h^{J}))|\mathcal{F}_{t_{J-1}}\Big] \\
&\quad + k\sum_{i'=1}^{n}(\mathbf{Mass}+k\,\mathbf{Stiff})^{-1}\,\mathbf{Mass}_{\nu^{i'}}^{J-1}\,\overrightarrow{\mathbf{Z}}_h^{\star,i',J-1} \\
&= \mathbb{E}\Big[(\mathbf{Mass}+k\,\mathbf{Stiff})^{-1}\,\mathbf{Mass}(\boldsymbol{A}_{\boldsymbol{Y}^{\star,J}}\overrightarrow{\mathbf{X}}_h^{\star,J}+\overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,J}}+k(\overrightarrow{\mathbf{X}}_h^{\star,J}-\overrightarrow{\widetilde{\boldsymbol{X}}}_h^{J}))|\mathcal{F}_{t_{J-1}}\Big] \\
&\quad + k\sum_{i'=1}^{n}(\mathbf{Mass}+k\,\mathbf{Stiff})^{-1}\,\mathbf{Mass}_{\nu^{i'}}^{J-1}\,\boldsymbol{A}_{\boldsymbol{Y}^{\star,J}}(\mathbf{Mass}+k\,\mathbf{Stiff})^{-1}\,\mathbf{Mass}_{\sigma^{i'}}^{J-1}\,\overrightarrow{\mathbf{X}}_h^{\star,J-1} \\
&\overset{!}{=} \boldsymbol{A}_{\boldsymbol{Y}^{\star,J-1}}\overrightarrow{\mathbf{X}}_h^{\star,J-1}+\overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,J-1}}.
\end{aligned}$$

Inserting (9.17) for $\overrightarrow{\mathbf{X}}_h^{\star,J}$ and then identifying $\boldsymbol{A}_{\boldsymbol{Y}^{\star,J-1}}$ and $\overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,J-1}}$ yields the result. Note that (9.18) in combination with the discrete version of (7.4) yield

$$\overrightarrow{\boldsymbol{u}}_h^{\star,j}(\overrightarrow{\mathbf{X}}_h^{\star,j}) := -\frac{1}{\alpha}\overrightarrow{\boldsymbol{\mathcal{Y}}}_h^{\star,j}(\overrightarrow{\mathbf{X}}_h^{\star,j}) = -\frac{1}{\alpha}\Big(\boldsymbol{A}_{\boldsymbol{Y}^{\star,j}}\overrightarrow{\mathbf{X}}_h^{\star,j}+\overrightarrow{\boldsymbol{V}}_{\boldsymbol{Y}^{\star,j}}\Big), \tag{9.19}$$

which is a discrete version of the optimal feedback law.

2. A different approach to approximate the optimal feedback control of the stochastic LQ problem (6.2)–(6.3) would be via (space-) time discretization of the stochastic Riccati equation; see [YZ99] for the stochastic LQ problem involving an SDE, and [KK91] for the deterministic LQ problem involving a PDE. The latter strategy leads to a discrete version of the optimal control law which approaches but is different from (9.19), which rests on the time discretization Scheme 9.4 and successive substitution as detailed in 1.

10. Computational studies

We report on computational studies for Schemes 9.1 and 9.4 which utilize the different mesh strategies from Chapter 9. We study discretization effects for the BSPDE (6.1) in Section 10.1. Section 10.2 addresses the simulation of the FBSPDE (FBSHE) focussing on the Picard type method and the stochastic gradient method.

10.1. Backward stochastic heat equation

Our focus is on stability properties of computed iterates, behavior of the different partitioning approaches, as well as empirical rates of convergence w.r.t. space and time discretization. The studies for Example 10.1 evidence that

- a CFL condition is needed for the explicit Euler scheme, but not for its implicit variant;
- the **(BTC)** method yields improved results if compared with the **(V)** method;
- (space) the empirical order of convergence is slightly better if compared with the results in Theorem 8.2;
- (time) the empirical order of convergence is 0.5, which is stable w.r.t. the dimension of the discretized state space;
- a rougher resolution of the triangulation of the forward equation (10.1) ($h_F > h_B$) yields a comparable approximation of the backward equation (10.2) while saving computational time.

Example 10.1
Let $D \subset \mathbb{R}$. Consider $(0 \le t \le T)$

$$\mathrm{d}X(t) = \mu \Delta X(t)\,\mathrm{d}t + \sum_{i=1}^{n} \sigma^i(t) X(t)\,\mathrm{d}W^i(t), \qquad X(0) = x_0, \tag{10.1}$$

and

$$\mathrm{d}Y(t) = \Big[-\delta \Delta Y(t) - \sum_{i=1}^{n} \nu^i(t) Z^i(t)\Big]\,\mathrm{d}t + \sum_{i=1}^{n} Z^i(t)\,\mathrm{d}W^i(t), \qquad Y(T) = g(X(T)), \tag{10.2}$$

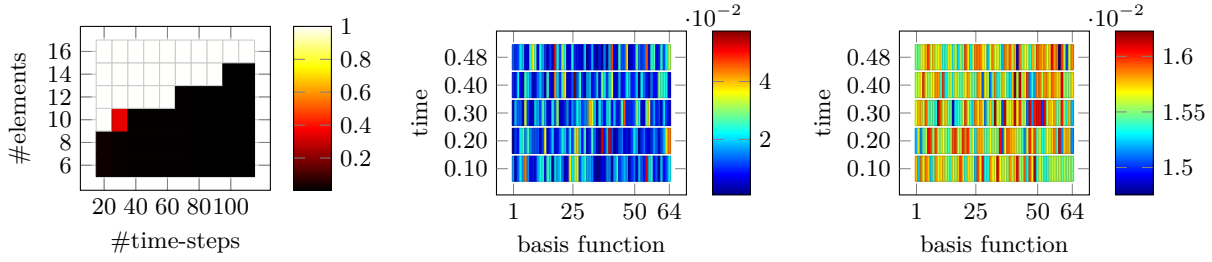
with homogeneous Dirichlet boundary conditions for both, X and Y.

Example 10.1 is studied for different choices of parameters, see Table 10.1: Setups **A** and **C** focus on multiplicative scalar noise, Setup **B** on a rough approximation of colored noise. If not specified differently, Scheme 9.1 together with the approaches **(V)** and **(BTC)** is used to simulate Example 10.1, where both equations (10.1)–(10.2) are discretized by $\mathbb{P}_1$-finite elements on the same triangulation with $h = 0.05$ and $k = 0.02$. For Setups **A** and **B**, the scheme from Remark 9.3 is used to provide reference solutions $(\widehat{\mathcal{Y}}_h^j(\cdot), \widehat{\mathcal{Z}}_h^{i,j}(\cdot))$.

Table 10.1. Parameter Setups **A**, **B**, and **C** for Example 10.1.

	D	T	μ	δ	$x_0(x)$	$g(x)$	n	$\sigma^i(t,x)$	$\nu^i(t,x)$
Setup A	$(0,1)$	0.50	0.20	0.20	$\sin(\pi x)$	$5x$	1	1.0	1.0
Setup B	$(0,1)$	0.50	0.20	0.20	$\sin(\pi x)$	$5x$	5	$\frac{1}{i}\sin(\pi x i)$	$\frac{1}{i}\sin(\pi x i)$
Setup C	$(0,1)$	0.50	0.20	0.20	$\sin(\pi x)$	$2x^3$	1	1.0	1.0

Scheme 9.1 is based on the implicit Euler method, which is more stable if compared with the explicit Euler method: the implicit and the explicit Euler method yield matchable results only in case the relation $k \leq h^2$ is met for the latter; see Figure 10.1 **(a)** which shows the difference $\mathbf{diff_{IE}} := \sup_{t_j}\big(\mathbb{E}[\|\mathcal{Y}_{h,R}^{j,IE}(X_h^j(\omega)) - \mathcal{Y}_{h,R}^{j,EE}(X_h^j(\omega))\|_{\mathbb{L}^2}^2]\big)^{1/2}$. Moderate values of the time discretization parameter k are preferred, due to the high computational demands, which is why the implicit Euler method is chosen below. Figures 10.1 **(b)**, **(c)** show how sampled realizations distribute in the regions C_r^j. A proper sampling is important since in each region C_r^j a (local) expectation value has to be approximated.

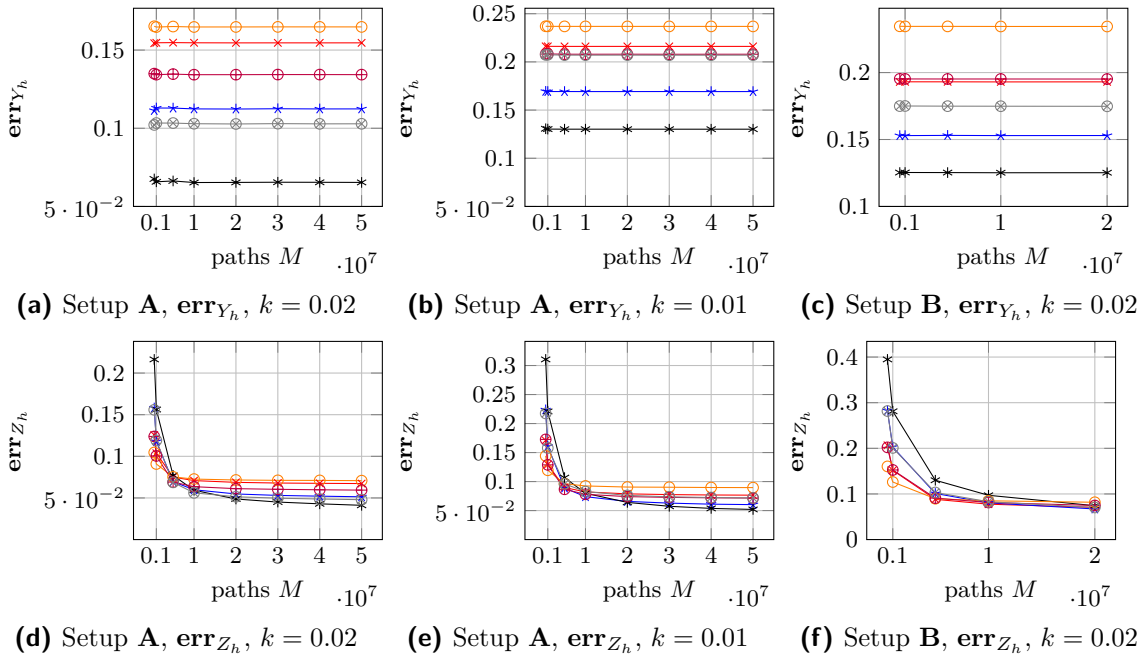


(a) $\mathbf{diff_{IE}} \wedge 1$ **(b)** **(V)** $R = 64$ **(c)** **(BTC)** $R = 64$

Figure 10.1. Empirical stability of the explicit Euler scheme using Setup **A** together with **(V)**, $R = 32$, and $M = 1.0 \cdot 10^6$ is shown in **(a)**. Frequency of the regions C_r^j during the computation of Setup **A** with $R = 64$ is shown in **(b)** and **(c)**.

Figure 10.2 shows that an accurate approximation of the BSDPE (10.2) can be achieved by using Scheme 9.1 and the partition estimation method. Moreover it shows that by increasing both, the amount of paths M as well as the amount of basis functions R, Scheme 9.1 with approaches **(V)** or **(BTC)** converge to the reference solution. This is quantified by analyzing the errors $\mathbf{err}_{Y_h} := \sup_{t_j}\big(\mathbb{E}[\|\widehat{\mathcal{Y}}_h^j(X_h^j) - \mathcal{Y}_{h,R}^j(X_h^j)\|_{\mathbb{L}^2}^2]\big)^{1/2}$ and $\mathbf{err}_{Z_h} := \big(\mathbb{E}[k\sum_{j=0}^{J-1}\sum_{i=1}^{n}\|\widehat{\mathcal{Z}}_h^{i,j}(X_h^j) - \mathcal{Z}_{h,R}^{i,j}(X_h^j)\|_{\mathbb{L}^2}^2]\big)^{1/2}$. The studies show that increasing the amount of paths M only improves the approximation of $\mathcal{Y}_h^j(\cdot)$ and $\mathcal{Z}_h^{i,j}(\cdot)$ up to a certain degree, which may then only be improved by increasing the number R of basis functions

and the amount of paths M simultaneously.



(a) Setup **A**, $\mathbf{err}_{Y_h}$, $k = 0.02$ **(b)** Setup **A**, $\mathbf{err}_{Y_h}$, $k = 0.01$ **(c)** Setup **B**, $\mathbf{err}_{Y_h}$, $k = 0.02$

(d) Setup **A**, $\mathbf{err}_{Z_h}$, $k = 0.02$ **(e)** Setup **A**, $\mathbf{err}_{Z_h}$, $k = 0.01$ **(f)** Setup **B**, $\mathbf{err}_{Z_h}$, $k = 0.02$

Figure 10.2. Behavior of the errors $\mathbf{err}_{Y_h}$ and $\mathbf{err}_{Z_h}$ of the approximation of Example 10.1 using the approach **(BTC)** $R = 64$ (—×—), $R = 128$ (—⋆—), $R = 256$ (—∗—), the approach **(V)** with $R = 32$ (—○—), $R = 64$ (—⊕—), and $R = 128$ (—⊗—).

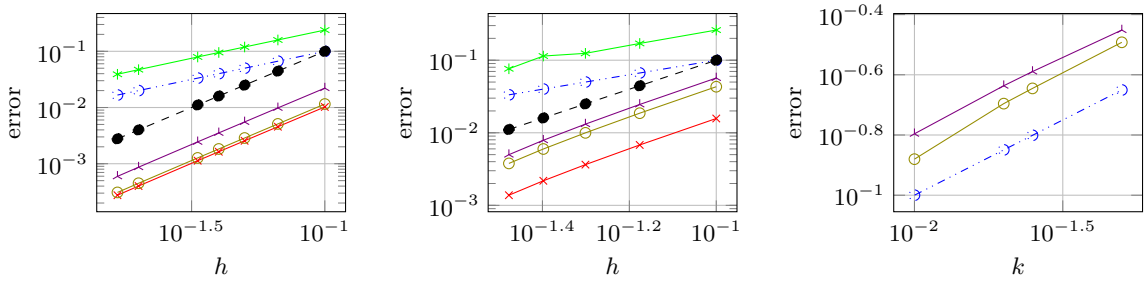
Next, we study the convergence behavior of (Y_h, Z_h) with respect to the space discretization parameter. For this purpose, we simulate paths to approximately solve Example 10.1

a) in the case of Setup **B** with the help of the formulas in Remark 9.3 using different space discretization parameters $h \in \{\frac{1}{10}, \frac{1}{15}, \frac{1}{20}, \frac{1}{25}, \frac{1}{30}, \frac{1}{50}, \frac{1}{60}\}$ and compare them with a reference solution $(\widehat{\mathcal{Y}}_h^j(\cdot), \widehat{\mathcal{Z}}_h^{i,j}(\cdot))$ which is simulated using $h^\star = 1/300$;

b) in the case of Setup **C** with the help of method **(V)** ($R = 128$, $M = 1.0 \cdot 10^7$) using different space discretization parameters $h \in \{\frac{1}{10}, \frac{1}{15}, \frac{1}{20}, \frac{1}{25}, \frac{1}{30}\}$ and compare them with a reference solution $(\widehat{\mathcal{Y}}_{h,R}^j(\cdot), \widehat{\mathcal{Z}}_{h,R}^{i,j}(\cdot))$ which is simulated using $h^\star = 1/60$.

Calculated errors which use 20.000 paths are illustrated in Figure 10.3. The empirical rate of convergence for $\left(\mathbb{E}\left[k \sum_{j=0}^J \|\widehat{\mathcal{Y}}_h^j(X_h^j) - \mathcal{Y}_h^j(X_h^j)\|_{\mathbb{L}^2}^2\right]\right)^{1/2}$ is 2.01, which coincides with Theorem 8.2. Surprisingly, the rate for Z obtained in the computational study is twice the rate that is given in Theorem 8.2.

The convergence behavior of (Y_h, Z_h) with respect to the time discretization parameter k is displayed in Figure 10.3 **(c)**. We simulate Example 10.1 in the case of Setup **B** with the help of the formulas in Remark 9.6 using different choices $k_i = T/N_i$ ($N_i \in \{10, 20, 25, 50\}$), and compare them with a reference solution $(\widehat{\mathcal{Y}}_h^j(\cdot), \widehat{\mathcal{Z}}_h^{i,j}(\cdot))$ which is simulated with $N^\star = 100$. We obtain an empirical rate close to 0.5 for both, the error in Y and in Z, which is stable w.r.t. refined spatial meshes. This observation complements the theoretical studies on time discretization in [Zha04], which motivates dependence on the dimension of the state space.



(a) Setup **B**, space **(b)** Setup **C**, space **(c)** Setup **B**, time

Figure 10.3. Rates of convergence of the space discretization **(a)**, **(b)** and time discretization **(c)** indicating the behavior of $(\mathbb{E}[k\sum_{j=0}^{J}\|\widehat{\mathcal{Y}}_h^j(X_h^j)-\mathcal{Y}_h^j(X_h^j)\|_{\mathbb{L}^2}^2])^{1/2}$ (—×—), $(\mathbb{E}[\sup_{t_j}\|\widehat{\mathcal{Y}}_h^j(X_h^j)-\mathcal{Y}_h^j(X_h^j)\|_{\mathbb{L}^2}^2])^{1/2}$ (—×—), $(\mathbb{E}[k\sum_{j=0}^{J}\|\nabla[\widehat{\mathcal{Y}}_h^j(X_h^j)-\mathcal{Y}_h^j(X_h^j)]\|_{\mathbb{L}^2}^2])^{1/2}$ (—×—), $(\mathbb{E}[k\sum_{j=0}^{J-1}\|\widehat{\mathcal{Z}}_h^j(X_h^j)-\mathcal{Z}_h^j(X_h^j)\|_{\mathbb{L}^2}^2])^{1/2}$ (—○—), as well as reference slopes h^1 (-·×·-), h^2 (-·●·-), $k^{1/2}$ (-·×·-).

In the simulations discussed so far, the same triangulation is used for both, the space discretization of the forward and for the backward equation. However, by choosing a rougher triangulation for the forward equation $h_F > h_B$, a significant amount of simulation time as well as memory can be saved, while simulations are comparable. This is pointed out in Table 10.2 for Setups **A**,**B**, and **C** which displays the differences

$$\mathbf{diff}_Y := \left(\mathbb{E}\left[\sup_{j=0,\dots,J}\frac{\|\mathcal{Y}^j_{(h_F,h_B),R}(X^j_{h_F})-\mathcal{Y}^j_{h_B,R}(X^j_{h_B})\|^2_{\mathbb{L}^2}}{\|\mathcal{Y}^j_{h_B,R}(X^j_{h_B})\|^2_{\mathbb{L}^2}}\right]\right)^{1/2},$$

$$\mathbf{diff}_Z := \left(\mathbb{E}\left[k\sum_{j=0}^{J-1}\frac{\sum_{i=1}^n\|\mathcal{Z}^{i,j}_{(h_F,h_B),R}(X^j_{h_F})-\mathcal{Z}^{i,j}_{h_B,R}(X^j_{h_B})\|^2_{\mathbb{L}^2}}{\sum_{i=1}^n\|\mathcal{Z}^{i,j}_{h_B,R}(X^j_{h_B})\|^2_{\mathbb{L}^2}}\right]\right)^{1/2}$$

for $(h_F,h_B)=(\frac{1}{10},\frac{1}{20})$, **(BTC)**, and **(V)** using $M=2.0\cdot10^7$ paths.

Table 10.2. Different mesh sizes in the triangulation of the forward paths: Differences $\mathbf{diff}_Y$ and $\mathbf{diff}_Z$, as well as the absolute simulation time computed on an Intel Core i5-4670 3.40GHz processor with 16GB RAM in double precision arithmetic.

			$\mathbf{diff}_Y$	$\mathbf{diff}_Z$	h_B time	(h_F,h_B) time
Setup A	**(BTC)**	$R=256$	0.0069	0.0014	5h 38min	3h 07min
	(V)	$R=128$	0.0069	0.0014	10h 21min	5h 44min
Setup B	**(BTC)**	$R=256$	0.0145	0.0147	12h 07min	6h 58min
	(V)	$R=128$	0.0551	0.0064	16h 49min	9h 34min
Setup C	**(BTC)**	$R=256$	0.0204	0.0041	5h 35min	3h 07min
	(V)	$R=128$	0.0204	0.0042	10h 20min	5h 42min

10.2. Forward-backward stochastic heat equation

In this section, we present computational studies for the forward-backward stochastic heat equation (FBSHE), which use the Picard type (ODP) scheme (i.e., Scheme 9.4, in combination with **(V)**), the stochastic gradient method (i.e., Scheme 9.5), and a direct computation avoiding the computation of the conditional expectations (Remark 9.6) in the case of the stochastic LQ problem.

The studies carried out for Example 10.2 show that

- to use the deterministic optimal control $U_h^{\star,det}$ as initial value U_h^{init} significantly reduces the number of iterations for both schemes;
- Scheme 9.5 returns the same solution as Scheme 9.4 in cases where Scheme 9.4 converges. Scheme 9.4 only converges for short durations $T > 0$, while Scheme 9.5 terminates for general T;
- the use of the regions $C_r^{(v-1),j}$ for the computation of the forward equation $X_h^{(v),j}$ in the v-th Picard iteration step or gradient iteration step does not crucially affect the simulations in both Schemes 9.4–9.5.
- (space) the empirical order of convergence is slightly better if compared with the results in Theorem 8.4;
- (time) the empirical order of convergence is slightly less then 0.5.

Example 10.2 (Stochastic optimal control problem)
Let $D \subset \mathbb{R}$. Denote by $\mu > 0$ a constant in front of the Laplacian in the state equation (6.3). Find a minimum of (6.2) subject to (6.3).

We consider three different approaches for its simulation: the Picard type (ODP) scheme (Scheme 9.4) with the partition estimation method, the stochastic gradient method in Scheme 9.5, both schemes using $R = 128$ Voronoi regions and $M = 1.0 \cdot 10^6$ paths, and the scheme from Remark 9.6 in the case of the stochastic LQ problem (i.e., $g(\cdot)$ quadratic in (6.2)–(6.3)). In each approach, we use $\mathbb{P}_1$-finite elements and the same triangulation for both, the state and adjoint equation. If not specified differently, we choose $h = 1/20$, $k = T/15$, and $g(x) := \frac{\kappa}{2}\|x - \widetilde{X}(T)\|_{\mathbb{L}^2}^2$ in our simulations. In all simulations, we approximate the cost functional $\mathcal{J}(X_h, U_h)$ with the help of $\widetilde{M} = 100.000$ paths, and the parameters of the Armijo rule are set to $\sigma = 0.01$, $\beta = 0.5$, and $\sigma^\star \in \{5, 10, 100\}$, depending of the setup.

Table 10.3. Parameter Setups **D**, **E**, **F**, and **G** for Example 10.2.

