De Gruyter Studies in Mathematics 58

Ľubomír Baňas, Zdzisław Brzeźniak,
Mikhail Neklyudov, Andreas Prohl

Stochastic Ferromagnetism

Analysis and Numerics

De Gruyter

Mathematics Subject Classification 2010: 35R60, 60H15, 60H35, 65Cxx, 91B51, 37A25, 37A30, 37C40, 37M25, 60J05, 60J22, 60M12, 65N06, 65N30.

Authors:

Ľubomír Baňas
Department of Mathematics
Universität Bielefeld
Postfach 100131
33501 Bielefeld
Germany
banas@math.uni-bielefeld.de

Zdzisław Brzeźniak
University of York
Department of Mathematics
Heslington
York YO10 5DD
UK
zdzislaw.brzezniak@york.ac.uk

Mikhail Neklyudov
School of Mathematics and Statistics
Carslaw Building (F07)
University of Sydney NSW 2006
Australia
mikhail.neklyudov@sydney.edu.au

Andreas Prohl
Universität Tübingen
FB Mathematik
Auf der Morgenstelle 10
72076 Tübingen
Germany
prohl@na.uni-tuebingen.de

ISBN 978-3-11-030699-6
e-ISBN 978-3-11-030710-8
Set-ISBN 978-3-11-030711-5
ISSN 0179-0986

Library of Congress Cataloging-in-Publication Data

A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Printing and binding: Hubert & Co. GmbH & Co. KG, Göttingen
∞ Printed on acid-free paper

Printed in Germany

www.degruyter.com

Contents

Introduction

We study thermally activated magnetization dynamics of ferromagnetic nanostructures. A classical microscopic description of an interacting spin system which couples with the surrounding microscopic degrees of freedom (i.e. phonons, conducting electrons, nuclear spins, etc.) is based on the principles of Hamiltonian mechanics [61, Chapter 6]. A mesoscopic description of the statistical properties can be motivated from these equations to reduce the complexity of the model: a general Langevin type model which describes the interaction of atomistic ferromagnetic N-spin ensembles $\mathbf{X} \equiv (X_1, \dots, X_N) : \mathbb{R}^+ \times \boldsymbol{\Omega} \to (\mathbb{S}^2)^N$ in a heat bath is the stochastic Landau-Lifshitz-Gilbert equation (SLLG), see [24, 62, 82, 22, 61],

$$\frac{\partial \mathbf{X}}{\partial t} = \mathbf{X} \times \big(\mathbf{H}_{\text{eff}} + \mathbf{H}_{\text{thm}}\big) - \alpha\, \mathbf{X} \times \big(\mathbf{X} \times \mathbf{H}_{\text{eff}}\big)\,. \tag{1}$$

The deterministic version of this equation (i.e. $\mathbf{H}_{\text{thm}} \equiv \mathbf{0}$) has been introduced in 1935 by Landau and Lifshitz as a phenomenological equation to describe the magnetization at positive temperatures. It was extended to the form (1) by W.F. Brown [24] to account for thermal effects in the case of a single spin ($N = 1$). Here, $\mathbf{H}_{\text{eff}} \equiv \mathbf{H}_{\text{eff}}(\mathbf{X}) = -\nabla \mathcal{E}(\mathbf{X})$ denotes the effective field which acts on spins in the ensemble and which is governed by the total energy of the system $\mathcal{E} : (\mathbb{S}^2)^N \to \mathbb{R}$. This energy is the sum of the exchange energy $\mathcal{E}_{\text{exch}}$ to describe spin-spin interactions, the anisotropy energy $\mathcal{E}_{\text{ani}}$ to model energetically favored alignment of spins with crystallographic axes with the help of the density $\phi : \mathbb{S}^2 \to \mathbb{R}_0^+$, and the external energy $\mathcal{E}_{\text{ext}}$ to account for applied forces $\mathbf{h}_{\text{ext}}$,

$$\mathcal{E}(\mathbf{X}) = \frac{A}{2} \sum_{m,l=1}^{N} J_m^l \langle X_l, X_m \rangle + \frac{K}{2} \sum_{i=1}^{N} \phi(X_i) - \langle\!\langle \mathbf{h}_{\text{ext}}, \mathbf{X} \rangle\!\rangle\,. \tag{2}$$

Here $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ denotes the scalar product in $(\mathbb{R}^3)^N$, $\langle \cdot, \cdot \rangle$ stands for the scalar product in $\mathbb{R}^3$, and $\mathbf{J} = (J_m^l)_{m,l=1}^N \in \mathbb{R}^{N \times N}$ is some given symmetric positive definite matrix. The dynamics of magnetic nanostructures in a heat bath may not be described by classical thermodynamics which is used for macroscopic systems, and where the behavior is reproducible; instead, their modelling is based on non-equilibrium stochastic thermodynamics [102], where irreversible heat losses between the system and the surrounding heat bath are described by relevant thermal fluctuations far away from the equilibrium. In the above model (1), the stochastic field $\mathbf{H}_{\text{thm}} : \mathbb{R}^+ \times \boldsymbol{\Omega} \to (\mathbb{R}^3)^N$ accounts for the interaction of the spin system with thermal fluctuations which allows the system to overcome energy barriers, and to realize related relaxation dynamics. In order to model non-equilibrium thermodynamics, it is customarily assumed that $\mathbf{H}_{\text{thm}} \equiv (H^1_{\text{thm}}, \ldots, H^N_{\text{thm}})$ is Gaussian noise which is uncorrelated in space and time $(t, s \geq 0)$, i.e.,

$$\mathbb{E}\big[H^i_{\text{thm}}(t)\big] = 0\,, \qquad \mathbb{E}\big[\big\langle H^i_{\text{thm}}(t), H^j_{\text{thm}}(s)\big\rangle\big] = \nu^2\, \delta_{ij}\delta(t-s)\,, \tag{3}$$

for all $1 \leq i, j \leq N$. Here $\nu \equiv \nu(\tau) \propto \tau > 0$ is a temperature dependent constant to scale the intensity of thermal fluctuations relative to dissipative effects. The intensity obeys a fluctuation-dissipation relation such that the coupled system converges towards a thermal equilibrium which is described by a Gibbs distribution; see Chapter 1 for further details.

A practically relevant task is to study relaxation dynamics towards thermal equilibrium at elevated temperatures, which often goes along with a spontaneous magnetization reversal to migrate from a metastable magnetic state to another one with lower energy; the quantitative behavior then depends on the intensity $\nu \equiv \nu(\tau) > 0$ in (3). Different approaches by Neel and Brown for single spins [104] provide strong evidence that probabilities for a corresponding thermally induced magnetization reversal to overcome an energy barrier $\Delta\mathcal{E}$ follow the Arrhenius law, which to leading order is proportional to $\sqrt{\tau}e^{-\frac{\Delta\mathcal{E}}{k_B\tau}}$. However, the thermodynamic properties of non-uniform magnetization reversal for general energies $\mathcal{E}$ from (2) are more involved, which is why less is known about corresponding energy barriers. In this case, computational studies may provide valuable insight in the coupling dynamics. A better understanding of the magnetization dynamics at elevated temperatures helps to develop improved nano-scale data storage devices, where too short relaxation times may result in a loss of initially stored data: the smaller memory elements are, the more relevant becomes thermal noise, and its ability to trigger noise-induced magnetization reversal. Another application is heat-assisted magnetic recording to alleviate magnetization reversal on hard-disks by laser pulses, and a corresponding study of the response of spins depending on the temperature.

