Limiting Behaviors of High Dimensional Stochastic Spin Ensemble

Y. Gao[1], K. Kirkpatrick[2], J. Marzuola[1], J. Mattingly[3], and K. Newhall[1]

[1]Department of Mathematics, University of North Carolina at Chapel Hill
[2]Department of Mathematics, University of Illinois at Urbana-Champaign
[3]Department of Mathematics, Duke University

Abstract

Lattice spin models in statistical physics are used to understand magnetism. Their Hamiltonians are a discrete form of a version of a Dirichlet energy, signifying a relationship to the Harmonic map heat flow equation. The Gibbs distribution, defined with this Hamiltonian, is used in the Metropolis-Hastings (M-H) algorithm to generate dynamics tending towards an equilibrium state. In the limiting situation when the inverse temperature is large, we establish the relationship between the discrete M-H dynamics and the continuous Harmonic map heat flow associated with the Hamiltonian. We show the convergence of the M-H dynamics to the Harmonic map heat flow equation in two steps: First, with fixed lattice size and proper choice of proposal size in one M-H step, the M-H dynamics acts as gradient descent and will be shown to converge to a system of Langevin stochastic differential equations (SDE). Second, with proper scaling of the inverse temperature in the Gibbs distribution and taking the lattice size to infinity, it will be shown that this SDE system converges to the deterministic Harmonic map heat flow equation. Our results are not unexpected, but show remarkable connections between the M-H steps and the SDE Stratonovich formulation, as well as reveal trajectory-wise out of equilibrium dynamics to be related to a canonical PDE system with geometric constraints.

Contents

1 Introduction

The Metropolis-Hastings (M-H) algorithm [13] is widely used in particle statistics for model estimations [24, 4, 21, 3, 22]. It constructs a discrete-time Markov chain to sample a desired probability distribution by accepting or rejecting proposed states. For applications in statistical physics, it is often the Gibbs or canonical distribution that is to be sampled. In this case, the algorithm accepts all the proposed new states with lower energy and often rejects the proposals with higher energy. Similar behavior can be obtained from a Langevin Stochastic differential equation (SDE) that performs gradient descent with noise; it too has the Gibbs distribution as its steady-state distribution. This suggests that the Langevin SDE might be the optimal M-H algorithm in which all proposals are accepted.

For certain forms of probability distributions, the diffusion limit and therefore optimal scaling, of the random walk M-H algorithm has been obtained in [27, 5, 23]. Specifically, for product measures in [27] and the Gibbs distribution of a lattice model in [5], the weak convergence to Langevin diffusions have been shown by comparing generator functions. For non-product form measures the weak convergence to a stochastic partial differential equation was shown in [23]. These works consider the weak convergence only in equilibrium. Subsequent works [15, 14] consider scaling limits of out of equilibrium systems approaching equilibrium.

To address the question of trajectory-wise convergence, we study the XY and the classical Heisenberg lattice spin models [28] that play an important role in statistical physics to understand phase transitions and other phenomena including superconductivity [20, 7]. The XY and classical Heisenberg models are defined on a periodic d-dimensional lattice $\mathbb{T}^d$ with $\delta x = \frac{1}{N}$ the distance between adjacent vertices. Each spin sits at a lattice point and is described by a unit vector $\sigma_i : \mathbb{T}^d \to \mathbb{S}^n$, where $n = 1$ for the XY model and $n = 2$ for the classical Heisenberg model. The Hamiltonian of the system,

$$H = J \sum_{<i,j>} \|\sigma_i - \sigma_j\|^2, \tag{1}$$

gives energy to misaligned neighboring spins where $< i, j >$ represents nearest neighbors and $J = N^{2-d}$ is a scaling factor. Denote σ as the total spin configuration of $\sigma_i, i \in \mathbb{T}^d$, the M-H algorithm accepts/rejects based on the Gibbs distribution defined as

$$\rho(\sigma) = Z^{-1} \exp(-\beta H(\sigma)), \tag{2}$$

where $\beta = (k_B T)^{-1}$ is the inverse temperature and Z is the normalizing factor (aka partition function). The distribution is unaware of the confining geometry that the spins must remain in $\mathbb{S}^n$. Rather, it is included in the proposal step of the M-H algorithm. As the nearest neighbor coupling is encoded in H, it follows naturally from our analysis here that with no accept/reject step each spin behaves like a Brownian motion on the surface of $\mathbb{S}^n$. It follows similarly from our analysis in the proof. Since the XY model and the classical Heisenberg model are widely used to study superconductors and ferromagnets, their critical properties are of interest. Asymptotic results on the total spin of the mean-field XY and classical Heisenberg models have been studied by large deviation theory and Stein's method in [17, 18]. Numerically, Monte Carlo methods are used to verify analytical results about XY model in [22, 3] and classical Heisenberg model in [26, 6].

We will show the M-H algorithm applied to the above lattice system produces equivalent trajectories to the overdamped Langevin equation,

$$\mathrm{d}\sigma_i = \mathrm{P}^{\perp}_{\sigma_i}(\Delta_N \sigma_i)\mathrm{d}t + \mathrm{P}^{\perp}_{\sigma_i}\left(\sqrt{\frac{N}{\beta}}\mathrm{d}W_i\right), \tag{3}$$

(interpreted in the Stratonovich sense) in the limit of small perturbations to create the proposal where

$$\Delta_N \sigma_i = -N^2(2\sigma_i - \sigma_{i+1} - \sigma_{i-1}),$$

is the discrete Laplacian and $\mathrm{P}_x^\perp(y) = y - (x \cdot y)x$ for $\|x\| = 1$ is the projection of y onto the tangent plane of x. The Stratonovich understanding of (3) is essential to keep the σ_i as unit vectors, and for more on this equation see [1]. Under the Itô understanding, an Itô correction term will drop out in (3) and show naturally in the proof in section 3. This system is the Langevin system that performs gradient descent on the energy defined by (1) with the added constraint that σ_i is confined to $\mathbb{S}^n, n = 1, 2$. In $\mathbb{S}^2$, the classical Heisenberg model, this system is an SDE representation of the overdamped Landau-Lifshitz-Gilbert equation that has the Gibbs distribution as its invariant measure [1, 19].

Taking the number of lattice points, N, to infinity or equivalently the lattice spacing $\delta x = \frac{1}{N}$ to zero, the limit of the deterministic part of (3) is the partial differential equation (PDE) called the harmonic map heat flow equation

$$\partial_t \sigma = \mathrm{P}_\sigma^\perp(\Delta\sigma). \tag{4}$$

In the $\mathbb{S}^2$ case, (4) is in the form of the overdamped Landau-Lifshitz equation [9]

$$\partial_t \sigma = -\sigma \times (\sigma \times \Delta\sigma). \tag{5}$$

In [10] this Landau-Lifshitz equation was shown to be equivalent to the Harmonic map heat flow from $\mathbb{T}^d \to \mathbb{S}^2$. With the scaling $J = N^{2-d}$, the Hamiltonian in (1) is the discrete form of the Dirichlet energy, $\int_\Omega |\nabla\sigma|^2 \mathrm{d}\Omega$, for this harmonic map heat flow. This suggests that by decreasing the temperature, the out of equilibrium dynamics of the M-H algorithm converge to the deterministic flow of (5) with large N for the classical Heisenberg model. We will show this equivalence by showing the convergence of the system of SDE (3) to the PDE (4) in the limit of large N with an appropriate scaling of the temperature to zero with N.

One method to obtain the deterministic limit of a stochastic system is to consider the hydrodynamic limit with relative entropy bound [11, 30, 8]. Due to the geometric constraint in XY and classical Heisenberg model, it is difficult to calculate the averages with respect to the Gibbs states as in [11, 30, 8] if the spin is expressed in Cartesian coordinates. One might try to use polar coordinates to do window averaging but the potential is not convex as in [8]. Since the hydrodynamic limit for XY and classical Heisenberg model are not fully understood, we choose an alternative way of taking inverse temperature β to infinity along with particle number $N \to \infty$.

One difficulty in the proof comes from the constraint of the spins staying as unit vectors. This requires a normalising step in the M-H algorithm and makes the calculation complicated. We take the Taylor expansion of the M-H step and approximate it as a linear step. This truncation of the spin vector does not stay on the sphere but the error for the subsequent steps is shown to converge in the limit as $N \to \infty$ with our system size dependent choice of parameters. Moreover, in the weak convergence result of M-H dynamics to diffusion process [27, 5, 23], the assumption of equilibrium is essential to bound the error terms. The result here only assumes that we are starting the M-H dynamics (and thus the SDE system) from a deterministic initial condition satisfying a certain regularity condition and then evolving into equilibrium. To bound the error terms, the scaling chosen here is worse than in the previously mentioned papers and is likely not be optimal. We will use numerical simulations to explore how tight these bounds appear to be.

The remainder of the paper is as follows. In Section 2 we present the main results in two parts. First, the convergence of M-H dynamics to the SDE system (3) as the proposal size of M-H step goes to zero is stated, then the convergence of the SDE system (3) to the deterministic PDE (4) as the lattice size goes to infinity and temperature to zero is stated. The key steps of the proof are given in Sections 3 and 4 for the more complicated classical Heisenberg model from $\mathbb{T}^1 \to \mathbb{S}^2$ with details appearing in the Appendix. The proof for XY model follows similarly. For the M-H to SDE (3) proof in Section 3, we apply a similar approach as in [23], by first Taylor expanding the M-H step, keeping only the first three terms, then computing the required conditional expectations with respect to the Gaussian random variables to obtain the drift and diffusion terms of an Euler step for the diffusion process. Then, the difference between the M-H and SDE dynamics in L^2 norm is bounded by a Grönwall inequality. For the SDE (3) to PDE (5) proof in Section 4, we compare the SDE system with the finite difference approximation of the Landau-Lifshitz equation. The difference between the SDE and ODE system is governed by another diffusion process. We will rescale this process and show the rescaled error is bounded for a long time using stopping time. These convergence results are supported by numerical simulations of the systems in Section 5. Conclusions are presented in Section 6.

Remark 1.1. *We only show the case $\mathbb{T}^1 \to \mathbb{S}^2$. The calculation could be generalised for other cases of $\mathbb{T}^d \to \mathbb{S}^2$ quite similarly.*

Acknowledgements

Y.G. and J.L.M were supported in part by NSF Applied Math Grant DMS-1312874 and NSF CAREER Grant DMS-1352353. K.K. was supported in part by NSF CAREER Award DMS-1254791 and a Simons Foundation Fellowship. J.D.M. was partially supported by DMS-1613337 from the National Science Foundation and a Simons Travel grant. The authors thank Jonathan Weare and Jianfeng Lu for helpful conversations during the preparation of this work. Also, we thank MSRI for hosting three of the authors during Fall 2015.

2 Main Results

In this section we will explain how we apply M-H algorithm to the XY and classical Heisenberg models, and state our main results. Our first result is that the M-H dynamics is close to a stochastic Euler scheme for the SDE (3) in Itô understanding. The bound on the error between the M-H dynamics and the SDE (3) is accomplished using arguments similar to the convergence of the stochastic Euler method. Our second result bounds the error between the SDE system and the finite difference approximation of the harmonic map heat flow equation (4).

2.1 Metropolis-Hastings step

Here, we explicitly state the M-H dynamics for XY and classical Heisenberg models we consider with Hamiltonian given by (1) for the case $d = 1$.

Consider a set of spins evolving in time, σ_i^n for particle $i = 1 \dots N$ and time step $n \geq 0$ with time step size δt. To create the proposal, take the normal random vector

$$w_i^n = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \quad \text{with } z_1, z_2 \sim \mathcal{N}(0,1)$$

for the XY model and three-dimensional normal random vector

$$w_i^n = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}, \quad \text{with } z_1, z_2, z_3 \sim \mathcal{N}(0,1),$$

for the classical Heisenberg model. Then project to the tangent plane of σ_i^n to get the random vector $\nu_i^n = P^{\perp}_{\sigma_i^n}(w_i^n) = w_i^n - (w_i^n, \sigma_i^n)\sigma_i^n$. Since we are trying to get a trajectory-wise convergence result, it is convient to imbed the M-H algorithm and the SDE dynamics in the same probability space. To this end, we define

$$w_i^n \equiv \frac{W_i((n+1)\delta t) - W_i(n\delta t)}{\sqrt{\delta t}},$$

where $W_i, 1 \leq i \leq N$ are the Brownian motion in (3). At time step n the proposal for next time step is

$$\tilde{\sigma}_i^n = \exp_{\sigma_i^n}(\varepsilon \nu_i^n), \quad 1 \leq i \leq N, \tag{6}$$

with $\exp_{\sigma_i^n}$ the exponential map and ε the proposal size. The values σ^n and $\tilde{\sigma}^n$ are used to denote the total spin configuration $\sigma_i^n, 1 \leq i \leq N$ at time step n and the total proposal spin configuration $\tilde{\sigma}_i^n, 1 \leq i \leq N$. The proposal $\tilde{\sigma}^n$ is accepted with probability

$$\alpha = 1 \wedge e^{-\beta \delta H}, \tag{7}$$

and rejected otherwise, where

$$\delta H = H(\tilde{\sigma}^n) - H(\sigma^n) = \sum_{j=1}^{N} \frac{\partial H}{\partial \sigma_j^n} \cdot (\tilde{\sigma}_j^n - \sigma_j^n) + 4J \sum_{j=1}^{N} (\tilde{\sigma}_j^n - \sigma_j^n) \cdot (\tilde{\sigma}_j^n - \sigma_j^n) - 2J \sum_{j=1}^{N} (\tilde{\sigma}_j^n - \sigma_j^n) \cdot (\tilde{\sigma}_{j+1}^n - \sigma_{j+1}^n) \tag{8}$$

is the difference between the Hamiltonian (1) of the proposal $\tilde{\sigma}^n$ and of the current spin configuration σ^n. Then

$$\sigma^{n+1} = \kappa_n \tilde{\sigma}^n + (1 - \kappa_n)\sigma^n, \quad \kappa_n \sim \text{Bernoulli}(\alpha(\tilde{\sigma}^n, \sigma^n)).$$

Repeating this step, we create a discrete Markov process at time steps $n+1, n+2, \ldots$ and we will show the convergence of the Markov chain to the solution to the Langevin SDE system (3).

Remark 2.1. *In fact, either choice of the following projection gives us the same result for the classical Heisenberg model*

$$P^{\perp}_{\sigma_i^n}(w_i^n) = \begin{cases} \sigma_i^n \times w_i^n \\ -\sigma_i^n \times (\sigma_i^n \times w_i^n) = w_i^n - \sigma_i^n (\sigma_i^n)^T w_i^n \end{cases}$$

as both lead to random walk on the sphere (see Appendix C).

2.2 Convergence of Metropolis dynamics to SDE system

First we are going to show the convergence from M-H dynamics to Langevin SDEs with fixed number of particles N as the proposal size $\varepsilon \to 0$. Intuitively, using the Taylor series truncation of the proposal, the approximation of one M-H step leads to an expression that looks like one Euler step for simulating the SDE (3) in Itô sense.

Let $\mathcal{F}_t$ denote the filtration generated by the Brownian motion W_i in (3) and Bernoulli random variables κ_n at $n\delta t$, we denote the conditional expectation $\mathbb{E}\left[\cdot|\mathcal{F}_{n\delta t}\right]$ by $\mathbb{E}_n\left[\cdot\right]$.

The drift over one step of the Metropolis-Hastings algorithm for the i-th particle for small ε is approximated by

$$\mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] \approx -\frac{1}{2}\beta\varepsilon^2 \mathrm{P}^{\perp}_{\sigma_i^n}\left(\frac{\partial H}{\partial \sigma_i^n}\right) - \varepsilon^2 \sigma_i^n, \tag{9}$$

where $\mathrm{P}^{\perp}_{\sigma_i^n} = I - \sigma_i^n(\sigma_i^n)^T$ is the projection onto the tangent plane of σ_i^n.

Denoting the noise contribution over one step as

$$\Gamma_i^n \equiv \sigma_i^{n+1} - \sigma_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right], \tag{10}$$

it is approximated by

$$\Gamma_i^n \approx \varepsilon \nu_i^n = \varepsilon \mathrm{P}^{\perp}_{\sigma_i^n}(w_i^n). \tag{11}$$

Thus, one step of the Metropolis-Hastings algorithm is approximately given by

$$\sigma_i^{n+1} - \sigma_i^n \approx -\frac{1}{2}\beta\varepsilon^2 \mathrm{P}^{\perp}_{\sigma_i^n}\left(\frac{\partial H}{\partial \sigma_i^n}\right) - \varepsilon^2\sigma_i^n + \mathrm{P}^{\perp}_{\sigma_i^n}(\varepsilon w_i^n). \tag{12}$$

Defining $\beta\varepsilon^2 = N\delta t$ where δt is the time step size, the above equation changes to

$$\sigma_i^{n+1} \approx \sigma_i^n - \frac{1}{2}N\mathrm{P}^{\perp}_{\sigma_i^n}\left(\frac{\partial H}{\partial \sigma_i^n}\right)\delta t - \frac{N}{\beta}\sigma_i^n\delta t + \mathrm{P}^{\perp}_{\sigma_i^n}\left(\sqrt{\frac{N}{\beta}}w_i^n\sqrt{\delta t}\right). \tag{13}$$

Since $\frac{\partial H}{\partial \sigma_i^n} = 2J(2\sigma_i^n - \sigma_{i+1}^n - \sigma_{i-1}^n)$ and $J = N$ when $d = 1$, the above is the Euler step for the Langevin SDE (3) in Itô interpretation

$$\mathrm{d}\sigma_i = \mathrm{P}^{\perp}_{\sigma_i}\left(\Delta_N \sigma_i\right)\mathrm{d}t - \frac{N}{\beta}\sigma_i \mathrm{d}t + \mathrm{P}^{\perp}_{\sigma_i}\left(\sqrt{\frac{N}{\beta}}\mathrm{d}W_i\right). \tag{14}$$

This intuitive idea leads to the first result:

Theorem 2.1. *Define the piecewise constant interpolation of M-H dynamics as $\bar{\sigma}_i(t)$,*

$$\bar{\sigma}_i(t) = \sigma_i^n \quad n\delta t \le t < (n+1)\delta t, \tag{15}$$

and $\sigma_i(t)$ as the solution for the Langevin SDE system (14) with initial condition $\|\sigma_i(0)\| = 1, 1 \le i \le N$. If we think of the proposal in M-H step coming from the noise $\varepsilon w_i^n = \sqrt{N\beta^{-1}}\left[W_i((n+1)\delta t) - W_i(n\delta t)\right]$, then we have the following strong convergence result:

$$\mathbb{E}\left[\sup_{0\le s\le t}\|\sigma_i(s) - \bar{\sigma}_i(s)\|^2\right] \le C_1\sqrt{\delta t}\exp(C_2 T), \quad t \in [0,T], 1 \le i \le N, \tag{16}$$

for any $T \in (0,\infty)$, where C_1, C_2 are functions of N, β, J, T and independent of the choice of i and δt.

Remark 2.2. *The equation* (14) *is equivalent to the SDE in Stratonovich sense* (3) *which gives* $d\|\sigma_i\|^2 = 2\sigma_i \cdot d\sigma_i = 0$ *to make* σ_i *stay on the unit sphere.*

Remark 2.3. *Theorem 2.1 is a trajectory-wise convergence result.*

2.3 Convergence of SDE system to the Landau-Lifshitz equation

Notice in the SDE (14), if β is chosen to be $\beta = N^\gamma, \gamma > 1$, formally the noise part disappears with $N \to \infty$. This gives the idea of the second result:

Theorem 2.2. *For the harmonic map heat flow equation* (4) *with periodic boundary condition and initial condition satisfing*

$$\|\sigma(\cdot, 0)\| = 1, \|\nabla\sigma(\cdot, 0)\| \leq \lambda, \tag{17}$$

for some λ *as in [10], the solution exists and is smooth. Denote the finite difference approximation of* (4) *as*

$$d\tilde{\sigma}_i = P^{\perp}_{\tilde{\sigma}_i}(\Delta_N \tilde{\sigma}_i), \quad \|\tilde{\sigma}_i\| = 1 \tag{18}$$

and $\|\tilde{\sigma}_i(t) - \sigma(i\delta x, t)\| \to 0$ *on any fixed time interval where the solution remains well defined, when the space discretization* $\delta x = \frac{1}{N}$ *goes to zero [29, Theorem 1].*

For any $0 < p < \frac{1}{2}$, *there exist a constant* $\gamma > 1, \beta = N^\gamma$ *and constants* C_1, C_2 *independent of* N, *such that if*

$$\left(\left(\frac{N}{\beta}\right)^{1-p} T + C_1 \frac{1}{N}\left(\frac{N}{\beta}\right)^{1-2p}\right) e^{C_2 T} \leq 1,$$

then the difference between the SDE (14) *and the finite difference approximation* (18) *has the following bound*

$$\mathbb{E}\left[\sup_{0\leq s\leq t} \frac{1}{N}\sum_{i=1}^{N} \|\sigma_i(s) - \tilde{\sigma}_i(s)\|^2\right] \leq \left(\frac{N}{\beta}\right)^{p/2}, \quad t \in [0, T]. \tag{19}$$

As $\gamma > 1$, $\frac{N}{\beta}$ is small when N is large, so the difference between the SDE system and finite difference approximation of the PDE is bounded by a small term for a long time T that goes to ∞ with $N \to \infty$. The solution for PDE is smooth so the finite difference approximation is close to the PDE solution as shown in [29].