	D	T	μ	$x_0(x)$	n	$\nu^i(t,x)$	(δ, α, κ)	$\widetilde{X}(t,x)$
Setup D	$(0,1)$	0.40	0.15	$2\sin(\pi x)$	20	$\sin(\pi x i)$	$\left(1, \frac{1}{10000}, 0\right)$	$\Pi_h[\frac{T-t}{T}x_0]$
Setup E	$(0,1)$	0.05	0.20	$2\sin(\pi x)$	5	$\frac{2}{i}\sin(\pi x i)$	$\left(1, \frac{1}{5000}, 0\right)$	$\Pi_h[2]$
Setup F	$(0,1)$	0.25	0.15	$2\sin(\pi x)$	5	$\frac{2}{i}\sin(\pi x i)$	$\left(\frac{1}{2}, \frac{1}{5000}, \frac{1}{2}\right)$	$\Pi_h[\frac{T-t}{T}x_0]$
Setup G	$(0,1)$	0.25	0.15	$2\sin(\pi x)$	5	$\frac{2}{i}\sin(\pi x i)$	$\left(0, \frac{1}{5000}, 1\right)$	$\Pi_h[1]$

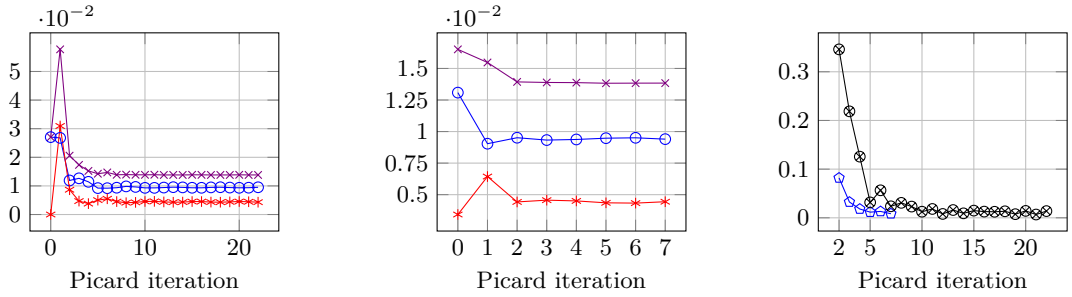
We consider different choices of parameters, see Table 10.3: Setup **E** focusses on a stochastic optimal control problem for small time durations s.t. the Picard algorithm terminates. Setups **F** and **G** reflect a more general setting, in which only the stochastic gradient method or the direct computation terminate.

One trajectory of the approximation of the optimal stochastic control $U_h^{\star,j}(\omega)$ and the corresponding optimal state $X_h^{\star,j}(\omega)$ is illustrated in Figure 6.1 using the formulas in Remark 9.6 for Setup **D** with $h = 1/60$ and $k = T/40$.

10.2.1. Simulations for short time durations $T > 0$

We choose the scheme from Remark 9.6 to compute a reference solution for Setup **E**; its cost functional is $\mathcal{J}(X_h^\star, U_h^\star) = 0.01374$, where $\mathbb{E}\big[\frac{\delta}{2} k \sum_{j=0}^J \|X_h^{\star,j} - \widetilde{X}(t_j)\|_{\mathbb{L}^2}^2\big] = 0.00932$, and $\mathbb{E}\big[\frac{\alpha}{2} k \sum_{j=0}^J \|U_h^{\star,j}\|_{\mathbb{L}^2}^2\big] = 0.00442$.

The first method under consideration is the Picard type (ODP) scheme. The Picard iteration stops if the distance of two consecutive cost functionals drops below a given tolerance $\texttt{TOL} = 5.0 \cdot 10^{-6}$. Figure 10.4 shows the decay $v \mapsto \mathcal{J}(X_h^{(v)}, \mathcal{U}_{h,R}^{(v)}(X_h^{(v)}))$ as the number of iterations increases. Two different initial controls $U_h^{j,init}$ are used for the simulation: the trivial control ($U_h^{j,init} \equiv 0$), and the optimal deterministic control ($U_h^{j,init} \equiv U_h^{\star,j,det}$) which we first compute by a steepest descent method; see e.g. [HPUU08]. Both initial values of the Picard iteration return almost the same approximation of the stochastic optimal control, which is close to the reference solution. However, to use the optimal deterministic control as initial value for $U_h^\star$ significantly reduces required Picard iterations (for Setup **F**: 7 steps compared with 22 steps).

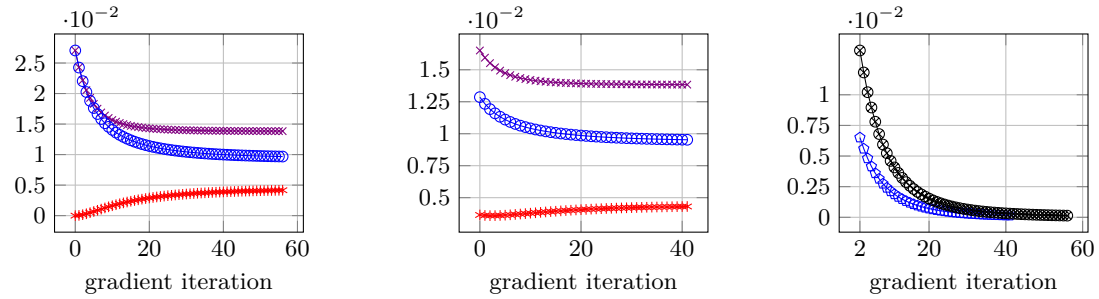


(a) $U_h^{j,init} \equiv 0$ **(b)** $U_h^{j,init} \equiv U_h^{\star,j,det}$ **(c)** Distance $\mathcal{V}_D^{(v)}$

Figure 10.4. The behavior of the cost functional $\mathcal{J}(X_h^{(v)}, \mathcal{U}_{h,R}^{(v)}(X_h^{(v)}))$ (—×—), its parts $\mathbb{E}\big[\frac{\delta}{2} k \sum_{j=0}^J \|X_h^{(v),j} - \widetilde{X}(t_j)\|_{\mathbb{L}^2}^2\big]$ (—○—), and $\mathbb{E}\big[\frac{\alpha}{2} k \sum_{j=0}^J \|\mathcal{U}_{h,R}^{(v),j}(X_h^{(v),j})\|_{\mathbb{L}^2}^2\big]$ (—✳—) in the simulation of Setup **E** by Scheme 9.4. The distance $\mathcal{V}_D^{(v)}$ is shown in part **(c)** for both initial controls ((—○—) for $U_h^{\star,j,det}$; (—⊗—) for the trivial control).

The stochastic gradient method (Scheme 9.5) returns a similar result as Scheme 9.4, and iterates monotonically decrease the functional value (Figure 10.5). The gradient iteration

stops if the "squared norm of the gradient" $\mathcal{G}^{(v)} := \frac{1}{R}\sum_{r=1}^{R} k \sum_{j=0}^{J} \|\mathbf{g}_{h,r}^{(v-1),j}\|_{\mathbb{L}^2}^2$ is less than a given tolerance $\texttt{TOL} = 1.0 \cdot 10^{-6}$.



(a) $U_h^{j,init} \equiv 0$ **(b)** $U_h^{j,init} \equiv U_h^{\star,j,det}$ **(c)** Distance $\mathcal{V}_D^{(v)}$

Figure 10.5. The behavior of the cost functional, its parts, and the distance $\mathcal{V}_D^{(v)}$ in the simulation of Setup **E** by Scheme 9.5.

Smallness of the variation between two consecutive Voronoi partitions is motivated in Figures 10.4 **(c)** and 10.5 **(c)** by quantifying $\mathcal{V}_D^{(v)} := \big(\frac{1}{R}\sum_{r=1}^{R} k \sum_{j=1}^{J-1} \|\widehat{X}_{h,r}^{(v),j} - \widehat{X}_{h,r}^{(v-1),j}\|_{\mathbb{L}^2}^2\big)^{1/2}$: we observe that constructing the coefficients and the basis functions in $\mathcal{Y}_{h,R}^{(v-1),j}$, resp. $\mathcal{U}_{h,R}^{(v-1),j}$, according to the law of $X_h^{(v-1),j}$ in (9.12), resp. (9.15), rather than $X_h^{(v),j}$ does not seriously affect the simulation results.

10.2.2. Simulations for general settings

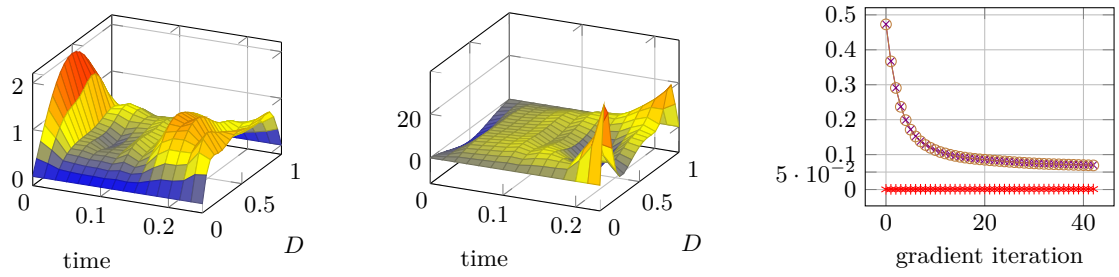
Next we consider the linear-convex stochastic optimal control problem (6.2)–(6.3) for Setups **F** and **G**, where the Picard type (ODP) scheme (Scheme 9.4) diverges. Table 10.4 contains the resulting cost functionals and parts in the case of Setup **F** simulated using the stochastic gradient method and the formulas of Remark 9.6. The approximation by using stochastic gradient method improves by increasing the amount of regions R.

Table 10.4. Comparison of the resulting cost functionals using different controls and methods for Setup **F**.

	$\mathcal{J}(X_h^\star, U_h^\star)$	$\mathbb{E}\big[\frac{\delta k}{2}\sum_j \|X_h^{\star,j} - \widetilde{X}(t_j)\|_{\mathbb{L}^2}^2\big]$	$\mathbb{E}\big[\frac{\alpha k}{2}\sum_j \|U_h^{\star,j}\|_{\mathbb{L}^2}^2\big]$	$\mathbb{E}\big[g(X_h^{\star,J})\big]$
$U_h^\star \equiv 0$	0.112709	0.013530	0.0	0.099170
$U_h^\star \equiv U_h^{\star,det}$	0.085636	0.011064	0.000055	0.074518
Scheme 9.5 $R = 32$	0.011940	0.003437	0.000391	0.008112
Scheme 9.5 $R = 64$	0.011681	0.003353	0.000386	0.007942
Scheme 9.5 $R = 128$	0.011407	0.003371	0.000375	0.007661
Remark 9.6	0.009969	0.002004	0.000681	0.007284

Consider Setup **G** where $g(x) := \kappa\big(\int_D (1 + |(x(r) - \widetilde{X}(T,r)|^2)^{p/2}\,\mathrm{d}r - 1\big)$ with $p = 1.5$. This problem is not of LQ-type, and thus excludes its numerical approximation by the scheme from Remark 9.6. The terminal time $T > 0$ is large such that Scheme 9.4 fails to converge,

while Scheme 9.5 does. One path of the (approximate) optimal solution, and a plot of the decay of the functional along the computed iterates are shown in Figure 10.6.



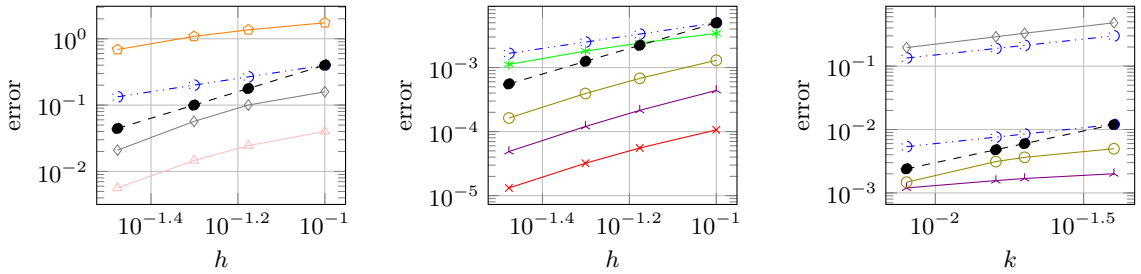
(a) $t_j \mapsto X_h^{\star,j}(\omega)$ **(b)** $t_j \mapsto \mathcal{U}_{h,R}^{\star,j}(X_h^{\star,j}(\omega))$ **(c)** cost functional

Figure 10.6. One path of the optimal state $X_h^{\star,j}$ and the optimal control $\mathcal{U}_{h,R}^{\star,j}(X_h^{\star,j}(\omega))$ for Setup **G** using the stochastic gradient method. The cost functional and its parts is shown in part **(c)**, where $\mathbb{E}[g(X_h^{(v),J})]$ is visualized by (—○—).

Rates of convergence

We study the convergence behavior of $(X_h^\star, Y_h^\star, Z_h^\star)$ with respect to time and space discretization using Example 10.2 with Setup **D** and fixed $k = T/100$. We simulate paths to approximately solve Example 10.2 with the help of the formulas in Remark 9.6 using different space discretization parameters $h \in \{\frac{1}{10}, \frac{1}{15}, \frac{1}{20}, \frac{1}{25}, \frac{1}{30}\}$, and compare them with a reference solution $(\widehat{X}_h^{\star,j}, \widehat{\mathcal{Y}}_h^{\star,j}(\widehat{X}_h^{\star,j}), \widehat{\mathcal{Z}}_h^{\star,i,j}(\widehat{X}_h^{\star,j}))$, which is simulated using $h^\star = 1/60$. Calculated errors which use 20.000 paths are illustrated in Figure 10.7. We find second order of convergence for $(\mathbb{E}[k\sum_{j=0}^J \|\widehat{\mathcal{Y}}_h^{\star,j}(X_h^{\star,j}) - \mathcal{Y}_h^{\star,j}(X_h^{\star,j})\|_{\mathbb{L}^2}^2])^{1/2}$ and $(\mathbb{E}[k\sum_{j=0}^J \|\widehat{X}_h^{\star,j} - X_h^{\star,j}\|_{\mathbb{L}^2}^2])^{1/2}$ as is stated in Theorem 8.4. Similarly as in the case of the BSHE, the observed rate for $Z^\star$ is twice the rate of what is given in Theorem 8.4.

Finally, we turn to time discretization errors which so far lack a theoretical analysis. For empirical evidence of related rates, we fix $h = 1/40$ and simulate paths to approximately solve Example 10.2 with the help of the formulas in Remark 9.6 using different discretization parameters $k_i = T/N_i$ with $N_i \in \{10, 20, 25, 50\}$, and compare them with a reference solution $(\widehat{X}_h^{\star,j}, \widehat{\mathcal{Y}}_h^{\star,j}(\widehat{X}_h^{\star,j}), \widehat{\mathcal{Z}}_h^{\star,i,j}(\widehat{X}_h^{\star,j}))$, which is simulated using $N^\star = 100$. We observe a rate close to 0.5 for both, state and adjoint equation, while the rate improves for $Z^\star$.



(a) Setup **D**, $X_h^\star$, space **(b)** Setup **D**, $(Y_h^\star, Z_h^\star)$, space **(c)** Setup **D**, time

Figure 10.7. Rates of convergence of the space discretization **(a)**, **(b)** and time discretization **(c)** indicating the behavior of $(\mathbb{E}[k\sum_{j=0}^{J}\|\widehat{X}_h^{\star,j} - X_h^{\star,j}\|_{\mathbb{L}^2}^2])^{1/2}$ (—△—), $(\mathbb{E}[\sup_{t_j}\|\widehat{X}_h^{\star,j} - X_h^{\star,j}\|_{\mathbb{L}^2}^2])^{1/2}$ (—◇—), $(\mathbb{E}[k\sum_{j=0}^{J}\|\nabla[\widehat{X}_h^{\star,j} - X_h^{\star,j}]\|_{\mathbb{L}^2}^2])^{1/2}$ (—⬠—), $(\mathbb{E}[k\sum_{j=0}^{J}\|\widehat{\mathcal{Y}}_h^{\star,j}(X_h^{\star,j}) - \mathcal{Y}_h^{\star,j}(X_h^{\star,j})\|_{\mathbb{L}^2}^2])^{1/2}$ (—×—), $(\mathbb{E}[\sup_{t_j}\|\widehat{\mathcal{Y}}_h^{\star,j}(X_h^{\star,j}) - \mathcal{Y}_h^{\star,j}(X_h^{\star,j})\|_{\mathbb{L}^2}^2])^{1/2}$ (—+—), $(\mathbb{E}[k\sum_{j=0}^{J}\|\nabla[\widehat{\mathcal{Y}}_h^{\star,j}(X_h^{\star,j}) - \mathcal{Y}_h^{\star,j}(X_h^{\star,j})]\|_{\mathbb{L}^2}^2])^{1/2}$ (—✳—), $(\mathbb{E}[k\sum_{j=0}^{J-1}\|\widehat{\mathcal{Z}}_h^{\star,j}(X_h^{\star,j}) - \mathcal{Z}_h^{\star,j}(X_h^{\star,j})\|_{\mathbb{L}^2}^2])^{1/2}$ (—○—), as well as reference slopes h^1 (··›··), h^2 (-●-), $k^{1/2}$ (··›··), k^1 (-●-).

Part III.

Stochastic optimal control of finite ensembles of nanomagnets

11. Introduction

We consider the magnetization process $\boldsymbol{m} := (\boldsymbol{m}_1, \dots, \boldsymbol{m}_N) : [0,T] \times \Omega \to (\mathbb{S}^2)^N$ within the finite network of $N \in \mathbb{N}$ interacting ferromagnetic particles in the presence of thermal fluctuations. A relevant application of this setting are single-molecule nanomagnets [FS10] where the dynamics of each particle is described by the magnetization $\boldsymbol{m}_i : [0,T] \times \Omega \to \mathbb{S}^2$, which is coupled with the remaining ones via the SDE system ($\alpha > 0$; $\nu \geq 0$)

$$\begin{aligned} \mathrm{d}\boldsymbol{m}_i &= \Big(\boldsymbol{m}_i \times \boldsymbol{\mathcal{H}}_{\mathit{eff},i}(\boldsymbol{m}, \boldsymbol{u}_i) - \alpha \boldsymbol{m}_i \times \big(\boldsymbol{m}_i \times \boldsymbol{\mathcal{H}}_{\mathit{eff},i}(\boldsymbol{m}, \boldsymbol{u}_i)\big)\Big)\,\mathrm{d}t + \nu \boldsymbol{m}_i \times \circ\, \mathrm{d}\boldsymbol{W}_i, \\ \boldsymbol{m}_i(0) &= \boldsymbol{m}_{0,i}. \end{aligned} \tag{11.1}$$

where $\boldsymbol{\mathcal{H}}_{\mathit{eff},i}(\boldsymbol{m}, \boldsymbol{u}_i) = \boldsymbol{\mathcal{H}}_{\mathit{ani},i}(\boldsymbol{m}_i) + \boldsymbol{\mathcal{H}}_{d,i}(\boldsymbol{m}_i) + \boldsymbol{\mathcal{H}}_{\mathit{exch},i}(\boldsymbol{m}) + \boldsymbol{\mathcal{H}}_{\mathit{ext},i}(\boldsymbol{u}_i)$ denotes the effective field, which combines forces related to

- the anisotropy energy where $\boldsymbol{\mathcal{H}}_{\mathit{ani},i}(\boldsymbol{m}_i) = C_{\mathit{ani}} \boldsymbol{A} \boldsymbol{m}_i$, with $\boldsymbol{A} \in \mathbb{R}^{3\times 3}_{\mathit{diag}}$, to favor magnetizations $\boldsymbol{m}_i$ which are (anti-)parallel to the easy axis $\boldsymbol{e} \in \mathbb{R}^3$,
- the stray field, which prefers magnetizations $\boldsymbol{m}_i$ without surface charges; for simplicity we choose $\boldsymbol{\mathcal{H}}_{d,i}(\boldsymbol{m}_i) = -C_d \boldsymbol{B}_i \boldsymbol{m}_i$, with $\boldsymbol{B}_i \in \mathbb{R}^{3\times 3}_{\mathit{diag}}$; c.f. [AB09, ACLP11],
- the exchange energy which penalizes non-alignment of (neighboring) magnetizations via $\boldsymbol{\mathcal{H}}_{\mathit{exch},i}(\boldsymbol{m}) = C_{\mathit{exch}} (\boldsymbol{C}\boldsymbol{m})_i$, for some positive definite $\boldsymbol{C} \in \mathbb{R}^{3N\times 3N}_{\mathit{sym}}$, and
- the external force $\boldsymbol{\mathcal{H}}_{\mathit{ext},i}(\boldsymbol{u}_i) = \boldsymbol{u}_i$.

The constants $C_{\mathit{ani}}, C_d, C_{\mathit{exch}} \geq 0$ account for the strength of these forces. We refer to [BBNP13,BMS09,NP13] for further details on the model. In (11.1), let $\boldsymbol{W} := (\boldsymbol{W}_1, \dots, \boldsymbol{W}_N)$ denote an $(\mathbb{R}^3)^N$-valued Wiener process on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ to represent uncontrolled thermal fluctuations from a surrounding heat bath. A practically relevant task is to control switching dynamics of ferromagnetic spins; for example, controlled precessional switching requires to properly adjust the intensity and duration to initially overcome anisotropic forces of magnetizations which are aligned with the easy axis, and to later reduce this field to eventually allow relaxation forces to take over; see Figure 15.1, and [BMS09] for further details on this problem. This crossing of energy barriers is alleviated at positive temperatures which in the model is represented by the stochastic forcing term.

First studies concerning the deterministic optimal control of ferromagnetic dynamics are carried out in [AB09, ACLP11] in the case of an finite ensemble of nanomagnetic particles, and in [DKPS15] for infinitely many particles, i.e. (11.1) for $\nu = 0$ is a PDE. To our knowledge, this is the first work which deals with the related stochastic control problem to optimally control magnetization dynamics in the presence of thermal noise.

Problem 11.1 (Stochastic control problem)
Let $\delta, \kappa, \nu \geq 0$, $T, \alpha, \lambda > 0$, and $N \in \mathbb{N}$. Let $\mathbb{U} \subset (\mathbb{R}^3)^N$, $\widetilde{\boldsymbol{m}} \equiv (\widetilde{\boldsymbol{m}}_1, \dots, \widetilde{\boldsymbol{m}}_N) \in L^2(0,T;(\mathbb{S}^2)^N)$, $h \in C^1((\mathbb{S}^2)^N;\mathbb{R})$, and $\boldsymbol{m}_0 \equiv (\boldsymbol{m}_{0,1}, \dots, \boldsymbol{m}_{0,N}) \in (\mathbb{S}^2)^N$ be given. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space, and $\boldsymbol{W}$ an $(\mathbb{R}^3)^N$-valued Wiener process. Find a tuple

$$(\boldsymbol{m}^\star, \boldsymbol{u}^\star) := (\boldsymbol{m}_1^\star, \dots, \boldsymbol{m}_N^\star; \boldsymbol{u}_1^\star, \dots, \boldsymbol{u}_N^\star) \in L^2_{\mathbb{F}}\Big(\Omega; C\big([0,T];(\mathbb{S}^2)^N\big) \times L^2\big(0,T;\mathbb{U}\big)\Big)$$

which minimizes

$$\mathcal{J}_{sto}(\boldsymbol{m}, \boldsymbol{u}) := \frac{1}{2}\mathbb{E}\Big[\int_0^T \Big(\delta\|\boldsymbol{m}(s) - \widetilde{\boldsymbol{m}}(s)\|^2_{(\mathbb{R}^3)^N} + \lambda\|\boldsymbol{u}(s)\|^2_{(\mathbb{R}^3)^N}\Big)\,\mathrm{d}s + \kappa h\big(\boldsymbol{m}(T)\big)\Big] \tag{11.2}$$

subject to equation (11.1).