Chapter 1 addresses finitely many interacting spins and related long-time dynamics, which is inspired by the early work [24] for a single ferromagnetic spin. A question of considerable interest is whether these results also hold for a system which consists of *infinitely* many spins, cf. [21], and if e.g. the corresponding $\mathbb{L}^2(\mathcal{O}, \mathbb{R}^3)$-valued noise may be correlated in space or not to allow for thermodynamically consistent long-time dynamics. For systems which consist of infinitely many spins occupying the ferromagnetic body $\mathcal{O} \subset \mathbb{R}^n$, $n \leq 3$, the following mesoscopic continuum model describes the magnetization process $\boldsymbol{m} : \mathbb{R}^+ \times \mathcal{O} \times \boldsymbol{\Omega} \to \mathbb{S}^2$ at elevated temperatures,

$$\begin{aligned} \frac{\partial \boldsymbol{m}}{\partial t} = \boldsymbol{m} \times \big(\mathbf{H}_{\text{eff}} + \mathbf{H}_{\text{thm}}\big) - \alpha\, \boldsymbol{m} \times (\boldsymbol{m} \times \mathbf{H}_{\text{eff}}) \qquad & \text{on } \mathbb{R}^+ \times \mathcal{O} \times \boldsymbol{\Omega} \\ \frac{\partial \boldsymbol{m}}{\partial \mathbf{n}} = 0 \qquad & \text{on } \mathbb{R}^+ \times \partial\mathcal{O} \times \boldsymbol{\Omega} \\ \boldsymbol{m}(0, \cdot) = \boldsymbol{m}_0 \qquad & \text{on } \mathcal{O} \times \boldsymbol{\Omega}\,, \end{aligned} \tag{4}$$

where $\mathbf{H}_{\text{eff}}(\boldsymbol{m}) = -D\mathcal{E}(\boldsymbol{m})$, and

$$\mathcal{E}(\boldsymbol{m}) = \frac{A}{2} \int_{\mathcal{O}} |\nabla \boldsymbol{m}|^2 \, \mathbf{dx} + \frac{K}{2} \int_{\mathbb{R}^d} \phi(\boldsymbol{m}) \, \mathbf{dx} - \int_{\mathcal{O}} \langle \mathbf{h}_{\text{ext}}, \boldsymbol{m} \rangle \, \mathbf{dx}\,. \tag{5}$$

There is again physical evidence [37] that the related deterministic LLG model (i.e. $\mathbf{H}_{\text{thm}} \equiv \mathbf{0}$) which describes the dynamics of magnetizations $\mathbf{m} : \mathbb{R}^+ \times \mathcal{O} \to \mathbb{S}^2$ requires a modification at elevated temperatures: in this case, an enhanced damping property of the spin system is observed in experiments, as well as a non-constant (sample averaged) magnetization magnitude in both, space and time, which may not be explained by the deterministic model. As a consequence, a stochastic version of the deterministic LLG model is used to statistically describe small-scale effects which are too complex to be described in detail by a microscopic model. From a mathematical viewpoint, problem (4)–(5) is a stochastic nonlinear partial differential equation where the solution process is $\mathbb{S}^2$-valued, see also [32]. The related deterministic LLG model has been analyzed in the literature for $n = 2, 3$: global weak solutions are known to exist, and the possible formation of singularities at finite times from smooth initial data ($n \geq 2$) is motivated by the numerical studies in [17, 10]. In Chapter 2, we evidence a regularizing effect on solutions of (4)–(5) in the case of space-time white noise in (4) by means of computational experiments which are obtained from a convergent space-time discretization.

Simulations to obtain relevant statistical information from (4)–(5) are in general based on Monte-Carlo methods, and are computationally intensive. Hence, a major goal is to derive effective macroscopic equations of motion for averaged magnetizations which accurately account for thermal effects. A phenomenological description for a single macro-spin $\mathbf{m} = \mathbb{E}[\boldsymbol{m}]$ which allows for proper relaxation dynamics has been derived in [59] within a mean-field approximation, which is based on the following consequence of equation (4),

$$\frac{\partial \mathbf{m}}{\partial t} = \boldsymbol{\Lambda}_{\mathtt{N}}\, \mathbf{m} + \mathbb{E}\big[\boldsymbol{m} \times \mathbf{H}_{\text{eff}}\big] - \alpha\, \mathbb{E}\big[\boldsymbol{m} \times (\boldsymbol{m} \times \mathbf{H}_{\text{eff}})\big] \qquad \text{on } \mathbb{R}^+ \times \mathcal{O}\,, \tag{6}$$

where $\boldsymbol{\Lambda}_{\mathtt{N}} \equiv \boldsymbol{\Lambda}_{\mathtt{N}}(\tau) \propto \tau$ in front of the Bloch relaxation term is known as the Neel time; see [59]. It is this term which allows for a varying length of the macro-spin $\mathbf{m}$ for different temperatures. An approximation of the nonlinear terms in (6) which yields a closed effective equation for the macro-spin $\mathbf{m}$ is then often referred to as Landau-Lifshitz-Bloch equation (LLB). It has been shown to properly describe domain wall motion in the presence of a non-constant magnetization length, macroscopic magnetization magnitudes, observed enhanced macroscopic damping, or longitudinal — next to transverse — relaxation dynamics at elevated temperatures.

A different approach to construct effective magnetization models which account for thermal activation is proposed in [14], where mutual orthogonality of vectors $\mathbf{m}$, $\mathbf{m} \times \mathbf{H}_{\text{eff}}$, and $\mathbf{m} \times (\mathbf{m} \times \mathbf{H}_{\text{eff}})$ is used to describe the temperature-dependent damped gyroscopic precession by means of

$$\begin{aligned} \frac{\partial \mathbf{m}}{\partial t} = \kappa\, \mathbf{m} + \mathbf{m} \times \mathbf{H}_{\text{eff}} - \frac{\widehat{\alpha}}{m} \mathbf{m} \times (\mathbf{m} \times \mathbf{H}_{\text{eff}}) \qquad & \text{on } \mathbb{R}^+ \times \mathcal{O} \\ \frac{\partial \mathbf{m}}{\partial \mathbf{n}} = 0 \qquad & \text{on } \mathbb{R}^+ \times \partial\mathcal{O} \\ \mathbf{m}(0,\cdot) = \mathbf{m}_0 \qquad & \text{on } \mathcal{O}\,. \end{aligned} \tag{7}$$

The leading Bloch relaxation term again allows for shrinking ($\kappa < 0$), extension ($\kappa > 0$), and conservation ($\kappa = 0$) of the magnetization length at finite temperatures, where the function κ is chosen to meet the following phenomenological power-law by Landau below the Curie temperature τ_C,

$$\widetilde{m}(\tau) = \widetilde{m}_0 \big(1 - \frac{\tau}{\tau_C}\big)^{\beta}\,,$$

for some $\beta > 0$. Both models, the one which comes from a moment closure approximation of (6), and (7) are phenomenological and lack a rigorous derivation from the mesoscopic model (4)–(5), so that it remains unclear whether these descriptions properly describe the magnetization dynamics at elevated temperatures which is governed by (4)–(5). A major advantage of both macro-spin models is their capability to describe space-time multiscale magnetization, while the spin model (4)–(5) is practically restricted to nanometer scales; a major disadvantage, however, is that additional material functions are needed, such as e.g. $\widehat{\alpha} \equiv \widehat{\alpha}(\tau)$ in order to reliably model microscopic dissipative effects on a macroscopic scale for temperatures $\tau \in [0, \tau_C)$.