Remark 2.4. *The choice of* γ *depends on* p *with the following relation*

$$\left(\frac{N}{\beta}\right)^{p/2} N^3 \leq 1.$$

For a uniform bound in $0 \leq t \leq T$, *we need* $p < \frac{1}{2}$ *so* $\gamma > 13$. *For a bound with some fixed* $t \in [0, T]$, *we only need* $p < 1$ *and* $\gamma > 7$. *We do not believe this bound is sharp for the convergence result at all, which will be addressed in Section 5 when we perform numerical simulations of these models. We find that* $\gamma = \frac{3}{2}$ *is enough to see convergence in our numerical simulations.*

3 Metropolis-Hastings dynamics to SDE system

In this section the convergence of the M-H algorithm to the SDE (14) for the classical Heisenberg model will be shown by calculating the drift and diffusion of one M-H step, which is approximately a stochastic Euler step for (14). Then the error estimation of stochastic Euler's method is used to give a bound on the difference between M-H and SDE dynamics with proposal size $\varepsilon \to 0$. Here the basic steps are outlined, the detail of error estimation is given in Appendix A.

Remark 3.1. *The proof for the XY model will be similar, one only needs to change the random vector* ν_i^n *on the tangent plane as a two-dimensional vector.*

3.1 Set-up

In the calculation to follow, we have the following assumptions and notations.

The number of the particles N on unit length is fixed and the limiting case $\varepsilon \to 0$ is considered. We have β, J as functions of N so they are also regarded as constant.

In the calculation, the proposal $\tilde{\sigma}_i^n$ is approximated by normalizing $\sigma_i^n + \varepsilon\nu_i^n$

$$\tilde{\sigma}_i^n = \exp_{\sigma_i^n}(\varepsilon\nu_i^n) \approx \frac{\sigma_i^n + \varepsilon\nu_i^n}{\|\sigma_i^n + \varepsilon\nu_i^n\|}.$$

By Taylor expanding $\frac{\sigma_i^n + \varepsilon\nu_i^n}{\|\sigma_i^n + \varepsilon\nu_i^n\|}$, the proposal $\tilde{\sigma}_i^n$ can be approximated by order ε and ε^2 expansion

$$\begin{aligned} \tilde{\sigma}_i^n &\approx \sigma_i^n + \varepsilon\nu_i^n, \\ \tilde{\sigma}_i^n &\approx \sigma_i^n + \varepsilon\nu_i^n - \frac{1}{2}\varepsilon^2(\nu_i^n \cdot \nu_i^n)\sigma_i^n. \end{aligned} \tag{20}$$

The proof of the following Lemma is shown in Appendix A.

Lemma 3.1. *Denote*

$$\begin{aligned} a_i^n &\equiv \tilde{\sigma}_i^n - \frac{\sigma_i^n + \varepsilon\nu_i^n}{\|\sigma_i^n + \varepsilon\nu_i^n\|}, \\ c_i^n &\equiv \tilde{\sigma}_i^n - (\sigma_i^n + \varepsilon\nu_i^n), \\ d_i^n &\equiv \tilde{\sigma}_i^n - \left(\sigma_i^n + \varepsilon\nu_i^n - \frac{1}{2}\varepsilon^2(\nu_i^n \cdot \nu_i^n)\sigma_i^n\right). \end{aligned}$$

Then $\mathbb{E}\left[\|a_i^n\|^k\right] \le A_k\varepsilon^{3k}, \mathbb{E}\left[\|c_i^n\|^k\right] \le C_k\varepsilon^{2k}$ *and* $\mathbb{E}\left[\|d_i^n\|^k\right] \le D_k\varepsilon^{3k}$.

Using the approximation (20), δH in (8) can be written as

$$\begin{aligned} \delta H &= \varepsilon\frac{\partial H}{\partial \sigma_i^n} \cdot \nu_i^n + R_i^n + h_i^n \approx O(\varepsilon), \\ R_i^n &\equiv \varepsilon\sum_{j\neq i}\frac{\partial H}{\partial \sigma_j^n} \cdot \nu_j^n \approx O(\varepsilon), \\ h_i^n &\equiv \sum_j \frac{\partial H}{\partial \sigma_j^n} \cdot c_j^n + 4J\sum_j \delta\sigma_j^n \cdot \delta\sigma_j^n - 2J\sum_j \delta\sigma_j^n \cdot \delta\sigma_{j+1}^n \approx O(\varepsilon^2), \end{aligned} \tag{21}$$

and we only keep the ε term in δH in the following calculation so δH is approximated by a normal random variable. We are going to show the calculation for one specific particle i so we take i-th term $\frac{\partial H}{\partial \sigma_i^n} \cdot \nu_i^n$ and the summation of $j \neq i$ terms as a single term R_i^n.

3.2 Drift

Proposition 3.1. *Let $\{\sigma^n\}$ be the Markov chain given by the Metropolis-Hastings algorithm, and $\{\sigma_i^n\}$ the spin for i-th particle at time step n. Then*

$$\mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] = -\frac{1}{2}\beta\varepsilon^2 P_{\sigma_i^n}^{\perp}\left(\frac{\partial H}{\partial \sigma_i^n}\right) - \varepsilon^2\sigma_i^n + \theta_i^n, \tag{22}$$

where the error term

$$\theta_i^n \equiv \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] - \left(-\frac{1}{2}\beta\varepsilon^2 P_{\sigma_i^n}^{\perp}\left(\frac{\partial H}{\partial \sigma_i^n}\right) - \varepsilon^2\sigma_i^n\right) \tag{23}$$

satisfies $\mathbb{E}\left[\|\theta_i^n\|^2\right] \le C\varepsilon^6$.

In the calculation we keep the order ε^2 term. The remainder is order ε^3 and will be shown to be bounded in the error estimation for M-H and SDE dynamics. The basic steps are given in the following calcuation, for details of the error estimation see Appendix A.

Since $\sigma_i^{n+1} = \exp_{\sigma_i^n}(\varepsilon\nu_i^n)$ with probability $1 \wedge e^{-\beta\delta H}$ and stay σ_i^n otherwise,

$$\begin{aligned}\mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] &= \mathbb{E}_n\left[\left(\exp_{\sigma_i^n}(\varepsilon\nu_i^n) - \sigma_i^n\right)\left(1 \wedge e^{-\beta\delta H}\right)\right] \\ &\approx \mathbb{E}_n\left[\left(\varepsilon\nu_i^n - \frac{\varepsilon^2}{2}(\nu_i^n\cdot\nu_i^n)\sigma_i^n + d_i^n\right)\left(1 \wedge e^{-\beta\delta H}\right)\right] \\ &= \varepsilon\mathbb{E}_n\left[\nu_i^n\left(1 \wedge e^{-\beta\delta H}\right)\right] - \frac{\varepsilon^2}{2}\mathbb{E}_n\left[(\nu_i^n\cdot\nu_i^n)\sigma_i^n\left(1 \wedge e^{-\beta\delta H}\right)\right] + \mathbb{E}_n\left[d_i^n\left(1 \wedge e^{-\beta\delta H}\right)\right].\end{aligned} \tag{24}$$

We drop the third term in the last line of (24) as it is an ε^3 term:

$$\mathbb{E}\left[\left\|d_i^n\left(1 \wedge e^{-\beta\delta H}\right)\right\|\right] \leq \mathbb{E}\left[\|d_i^n\|\right] \leq C\varepsilon^3,$$

since $0 < |1 \wedge e^{-\beta\delta H}| < 1$.

For the second term in the last line of (24), since $1 \wedge e^{-\beta\delta H} \approx 1 + O(\varepsilon)$ we have that

$$\frac{\varepsilon^2}{2}\mathbb{E}_n\left[(\nu_i^n\cdot\nu_i^n)\sigma_i^n\left(1 \wedge e^{-\beta\delta H}\right)\right] = \mathbb{E}_n\left[\frac{\varepsilon^2}{2}(\nu_i^n\cdot\nu_i^n)\sigma_i^n\right] + O(\varepsilon^3) = \varepsilon^2\sigma_i^n + O(\varepsilon^3).$$

This corresponds to the Itô correction for (3).

The first term in the last line of (24) is the most difficult one to approximate. Using the notation in (21)

$$1 \wedge e^{-\beta\delta H} = 1 \wedge e^{-\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n + R_i^n + h_i^n\right)} \approx 1 \wedge e^{-\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n + R_i^n\right)} + O(\varepsilon^2),$$

since $h_i^n \approx O(\varepsilon^2)$, to write it as

$$\mathbb{E}_n\left[\varepsilon\nu_i^n\left(1 \wedge e^{-\beta\delta H}\right)\right] = \mathbb{E}_n\left[\varepsilon\nu_i^n\left(1 \wedge e^{-\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n + R_i^n\right)}\right)\right] + O(\varepsilon^3).$$

For any orthonormal basis $\{b_1, b_2, b_3\}$ in $\mathbb{R}^3$, the normal random vector w_i^n can be expressed as

$$w_i^n = (w_i^n\cdot b_1)b_1 + (w_i^n\cdot b_2)b_2 + (w_i^n\cdot b_3)b_3$$

and $(w_i^n\cdot b_1), (w_i^n\cdot b_2), (w_i^n\cdot b_3)$ are independent standard normal random variables. Denote $r_1 = (w_i^n\cdot b_1), r_2 = (w_i^n\cdot b_2), r_3 = (w_i^n\cdot b_3)$,$w_i^n = r_1b_1 + r_2b_2 + r_3b_3$. Choose b_1, b_2 two orthonormal vectors on the tangent plane of σ_i^n and $b_3 = \sigma_i^n$,

$$\nu_i^n = \mathrm{P}_{\sigma_i^n}(w_i^n) = r_1b_1 + r_2b_2,$$

where $r_1, r_2 \sim \mathcal{N}(0,1)$ are independent. Then,

$$\begin{aligned}\mathbb{E}_n\left[\varepsilon\nu_i^n\left(1 \wedge e^{-\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n + R_i^n\right)}\right)\right] =& \mathbb{E}_n\left[\varepsilon r_1b_1\left(1 \wedge e^{-\beta\left(\varepsilon r_1\frac{\partial H}{\partial\sigma_i^n}\cdot b_1 + \varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2 + R_i^n\right)}\right)\right] \\ &+ \mathbb{E}_n\left[\varepsilon r_2b_2\left(1 \wedge e^{-\beta\left(\varepsilon r_1\frac{\partial H}{\partial\sigma_i^n}\cdot b_1 + \varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2 + R_i^n\right)}\right)\right].\end{aligned} \tag{25}$$

The two terms on the right are similar in form so we only show the calculation for the first one and the second one follows similarly.

Remark 3.2. *For the XY model, the projection of the normal random vector onto the tangent plane of σ_i^n is represented by the form r_1b_1, where $r_1 \sim N(0,1)$. The other parts of the calculation basically stays the same.*

Using tower property of conditional expectation for the first term on the RHS of (25), we have

$$\begin{aligned}
&\mathbb{E}_n\left[\varepsilon r_1 b_1\left(1\wedge e^{-\beta\left(\varepsilon r_1\frac{\partial H}{\partial\sigma_i^n}\cdot b_1+\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n\right)}\right)\right]\\
&\qquad=\mathbb{E}_n\left\{\mathbb{E}_n\left[\varepsilon r_1 b_1\left(1\wedge e^{-\beta\left(\varepsilon r_1\frac{\partial H}{\partial\sigma_i^n}\cdot b_1+\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n\right)}\right)\Big|r_2,R_i^n\right]\right\}.
\end{aligned}$$

We recall the following Lemma 2.4 in [23]. (See also [27].)

Lemma 3.2. *For $z\sim\mathcal{N}(0,1)$,*

$$\mathbb{E}\left[z\left(1\wedge e^{az+b}\right)\right]=ae^{\frac{a^2}{2}+b}\Phi\left(-\frac{b}{|a|}-|a|\right),\tag{26}$$

for any real constants a,b, and $\Phi(\cdot)$ is the CDF for the standard normal random variable.

The proof of this Lemma is the direct result of the integration for the expectation. And the Lemma gives

$$\begin{aligned}
&\mathbb{E}_n\left[\varepsilon r_1 b_1\left(1\wedge e^{-\beta\left(\varepsilon r_1\frac{\partial H}{\partial\sigma_i^n}\cdot b_1+\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n\right)}\right)\Big|r_2,R_i^n\right]\\
&\quad=-\beta\varepsilon^2\left(\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)b_1 e^{\frac{\left(\beta\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)^2}{2}-\beta\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2-\beta R_i^n}\Phi\left(\frac{\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n}{\left|\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|}-\left|\beta\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|\right).
\end{aligned}\tag{27}$$

Before taking the expectation over r_2, we further simplify this expression by noting that $e^{O(\varepsilon)}=1+O(\varepsilon)$ resulting in

$$\begin{aligned}
&\mathbb{E}_n\left[\varepsilon r_1 b_1\left(1\wedge e^{-\beta\left(\varepsilon r_1\frac{\partial H}{\partial\sigma_i^n}\cdot b_1+\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n\right)}\right)\Big|r_2,R_i^n\right]\\
&\quad\approx\left(-\beta\varepsilon^2\left(\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)b_1+O(\varepsilon^3)\right)\left[\Phi\left(\frac{\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n}{\left|\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|}\right)+O(\varepsilon)\right].
\end{aligned}\tag{28}$$

For a mean zero Gaussian random variable z, we know

$$\mathbb{E}\left[\Phi(z)\right]=\mathbb{E}\left[\Phi(z)-\frac{1}{2}+\frac{1}{2}\right]=\int_{-\infty}^{\infty}\left(\Phi(z)-\frac{1}{2}+\frac{1}{2}\right)p(z)\mathrm{d}z=\frac{1}{2},$$

as $\Phi(z)-\frac{1}{2}$ is an odd function and the probability density function $p(z)$ is even.

Notice that $R_i^n=\varepsilon\sum_{j\neq i}\frac{\partial H}{\partial\sigma_j^n}\cdot\nu_j^n$ is a sum of independent mean zero Gaussian random variables, so $\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n$ is a Gaussian random variable with mean 0, therefore

$$\mathbb{E}_n\left[-\beta\varepsilon^2\left(\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)b_1\Phi\left(\frac{\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n}{\left|\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|}\right)\right]=-\frac{1}{2}\beta\varepsilon^2\left(\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)b_1.$$

The second term on the RHS of (25) follows similarly,

$$\mathbb{E}_n\left[\varepsilon r_2 b_2\left(1\wedge e^{-\beta\left(\varepsilon r_1\frac{\partial H}{\partial\sigma_i^n}\cdot b_1+\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n\right)}\right)\right]=-\frac{1}{2}\beta\varepsilon^2\left(\frac{\partial H}{\partial\sigma_i^n}\cdot b_2\right)b_2+O(\varepsilon^3).$$

Combining the above

$$\mathbb{E}_n\left[\sigma_i^{n+1}-\sigma_i^n\right]\approx-\frac{1}{2}\beta\varepsilon^2\mathrm{P}_{\sigma_i^n}^{\perp}\left(\frac{\partial H}{\partial\sigma_i^n}\right)-\varepsilon^2\sigma_i^n,$$

where $\frac{\partial H}{\partial\sigma_i^n}=\frac{J}{N^2}\Delta_N\sigma_i^n$ and $\Delta_N\sigma_i^n=N^2(\sigma_{i+1}^n+\sigma_{i-1}^n-2\sigma_i^n)$ denotes the discrete Laplacian.

3.3 Diffusion

Recall Γ_i^n in (10),

$$\Gamma_i^n = \begin{cases} \varepsilon\nu_i^n + c_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] & \text{with probability } \alpha \\ -\mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] & \text{with probability } 1-\alpha \end{cases}$$

with accept rate α in (7). Since $\mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right]$ is an order ε^2 term and $\alpha \approx 1$ with small ε, we are going to show

$$\Gamma_i^n \approx \varepsilon\nu_i^n.$$

Proposition 3.2. *The diffusion term*

$$\Gamma_i^n = \sigma_i^{n+1} - \sigma_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] = \varepsilon\nu_i^n + \phi_i^n, \tag{29}$$

where

$$\phi_i^n \equiv \Gamma_i^n - \varepsilon\nu_i^n = \sigma_i^{n+1} - \sigma_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] - \nu_i^n \tag{30}$$

is a random variable with mean $\mathbb{E}\left[\phi_i^n\right] = 0$, variance $\mathbb{E}\left[\|\phi_i^n\|^2\right] \le C\varepsilon^3$, and covariance $\mathbb{E}\left[\phi_i^n \cdot \phi_i^m\right] = 0$ for $n \neq m$.

Proof. For the mean

$$\phi_i^n = \sigma_i^{n+1} - \sigma_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] - \varepsilon\nu_i^n,$$

then $\mathbb{E}\left[\phi_i^n\right] = \mathbb{E}\left[\sigma_i^{n+1} - \sigma_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] - \varepsilon\nu_i^n\right] = 0$.

For the variance,

$$\begin{aligned}
\mathbb{E}\left[\|\phi_i^n\|^2\right] =& \mathbb{E}\left[\left\|\exp_{\sigma_i^n}(\varepsilon\nu_i^n) - \sigma_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] - \varepsilon\nu_i^n\right\|^2 \left(1 \wedge e^{-\beta\delta H}\right)\right] \\
&+ \mathbb{E}\left[\left\|-\varepsilon\nu_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right]\right\|^2 \left(1 - \left(1 \wedge e^{-\beta\delta H}\right)\right)\right] \\
=& \mathbb{E}\left[\left\|c_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right]\right\|^2 \left(1 \wedge e^{-\beta\delta H}\right)\right] \\
&+ \mathbb{E}\left[\left\|-\varepsilon\nu_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right]\right\|^2 \left(1 - \left(1 \wedge e^{-\beta\delta H}\right)\right)\right].
\end{aligned} \tag{31}$$

The first term in the last line of (31)

$$\begin{aligned}
\mathbb{E}\left[\left\|c_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right]\right\|^2 \left(1 \wedge e^{-\beta\delta H}\right)\right] \le& \mathbb{E}\left[\left\|c_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right]\right\|^4\right]^{\frac{1}{2}} \mathbb{E}\left[\left(1 \wedge e^{-\beta\delta H}\right)^2\right]^{\frac{1}{2}} \\
& C\left(\mathbb{E}\left[\|c_i^n\|^4\right] + \mathbb{E}\left[\left\|\mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right]\right\|^4\right]\right)^{\frac{1}{2}} \\
\le& C\varepsilon^4,
\end{aligned}$$

as $\mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] = -\frac{1}{2}\beta\varepsilon^2 \mathrm{P}_{\sigma_i^n}^{\perp}\left(\frac{\partial H}{\partial \sigma_i^n}\right) - \varepsilon^2\sigma_i^n + O(\varepsilon^3)$ and $\mathbb{E}\left[\|c_i^n\|^4\right] \le C\varepsilon^8$ shown in Appendix A.

For the second term in the last line of (31), since $\left|1 - \left(1 \wedge e^{-\beta\delta H}\right)\right| = \left|e^0 - e^{0\wedge(-\beta\delta H)}\right| \le |\beta\delta H|$, we observe

$$\mathbb{E}\left[\left\|-\varepsilon\nu_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right]\right\|^2 \left(1 - \left(1 \wedge e^{-\beta\delta H}\right)\right)\right] \le \mathbb{E}\left[\left\|-\varepsilon\nu_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right]\right\|^4\right]^{\frac{1}{2}} \mathbb{E}\left[|\beta\delta H|^2\right]^{\frac{1}{2}} \le C\varepsilon^3,$$

for some constant C, since $-\varepsilon\nu_i^n - \mathbb{E}_n\left[\sigma_i^{n+1} - \sigma_i^n\right] = -\varepsilon\nu_i^n + O(\varepsilon^2)$ and $\delta H = \varepsilon\sum_j \frac{\partial H}{\partial \sigma_j^n} \cdot \nu_j^n + O(\varepsilon^2)$ are both order ε term.

Combining the above, the variance in (31) is bounded by $\mathbb{E}\left[\|\phi_i^n\|^2\right] \le C\varepsilon^3$.