The dependence on the control in the functional is quadratic, while it is linear in the drift part of the SDE. This observation is relevant in Chapter 12, where we sketch the standard construction of a weak solution $\boldsymbol{\pi} := (\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \boldsymbol{W}, \boldsymbol{u}^\star)$ of Problem 11.1 for Assumption C_1 via relaxation using Young measure-valued $\mathbb{F}$-adapted controls. The computational simulation algorithm in Chapter 14 then bases on a discretization of the related necessary optimality conditions with $(\mathbb{S}^2)^N$-valued magnetization iterates, which is a coupled system of forward-backward stochastic differential equations. The stochastic gradient method (see Part II of the thesis) is then adopted to the present nonlinear SDE (11.1) to decrease the energy $\mathcal{J}_{sto}$ along a finite sequence of controls, where updated search directions are obtained via representations by approximate regression functions, which in turn are computed by a least squares Monte-Carlo method to approximate involved conditional expectations. Computational studies in Chapter 15 are reported which control switching dynamics of single resp. multiple ferromagnetic chains in a surrounding heat bath to e.g. evidence different high- vs. low-dimensional optimal controls.

12. Optimal control of a ferromagnetic N-particle system

We prove existence of a weak stochastic control which solves Problem 11.1 via studying the relaxed stochastic control problem, following the general setup in [YZ99]. Therefore we need the following assumption.

Assumption $\mathbf{C_1}$
Let $\mathbf{0} \in \mathbb{U} \subset (\mathbb{R}^3)^N$ be compact.

Let $(\mathbb{X}, \|\cdot\|_{\mathbb{X}})$ be the Polish space $\big(C([0,T];(\mathbb{R}^3)^N), \|\cdot\|_{sup}\big)$ or $\big(L^2(0,T;(\mathbb{R}^3)^N), \|\cdot\|_{L^2}\big)$. Below, let $L^p_{\mathbb{F}}(\Omega;\mathbb{X})$ denote the space of all $\mathbb{F}$-adapted $\mathbb{X}$-valued random variables X, such that $\mathbb{E}[\|X\|^p_{\mathbb{X}}] < \infty$. We denote by $\mathcal{V}(T;\mathbb{U})$ the space of all non-negative Radon measures $\boldsymbol{\lambda}$ on $[0,T] \times \mathbb{U}$ such that

$$\boldsymbol{\lambda}(B \times \mathbb{U}) = |B| \quad \forall B \in \mathscr{B}([0,T]).$$

Then, $\boldsymbol{\lambda} \in \mathcal{V}(T;\mathbb{U})$ can be represented by $\boldsymbol{\lambda}(\,\mathrm{d}t,\,\mathrm{d}u) = \boldsymbol{\nu}(t,\,\mathrm{d}u)\,\mathrm{d}t$ for almost all $t \in [0,T]$, where $\boldsymbol{\nu}(t,\cdot) \in \mathcal{P}(\mathbb{U})$ denotes the Radon-Nikodym derivative of $\boldsymbol{\lambda}$. By Assumption C_1, $(\mathcal{V}(T;\mathbb{U}), \overset{\star}{\rightharpoonup})$ is a compact Polish space.
A weak stochastic admissible control for Problem 11.1 is defined as follows:

Definition 12.1 (Weak stochastic admissible control)
The 6-tuple $\boldsymbol{\pi} := \big(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \boldsymbol{W}, \boldsymbol{u}\big)$ is called a weak stochastic admissible control, if

1. $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is a filtered probability space satisfying the usual assumptions;
2. $\boldsymbol{W}$ is an $\mathbb{F}$-adapted $(\mathbb{R}^3)^N$-valued Wiener process on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$;
3. $\boldsymbol{u}$ is an $\mathbb{F}$-adapted control, such that $\boldsymbol{u} \in L^2_{\mathbb{F}}(\Omega; L^2(0,T;\mathbb{U}))$;
4. there exists an $\mathbb{F}$-adapted unique strong solution $\boldsymbol{m} \in L^2_{\mathbb{F}}(\Omega; C([0,T];(\mathbb{R}^3)^N))$ of the state equation (11.1) under $\boldsymbol{u}$ and $\boldsymbol{m}_0$;
5. $(\boldsymbol{m}, \boldsymbol{u})$ satisfies $\mathcal{J}_{sto}(\boldsymbol{m}, \boldsymbol{u}) < \infty$.

By $\mathfrak{U}_{ad}(\boldsymbol{m}_0; T)$ we denote the space of all weak stochastic admissible controls.

Then we can precise Problem 11.1: Find $\boldsymbol{\pi}^\star \in \mathfrak{U}_{ad}(\boldsymbol{m}_0; T)$ which minimizes

$$\mathfrak{J}_{sto}(\boldsymbol{\pi}) := \frac{1}{2}\mathbb{E}\Big[\int_0^T \Big(\delta\|\boldsymbol{m}(s) - \widetilde{\boldsymbol{m}}(s)\|^2_{(\mathbb{R}^3)^N} + \lambda\|\boldsymbol{u}(s)\|^2_{(\mathbb{R}^3)^N}\Big)\,\mathrm{d}s + \kappa h\big(\boldsymbol{m}(T)\big)\Big]. \tag{12.1}$$

Theorem 12.2 (Existence of a weak stochastic control)
Let $T > 0$ and $\boldsymbol{m}_0 \in (\mathbb{S}^2)^N$ be fixed and Assumption C_1 be fulfilled. There exists a weak optimal control $\boldsymbol{\pi}^\star \in \mathfrak{U}_{ad}(\boldsymbol{m}_0; T)$, i.e.

$$\mathfrak{J}_{sto}(\boldsymbol{\pi}^\star) = \inf_{\boldsymbol{\pi} \in \mathfrak{U}_{ad}(\boldsymbol{m}_0;T)} \mathfrak{J}_{sto}(\boldsymbol{\pi}).$$

Proof
The proof follows the general setup of [YZ99, Chapter 2, Theorem 5.3], which is not directly applicable since assumption (SE3), known as "Roxin's convexity condition", is not valid here.
Choose a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and let $\boldsymbol{W}$ be an $\mathbb{F}$-adapted $(\mathbb{R}^3)^N$-valued Wiener process on it. We have $\mathfrak{U}_{ad}(\boldsymbol{m}_0; T) \neq \varnothing$, since for $\boldsymbol{u} \equiv \mathbf{0}$ there exists an $\mathbb{F}$-adapted unique strong solution $\boldsymbol{m} \in L^2_{\mathbb{F}}(\Omega; C([0,T]; (\mathbb{R}^3)^N))$ of equation (11.1).
Let now $\boldsymbol{\pi}^n = (\Omega^n, \mathcal{F}^n, \mathbb{F}^n, \mathbb{P}^n, \boldsymbol{W}^n, \boldsymbol{u}^n)$ be a minimizing sequence of weak admissible controls for (12.1), that is

$$\lim_{n\to\infty} \mathfrak{J}_{sto}(\boldsymbol{\pi}^n) = \inf_{\boldsymbol{\pi} \in \mathfrak{U}_{ad}(\boldsymbol{m}_0;T)} \mathfrak{J}_{sto}(\boldsymbol{\pi}) =: \mathfrak{M} > -\infty.$$

We first rewrite equation (11.1). The noise term in equation (11.1) is in Stratonovich form, which can be reformulated to $(1 \leq i \leq N)$

$$\nu \mathbf{m}_i \times \circ \, \mathrm{d}\boldsymbol{W}_i = \frac{\nu^2}{2} \sum_{l=1}^{3} (\boldsymbol{m}_i \times \mathbf{e}_l) \times \mathbf{e}_l \, \mathrm{d}t + \nu \boldsymbol{m}_i \times \mathrm{d}\boldsymbol{W}_i, \tag{12.2}$$

where $\mathbf{e}_l \in \mathbb{R}^3$ are unit vectors for $l = 1, 2, 3$. Then equation (11.1) can be rewritten as

$$\mathrm{d}\boldsymbol{m}^n = \big(\mathbf{b}_I(\boldsymbol{m}^n) + \mathbf{b}_{II}(\boldsymbol{m}^n, \boldsymbol{u}^n)\big) \, \mathrm{d}t + \boldsymbol{\sigma}(\boldsymbol{m}^n) \, \mathrm{d}\boldsymbol{W}^n, \qquad \boldsymbol{m}^n(0) = \boldsymbol{m}_0, \tag{12.3}$$

where $\mathbf{b}_I(\cdot)$ combines the drift terms of (11.1) together with the Itô correction term from (12.2), which all are independent of the control $\boldsymbol{u}$. The control dependent terms are composed in $\mathbf{b}_{II}(\cdot, \cdot)$, such that for all $i = 1, \ldots, N$

$$\begin{aligned} \mathbf{b}_{I,i}(\boldsymbol{m}^n) + \mathbf{b}_{II,i}(\boldsymbol{m}^n, \boldsymbol{u}^n) = \boldsymbol{m}_i^n \times \Big(\boldsymbol{\mathcal{H}}_{\text{eff},i}(\boldsymbol{m}^n, \boldsymbol{u}_i^n) - \alpha \boldsymbol{m}_i^n \times \boldsymbol{\mathcal{H}}_{\text{eff},i}(\boldsymbol{m}^n, \boldsymbol{u}_i^n) \Big) \\ + \frac{\nu^2}{2} \sum_{l=1}^{3} (\boldsymbol{m}_i^n \times \mathbf{e}_l) \times \mathbf{e}_l. \end{aligned}$$

The matrix $\boldsymbol{\sigma}(\cdot)$ in (12.3) consists of the 3×3 blocks on the diagonal containing $\boldsymbol{\sigma}_{il}(\boldsymbol{m}^n) = \nu(\boldsymbol{m}_i^n \times \mathbf{e}_l)^T$ for each $i = 1, \ldots, N$ and $l = 1, 2, 3$. We obtain $\mathbb{P}^n$-a.s.

$$\|\boldsymbol{m}^n(t)\|^2_{(\mathbb{R}^3)^N} = \|\boldsymbol{m}_0\|^2_{(\mathbb{R}^3)^N} \quad \text{for all } t \in [0,T] \tag{12.4}$$

by applying Itô's formula using the functional $\mathbf{x} \mapsto \|\mathbf{x}\|^2_{(\mathbb{R}^3)^N}$. By using this property and Assumption C_1, we conclude that $\mathbf{b}_I$, $\mathbf{b}_{II}$, resp. $\boldsymbol{\sigma}$ are Lipschitz continuous functions on $(\mathbb{S}^2)^N \times \mathbb{U}$ resp. $(\mathbb{S}^2)^N$. Thus there exists an $\mathbb{F}^n$-adapted unique continuous global strong solution $\boldsymbol{m}^n \in L^2_{\mathbb{F}^n}(\Omega^n; C([0,T]; (\mathbb{R}^3)^N))$ of (11.1) under $\boldsymbol{u}^n \in L^2_{\mathbb{F}^n}(\Omega^n; L^2(0,T; \mathbb{U}))$ on the filtered probability space $(\Omega^n, \mathcal{F}^n, \mathbb{F}^n, \mathbb{P}^n)$. We define next

$$\mathbf{B}_I^n(t) := \int_0^t \mathbf{b}_I(\boldsymbol{m}^n(s)) \, \mathrm{d}s, \qquad \mathbf{B}_{II}^n(t) := \int_0^t \mathbf{b}_{II}(\boldsymbol{m}^n(s), \boldsymbol{u}^n(s)) \, \mathrm{d}s,$$

$$\boldsymbol{\Sigma}^n(t) := \int_0^t \boldsymbol{\sigma}(\boldsymbol{m}^n(s))\,\mathrm{d}\boldsymbol{W}^n(s),$$
$$\mathbf{F}^n(t) := \int_0^t \Big(\delta\|\boldsymbol{m}^n(s) - \widetilde{\boldsymbol{m}}(s)\|^2_{(\mathbb{R}^3)^N} + \lambda\|\boldsymbol{u}^n(s)\|^2_{(\mathbb{R}^3)^N}\Big)\,\mathrm{d}s.$$

Because of property (12.4) and Assumption C_1, a simple argument then shows for

$$\mathbf{X}^n := (\boldsymbol{m}^n, \mathbf{B}^n_I, \mathbf{B}^n_{II}, \boldsymbol{\Sigma}^n, \boldsymbol{W}^n, \mathbf{F}^n)$$

the existence of a constant $C > 0$ such that for all $p \geq 2$

$$\sup_{n\geq 1}\Big(\mathbb{E}^n\big[\|\mathbf{X}^n(t) - \mathbf{X}^n(s)\|^{2p}_{\cdot}\big]\Big) \leq C|t-s|^p \quad \text{for all } s,t \in [0,T]. \tag{12.5}$$

By the stochastic version of the Arzela-Ascoli theorem, which is applicable to $\{\mathbf{X}^n;\, n \in \mathbb{N}\}$, and the compactness of $\mathcal{V}(T;\mathbb{U})$, we obtain that the sequence of laws for $(\mathbf{X}^n, \boldsymbol{\delta}_{\boldsymbol{u}^n})$ is tight on $\big([C([0,T];(\mathbb{R}^3)^N)]^5 \times C([0,T];\mathbb{R})\big)\times\mathcal{V}(T;\mathbb{U})$-valued random variables. By Prohorov's lemma and Skorokhod's theorem, there exist a probability space $(\overline{\Omega},\overline{\mathcal{F}},\overline{\mathbb{P}})$, and a subsequence $\{(\overline{\mathbf{X}}^{n_r}, \overline{\boldsymbol{\lambda}}^{n_r});\, r \in \mathbb{N}\}$ as well as $(\overline{\mathbf{X}}, \overline{\boldsymbol{\lambda}})$ on it, such that for $r \geq 1$

$$\mathbb{P}^{n_r}_{\mathbf{X}^{n_r}} = \overline{\mathbb{P}}_{\overline{\mathbf{X}}^{n_r}}, \qquad \text{and} \qquad \mathbb{P}^{n_r}_{\boldsymbol{\delta}_{\boldsymbol{u}^{n_r}}} = \overline{\mathbb{P}}_{\overline{\boldsymbol{\lambda}}^{n_r}}, \tag{12.6}$$

and, moreover, $\overline{\mathbb{P}}$-a.s.

$$\overline{\boldsymbol{m}}^{n_r} \to \overline{\boldsymbol{m}} \qquad \text{in } C([0,T];(\mathbb{R}^3)^N), \tag{12.7a}$$
$$\overline{\mathbf{B}}^{n_r}_I \to \overline{\mathbf{B}}_I, \quad \overline{\mathbf{B}}^{n_r}_{II} \to \overline{\mathbf{B}}_{II}, \quad \overline{\boldsymbol{\Sigma}}^{n_r} \to \overline{\boldsymbol{\Sigma}} \qquad \text{in } C([0,T];(\mathbb{R}^3)^N), \tag{12.7b}$$
$$\overline{\boldsymbol{W}}^{n_r} \to \overline{\boldsymbol{W}} \qquad \text{in } C([0,T];(\mathbb{R}^3)^N), \tag{12.7c}$$
$$\overline{\mathbf{F}}^{n_r} \to \overline{\mathbf{F}} \qquad \text{in } C([0,T];\mathbb{R}), \tag{12.7d}$$
$$\overline{\boldsymbol{\lambda}}^{n_r} \overset{\star}{\rightharpoonup} \overline{\boldsymbol{\lambda}} \qquad \text{in } \mathcal{V}(T;\mathbb{U}) \tag{12.7e}$$

holds for $r \to \infty$. By following the proof of [YZ99, Chapter 2, Theorem 5.3], filtrations $\overline{\mathbb{F}}^{n_r}$ and $\overline{\mathbb{F}}$ can be constructed such that $\overline{\boldsymbol{W}}^{n_r}$ is an $\overline{\mathbb{F}}^{n_r}$-adapted Wiener process on $(\overline{\Omega},\overline{\mathcal{F}},\overline{\mathbb{P}})$. By using (12.6) we obtain the following stochastic differential equation

$$\begin{aligned}\overline{\boldsymbol{m}}^{n_r}(t) &= \overline{\mathbf{B}}^{n_r}_I(t) + \overline{\mathbf{B}}^{n_r}_{II}(t) + \overline{\boldsymbol{\Sigma}}^{n_r}(t)\\ &= \boldsymbol{m}_0 + \int_0^t \mathbf{b}_I(\overline{\boldsymbol{m}}^{n_r}(s))\,\mathrm{d}s + \int_0^t\int_{\mathbb{U}} \mathbf{b}_{II}(\overline{\boldsymbol{m}}^{n_r}(s),\mathbf{u})\,\overline{\boldsymbol{\nu}}^{n_r}(s,\mathrm{d}\mathbf{u})\,\mathrm{d}s\\ &\quad + \int_0^t \boldsymbol{\sigma}(\overline{\boldsymbol{m}}^{n_r}(s))\,\mathrm{d}\overline{\boldsymbol{W}}^{n_r}(s)\end{aligned} \tag{12.8}$$

on $(\overline{\Omega},\overline{\mathcal{F}},\overline{\mathbb{P}})$. The results in (12.7a)–(12.7b) may now be used to pass to the limit in the drift term which involves $\mathbf{b}_I$, while (12.7e) allows to pass to the limit in the relaxed formulation of the drift that involves $\mathbf{b}_{II}$. We obtain as limiting equation

$$\overline{\boldsymbol{m}}(t) = \boldsymbol{m}_0 + \int_0^t \mathbf{b}_I(\overline{\boldsymbol{m}}(s))\,\mathrm{d}s + \int_0^t\int_{\mathbb{U}} \mathbf{b}_{II}(\overline{\boldsymbol{m}}(s),\mathbf{u})\,\overline{\boldsymbol{\nu}}(s,\mathrm{d}\mathbf{u})\,\mathrm{d}s + \overline{\boldsymbol{\Sigma}}(t). \tag{12.9}$$

Moreover, we obtain by using (12.6) the following cost functional

$$\overline{\mathbb{E}}[\overline{\mathbf{F}}^{n_r}(T)] = \frac{1}{2}\overline{\mathbb{E}}\bigg[\int_0^T \Big(\delta\|\overline{\boldsymbol{m}}^{n_r}(s) - \widetilde{\boldsymbol{m}}(s)\|^2_{(\mathbb{R}^3)^N} + \lambda\int_{\mathbb{U}}\|\mathbf{u}\|^2_{(\mathbb{R}^3)^N}\,\overline{\boldsymbol{\nu}}^{n_r}(s,\mathrm{d}\mathbf{u})\Big)\,\mathrm{d}s$$

$$+ \kappa h\big(\overline{\boldsymbol{m}}^{n_r}(T)\big)\Big]$$
$$= \mathfrak{J}_{sto}(\boldsymbol{\pi}^{n_r}).$$

By using that $\{\boldsymbol{\pi}^{n_r};\, r \in \mathbb{N}\}$ is a minimizing sequence, and (12.7a), (12.7e), employing the compactness of $\mathbb{U}$, passing to the limit in all terms yields

$$\begin{aligned}\overline{\mathbb{E}}[\overline{\mathbf{F}}(T)] &= \frac{1}{2}\overline{\mathbb{E}}\Big[\int_0^T \Big(\delta\|\overline{\boldsymbol{m}}(s) - \widetilde{\boldsymbol{m}}(s)\|^2_{(\mathbb{R}^3)^N} + \lambda \int_{\mathbb{U}} \|\mathbf{u}\|^2_{(\mathbb{R}^3)^N} \overline{\boldsymbol{\nu}}(s,\, \mathrm{d}\mathbf{u})\Big)\, \mathrm{d}s + \kappa h\big(\overline{\boldsymbol{m}}(T)\big)\Big] \\ &= \mathfrak{M}. \end{aligned} \tag{12.10}$$

Following further the proof of [YZ99, Chapter 2, Theorem 5.3], it can be shown that $\overline{\boldsymbol{\Sigma}}(t)$ and $(\overline{\boldsymbol{\Sigma}}\,\overline{\boldsymbol{\Sigma}}^T(t) - \int_0^t \boldsymbol{\sigma}\boldsymbol{\sigma}^T(\overline{\boldsymbol{m}}(s))\,\mathrm{d}s)$ are both $\overline{\mathbb{F}}$-martingales. Using a martingale representation theorem we can then extend the filtered probability space $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{F}}, \overline{\mathbb{P}})$ to $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{F}}, \widehat{\mathbb{P}})$ and obtain an $\widehat{\mathbb{F}}$-adapted Wiener process $\widehat{\boldsymbol{W}}$ on it, such that

$$\overline{\boldsymbol{\Sigma}}(t) = \int_0^t \boldsymbol{\sigma}(\overline{\boldsymbol{m}}(s))\, \mathrm{d}\widehat{\boldsymbol{W}}(s). \tag{12.11}$$

Using the extension on $\overline{\boldsymbol{m}}$ and $\overline{\boldsymbol{\lambda}}$ in (12.9) and (12.10), and inserting equation (12.11), we have a filtered probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{F}}, \widehat{\mathbb{P}})$, an $\widehat{\mathbb{F}}$-adapted Wiener process $\widehat{\boldsymbol{W}}$, and an $\widehat{\mathbb{F}}$-adapted $\mathcal{V}(T;\mathbb{U})$-valued relaxed control $\overline{\boldsymbol{\lambda}}$, minimizing the cost functional (12.10), and, moreover, $\overline{\boldsymbol{m}}$ is a solution of (12.9), where (12.11).
Finally, it remains to show that there exists an $\widehat{\mathbb{F}}$-adapted $\mathbb{U}$-valued optimal control $\overline{\boldsymbol{u}}$. Therefore, we consider the control term in (12.10) and obtain using Jensen's inequality

$$\frac{\lambda}{2}\int_0^T \int_{\mathbb{U}} \|\mathbf{u}\|^2_{(\mathbb{R}^3)^N}\, \overline{\boldsymbol{\nu}}(s, \mathrm{d}\mathbf{u})\, \mathrm{d}s \geq \frac{\lambda}{2}\int_0^T \Big\| \int_{\mathbb{U}} \mathbf{u}\, \overline{\boldsymbol{\nu}}(s, \mathrm{d}\mathbf{u})\Big\|^2_{(\mathbb{R}^3)^N} \mathrm{d}s \geq \frac{\lambda}{2}\int_0^T \|\overline{\boldsymbol{u}}(s)\|^2_{(\mathbb{R}^3)^N}\, \mathrm{d}s,$$

where we define $\overline{\boldsymbol{u}}(s) := \int_{\mathbb{U}} \mathbf{u}\, \overline{\boldsymbol{\nu}}(s, \mathrm{d}\mathbf{u})$. Thus, $(\overline{\boldsymbol{m}}, \overline{\boldsymbol{u}})$ minimizes the cost functional, moreover $(\overline{\boldsymbol{m}}, \overline{\boldsymbol{u}})$ is admissible: due to the linearity in $\mathbf{u}$ of $\mathbf{b}_{II}(\overline{\boldsymbol{m}}, \mathbf{u}) = \mathbf{A}_{II}(\overline{\boldsymbol{m}}(s))\cdot\mathbf{u}$, where $\mathbf{A}_{II}(\overline{\boldsymbol{m}}(s))$ is matrix-valued, we obtain

$$\begin{aligned}\int_{\mathbb{U}} \mathbf{b}_{II}(\overline{\boldsymbol{m}}(s), \mathbf{u})\, \overline{\boldsymbol{\nu}}(s, \mathrm{d}\mathbf{u}) &= \int_{\mathbb{U}} \mathbf{A}_{II}(\overline{\boldsymbol{m}}(s)) \cdot \mathbf{u}\, \overline{\boldsymbol{\nu}}(s, \mathrm{d}\mathbf{u}) = \mathbf{A}_{II}(\overline{\boldsymbol{m}}(s)) \cdot \int_{\mathbb{U}} \mathbf{u}\, \overline{\boldsymbol{\nu}}(s, \mathrm{d}\mathbf{u}) \\ &= \mathbf{A}_{II}(\overline{\boldsymbol{m}}(s)) \cdot \overline{\boldsymbol{u}}(s) = \mathbf{b}_{II}\big(\overline{\boldsymbol{m}}(s), \overline{\boldsymbol{u}}(s)\big).\end{aligned}$$

As a consequence, $\overline{\boldsymbol{m}}$ is a strong solution of the state equation (11.1) on $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{F}}, \widehat{\mathbb{P}})$ under $\overline{\boldsymbol{u}}$. □

13. Characterization of the optimal control problems

We start with the generalized Hamiltonian system as necessary optimality conditions for the deterministic and stochastic optimal control problems. Below, we consider the following exchange field $\boldsymbol{\mathcal{H}}_{exch}$, stray field $\boldsymbol{\mathcal{H}}_{d}$, and anisotropy energy $\boldsymbol{\mathcal{H}}_{ani}$ in (11.1).