This work reports on recent developments concerning the analytical and the numerical treatment of the SLLG equation, addressing in particular the following questions:

(i) *Finite ensembles: long-time behavior.* The fluctuation-dissipation relation from physics determines the noise in dissipative non-equilibrium systems. For one spin and a simplified field $\mathbf{H}_{\text{eff}}$, formal arguments in [62] show that the stationary distribution of (1) is Gibbsian. We remark that (1) is not a gradient system with additive noise, for which the invariant measure is known to be of this type. In Chapter 1, we show uniqueness, and exponential ergodicity of an invariant measure of Gibbs type for (1). A structure-preserving numerical discretization is proposed which yields $\mathbb{S}^2$-valued iterates, inherits the Lyapunov structure, as well as the discrete ergodicity property, and thus converges to the solution of the SLLG equation both, at finite and infinite times. We remark that to construct a convergent discretization is non-trivial because of the Stratonovich stochastic integral and the weak coercivity properties of the nonlinear drift function in (1), which is why general time-explicit integrators may even not converge at finite times [84, 62], or may fail to be ergodic; see e.g. [91, 88, 92]. Another subject which is addressed is the interplay of stochasticity and $(\mathbb{R}^3)^N$-valued solutions which approximate $(\mathbb{S}^2)^N$ by a penalization strategy: a main observation here is that such an approximation also requires a modification of the noise term in order to ensure a proper long-time dynamics.

(ii) *Infinite ensembles: blow-up behavior and long-time dynamics at elevated temperatures.* Possible finite-time finite-energy blow-up behavior of initially smooth solutions of the deterministic Landau-Lifshitz-Gilbert equation on bounded domains $\mathcal{O} \subset \mathbb{R}^2$ is motivated by computational studies in [17]. In Chapter 2, an implementable finite element based space-time discretization is proposed for bounded Lipschitz domains $\mathcal{O} \subset \mathbb{R}^n$ $(n \leq 3)$, where iterates construct a weak martingale solution of the SLLG equation (4)–(5) for vanishing discretization parameters. This discretization is structure-preserving, i.e. solutions satisfy a (pointwise) sphere-property, as well as an energy estimate. Computational studies motivate possible pathwise blow-up of solutions, but a smooth evolution of related expectations in the presence of space-time white noise.

(iii) *Effective macro-spin magnetization dynamics in a heat bath.* A challenging goal is to derive macroscopic equations to properly describe macro-spin magnetization dynamics for a broad range of temperatures. Macro-spin magnetizations $\mathbf{m} = \mathbb{E}[\boldsymbol{m}]$ are first moments of solutions of the infinite-dimensional SLLG equation (4)–(5), which in the case of the single spin model is approximated in [59] by the solution of the phenomenological LLB model. An independent, simple description of magnetization dynamics leads to (7), where changes of the magnetizations are described in terms of the current magnetization, the torque, and the damping term, together with the Landau power law to account for temperature effects on the saturation magnetization. Comparative computational studies for both, the stochastic mesoscopic system (4)–(5), and the macroscopic model (7) are provided, which motivate increased dissipativity for system (7) at elevated temperatures.

The main goal in this work is to use constructive methods to verify mathematical results — for instance, to construct an invariant measure in part (i) by a structure-preserving time discretization, and a weak martingale solution in (ii) by finite element based space-time discretizations. This approach then provides a theoretical foundation for computational simulations with such schemes to study phenomena which so far lack a rigorous analytical understanding — such as e.g. the (long-time) dynamics of the stochastic partial differential equation (4)–(5) with space-time white noise. The following three chapters address items (i) to (iii). Chapters 1 and 2 each start with a preliminary section which provides relevant background material. Numerical schemes are proposed in the main parts of the different chapters in order to construct strong SDE-solutions and corresponding invariant measures (Chapter 1), a weak martingale solution for the SPDE (4)–(5) (Chapter 2), and a weak resp. strong solution of (7) (Chapter 3). These schemes are implemented, and corresponding simulations are discussed in each chapter to complement our theoretical results.

Most of the work was done when the third author (M. N.) was affiliated with the Universität Tübingen and supported by the DFG-project: ‘Long-time dynamics of the Landau-Lifshitz-Gilbert equation’ (2011–2013); partial support by the ARC Discovery grant DP120101886 is also gratefully acknowledged. Chapters 1 resp. 2 base on ideas from [95] resp. [12], which are considerably extended here. The analysis of the model in Chapter 3 uses concepts from [14].

Chapter 1

The role of noise in finite ensembles of nanomagnetic particles

We study the effect of noise on a ferromagnetic chain consisting of N spins, where the magnetization process $\mathbf{X} : \mathbb{R}^+ \times \boldsymbol{\Omega} \to (\mathbb{S}^2)^N$ evolves according to

$$\frac{\partial \mathbf{X}}{\partial t} = \mathbf{X} \times \big(\mathbf{H}_{\text{eff}} + \mathbf{H}_{\text{thm}}\big) - \alpha\, \mathbf{X} \times \big(\mathbf{X} \times \mathbf{H}_{\text{eff}}\big)\,. \tag{1.1}$$

Thermal fluctuations are usually taken into account by augmenting the effective field $\mathbf{H}_{\text{eff}} \equiv \mathbf{H}_{\text{eff}}(\mathbf{X}) = -\nabla\mathcal{E}(\mathbf{X})$ (see (2)) in the Landau-Lifshitz-Gilbert equation with an isotropic Gaussian white noise field to represent different processes involving magnon, phonon and electron interactions; cf. (1)–(3). This model has been suggested by Brown [24] to study transition states for thermally activated magnetization reversal of a *single* spin. Empirical studies indicate a different magnetization dynamics at elevated temperatures, such as enhanced damping, increased relaxation rates, and a shrinking saturation magnetization for increasing temperatures [38, 62]. However, starting with [24], most works study only a single spin and the anisotropic energy $\mathcal{E} = \mathcal{E}_{\text{ani}}$ in (2), see e.g. [82, 62], and the references therein. In particular, it has been shown formally in [62] that the Gibbs distribution (with $N = 1$)

$$\mu[\mathbf{dx}] = \frac{e^{-\frac{2\alpha}{\nu^2}\mathcal{E}(\mathbf{x})}\mathbf{dx}}{\int\limits_{(\mathbb{S}^2)^N} e^{-\frac{2\alpha}{\nu^2}\mathcal{E}(\mathbf{x})}\,\mathbf{dx}} \tag{1.2}$$