For the covariance of ϕ_i^n, ϕ_i^m at different time steps $n > m$, and $\zeta = x, y, z$ denotes the coordinates of the vector,

$$\mathbb{E}\left[\phi_{i,\zeta}^n \phi_{i,\zeta}^m\right] = \mathbb{E}\left[\mathbb{E}_n\left[\phi_{i,\zeta}^n \phi_{i,\zeta}^m\right]\right] = \mathbb{E}\left[\phi_{i,\zeta}^m \mathbb{E}_n\left[\phi_{i,\zeta}^n\right]\right] = \mathbb{E}\left[\phi_{i,\zeta}^m 0\right] = 0.$$

□

Remark 3.3. *In fact, the error term is $\mathbb{E}\left[\|\phi_i^n\|^2\right] \sim O(\varepsilon^3)$ and this determines the order of the convergence in Theorem 2.1. The detail of calculation is given in Appendix A.3.*

3.4 Error Estimation

For the error estimation, we apply similar techniques as in the proof of stochastic Euler's method.

Take $\sigma_i, \bar{\sigma}_i$ as in Theorem 2.1. For simplicity we denote $\mu_i(\sigma) = \mathrm{P}^{\perp}_{\sigma_i}(\Delta_N \sigma_i) - \frac{N}{\beta}\sigma_i, \psi_i(\sigma) = \sqrt{N}\beta(I - \sigma_i\sigma_i^T)$ the coefficients in (14). When N, J, β are fixed and $\|\sigma_i\| = 1$, the coefficient μ, ψ are Lipschitz continuous in each coordinates of x. From Theorem 5.2.1 in [25], the SDE system has a unique solution.

Now we have the following estimate on the error.

Proposition 3.3. *Define the error $e(t)$ between M-H interpolation $\bar{\sigma}_i$ and SDE (14) solution σ_i as*

$$e(t) \equiv \sup_{1 \le i \le N, 0 \le s \le t} \mathbb{E}\left[\|\sigma_i(s) - \bar{\sigma}_i(s)\|^2\right]. \tag{32}$$

For any fixed $T > 0$, $e(t)$ is bounded by

$$e(t) \le C(N, J, \beta, T)\sqrt{\delta t} \quad t \in [0, T]. \tag{33}$$

Proof. For the proof we are going to show $e(t)$ satisfies the Grönwall inequality (C_i denotes some constant bound):

$$e(t) \le (C_1 T + C_2)\int_0^t e(s)\mathrm{d}s + \left(C_3\sqrt{\delta t} + C_4 \delta t + C_5 \delta_t^2\right), \tag{34}$$

so $e(t) \le \left(C_3\sqrt{\delta t} + C_4\delta t + C_5 \delta t^2\right)\exp\left(C_1 T(T + C_2)\right)$.

Since $\bar{\sigma}_i(t) = \sigma_i^{\lfloor \frac{t}{\delta t} \rfloor} = \sigma_i^0 + \sum_{j=0}^{\lfloor \frac{s}{\delta t} \rfloor - 1}\left(\sigma_i^{j+1} - \sigma_i^j\right)$ and both $\sigma_i, \bar{\sigma}_i$ start from the same initial condition, from definition of $e(t)$ and θ_i^n in (23), Γ_i^n in (10), ϕ_i^n in (30):

$$\begin{aligned} e(t) &= \sup_{1 \le i \le N, 0 \le s \le t} \mathbb{E}\left[\|\sigma_i(s) - \bar{\sigma}_i(s)\|^2\right] \\ &= \sup_{1 \le i \le N, 0 \le s \le t} \mathbb{E}\left[\left\|\int_0^s \mu_i\big(\sigma(u)\big)\mathrm{d}u + \int_0^s \psi_i\big(\sigma(u)\big)\mathrm{d}W_i(u) - \sum_{j=0}^{\lfloor \frac{s}{\delta t}\rfloor - 1}\left(\mathbb{E}_n\left[\sigma_i^{j+1} - \sigma_i^j\right] + \Gamma_i^j\right)\right\|^2\right] \\ &= \sup_{1 \le i \le N, 0 \le s \le t} \mathbb{E}\left[\left\|\int_0^s \mu_i\big(\sigma(u)\big)\mathrm{d}u + \int_0^s \psi_i\big(\sigma(u)\big)\mathrm{d}W_i(u) - \sum_{j=0}^{\lfloor \frac{s}{\delta t}\rfloor - 1}\left(\mu_i(\sigma^j)\delta t + \varepsilon\nu_i^j + \theta_i^j + \phi_i^j\right)\right\|^2\right] \end{aligned} \tag{35}$$

where $\mu_i(\sigma^j)\delta t = \int_{j\delta t}^{(j+1)\delta t}\mu_i\big(\bar{\sigma}(u)\big)\mathrm{d}u$ and $\varepsilon\nu_i^j = \int_{j\delta t}^{(j+1)\delta t}\psi_i\big(\bar{\sigma}(u)\big)\mathrm{d}W_i$. Applying Hölder's inequality and $\mathbb{E}\left[|X + Y|^2\right] \le 2\mathbb{E}\left[X^2 + Y^2\right]$ produces

$$\begin{aligned} e(t) \le C \sup_{1 \le i \le N, 0 \le s \le t} \mathbb{E}\Bigg[&\left\|\int_0^{\lfloor \frac{s}{\delta t}\rfloor \delta t}\mu_i\big(\sigma(u)\big) - \mu_i\big(\bar{\sigma}(u)\big)\mathrm{d}u\right\|^2 + \left\|\int_0^{\lfloor \frac{s}{\delta t}\rfloor \delta t}\Big(\psi_i\big(\sigma(u)\big) - \psi_i\big(\bar{\sigma}(u)\big)\Big)\mathrm{d}W_i\right\|^2 \\ &+ \left\|\int_{\lfloor \frac{s}{\delta t}\rfloor \delta t}^s \mu_i\big(\sigma(s)\big)\mathrm{d}s\right\|^2 + \left\|\int_{\lfloor \frac{s}{\delta t}\rfloor \delta t}^s \psi_i\big(\sigma_i(s)\big)\mathrm{d}W_i\right\|^2 + \left\|\sum_{j=0}^{\lfloor \frac{s}{\delta t}\rfloor - 1}\theta_i^j\right\|^2 + \left\|\sum_{j=0}^{\lfloor \frac{s}{\delta t}\rfloor - 1}\phi_i^j\right\|^2\Bigg]. \end{aligned} \tag{36}$$

Using Hölder inequality for the first term in (36) with the coordinate $\zeta = x, y, z$

$$\left|\int_0^{\lfloor \frac{s}{\delta t}\rfloor \delta t}\mu_{i,\zeta}\big(\sigma(u)\big) - \mu_{i,\zeta}\big(\bar{\sigma}(u)\big)\mathrm{d}u\right|^2 \le \int_0^{\lfloor \frac{s}{\delta t}\rfloor \delta t}\Big(\mu_{i,\zeta}\big(\sigma(u)\big) - \mu_{i,\zeta}\big(\bar{\sigma}(u)\big)\Big)^2\mathrm{d}u\int_0^{\lfloor \frac{s}{\delta t}\rfloor \delta t}1^2\mathrm{d}u,$$

since μ_i is Lipschitz,

$$\int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} \Big(\mu_{i,\zeta}\big(\sigma(u)\big) - \mu_{i,\zeta}\big(\bar{\sigma}(u)\big)\Big)^2 \mathrm{d}u \leq C_1 \int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} \big(\sigma_{i,\zeta}(u) - \bar{\sigma}_{i,\zeta}(u)\big)^2 \mathrm{d}u,$$

combine $\zeta = x, y, z$ terms,

$$\sup_{1\leq i\leq N, 0\leq s\leq t} \mathbb{E}\left[\left\| \int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} \mu_i\big(\sigma(u)\big) - \mu_i\big(\bar{\sigma}(u)\big)\mathrm{d}u \right\|^2\right] \leq C_1 t \int_0^t e(s)\mathrm{d}s.$$

Applying Itô isometry to the second term of (36) for the x coordinate

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} \psi_i\big(\sigma(u)\big) - \psi_i\big(\bar{\sigma}(u)\big)\mathrm{d}W_i(u)\right)_x^2\right] \\
&= \mathbb{E}\left[\left(\int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} \bar{\sigma}_i^x(\bar{\sigma}_i^x \mathrm{d}W_i^x + \bar{\sigma}_i^y \mathrm{d}W_i^y + \bar{\sigma}_i^z \mathrm{d}W_i^z) - \sigma_i^x(\sigma_i^x \mathrm{d}W_i^x + \sigma_i^y \mathrm{d}W_i^y + \sigma_i^z \mathrm{d}W_i^z)\right)^2\right] \\
&= \mathbb{E}\left[\left(\int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} (\bar{\sigma}_i^x)^2 - (\sigma_i^x)^2 \mathrm{d}W_i^x\right)^2 + \left(\int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} \bar{\sigma}_i^x\bar{\sigma}_i^y - \sigma_i^x\sigma_i^y \mathrm{d}W_i^y\right)^2 + \left(\int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} \bar{\sigma}_i^x\bar{\sigma}_i^z - \sigma_i^x\sigma_i^z \mathrm{d}W_i^z\right)^2\right] \\
&= \mathbb{E}\left[\int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} \big((\bar{\sigma}_i^x)^2 - (\sigma_i^x)^2\big)^2 \mathrm{d}u + \int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} (\bar{\sigma}_i^x\bar{\sigma}_i^y - \sigma_i^x\sigma_i^y)^2 \mathrm{d}u + \int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} (\bar{\sigma}_i^x\bar{\sigma}_i^z - \sigma_i^z\sigma_i^z)^2 \mathrm{d}u\right].
\end{aligned}$$

Since

$$\begin{aligned}
&\big((\bar{\sigma}_i^x)^2 - (\sigma_i^x)^2\big)^2 \leq (|\bar{\sigma}_i^x| + |\sigma_i^x|)^2(\bar{\sigma}_i^x - \sigma_i^x)^2 \leq 4(\bar{\sigma}_i^x - \sigma_i^x)^2, \\
&(\bar{\sigma}_i^x\bar{\sigma}_i^y - \sigma_i^x\sigma_i^y)^2 = \big((\bar{\sigma}_i^x - \sigma_i^x)\bar{\sigma}_i^y + \sigma_i^x(\bar{\sigma}_i^y - \sigma_i^y)\big)^2 \leq 2(\bar{\sigma}_i^x - \sigma_i^x)^2(\bar{\sigma}_i^y)^2 + 2(\sigma_i^x)^2(\bar{\sigma}_i^y - \sigma_i^y)^2, \\
&(\bar{\sigma}_i^x\bar{\sigma}_i^z - \sigma_i^x\sigma_i^z)^2 = \big((\bar{\sigma}_i^x - \sigma_i^x)\bar{\sigma}_i^z + \sigma_i^x(\bar{\sigma}_i^z - \sigma_i^z)\big)^2 \leq 2(\bar{\sigma}_i^x - \sigma_i^x)^2(\bar{\sigma}_i^z)^2 + 2(\sigma_i^x)^2(\bar{\sigma}_i^z - \sigma_i^z)^2,
\end{aligned}$$

then

$$\mathbb{E}\left[\left(\int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} \psi_i\big(\sigma(u)\big) - \psi_i\big(\bar{\sigma}(u)\big)\mathrm{d}W_i(u)\right)_x^2\right] \leq C\mathbb{E}\left[\int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} (\bar{\sigma}_i^x - \sigma_i^x)^2 + (\bar{\sigma}_i^y - \sigma_i^y)^2 + (\bar{\sigma}_i^z - \sigma_i^z)^2 \mathrm{d}s\right]$$

and y, z coordinates of the second term in (36) are similar. Summing up x, y, z coordinates, the second term in (36) is bounded by

$$\sup_{1\leq i\leq N, 0\leq s\leq t} \mathbb{E}\left[\left\| \int_0^{\lfloor \frac{s}{\delta t} \rfloor \delta t} \Big(\psi_i\big(\sigma(u)\big) - \psi_i\big(\bar{\sigma}(u)\big)\Big)\mathrm{d}W_i \right\|^2\right] \leq C_2 \int_0^t e(s)\mathrm{d}s.$$

For the third term in (36)

$$\left\| \int_{\lfloor \frac{s}{\delta t} \rfloor \delta t}^s \mu_i\big(\sigma(s)\big)\mathrm{d}s \right\|^2 \leq C_3 \delta t^2,$$

since $\|\sigma_i\| = 1$ and $s - \lfloor \frac{s}{\delta t} \rfloor \delta t \leq \delta t$.

Apply Itô isometry again for the fourth term in (36),

$$\mathbb{E}\left[\left\| \int_{\lfloor \frac{s}{\delta t} \rfloor \delta t}^s \psi_i\big(\sigma_i(s)\big)\mathrm{d}W_i \right\|^2\right] \leq C_4 \delta t.$$

From Cauchy inequality and $\mathbb{E}\left[\left\|\theta_i^j\right\|^2\right] \le C\varepsilon^6$ in Proposition 3.1, the fifth term in (36) is bounded by

$$\mathbb{E}\left[\left\|\sum_{j=0}^{\lfloor\frac{s}{\delta t}\rfloor-1}\theta_i^j\right\|^2\right] \le \left\lfloor\frac{s}{\delta t}\right\rfloor \sum_{j=0}^{\lfloor\frac{s}{\delta t}\rfloor-1}\mathbb{E}\left[\left\|\theta_i^j\right\|^2\right] \le C_5\left(\left\lfloor\frac{t}{\delta t}\right\rfloor\right)^2\varepsilon^6 = C_6\delta t.$$

From Proposition 3.2, $\mathbb{E}\left[\phi_i^j\cdot\phi_i^k\right] \le C\delta_{jk}\varepsilon^3$ with δ_{jk} the Kronecker delta, the sixth term in (36)

$$\mathbb{E}\left[\left\|\sum_{j=0}^{\lfloor\frac{s}{\delta t}\rfloor-1}\phi_i^j\right\|^2\right] = \sum_{j=0}^{\lfloor\frac{s}{\delta t}\rfloor-1}\mathbb{E}\left[\left\|\phi_i^j\right\|^2\right] \le C_7\left\lfloor\frac{t}{\delta t}\right\rfloor\varepsilon^3 \le C_8\sqrt{\delta t}.$$

Combining all above, we get the Grönwall inequality (34). ☐

Remark 3.4. *In Grönwall inequality, the $C_3\sqrt{\delta t}$ term decides the order of convergence. It comes from $\mathbb{E}\left[\left\|\sum_{j=0}^{\lfloor\frac{s}{\delta t}\rfloor-1}\phi_i^j\right\|^2\right]$, which we show as $O(\varepsilon^3)$ term in A.3.*

With Proposition 3.3, we can get a uniform bound by using Doob's martingale inequality in [25] for a nonnegative submartingale X_t and constant $p>1$:

$$\mathbb{E}\left[\left|\sup_{0\le s\le t}X_s\right|^p\right]^{1/p} \le \frac{p}{p-1}\mathbb{E}\left[|X_t|^p\right]^{1/p}. \tag{37}$$

Proposition 3.4. *Define the error between M-H and SDE dynamics for i-th spin as*

$$e_i(t) = \mathbb{E}\left[\sup_{0\le s\le t}\|\sigma_i(s)-\bar{\sigma}_i(s)\|^2\right] \quad t\in[0,T], 1\le i\le N, \tag{38}$$

for any $T\in\mathbb{R}^+$. There exists some constant C as a function of T, N, J, β and independent of $i, \delta t$,

$$e_i(t) \le C\sqrt{\delta t}. \tag{39}$$

for any $t\in[0,T]$.

Proof. Similar to (36) in Proposition 3.3, we have

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le s\le t}\|\sigma_i-\bar{\sigma}_i\|^2\right] \le C\mathrm{E}\Bigg[&\sup_{0\le s\le t}\left\|\int_0^{\lfloor\frac{s}{\delta t}\rfloor\delta t}\mu_i(\sigma)-\mu_i(\bar{\sigma})\mathrm{d}u\right\|^2 + \sup_{0\le s\le t}\left\|\int_{\lfloor\frac{s}{\delta t}\rfloor\delta t}^{s}\mu_i(\bar{\sigma})\mathrm{d}u\right\|^2 \\ &+ \sup_{0\le s\le t}\left\|\int_0^{\lfloor\frac{s}{\delta t}\rfloor\delta t}\psi_i(\sigma)-\psi_i(\bar{\sigma})\mathrm{d}W_i + \int_{\lfloor\frac{s}{\delta t}\rfloor\delta t}^{s}\psi_i(\sigma)\mathrm{d}W_i\right\|^2 + \sup_{0\le s\le t}\left(\sum_{j=0}^{\lfloor\frac{s}{\delta t}\rfloor-1}\theta_i^j\right)^2 + \sup_{0\le s\le t}\left(\sum_{j=0}^{\lfloor\frac{s}{\delta t}\rfloor-1}\phi_i^j\right)^2\Bigg],\end{aligned} \tag{40}$$

and each term on the RHS will be bounded by $C\sqrt{\delta t}$ for some constant C.

Applying Hölder's inequality for the first term in (40):

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le s\le t}\left\|\int_0^{\lfloor\frac{s}{\delta t}\rfloor\delta t}\mu_i(\sigma)-\mu_i(\bar{\sigma})\mathrm{d}u\right\|^2\right] &\le \mathbb{E}\left[\sup_{0\le s\le t} t\int_0^{\lfloor\frac{s}{\delta t}\rfloor\delta t}\|\mu_i(\sigma)-\mu_i(\bar{\sigma})\|^2\mathrm{d}u\right] \\ &\le \mathbb{E}\left[t\int_0^t\|\mu_i(\sigma)-\mu_i(\bar{\sigma})\|^2\mathrm{d}u\right] \le Ct^2\sum_{j=i-1}^{i+1}\sup_{0\le s\le t}\mathbb{E}\left[\|\sigma_i(s)-\bar{\sigma}_i(s)\|^2\right] \\ &\le Ct^2\sqrt{\delta t}.\end{aligned}$$

The last inequality is from $\sup_{0\le s\le t, 1\le i\le N} \mathbb{E}\left[\|\sigma_i(s)-\bar{\sigma}_i(s)\|^2\right] \le C\sqrt{\delta t}$ in Proposition 3.3.

For the second term in (40), the length of integral is smaller than δt and $\|\mu_i(\sigma)\| \le C$ as $\|\sigma_i\| = 1$, so

$$\sup_{0\le s\le t}\left\|\int_{\lfloor \frac{s}{\delta t}\rfloor \delta t}^{s} \mu_i(\bar{\sigma})\mathrm{d}u\right\|^2 \le C\delta t^2. \tag{41}$$

The integral in the third term of (40) is a martingale. For $\zeta = x, y, z$, denote

$$M_{t,\zeta} \equiv \left(\int_0^{\lfloor \frac{s}{\delta t}\rfloor \delta t} \psi_i(\sigma) - \psi_i(\tilde{\sigma})\mathrm{d}W_i - \int_{\lfloor \frac{s}{\delta t}\rfloor \delta t}^{s} \psi_i(\tilde{\sigma}_i)\mathrm{d}W_i\right)_\zeta.$$

This is a Martingale and $X_t = |M_{t,\zeta}|$ is a nonnegative submartingale, hence by Doob's inequality (37)

$$\mathbb{E}\left[\|\sup_{0\le s\le t} X_s\|^2\right] \le 4\mathbb{E}\left[\|X_t\|^2\right].$$

For a nonnegative submartingale X_t

$$\mathbb{E}\left[\sup_{0\le s\le t} X_s^2\right] = \mathbb{E}\left[\left(\sup_{0\le s\le t} X_s\right)^2\right] \le C\mathbb{E}\left[X_t^2\right] = C\mathbb{E}\left[M_{t,\zeta}^2\right].$$

Applying Itô isometry for the last term similar to the proof of Proposition 3.3 and summing for all coordinates $\zeta = x, y, z$:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\le s\le t}\left\|\int_0^{\lfloor \frac{s}{\delta t}\rfloor \delta t} \psi_i(\sigma) - \psi_i(\tilde{\sigma})\mathrm{d}W_i - \int_{\lfloor \frac{s}{\delta t}\rfloor \delta t}^{s} \psi_i(\tilde{\sigma}_i)\mathrm{d}W_i\right\|^2\right] \\
&\le C\mathbb{E}\left[\left\|\int_0^{\lfloor \frac{t}{\delta t}\rfloor \delta t} \psi_i(\sigma) - \psi_i(\tilde{\sigma})\mathrm{d}W_i - \int_{\lfloor \frac{t}{\delta t}\rfloor \delta t}^{s} \psi_i(\tilde{\sigma}_i)\mathrm{d}W_i\right\|^2\right] \\
&\le C_1\mathbb{E}\left[\int_0^{\lfloor \frac{t}{\delta t}\rfloor \delta t} \|\sigma_i - \bar{\sigma}_i\|^2 \mathrm{d}u\right] + C_2\mathbb{E}\left[\int_{\lfloor \frac{t}{\delta t}\rfloor \delta t}^{s} \|\sigma_i\|^2\mathrm{d}u\right] \\
&\le C_1\sqrt{\delta t} + C_2\delta t.
\end{aligned}$$

From Cauchy inequality, the fourth term in (40)

$$\sup_{0\le s\le t}\left(\sum_{j=0}^{\lfloor \frac{s}{\delta t} - 1\rfloor} \theta_i^j\right)^2 \le \sup_{0\le s\le t}\left\lfloor \frac{s}{\delta t}\right\rfloor \sum_{j=0}^{\lfloor \frac{s}{\delta t}\rfloor - 1} \|\theta_i^j\|^2 \le \left\lfloor \frac{t}{\delta t}\right\rfloor \sum_{j=0}^{\lfloor \frac{t}{\delta t}\rfloor - 1} \|\theta_i^j\|^2,$$

from 3.1 $\mathbb{E}\left[\|\theta_i^j\|^2\right] \le C\varepsilon^6$ so the last expectation is bounded by $C\delta t$.