Assumption $\mathsf{C_2}$
Let

$$\boldsymbol{\mathcal{H}}_{exch,i}(\boldsymbol{m}) = C_{exch}(\boldsymbol{m}_{i-1} - 2\boldsymbol{m}_i + \boldsymbol{m}_{i+1}),$$

with $\boldsymbol{m}_0 := \boldsymbol{m}_N$ and $\boldsymbol{m}_{N+1} := \boldsymbol{m}_1$. Moreover, combine

$$\boldsymbol{\mathcal{H}}_{d,ani,i}(\boldsymbol{m}_i) := \boldsymbol{\mathcal{H}}_{ani,i}(\boldsymbol{m}_i) + \boldsymbol{\mathcal{H}}_{d,i}(\boldsymbol{m}_i) = C_{ani}\boldsymbol{A}\boldsymbol{m}_i - C_d\boldsymbol{B}_i\boldsymbol{m}_i = -C_{d,ani}\boldsymbol{D}_i\boldsymbol{m}_i,$$

with $\boldsymbol{D}_i \in \mathbb{R}^{3\times 3}_{diag}$ and $C_{d,ani} \in \mathbb{R}$.

13.1. Deterministic optimal control

We start with the necessary optimality conditions for the corresponding deterministic optimal control problem (Problem 11.1 with $\nu \equiv 0$). By Pontryagin's maximum principle there exists a triple $(\boldsymbol{m}^\star, \boldsymbol{p}^\star, \boldsymbol{u}^\star) \in \big[C([0,T];(\mathbb{R}^3)^N)\big]^2 \times L^2(0,T;\mathbb{U})$ with

$$\boldsymbol{p}^\star := (\boldsymbol{p}_1^\star, \dots, \boldsymbol{p}_N^\star) \in C\big([0,T];(\mathbb{R}^3)^N\big)$$

which satisfies the (deterministic version of the) state equation (11.1) together with

$$\begin{aligned} \mathrm{d}\boldsymbol{p}_i^\star &= -\big(\mathbf{f}_i(\boldsymbol{m}^\star, \boldsymbol{u}_i^\star, \boldsymbol{p}^\star) - \delta(\boldsymbol{m}_i^\star - \widetilde{\boldsymbol{m}}_i)\big)\,\mathrm{d}t, \\ \boldsymbol{p}^\star(T) &= -\kappa\frac{\partial h}{\partial \boldsymbol{x}}(\boldsymbol{m}^\star(T)), \end{aligned} \tag{13.1}$$

where $(1 \le i \le N)$

$$\begin{aligned} \mathbf{f}_i(\boldsymbol{m}, \boldsymbol{u}_i, \boldsymbol{p}) := &-\boldsymbol{p}_i \times \Big(\boldsymbol{H}_{\mathit{eff},i}(\boldsymbol{m}, \boldsymbol{u}_i) - \alpha\boldsymbol{m}_i \times \boldsymbol{H}_{\mathit{eff},i}(\boldsymbol{m}, \boldsymbol{u}_i)\Big) \\ &+ \alpha(\boldsymbol{p}_i \times \boldsymbol{m}_i) \times \boldsymbol{H}_{\mathit{eff},i}(\boldsymbol{m}, \boldsymbol{u}_i) \\ &- C_{d,ani}\boldsymbol{D}_i(\boldsymbol{p}_i - \alpha(\boldsymbol{p}_i \times \boldsymbol{m}_i)) \times \boldsymbol{m}_i \\ &- C_{exch}\big[\boldsymbol{m}_{i-1} \times (\boldsymbol{p}_{i-1} - \alpha\boldsymbol{p}_{i-1} \times \boldsymbol{m}_{i-1}) - 2\boldsymbol{m}_i \times (\boldsymbol{p}_i - \alpha\boldsymbol{p}_i \times \boldsymbol{m}_i) \\ &+ \boldsymbol{m}_{i+1} \times (\boldsymbol{p}_{i+1} - \alpha\boldsymbol{p}_{i+1} \times \boldsymbol{m}_{i+1})\big], \end{aligned} \tag{13.2}$$

and the maximum condition

$$\mathcal{H}_{det}(\boldsymbol{m}^\star(t),\boldsymbol{u}^\star(t),\boldsymbol{p}^\star(t)) = \max_{\boldsymbol{u}\in\mathbb{U}} \mathcal{H}_{det}(\boldsymbol{m}^\star(t),\boldsymbol{u},\boldsymbol{p}^\star(t)) \quad \text{for a.e. } t\in[0,T] \tag{13.3}$$

holds, where

$$\begin{aligned}\mathcal{H}_{det}(\boldsymbol{m},\boldsymbol{u},\boldsymbol{p}) := \sum_{i=1}^N \Big(&\Big\langle \boldsymbol{p}_i, \boldsymbol{m}_i \times (\boldsymbol{\mathcal{H}}_{\text{eff},i}(\boldsymbol{m},\boldsymbol{u}_i) - \alpha \boldsymbol{m}_i \times \boldsymbol{\mathcal{H}}_{\text{eff},i}(\boldsymbol{m},\boldsymbol{u}_i))\Big\rangle_{\mathbb{R}^3} \\ &- \frac{1}{2}\Big(\delta\|\boldsymbol{m}_i - \widetilde{\boldsymbol{m}}_i\|^2_{\mathbb{R}^3} - \lambda\|\boldsymbol{u}_i\|^2_{\mathbb{R}^3}\Big)\Big).\end{aligned} \tag{13.4}$$

13.2. Stochastic optimal control

Existence of a weak stochastic control $\boldsymbol{\pi} := (\Omega,\mathcal{F},\mathbb{F},\mathbb{P},\boldsymbol{W},\boldsymbol{u}^\star)$ is proven in Chapter 12 by abstract arguments. For the simulations of the stochastic optimal control problem below, however, we assume a complete filtered probability space $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$, where that the filtration $\mathbb{F}$ is generated by the $(\mathbb{R}^3)^N$-valued Wiener process $\boldsymbol{W} = \{\boldsymbol{W}(t);\, t\in[0,T]\}$. By applying the stochastic maximum principle, see e.g. [YZ99, Chapter 3, Theorem 3.2], we then obtain the first order optimality conditions of Problem 11.1: Find $(\boldsymbol{m}^\star,\boldsymbol{u}^\star) \in L^2_{\mathbb{F}}(\Omega;C([0,T];(\mathbb{R}^3)^N)) \times L^2_{\mathbb{F}}(\Omega;L^2(0,T;\mathbb{U}))$ and moreover

$$(\boldsymbol{p}^\star,\boldsymbol{q}^\star) := (\boldsymbol{p}_1^\star,\dots,\boldsymbol{p}_N^\star;\boldsymbol{q}_1^\star,\dots,\boldsymbol{q}_N^\star) \in L^2_{\mathbb{F}}\Big(\Omega;C\Big([0,T];(\mathbb{R}^3)^N\Big)\times L^2\Big(0,T;(\mathbb{R}^{3\times 3N})^N\Big)\Big),$$

such that (11.1) and the adjoint equation ($1\le i\le N$)

$$\begin{aligned}\mathrm{d}\boldsymbol{p}_i^\star &= -\Big(\mathbf{f}_i(\boldsymbol{m}^\star,\boldsymbol{u}_i^\star,\boldsymbol{p}^\star) + \frac{\nu^2}{2}\sum_{l=1}^3 \mathbf{e}_l\times(\mathbf{e}_l\times\boldsymbol{p}_i^\star) - \nu\sum_{l=1}^3 \boldsymbol{q}^\star_{i,3(i-1)+l}\times\mathbf{e}_l - \delta(\boldsymbol{m}_i^\star - \widetilde{\boldsymbol{m}}_i)\Big)\,\mathrm{d}t \\ &\quad + \boldsymbol{q}_i^\star\,\mathrm{d}\boldsymbol{W}, \\ \boldsymbol{p}^\star(T) &= -\kappa\frac{\partial h}{\partial \boldsymbol{x}}(\boldsymbol{m}^\star(T))\end{aligned} \tag{13.5}$$

are satisfied, where $\mathbf{f}$ was given in (13.2). Moreover, the maximum condition holds $\mathbb{P}$-a.s.

$$\begin{aligned}&\mathcal{H}_{sto}(\boldsymbol{m}^\star(t),\boldsymbol{u}^\star(t),\boldsymbol{p}^\star(t),\boldsymbol{q}^\star(t)) \\ &= \max_{\boldsymbol{u}\in\mathbb{U}} \mathcal{H}_{sto}(\boldsymbol{m}^\star(t),\boldsymbol{u},\boldsymbol{p}^\star(t),\boldsymbol{q}^\star(t)) \quad \text{for a.e. } t\in[0,T],\end{aligned} \tag{13.6}$$

where

$$\begin{aligned}\mathcal{H}_{sto}(t,\boldsymbol{m},\boldsymbol{u},\boldsymbol{p},\boldsymbol{q}) := \sum_{i=1}^N \Big(&\Big\langle \boldsymbol{p}_i, \boldsymbol{m}_i \times (\boldsymbol{\mathcal{H}}_{\text{eff},i}(\boldsymbol{m},\boldsymbol{u}_i) - \alpha \boldsymbol{m}_i \times \boldsymbol{\mathcal{H}}_{\text{eff},i}(\boldsymbol{m},\boldsymbol{u}_i))\Big\rangle_{\mathbb{R}^3} \\ &- \frac{\nu^2}{2}\sum_{l=1}^3 \|\mathbf{e}_l\times\boldsymbol{p}_i\|^2_{\mathbb{R}^3} + tr[\boldsymbol{q}_i^T\boldsymbol{\sigma}_{ii}(\boldsymbol{m})] \\ &- \frac{1}{2}\Big(\delta\|\boldsymbol{m}_i - \widetilde{\boldsymbol{m}}_i\|^2_{\mathbb{R}^3} - \lambda\|\boldsymbol{u}_i\|^2_{\mathbb{R}^3}\Big)\Big).\end{aligned} \tag{13.7}$$

14. Simulation of the optimal control problems

We discretize and simulate the generalized Hamiltonian systems which are related to the deterministic optimal and the stochastic optimal control problem. Throughout this work, we consider a uniform time-grid $\{t_j\}_{j=0}^J$ of mesh size $k := t_j - t_{j-1} > 0$ which covers $[0,T]$. For the simulations, we allow general controls to avoid the control constraint above, which would otherwise require a projected gradient method to approximate (13.6). Under these assumptions, the maximum condition in (13.6) then reduces to $\mathbb{P}$-a.s. solving

$$-\alpha(\boldsymbol{p}_i(t) \times \boldsymbol{m}_i(t)) \times \boldsymbol{m}_i(t) + \boldsymbol{p}_i(t) \times \boldsymbol{m}_i(t) - \lambda \boldsymbol{u}_i(t) = 0 \quad \text{for a.e. } t \in [0,T]. \tag{14.1}$$

Note that this condition causes optimal admissible pairings $\{(\boldsymbol{m}_i, \boldsymbol{u}_i);\, 1 \le i \le N\}$ to be $\mathbb{P}$-a.s. orthogonal to each other.

14.1. Stochastic optimal control

In order to account for the forward-backward character of the generalized Hamiltonian system (11.1), (13.5), (14.1), the following Picard type algorithm is motivated from [BZ08]:

Scheme 14.1 (Picard-scheme for the stochastic control problem)

(1) Set $\boldsymbol{u}^{(1),j} \equiv \boldsymbol{u}_{init}^j \in \mathbb{U}$ for each $j = 0, \dots, J$ in the first Picard iteration step.

(2) Iterate $v = 1, 2, \dots$ until a stopping criterion is met:

(i) **FSDE:** Set $\boldsymbol{m}^{(v),0} = \boldsymbol{m}_0$. For each $j = 1, \dots, J$, compute $\boldsymbol{u}^{(v),j-1}$ using $\boldsymbol{p}^{(v-1),j-1}$ and $\boldsymbol{m}^{(v),j-1}$ according to (14.1). Simulate the $(\mathbb{R}^3)^N$-valued random variable $\boldsymbol{m}^{(v),j}$ by a time discretization of (11.1) using $\boldsymbol{u}^{(v),j-1}$ and $\boldsymbol{m}^{(v),j-1}$.

(ii) **BSDE:** Set $\boldsymbol{p}^{(v),J} = -\kappa \frac{\partial h}{\partial \boldsymbol{x}}(\boldsymbol{m}^{(v),J})$. For each $j = J-1, \dots, 0$, compute first $\boldsymbol{u}^{(v),j+1}$ from (14.1) using $\boldsymbol{p}^{(v),j+1}$ and $\boldsymbol{m}^{(v),j+1}$. Simulate the $\mathbb{R}^{3N \times 3N}$ and $(\mathbb{R}^3)^N$-valued random variables $\boldsymbol{q}^{(v),j}$ and $\boldsymbol{p}^{(v),j}$ by a time discretization of (13.5) using $\boldsymbol{u}^{(v),j+1}$, $\boldsymbol{p}^{(v),j+1}$, $\boldsymbol{q}^{(v),j+1}$, $\boldsymbol{m}^{(v),j+1}$ and $\boldsymbol{m}^{(v),j}$.

(iii) Evaluate if a stopping criterion is met.

Note that the control $\boldsymbol{u}^{(v)} = (\boldsymbol{u}^{(v),0}, \dots, \boldsymbol{u}^{(v),J})$ in Scheme 14.1 is treated as known quantity in the state equation (FSDE) as well as in the adjoint equation (BSDE), and is computed via (14.1). Moreover, the solution of the adjoint equation $\boldsymbol{p}^{(v-1)} = (\boldsymbol{p}^{(v-1),0}, \dots, \boldsymbol{p}^{(v-1),J})$ of the previous Picard iteration step $v-1$ is used in the state equation. Precise formulas how

to compute a single time-step of the state and adjoint equation are given in Schemes 14.2 and 14.3 below.

We use the following scheme to approximate the state equation (11.1) in the v-th step of the Picard iteration which corresponds to the semi-implicit Scheme B from [MTF$^+$10].

Scheme 14.2 (Time discretization of the FSDE)
Let $\boldsymbol{u}^{(v),j-1}$ and $\boldsymbol{m}^{(v),j-1}$ be given. Compute $\boldsymbol{m}^{(v),j} \in L^2_{\mathcal{F}_{t_j}}(\Omega;(\mathbb{S}^2)^N)$ by the following two-step method: Compute first $\boldsymbol{w}^{(v),j} \in L^2_{\mathcal{F}_{t_j}}(\Omega;(\mathbb{S}^2)^N)$ using for each $i = 1,\dots,N$

$$\begin{aligned}\boldsymbol{w}_i^{(v),j} &= \boldsymbol{m}_i^{(v),j-1} \\ &\quad + k\frac{\boldsymbol{m}_i^{(v),j-1}+\boldsymbol{w}_i^{(v),j}}{2} \\ &\qquad \times \Big(\boldsymbol{\mathcal{H}}_{\text{eff},i}(\boldsymbol{m}^{(v),j-1},\boldsymbol{u}_i^{(v),j-1}) - \alpha \boldsymbol{m}_i^{(v),j-1}\times \boldsymbol{\mathcal{H}}_{\text{eff},i}(\boldsymbol{m}^{(v),j-1},\boldsymbol{u}_i^{(v),j-1})\Big) \\ &\quad + \nu\frac{\boldsymbol{m}_i^{(v),j-1}+\boldsymbol{w}_i^{(v),j}}{2}\times \Delta_{j-1}\boldsymbol{W}_i,\end{aligned}$$

where $\Delta_{j-1}\boldsymbol{W}_i := \boldsymbol{W}_i(t_j) - \boldsymbol{W}_i(t_{j-1}) \sim \mathcal{N}(\boldsymbol{0}, k\mathbb{1}_{3\times 3})$.
Set $\boldsymbol{z}^{(v),j} := \frac{1}{2}(\boldsymbol{m}^{(v),j-1}+\boldsymbol{w}^{(v),j})$, and compute then for each $i = 1,\dots,N$

$$\begin{aligned}\boldsymbol{m}_i^{(v),j} &= \boldsymbol{m}_i^{(v),j-1} \\ &\quad + k\frac{\boldsymbol{m}_i^{(v),j-1}+\boldsymbol{m}_i^{(v),j}}{2} \\ &\qquad \times \Big(\boldsymbol{\mathcal{H}}_{\text{eff},i}(\boldsymbol{z}^{(v),j},\boldsymbol{u}_i^{(v),j-1}) - \alpha \boldsymbol{z}_i^{(v),j}\times \boldsymbol{\mathcal{H}}_{\text{eff},i}(\boldsymbol{z}^{(v),j},\boldsymbol{u}_i^{(v),j-1})\Big) \\ &\quad + \nu\frac{\boldsymbol{m}_i^{(v),j-1}+\boldsymbol{m}_i^{(v),j}}{2}\times \Delta_{j-1}\boldsymbol{W}_i.\end{aligned}$$

The use of the semi-implicit Scheme B from [MTF$^+$10] has the advantage that the sphere constraint is preserved, i.e., $\boldsymbol{m}^{(v),j}$ takes values in $(\mathbb{S}^2)^N$. Moreover, its computation can be performed without the application of any root-finding method, which makes this scheme more preferable for our purposes of simulating finitely many particles, if e.g. compared with the midpoint scheme; see [NP13] and Part I.

We apply an explicit time discretization together with the one-step dynamic programming scheme for the simulation of the backward SDE (13.5); see e.g. [Zha04].

Scheme 14.3 (Time discretization of the BSDE)
Let $\boldsymbol{u}^{(v),j+1}$, $\boldsymbol{m}^{(v),j+1}$, $\boldsymbol{p}^{(v),j+1}$, and $\boldsymbol{q}^{(v),j+1}$ be given. Compute $\boldsymbol{q}_l^{(v),j} \in L^2_{\mathcal{F}_{t_j}}(\Omega;(\mathbb{R}^3)^N)$ and $\boldsymbol{p}^{(v),j} \in L^2_{\mathcal{F}_{t_j}}(\Omega;(\mathbb{R}^3)^N)$ by the following method: For each $i = 1,\dots,N$, compute

$$\boldsymbol{q}_{i,l}^{(v),j} = \frac{1}{k}\mathbb{E}\Big[\boldsymbol{p}_i^{(v),j+1}\Delta_j W_l\Big|\mathcal{F}_{t_j}\Big] \qquad \forall l = 1,\dots,3N, \tag{14.2}$$

and

$$\begin{aligned}\boldsymbol{p}_i^{(v),j} = \mathbb{E}\Big[\boldsymbol{p}_i^{(v),j+1} + k\mathbf{f}_i(\boldsymbol{m}^{(v),j+1}, \boldsymbol{u}^{(v),j+1}, \boldsymbol{p}^{(v),j+1}) - \frac{\nu^2}{2}k\sum_{l=1}^{3}\mathbf{e}_l \times (\mathbf{e}_l \times \boldsymbol{p}_i^{(v),j+1}) \\ + \delta k(\boldsymbol{m}_i^{(v),j+1} - \widetilde{\boldsymbol{m}}_i(t_{j+1}))\Big|\mathcal{F}_{t_j}\Big] + \nu k\sum_{l=1}^{3}\boldsymbol{q}_{i,3(i-1)+l}^{(v),j} \times \mathbf{e}_l.\end{aligned} \tag{14.3}$$

Again, an explicit time-stepping scheme is chosen to reduce the computational effort. A nonlinear equation in (14.3) is omitted by the explicit treatment of the control, where $\boldsymbol{u}^{(v),j+1}$ can be computed according to (14.1) using $\boldsymbol{p}^{(v),j+1}$ and $\boldsymbol{m}^{(v),j+1}$.

By using the Markov chain property of the time discretization $\{\boldsymbol{m}^{(v),j};\, j = 0,\dots,J\}$ at Picard iteration step v we can represent the solution of (14.2)–(14.3) by two measurable, deterministic, but unknown functions $\mathcal{P}^{(v),j} : (\mathbb{S}^2)^N \to (\mathbb{R}^3)^N$ and $\mathcal{Q}_l^{(v),j} : (\mathbb{S}^2)^N \to (\mathbb{R}^3)^N$ for all $l = 1,\dots,3N$, such that

$$\boldsymbol{q}_l^{(v),j} = \mathcal{Q}_l^{(v),j}(\boldsymbol{m}^{(v),j}), \qquad \boldsymbol{p}^{(v),j} = \mathcal{P}^{(v),j}(\boldsymbol{m}^{(v),j}), \tag{14.4}$$

where for each $i = 1,\dots,N$

$$\mathcal{Q}_{i,l}^{(v),j}(\boldsymbol{x}) = \frac{1}{k}\mathbb{E}\Big[\mathcal{P}_i^{(v),j+1}(\boldsymbol{m}_{|j,\boldsymbol{x}}^{(v),j+1})\Delta_j W_l\Big] \qquad \forall l = 1,\dots,3N, \tag{14.5}$$

$$\begin{aligned}\mathcal{P}_i^{(v),j}(\boldsymbol{x}) = \mathbb{E}\Big[\mathcal{P}_i^{(v),j+1}(\boldsymbol{m}_{|j,\boldsymbol{x}}^{(v),j+1}) + k\mathbf{f}_i(\boldsymbol{m}_{|j,\boldsymbol{x}}^{(v),j+1}, \boldsymbol{u}^{(v),j+1}, \mathcal{P}^{(v),j+1}(\boldsymbol{m}_{|j,\boldsymbol{x}}^{(v),j+1})) \\ - \frac{\nu^2}{2}k\sum_{l=1}^{3}\mathbf{e}_l \times (\mathbf{e}_l \times \mathcal{P}_i^{(v),j+1}(\boldsymbol{m}_{|j,\boldsymbol{x}}^{(v),j+1})) + \delta k(\boldsymbol{m}_{i|j,\boldsymbol{x}}^{(v),j+1} - \widetilde{\boldsymbol{m}}_i(t_{j+1}))\Big] \\ + \nu k\sum_{l=1}^{3}\mathcal{Q}_{i,3(i-1)+l}^{(v),j}(\boldsymbol{x}) \times \mathbf{e}_l.\end{aligned} \tag{14.6}$$

Here, $\boldsymbol{m}_{|j,\boldsymbol{x}}^{(v),j+1}$ denotes the value of the Markov chain at time t_{j+1}, which has started at t_j at state $\boldsymbol{x} \in (\mathbb{S}^2)^N$. The control $\boldsymbol{u}^{(v),j+1}$ in (14.6) is computed using $\boldsymbol{m}_{|j,\boldsymbol{x}}^{(v),j+1}$ and $\mathcal{P}^{(v),j+1}(\boldsymbol{m}_{|j,\boldsymbol{x}}^{(v),j+1})$ according to equation (14.1).