is the stationary distribution of the stochastic Landau-Lifschitz-Gilbert equation (SLLG) (1.1). Here $\mathbf{H}_{\text{thm}} = \nu\,\dot{\mathbf{W}}$, where $\mathbf{W}$ denotes an $(\mathbb{R}^3)^N$-valued Wiener process, and the Stratonovich form of the stochastic integrals is used in (1.1). On the other hand, according to statistical mechanics, a system in thermal equilibrium is described by the Maxwell-Boltzmann statistics, and consequently the stationary distribution has the form

$$\frac{e^{-\frac{1}{k_B\tau}\mathcal{E}(\mathbf{x})}\mathbf{dx}}{\int\limits_{(\mathbb{S}^2)^N} e^{-\frac{1}{k_B\tau}\mathcal{E}(\mathbf{y})}\,\mathbf{dy}}\,,$$

where k_B is the Boltzmann constant, and $\tau \geq 0$ denotes the temperature of the system. Thus we can deduce the following fluctuation–dissipation relation

$$\frac{2\alpha}{\nu^2} = \frac{1}{k_B\tau}\,, \tag{1.3}$$

which determines the constant $\nu > 0$ in terms of the temperature in (3). We recall that the basis of this relation to hold is a separation of time scales, where the relaxation time of the heat bath is assumed to be much faster than that of the spin system.

Next to thermodynamically consistent equilibria, a physically relevant quantity in the modelling of non-uniform magnetization reversal is the relaxation time, which is the characteristic time for the N-spin system to reach an equilibrium. In Theorem 1.7, we state exponentially fast relaxation of (1.1) to its unique equilibrium for finite ensembles of nanomagnetic particles, i.e. we prove ergodicity of the Gibbs distribution (1.2), with exponential rate of convergence $\rho > 0$. For simplicity, we consider only the exchange energy $\mathcal{E}_{\text{exch}}$, but the other two energies $\mathcal{E}_{\text{ani}}, \mathcal{E}_{\text{ext}}$ may easily be added and do not alter the result; cf. Remark 1.26 for the general case. The rate ρ is related to the Néel-Brown relaxation time $\tau_{NB} = \frac{1}{\rho}$ of the system, which is the subject of a vast number of physical papers, see e.g. [1, 24, 42], and others.

The technical difficulty of the result stems from the fact that the noise is degenerate if we consider the evolution of the system in $(\mathbb{R}^3)^N$. Consequently, we need to incorporate the 'sphere-property' of each single spin into the configuration space of the system, and hence study the evolution of the system on the compact Riemannian manifold $(\mathbb{S}^2)^N$. Another difficulty, when compared to results in [88], lies in the fact that the noise is multiplicative and, consequently, control-type arguments as in [88, Lemma 3.4] to establish the irreducibility of the system are not applicable. Indeed, it is well–known that in general the solution of a SDE is not a continuous function of the driving process in the topology of the space of continuous functions. To circumvent these issues, we apply instead the Girsanov theorem to find a proper representation of transition semigroups, which allows to conclude its irreducibility by the one of the corresponding Wiener process. Furthermore, we show that the energy $\mathcal{E}$ is a Lyapunov function of (1.1), and that the transition semigroup satisfies certain regularity properties. The Lyapunov property also proves to be important in Chapter 2 about the corresponding stochastic PDE (4): it appears as an energy inequality in the infinite dimensional situation and allows to show convergence of the numerical scheme which is discussed in this chapter.

The result in Theorem 1.7 does not provide a precise rate of convergence towards the equilibrium, which motivates Theorem 1.8 where we show that the exponential rate of convergence in the weaker $\mathbb{L}^2\big((\mathbb{S}^2)^N;\mu\big)$-topology is estimated from below by

$$\beta = \nu^2 N \kappa e^{-\frac{2\,\mathrm{osc}(\mathcal{E})}{k_B \tau}}, \tag{1.4}$$

where N is the number of spins, κ is the spectral gap of the Laplace-Beltrami operator on the sphere $\mathbb{S}^2$, and $\mathrm{osc}(\mathcal{E}) = \sup \mathcal{E} - \inf \mathcal{E}$. Notice that β is an increasing function of the temperature τ, and hence a decreasing function with respect to the damping parameter α by (1.3), which contradicts the intuition that the more damping we put on the system the faster becomes the convergence to equilibrium (see Figure 1.1 for numerical simulations of the rate of convergence).

We develop several strategies to approximate problem (1.1). We start in Subsection 1.2.4 with an 'outer approximation' in $(\mathbb{R}^3)^N$ with the help of the Ginzburg-Landau penalization term. This approach is motivated by numerical demands where the restriction to schemes with sphere-valued solutions requires to construct non-standard discretizations, and rules out many well-known (high-order) discretizations; another motivation is to study the impact of stochastic forcing onto approximately sphere-valued solutions, including the asymptotic regime $t \to \infty$. A main observation is that a relaxation of the sphere property of solutions has to go together with a modification of the noise in order to ensure proper (approximate) long-time dynamics. For this purpose, we compare two approximate problems (cases $\delta = 0$ and $\delta > 0$ in system (1.55)). In the first one the noise of the penalized system is the same as for the limiting system. The second one has an additional additive noise. Our results show that the system (1.55) with conservative noise ($\delta = 0$) behaves better on the finite time interval. The problem inherits a natural energy inequality (Proposition 1.37), the solution stays in the unit ball, and converges on each finite time interval to the solution of the non-penalized system if initial data are sphere-valued (Theorem 1.10). The system allows for several natural choices of a configuration space but, as discussed in the Remark 1.33, neither of them justifies the strong irreducibility property; thus, ergodicity of the system is not clear. The modification of the noise (case $\delta > 0$) improves control over the long-time dynamics of the system (1.55): we are then able to show the geometric ergodicity of the system, i.e. the system exponentially converges to the unique invariant measure (Theorem 1.9). Furthermore, if the additive noise is sufficiently small, the solution converges to the solution of the non-penalized system (Corollary 1.39). Thus, we see that the behavior of the system is very sensitive with respect to the type of the used noise, and convergence to the limiting system for finite times is guaranteed only for sufficiently small $\delta > 0$. These issues motivate a second approximation strategy, which uses discretization in time of the system (1.1) where the geometric constraint is preserved at each step of the simulation. In Section 1.3, we present two numerical schemes to simulate system (1.1); as it is well-known, naive time discretizations of SDEs may easily loose not only the geometric rate of convergence, but overall asymptotic convergence properties; see Subsection 1.4.2 for computational evidence. The first scheme (Scheme 1.11) is nonlinear implicit and yields an $(\mathbb{S}^2)^N$-valued discrete Markov chain, which inherits the Lyapunov function property from the limiting system. As a consequence, we may show geometric exponential ergodicity of the system with the same method which is used to verify Theorem 1.7. Furthermore, we show local in time strong rates of convergence towards the continuous process for corresponding iterates. This result, together with the geometric ergodicity property for the limiting equation (1.1) from Theorem 1.7 implies convergence of invariant measures from the numerical scheme to the Gibbs measure (1.2), as a consequence of the general results of Shardlow & Stuart in [101]. The second scheme (Scheme 1.16) is linear implicit, and hence computationally more efficient. Iterates of this discrete Markov chain are again $(\mathbb{S}^2)^N$-valued, but the discrete Lyapunov condition is not available any more. As a consequence, tools for the first scheme do not apply to verify geometric ergodicity. However, we are able to show convergence of invariant measures to the unique time-asymptotic Gibbs distribution (1.2) of (1.1) for a vanishing discretization parameter by the perturbation result of Shardlow & Stuart in [101]. We also show an optimal rate of weak convergence for finite times. These results are complemented by computational studies in Section 1.4, where evidence is provided that numerical schemes may fail to approximate proper long-time dynamics if the sphere-property of iterates is not accounted for; another series of experiments studies the effect of penalization. Furthermore, computational studies with different projection methods are reported which are related to penalization concepts, and are often used to solve the related deterministic problem (LLG).