In the fifth term of (40), $\sum_{j=0}^{\lfloor \frac{s}{\delta t}\rfloor - 1} \phi_i^j$ is a discrete martingale. Again using martingale inequality for each coordinate and then summing up,

$$\mathbb{E}\left[\sup_{0\le s\le t}\left\|\sum_{j=1}^{\lfloor \frac{s}{\delta t}\rfloor} \phi_i^j\right\|^2\right] \le \mathbb{E}\left[\left\|\sum_{j=1}^{\lfloor \frac{t}{\delta t}\rfloor} \phi_i^j\right\|^2\right], \tag{42}$$

from 3.2 $\mathbb{E}\left[\phi_i^j \cdot \phi_i^k\right] \le \delta_{jk} C\varepsilon^3$, it is bounded by

$$\mathbb{E}\left[\sum_{j=1}^{\lfloor \frac{t}{\delta t}\rfloor} \|\phi_i^j\|^2\right] \le C\left\lfloor \frac{t}{\delta t}\right\rfloor \varepsilon^3 \le C\sqrt{\delta t}. \tag{43}$$

Combining above, (39) is obtained. □

4 From SDE system to deterministic PDE

In this section, we explain the convergence from the SDE system (14) to the deterministic Landau-Lifshitz equation without dispersion term (5) with proper choice of $\beta = N^\gamma$ and number of particles $N \to \infty$.

Remark 4.1. *For the XY model from $\mathbb{T}^1 \to \mathbb{S}^1$, the convergence from Langevin equation* (3) *to harmonic map heat flow equation* (4) *can be shown similarly by taking $P_x^\perp(y) = y - (x,y)x$ and the rest of the proof stays the same.*

From [10], for sufficiently regular initial data, there exists a global smooth solution to the Landau-Lifshitz equation with periodic boundary conditions. We will assume such a solution exists in all contexts below. Since the finite difference approximation (18) will converge to the Landau-Lifshitz equation (5) as $N \to \infty$ [29], we only need to compare the SDE system (14) and the ODE system (18).

In the following the error between (14) and (18) is calculated. Since $\beta = N^\gamma, \gamma > 1$ and $N \to \infty$, we denote

$$\epsilon \equiv \sqrt{\frac{N}{\beta}} \tag{44}$$

as a small parameter going to zero with $N \to \infty$. The SDE is then written as

$$\mathrm{d}\sigma_i = \mathrm{P}_{\sigma_i}^\perp (\Delta_N \sigma_i)\,\mathrm{d}t - \epsilon^2 \sigma_i \mathrm{d}t - \epsilon \mathrm{P}_{\sigma_i}^\perp (\mathrm{d}W_i)\,. \tag{45}$$

Lemma 4.1. *Define the error between SDE* (45) *and ODE* (18) *for i-th spin as $\tilde{e}_i \equiv \sigma_i - \tilde{\sigma}_i$ and define $e_i \equiv \epsilon^{-p}\tilde{e}_i$ for $0 < p < 1$. Define*

$$e = \frac{1}{N}\sum_i \|e_i\|^2, \tag{46}$$

we have the following inequality

$$e(t) \le \int_0^t \left(C_1 e^{3/2} + C_2 e\right) ds + C_3 \epsilon^{2-2p} t + \int_0^t \epsilon^{1-p} \frac{1}{N} \sum_i \left(P_{\sigma_i}^\perp (dW_i(t), e_i)\right), \tag{47}$$

with ϵ small enough so that $\epsilon^p N^{5/2} \le 1$.

Proof. Taking the projection given by

$$\mathrm{P}_{\sigma_i}^\perp(\Delta\sigma_i) = -\sigma_i \times (\sigma_i \times \Delta\sigma_i) = \Delta\sigma_i - (\Delta\sigma_i, \sigma_i)\sigma_i$$

together with $\|\sigma_i\| = 1$, we have that

$$(\Delta_N \sigma_i, \sigma_i) = -\frac{1}{2}(\|\nabla_N^+ \sigma_i\|^2 + \|\nabla_N^- \sigma_i\|^2),$$

where $\nabla_N^+ \sigma_i = N(\sigma_{i+1} - \sigma_i), \nabla_N^- \sigma_i = N(\sigma_i - \sigma_{i-1})$. The SDE system (45) can then be written as

$$\mathrm{d}\sigma_i = \left(\Delta_N \sigma_i + \frac{1}{2}(\|\nabla_N^+ \sigma_i\|^2 + \|\nabla_N^- \sigma_i\|^2)\sigma_i\right)\mathrm{d}t - \epsilon^2 \sigma_i \mathrm{d}t - \epsilon \mathrm{P}_{\sigma_i}^\perp (\mathrm{d}W_i)$$

and the ODE system (18) can similarly be written as

$$\mathrm{d}\tilde{\sigma}_i = \left(\Delta_N \tilde{\sigma}_i + \frac{1}{2}(\|\nabla_N^+ \tilde{\sigma}_i\|^2 + \|\nabla_N^- \tilde{\sigma}_i\|^2)\tilde{\sigma}_i\right)\mathrm{d}t.$$

By definition $\tilde{e}_i = \sigma_i - \tilde{\sigma}_i$ satisfies the following equation

$$\begin{aligned}\mathrm{d}\tilde{e}_i =& \Delta_N \tilde{e}_i \mathrm{d}t + \frac{1}{2}\left[2(\nabla_N^+ \tilde{\sigma}_i, \nabla_N^+ \tilde{e}_i)\sigma_i + \|\nabla_N^+ \tilde{e}_i\|^2 \sigma_i + 2(\nabla_N^- \tilde{\sigma}_i, \nabla_N^- \tilde{e}_i)\sigma_i + \|\nabla_N^- \tilde{e}_i\|^2 \sigma_i\right]\mathrm{d}t \\ &+ \frac{1}{2}\left(\|\nabla_N^+ \tilde{\sigma}_i\|^2 + \|\nabla_N^- \tilde{\sigma}_i\|^2\right)\tilde{e}_i \mathrm{d}t - \epsilon^2 \sigma_i \mathrm{d}t + \epsilon \mathrm{P}_{\sigma_i}^\perp (dW_i(t))\,.\end{aligned}$$

Since $\tilde{e}_i = \epsilon^p e_i$ with $0 < p < 1$,

$$\begin{aligned} \mathrm{d}e_i =& \Delta_N e_i \mathrm{d}t + \frac{1}{2}\left[2(\nabla_N^+\tilde{\sigma}_i, \nabla_N^+ e_i)\sigma_i + \epsilon^p\|\nabla_N^+ e_i\|^2\sigma_i + 2(\nabla_N^-\tilde{\sigma}_i, \nabla_N^- e_i)\sigma_i + \epsilon^p\|\nabla_N^- e_i\|^2\sigma_i\right]\mathrm{d}t \\ &+ \frac{1}{2}\left(\|\nabla_N^+\tilde{\sigma}_i\|^2 + \|\nabla_N^-\tilde{\sigma}_i\|^2\right)e_i\mathrm{d}t - \epsilon^{2-p}\sigma_i\mathrm{d}t + \epsilon^{1-p}\mathrm{P}_{\sigma_i}^{\perp}\left(dW_i(t)\right). \end{aligned} \tag{48}$$

Applying Itô's formula to $\frac{1}{2}\mathrm{d}\|e_i\|^2$, we have that

$$\begin{aligned} &\frac{1}{2}\mathrm{d}\|e_i\|^2 \\ &\quad = (e_i, \mathrm{d}e_i) + \mathcal{I}_i \\ &\quad = (\Delta_N e_i, e_i)\mathrm{d}t + \frac{1}{2}\left[2(\nabla_N^+\tilde{\sigma}_i, \nabla_N^+ e_i) + \epsilon^p\|\nabla_N^+ e_i\|^2 + 2(\nabla_N^-\tilde{\sigma}_i, \nabla_N^- e_i) + \epsilon^p\|\nabla_N^- e_i\|^2\right](\sigma_i, e_i)\mathrm{d}t \\ &\qquad + \frac{1}{2}\left(\|\nabla_N^+\tilde{\sigma}_i\|^2 + \|\nabla_N^-\tilde{\sigma}_i\|^2\right)\|e_i\|^2\mathrm{d}t - \epsilon^{2-p}(\sigma_i, e_i)\mathrm{d}t + \epsilon^{1-p}\left(\mathrm{P}_{\sigma_i}^{\perp}\left(\mathrm{d}W_i(t), e_i\right)\right) + \mathcal{I}_i, \end{aligned}$$

where $\mathcal{I}_i$ represents the Itô correction term of order $O(\epsilon^{2-2p})$ as shown in the following computation.

To calculate the Itô correction $\mathcal{I}_i$ we consider an SDE system for both e_i in (48) and σ_i in (14). The Itô correction $\mathcal{I}_i$ for $\mathrm{d}\|e_i\|^2$ combines three parts corresponding to $\frac{\partial^2 e_i^2}{\partial e_i^2}$, $\frac{\partial^2 e_i^2}{\partial\sigma_i\partial e_i}$ and $\frac{\partial^2 e_i^2}{\partial\sigma_i^2}$. Since $\|\sigma_i\| = 1$ we take $\left\|\mathrm{P}_{\sigma_i}^{\perp}\left(\mathrm{d}W_i(t)\right)\right\|^2$ as bounded by $C\mathrm{d}t$. The first term, $\frac{\partial^2 e_i^2}{\partial e_i^2}$, is a constant and $\left\|\epsilon^{1-p}\mathrm{P}_{\sigma_i}^{\perp}\left(\mathrm{d}W_i(t)\right)\right\|^2 \leq C\epsilon^{2-2p}\mathrm{d}t$. The second term, $\frac{\partial^2 e_i^2}{\partial\sigma_i\partial e_i} = \frac{\partial^2 e_i\epsilon^{-p}(\sigma_i-\tilde{\sigma}_i)}{\partial\sigma_i\partial e_i}$, is order ϵ^{-p} but $\epsilon^{1-p}\mathrm{P}_{\sigma_i}^{\perp}\left(\mathrm{d}W_i(t)\right)\cdot\epsilon\mathrm{P}_{\sigma_i}^{\perp}\left(\mathrm{d}W_i(t)\right)$ is order $\epsilon^{2-p}\mathrm{d}t$ so the Itô correction for the second term is also $O(\epsilon^{2-2p}\mathrm{d}t)$. For the third term $\frac{\partial^2 e_i^2}{\partial\sigma_i^2} = \epsilon^{-2p}\frac{\partial^2(\sigma_i-\tilde{\sigma}_i)^2}{\partial\sigma_i^2}$ but $\left\|\epsilon\mathrm{P}_{\sigma_i}^{\perp}\left(\mathrm{d}W_i(t)\right)\right\|^2 \leq C\epsilon^2\mathrm{d}t$ so the Itô correction for the third term is also $O(\epsilon^{2-2p}\mathrm{d}t)$.

From periodic boundary condition, we know that

$$\sum_{i=1}^{N}(\Delta_N e_i, e_i) = -\sum_i\|\nabla_N^+ e_i\|^2 = -\sum_i\|\nabla_N^- e_i\|^2,$$

hence summing up $\mathrm{d}\|e_i\|^2$ we have that

$$\begin{aligned} &\mathrm{d}\left(\sum_i\|e_i\|^2\right) \\ &= -\sum_i\left(\|\nabla_N^+ e_i\|^2 + \|\nabla_N^- e_i\|^2\right)\mathrm{d}t + \sum_i\left[2(\nabla_N^+\tilde{\sigma}_i, \nabla_N^+ e_i) + \epsilon^p\|\nabla_N^+ e_i\|^2 + 2(\nabla_N^-\tilde{\sigma}_i, \nabla_N^- e_i) + \epsilon^p\|\nabla_N^- e_i\|^2\right](\sigma_i, e_i)\mathrm{d}t \\ &\quad + \sum_i\left(\|\nabla_N^+\tilde{\sigma}_i\|^2 + \|\nabla_N^-\tilde{\sigma}_i\|^2\right)\|e_i\|^2\mathrm{d}t - 2N\epsilon^{2-p}(\sigma_i, e_i)\mathrm{d}t + 2\sum_i\epsilon^{1-p}\left(\mathrm{P}_{\sigma_i}^{\perp}\left(\mathrm{d}W_i(t), e_i\right)\right) + 2\sum_i\mathcal{I}_i \\ &\leq -\sum_i\left(\|\nabla_N^+ e_i\|^2 + \|\nabla_N^- e_i\|^2\right)\mathrm{d}t + \sum_i\left[2(\nabla_N^+\tilde{\sigma}_i, \nabla_N^+ e_i) + \epsilon^p\|\nabla_N^+ e_i\|^2 + 2(\nabla_N^-\tilde{\sigma}_i, \nabla_N^- e_i) + \epsilon^p\|\nabla_N^- e_i\|^2\right](\sigma_i, e_i)\mathrm{d}t \\ &\quad + \sum_i\left(\|\nabla_N^+\tilde{\sigma}_i\|^2 + \|\nabla_N^-\tilde{\sigma}_i\|^2\right)\|e_i\|^2\mathrm{d}t - 2N\epsilon^{2-p}(\sigma_i, e_i)\mathrm{d}t + 2\sum_i\epsilon^{1-p}\left(\mathrm{P}_{\sigma_i}^{\perp}\left(\mathrm{d}W_i(t), e_i\right)\right) + cN\epsilon^{2-2p}\mathrm{d}t \end{aligned} \tag{49}$$

For the second term of (49), from the Cauchy-Schwarz inequality, we observe that

$$\begin{aligned} \left|(\nabla_N^+\tilde{\sigma}_i, \nabla_N^+ e_i)(\sigma_i, e_i)\right| &\leq \frac{1}{2}\left(\|\nabla_N^+ e_i\|^2 + \|\nabla_N^+\tilde{\sigma}_i\|^2\|e_i\|^2\right), \\ \left|(\nabla_N^-\tilde{\sigma}_i, \nabla_N^- e_i)(\sigma_i, e_i)\right| &\leq \frac{1}{2}\left(\|\nabla_N^- e_i\|^2 + \|\nabla_N^-\tilde{\sigma}_i\|^2\|e_i\|^2\right). \end{aligned}$$

Since the solution for Landau-Lifshitz equation is smooth, in the third term of (49), $\|\nabla_N^+\tilde{\sigma}_i\|^2 + \|\nabla_N^-\tilde{\sigma}_i\|^2$ can be bounded by some constant C. For the fourth term in (49) $|\epsilon^{2-p}(\sigma_i, e_i)| = |\epsilon^{2-2p}(\sigma_i, \epsilon^p e_i)| \leq 2\epsilon^{2-2p}$

since $\epsilon^p e_i = \sigma_i - \tilde{\sigma}_i$. Hence from (49),

$$\sum_i \|e_i\|^2$$

$$\leq \int_0^t \epsilon^p \sum_i \left(\|\nabla_N^+ e_i\|^2 + \|\nabla_N^- e_i\|^2\right) \|e_i\| \mathrm{d}s + C_1 \int_0^t \sum_i \|e_i\|^2 \mathrm{d}s + C_2 N \epsilon^{2-2p} t + \int_0^t \epsilon^{1-p} \sum_i \left(\mathrm{P}_{\sigma_i}^{\perp} \left(dW_i(t), e_i\right)\right).$$

Using the assumption that $\epsilon^p N^{5/2} \leq 1$ we bound the $\sum_i \left(\epsilon^p \|\nabla_N^+ e_i\|^2 + \epsilon^p \|\nabla_N^- e_i\|^2\right) \|e_i\|$ term by $C \sum_i \frac{1}{\sqrt{N}} \|e_i\|^3$. Hence,

$$\frac{1}{N} \sum_i \left(\epsilon^p \|\nabla_N^+ e_i\|^2 + \epsilon^p \|\nabla_N^- e_i\|^2\right) \|e_i\| \leq C \frac{1}{\sqrt{N^3}} \sum_i \|e_i\|^3 \leq C \left(\frac{1}{N} \sum_i \|e_i\|^2\right)^{\frac{3}{2}} \tag{50}$$

as the p-norm is decreasing.

As $e = \frac{1}{N} \sum_i \|e_i\|^2$, we arrive at (47).

□

Remark 4.2. *If we choose the parameters such that the small Itô correction term from martingale $CN\epsilon^{2-2p}t$ is lower order in N and $\epsilon^p N^3 \sim O(1)$, we could show a similar result for $e = \sum_i \|e_i\|^2$ instead of $e = \frac{1}{N} \sum_i \|e_i\|^2$. And this way we could bound $\mathbb{E}_n \left[\sum_i \|e_i\|^2\right]$.*

Intuitively the inequality from Lemma 4.1 is of Grönwall type and the martingale part,

$$2 \int_0^t \epsilon^{1-p} \frac{1}{N} \sum_i \left(\mathrm{P}_{\sigma_i}^{\perp} \left(dW_i(t), e_i\right)\right)$$

has small coefficient ϵ^{1-p}. This means with large probability

$$e(t) \leq \int_0^t \left(C_1 e^{3/2} + C_2 e\right) \mathrm{d}s + C_3 \epsilon^{2-2p} t + \text{small term from martingale}$$

and $e(t)$ is bounded for a long time interval. We use this idea to show the following proposition.

Proposition 4.1.

$$\mathbb{P}\left(e(t) \leq \epsilon^{2-2p} \exp(Ct)\right) \geq 1 - e^{-\frac{N}{2}}, \quad t \in [0, T] \tag{51}$$

for some constant C and T satisfying $\epsilon^{1-p} e^{\frac{1}{2}CT} \leq 1$.

To prove Proposition 4.1, we will use the exponential martingale inequality for continuous L^2 martingale M_t as in [12, p. 25] or [2] :

$$\mathbb{P}(\sup_t \left(M_t - a/2\langle M\rangle_t\right) > b) \leq e^{-ab} \tag{52}$$

where $\langle M\rangle_t$ is the quadratic variation for M_t.

For the martingale in (46), we calculate its quadratic variation in the following lemma.