We approximate the measurable, deterministic functions $(\mathcal{P}^{(v),j}(\cdot), \mathcal{Q}_l^{(v),j}(\cdot))$ using the partitioning estimation method, which is a special case of the least squares Monte-Carlo method, see e.g. [BZ08, GLW05]. The idea of the least squares Monte-Carlo method is based on the representation $\mathbb{E}[\boldsymbol{\Theta}|\boldsymbol{m}^{(v),j} = \boldsymbol{x}] = \boldsymbol{v}^{(v),j}(\boldsymbol{x})$, which minimizes $\mathbb{E}\big[\big|\boldsymbol{v}^{(v),j}(\boldsymbol{x}) - \boldsymbol{\Theta}\big|^2\big]$ among all $\mathcal{F}_{t_j}$-measurable functions $\boldsymbol{v}^{(v),j} : (\mathbb{S}^2)^N \to \mathbb{R}$, such that $\mathbb{E}\big[|\boldsymbol{v}^{(v),j}(\boldsymbol{x})|^2\big] < \infty$. The function $\boldsymbol{v}^{(v),j}(\cdot)$ is approximated in two steps: Firstly, we approximate it by a function

$$\boldsymbol{v}_R^{(v),j}(\cdot) = \sum_{r=1}^{R}\overline{\mathbf{a}}_r^{(v),j}\eta_r^{(v),j}(\cdot)$$

in the finite dimensional subspace $\operatorname{span}\{\eta_r^{(v),j}(\cdot);\, r = 1,\dots,R\}$ of $L^2(\Omega;\mathbb{P})$. Secondly, we approximate the coefficients $\{\overline{\mathbf{a}}_r^{(v),j};\, r = 1,\dots,R\}$ of $\boldsymbol{v}_R^{(v),j}(\cdot)$ by $\{\mathbf{a}_r^{(v),j};\, r = 1,\dots,R\}$,

which can be computed from the following least squares problem

$$\operatorname*{argmin}_{\mathbf{a}^{(v),j}\in\mathbb{R}^R} \frac{1}{M}\sum_{m=1}^{M}\Big|\sum_{r=1}^{R}\mathbf{a}_r^{(v),j}\eta_r^{(v),j}(\boldsymbol{x}_m)-\boldsymbol{\Theta}_m\Big|^2, \tag{14.7}$$

where $M \gg R$ many independent samples $(\boldsymbol{x}_m, \boldsymbol{\Theta}_m)$ of $(\boldsymbol{x}, \boldsymbol{\Theta})$ are used. In the case of partitioning estimation we construct the finite dimensional subspace $\operatorname{span}\{\eta_r^{(v),j}(\cdot);\, r = 1,\dots,R\}$ by indicator functions $\eta_r^{(v),j}(\cdot) = \mathbb{1}_{C_r^{(v),j}}(\cdot)$ generated from a disjoint decomposition of the state space $\bigcup_{r=1}^R C_r^{(v),j} = (\mathbb{R}^3)^N$, i.e., since iterates of Scheme 14.2 are $\mathbb{S}^2$-valued, we have to partition $\bigcup_{r=1}^R C_r^{(v),j} = (\mathbb{S}^2)^N$. By this choice of basis functions, we can compute the coefficients in problem (14.7) according to

$$\mathbf{a}_r^{(v),j} = \frac{1}{\#\{\boldsymbol{x}_m \in C_r^{(v),j}\}}\sum_{m=1}^{M}\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x}_m)\boldsymbol{\Theta}_m \approx \mathbb{E}\Big[\boldsymbol{\Theta}\Big|\boldsymbol{x}\in C_r^{(v),j}\Big], \tag{14.8}$$

where the convention $\frac{0}{0} = 0$ is used, and the partition estimation function $\boldsymbol{v}_R^{(v),j}(\cdot)$ can be expressed as

$$\boldsymbol{v}_R^{(v),j}(\boldsymbol{x}) \approx \sum_{r=1}^{R}\Big(\frac{1}{\#\{\boldsymbol{x}_m \in C_r^{(v),j}\}}\sum_{m=1}^{M}\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x}_m)\boldsymbol{\Theta}_m\Big)\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x}).$$

In other word, for every $\boldsymbol{x} \in (\mathbb{S}^2)^N$, the partition estimation function $\boldsymbol{v}_R^{(v),j}(\boldsymbol{x})$ returns the local average of those $\boldsymbol{\Theta}_m$ whose $\boldsymbol{x}_m$ has been contained in the same region $C_\cdot^{(v),j}$ as $\boldsymbol{x}$.

We use Voronoi meshes for the partition $\bigcup_{r=1}^R C_r^{(v),j} = (\mathbb{S}^2)^N$, where the idea of the construction uses a sample based approximation of the distribution of $\boldsymbol{m}^{(v),j}$:

(1) Simulate R additional realizations $\{\widehat{\boldsymbol{m}}_r^{(v),j};\, r = 1,\dots,R\}$ of the $(\mathbb{S}^2)^N$-valued random variable $\boldsymbol{m}^{(v),j}$.

(2) Define the region $C_r^{(v),j}$ by

$$C_r^{(v),j} := \Big\{\boldsymbol{x} \in (\mathbb{S}^2)^N;\, \|\boldsymbol{x}-\widehat{\boldsymbol{m}}_r^{(v),j}\|_{(\mathbb{R}^3)^N} < \inf_{r\neq s}\|\boldsymbol{x}-\widehat{\boldsymbol{m}}_s^{(v),j}\|_{(\mathbb{R}^3)^N}\Big\}.$$

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

This strategy resolves the state space $(\mathbb{S}^2)^N$ according to the distribution of $\boldsymbol{m}^{(v),j}$, and creates more regions in areas where $\boldsymbol{m}^{(v),j}$ is more likely to take values, and may be quickly realized in actual simulations. This property is important since a new partition according to $\boldsymbol{m}^{(v),j}$ has to be constructed for each Picard-iteration step v.

The procedure to approximate the conditional expectations $(\mathcal{P}^{(v),j}(\cdot), \mathcal{Q}_l^{(v),j}(\cdot))$ in (14.5)–(14.6) by $(\mathcal{P}_R^{(v),j}(\cdot), \mathcal{Q}_{R,l}^{(v),j}(\cdot))$ which are applied to Schemes 14.1, 14.3 is summarized in Algorithm 1.

Algorithm 1 Picard type (ODP) scheme for the stochastic control problem

Simulate and store the increments $\Delta\widehat{\boldsymbol{W}} := \{(\Delta_j \widehat{\boldsymbol{W}}_r)_{j=0}^{J-1};\, r = 1, \dots, R\}$ for the construction of the basis regions.
Set $v = 0$. For the first iteration step, fix the initial control $\boldsymbol{u}^{(1),j} \equiv \boldsymbol{u}_{init}^{j}$ for each $j = 0, \dots, J$.
while Stopping criterion fails **do**
    Set $v = v + 1$.
    Construct the basis regions $\{C_r^{(v),j};\, r = 1, \dots, R\}$ using $\Delta\widehat{\boldsymbol{W}}$.
    Set $\mathcal{P}^{(v),J}(\boldsymbol{x}) = -\kappa \frac{\partial h}{\partial \boldsymbol{x}}(\boldsymbol{x})$.
    for $j = J - 1$ **to** 0 **do**
        Simulate M independent paths of the state equation and store $\boldsymbol{m}_m^{(v),j}$ and $\boldsymbol{m}_m^{(v),j+1}$.
        Compute the vector $\boldsymbol{\alpha}^{(v),j}$ with entries $\boldsymbol{\alpha}_r^{(v),j} := \#\{\boldsymbol{m}_m^{(v),j} \in C_r^{(v),j}\}$.
        Compute the vector $\boldsymbol{\beta}^{(v),j}$ with entries $\boldsymbol{\beta}_m^{(v),j} := \{r;\, C_r^{(v),j} \ni \boldsymbol{m}_m^{(v),j}\}$.
        for $l = 1$ **to** $3N$ **do**
            Compute the vector $\boldsymbol{H}$ with entries $\boldsymbol{H}_m := \mathcal{P}_R^{(v),j+1}(\boldsymbol{m}_m^{(v),j+1}) \frac{\Delta_j \boldsymbol{W}_{l,m}}{k}$.
            Compute for each $r = 1, \dots, R$ the coefficients

$$\mathbf{b}_{l,r}^{(v),j} := \frac{1}{\boldsymbol{\alpha}_r^{(v),j}} \sum_{m=1}^{M} \mathbb{1}_{\{r\}}(\boldsymbol{\beta}_m^{(v),j}) \boldsymbol{H}_m.$$

            Define $\mathcal{Q}_{R,l}^{(v),j}(\boldsymbol{x}) := \sum_{r=1}^{R} \mathbf{b}_{l,r}^{(v),j} \mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x})$.
        end for
        Compute the control $\boldsymbol{u}_m^{(v),j+1}$ using $\mathcal{P}_R^{(v),j+1}(\cdot)$ and $\boldsymbol{m}_m^{(v),j+1}$ by equation (14.1).
        Compute the vector $\boldsymbol{H}$ with entries

$$\begin{aligned}\boldsymbol{H}_m := \mathcal{P}_R^{(v),j+1}(\boldsymbol{m}_m^{(v),j+1}) + k\mathbf{f}\big(\boldsymbol{m}_m^{(v),j+1}, \boldsymbol{u}_m^{(v),j+1}, \mathcal{P}_R^{(v),j+1}(\boldsymbol{m}_m^{(v),j+1})\big) \\ - \frac{\nu^2}{2} k \sum_{l=1}^{3} \mathbf{e}_l \times \big(\mathbf{e}_l \times \mathcal{P}_R^{(v),j+1}(\boldsymbol{m}_m^{(v),j+1})\big) + \delta k\big(\boldsymbol{m}_m^{(v),j+1} - \widetilde{\boldsymbol{m}}(t_{j+1})\big).\end{aligned}$$

        Compute for each $r = 1, \dots, R$ the coefficients

$$\mathbf{a}_r^{(v),j} := \frac{1}{\boldsymbol{\alpha}_r^{(v),j}} \sum_{m=1}^{M} \mathbb{1}_{\{r\}}(\boldsymbol{\beta}_m^{(v),j}) \boldsymbol{H}_m.$$

        Define $\mathcal{P}_R^{(v),j}(\boldsymbol{x}) := \sum_{r=1}^{R} \big[\mathbf{a}_r^{(v),j} + \nu k \sum_{l=1}^{3} \mathbf{b}_{l,r}^{(v),j} \times \mathbf{e}_l\big] \mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x})$.
    end for
    Evaluate the cost function $\mathcal{J}_{sto}^{(v)}(\cdot)$ and its parts to decide whether to stop or not.
end while

As a result of Algorithm 1, we obtain the control function $\mathcal{U}^{(v),j}(\boldsymbol{m}^{(v),j})$ according to equation (14.1) for each $i = 1, \dots, N$ of the form

$$\mathcal{U}_i^{(v),j}(\boldsymbol{m}^{(v),j}) = \frac{1}{\lambda}\Big(-\alpha(\mathcal{P}_{R,i}^{(v),j}(\boldsymbol{m}^{(v),j}) \times \boldsymbol{m}_i^{(v),j}) \times \boldsymbol{m}_i^{(v),j} + \mathcal{P}_{R,i}^{(v),j}(\boldsymbol{m}^{(v),j}) \times \boldsymbol{m}_i^{(v),j}\Big). \tag{14.9}$$

By evaluating this function, we have an optimal response (forcing) for a observed realization of the state $\boldsymbol{m}(\omega)$.

However, this algorithm works only for very restrictive choices of the parameters (small final times T accompanied by large cost constants λ) and hence is not suitable to study switching dynamics of ferromagnetic particles, in particular. To overcome this problem, a stochastic gradient method was proposed in Part II of this thesis: in the present context, the idea is based on the following approximation of $\mathcal{U}^{(v),j}$ in (14.9) by a function $\mathcal{U}_R^{(v),j}$ in the linear subspace $\operatorname{span}\{\mathbb{1}_{C_r^{(v),j}}(\cdot);\, r=1,\dots,R\}$. By approximating the state $\boldsymbol{m}^{(v),j}$ in (14.9) by means of the (additional) Voronoi realization which assembles the region $C_r^{(v),j}$ where the realization take values, we obtain the expression

$$\begin{aligned}
\mathcal{U}_R^{(v),j}(\boldsymbol{x}) &= \frac{1}{\lambda}\Bigg(-\alpha\Big(\mathcal{P}_R^{(v),j}(\boldsymbol{x}) \times \sum_{r=1}^{R} \widehat{\mathbf{m}}_r^{(v),j}\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x})\Big) \times \sum_{r=1}^{R} \widehat{\mathbf{m}}_r^{(v),j}\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x}) \\
&\qquad + \mathcal{P}_R^{(v),j}(\boldsymbol{m}^{(v),j}) \times \sum_{r=1}^{R} \widehat{\mathbf{m}}_r^{(v),j}\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x})\Bigg) \\
&= \frac{1}{\lambda}\Bigg[-\alpha \sum_{r=1}^{R}\Bigg(\Big(\big[\mathbf{a}_r^{(v),j} + \nu k \sum_{l=1}^{3} \mathbf{b}_{l,r}^{(v),j} \times \mathbf{e}_l\big] \times \widehat{\mathbf{m}}_r^{(v),j}\Big) \times \widehat{\mathbf{m}}_r^{(v),j}\Bigg)\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x}) \\
&\qquad + \sum_{r=1}^{R}\Big(\big[\mathbf{a}_r^{(v),j} + \nu k \sum_{l=1}^{3} \mathbf{b}_{l,r}^{(v),j} \times \mathbf{e}_l\big] \times \widehat{\mathbf{m}}_r^{(v),j}\Big)\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x})\Bigg] \\
&=: \sum_{r=1}^{R} \mathbf{u}_r^{(v),j}\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x}),
\end{aligned} \tag{14.10}$$

using

$$\mathcal{P}_R^{(v),j}(\boldsymbol{x}) := \sum_{r=1}^{R} \mathbf{p}_r^{(v),j}\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x}) = \sum_{r=1}^{R}\big[\mathbf{a}_r^{(v),j} + \nu k \sum_{l=1}^{3} \mathbf{b}_{l,r}^{(v),j} \times \mathbf{e}_l\big]\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x}), \tag{14.11}$$

$$\boldsymbol{m}^{(v),j} \approx \sum_{r=1}^{R} \widehat{\mathbf{m}}_r^{(v),j}\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{m}^{(v),j}) \tag{14.12}$$

with the coefficients $\mathbf{a}^{(v),j}$, $\mathbf{b}^{(v),j}$ defined in Algorithm 1. The representation in approximation (14.10) is the starting point for a stochastic gradient method to approximatively solve Problem 11.1. The control $\boldsymbol{u}^{(v),j}$ in the forward equation, the backward equation, and in the cost functional are replaced by $\mathcal{U}_R^{(v),j}(\boldsymbol{m}^{(v),j}) = \sum_{r=1}^{R} \mathbf{u}_r^{(v),j}\mathbb{1}_{C_r^{(v),j}}(\boldsymbol{m}^{(v),j})$, and the focus is now on approximating the (deterministic) coefficients $\mathbf{u}^{(v)}$. By inserting the representations (14.10), (14.11), and (14.12) in the maximum condition (13.6) and differentiating according to the coefficients $\mathbf{u}^{(v)}$ (neglecting effects of $C_r^{(v),j}$), we obtain the descent direction $\mathbf{g}^{(v-1)}$

$$\begin{aligned}
\mathbf{g}^{(v-1)} &:= -\nabla_{\mathbf{u}}\mathcal{H}_{sto}(t, \widehat{\mathbf{m}}^{(v)}, \mathbf{u}^{(v-1)}, \mathbf{p}^{(v)}, \mathbf{q}^{(v)}) \\
&= \alpha(\mathbf{p}^{(v)} \times \widehat{\mathbf{m}}^{(v)}) \times \widehat{\mathbf{m}}^{(v)} - \mathbf{p}^{(v)} \times \widehat{\mathbf{m}}^{(v)} + \lambda \mathbf{u}^{(v-1)}.
\end{aligned} \tag{14.13}$$

Note that the objects $\mathbf{u}^{(v)}$, $\mathbf{p}^{(v)}$, $\widehat{\mathbf{m}}^{(v)}$, and $\mathbf{g}^{(v-1)}$ in (14.13) involve all regions and time steps $\{C_r^{(v),j};\, r=1,\dots,R,\, j=0,\dots,J\}$.

Scheme 14.4 (Stochastic gradient method for the stochastic control problem)

(1) Set $\mathbf{u}_r^{(0),j} \equiv \mathbf{u}_{init}^j$ for each $j = 0, \dots, J$, and for each basis region indexed by $r = 1, \dots, R$ in the first gradient iteration step.

(2) Iterate $v = 1, 2, \dots$ until a stopping criterion is met:

(i) **FSDE:** For each $j = 0, \dots, J-1$, compute $\mathcal{U}_R^{(v-1),j}(\boldsymbol{m}^{(v),j})$ according to (14.10). Simulate the $(\mathbb{R}^3)^N$-valued random variable $\boldsymbol{m}^{(v),j+1}$ by a time discretized scheme of (11.1) using $\mathcal{U}_R^{(v-1),j}(\boldsymbol{m}^{(v),j})$ and $\boldsymbol{m}^{(v),j}$.

(ii) **BSDE:** Set $\mathcal{P}^{(v),J}(\boldsymbol{x}) = -\kappa \frac{\partial h}{\partial \boldsymbol{x}}(\boldsymbol{m}^{(v),J})$. For each $j = J-1, \dots, 0$, approximate $\mathcal{Q}^{(v),J}(\cdot)$ and $\mathcal{P}^{(v),J}(\cdot)$ from (14.5)–(14.6) using the least squares Monte-Carlo method, $\mathcal{U}_R^{(v-1),j+1}(\cdot)$, $\mathcal{P}_R^{(v),j+1}(\cdot)$, $\boldsymbol{m}^{(v),j+1}$, and $\boldsymbol{m}^{(v),j}$.
Obtain $\mathcal{P}_R^{(v),j}(\boldsymbol{x})$, resp. $\mathcal{Q}_R^{(v),j}(\boldsymbol{x})$ with coefficients $\mathbf{p}^{(v),j}$, resp. $\mathbf{q}^{(v),j}$.

(iii) **Gradient step:** The coefficients $\mathbf{u}^{(v)}$ of the regression function $\mathcal{U}_R^{(v),\cdot}(\cdot) = \sum_{r=1}^R \mathbf{u}_r^{(v),\cdot} \mathbb{1}_{C_r^{(v),\cdot}}(\cdot)$ are computed according to

$$\mathbf{u}^{(v)} = \mathbf{u}^{(v-1)} - \sigma^{(v)} \mathbf{g}^{(v-1)}, \tag{14.14}$$

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

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

For the computation of the step size $\sigma^{(v)}$ in equation (14.14) we use a modification of Armijo's rule:

- Approximate the current cost function $\mathcal{J}_{sto}^{(v-1)}$ using the coefficients $\mathbf{u}^{(v-1)}$.
- Iterate $s = 0, 1, 2, \dots$ until a stopping criterion is met:
  - Compute $\mathbf{u}^{(v),s} = \mathbf{u}^{(v-1)} - \sigma^\star \beta^s \mathbf{g}^{(v-1)}$.
  - Approximate the cost function $\mathcal{J}_{sto}^{(v-1),s}$ using the coefficients $\mathbf{u}^{(v),s}$.
  - Stop, if $\mathcal{J}_{sto}^{(v-1),s} - \mathcal{J}_{sto}^{(v-1)} \leq -\underline{\sigma} \sigma^\star \beta^s \sum_{j=0}^J \sum_{r=1}^R \|\mathbf{g}_r^{(v-1),j}\|^2_{(\mathbb{R}^3)^N}$.

Algorithm 1 and 2 return a similar approximation in cases where Algorithm 1 is applicable. However, Algorithm 2 provides an approximation of the solution for less restrictive choices of the parameters and will therefore be used in Chapter 15.

Algorithm 2 Stochastic gradient method for the stochastic control problem

1: Simulate and store the increments $\Delta\widehat{\boldsymbol{W}} := \{(\Delta_j \widehat{\boldsymbol{W}}_r)_{j=0}^{J-1};\, r = 1, \dots, R\}$ for the construction of the basis regions.
2: Set $v = 0$. For the first iteration step, fix the initial control $\mathbf{u}^{(1),j} \equiv \mathbf{u}_{init}^{j}$ for each $j = 0, \dots, J$.
3: **while** Stopping criterion fails **do**
4: Follow lines 4–20 of Algorithm 1 using $\mathcal{U}_R^{(v-1),j}(\boldsymbol{m}_m^{(v),j})$, resp. $\mathcal{U}_R^{(v-1),j+1}(\boldsymbol{m}_m^{(v),j+1})$ in the computation of the forward, resp. backward equation.
5: Evaluate the cost function $\mathcal{J}_{sto}^{(v-1)}$ using the coefficients $\mathbf{u}^{(v-1)}$.
6: Set $s = 0$.
7: Compute $\mathbf{g}^{(v-1)}$ via

$$\begin{aligned}\mathbf{g}_r^{(v-1),j} := \alpha\Big(\big[\mathbf{a}_r^{(v),j} + \nu k \sum_{l=1}^{3} \mathbf{b}_{l,r}^{(v),j} \times \mathbf{e}_l\big] \times \widehat{\mathbf{m}}_r^{(v),j}\Big) \times \widehat{\mathbf{m}}_r^{(v),j} \\ - \big[\mathbf{a}_r^{(v),j} + \nu k \sum_{l=1}^{3} \mathbf{b}_{l,r}^{(v),j} \times \mathbf{e}_l\big] \times \widehat{\mathbf{m}}_r^{(v),j} + \lambda \mathbf{u}_r^{(v-1),j}.\end{aligned}$$

8: **while** Stopping criterion fails **do**
9: Compute $\mathbf{u}^{(v),s} = \mathbf{u}^{(v-1)} - \sigma^\star \beta^s \mathbf{g}^{(v-1)}$.
10: Evaluate the cost function $\mathcal{J}_{sto}^{(v-1),s}$ using the coefficients $\mathbf{u}^{(v),s}$.
11: Stop, if $\mathcal{J}_{sto}^{(v-1),s} - \mathcal{J}_{sto}^{(v-1)} \leq -\underline{\sigma}\sigma^\star\beta^s \sum_{j=0}^{J}\sum_{r=1}^{R} \|\mathbf{g}_r^{(v-1),j}\|^2_{(\mathbb{R}^3)^N}$.
12: Set $s = s + 1$.
13: **end while**
14: Fix $\mathbf{u}^{(v)} := \mathbf{u}^{(v),s}$ and define $\mathcal{U}_R^{(v),j}(\boldsymbol{x}) := \sum_{r=1}^{R} \mathbf{u}_r^{(v),j} \mathbb{1}_{C_r^{(v),j}}(\boldsymbol{x})$.
15: Evaluate the cost function $\mathcal{J}_{sto}^{(v)}(\cdot)$ and its parts to decide whether to stop or not.
16: **end while**

14.2. Deterministic optimal control

Similarly to Scheme 14.3, the explicit Euler method is used for the time discretization of the (backward) ODE in equation (13.1).