The chapter is organized as follows: in Section 1.1, we collect background material on ergodic properties of Markov chains, which is used in Section 1.2 to verify exponential ergodicity of the invariant Gibbs measure (1.2) for (1.1). Time discretization schemes, and penalization methods to approximate (1.1) are studied in Section 1.3. Computational studies are reported in Section 1.4.

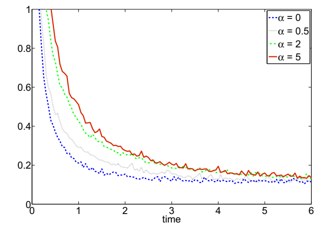

Figure 1.1. Scheme 1.11: Speed of convergence to the stationary distribution of (1) for different values of the parameter $\alpha \in \{0, 0.5, 2, 5\}$.

1.1 Preliminaries

We collect some results on geometric ergodicity of Markov chains in Subsection 1.1.1. In Subsection 1.1.2 we recall strategies to conclude ergodicity with rates for solutions of SDEs. Subsection 1.1.3 surveys different convergent discretizations of the deterministic LLG equation.

1.1.1 Geometric ergodicity of Markov chains

Here we recall the Meyn-Tweedie theory [89]. We follow the presentation from [88].

Let $\mathbb{X} \subset \mathbb{R}^d$ be a smooth Riemannian manifold, and $\mathbb{T}$ be either $\mathbb{R}^+$ or $\mathbb{Z}^+$. Let $\mathbf{X} := \{\mathbf{X}(t);\, t \in \mathbb{T}\}$ be a Markov process (or a Markov chain) on a state space $\big(\mathbb{X}, \mathscr{B}(\mathbb{X})\big)$, where $\mathscr{B}(\mathbb{X})$ is the σ-field of Borel subsets of $\mathbb{X}$. Let

$$P(t, \mathbf{x}, \mathscr{A}) := \mathbb{P}\big[\{\mathbf{X}(t) \in \mathscr{A};\, \mathbf{X}(0) = \mathbf{x}\}\big] \qquad \forall\, t \in \mathbb{T} \quad \forall\, \mathbf{x} \in \mathbb{X} \quad \forall\, \mathscr{A} \in \mathscr{B}(\mathbb{X})$$

be the transition kernel of the process $\mathbf{X}$. Let $\mathbb{B}_b(\mathbb{X})$ denote the set of Borel measurable bounded real-valued functions. Define the semigroup $P_t : \mathbb{B}_b(\mathbb{X}) \to \mathbb{B}_b(\mathbb{X})$ for $t \in \mathbb{T}$, which is associated with the process $\mathbf{X}$ by its values on the indicator function of Borel subsets of $\mathbb{X}$:

$$P_t \mathbb{1}_{\mathscr{A}}(\mathbf{x}) := P(t, \mathbf{x}, \mathscr{A}) \qquad \forall\, t \in \mathbb{T} \quad \forall\, \mathbf{x} \in \mathbb{X} \quad \forall\, \mathscr{A} \in \mathscr{B}(\mathbb{X})\,.$$

If $\mathbb{T} := \mathbb{R}^+$ then we denote by $\mathcal{L}$ the infinitesimal generator of the semigroup $\{P_t;\, t \in \mathbb{T}\}$. Let $B_\delta(\mathbf{x}) \subset \mathbb{R}^d$ denote a closed ball around $\mathbf{x}$ of radius $\delta > 0$.

Definition 1.1. A Markov process (or chain) $\mathbf{X}$ with transition probability $P(t, \cdot, \cdot)$ is weakly irreducible iff there exists a compact set $\mathscr{C} \subset \mathbb{X}$ with non-empty interior such that for some $\mathbf{y}^* \in \mathbf{Int}(\mathscr{C})$, for any $\delta > 0$, there exists $t \equiv t(\delta, \mathbf{y}^*) \in \mathbb{T}$ such that $P\big(t, \mathbf{x}, B_\delta(\mathbf{y}^*)\big) > 0$ for all $\mathbf{x} \in \mathscr{C}$. A Markov process (or chain) $\mathbf{X}$ with transition probability $P(t, \cdot, \cdot)$ is strongly irreducible iff for any $\mathbf{y} \in \mathbb{X}$, $t > 0$ and any open set $\mathscr{A} \subset \mathscr{B}(\mathbb{X})$ we have $P(t, \mathbf{y}, \mathscr{A}) > 0$.

Definition 1.2. A Markov process (or chain) $\mathbf{X}$ with transition probability $P(t, \cdot, \cdot)$ is regular iff the transition kernel has a nonnegative density $\big\{p(t, \mathbf{x}, \mathbf{y});\, t \in \mathbb{T},\, \mathbf{x}, \mathbf{y} \in \mathbb{X}\big\}$, such that

$$P(t, \mathbf{x}, \mathscr{A}) = \int_{\mathscr{A}} p(t, \mathbf{x}, \mathbf{y})\, \mathbf{dy} \qquad \forall\, t \in \mathbb{T} \quad \forall\, \mathbf{x} \in \mathbb{X} \quad \forall\, \mathscr{A} \in \mathscr{B}(\mathbb{X})\,,$$

where $p(t, \cdot, \cdot) \in \mathbb{C}(\mathbb{X}^2)$ for any $t \in \mathbb{T}$.