Lemma 4.2. *The quadratic variation for the martingale $M_t = 2 \int_0^t \epsilon^{1-p} \frac{1}{N} \sum_i \left(P_{\sigma_i}^{\perp} \left(dW_i(t), e_i\right)\right)$ is*

$$\langle M\rangle_t = 4 \int_0^t \frac{\epsilon^{2-2p}}{N^2} \sum_i \|e_i\|^2 - (\sigma_i, e_i)^2 ds. \tag{53}$$

Proof. The quadratic variation for M_t is captured by a direct summation of the square of the coefficients of the white noise, namely

$$
\begin{aligned}
&\left\langle 2\int_0^t \epsilon^{1-p}\frac{1}{N}\sum_i \left(\mathrm{P}^{\perp}_{\sigma_i}(dW_i(t), e_i)\right)\right\rangle_t = \frac{4\epsilon^{2-2p}}{N^2}\left\langle \int_0^t \sum_i \left(\mathrm{P}^{\perp}_{\sigma_i}(\mathrm{d}W_i), e_i\right)\right\rangle_t \\
&= \frac{4\epsilon^{2-2p}}{N^2}\left\langle \int_0^t \sum_i \mathrm{d}W_i\cdot e_i - (\mathrm{d}W_i,\sigma_i)\sigma_i\cdot e_i\right\rangle_t = \frac{4\epsilon^{2-2p}}{N^2}\left\langle \int_0^t \sum_i \sum_{\zeta=x,y,z}\left(e_i^\zeta - (\sigma_i, e_i)\sigma_i^\zeta\right)\mathrm{d}W_i^\zeta\right\rangle_t \\
&= \frac{4\epsilon^{2-2p}}{N^2}\int_0^t \sum_i \sum_{\zeta=x,y,z}\left(e_i^\zeta - (\sigma_i, e_i)\sigma_i^\zeta\right)^2\mathrm{d}t = \frac{4\epsilon^{2-2p}}{N^2}\int_0^t \sum_i \sum_{\zeta=x,y,z}(e_i^\zeta)^2 + (\sigma_i,e_i)^2(\sigma_i^\zeta)^2 - 2e_i^\zeta\sigma_i^\zeta(\sigma_i,e_i)\mathrm{d}t \\
&= \frac{4\epsilon^{2-2p}}{N^2}\int_0^t \sum_i \|e_i\|^2 + (\sigma_i,e_i)^2\|\sigma_i\|^2 - 2(\sigma_i,e_i)(\sigma_i,e_i)\mathrm{d}t \\
&= \frac{4\epsilon^{2-2p}}{N^2}\int_0^t \sum_i \|e_i\|^2 - (\sigma_i,e_i)^2\mathrm{d}t.
\end{aligned}
$$

□

Now taking $a = \frac{N}{2\epsilon^{2-2p}}$ and $b = \epsilon^{2-2p}$ in the inequality (52), we have that

$$
\mathbb{P}\left(\sup_{0\le s\le t}\left(2\int_0^s \epsilon^{1-p}\frac{1}{N}\sum_i\left(\mathrm{P}^{\perp}_{\sigma_i}(dW_i(t),e_i)\right)\mathrm{d}u - \int_0^s \frac{1}{N}\sum_i\left(\|e_i\|^2 - (\sigma_i,e_i)^2\right)\mathrm{d}u\right) > \epsilon^{2-2p}\right) \le e^{-N/2}. \tag{54}
$$

Thus, for probability $P \ge 1 - e^{-N/2}$,

$$
\begin{aligned}
&2\int_0^t \epsilon^{1-p}\frac{1}{N}\sum_i\left(\mathrm{P}^{\perp}_{\sigma_i}(dW_i(t),e_i)\right) - \int_0^t \frac{1}{N}\sum_i\left(\|e_i\|^2 - (\sigma_i,e_i)^2\right)\mathrm{d}s \\
&\quad\le \sup_{0\le s\le t}\left(2\int_0^s \epsilon^{1-p}\frac{1}{N}\sum_i\left(\mathrm{P}^{\perp}_{\sigma_i}(dW_i(t),e_i)\right)\mathrm{d}u - \int_0^s \frac{1}{N}\sum_i\left(\|e_i\|^2 - (\sigma_i,e_i)^2\right)\mathrm{d}u\right) \\
&\quad\le \epsilon^{2-2p}.
\end{aligned}
$$

Combining this with (47), we observe

$$
\begin{aligned}
e(t) &\le \int_0^t\left(C_1 e^{3/2} + C_2 e\right)\mathrm{d}s + C_3\epsilon^{2-2p}t + 2\int_0^t \epsilon^{1-p}\frac{1}{N}\sum_i\left(\mathrm{P}^{\perp}_{\sigma_i}(dW_i(t),e_i)\right) \\
&\le \int_0^t\left(C_1 e^{3/2} + C_2 e\right)\mathrm{d}s + C_3\epsilon^{2-2p}t + \int_0^t \frac{1}{N}\sum_i\left(\|e_i\|^2 - (\sigma_i,e_i)^2\right)\mathrm{d}s + \epsilon^{2-2p} \\
&\le \int_0^t\left(C_1 e^{3/2} + (C_2+1)e\right)\mathrm{d}s + C_3\epsilon^{2-2p}t + \epsilon^{2-2p}
\end{aligned} \tag{55}
$$

using the definition of $e(t)$ and $(\sigma_i, e_i)^2 \le \|e_i\|^2$.

Since (55) is a Grönwall type inequality, we build a special upper solution $u = \epsilon^{2-2p}e^{(C_1+C_2+C_3+1)t}$. Then

$$
\mathrm{d}u = (C_1+C_2+C_3+1)u \ge C_1u^{3/2} + (C_2+1)u + C_3\epsilon^{2-2p},
$$

where $C_1u \ge C_1u^{3/2}$ when $\epsilon^{1-p}e^{\frac{1}{2}(C_1+C_2+C_3+2)t} \le 1$. We observe that

$$
e^{(C_1+C_2+C_3+1)t} \ge 1 + (C_1+C_2+C_3+1)t,
$$

so $C_3u \ge C_3\epsilon^{2-2p}$. As $u(0) = \epsilon^{2-2p}$, we have that

$$
u(t) \ge \int_0^t\left(C_1u^{3/2} + (C_2+1)u\right)\mathrm{d}s + C_3\epsilon^{2-2p}t + \epsilon^{2-2p} \tag{56}
$$

and $u(t)$ is an upper bound for $e(t)$.

When ϵ is small enough we have that $\epsilon^{1-p}e^{\frac{1}{2}(C_1+C_2+C_3+2)T} \leq 1$ and $u(t)$ is an upper bound for $e(t)$. We observe

$$\mathbb{P}\left(e(t) \leq u(t) = \epsilon^{2-2p}e^{(C_1+C_2+C_3+1)t}\right) \geq 1 - e^{-N/2}, \quad t \in [0,T]. \tag{57}$$

In fact, as $e_i = \epsilon^{-p}(\sigma_i - \tilde{\sigma}_i)$, the bound $e(t) = \frac{1}{N}\sum_i \|e_i\|^2 \leq C$ by some constant C is sufficient. Appealing to a stopping time argument similar to the strong uniqueness proof of the SDE with locally Lipschitz continuous coefficients (see e.g. [16, Chapter 5.2,Theorem 2.5]), we have the following result.

Proposition 4.2. *Given $\epsilon, N > 0$, there exists a constant C independent of ϵ, N such that if $\epsilon^{2-2p}Te^{CT} \leq 1$, then*

$$\mathbb{E}\left[e(t)\right] \leq 1, \quad t \in [0,T].$$

Proof. Define a deterministic time $T \equiv \min\{t \in [0,\infty] : \mathbb{E}\left[e(t)\right] \geq 1\}$ and a stopping time $\tau \equiv \min\{t \in [0,T] : e(t) \geq 1\}$. Since $e(0) = 0$, we have both $T > 0$ and $\tau > 0$. If T is infinite then we are done, so assume T is bounded.

From Lemma 4.1,

$$e(t\wedge\tau) \leq \epsilon^{2-2p}(t\wedge\tau) + \int_0^{t\wedge\tau} C_1 e^{3/2}(s) + C_2 e(s)\mathrm{d}s + \int_0^{t\wedge\tau} \epsilon^{1-p}\frac{1}{N}\sum_i \mathrm{P}^{\perp}_{\sigma_i}\left(\mathrm{d}W_i\right). \tag{58}$$

As $e \geq 0$, we have

$$e(t\wedge\tau) \leq \epsilon^{2-2p}(t\wedge\tau) + \int_0^{t} C_1 e^{3/2}(s\wedge\tau) + C_2 e(s\wedge\tau)\mathrm{d}s + \int_0^{t\wedge\tau} \epsilon^{1-p}\frac{1}{N}\sum_i \mathrm{P}^{\perp}_{\sigma_i}\left(\mathrm{d}W_i\right) \tag{59}$$

and $e^{3/2}(s\wedge\tau) \leq e(s\wedge\tau)$ as $e(s\wedge\tau) \leq 1$, so

$$e(t\wedge\tau) \leq \epsilon^{2-2p}(t\wedge\tau) + \int_0^{t} (C_1 + C_2) e(s\wedge\tau)\mathrm{d}s + \int_0^{t\wedge\tau} \epsilon^{1-p}\frac{1}{N}\sum_i \mathrm{P}^{\perp}_{\sigma_i}\left(\mathrm{d}W_i\right). \tag{60}$$

Taking the expectation on both sides,

$$\mathbb{E}\left[e(t\wedge\tau)\right] \leq \epsilon^{2-2p}\mathbb{E}_n\left[(t\wedge\tau)\right] + \int_0^{t} (C_1 + C_2)\mathbb{E}\left[e(s\wedge\tau)\right]\mathrm{d}s. \tag{61}$$

The expectation $\mathbb{E}\left[\int_0^{t\wedge\tau} \mathrm{P}^{\perp}_{\sigma_i}\left(\mathrm{d}W_i\right)\right] = 0$ can be deduced from optional stopping theorem for continuous time or take $\int_0^{t\wedge\tau} \mathrm{P}^{\perp}_{\sigma_i}\left(\mathrm{d}W_i\right) = \int_0^T \mathbb{1}(s < t\wedge\tau)\mathrm{P}^{\perp}_{\sigma_i}\left(\mathrm{d}W_i(s)\right)$. Notice $\mathbb{E}\left[(t\wedge\tau)\right] \leq T$, so

$$\mathbb{E}\left[e(t\wedge\tau)\right] \leq \int_0^{t} (C_1 + C_2)\mathbb{E}\left[e(s\wedge\tau)\right]\mathrm{d}s + \epsilon^{2-2p}T$$

and by a Grönwall's inequality,

$$\mathbb{E}\left[e(t\wedge\tau)\right] \leq \epsilon^{2-2p}Te^{(C_1+C_2)t}. \tag{62}$$

Choosing $C > C_1 + C_2$, the result follows.

□

By similar arguments, a uniform bound on $e(t)$ can be obtained with a weaker condition on T.

Proposition 4.3. *Given $\epsilon, N > 0$, there exist constants $\tilde{C}_1, \tilde{C}_2$ independent of ϵ, N such that if $(\epsilon^{2-2p}T + \tilde{C}_1\frac{\epsilon^{2-4p}}{N})e^{\tilde{C}_2 T} \leq 1$, then*

$$\mathbb{E}\left[\sup_{0\leq s\leq t} e(s)\right] \leq 1, \quad t \in [0,T].$$

Proof. Define $T \equiv \min\{t : \mathbb{E}\left[\sup_{0\le s\le t} e(s)\right] \ge 1\}$ and $\tau \equiv \min\{t \in [0,T] : e(t) \ge 1\}$. Since $e(0) = 0$ we still have $T > 0, \tau > 0$.

Again from Lemma 4.1, we observe

$$e(t\wedge\tau) \le \int_0^t C_1 e^{3/2}(s\wedge\tau) + C_2 e(s\wedge\tau)\mathrm{d}s + \epsilon^{2-2p}(t\wedge\tau) + \int_0^{t\wedge\tau} \epsilon^{1-p}\frac{1}{N}\sum_i \left(\mathrm{P}^{\perp}_{\sigma_i}(\mathrm{d}W_i), e_i\right). \tag{63}$$

Taking the supremum on both sides,

$$\begin{aligned}\sup_{0\le s\le t} e(s\wedge\tau) \le \sup_{0\le s\le t}\int_0^s C_1 e^{3/2}(u\wedge\tau)\\ + C_2 e(u\wedge\tau)\mathrm{d}u + \sup_{0\le s\le t}\epsilon^{2-2p}(s\wedge\tau) + \sup_{0\le s\le t}\int_0^{s\wedge\tau}\epsilon^{1-p}\frac{1}{N}\sum_i\left(\mathrm{P}^{\perp}_{\sigma_i}(\mathrm{d}W_i), e_i\right)(u)\end{aligned} \tag{64}$$

and $0 \le e(t\wedge\tau) \le 1$ and $t\wedge\tau \le T$ gives

$$\sup_{0\le s\le t} e(s\wedge\tau) \le \int_0^t (C_1+C_2)e(u\wedge\tau)\mathrm{d}u + \epsilon^{2-2p}T + \sup_{0\le s\le t}\int_0^{s\wedge\tau}\epsilon^{1-p}\frac{1}{N}\sum_i\left(\mathrm{P}^{\perp}_{\sigma_i}(\mathrm{d}W_i), e_i\right)(u). \tag{65}$$

Taking expectations on both sides

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le s\le t} e(s\wedge\tau)\right] \le \int_0^t (C_1+C_2)\mathbb{E}\left[e(u\wedge\tau)\right]\mathrm{d}u\\ + \epsilon^{2-2p}T + \mathbb{E}\left[\sup_{0\le s\le t}\int_0^{s\wedge\tau}\epsilon^{1-p}\frac{1}{N}\sum_i\left(\mathrm{P}^{\perp}_{\sigma_i}(\mathrm{d}W_i), e_i\right)(u)\right].\end{aligned} \tag{66}$$

The first integral on the right hand side $\int_0^t (C_1+C_2)\mathbb{E}\left[e(u\wedge\tau)\right]\mathrm{d}u \le \int_0^t (C_1+C_2)\mathbb{E}_n\left[\sup_{0\le s\le u} e(s\wedge\tau)\right]\mathrm{d}u$. Doob's Martingale inequality (37) is used to give a bound of $C\epsilon^{1-2p}$ for the last expecation in (66).

Denote

$$M_s = \int_0^{s\wedge\tau}\epsilon^{1-p}\frac{1}{N}\sum_i\left(\mathrm{P}^{\perp}_{\sigma_i}(\mathrm{d}W_i), e_i\right) = \int_0^s \mathbb{1}(u\le\tau)\epsilon^{1-p}\frac{1}{N}\sum_i\left(\mathrm{P}^{\perp}_{\sigma_i}(\mathrm{d}W_i), e_i\right)(u), \tag{67}$$

since $\mathbb{1}(u\le\tau) \in \mathcal{F}_u$ it is a martingale. Using Doob's martingale inequality (37)

$$\mathbb{E}\left[\left(\sup_{0\le s\le t} M_s\right)^2\right] \le 4\mathbb{E}\left[M_t^2\right] = 4\int_0^t \mathbb{E}\left[\frac{\epsilon^{2-2p}}{N^2}\sum_i \|e_i\|^2 - (\sigma_i, e_i)^2\mathrm{d}u\right] \le C\frac{\epsilon^{2-4p}}{N}. \tag{68}$$

The second equality is from Itô isometry and the third inequality is because $\|\sigma_i\|, \|\tilde{\sigma}_i\| = 1$ so $|(\sigma_i, e_i)| \le \|\sigma_i\|$ and $\|\epsilon^p e_i\| = \|\sigma_i - \tilde{\sigma}_i\| \le 2$.

Now a Grönwall's inequality gives

$$\mathbb{E}\left[\sup_{0\le s\le t} e(s\wedge\tau)\right] \le \int_0^t (C_1+C_2)\mathbb{E}\left[\sup_{0\le s\le u} e(s\wedge\tau)\right]\mathrm{d}u + \epsilon^{2-2p}T + C_3\frac{\epsilon^{2-4p}}{N} \tag{69}$$

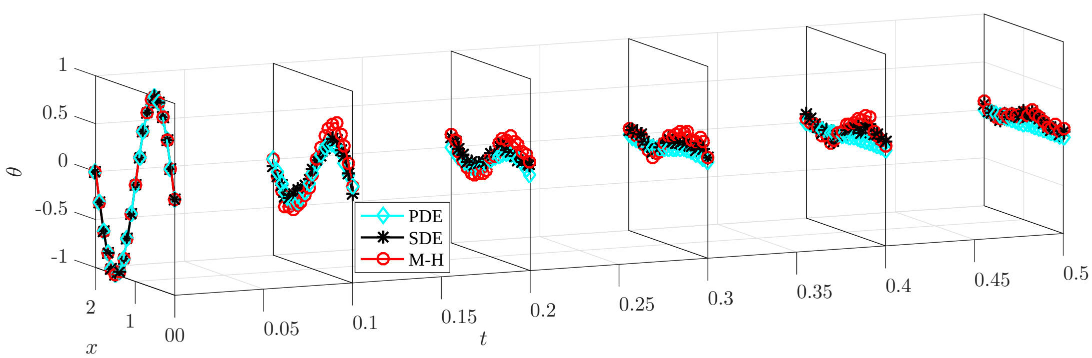
and $\mathbb{E}\left[\sup_{0\le s\le t} e(s\wedge\tau)\right] \le \left(\epsilon^{2-2p}T + C_3\frac{\epsilon^{2-4p}}{N}\right)e^{(C_1+C_2)t}$. As in the proof of the previous Proposition, choosing $\tilde{C}_2 > C_1 + C_2$ and $\tilde{C}_1 = C_3$, the result thus follows similarly.

□

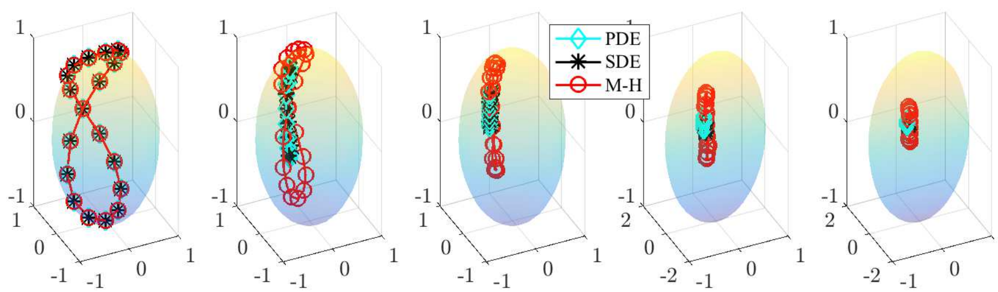
5 Numerical results

In this section, we support our convergence results using numerical simulations of the systems, showing both the temporal dynamics and the order of convergence. The convergence tests indicate that the error decays at least as well as predicted in Theorems 2.1 and 2.2. We check both the cases of the XY model and the classical Heisenberg model. The dynamics of the M-H algorithm are simulated as explained in Sec. 2.1. To simulate the SDE (14), written in the Itô sense, we use the stochastic Euler's method combined with a normalizing step to project the spin back onto the sphere after each time step for both the XY model and the classical Heisenberg model. The PDE (5) is numerically integrated by discretizing in space and using the Euler's method presented in [29] which includes a normalization step.

The out-of-equilibrium to equilibrium dynamics of the M-H algorithm, SDE, and discretized PDE are shown in Figure 1. Figure 1a shows the $\mathbb{T}^1 \to \mathbb{S}^1$ case of the XY model in terms of the polar coordinate θ of each spin. Figure 1b shows the $\mathbb{T}^1 \to \mathbb{S}^2$ case of the classical Heisenberg model with each spin plotted on the same unit sphere; nearest neighbors are connected by a solid line. In both cases, the M-H dynamics tend to lag behind the SDE and PDE which more closely follow each other. This suggests the error between the M-H algorithm and the PDE is dominated by the error between the M-H algorithm and the SDE. Thus the order of convergence between M-H algorithm and Landau-Lifshitz equation should almost follow the order of convergence in Theorem 2.1.



(a) $\mathbb{T}^1 \to \mathbb{S}^1$



(b) $\mathbb{T}^1 \to \mathbb{S}^2$

Figure 1: Dynamics of M-H algorithm (red circles), Langevin equation (black stars) and Landau-Lifshitz equation (cyan diamonds) at various instances of time. They follow each other to converge to equilibrium. In both panels: lattice length $L = 2$, space discretization $\delta x = \frac{1}{N} = \frac{1}{10}$, time step size for M-H algorithm $\delta t = \frac{1}{N^3} = 0.001$, inverse temperature $\beta = N^{3/2} \approx 31.6$, proposal size $\varepsilon = \sqrt{\frac{N\delta t}{\beta}} \approx 0.0178$.

Figures 2 and 3 show the order of convergence for the error between the M-H algorithm and the Langevin equation with respect to the time step size δt, for which the equivalent M-H proposal size is $\varepsilon^2 = \delta t \frac{N}{J\beta}$. The error is calculated at a fixed time T as

$$\mathbb{E}\left[\sqrt{\frac{1}{N}\sum_{i=1}^{N}|\sigma_i^{\mathrm{MH}}(T)-\sigma_i^{\mathrm{SDE}}(T)|^2}\right] \tag{70}$$

where the expectation is taken over multiple realizations. All four frames support that the convergence is at least as good as $\delta t^{1/4}$, which is equivalent to the $\sqrt{\delta t}$ convergence given in theorem 2.1, since the 2-norm is used in the numerical experiments (thus the error is expected to be of order $(\sqrt{\delta t})^{\frac{1}{2}} = \delta t^{\frac{1}{4}}$). The faster convergence of order $\delta t^{1/2}$ in panels (b) and (c) of Fig. 2 we suspect is due to the fact that these out-of-equilibrium dynamics are dominated by the deterministic part of the SDE, and this part has different error scaling from the noisy dynamics. In equilibrium, the deterministic term, $P_{\sigma_i}^{\perp}(\delta_N\sigma_i)$, is small since it is zero at the minimum of the Hamiltonian (maximum of the Gibbs distribution), and the noisy part of the dynamics dominate.