Scheme 14.5 (Time discretization of the BODE)
Let $\boldsymbol{u}^{(v),j+1}$, $\boldsymbol{m}^{(v),j+1}$, and $\boldsymbol{p}^{(v),j+1}$ be given. Compute $\boldsymbol{p}^{(v),j} \in (\mathbb{R}^3)^N$ by the following method: For each $i = 1, \dots, N$, compute

$$\boldsymbol{p}_i^{(v),j} = \boldsymbol{p}_i^{(v),j+1} + k\mathbf{f}_i(\boldsymbol{m}^{(v),j+1}, \boldsymbol{u}^{(v),j+1}, \boldsymbol{p}^{(v),j+1}) + \delta k(\boldsymbol{m}_i^{(v),j+1} - \widetilde{\boldsymbol{m}}_i(t_{j+1})). \tag{14.15}$$

For time discretization, we consider the deterministic version of Scheme 14.2. Then a (standard) gradient method is used to solve the (deterministic) optimal control problem.

15. Computational studies

We present numerical experiments concerning the deterministic and the stochastic optimal control problem. Therefore, two different scenarios are considered: First, the behavior of the deterministic and stochastic control for the switching of a single nanomagnetic particle is under investigation, where we consider the case of a restricted control, that is a control which only can attain values in some submanifold of $\mathbb{R}^3$, i.e., $\mathbb{R}^2$ or $\mathbb{R}^1$. In the second part our focus is on controlling the magnetization of a finite ensemble of nanomagnetic particles. Here, different scenarios are considered which involve exchange energies and the case where a concrete time evolution of the desired target profile $\widetilde{\boldsymbol{m}}$ is of interest.

In order to simulate both settings we approximate the necessary optimality conditions (11.1), (13.5), (14.1), by the stochastic gradient method (Algorithm 2) using the stopping criterion $\frac{1}{R}\sum_{r=1}^{R} k \sum_{j=0}^{J} \|\mathbf{g}_r^{(v-1),j}\|^2_{(\mathbb{R}^3)^N} \leq \texttt{TOL}$. The parameters $(\beta, \sigma^\star, \underline{\sigma}, \texttt{TOL})$ of the Armijo rule are adjusted to each specific configuration to avoid unnecessary iteration steps and to save simulation time. In the following, we denote by $\boldsymbol{u}_{sto}^{(v),j} := \mathcal{U}_R^{(v),j}(\boldsymbol{m}_{sto}^{(v),j})$ resp. $\boldsymbol{u}_{sto}^{\star,j} := \mathcal{U}_R^{\star,j}(\boldsymbol{m}_{sto}^{\star,j})$ the approximation of the optimal control.

15.1. Single nanomagnetic particle

Consider a single nanomagnetic particle ($N \equiv 1$). Our interest lies in the optimal magnetization switching from $\boldsymbol{m}_0$ (at initial time) to $(-\boldsymbol{m}_0)$ at final time T:

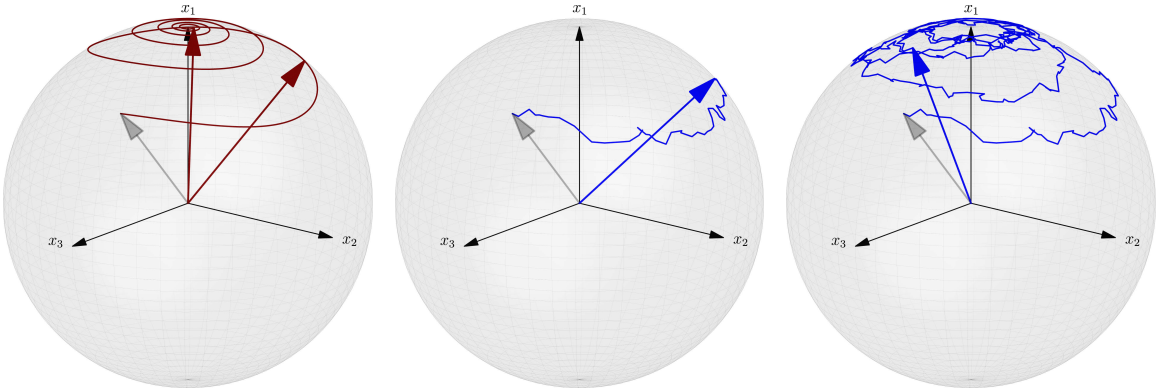
Example 15.1
Fix $T > 0$, $\alpha, \lambda, \kappa > 0$, and set $\delta = 0$, $N = 1$. Solve Problem 11.1 where $h(\boldsymbol{m}(T)) = \|\boldsymbol{m}(T) - (-\boldsymbol{m}_0)\|^2_{\mathbb{R}^3}$ in (11.2). The effective field $\boldsymbol{\mathcal{H}}_{eff}$ in equation (11.1) consists of the external field $\boldsymbol{\mathcal{H}}_{ext}$, and a demagnetization as well as an anisotropy field in $\boldsymbol{\mathcal{H}}_{d,ani}$.

Table 15.1. Parameter Setups **A** and **B** for the stochastic control problem considered in Example 15.1.

	T	α	ν	(λ, κ)	$\boldsymbol{m}_0$	$C_{d,ani}$	$(d_{x_1}, d_{x_2}, d_{x_3})$
Setup A	0.5	0.1	0.3	$(0.001, 1.0)$	$\mathbf{e}_1$	0.0	–
Setup B	0.5	0.1	0.3	$(0.001, 1.0)$	$\mathbf{e}_1$	5.0	$(-1.0, 0.2, 0.7)$

We consider two different parameter setups which are proposed in Table 15.1 for the simulation of Example 15.1. Setup **A** avoids the demagnetization and anisotropy energy ($C_{d,ani} \equiv 0.0$), moreover in Setup **B** the axis $\mathbf{x}_1$ where $\boldsymbol{m}_0$ is located is more likely; see

Figure 15.1, where the trajectory of a path of equation (11.1) without control and initial value $\boldsymbol{m}_0 = (\sqrt{0.5}, 0.1, 0.7)^T$ is shown. Figure 15.1 **(a)** shows asymptotic alignment with the easy axis ($\mathbf{e}_1 \in \mathbb{R}^3$) due to anisotropy effects. Starting with $\boldsymbol{m}_0 = (-1, 0, 0)^T$ an appropriate switching strategy might be to move the magnetization to the northern hemisphere, and then allow anisotropy effects to complete alignment with $\mathbf{e}_1$. In this framework, if the final time T is fixed or noise due to an existing heat bath is relevant, a corresponding optimal switching strategy is quite complex, which is why we solve/simulate the stochastic optimal control problem (Example 15.1).



(a) $t \mapsto \boldsymbol{m}_{det}$ in $[0, 5]$ **(b)** $t \mapsto \boldsymbol{m}_{sto}$ in $[0, 0.5]$ **(c)** $t \mapsto \boldsymbol{m}_{sto}$ in $[0, 5]$

Figure 15.1. One solution path $\boldsymbol{m}_{det}$ resp. $\boldsymbol{m}_{sto}$ of equation (11.1) without control ($\boldsymbol{u} \equiv 0$), starting in $\boldsymbol{m}_0 = (\sqrt{0.5}, 0.1, 0.7)^T$ and Setup **B** using $T \in \{0.5, 5\}$.

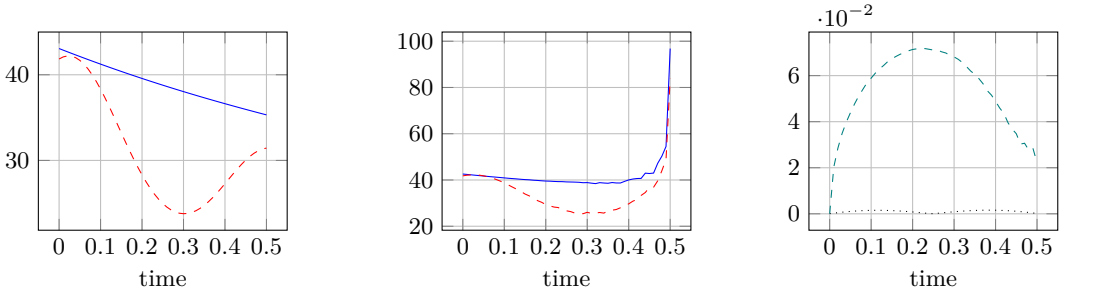
For the simulation of the stochastic optimal control problem we use Algorithm 2 with the parameters given in Table 15.2, where $\widetilde{M} = 1.0 \cdot 10^6$ independent paths for the evaluation of the cost function $\mathcal{J}_{sto}$ (also independent of the M paths which are used in the computation of the stochastic backward equation) are considered.

Table 15.2. Discretization parameters used to simulate Example 15.1.

J	M	$\widetilde{M}$	R
50	$5.0 \cdot 10^6$	$1.0 \cdot 10^6$	100

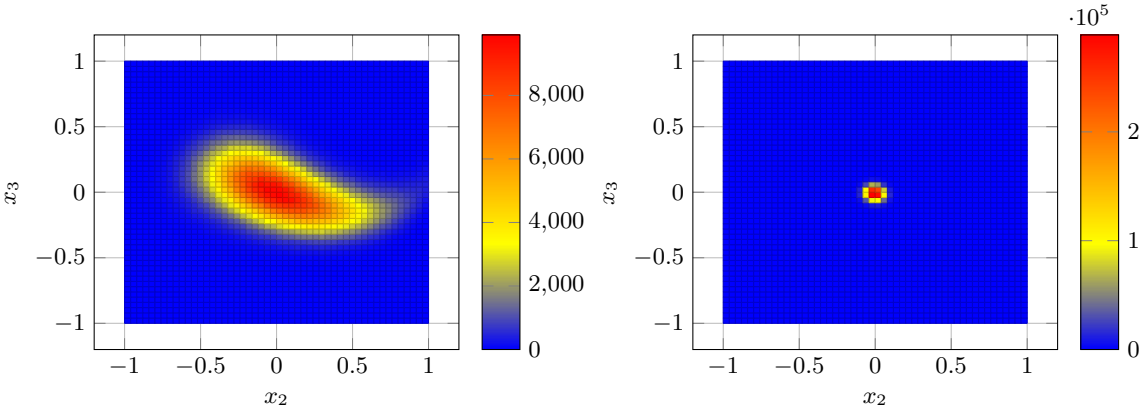
We start with an optimal control of the corresponding deterministic control problem for Setups **A** and **B** (with $\nu \equiv 0$). For both Setups **A** and **B**, we obtain comparable cost functionals: for Setup **A**, the cost functional is $\mathcal{J}^\star_{det} = 0.0099$, with $k \sum_{j=0}^{J} \|\boldsymbol{u}^{\star,j}_{det}\|^2_{\mathbb{R}^3} = 19.85$ and $\|\boldsymbol{m}^{\star,J}_{det} + \boldsymbol{m}_0\|^2_{\mathbb{R}^3} = 3.494 \cdot 10^{-5}$; in Setup **B**, we allow for less control, and hence $k \sum_{j=0}^{J} \|\boldsymbol{u}^{\star,j}_{det}\|^2_{\mathbb{R}^3} = 15.82$ while obtaining a similar accuracy of $\boldsymbol{m}^J$ at the final time ($\|\boldsymbol{m}^{\star,J}_{det} + \boldsymbol{m}_0\|^2_{\mathbb{R}^3} = 3.108 \cdot 10^{-5}$) and thus a slightly smaller cost functional $\mathcal{J}^\star_{det} = 0.0079$. There, the time evolution of the deterministic control differs: in the beginning more external force is necessary to escape the easy axis and initiate switching; in the end less control is needed due to the attraction of the easy axis; see Figure 15.2(A).

The optimal control $\boldsymbol{u}^{\star}_{det}$ for the deterministic control problem seems to be a promising candidate as initial value $\boldsymbol{u}^{(0)}_{sto}$ of the stochastic gradient method. For Setup **B**, we hence start with $\mathcal{J}^{(0)}_{sto} = 0.0799$. Then, most optimal trajectories realize a movement of the state $\boldsymbol{m}^{(0)}_{sto}$ to the northern hemisphere $\mathbf{x}_1 > 0$ (i.e., 99.978% of the simulated paths); however the magnetization of the particle might at final time not be close to $(-\boldsymbol{m}_0)$; see e.g. Figure 15.3(A), where the distribution of the $(\mathbf{x}_2, \mathbf{x}_3)$ coordinates of $\boldsymbol{m}^{(0),J}_{sto}$ are illustrated.



(a) $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{det}\|^2_{\mathbb{R}^3}$ **(b)** $t_j \mapsto \mathbb{E}\big[\|\boldsymbol{u}^{\star,j}_{sto}\|^2_{\mathbb{R}^3}\big]$ **(c)** $t_j \mapsto \ll \boldsymbol{u}^{\star,j}_{\cdot}, \boldsymbol{m}^{\star,j}_{\cdot} \gg$

Figure 15.2. Intensities $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{det}\|^2_{\mathbb{R}^3}$, resp. $t_j \mapsto \mathbb{E}\big[\|\boldsymbol{u}^{\star,j}_{sto}\|^2_{\mathbb{R}^3}\big]$ for Setup **A** (——) and **B** (- - -). Time evolution of the angle $t_j \mapsto \ll \boldsymbol{u}^{\star,j}_{\cdot}, \boldsymbol{m}^{\star,j}_{\cdot} \gg$ for the deterministic (·····) and stochastic (- - -) problem using Setup **A**.



(a) Distribution of $\boldsymbol{m}^{(0),J}_{sto}$ **(b)** Distribution of $\boldsymbol{m}^{\star,J}_{sto}$

Figure 15.3. Distribution of the $(\mathbf{x}_2, \mathbf{x}_3)$ values of $\boldsymbol{m}^{(0),J}_{sto}$ resp. $\boldsymbol{m}^{\star,J}_{sto}$ for Example 15.1 and Setup **B** (if $\mathbf{x}_1 \geq 0$).

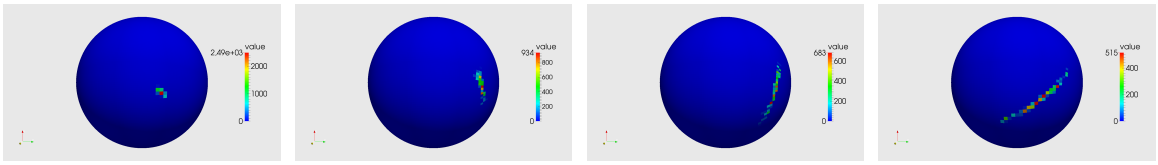
By applying Algorithm 2 we are able to find an optimal control $\mathcal{U}^{\star,j}_{sto}(\cdot)$, which for Example 15.1 with Setup **B** returns a cost function $\mathcal{J}^{\star}_{sto} = 0.0100$ with parts $k\sum_{j=0}^{J}\mathbb{E}\big[\|\boldsymbol{u}^{\star,j}_{sto}\|^2_{\mathbb{R}^3}\big] = 17.09$, and $\mathbb{E}\big[\|\boldsymbol{m}^{\star,J}_{sto} + \boldsymbol{m}_0\|^2_{\mathbb{R}^3}\big] = 0.0028$. Here, the behavior of the control process differs if compared with the deterministic case; see Figure 15.2 **(b)**: in the beginning it behaves in average like the deterministic control function; but close to the terminal time T more control is used to ensure that especially at the final time the state is close to the desired direction. The improvement at terminal time is pointed out in Figure 15.3 **(b)**, where the

distribution of the $(\mathbf{x}_2, \mathbf{x}_3)$ coordinates of $\boldsymbol{m}_{sto}^{\star,J}$ are displayed.

Equation (14.1) states that the optimal control $\boldsymbol{m}_{\cdot}^{\star}$ is orthogonal to $\boldsymbol{u}_{\cdot}^{\star}$. Figure 15.2 **(c)** indicates this behavior for the iterates of the time discretization. There, the temporal evolution

$$t_j \mapsto \ll \boldsymbol{u}_{\cdot}^{\star,j}, \boldsymbol{m}_{\cdot}^{\star,j} \gg := \mathbb{E}\Big[\frac{|\langle \boldsymbol{u}_{\cdot}^{\star,j}, \boldsymbol{m}_{\cdot}^{\star,j}\rangle_{\mathbb{R}^3}|}{\|\boldsymbol{u}_{\cdot}^{\star,j}\|_{\mathbb{R}^3}\|\boldsymbol{m}_{\cdot}^{\star,j}\|_{\mathbb{R}^3}}\Big]$$

of the angle of $\boldsymbol{u}_{\cdot}^{\star,j}$ and $\boldsymbol{m}_{\cdot}^{\star,j}$ is illustrated, showing that control and state are almost orthogonal for Setup **A**; this relation also holds for Setup **B**.



(a) $t_j = 0.1$ **(b)** $t_j = 0.2$ **(c)** $t_j = 0.3$ **(d)** $t_j = 0.4$

Figure 15.4. Density of the direction of the optimal control $\boldsymbol{u}_{sto}^{\star,j}\|\boldsymbol{u}_{sto}^{\star,j}\|_{\mathbb{R}^3}^{-1}$ at certain time points $t_j \in \{0.1, 0.2, 0.3, 0.4\}$ for Example 15.1 and Setup **B**.

Figure 15.4 displays the density of the direction of the stochastic optimal control, showing that for Setup **B** the direction of the optimal control is close to the $\mathbf{x}_2$–axis, with variations close to final time T.

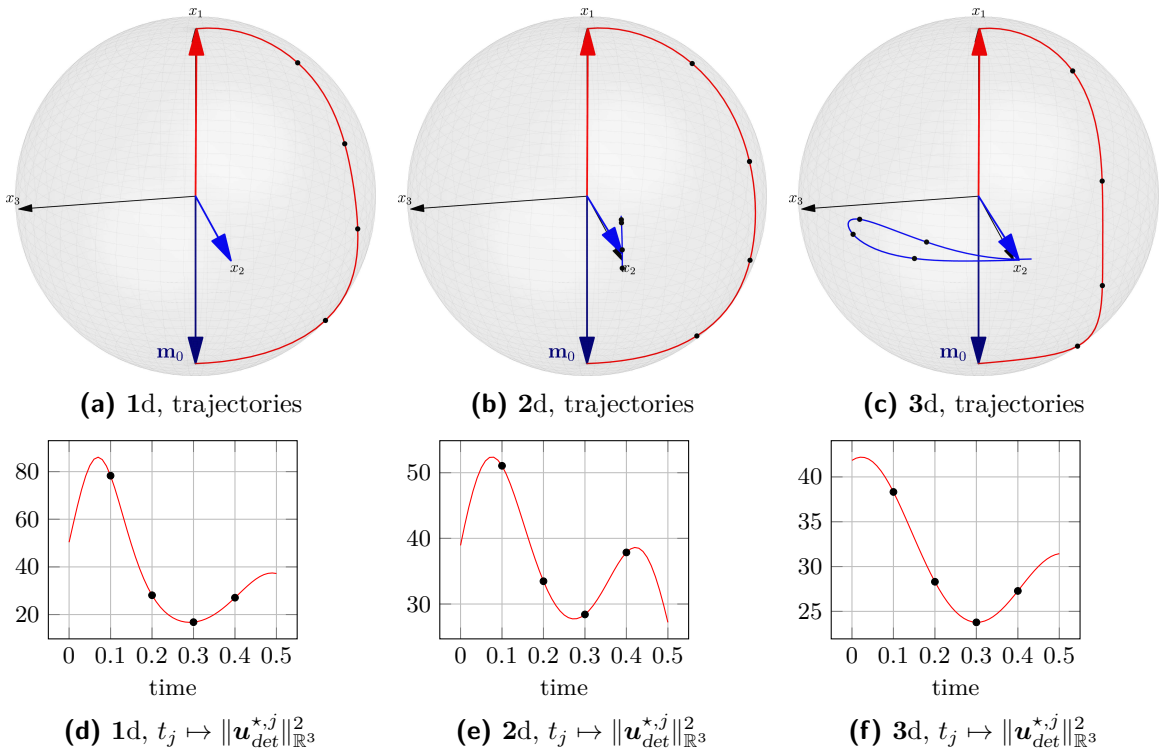
Next, we study Example 15.1 with Setup **B** and reduced (low dimensional) control, i.e., the cases where the control can only attain values in the $\mathbf{x}_1$–$\mathbf{x}_2$ plain (denoted by **2**d control), or only values in the $\mathbf{x}_2$-direction (denoted by **1**d control). Changes in the cost functional and its parts are shown in Table 15.3 for Setup **B**, indicating that for the deterministic optimal control problem only slight differences occur by restricting the control to the $\mathbf{x}_2$-axis resp. the $\mathbf{x}_1$–$\mathbf{x}_2$ plain.

Table 15.3. Comparison of the resulting cost functionals for different dimensions of the control for Example 15.1 and Setup **B**.

	Deterministic Control Problem:			**Stochastic Control Problem:**		
	$\mathcal{J}_{det}^{\star}$	$k\sum_j \|\boldsymbol{u}_{det}^{\star,j}\|_{\mathbb{R}^3}^2$	$\|\boldsymbol{m}_{det}^{\star,J} + \boldsymbol{m}_0\|_{\mathbb{R}^3}^2$	$\mathcal{J}_{sto}^{\star}$	$k\sum_j \mathbb{E}\big[\|\boldsymbol{u}_{sto}^{\star,j}\|_{\mathbb{R}^3}^2\big]$	$\mathbb{E}\big[\|\boldsymbol{m}_{sto}^{\star,J} + \boldsymbol{m}_0\|_{\mathbb{R}^3}^2\big]$
1d	**0.0103**	20.59	$3.699 \cdot 10^{-5}$	**0.0184**	24.15	0.0127
2d	**0.0096**	19.23	$3.786 \cdot 10^{-5}$	**0.0142**	21.02	0.0075
3d	**0.0079**	15.82	$3.108 \cdot 10^{-5}$	**0.0101**	17.49	0.0027

The time evolution of the deterministic optimal state and the deterministic optimal control are shown in Figure 15.5. The intensity of the optimal control differs: in the **1**d case a large intensity of control is needed in the beginning to transfer the state to the northern hemisphere, where then a combination of anisotropy and control steer the state in the desired target position. This behavior is similar for the **2**d and **3**d cases, where the control is free to choose different directions of the control, and thus a less intense control is needed.