Definition 1.3. A Markov process (or chain) $\mathbf{X}$ satisfies the minorization condition iff there exist $s \in \mathbb{T}$, an $\eta > 0$, a compact set $\mathscr{C} \subset \mathbb{X}$, and a probability measure ν on $\big(\mathscr{C}, \mathscr{B}(\mathscr{C})\big)$ such that

$$P(s, \mathbf{x}, \mathscr{A}) \geq \eta\, \nu[\mathscr{A}] \qquad \forall\, \mathscr{A} \in \mathscr{B}(\mathscr{C}) \quad \forall\, \mathbf{x} \in \mathscr{C}\,.$$

Lemma 1.4. *If a Markov process* $\mathbf{X}$ *is weakly irreducible and regular then it satisfies the minorization condition.*

Proof. Step 1. Local discussion. By the irreducibility assumption, there exist a compact set $\mathscr{C} \subset \mathbb{X}$, and $\mathbf{y}^* \in \mathsf{Int}(\mathscr{C})$ such that for any $\delta > 0$ there exists a time $t = t(\delta) > 0$ such that $P\big(t, \mathbf{y}^*, B_\delta(\mathbf{y}^*)\big) > 0$. Then we can find a possibly smaller neighborhood $B_{\delta_1}(\mathbf{y}^*) \subset \mathscr{C}$, and $t_1 > 0$ such that $P\big(t_1, \mathbf{y}^*, B_{\delta_1}(\mathbf{y}^*)\big) > 0$. Indeed, since $\mathbf{y}^*$ is in the interior of $\mathscr{C}$, there exists a $\gamma > 0$ such that $B_\gamma(\mathbf{y}^*) \subset \mathscr{C}$ and we may take $\delta_1 := \frac{\gamma}{2}$ and $t_1 := t(\frac{\gamma}{2})$. The existence of a density implies that there exists $\mathbf{z}^* \in B_{\delta_1}(\mathbf{y}^*) \subset \mathscr{C}$, and some $\epsilon > 0$ such that

$$p(t_1, \mathbf{y}^*, \mathbf{z}^*) \geq 2\epsilon > 0\,.$$

By the regularity assumption, there exist neighborhoods $B_r(\mathbf{y}^*), B_r(\mathbf{z}^*) \subset \mathscr{C}$ such that

$$p(t_1, \mathbf{y}, \mathbf{z}) \geq \epsilon \qquad \forall\, \mathbf{y} \in B_r(\mathbf{y}^*) \quad \forall\, \mathbf{z} \in B_r(\mathbf{z}^*)\,.$$

Hence we have that

$$\begin{aligned} P(t_1, \mathbf{y}, \mathscr{A}) &= \int\limits_{\mathscr{A}} p(t_1, \mathbf{y}, \mathbf{z})\, \mathbf{dz} \geq \int\limits_{\mathscr{A} \cap B_r(\mathbf{z}^*)} p(t_1, \mathbf{y}, \mathbf{z})\, \mathbf{dz} \\ &\geq \epsilon\, \mathsf{Leb}\big[\mathscr{A} \cap B_r(\mathbf{z}^*)\big] \qquad \forall\, \mathbf{y} \in B_r(\mathbf{y}^*) \quad \forall\, \mathscr{A} \in \mathscr{B}(\mathbb{X})\,, \end{aligned} \tag{1.5}$$

where $\mathsf{Leb} : \mathscr{B}(\mathbb{X}) \to \mathbb{R}^+$ is the Riemannian volume measure on $\mathbb{X}$.

Step 2. Global discussion in $\mathscr{C}$. By the irreducibility assumption, there exists a time $t_2 > 0$ such that

$$P\big(t_2, \mathbf{x}, B_r(\mathbf{y}^*)\big) > 0 \qquad \forall\, \mathbf{x} \in \mathscr{C}\,.$$

Furthermore, by the regularity assumption, the function $P\big(t_2, \cdot, B_r(\mathbf{y}^*)\big)$ is continuous on the compact set $\mathscr{C}$. Thus,

$$\min_{\mathbf{x} \in \mathscr{C}} P\big(t_2, \mathbf{x}, B_r(\mathbf{y}^*)\big) \geq \gamma_1 > 0\,.$$

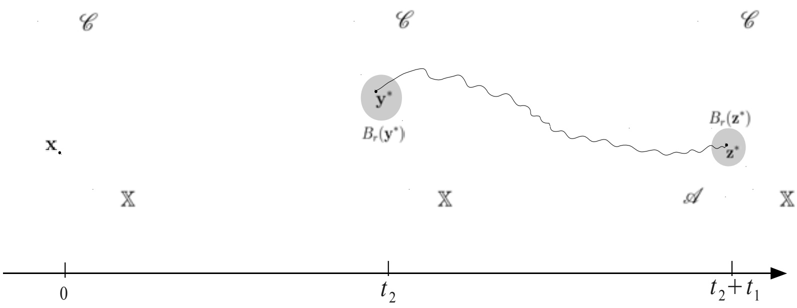

Figure 1.2. Illustration of the application of the Kolmogorov-Chapman equation in formula (1.6)

Consequently, by the Chapman-Kolmogorov equation and (1.5) we find (see Figure 1.2)

$$\begin{aligned} P\big(t_1+t_2,\mathbf{x},\mathscr{A}\big) &\geq \int\limits_{B_r(\mathbf{y}^*)} p(t_2,\mathbf{x},\mathbf{w}) P(t_1,\mathbf{w},\mathscr{A})\,\mathbf{dw} \\ &\geq \epsilon\,\mathrm{Leb}\big[\mathscr{A}\cap B_r(\mathbf{z}^*)\big] \int\limits_{B_r(\mathbf{y}^*)} p(t_2,\mathbf{x},\mathbf{w})\,\mathbf{dw} \\ &\geq \gamma_1\epsilon\,\mathrm{Leb}\big[\mathscr{A}\cap B_r(\mathbf{z}^*)\big] \qquad (1.6) \\ &= \gamma_1\epsilon\,\mathrm{Leb}\big[B_r(\mathbf{z}^*)\big]\frac{\mathrm{Leb}\big[\mathscr{A}\cap B_r(\mathbf{z}^*)\big]}{\mathrm{Leb}[B_r(\mathbf{z}^*)]}\,. \end{aligned}$$

We may now put $\eta := \gamma_1\epsilon\,\mathrm{Leb}\big[B_r(\mathbf{z}^*)\big]$, and $\nu := \frac{\mathrm{Leb}[\cdot\cap B_r(\mathbf{z}^*)]}{\mathrm{Leb}[B_r(\mathbf{z}^*)]}$. □

Definition 1.5. The mapping $V : \mathbb{X} \to [1,\infty)$ is a Lyapunov function for the Markov chain $\{\mathbf{X}^j\}_{j=0}^{\infty}$ if there exist numbers $\alpha \in (0,1)$, and $\beta \in [0,\infty)$ such that

$$\mathbb{E}\big[V(\mathbf{X}^{j+1})\big|\,\sigma\big(\{\mathbf{X}^0,\mathbf{X}^1,\ldots,\mathbf{X}^j\}\big)\big] \leq \alpha V(\mathbf{X}^j)+\beta\,,$$

and V is unbounded if the set $\mathbb{X}$ is unbounded, i.e.,

$$\lim_{\mathrm{dist}(\mathbf{y},\mathbf{x})\to\infty} V(\mathbf{y}) = \infty \qquad \forall\,\mathbf{x}\in\mathbb{X}\,.$$

Furthermore, we assume that level sets $\{\mathbf{y}\in\mathbb{X};\, V(\mathbf{y})\leq a\}$, $a>1$ are either compact or contain compact subsets such that their union (over a) is $\mathbb{X}$.

If $\mathbb{T}=\mathbb{R}^+$, we can reformulate Definition 1.5 in terms of the infinitesimal generator $\mathcal{L}$ of the semigroup $\{P_t;\, t\in\mathbb{T}\}$ associated with $\mathbf{X}$.