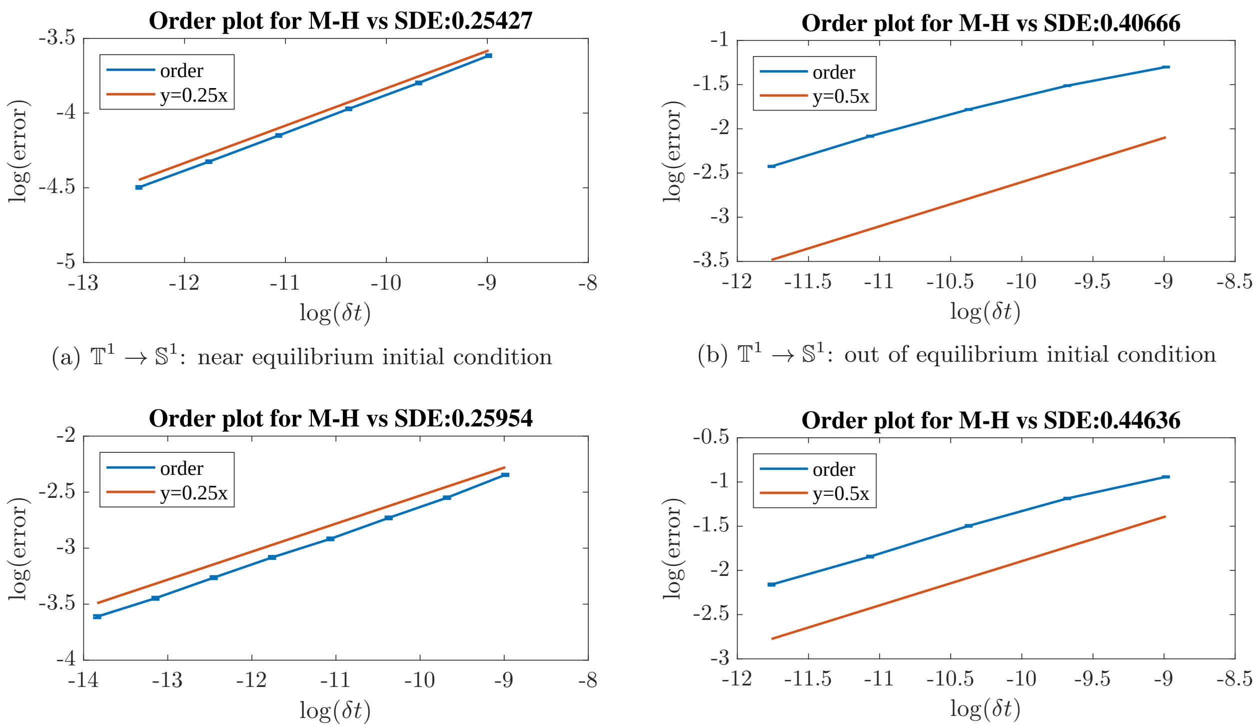
Proposition 3.1 states the error on the deterministic drift of one Metropolis step, θ_i^n, is of size ϵ^3. Dividing by a time-step δt that is proportional to ϵ^2 so that the left-hand side approximates a derivative for the SDE, the resulting error is $O(\epsilon)$ or equivalently $O(\delta t^{1/2})$ as seen in the numerical simulations. Similarly, from Proposition 3.2 the error on the stochastic diffusion of one Metropolis step, ϕ_i^n, is of size $\epsilon^{3/2}$ implying error of order $O(\delta t^{1/4})$, after dividing by the size of the first order term, ϵ. To further test if the difference in convergence order is from the deterministic terms dominating, we increase the size of the noise, $\sqrt{\frac{N}{\beta}}$, in equation (14) by decreasing β to make the noisy dynamics dominate. The out-of-equilibrium error with small $\beta = 1$ shown in Fig. 3 has $\delta t^{1/4}$ convergence, confirming the original statement of Theorem 2.1. We therefore conclude that the error bound of $\delta t^{1/4}$ is tight, and this error comes from the noisy part of the dynamics.

Figure 4 shows the convergence test for the error between the M-H and the PDE dynamics with respect to $\delta x = \frac{1}{N}$. The discrete version of the PDE is simulated with the time-step scaling of $\delta t = \frac{1}{N^4}$ and $\beta = N^{3/2}$, which are also used with $\varepsilon = \sqrt{\frac{N\delta t}{\beta}}$ in the M-H algorithm. These scalings give the order of convergence to be approximately 1, better than our analytical result in Theorem 2.2. A possible explanation is that the error from Theorem 2.1 dominates. As discussed above, Fig. 1 implies the error between the M-H algorithm and the Langevin SDE dominates over the error between the Langevin SDE and the Landau-Lifshitz equation, thus we would expect error of $\delta t^{\frac{1}{4}}$ in theorem 2.1 to dominate. Since we choose the scaling of $\delta t = \frac{1}{N^4} = \delta x^4$ this order of convergence with respect to $\delta x = \frac{1}{N}$ is expected to be $\delta t^{\frac{1}{4}} = \delta x$, or order one. We also point out that the scalings of $\delta t = \frac{1}{N^4}$ and $\beta = N^{3/2}$ are better than the scalings one might guess from theorem 2.1 ($\sqrt{\delta t}$ smaller than the order of e^{-C_2} with C_2 an increasing function of N) and Remark 2.4 ($\beta \gg N^7$). We suspect from the numerical experiments that the scalings of $\beta = N^{\frac{3}{2}}, \delta t = \frac{1}{N^4}$ are tight bounds resulting in order one convergence, but do not have a proof as of yet.

6 Conclusion

We have shown with proposal size $\varepsilon \to 0$ in the Metropolis Hastings algorithm, the Metropolis dynamics converges to the Langevin stochastic differential equation system. With proper scaling of $\beta = N^{\alpha}, \alpha > 1$ and the number of particles $N \to \infty$, the SDE system converges to the deterministic harmonic map heat flow equation.

Several future works are suggested by the results we have obtained. First, the scaling in the analysis is not optimal as suggested by the numerical simulations. One thought to improve the scaling in the calculation is to try to divide it into two situations: near equilibrium and out of equilibrium. When it is out of equilibrium, the drift part $\mathrm{P}_{\sigma_i}^{\perp}(\Delta_N\sigma_i)$ of the SDE (14) should dominate. The Metropolis dynamics would also have a large probability of choosing a lower energy state as the proposal. In this sense both dynamics are close to the deterministic gradient descent. When it's near equilibrium, the drift $\mathrm{P}_{\sigma_i}^{\perp}(\Delta_N\sigma_i)$ in SDE is approximately zero and should behave like a Brownian motion in the neighborhood of equilibrium. The Metropolis dynamics with small proposal size would also stay in the neighborhood of the equilibrium state for a long time. We hope this kind of intuition could help lead to a better scaling in the future work.



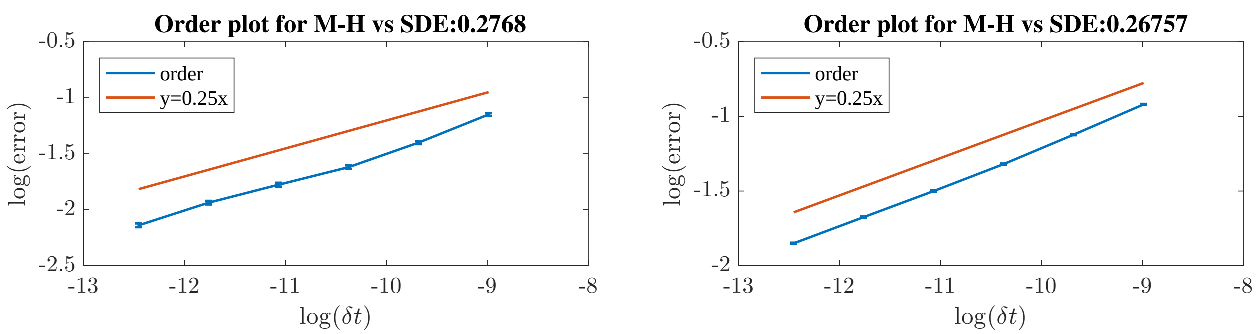
(a) $\mathbb{T}^1 \to \mathbb{S}^1$: near equilibrium initial condition

(b) $\mathbb{T}^1 \to \mathbb{S}^1$: out of equilibrium initial condition

(c) $\mathbb{T}^1 \to \mathbb{S}^2$: near equilibrium initial condition

(d) $\mathbb{T}^1 \to \mathbb{S}^2$: out of equilibrium initial condition

Figure 2: Order of convergence for the error between M-H algorithm and Langevin equation with respect to time step size $\delta t = \frac{J\beta}{N^2}\varepsilon^2$ for $\beta = N^{3/2}$. When the initial condition is near equilibrium, the order of convergence is approximately 0.25 as predicted in theorem 2.1. When the initial condition is out of equilibrium, the order is better than 0.25 and close to 0.5. In all four panels: lattice length $L = 2$, space discretization $\delta x = \frac{1}{N} = \frac{1}{10}$, time step size for M-H algorithm $\delta t = \frac{1}{N^3}$, inverse temperature $\beta = N^{3/2}$, proposal size $\varepsilon = \sqrt{\frac{N\delta t}{\beta}}$.

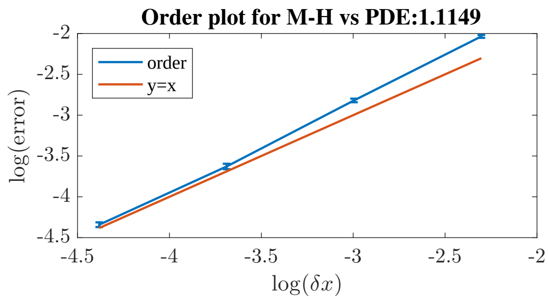


(a) $\mathbb{T}^1 \to \mathbb{S}^1$

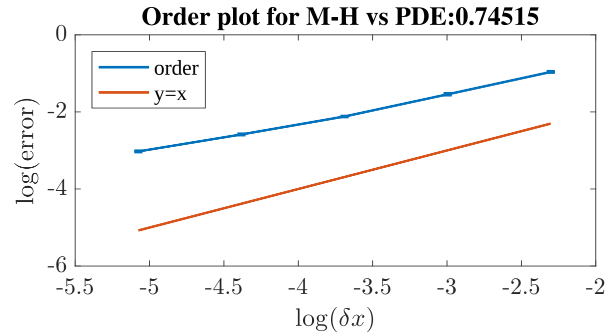
(b) $\mathbb{T}^1 \to \mathbb{S}^2$

Figure 3: Order of convergence for the error between M-H algorithm and Langevin equation with respect to time step size $\delta t = \frac{J\beta}{N^2}\varepsilon^2$ for $\beta = 1$. The order of convergence is approximately 0.25 as predicted in theorem 2.1 with out of equilibrium initial condition. In both panels: lattice length $L = 2$, space discretization $\delta x = \frac{1}{N} = \frac{1}{10}$, time step size for M-H algorithm $\delta t = \frac{1}{N^3}$, inverse temperature $\beta = 1$, proposal size $\varepsilon = \sqrt{\frac{N\delta t}{\beta}}$.

Secondly, the SDE system we get is (3) in Stratonovich sense. If we scale $\beta = N^\alpha$, intuitively $\alpha = 0$ the noise part scales like $\sqrt{\delta x}\sqrt{\delta t}$ as time-space white noise and $0 < \alpha < 1$ gives colored noise. We would



(a) $\mathbb{T}^1 \to \mathbb{S}^1$



(b) $\mathbb{T}^1 \to \mathbb{S}^2$

Figure 4: Order of convergence for the error between M-H algorithm and Landau-Lifshitz equation with respect to lattice discretization size $\delta x = \frac{1}{N}$. From Figure 1 the error between M-H algorithm and Langevin equation dominates over the error between Langevin equation and Landau-Lifshitz equation. The order of convergence is expected to be $\delta t^{\frac{1}{4}} = \delta x$ since we choose $\delta t = \delta x^4$ and it is approximately the case in 4a and 4b. The analytical results in thoerem 2.1 and 2.2 does not give as good convergence rate and demands worse scaling of $\delta t, \beta$ as function of N. The following parameters tested in numerical experiments are enough: Lattice length $L = 2$, initial space discretization $\delta x = \frac{1}{N} = \frac{1}{10}$, time step size for the M-H algorithm $\delta t = \frac{1}{N^3}$, inverse temperature $\beta = 1$, proposal size $\varepsilon = \sqrt{\frac{N\delta t}{\beta}}$.

guess different scale of β would lead the convergence result of a stochastic partial differential equation with white noise or colored noise. And as in [27, 5, 23] this convergence result might imply the optimal scaling of proposal size ε in the Metropolis-Hastings algorithm. We note that as in [23], the drift term might become non-local through the influence of the colored noise, so this approach must be taken with some care.

A Drift and diffusion calculation

In this part we give the details of the error estimations.

Here we state two simple inequalities that is used later. The first is

$$\mathbb{E}\left[|X+Y|^k\right] \le C_k(\mathbb{E}\left[|X|^k\right] + \mathbb{E}\left[|Y|^k\right])$$

for some constant C_k as $(X+Y)^k \le \mathbb{I}(|X| \le |Y|)2^k|Y|^k + \mathbb{I}(|X| > |Y|)2^k|X|^k$. Furthermore, this is also true for vectors

$$\mathbb{E}\left[\|X+Y\|^k\right] \le C_k(\mathbb{E}\left[\|X\|^k\right] + \mathbb{E}\left[\|Y\|^k\right])$$

since $\mathbb{E}\left[\|X+Y\|^k\right] = \mathbb{E}\left[(\|X+Y\|^2)^{\frac{k}{2}}\right] \le \mathbb{E}\left[(2\|X\|^2 + 2\|Y\|^2)^{\frac{k}{2}}\right] \le 2^{\frac{k}{2}}C_k(\mathbb{E}\left[\|X\|^k\right] + \mathbb{E}_n\left[\|Y\|^k\right])$.

The second is Hölder inequality

$$\mathbb{E}\left[XY\right] \le \mathbb{E}\left[|XY|\right] \le \mathbb{E}\left[|X|^p\right]^{\frac{1}{p}}\mathbb{E}\left[|Y|^q\right]^{\frac{1}{q}}$$

with $\frac{1}{p} + \frac{1}{q} = 1$.

A.1 Exponential map

In Section 3.1 we use the notation:

$$\exp_{\sigma_i^n}(\varepsilon\nu_i^n) = \sigma_i^n + \varepsilon\nu_i^n + c_i^n = \sigma_i^n + \varepsilon\nu_i^n - \frac{1}{2}\varepsilon^2(\nu_i^n \cdot \nu_i^n)\sigma_i^n + d_i^n.$$

Now we estimate c_i^n, d_i^n as:

$$\mathbb{E}\left[\|c_i^n\|^k\right] \le C_k\varepsilon^{2k}, \mathbb{E}\left[\|d_i^n\|^k\right] \le D_k\varepsilon^{3k}$$

for any postive integer k and some constants C_k, D_k independent of i, n.

Notice

$$\exp_{\sigma_i^n}(\varepsilon\nu_i^n) = \frac{\sigma_i^n + \varepsilon\nu_i^n}{\|\sigma_i^n + \varepsilon\nu_i^n\|} + a_i^n = \frac{\sigma_i^n + \varepsilon\nu_i^n}{\sqrt{\|\sigma_i^n\|^2 + \varepsilon^2\|\nu_i^n\|^2 - 2\varepsilon(\sigma_i^n \cdot \nu_i^n)}} + a_i^n = \frac{\sigma_i^n + \varepsilon\nu_i^n}{\sqrt{1 + \varepsilon^2\|\nu_i^n\|^2}} + a_i^n.$$

Taylor expanding $\frac{1}{\sqrt{1+\varepsilon^2\|\nu_i^n\|^2}}$, we have

$$\exp_{\sigma_i^n}(\varepsilon\nu_i^n) = (\sigma_i^n + \varepsilon\nu_i^n)\left[1 - \frac{\varepsilon^2}{2}\|\nu_i^n\|^2 + \eta_i^n\right] + a_i^n,$$

where η_i^n is the remainder of the Taylor expansion for $\frac{1}{\sqrt{1+\varepsilon^2\|\nu_i^n\|^2}}$. Then,

$$d_i^n = -\frac{\varepsilon^3}{2}\|\nu_i^n\|^2\nu_i^n + (\sigma_i^n + \varepsilon\nu_i^n)\eta_i^n + a_i^n, \tag{71}$$

$$c_i^n = -\frac{\varepsilon^2}{2}\|\nu_i^n\|^2\sigma_i^n + d_i^n + a_i^n. \tag{72}$$

Let us first deal with the term a_i^n, since the geodesic on the unit sphere is the great circle, $\exp_{\sigma_i^n}(\varepsilon\nu_i^n)$, $\frac{\sigma_i^n+\varepsilon\nu_i^n}{\|\sigma_i^n+\varepsilon\nu_i^n\|}$ are on the same great circle. The arc length of geodesic is $\|\varepsilon\nu_i^n\|$, and the arc length of $\frac{\sigma_i^n+\varepsilon\nu_i^n}{\|\sigma_i^n+\varepsilon\nu_i^n\|}$ is $\arctan(\|\varepsilon\nu_i^n\|)$. The vector a_i^n is the straight line connecting the two points of the difference between these two arcs, and is bounded by the difference of arc lengths:

$$\|a_i^n\| \le \|\varepsilon\nu_i^n\| - \arctan(\|\varepsilon\nu_i^n\|).$$

Taylor expanding for $\arctan x$,

$$\arctan x = x - \frac{x^3}{3} + r, r = \frac{f^{(4)}(\xi)}{4!}x^4, f^{(4)}(x) = \frac{24x(1-x^2)}{(1+x^2)^4},$$

and when $x = \|\varepsilon\nu_i^n\| \ge 0, |f^{(4)}(x)| \le |\frac{24x}{(1+x^2)}\frac{1-x^2}{1+x^2}\frac{1}{(1+x^2)^2}| \le 24$, $|r| \le x^4$. Hence, from $x - \arctan x = \frac{x^3}{3} - r$,

$$\|a_i^n\| \le \|\varepsilon\nu_i^n\| - \arctan(\|\varepsilon\nu_i^n\|) \le \frac{\|\varepsilon\nu_i^n\|^3}{3} + \|\varepsilon\nu_i^n\|^4,$$

so

$$\mathbb{E}\left[\|a_i^n\|^k\right] \le C\mathbb{E}_n\left[\|\varepsilon\nu_i^n\|^{3k}\right] \le C\varepsilon^{3k}$$

In the Taylor expansion for $f(x) = \frac{1}{\sqrt{1+x}}, x \ge 0$, the remainder $r = f(x) - (1 - \frac{x}{2})$ is given by

$$r = \frac{f''(\xi)}{2}x^2 = \frac{3}{8}(1+\xi)^{-\frac{5}{2}}x^2, \xi \in [0, \infty)$$

and $|r| \le \frac{3}{8}x^2$. Applying the above estimates for $r = \eta_i^n$ with $x = \varepsilon^2\|\nu_i^n\|^2 \ge 0$, we observe

$$|\eta_i^n| \le \frac{3}{8}\varepsilon^4\|\nu_i^n\|^4.$$

Now we could get the bound for the terms c_i^n, d_i^n. The first term for d_i^n in (71) is bounded by

$$\mathbb{E}\left[\left\|\frac{\varepsilon^3}{2}\|\nu_i^n\|^2\nu_i^n\right\|^k\right] \le C\mathbb{E}\left[\varepsilon^{3k}\|\nu_i^n\|^{3k}\right] \le C\varepsilon^{3k}.$$

The second term in (71) gives

$$\mathbb{E}\left[\|(\sigma_i^n + \varepsilon\nu_i^n)\eta_i^n\|^k\right] \le \mathbb{E}_n\left[\|\sigma_i^n + \varepsilon\nu_i^n\|^{2k}\right]^{\frac{1}{2}}\mathbb{E}_n\left[|\eta_i^n|^{2k}\right]^{\frac{1}{2}} \le c\varepsilon^{4k}$$

by Hölder's inequality. This is because the first term in the right hand side is bounded by

$$\mathbb{E}\left[\|\sigma_i^n+\varepsilon\nu_i^n\|^{2k}\right]\le C_1\left(\mathbb{E}\left[\|\sigma_i^n\|^{2k}\right]+\varepsilon^{2k}\mathbb{E}\left[\|\nu_i^n\|^{2k}\right]\right)\le C_1+C_2\varepsilon^{2k}$$

and the second term in the right hand side is bounded by

$$\mathbb{E}\left[|\eta_i^n|^{2k}\right]\le C\mathbb{E}\left[\varepsilon^{8k}\|\nu_i^n\|^{8k}\right]\le C\varepsilon^{8k}.$$

The first term for c_i^n in (72)

$$\mathbb{E}\left[\left\|\frac{\varepsilon^2}{2}\|\nu_i^n\|^2\sigma_i^n\right\|^k\right]\le C\varepsilon^{2k}\mathbb{E}\left[\|\nu_i^n\|^{2k}\|\sigma_i^n\|^k\right]\le C\varepsilon^{2k}.$$

The bound for $\mathbb{E}\left[\|c_i^n\|^k\right],\mathbb{E}\left[\|d_i^n\|^k\right]$ are found using the inequalities

$$\mathbb{E}\left[\|X+Y\|^k\right]\le C_k(\mathbb{E}\left[\|X\|^k\right]+\mathbb{E}\left[\|Y\|^k\right])$$

with the above bounds for the terms in c_i^n, d_i^n.

A.2 Drift

Now we give the error estimation for the drift calculation. Denote

$$\theta_i^n\equiv\mathbb{E}_n\left[\sigma_i^{n+1}-\sigma_i^n\right]-\left(-\frac{1}{2}\beta\varepsilon^2\mathrm{P}^{\perp}_{\sigma_i^n}\left(\frac{\partial H}{\partial\sigma_i^n}\right)-\varepsilon^2\sigma_i^n\right).$$

We will show

$$\mathbb{E}\left[\|\theta_i^n\|^2\right]\le C\varepsilon^6.$$

For simplicity we write $\theta_i^n=\sum_k\theta_k$, where θ_k denotes the error for each step of the drift calculation in Section 3.2. And we will show each $\mathbb{E}\left[\|\theta_k\|^2\right]\le C\varepsilon^6$.