For $\nu \neq 0$, we observe again a sharp increase of the control at terminal times in all three cases. The increase in the intensity of the control is accompanied with changes in the direction of the control close to the final time; see Figure 15.6.

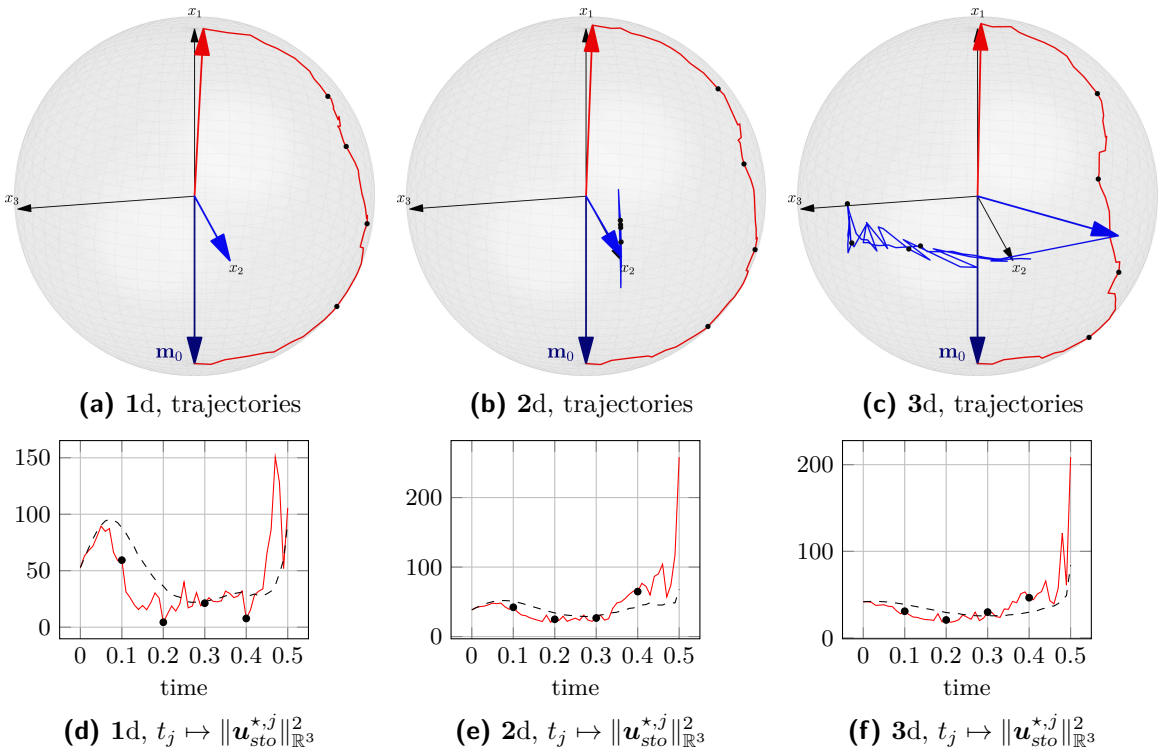


(a) **1**d, trajectories **(b)** **2**d, trajectories **(c)** **3**d, trajectories

(d) **1**d, $t_j \mapsto \|\boldsymbol{u}_{det}^{\star,j}\|_{\mathbb{R}^3}^2$ **(e)** **2**d, $t_j \mapsto \|\boldsymbol{u}_{det}^{\star,j}\|_{\mathbb{R}^3}^2$ **(f)** **3**d, $t_j \mapsto \|\boldsymbol{u}_{det}^{\star,j}\|_{\mathbb{R}^3}^2$

Figure 15.5. Time evolution of the optimal state $t_j \mapsto \boldsymbol{m}_{det}^{\star,j}$ (red) and the direction of the optimal control $t_j \mapsto \boldsymbol{u}_{det}^{\star,j}\|\boldsymbol{u}_{det}^{\star,j}\|_{\mathbb{R}^3}^{-1}$ (blue), its magnitude of control $t_j \mapsto \|\boldsymbol{u}_{det}^{\star,j}\|_{\mathbb{R}^3}^2$ (——) for different dimensions of the control (**1**d, **2**d, **3**d) using Example 15.1 and Setup **B** in the deterministic case ($\nu \equiv 0$).

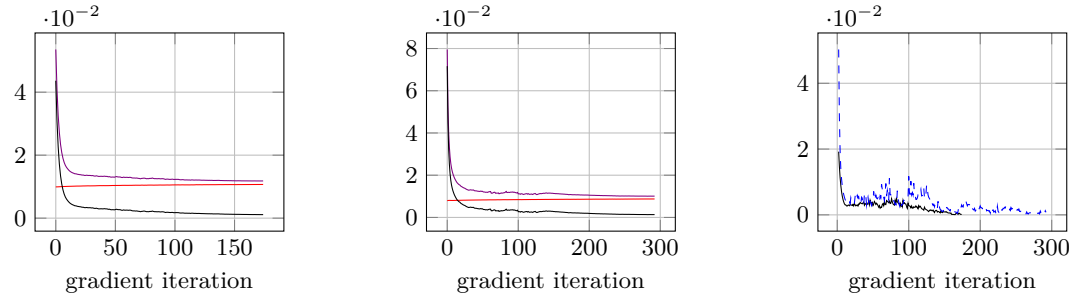
The characteristics of the stochastic gradient method is illustrated in Figure 15.7 where the evolution of the cost functional and its parts are shown in **(a)**–**(b)**, indicating convergence of the stochastic gradient method. The distance of two consecutive Voronoi partitions is quantified by $\mathcal{V}_D^{(v)} := \big(\frac{1}{R}\sum_{r=1}^{R} k \sum_{j=1}^{J-1} \|\widehat{\boldsymbol{m}}_{sto}^{(v),j} - \widehat{\boldsymbol{m}}_{sto}^{(v-1),j}\|_{\mathbb{R}^3}^2\big)^{\frac{1}{2}}$, where we observe that the construction of the basis functions in $\mathcal{U}_R^{(v-1),j}$ according to the law of $\boldsymbol{m}^{(v-1),j}$ while being evaluated at $\boldsymbol{m}^{(v),j}$ does not affect the results.



(a) **1**d, trajectories **(b)** **2**d, trajectories **(c)** **3**d, trajectories

(d) **1**d, $t_j \mapsto \|\boldsymbol{u}_{sto}^{\star,j}\|_{\mathbb{R}^3}^2$ **(e)** **2**d, $t_j \mapsto \|\boldsymbol{u}_{sto}^{\star,j}\|_{\mathbb{R}^3}^2$ **(f)** **3**d, $t_j \mapsto \|\boldsymbol{u}_{sto}^{\star,j}\|_{\mathbb{R}^3}^2$

Figure 15.6. One trajectory of the optimal state $t_j \mapsto \boldsymbol{m}_{sto}^{\star,j}$ (red) and the direction of the optimal control $t_j \mapsto \boldsymbol{u}_{sto}^{\star,j}\|\boldsymbol{u}_{sto}^{\star,j}\|_{\mathbb{R}^3}^{-1}$ (blue), its magnitude of control $t_j \mapsto \|\boldsymbol{u}_{sto}^{\star,j}\|_{\mathbb{R}^3}^2$ (——) and expectation (- - -) for different dimensions of the control (**1**d, **2**d, **3**d) using Example 15.1 and Setup **B**.



(a) Setup **A**, $v \mapsto \mathcal{J}_{sto}^{(v)}$ **(b)** Setup **B**, $v \mapsto \mathcal{J}_{sto}^{(v)}$ **(c)** $v \mapsto \mathcal{V}_D^{(v)}$

Figure 15.7. Behavior of the cost functional $\mathcal{J}_{sto}^{(v)}$ (——) and its parts $\mathbb{E}\big[\frac{\lambda}{2}k\sum_{j=0}^{J}\|\boldsymbol{u}_{sto}^{(v),j}\|_{\mathbb{R}^3}^2\big]$ (——) and $\mathbb{E}\big[\frac{\kappa}{2}\|\boldsymbol{m}_{sto}^{(v),J} + \boldsymbol{m}_0\|_{\mathbb{R}^3}^2\big]$ (——) in the simulation of Example 15.1. Distance $\mathcal{V}_D^{(v)}$ of two consecutive Voronoi partitions for Setups **A** (——) and **B** (- - -).

15.2. Finite ensemble of nanomagnetic particles

Next, we focus on the behavior of an ensemble of $N = 3$ nanomagnetic particles, which additionally are subjected to exchange forces. Our interest lies again in the switching control for one ($i = 2$) of these particles from $\boldsymbol{m}_{0,2}$ (at initial time) to $(-\boldsymbol{m}_{0,2})$ at given final time T:

Example 15.2
Let $T > 0$, $\alpha, \delta, \kappa \geq 0$, and $\lambda > 0$ be fixed. Solve Problem 11.1, where $h(\boldsymbol{m}(T)) := \|\boldsymbol{m}(T) - \widetilde{\boldsymbol{m}}(T)\|^2_{(\mathbb{R}^3)^N}$ in (11.2). The effective field $\boldsymbol{\mathcal{H}}_{\mathit{eff}}$ in equation (11.1) consists of the external field $\boldsymbol{\mathcal{H}}_{\mathit{ext}}$, the exchange $\boldsymbol{\mathcal{H}}_{\mathit{exch}}$, and a demagnetization as well as an anisotropy in $\boldsymbol{\mathcal{H}}_{d,ani}$.

Corresponding parameter setups are given in Table 15.4, with varying values of C_{exch}. Setups **C**, **D**, and **E** only differ in the considered parts of the cost functional $\mathcal{J}_{\mathit{sto}}$: In Setup **C** ($\delta = 0, \kappa = 1$) the ensemble of particles is only required to be close to the desired target profile at terminal time; however, in Setup **D** ($\delta = 1, \kappa = 0$) we control the whole evolution of the state, and Setup **E** ($\delta = 0.5, \kappa = 0.5$) is a combination of both goals.
The additional influence of the exchange is illustrated in Figure 15.8, where the trajectory of a path of equation (11.1) without control and initial value $\boldsymbol{m}_{0,1} = \boldsymbol{m}_{0,3} = \mathbf{e}_1$, $\boldsymbol{m}_{0,2} = (\sqrt{0.5}, 0.1, 0.7)^T$ is shown. It is due to exchange $\boldsymbol{m}_2$ processes faster to $\mathbf{e}_1$, while $\boldsymbol{m}_1$ and $\boldsymbol{m}_3$ rotate out of $\mathbf{e}_1$.

Table 15.4. Parameter Setups **C**, **D**, and **E** for the stochastic control problem considered in Example 15.2.

	T	α	ν	$(\delta, \lambda, \kappa)$	$\boldsymbol{m}_0$	$C_{d,ani}$	$(d_{x_1}, d_{x_2}, d_{x_3})$
Setup C	0.5	0.1	0.3	$(0.0, 0.001, 1.0)$	$(\mathbf{e}_1, -\mathbf{e}_1, \mathbf{e}_1)$	5.0	$(-1.0, 0.2, 0.7)$
Setup D	0.5	0.1	0.3	$(1.0, 0.001, 0.0)$	$(\mathbf{e}_1, -\mathbf{e}_1, \mathbf{e}_1)$	5.0	$(-1.0, 0.2, 0.7)$
Setup E	0.5	0.1	0.3	$(0.5, 0.001, 0.5)$	$(\mathbf{e}_1, -\mathbf{e}_1, \mathbf{e}_1)$	5.0	$(-1.0, 0.2, 0.7)$

For the control problem, we use again Algorithm 2 for the discretization parameters given in Table 15.5, and start the stochastic gradient method with $\boldsymbol{u}^\star_{\mathit{det}}$ for $\nu = 0$.

Table 15.5. Discretization parameters used to simulate Example 15.2.

J	M	$\widetilde{M}$	R
50	$5.0 \cdot 10^6$	$5.0 \cdot 10^5$	100

We start with the case of no exchange $C_{\mathit{exch}} = 0.0$. In the corresponding deterministic problem (i.e., $\nu = 0$) the same result as in the single particle case (Example 15.1, Setup **B**) is obtained ($\mathcal{J}^\star_{\mathit{det}} = 0.0078$), since there is no interaction between the particles and the first and third particle starts already in the desired state, thus there is no need to control these particles. This behavior changes for $\nu \neq 0$: control is needed also for the first and the third particle due to the noise. This control is in the $\mathbf{x}_1$–$\mathbf{x}_2$ plain and balances the random

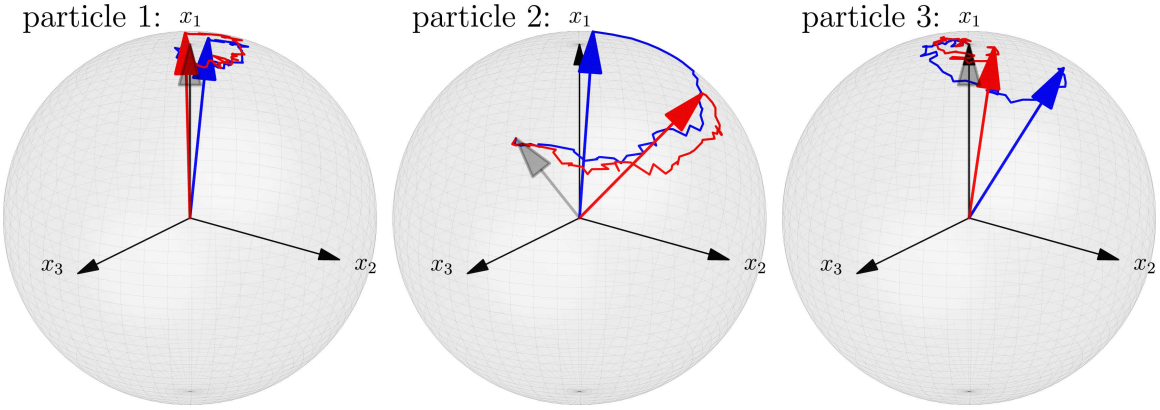


Figure 15.8. Time evolution of the solution $\boldsymbol{m}_{sto}$ of equation (11.1) without control ($\boldsymbol{u} \equiv 0$), starting from $\boldsymbol{m}_{0,1} = \mathbf{e}_1$, $\boldsymbol{m}_{0,2} = (\sqrt{0.5}, 0.1, 0.7)^T$, and $\boldsymbol{m}_{0,3} = \mathbf{e}_1$ for Setup **C** with exchange ($C_{exch} = 1.0$, blue), and without exchange ($C_{exch} = 0.0$, red).

influence. The control of the second particle is similar to the single spin case. Again, relevant control is needed at terminal time to steer the state into the desired target state; see Figure 15.9.

This behavior only slightly changes in the presence of exchange forces ($C_{exch} = 1.0$); see Figure 15.10: Here, more control is needed especially for the first and the third particle to match the target state.

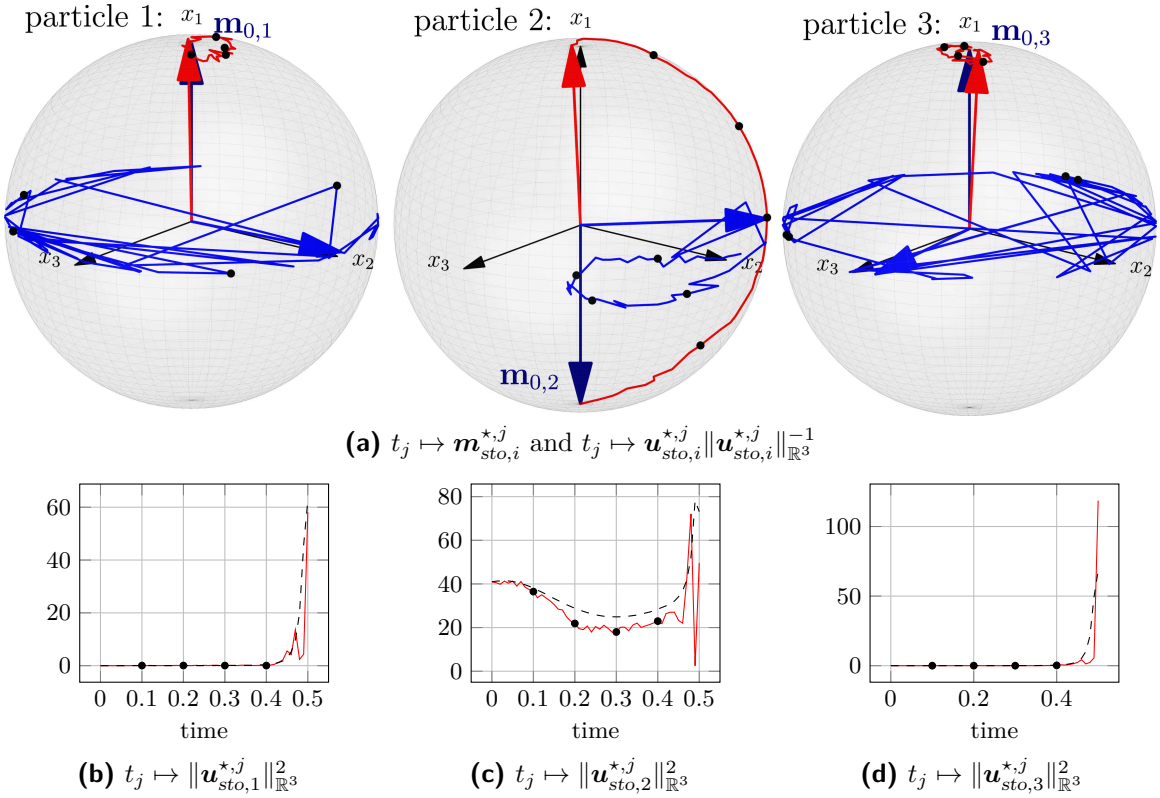
Next, we study Setups **D** and **E** with deterministic target profile $\widetilde{\boldsymbol{m}} : [0, T] \to (\mathbb{S}^2)^N$ defined by

$$\widetilde{\boldsymbol{m}}_1(t) \equiv \mathbf{e}_1, \qquad \widetilde{\boldsymbol{m}}_2(t) = \begin{pmatrix} -\cos(\pi \frac{t}{T}) \\ \sin(\pi \frac{t}{T}) \\ 0 \end{pmatrix}, \qquad \widetilde{\boldsymbol{m}}_3(t) \equiv \mathbf{e}_1,$$

which (continuously) rotates the magnetization of the second particle from $(-\mathbf{e}_1)$ to $\mathbf{e}_1$ in $[0, T]$. Since the state $\boldsymbol{m}^\star$ is now required to stay close to the desired target profile $\widetilde{\boldsymbol{m}}$, the optimal control acts in the whole time interval and not only with a high intensity at the end; see Figure 15.11.

Table 15.6. Comparison of the resulting cost functionals for different parameter setups of Example 15.2.

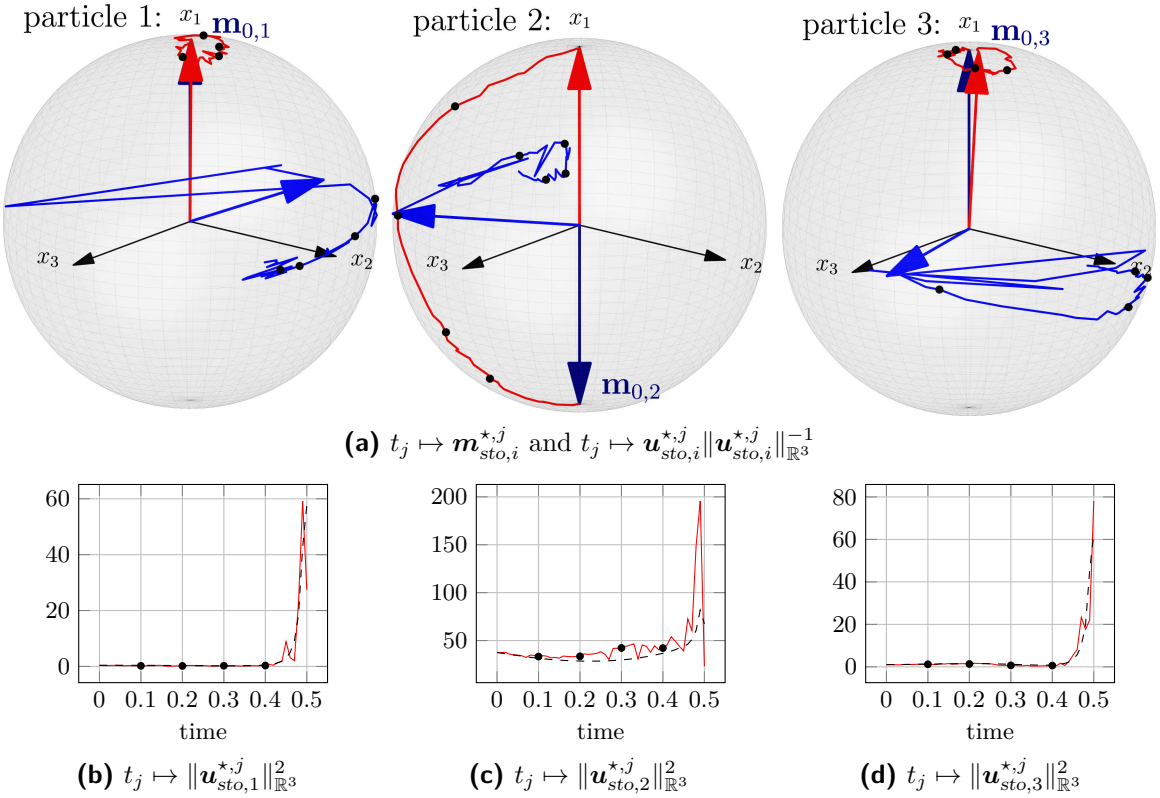
	C_{exch}	$\mathcal{J}^\star_{sto}$	$k \sum_j \mathbb{E}\big[\|\boldsymbol{m}^{\star,j}_{sto} - \widetilde{\boldsymbol{m}}^j\|^2_{(\mathbb{R}^3)^3}\big]$	$k \sum_j \mathbb{E}\big[\|\boldsymbol{u}^{\star,j}_{sto}\|^2_{(\mathbb{R}^3)^3}\big]$	$\mathbb{E}\big[\|\boldsymbol{m}^{\star,J}_{sto} - \widetilde{\boldsymbol{m}}^J\|^2_{(\mathbb{R}^3)^3}\big]$
Setup C	0.0	**0.0221**	–	20.53	0.0237
Setup C	1.0	**0.0238**	–	21.61	0.0260
Setup D	1.0	**0.0195**	0.0162	22.75	0.0642
Setup E	1.0	**0.0228**	0.0270	27.35	0.0153



(a) $t_j \mapsto \boldsymbol{m}^{\star,j}_{sto,i}$ and $t_j \mapsto \boldsymbol{u}^{\star,j}_{sto,i} \|\boldsymbol{u}^{\star,j}_{sto,i}\|^{-1}_{\mathbb{R}^3}$