Definition 1.6. A mapping $V : \mathbb{X} \to [1, \infty)$ is a Lyapunov function for the Markov process $\mathbf{X}$ with generator $\mathcal{L}$ if there exist constants $0 < c, d < \infty$, such that

$$\mathcal{L}V \leq -cV + d\,, \tag{1.7}$$

and V is unbounded if the set $\mathbb{X}$ is unbounded, i.e.,

$$\lim_{\mathrm{dist}(\mathbf{y},\mathbf{x})\to\infty} V(\mathbf{y}) = \infty \qquad \forall \mathbf{x} \in \mathbb{X}\,.$$

Furthermore, we assume that level sets $\{\mathbf{y} \in \mathbb{X};\, V(\mathbf{y}) \leq a\}$, $a > 1$ are either compact or contain compact subsets such that their union (over a) is $\mathbb{X}$.

Below, we collect a series of propositions which describe the behavior of a Markov process $\mathbf{X}$ under the assumption that there exists a Lyapunov function. First we show that the constant β in Definition 1.5 can be replaced by zero outside of a certain compact subset of $\mathbb{X}$ at the expense of an increased constant α. Let $\mathbb{1}_{\mathscr{C}} : \mathbb{X} \to \{0, 1\}$ denote the characteristic function of $\mathscr{C}$.

Proposition 1.7. *Assume that $\{\mathbf{X}^j\}_{j=0}^{\infty}$ is a Markov chain with Lyapunov function $V : \mathbb{X} \to [1, \infty)$. Let $\gamma \in (\alpha, 1)$, and $s \geq 1$, and denote*

$$\mathscr{C}(s, \gamma) := \Big\{\mathbf{x} \in \mathbb{X};\, V(\mathbf{x}) \leq \frac{s\beta}{\gamma - \alpha}\Big\}\,.$$

Then

$$\mathbb{E}\Big[V(\mathbf{X}^{j+1})\big|\sigma\big(\{\mathbf{X}^0, \mathbf{X}^1, \ldots, \mathbf{X}^j\}\big)\Big] \leq \gamma V(\mathbf{X}^j) + s\beta\, \mathbb{1}_{\mathscr{C}(s,\gamma)}(\mathbf{X}^j)\,.$$

Proof. Fix $j \geq 0$. The result is evident if $\mathbf{X}^j \in \mathscr{C}(s, \gamma)$. Otherwise, $V(\mathbf{X}^j) > \frac{s\beta}{\gamma-\alpha}$. Consequently $\gamma V(\mathbf{X}^j) > \alpha V(\mathbf{X}^j) + s\beta \geq \alpha V(\mathbf{X}^j) + \beta$, and the result follows. □

The next results asserts polynomial convergence to 0 of a Lyapunov function as time converges to infinity. Let $a \wedge b := \min\{a, b\}$.

Proposition 1.8. *Let $\{\mathbf{X}^j\}_{j=0}^{\infty}$ be a Markov process with Lyapunov function $V : \mathbb{X} \to [1, \infty)$ (with parameters α and β), that $\mathcal{F}_j := \sigma\big(\{\mathbf{X}^0, \mathbf{X}^1, \ldots, \mathbf{X}^j\}\big), j \geq 0$, J is a stopping time, $\gamma \in (\alpha, 1)$, and $\mathscr{C} := \mathscr{C}(2, \gamma)$. Then there exists some $C > 0$ such that*

$$\begin{aligned}
\mathbb{E}\big[V(\mathbf{X}^j)\mathbb{1}_{\{J>j\}}\big] &\leq \mathbb{E}\big[V(\mathbf{X}^j)\mathbb{1}_{\{J\geq j\}}\big] \\
&\leq C\gamma^j\Big(\mathbb{E}\big[V(\mathbf{X}^0)\big] + \mathbb{E}\Big[\sum_{l=1}^{j\wedge J} \gamma^{-l}\mathbb{1}_{\mathscr{C}}(\mathbf{X}^{l-1})\Big]\Big) \\
&\leq C\Big(\mathbb{E}\big[V(\mathbf{X}^0)\big] + 1\Big)\frac{\gamma^j}{1-\gamma} \qquad (j \geq 1)\,.
\end{aligned}$$

Proof. The first inequality is trivial. The third inequality immediately follows from the following elementary estimate

$$\gamma^j \sum_{l=1}^{j \wedge J} \gamma^{-l} \mathbb{1}_{\mathscr{C}}(\mathbf{X}^{l-1}) \leq \sum_{l=1}^{j} \gamma^{j-l} \leq \frac{1}{1-\gamma}\,.$$

To show the second inequality we consider a finite differences representation for the function $F : \mathbb{X} \times \mathbb{T} \to \mathbb{R}^+$, defined by $F(\mathbf{X}, j) := \gamma^{-j} V(\mathbf{X})$. We have

$$\begin{aligned} F\big(\mathbf{X}^{J\wedge j}, J \wedge j\big) &= F\big(\mathbf{X}^0, 0\big) + \sum_{l=1}^{J\wedge j} \Big(F\big(\mathbf{X}^l, l\big) - F\big(\mathbf{X}^{l-1}, l-1\big) \Big) \\ &= F\big(\mathbf{X}^0, 0\big) + \sum_{l=1}^{j} \mathbb{1}_{\{J > l-1\}} \Big(F\big(\mathbf{X}^l, l\big) - F\big(\mathbf{X}^{l-1}, l-1\big) \Big)\,. \end{aligned}$$

Taking the expectation and applying the tower property for conditional expectation leads to

$$\begin{aligned} \mathbb{E}\Big[F\big(\mathbf{X}^{J\wedge j}, J \wedge j\big)\Big] &= \mathbb{E}\big[F\big(\mathbf{X}^0, 0\big)\big] + \\ &+ \sum_{l=1}^{j} \mathbb{E}\Big[\mathbb{E}\big[\mathbb{1}_{\{J > l-1\}} \Big(F\big(\mathbf{X}^l, l\big) - F\big(\mathbf{X}^{l-1}, l-1\big) \Big) \big| \mathcal{F}_{l-1}\big]\Big]\,. \end{aligned}$$

Notice that the event $\{J > l-1\}$ is $\mathcal{F}_{l-1}$–measurable, and $F\big(\mathbf{X}^0, 0\big) = V(\mathbf{X}^0)$. Hence,

$$\begin{aligned} \mathbb{E}\big[F\big(\mathbf{X}^{J\wedge j}, J \wedge j\big)\big] &= \mathbb{E}\big[V(\mathbf{X}^0)\big] + \\ &+ \sum_{l=1}^{j} \mathbb{E}\Big[\mathbb{1}_{\{J > l-1\}} \mathbb{E}\big[F(\mathbf{X}^l, l) - F(\mathbf{X}^{l-1}, l-1) \big| \mathcal{F}_{l-1}\big]\Big]\,. \end{aligned} \tag{1.8}$$