Notice θ_k are conditional expectations in the form of $\theta_k=\mathbb{E}_n\left[\mathbf{X}\right]$. Since $f(x)=x^2$ is a convex function, we have $(\mathbb{E}_n\left[\mathbf{X}\right])^2\le\mathbb{E}_n\left[\|\mathbf{X}\|^{\mathbf{2}}\right]$, hence

$$\mathbb{E}\left[\|\theta_k\|^2\right]=\mathbb{E}\left[(\mathbb{E}_n\left[\mathbf{X}\right])^2\right]\le\mathbb{E}\left[\mathbb{E}_n\left[\|\mathbf{X}\|^{\mathbf{2}}\right]\right]=\mathbb{E}\left[\|\mathbf{X}\|^{\mathbf{2}}\right].$$

We will use this to bound $\mathbb{E}\left[\|\theta_k\|^2\right]$.

In the drift calculation, we first take the approximation $\mathbb{E}_n\left[\sigma_i^{n+1}-\sigma_i^n\right]\approx\mathbb{E}_n\left[\left(\varepsilon\nu_i^n-\frac{\varepsilon^2}{2}(\nu_i^n\cdot\nu_i^n)\sigma_i^n\right)\left(1\wedge e^{-\beta\delta H}\right)\right]$. Denote the difference of them as

$$\theta_1\equiv\mathbb{E}_n\left[\sigma_i^{n+1}-\sigma_i^n\right]-\mathbb{E}_n\left[\left(\varepsilon\nu_i^n-\frac{\varepsilon^2}{2}(\nu_i^n\cdot\nu_i^n)\sigma_i^n\right)\left(1\wedge e^{-\beta\delta H}\right)\right]=\mathbb{E}_n\left[d_i^n\left(1\wedge e^{-\beta\delta H}\right)\right],$$

then by Hölder's inequality

$$\mathbb{E}\left[\|\theta_1\|^2\right]\le\mathbb{E}\left[\left\|d_i^n\left(1\wedge e^{-\beta\delta H}\right)\right\|^2\right]\le C\mathbb{E}\left[\|d_i^n\|^4\right]^{\frac{1}{2}}\mathbb{E}\left[\left(1\wedge e^{-\beta\delta H}\right)^4\right]^{\frac{1}{2}}\le C\varepsilon^6.$$

For the second term in drift $\mathbb{E}_n\left[\left(-\frac{\varepsilon^2}{2}(\nu_i^n\cdot\nu_i^n)\sigma_i^n\right)\left(1\wedge e^{-\beta\delta H}\right)\right]\approx\mathbb{E}_n\left[\left(-\frac{\varepsilon^2}{2}(\nu_i^n\cdot\nu_i^n)\sigma_i^n\right)\right]$, denote the error term as

$$\theta_2\equiv\mathbb{E}_n\left[\left(-\frac{\varepsilon^2}{2}(\nu_i^n\cdot\nu_i^n)\sigma_i^n\right)\left(1\wedge e^{-\beta\delta H}\right)\right]-\mathbb{E}_n\left[\left(-\frac{\varepsilon^2}{2}(\nu_i^n\cdot\nu_i^n)\sigma_i^n\right)\right]=\mathbb{E}_n\left[\left(-\frac{\varepsilon^2}{2}(\nu_i^n\cdot\nu_i^n)\sigma_i^n\right)\left(1\wedge e^{-\beta\delta H}-1\right)\right].$$

We have $|(1\wedge e^x)-1|\le|x|$, and $|\delta H|\le C\sum_j\left(\|\varepsilon\nu_j^n\|+\|\varepsilon\nu_j^n\|^2+\|c_j^n\|\right)$ so by Hölder inequality

$$\mathbb{E}\left[\|\theta_2\|^2\right]\le\mathbb{E}\left[\left\|-\frac{\varepsilon^2}{2}(\nu_i^n\cdot\nu_i^n)\sigma_i^n\right\|^4\right]^{\frac{1}{2}}\mathbb{E}\left[\left(1\wedge e^{-\beta\delta H}-1\right)^4\right]^{\frac{1}{2}}\le C\mathbb{E}\left[\varepsilon^8\|\nu_i^n\|^8\|\sigma_i^n\|^4\right]^{\frac{1}{2}}\mathbb{E}\left[|\beta\delta H|^4\right]^{\frac{1}{2}}\le C\beta^2\varepsilon^6.$$

Next we replace δH by $\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot \nu_i^n + R_i^n$ in the drift calculation. Define

$$\theta_3 \equiv \mathbb{E}_n \left[\varepsilon \nu_i^n \left(1 \wedge e^{-\beta \delta H}\right) \right] - \mathbb{E}_n \left[\varepsilon \nu_i^n \left(1 \wedge e^{-\beta \left(\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot \nu_i^n + R_i^n \right)} \right) \right].$$

Notice that $\delta H = \varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot \nu_i^n + R_i^n + h_i^n$ and $|(1 \wedge e^{x+\delta x}) - (1 \wedge x)| \leq |\delta x|$, hence

$$\mathbb{E}\left[\|\theta_3\|^2\right] \leq \mathbb{E}\left[\|\varepsilon \nu_i^n\|^4\right]^{\frac{1}{2}} \mathbb{E}\left[(\beta h_i^n)^4\right]^{\frac{1}{2}}.$$

Since $|h_i^n| \leq C \sum_j \left(\|\varepsilon \nu_j^n\|^2 + \|c_j^n\| \right)$, we have $\mathbb{E}\left[(h_i^n)^4\right] \leq C\varepsilon^8$ and

$$\mathbb{E}\left[\|\theta_3\|^2\right] \leq \mathbb{E}\left[\|\varepsilon \nu_i^n\|^4\right]^{\frac{1}{2}} \mathbb{E}\left[(h_i^n)^4\right]^{\frac{1}{2}} \leq C\beta^2\varepsilon^6.$$

Then we write the drift term as in (25), and have the first term given by

$$\begin{aligned} &\mathbb{E}\left[\varepsilon r_1 b_1 \left(1 \wedge e^{-\beta\left(\varepsilon r_1 \frac{\partial H}{\partial \sigma_i^n} \cdot b_1 + \varepsilon r_2 \frac{\partial H}{\partial \sigma_i^n} \cdot b_2 + R_i^n\right)}\right) \middle| R_i^n, r_2 \right] \\ =& -\beta \varepsilon^2 \left(\frac{\partial H}{\partial \sigma_i^n} \cdot b_1 \right) b_1 e^{\frac{\left(\beta\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot b_1\right)^2}{2} + \beta\varepsilon r_2 \frac{\partial H}{\partial \sigma_i^n} \cdot b_2 + \beta R_i^n} \Phi\left(\frac{\varepsilon r_2 \frac{\partial H}{\partial \sigma_i^n} \cdot b_2 + R_i^n}{\left|\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot b_1\right|} - \left|\beta\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot b_1\right| \right). \end{aligned}$$

We approximate $e^{\frac{\left(\beta\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot b_1\right)^2}{2} + \beta\varepsilon r_2 \frac{\partial H}{\partial \sigma_i^n} \cdot b_2 + \beta R_i^n}$ by 1 in the above equation and bound the following term

$$\begin{aligned} \theta_4 \equiv& \mathbb{E}_n \left[-\beta \varepsilon^2 \left(\frac{\partial H}{\partial \sigma_i^n} \cdot b_1 \right) b_1 e^{\frac{\left(\beta\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot b_1\right)^2}{2} + \beta\varepsilon r_2 \frac{\partial H}{\partial \sigma_i^n} \cdot b_2 + \beta R_i^n} \Phi\left(\frac{\varepsilon r_2 \frac{\partial H}{\partial \sigma_i^n} \cdot b_2 + R_i^n}{\left|\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot b_1\right|} - \left|\beta\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot b_1\right| \right) \right] \\ &- \mathbb{E}_n \left[-\beta \varepsilon^2 \left(\frac{\partial H}{\partial \sigma_i^n} \cdot b_1 \right) b_1 \Phi\left(\frac{\varepsilon r_2 \frac{\partial H}{\partial \sigma_i^n} \cdot b_2 + R_i^n}{\left|\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot b_1\right|} - \left|\beta\varepsilon \frac{\partial H}{\partial \sigma_i^n} \cdot b_1\right| \right) \right] \end{aligned}$$

by

$$\mathbb{E}_n\left[\|\theta_4\|^2\right] \leq C\varepsilon^6.$$

For $z \sim \mathcal{N}(\mu, \sigma^2)$, we use

$$\mathbb{E}\left[|e^z - 1|^k\right] = \mathbb{E}\left[|e^z - 1|^k \mathbb{1}(z \leq 2)\right] + \mathbb{E}\left[|e^z - 1|^k \mathbb{1}(z > 2)\right]. \tag{73}$$

For the first term in (73), since $|e^z - 1| \leq e^2 z$ for $z \leq 2$, we have

$$\mathbb{E}\left[|e^z - 1|^k \mathbb{1}(z \leq 2)|\right] \leq e^{2k} \mathbb{E}\left[z^k\right].$$

When $k = 4$, $\mathbb{E}\left[z^k\right] = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$.

For the second term in (73), $|e^z - 1|^k \leq e^{kz}$ when $z > 2$, suppose $2 + \mu + k\sigma^2 \geq 1$ we have

$$\begin{aligned} \mathbb{E}_n\left[e^{kz} \mathbb{1}(z > 2)\right] &= \int_2^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{kz} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \mathrm{d}z \\ &= \int_2^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{k\mu + \frac{k^2\sigma^2}{2}} e^{-\frac{(z-\mu-k\sigma^2)^2}{2\sigma^2}} \mathrm{d}z = e^{k\mu + \frac{k^2\sigma^2}{2}} \int_{2+\mu+k\sigma^2}^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \mathrm{d}x \\ &\leq e^{k\mu + \frac{k^2\sigma^2}{2}} \int_{2+\mu+k\sigma^2}^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x}{2\sigma^2}} \mathrm{d}x = e^{k\mu + \frac{k^2\sigma^2}{2}} \sqrt{\frac{2}{\pi}} \sigma e^{-\frac{2+\mu+k\sigma^2}{2\sigma^2}} \\ &\leq e^{k\mu + \frac{k^2\sigma^2}{2}} \sqrt{\frac{2}{\pi}} \sigma e^{-\frac{1}{2\sigma^2}}. \end{aligned}$$

Notice that $\frac{\left(\beta\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)^2}{2}+\beta\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+\beta R_i^n\sim\mathcal{N}(c_1\varepsilon^2,c_2\varepsilon^2)$, when ε is small $2+\mu+k\sigma^2\geq 1$, hence we get

$$\mathbb{E}\left[\left(e^{\frac{\left(\beta\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)^2}{2}+\beta\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+\beta R_i^n}-1\right)^4\right]\leq e^8c_2^2\varepsilon^4+\sqrt{\frac{2c_2}{\pi}}\varepsilon e^{-\frac{1}{2c_2\varepsilon^2}}\leq C\varepsilon^4$$

as the term $e^{-\frac{1}{2c_2\varepsilon^2}}$ decays faster than any polynomial of ε as $\varepsilon\to 0$.

So we have

$$\begin{aligned}\mathbb{E}\left[\|\theta_4\|^2\right]=&\mathbb{E}\left[\left\|-\beta\varepsilon^2\left(\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)b_1\Phi\left(\frac{\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n}{\left|\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|}-\left|\beta\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|\right)\right\|^4\right]^{\frac{1}{2}}\\&\times\mathbb{E}\left[\left|e^{\frac{\left(\beta\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)^2}{2}+\beta\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+\beta R_i^n}-1\right|^4\right]^{\frac{1}{2}}\\\leq&C\varepsilon^4\varepsilon^2.\end{aligned}$$

Then we approximate $\Phi\left(\frac{\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n}{\left|\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|}-\left|\beta\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|\right)$ by $\Phi\left(\frac{\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n}{\left|\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|}\right)$ in the calculation. Denote

$$\begin{aligned}\theta_5\equiv\mathbb{E}&\left[-\beta\varepsilon^2\left(\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)b_1\Phi\left(\frac{\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n}{\left|\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|}-\left|\beta\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|\right)\right]\\&-\mathbb{E}_n\left[-\beta\varepsilon^2\left(\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)b_1\Phi\left(\frac{\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n}{\left|\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|}\right)\right].\end{aligned}$$

From $|\Phi(x+\delta x)-\Phi(x)|=|\Phi'(\xi)\delta x|=\left|e^{-\frac{(\xi-\mu)^2}{2\sigma^2}}\delta x\right|\leq|\delta x|$, we have

$$\begin{aligned}\mathbb{E}\left[\theta_5^2\right]=&\mathbb{E}\left[\left\|-\beta\varepsilon^2\left(\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right)b_1\right\|^4\right]^{\frac{1}{2}}\\&\times\mathbb{E}\left[\left|\Phi\left(\frac{\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n}{\left|\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|}-\left|\beta\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|\right)-\Phi\left(\frac{\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n}{\left|\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1\right|}\right)\right|\right]^{\frac{1}{2}}\\&\leq C\varepsilon^4\varepsilon^2.\end{aligned}$$

Similarly we would get the bound for

$$\mathbb{E}\left[\varepsilon r_2b_2\left(1\wedge e^{-\beta\left(\varepsilon r_1\frac{\partial H}{\partial\sigma_i^n}\cdot b_1+\varepsilon r_2\frac{\partial H}{\partial\sigma_i^n}\cdot b_2+R_i^n\right)}\right)\right]$$

and see every $\mathbb{E}\left[\|\theta_i\|^2\right]$ is bounded by $C\varepsilon^6$ so $\mathbb{E}\left[\|\theta_i^n\|^2\right]\leq C\varepsilon^6$.

A.3 Diffusion

Here we are show that $\mathbb{E}\left[\|\phi_i^n\|^2\right]\sim O(\varepsilon^3)$. In

$$\mathbb{E}\left[\|\phi_i^n\|^2\right]=\mathbb{E}\left[\left\|c_i^n-\mathbb{E}_n\left[\sigma_i^{n+1}-\sigma_i^n\right]\right\|^2\left(1\wedge e^{-\beta\delta H}\right)\right]+\mathbb{E}\left[\left\|-\varepsilon\nu_i^n-\mathbb{E}_n\left[\sigma_i^{n+1}-\sigma_i^n\right]\right\|^2\left(1-\left(1\wedge e^{-\beta\delta H}\right)\right)\right]$$

every term is order ε^4 except for $\mathbb{E}\left[\varepsilon^2\|\nu_i^n\|^2\left(1-\left(1\wedge e^{-\beta\delta H}\right)\right)\right]$. We show

$$\mathbb{E}\left[\|\nu_i^n\|^2\left(1\wedge e^{-\beta\delta H}\right)\right]=\mathbb{E}\left[\|\nu_i^n\|^2\right]+O(\varepsilon).$$

Indeed, since $|(1\wedge e^x)-(1\wedge(1+x))|\leq x^2$,

$$\begin{aligned}
&\mathbb{E}\left[\|\nu_i^n\|^2\left(1\wedge e^{-\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)}\right)\right]\\
&\quad\approx\mathbb{E}\left[\|\nu_i^n\|^2\left(1\wedge\left(1-\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)\right)\right)\right]+O(\varepsilon^2)\\
&\quad=\mathbb{E}\left[\|\nu_i^n\|^2\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)<0\right\}}\right]+\mathbb{E}\left[\|\nu_i^n\|^2\left(1-\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)\right)\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)>0\right\}}\right]\\
&\quad=\mathbb{E}\left[\|\nu_i^n\|^2\right]-\mathbb{E}\left[\|\nu_i^n\|^2\left(\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)\right)\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)>0\right\}}\right].
\end{aligned}$$

Then it remains to show $\mathbb{E}\left[\|\nu_i^n\|^2\left(\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)\right)\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)>0\right\}}\right]$ is approximately an $O(\varepsilon)$ term. As before take $\nu_i^n=r_1b_1+r_2b_2$ and we will only take care of r_1, since the calculation for r_2 is similar. Denote $R\equiv R_i^n+\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_2r_2\sim N(0,\varepsilon^2c_1^2), c_1\sim O(1)$. The expectation of the r_1 part is

$$-\beta\mathbb{E}\left[r_1^2\left(R+\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1\right)\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1+R\right)>0\right\}}\right].$$

To calculate, first condition on R and compute it over r_1. This involves expectations in forms of

$$\mathbb{E}\left[z^2\mathbb{1}_{\{(az+b)>0)\}}\right], \text{and } \mathbb{E}\left[z^3\mathbb{1}_{\{(az+b)>0\}}\right].$$

For $z\sim N(0,1)$, a direct calculation gives

$$\begin{aligned}
\mathbb{E}\left[z^2\mathbb{1}_{\{(az+b)>0\}}\right]&=\frac{-b}{\sqrt{2\pi}|a|}e^{-\frac{b^2}{2a^2}}+\Phi\left(\frac{b}{|a|}\right),\\
\mathbb{E}\left[z^3\mathbb{1}_{\{(az+b)>0\}}\right]&=\frac{1}{\sqrt{2\pi}}\left(2+\frac{b^2}{a^2}\right)e^{-\frac{b^2}{2a^2}}\operatorname{sign}(a).
\end{aligned}\tag{74}$$

Using the tower property, we have

$$\begin{aligned}
&\mathbb{E}\left[r_1^2\left(R+\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1\right)\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1+R\right)>0\right\}}\right]\\
&\quad=\mathbb{E}\left[\mathbb{E}_n\left[r_1^2R\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1+R\right)>0\right\}}\middle|R\right]\right]+\mathbb{E}\left[\mathbb{E}_n\left[\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1(r_1)^3\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1+R\right)>0\right\}}\middle|R\right]\right].
\end{aligned}\tag{75}$$

Denote $c_2=\frac{\partial H}{\partial\sigma_i^n}\cdot b_1$, for the first term in (75), the first formula in (74) gives

$$\mathbb{E}_n\left[r_1^2R\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1+R\right)>0\right\}}\middle|R\right]=\mathbb{E}_n\left[-\frac{R^2}{\sqrt{2\pi}|\varepsilon c_2|}e^{-\frac{R^2}{2\varepsilon^2c_2^2}}+R\Phi\left(\frac{R}{|\varepsilon c_2|}\right)\right].\tag{76}$$

For the second term in (75), the second formula in (74) gives

$$\begin{aligned}
&\mathbb{E}_n\left[\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1(r_1)^3\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1+R\right)>0\right\}}\middle|R\right]=\mathbb{E}_n\left[\varepsilon c_2\frac{1}{\sqrt{2\pi}}\left(2+\frac{R^2}{\varepsilon^2c_2^2}\right)e^{-\frac{R^2}{2\varepsilon^2c_2^2}}\operatorname{sign}(c_2)\right]\\
&\quad=\mathbb{E}_n\left[\frac{2\varepsilon|c_2|}{\sqrt{2\pi}}e^{-\frac{R^2}{2\varepsilon^2c_2^2}}+\frac{R^2}{\sqrt{2\pi}\varepsilon|c_2|}e^{-\frac{R^2}{2\varepsilon^2c_2^2}}\right].
\end{aligned}\tag{77}$$

Combining (76) and (77), we have

$$\mathbb{E}_n\left[r_1^2\left(R+\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1\right)\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1+R\right)>0\right\}}\middle|R\right]=\mathbb{E}_n\left[R\Phi\left(\frac{R}{|\varepsilon c_2|}\right)\right]+\mathbb{E}_n\left[\frac{2\varepsilon|c_2|}{\sqrt{2\pi}}e^{-\frac{R^2}{2\varepsilon^2c_2^2}}\right].$$

Since $R\left(\Phi\left(\frac{R}{|\varepsilon c_2|}\right)-\frac{1}{2}\right)\geq 0$, the first term $\mathbb{E}_n\left[R\Phi\left(\frac{R}{|\varepsilon c_2|}\right)\right]=\mathbb{E}_n\left[R\left(\Phi\left(\frac{R}{|\varepsilon c_2|}\right)-\frac{1}{2}\right)+\frac{1}{2}R\right]\geq 0$. In the second term $\frac{R}{\varepsilon c_2}\sim N(0,\frac{c_1^2}{c_2^2})$ and

$$\mathbb{E}_n\left[e^{-\frac{R^2}{2\varepsilon^2 c_2^2}}\right]=\frac{c_2}{\sqrt{c_1^2+c_2^2}}\sim O(1)$$

after a direct calculation. This shows

$$-\beta\mathbb{E}_n\left[r_1^2\left(R+\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1\right)\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot b_1r_1+R\right)>0\right\}}\right]\sim O(\varepsilon)$$

and the r_2 part follows similarly.

Thus we conclude $\mathbb{E}\left[\|\nu_i^n\|^2\left(\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)\right)\mathbb{1}_{\left\{\beta\left(\varepsilon\frac{\partial H}{\partial\sigma_i^n}\cdot\nu_i^n+R_i^n\right)>0\right\}}\right]$ is an ε term.