(b) $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,1}\|^2_{\mathbb{R}^3}$ **(c)** $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,2}\|^2_{\mathbb{R}^3}$ **(d)** $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,3}\|^2_{\mathbb{R}^3}$

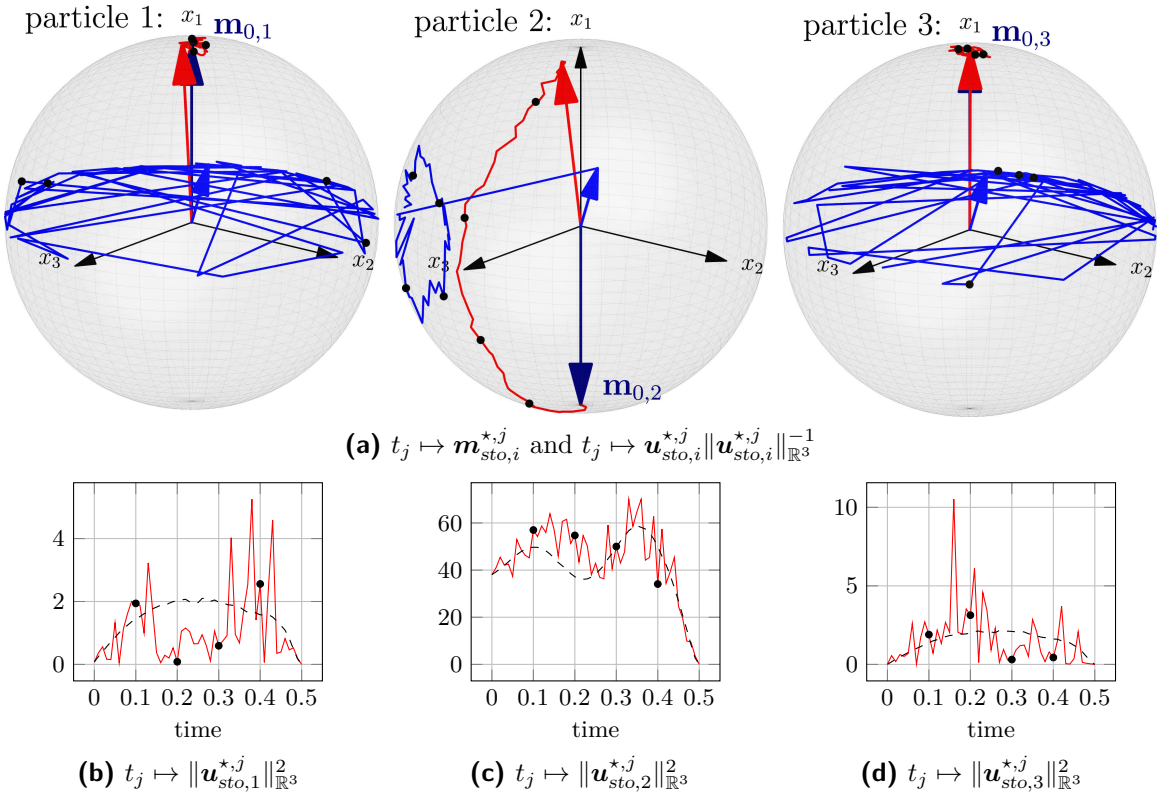
Figure 15.9. Time evolution of the optimal state $t_j \mapsto \boldsymbol{m}^{\star,j}_{sto,i}$ (red) and the direction of the optimal control $t_j \mapsto \boldsymbol{u}^{\star,j}_{sto,i} \|\boldsymbol{u}^{\star,j}_{sto,i}\|^{-1}_{\mathbb{R}^3}$ (blue), the magnitude of control $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,i}\|^2_{\mathbb{R}^3}$ (——), and the expectation (- - -) for each particle $i = 1, 2, 3$ using Example 15.2 and Setup **C** in the case where no exchange is considered ($C_{exch} = 0.0$).



(a) $t_j \mapsto \boldsymbol{m}^{\star,j}_{sto,i}$ and $t_j \mapsto \boldsymbol{u}^{\star,j}_{sto,i} \|\boldsymbol{u}^{\star,j}_{sto,i}\|^{-1}_{\mathbb{R}^3}$

(b) $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,1}\|^2_{\mathbb{R}^3}$ **(c)** $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,2}\|^2_{\mathbb{R}^3}$ **(d)** $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,3}\|^2_{\mathbb{R}^3}$

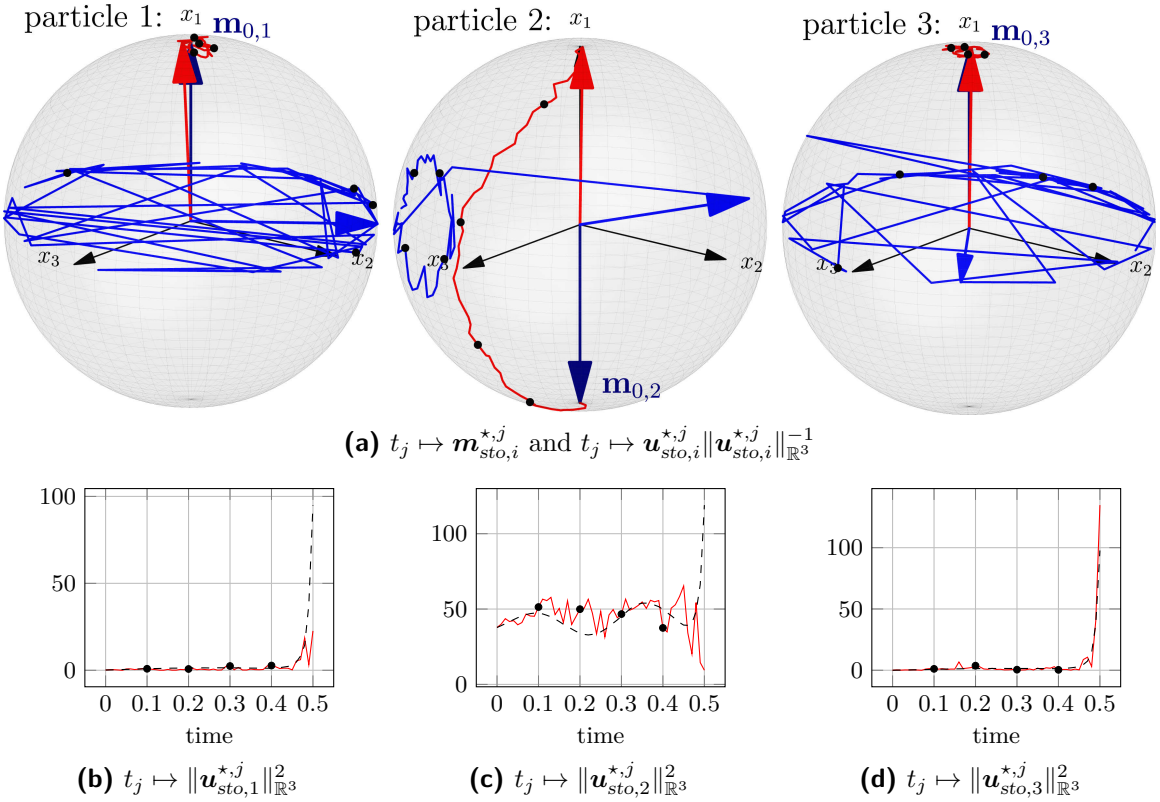
Figure 15.10. Time evolution of the optimal state $t_j \mapsto \boldsymbol{m}^{\star,j}_{sto,i}$ (red) and the direction of the optimal control $t_j \mapsto \boldsymbol{u}^{\star,j}_{sto,i} \|\boldsymbol{u}^{\star,j}_{sto,i}\|^{-1}_{\mathbb{R}^3}$ (blue), the magnitude of the control $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,i}\|^2_{\mathbb{R}^3}$ (——), and the expectation (- - -) for each particle $i = 1, 2, 3$ using Example 15.2 and Setup **C** in the case where exchange is considered ($C_{exch} = 1.0$).



(a) $t_j \mapsto \boldsymbol{m}^{\star,j}_{sto,i}$ and $t_j \mapsto \boldsymbol{u}^{\star,j}_{sto,i}\|\boldsymbol{u}^{\star,j}_{sto,i}\|^{-1}_{\mathbb{R}^3}$

(b) $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,1}\|^2_{\mathbb{R}^3}$ **(c)** $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,2}\|^2_{\mathbb{R}^3}$ **(d)** $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,3}\|^2_{\mathbb{R}^3}$

Figure 15.11. Time evolution of the optimal state $t_j \mapsto \boldsymbol{m}^{\star,j}_{sto,i}$ (red) and the direction of the optimal control $t_j \mapsto \boldsymbol{u}^{\star,j}_{sto,i}\|\boldsymbol{u}^{\star,j}_{sto,i}\|^{-1}_{\mathbb{R}^3}$ (blue), its magnitude of control $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,i}\|^2_{\mathbb{R}^3}$ (——) and expectation (- - -) for each particle $i = 1, 2, 3$ using Example 15.2 and Setup **D** in the case where exchange is considered ($C_{exch} = 1.0$).



(a) $t_j \mapsto \boldsymbol{m}^{\star,j}_{sto,i}$ and $t_j \mapsto \boldsymbol{u}^{\star,j}_{sto,i} \|\boldsymbol{u}^{\star,j}_{sto,i}\|^{-1}_{\mathbb{R}^3}$

(b) $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,1}\|^2_{\mathbb{R}^3}$ **(c)** $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,2}\|^2_{\mathbb{R}^3}$ **(d)** $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,3}\|^2_{\mathbb{R}^3}$

Figure 15.12. Time evolution of the optimal state $t_j \mapsto \boldsymbol{m}^{\star,j}_{sto,i}$ (red) and the direction of the optimal control $t_j \mapsto \boldsymbol{u}^{\star,j}_{sto,i} \|\boldsymbol{u}^{\star,j}_{sto,i}\|^{-1}_{\mathbb{R}^3}$ (blue), as well as the magnitude of the control $t_j \mapsto \|\boldsymbol{u}^{\star,j}_{sto,i}\|^2_{\mathbb{R}^3}$ (——) and its expectation (- - -) for each particle $i = 1, 2, 3$ using Example 15.2 and Setup **E** in the case where exchange is considered ($C_{exch} = 1.0$).

Bibliography

[AB09] F. Alouges and K. Beauchard, *Magnetization switching on small ferromagnetic ellipsoidal samples*, ESAIM Control Optim. Calc. Var. **15** (2009), no. 3, 676–711, doi:10.1051/cocv:2008047.

[ACLP11] S. Agarwal, G. Carbou, S. Labbe, and C. Prieur, *Control of a network of magnetic ellipsoidal samples*, Math. Control Relat. F. **1** (2011), no. 2, 129–147, doi:10.3934/mcrf.2011.1.129.

[AdBH14] F. Alouges, A. de Bouard, and A. Hocquet, *A semi-discrete scheme for the stochastic Landau-Lifshitz equation*, Stoch. Partial Differ. Equ. Anal. Comput. **2** (2014), no. 3, 281–315, doi:10.1007/s40072-014-0033-7.

[AJ06] F. Alouges and P. Jaisson, *Convergence of a finite element discretization for the Landau-Lifshitz equations in micromagnetism*, Math. Models Methods Appl. Sci. **16** (2006), no. 2, 299–316, doi:10.1142/S0218202506001169.

[AS92] F. Alouges and A. Soyeur, *On global weak solutions for Landau-Lifshitz equations: existence and nonuniqueness*, Nonlinear Anal. **18** (1992), no. 11, 1071–1084, doi:10.1016/0362-546X(92)90196-L.

[BBNP13] L. Baňas, Z. Brzeźniak, M. Neklyudov, and A. Prohl, *Stochastic Ferromagnetism - Analysis and Computation*, vol. 58, De Gruyter Studies in Mathematics, 2013.

[BBNP14] ———, *Convergent Finite Element based discretization of the stochastic Landau-Lifschitz-Gilbert equation*, IMA J. Numer. Anal. **34** (2014), no. 2, 502–549, doi:10.1093/imanum/drt020.

[BBP08] L. Baňas, S. Bartels, and A. Prohl, *A convergent implicit finite element discretization of the Maxwell-Landau-Lifshitz-Gilbert equation*, SIAM J. Numer. Anal. **46** (2008), no. 3, 1399–1422, doi:10.1137/070683064.

[BBP13] L. Baňas, Z. Brzeźniak, and A. Prohl, *Computational Studies for the stochastic Landau-Lifshitz-Gilbert equation*, SIAM J. Sci. Comput. **35** (2013), no. 1, 62–81, doi:10.1137/110856666.

[BD07] C. Bender and R. Denk, *A forward scheme for backward SDEs*, Stochastic Proc. Appl. **117** (2007), 1793–1823, doi:10.1016/j.spa.2007.03.005.

[Ben83] A. Bensoussan, *Stochastic Maximum Principle for Distributed Parameter Systems*, Journal of the Franklin Institute (1983), 387–406, doi:10.1016/0016-0032(83)90059-5.

[BGJ12] Z. Brzeźniak, B. Goldys, and T. Jegaraj, *Large deviations for a stochastic Landau-Lifschitz equation, extended version*, arXiv:1202.0370v1, 2012.

[BGJ13] ———, *Weak Solutions of a stochastic Landau-Lifshitz-Gilbert equation*, Appl. Math. Res. Express. AMRX **1** (2013), no. 1, 1–33, doi:10.1093/amrx/abs009.

[BKM+99] A.R. Bishop, T. Kamppeter, F.G. Mertens, E. Moro, and A. Sánchez, *Stochastic vortex dynamics in two-dimensional easy-plane ferromagnets: Multiplicative versus additive noise*, Physical Review B **59** (1999), no. 17, 11349–11357, doi:10.1103/PhysRevB.59.11349.

[BM10] C. Bender and T. Moseler, *Importance Sampling for Backward SDEs*, Stoch. Anal. Appl. **28** (2010), no. 2, 226–253, doi:10.1080/07362990903546405.

[BMS09] G. Bertotti, I. Mayergoyz, and C. Serpico, *Nonlinear magnetization dynamics in nanosystems*, Elsevier Series in Electromagnetism, Elsevier B.V., 2009.

[BP06] S. Bartels and A. Prohl, *Convergence of an Implicit Finite Element Method for the Landau-Lifshitz-Gilbert equation*, SIAM J. Numer. Anal. **44** (2006), no. 4, 1405–1419, doi:10.1137/050631070.

[Bre73] R.P. Brent, *Algorithms for minimization without derivatives*, 1st ed., Prentice Hall, 1973.

[Bro63] W.F. Brown Jr., *Thermal fluctuations of a single-domain particle*, Physical Review **130** (1963), no. 5, 1677–1686, doi:10.1103/PhysRev.130.1677.

[BSDDM05] M. Barton-Smith, A. Debussche, and L. Di Menza, *Numerical study of two-dimensional stochastic NLS equations*, Numer. Methods Partial Differential Eq. **21** (2005), no. 4, 810–842, doi:10.1002/num.20064.

[BT04] B. Bouchard and N. Touzi, *Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations*, Stochastic Proc. Appl. **111** (2004), 175–206, doi:10.1016/j.spa.2004.01.001.

[BW12] B. Bouchard and X. Warin, *Monte-Carlo valuation of American Options: facts and new algorithms to improve existing methods*, Numerical Methods in Finance, Springer Proc. Math., vol. 12, 2012, pp. 215–255.

[BZ08] C. Bender and J. Zhang, *Time Discretization and Markovian Iteration for Coupled FBSDEs*, Ann. Appl. Probab. **18** (2008), no. 1, 143–177, doi:10.1214/07-AAP448.

[Cho07] P.-L. Chow, *Stochastic Partial Differential Equations*, Chapman & Hall/CRC, 2007.

[Cim08] I. Cimrák, *A survey on the numerics and computations for the Landau-Lifshitz equation of micromagnetism*, Arch. Comput. Methods Eng **15** (2008), no. 3, 277–309, doi:10.1007/s11831-008-9021-2.

[CL93] R.W. Chantrell and A. Lyberatos, *Thermal fluctuations in a pair of magnetostatically coupled particles*, J. Appl. Phys. **73** (1993), no. 10, 6501–6503, doi:10.1063/1.352594.

[CM12] D. Crisan and K. Manolarakis, *Solving Backward Stochastic Differential Equations using the Cubature Method: Application to Nonlinear Pricing*, SIAM J. Financial Math. **3** (2012), 534–571, doi:10.1137/090765766.

[CP12] E. Carelli and A. Prohl, *Rates of convergence for discretisations of the stochastic incompressible Navier-Stokes equations*, SIAM J. Numer. Anal. **50** (2012), no. 5, 2467–2496, doi:10.1137/110845008.

[DBD04] A. De Bouard and A. Debussche, *A semi-discrete scheme for the stochastic nonlinear Schrödinger equation*, Numer. Math. **96** (2004), no. 4, 733–770, doi:10.1007/s00211-003-0494-5.

[DKPS15] T. Dunst, M. Klein, A. Prohl, and A. Schäfer, *Optimal control in evolutionary micromagnetism*, IMA J. Numer. Anal. **35** (2015), no. 3, 1342–1380, doi:10.1093/imanum/dru034.

[DMP96] J. Douglas, J. Ma, and P. Protter, *Numerical Methods for Forward-Backward Stochastic Differential Equations*, Ann. Appl. Probab. **6** (1996), no. 3, 940–968, doi:10.1214/aoap/1034968235.

[DP15] T. Dunst and A. Prohl, *The forward-backward stochastic heat equation: numerical analysis and simulation*, Accepted for publication in SIAM Journal on Scientific Computing, 2015.

[DP16] ———, *Stochastic optimal control of finite ensembles of nanomagnets*, (submitted) https://na.uni-tuebingen.de/pub/dunst/MagnetizationControl.pdf, 2016.

[DPZ92] G. Da Prato and J. Zabczyk, *Stochastic equations in infinite dimensions*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1992.

[DT12] K. Du and S. Tang, *Strong Solution of Backward Stochastic Partial Differential Equations in $\mathcal{C}^2$ Domains*, Probab. Theory. Rel. **154** (2012), 255–285, doi:10.1007/s00440-011-0369-0.

[Dun15] T. Dunst, *Convergence with rates for a time-discretization of the stochastic Landau-Lifschitz-Gilbert equation*, IMA J. Numer. Anal. **35** (2015), no. 2, 615–651, doi:10.1093/imanum/dru005.

[FS10] J.R. Friedman and M.P. Sarachik, *Single-Molecule Nanomagnets*, Annu. Rev. Condens. Matter Phys.. **1** (2010), 109–128, doi:10.1146/annurev-conmatphys-070909-104053.

[GH93] B.L. Guo and M.C. Hong, *The Landau-Lifshitz equation of the ferromagnetic spin chain and harmonic maps*, Calc. Var. Partial Differential Equations **1** (1993), no. 3, 311–334, doi:10.1007/BF01191298.

[Gil55] T.L. Gilbert, *A lagrangian formulation of the gyromagnetic equation of the magnetic field*, Physical Review (1955), 1243–1255.

[Gil04] ———, *A Phenomenological Theory of Damping in Ferromagnetic Materials*, IEEE Transactions on Magnetics **40** (2004), no. 6, 3443–3449, doi:10.1109/TMAG.2004.836740.

[GKKW02] L. Györfi, A. Krzyzak, M. Kohler, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer University Press, 2002.

[GLW05] E. Gobet, J. Lemor, and X. Warin, *A regression-based Monte Carlo Method to solve Backward Stochastic Differential equations*, Ann. Appl. Probab. **15** (2005), no. 3, 2172–2202, doi:10.1214/105051605000000412.

[GPL98] J.L. García-Palacios and F.J. Lázaro, *Langevin-dynamics study of the dynamical properties of small magnetic particles*, Phys. Rev. B **58** (1998), no. 22, 14937–14958, doi:10.1103/PhysRevB.58.14937.

[GT14] E. Gobet and P. Turkedjiev, *Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions*, hal-00642685, version 4, 2014.

[HPUU08] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Contraints*, 1st ed., Spinger University Press, 2008.

[Ich82] A. Ichikawa, *Stability of semilinear stochastic evolution equations*, J. Math. Anal. Appl. **90** (1982), no. 1, 12–44, doi:10.1016/0022-247X(82)90041-5.

[KK91] M. Kroller and K. Kunisch, *Convergence rates for the feedback operators arising in the linear quadratic regulator problem governed by parabolic equations*, SIAM J. Numer. Anal. **28** (1991), no. 5, 1350–1385, doi:10.1137/0728071.

[KP06] M. Kruzik and A. Prohl, *Recent Developments in the Modeling, Analysis, and Numerics of Ferromagnetism*, SIAM Rev. **48** (2006), no. 3, 439–483, doi:10.1137/S0036144504446187.

[LL35] L.D. Landau and E.M. Lifshitz, *On the theory of the dispersion of magnetic permeability of ferromagnetic bodies*, Phyz. Z. Sowj. **8** (1935), 153–169.

[MRT02] G.N. Milstein, Y.M. Repin, and M.V. Tretyakov, *Numerical methods for stochastic systems preserving symplectic structure*, SIAM J. Numer. Anal. **20** (2002), no. 4, 1583–1604, doi:10.1137/S0036142901395588.

[MT06] G. Milstein and M. Tretyakov, *Numerical Algorithms for Forward-Backward Stochastic Differential Equations*, SIAM J. Sci. Comput. **28** (2006), no. 2, 561–582, doi:10.1137/040614426.

[MTF+10] J.H. Mentink, M.V. Tretyakov, A. Fasolino, M.I. Katsnelson, and T. Rasing, *Stable and fast semi-implicit integration of the stochastic Landau-Lifshitz equation*, J. Phys.: Condens. Matter **22** (2010), 12pp, doi:10.1088/0953-8984/22/17/176001.

[MY07] J. Ma and J. Yong, *Forward-Backward Stochastic Differential Equations and their Applications*, 1st ed., Spinger University Press, 2007.

[Née46] L. Néel, *Bases d'une nouvelle théorie générale du champ coercitif*, Annales de l'université de Grenoble **22** (1946), 299–343.

[NP13] M. Neklyudov and A. Prohl, *The role of noise in finite ensembles of nanomagnetic particles*, Arch. Ration. Mech. Anal. **210** (2013), no. 2, 499–534, doi:10.1007/s00205-013-0654-4.

[Pri01] J. Printems, *On the discretization in time of parabolic stochastic partial differential equations*, M2AM **35** (2001), no. 6, 1055–1078, doi:10.1051/m2an:2001148.

[Pro01] A. Prohl, *Computational Micromagnetism*, 1st ed., Vieweg+Teubner Verlag, 2001.

[YZ99] J. Yong and X. Zhou, *Stochastic Controls. Hamiltonian Systems and HJB Equations*, 2nd ed., Spinger University Press, 1999.

[Zha04] J. Zhang, *A numerical scheme for BSDEs*, Ann. Appl. Probab. **14** (2004), 459–488, doi:10.1214/aoap/1075828058.

[Zho93] X.Y. Zhou, *On the Necessary Conditions of Optimal Control for Stochastic Partial Differential Equations*, SIAM J. Control Optim. **31** (1993), no. 6, 1462–1478, doi:10.1137/0331068.