B Quadratic variation

For an n dimensional process X

$$\mathrm{d}X=\mu\mathrm{d}t+\sigma\mathrm{d}W.$$

Itô's chain rule for a function $f(X)$ is

$$\mathrm{d}f(X_1,X_2,\ldots,X_n)=\sum_i\frac{\partial f}{\partial X_i}\mathrm{d}X_i+\frac{1}{2}\sum_i\sum_j\frac{\partial^2 f}{\partial X_i\partial X_j}\mathrm{d}[X_i,X_j](t),\tag{78}$$

where

$$\mathrm{d}[X_i,X_j](t)=a_{ij}\mathrm{d}t,\quad a=\sigma\sigma^T.$$

For $H=J\sum_{<i,j>}\|\sigma_i-\sigma_j\|^2$, denote i,j as i,j-th spin and $\alpha,\beta=x,y,z$ for the coordinates of the spin

$$\mathrm{d}H=\sum_{i,\alpha}\frac{\partial H}{\partial\sigma_i^\alpha}\mathrm{d}\sigma_i^\alpha+\frac{1}{2}\sum_{i,j,\alpha,\beta}\frac{\partial^2 H}{\partial\sigma_i^\alpha\partial\sigma_j^\beta}\mathrm{d}[\sigma_i^\alpha,\sigma_j^\beta]$$

where

$$\frac{\partial H}{\partial\sigma_i^x}=-\frac{2J}{N^2}\triangle_N\sigma_i^x,\quad\frac{\partial^2 H}{\partial(\sigma_i^x)^2}=\begin{cases}4J & j=i\\-2J & j=i\pm 1\\0 & \text{otherwise}\end{cases}$$

and $\frac{\partial^2 H}{\partial\sigma_i^x\partial\sigma_i^y}=0$. The results for y,z are similar.

The noise term in SDE is $\mathrm{P}_{\sigma_i}^{\perp}(\mathrm{d}W_i)=\mathrm{d}W_i-(\mathrm{d}W_i,\sigma_i)\sigma_i$. For $i\neq j$, W_i,W_j are independent and then $\mathrm{d}[\sigma_i^\alpha,\sigma_j^\beta]=0$, so only $\mathrm{d}[\sigma_i^x,\sigma_i^x]$ need to be calculated. Since W_i^x,W_i^y,W_i^z are also independent, the quadratic variation is calculated by summing up the coefficients before each $\mathrm{d}W_i^\alpha$

$$\mathrm{d}[\sigma_i^x,\sigma_i^x]=\epsilon^2\left[\left(1-(\sigma_i^x)^2\right)^2+(\sigma_i^x\sigma_i^y)^2+(\sigma_i^x\sigma_i^z)^2\right]\mathrm{d}t=\epsilon^2\left[1-(\sigma_i^x)^2\right]\mathrm{d}t,$$
$$\mathrm{d}[\sigma_i^y,\sigma_i^y]=\epsilon^2\left[\left(1-(\sigma_i^y)^2\right)^2+(\sigma_i^x\sigma_i^y)^2+(\sigma_i^y\sigma_i^z)^2\right]\mathrm{d}t=\epsilon^2\left[1-(\sigma_i^y)^2\right]\mathrm{d}t,$$
$$\mathrm{d}[\sigma_i^z,\sigma_i^z]=\epsilon^2\left[\left(1-(\sigma_i^z)^2\right)^2+(\sigma_i^x\sigma_i^z)^2+(\sigma_i^y\sigma_i^z)^2\right]\mathrm{d}t=\epsilon^2\left[1-(\sigma_i^z)^2\right]\mathrm{d}t.$$

As $(\sigma_i^x)^2+(\sigma_i^y)^2+(\sigma_i^z)^2=1$, the Itô correction is

$$\frac{1}{2}4J\epsilon^2\sum_i[1+1+1-(\sigma_i^x)^2-(\sigma_i^y)^2-(\sigma_i^y)^2]=\frac{1}{2}\sum 4J*2\epsilon^2=4JN\epsilon^2$$

and from (8)

$$\mathrm{d}H=\sum-\frac{2J}{N^2}\triangle_N\sigma_i\cdot\left[\left(\frac{JN}{N^2}\mathrm{P}_{\sigma_i}^{\perp}\left(\Delta_N\sigma_i(t)\right)dt-\epsilon^2\sigma_i(t)dt+\mathrm{P}_{\sigma_i}^{\perp}\left(\epsilon\mathrm{d}W_i(t)\right)\right)\right]+4JN\epsilon^2.\tag{79}$$

Denote M_t as the martingale given by

$$\mathrm{d}M_t = \sum_i -\frac{2J}{N^2}\triangle_N\sigma_i \cdot \mathrm{P}_{\sigma_i}^{\perp}\left(\epsilon \mathrm{d}W_i(t)\right) \tag{80}$$

and $\langle M\rangle_t$ the corresponding quadratic variation. The quantity $\mathrm{d}\langle M\rangle_t$ is calculated by summing the square of the coefficients of $\mathrm{d}W_i$ in x, y, z components:

$$\begin{aligned}
\mathrm{d}\langle M\rangle_t &= \mathrm{d}\left\langle \sum_i -\frac{2J}{N^2}\triangle_N\sigma_i \cdot \mathrm{P}_{\sigma_i}^{\perp}\left(\epsilon \mathrm{d}W_i(t)\right)\right\rangle_t = \mathrm{d}\left\langle \sum_i -\frac{2J\epsilon}{N^2}\triangle_N\sigma_i \cdot [\mathrm{d}W_i - (\mathrm{d}W_i\cdot\sigma_i)\sigma_i]\right\rangle_t \\
&= \mathrm{d}\left\langle \sum_i \left[\frac{-2J\epsilon}{N^2}\sum_{\alpha=x,y,z}\left(\triangle_N\sigma_i\cdot \mathrm{d}W_i - (\triangle_N\sigma_i\cdot\sigma_i)\sigma_i \mathrm{d}W_i\right)_\alpha\right]\right\rangle_t \\
&= \sum_i \frac{4J^2\epsilon^2}{N^4}\left[\sum_{\alpha=x,y,z}\left(\triangle_N\sigma_i - (\triangle_N\sigma_i\cdot\sigma_i)\sigma_i\right)_\alpha^2\right]\mathrm{d}t = \sum_i \frac{4J^2\epsilon^2}{N^4}\left\|\mathrm{P}_{\sigma_i}^{\perp}(\triangle_N\sigma_i)\right\|^2 \mathrm{d}t,
\end{aligned}$$

for the last step $\mathrm{P}_{\sigma_i}^{\perp}(\triangle_N\sigma_i) = \triangle_N\sigma_i - (\triangle_N\sigma_i\cdot\sigma_i)\sigma_i$.

Then the inequality for continuous L^2 martingale M_t

$$\mathbb{P}(\sup(M_t - \alpha/2\langle M\rangle_t) > \beta) \le e^{-\alpha\beta} \tag{81}$$

is used to get a bound on H.

Notice in (8), we have

$$\triangle_N\sigma_i \cdot \left(\mathrm{P}_{\sigma_i}^{\perp}\left(\Delta_N\sigma_i\right)\right) = \left\|\mathrm{P}_{\sigma_i}^{\perp}\left(\Delta_N\sigma_i\right)\right\|^2.$$

We can write it in the form of $\mathrm{d}\langle M\rangle_t$ by observing

$$\mathrm{d}H = \mathrm{d}M_t - \frac{N}{2\epsilon^2}\mathrm{d}\langle M\rangle_t + \left[\sum \epsilon^2\frac{2J}{N^2}\triangle_N\sigma_i\cdot\sigma_i + 4JN\epsilon^2\right]\mathrm{d}t$$

and the last term in the bracket

$$\sum \epsilon^2\frac{2J}{N^2}\triangle_N\sigma_i\cdot\sigma_i + 4JN\epsilon^2 = \sum 2J\epsilon^2(\sigma_{i+1}+\sigma_{i-1})\cdot\sigma_i \le 4JN\epsilon^2.$$

So

$$\begin{aligned}
H_N(t) &= H_N(0) + M_t - \frac{N}{2\epsilon^2}\langle M\rangle_t + \int\left[\sum \epsilon^2\frac{2J}{N^2}\triangle_N\sigma_i\cdot\sigma_i + 4JN\epsilon^2\right]\mathrm{d}s \\
&\le H_N(0) + M_t - \frac{N}{2\epsilon^2}\langle M\rangle_t + 4J\epsilon^2 Nt.
\end{aligned}$$

Take $\alpha = \frac{N}{\epsilon^2}$ in the inequality $\mathbb{P}(\sup(M_t - \alpha/2\langle M\rangle_t) > \beta) \le e^{-\alpha\beta}$,

$$\mathbb{P}\left(\sup_{t\le T} H_N \ge H_N(0) + 4JN\epsilon^2 T + \beta\right) \le \mathbb{P}\left(\sup_{t\le T} M_t - \frac{N}{2\epsilon^2}\langle M\rangle_t \ge \beta\right) \le e^{-\frac{\beta N}{\epsilon^2}}. \tag{82}$$

C Diffusion on sphere

We will use Fokker-Planck equation to show the Stratonovich SDE

$$\mathrm{d}\boldsymbol{x} = \mathrm{P}_{\boldsymbol{x}}^{\perp}(\mathrm{d}\boldsymbol{W}) \tag{83}$$

in $\mathbb{R}^2$ and $\mathbb{R}^3$ are describing Brownian motion on the unit circle and unit sphere. And in $\mathbb{R}^3$ it is regardless of the choice for $\mathrm{P}_{\boldsymbol{x}}^{\perp}(\boldsymbol{y}) = \boldsymbol{x}\times\boldsymbol{y}$ or $\mathrm{P}_{\boldsymbol{x}}^{\perp}(\boldsymbol{y}) = -\boldsymbol{x}\times(\boldsymbol{x}\times\boldsymbol{y}) = I - \boldsymbol{x}\boldsymbol{x}^T$.

C.1 Circle $\mathbb{S}^1$

For a Stratonovich SDE with the form

$$\mathrm{d}X_i = b_i \mathrm{d}t + \sum_j \sigma_{ij} \mathrm{d}W_j,$$

the corresponding Itô drift coefficient is

$$\tilde{b}_i = b_i + \frac{1}{2} \sum_j \sum_k \frac{\partial \sigma_{ij}}{\partial x_k} \sigma_{kj}. \tag{84}$$

On the circle $\mathrm{P}_{\boldsymbol{x}}^{\perp}(\mathrm{d}\boldsymbol{W}) = (I - \boldsymbol{x}\boldsymbol{x}^T)\mathrm{d}\boldsymbol{W} = \mathrm{d}\boldsymbol{W} - (\mathrm{d}\boldsymbol{W}, \boldsymbol{x})\boldsymbol{x}$. The corresponding Itô form for (83) is

$$\mathrm{d}\boldsymbol{x} = -\frac{1}{2}\boldsymbol{x} + (I - \boldsymbol{x}\boldsymbol{x}^T)\mathrm{d}\boldsymbol{W}. \tag{85}$$

For Itô SDE $\mathrm{d}\boldsymbol{x} = \mu(\boldsymbol{x}, t)\mathrm{d}t + \sigma(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{W}$, the Fokker-Planck equation is

$$\frac{\partial \rho}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}(\mu_i \rho) + \frac{1}{2} \sum_i \sum_j \frac{\partial^2}{\partial x_i \partial x_j}(D_{ij}\rho), \tag{86}$$

with diffusion tensor

$$D_{ij} = \sum_k \sigma_{ik} \sigma j k = (\sigma \sigma^T)_{ij}.$$

The Fokker-Planck equation for $\mathrm{d}\boldsymbol{x} = -\frac{1}{2}\boldsymbol{x} + (I - \boldsymbol{x}\boldsymbol{x}^T)\mathrm{d}\boldsymbol{W}$ is

$$\partial_t \rho = \frac{1}{2}\left[y^2 \partial_x^2 \rho + x^2 \partial_y^2 \rho - 2xy \partial_x \partial_y \rho - x \partial_x \rho - y \partial_y \rho\right]. \tag{87}$$

The Laplacian on the circle in polar coordinate is $\partial_t \rho = \partial_{\theta\theta}\rho$. Use transformation $x = r\cos\theta, y = r\sin\theta$,

$$\partial_{\theta\theta} = (-y\partial_x + x\partial_y)(-y\partial_x + x\partial_y) = y^2\partial_x^2 + x^2\partial_y^2 - 2xy\partial_x\partial_y - x\partial_x - y\partial_y$$

corresponding to the Fokker-Planck equation above.

C.2 Sphere $\mathbb{S}^2$

In the sphere case, the projection can take the following two forms

$$\mathrm{P}_{\boldsymbol{x}}^{\perp}(\boldsymbol{y}) = \begin{cases} \boldsymbol{x} \times \boldsymbol{y} \\ -\boldsymbol{x} \times (\boldsymbol{x} \times \boldsymbol{y}) = (I - \boldsymbol{x}\boldsymbol{x}^T)\boldsymbol{y} \end{cases}.$$

In both cases the Itô correction are the same as $-\boldsymbol{x}$. The Itô form for (83) is

$$\mathrm{d}\boldsymbol{x} = -\boldsymbol{x}\mathrm{d}t + \mathrm{P}_{\boldsymbol{x}}^{\perp}(\mathrm{d}W). \tag{88}$$

In the Fokker-Planck equation calculation, for both projections the diffusion tensor are the same

$$D = \begin{pmatrix} y^2 + z^2 & -xy & -xz \\ -xy & x^2 + z^2 & -yz \\ -xz & -yz & x^2 + y^2 \end{pmatrix}$$

using the fact $x^2 + y^2 + z^2 = 1$. The Fokker-Planck equation is

$$\begin{aligned} \partial_t \rho = \frac{1}{2} \big[&(y^2 + z^2)\rho_{xx} + (x^2 + z^2)\rho_{yy} + (x^2 + y^2)\rho_{zz} - \partial_x(xy\rho_y + xz\rho_z) \\ &- \partial_y(xy\rho_x + yz\rho_z) - \partial_z(xz\rho_x + yz\rho_y) \big]. \end{aligned}$$

The Laplacian on $\mathbb{S}^2$ in polar coordinate is

$$\frac{1}{\sin\theta}\partial_\theta(\sin\theta\partial_\theta) + \frac{1}{\sin^2\theta}\partial_\phi^2.$$

Using the change of coordinate

$$\begin{cases} x = r\sin\theta\cos\phi \\ y = r\sin\theta\sin\phi \\ z = r\cos\theta \end{cases},$$

we have

$$\begin{aligned}
\partial_\theta &= \frac{xz}{\sqrt{1-z^2}}\partial_x + \frac{yz}{\sqrt{1-z^2}}\partial_y - \sqrt{1-z^2}\partial_z \\
\partial_\theta^2 &= \frac{x^2z^2}{1-z^2}\partial_x^2 + \frac{y^2z^2}{1-z^2}\partial_y^2 + 2\frac{xyz^2}{1-z^2}\partial_{xy}^2 - 2xz\partial_{xz}^2 - 2yz\partial_{yz}^2 - x\partial_x - y\partial_y + (1-z^2)\partial_z^2 - z\partial_z \\
\partial_\phi^2 &= (-y\partial_x + x\partial_y)(-y\partial_x + x\partial_y) = y^2\partial_x^2 + x^2\partial_y^2 - 2xy\partial_x\partial_y - x\partial_x - y\partial_y.
\end{aligned}$$

As $x^2+y^2+z^2=1, \sin^2\theta = x^2+y^2, \frac{\cos\theta}{\sin\theta} = \frac{z}{\sqrt{x^2+y^2}}$, the equation

$$\partial_t\rho = \frac{1}{\sin\theta}\partial_\theta(\sin\theta\partial_\theta\rho) + \frac{1}{\sin^2\theta}\partial_\phi^2\rho$$

is corresponding to the Fokker-Planck equation above.

References

[1] Lubomir Banas, Zdzislaw Brzezniak, Mikhail Neklyudov, and Andreas Prohl. *Stochastic ferromagnetism: analysis and numerics*, volume 58. Walter de Gruyter, 2014.

[2] Richard F. Bass. *Diffusions and elliptic operators*. Probability and its Applications (New York). Springer-Verlag, New York, 1998.

[3] GG Batrouni. Metastable states in the planar two-dimensional xy model and dissipation in superfluid flow. *Physical Review B*, 70(18):184517, 2004.

[4] Kurt Binder, Dieter Heermann, Lyle Roelofs, A John Mallinckrodt, and Susan McKay. Monte carlo simulation in statistical physics. *Computers in Physics*, 7(2):156–157, 1993.

[5] Laird A Breyer and Gareth O Roberts. From Metropolis to diffusions: Gibbs states and optimal scaling. *Stochastic Processes and their Applications*, 90(2):181–206, 2000.

[6] Kun Chen, Alan M Ferrenberg, and DP Landau. Static critical behavior of three-dimensional classical Heisenberg models: A high-resolution Monte Carlo study. *Physical Review B*, 48(5):3249, 1993.

[7] Serena Eley, Sarang Gopalakrishnan, Paul M Goldbart, and Nadya Mason. Approaching zero-temperature metallic states in mesoscopic superconductor–normal–superconductor arrays. *Nature Physics*, 8(1):59, 2012.

[8] Tadahisa Funaki and Herbert Spohn. Motion by mean curvature from the Ginzburg-Landau interface model. *Communications in Mathematical Physics*, 185(1):1–36, 1997.

[9] Boling Guo and Shijin Ding. *Landau-Lifshitz Equations*, volume 1. World Scientific, 2008.

[10] Boling Guo and Min-Chun Hong. The Landau-Lifshitz equation of the ferromagnetic spin chain and harmonic maps. *Calculus of Variations and Partial Differential Equations*, 1(3):311–334, 1993.

[11] MZ Guo, GC Papanicolaou, and SRS Varadhan. Nonlinear diffusion limit for a system with nearest neighbor interactions. *Communications in Mathematical Physics*, 118(1):31–59, 1988.

[12] Martin Hairer and Jonathan C Mattingly. Spectral gaps in wasserstein distances and the 2d stochastic navier-stokes equations. *The Annals of Probability*, pages 2050–2091, 2008.

[13] W Keith Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[14] Benjamin Jourdain, Tony Lelièvre, and Bł ażej Miasojedow. Optimal scaling for the transient phase of Metropolis Hastings algorithms: the longtime behavior. *Bernoulli*, 20(4):1930–1978, 2014.

[15] Benjamin Jourdain, Tony Lelièvre, and Bł ażej Miasojedow. Optimal scaling for the transient phase of the random walk Metropolis algorithm: the mean-field limit. *Ann. Appl. Probab.*, 25(4):2263–2300, 2015.

[16] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

[17] Kay Kirkpatrick and Elizabeth Meckes. Asymptotics of the mean-field heisenberg model. *Journal of Statistical Physics*, 152(1):54–92, 2013.

[18] Kay Kirkpatrick and Tayyab Nawaz. Asymptotics of mean-field o (n) models. *Journal of Statistical Physics*, 165(6):1114–1140, 2016.

[19] Robert V Kohn, Maria G Reznikoff, and Eric Vanden-Eijnden. Magnetic elements at finite temperature and large deviation theory. *Journal of nonlinear science*, 15(4):223–253, 2005.

[20] John Michael Kosterlitz and David James Thouless. Ordering, metastability and phase transitions in two-dimensional systems. *Journal of Physics C: Solid State Physics*, 6(7):1181, 1973.

[21] David P Landau and Kurt Binder. *A guide to Monte Carlo simulations in statistical physics*. Cambridge university press, 2014.

[22] Ilaria Maccari, Andrea Maiorano, Enzo Marinari, and Juan Jesus Ruiz-Lorenzo. A numerical study of planar arrays of correlated spin islands. *The European Physical Journal B*, 89(5):127, 2016.

[23] Jonathan C Mattingly, Natesh S Pillai, Andrew M Stuart, et al. Diffusion limits of the random walk Metropolis algorithm in high dimensions. *The Annals of Applied Probability*, 22(3):881–930, 2012.

[24] MEJ Newman and GT Barkema. *Monte Carlo Methods in Statistical Physics chapter 1-4*. Oxford University Press: New York, USA, 1999.

[25] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 70–74. Springer, 2003.

[26] P Peczak, Alan M Ferrenberg, and DP Landau. High-accuracy Monte Carlo study of the three-dimensional classical Heisenberg ferromagnet. *Physical Review B*, 43(7):6087, 1991.

[27] Gareth O Roberts, Andrew Gelman, Walter R Gilks, et al. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The annals of applied probability*, 7(1):110–120, 1997.

[28] H Eugene Stanley. Dependence of critical properties on dimensionality of spins. *Physical Review Letters*, 20(12):589, 1968.

[29] E Weinan and Xiao-Ping Wang. Numerical methods for the Landau-Lifshitz equation. *SIAM Journal on Numerical Analysis*, pages 1647–1665, 2001.

[30] Horng-Tzer Yau. Relative entropy and hydrodynamics of Ginzburg-Landau models. *Letters in Mathematical Physics*, 22(1):63–80, 1991.