We apply Proposition 1.7 with $s = 2$ to conclude that

$$\begin{aligned} \mathbb{E}\big[F(\mathbf{X}^l, l) | \mathcal{F}_{l-1}\big] &= \gamma^{-l}\, \mathbb{E}\big[V(\mathbf{X}^l) \big| \mathcal{F}_{l-1}\big] \\ &\leq \gamma^{-l} \Big(\gamma V\big(\mathbf{X}^{l-1}\big) + 2\beta \mathbb{1}_{\mathscr{C}}(\mathbf{X}^{l-1}) \Big) \\ &= F\big(\mathbf{X}^{l-1}, l-1\big) + 2\gamma^{-l} \beta \mathbb{1}_{\mathscr{C}}(\mathbf{X}^{l-1})\,. \end{aligned} \tag{1.9}$$

We may combine identity (1.8) with inequality (1.9) to deduce that

$$\mathbb{E}\big[F(\mathbf{X}^{J\wedge j}, J \wedge j)\big] \leq \mathbb{E}\big[V(\mathbf{X}^0)\big] + 2\beta\, \mathbb{E}\Big[\sum_{l=1}^{j} \gamma^{-l} \mathbb{1}_{\{J > l-1\}} \mathbb{1}_{\mathscr{C}}(\mathbf{X}^{l-1})\Big]\,.$$

The result then follows from the estimate

$$\mathbb{E}\big[F(\mathbf{X}^{J\wedge j}, J \wedge j)\big] \geq \mathbb{E}\Big[F\big(\mathbf{X}^{J\wedge j}, J \wedge j\big) \mathbb{1}_{J \geq j}\Big] = \gamma^{-j}\, \mathbb{E}\Big[V(\mathbf{X}^j) \mathbb{1}_{\{J \geq j\}}\Big]\,.$$

□

We obtain the following estimates on the first return time $\tau_{\mathscr{C}} := \min\{j > 0; \mathbf{X}^j \in \mathscr{C}\}$ to the set $\mathscr{C}$.

Corollary 1.9. *Let $\{\mathbf{X}^j\}_{j=0}^{\infty}$ be a Markov process with Lyapunov function $V : \mathbb{X} \to [1,\infty)$, and $\mathscr{C} := \mathscr{C}(2,\gamma) \subset \mathbb{X}$ for $\gamma \in (\alpha,1)$. Then there exists $C > 0$ such that*

$$\begin{aligned} &\text{(i)} \qquad \mathbb{P}\big[\{\tau_{\mathscr{C}} > j\}\big] \le C\Big(\mathbb{E}\big[V(\mathbf{X}^0)\big] + 1\Big)\gamma^j \qquad (j > 0)\,, \\ &\text{(ii)} \qquad \mathbb{E}\big[\gamma^{-\tau_{\mathscr{C}}}\big] \le C\Big(\mathbb{E}\big[V(\mathbf{X}^0)\big] + 1\Big)\,. \end{aligned}$$

Proof. We apply the second inequality of Proposition 1.8 with stopping time $J = \tau_{\mathscr{C}}$. The definition of $\tau_{\mathscr{C}}$ implies that

$$\sum_{l=1}^{j\wedge\tau_{\mathscr{C}}} \gamma^{j-l}\mathbb{1}_{\mathscr{C}}(\mathbf{X}^{l-1}) = \gamma^{j-1}\mathbb{1}_{\mathscr{C}}(\mathbf{X}^0)\,.$$

Furthermore,

$$\mathbb{E}\big[V(\mathbf{X}^j)\mathbb{1}_{\tau_{\mathscr{C}}>j}\big] \ge \frac{2\beta}{\gamma-\alpha}\mathbb{E}[\mathbb{1}_{\tau_{\mathscr{C}}>j}] = \frac{2\beta}{\gamma-\alpha}\mathbb{P}\big[\{\tau_{\mathscr{C}} > j\}\big]\,.$$

Hence assertion (i) follows.

To show assertion (ii), we observe that

$$\mathbb{E}\big[\gamma^{-\tau_{\mathscr{C}}}\big] = \sum_{l=1}^{\infty}\gamma^{-l}\,\mathbb{P}\big[\{\tau_{\mathscr{C}} = l\}\big] \le \sum_{l=1}^{\infty}\gamma^{-l}\,\mathbb{P}\big[\{\tau_{\mathscr{C}} > l-1\}\big]\,.$$

Since $\gamma \in (\alpha,1)$ we can apply (i) with $\gamma' \in (\alpha,\gamma)$ to conclude that there exists $\kappa_1 > 0$ such that

$$\mathbb{E}\big[\gamma^{-\tau_{\mathscr{C}}}\big] \le \kappa_1\Big(\mathbb{E}\big[V(\mathbf{X}^0)\big] + 1\Big)\sum_{l=1}^{\infty}\left(\frac{\gamma'}{\gamma}\right)^l\,.$$

□

The previous Corollary can be generalized to estimate the time $\tau_r(\mathscr{C}) := \tau_{[r]}(\mathscr{C})$, $r \ge 0$ of the $[r]$-th visit to the set $\mathscr{C}$ (put $\tau_0 := 0$).

Corollary 1.10. *Assume that $\{\mathbf{X}^j\}_{j=0}^{\infty}$ is a Markov process with Lyapunov function $V : \mathbb{X} \to [1,\infty)$, and $\mathscr{C} := \mathscr{C}(2,\gamma) \subset \mathbb{X}$ for $\gamma \in (\alpha,1)$. There exists a positive constant $C \equiv C(\mathscr{C}, V)$ such that*

$$\mathbb{E}\big[\gamma^{-\tau_r(\mathscr{C})}\big] \le C^r\Big(\sup_{\mathscr{C}} V + 1\Big)^{r-1}\Big(\mathbb{E}\big[V(\mathbf{X}^0)\big] + 1\Big)\,.$$

Proof. By definition, we can assume that $r \in \mathbb{N}$. Denote $A(r) := \mathbb{E}\big[\gamma^{-\tau_r(\mathscr{C})}\big]$ for $r \in \mathbb{N}$. We have by elementary properties of the conditional expectation that for $r > 1$

$$\begin{aligned} A(r) &= \mathbb{E}\big[\gamma^{-\sum_{l=1}^{r}[\tau_l(\mathscr{C})-\tau_{l-1}(\mathscr{C})]}\big] \\ &= \mathbb{E}\Big[\mathbb{E}\big[\gamma^{-\sum_{l=1}^{r}[\tau_l(\mathscr{C})-\tau_{l-1}(\mathscr{C})]}\big|\mathcal{F}_{\tau_{r-1}(\mathscr{C})}\big]\Big] \\ &= \mathbb{E}\Big[\gamma^{-\sum_{l=1}^{r-1}[\tau_l(\mathscr{C})-\tau_{l-1}(\mathscr{C})]}\,\mathbb{E}\big[\gamma^{-[\tau_r(\mathscr{C})-\tau_{r-1}(\mathscr{C})]}\big|\mathcal{F}_{\tau_{r-1}(\mathscr{C})}\big]\Big] \\ &= \mathbb{E}\Big[\gamma^{-\sum_{l=1}^{r-1}[\tau_l(\mathscr{C})-\tau_{l-1}(\mathscr{C})]}\mathbb{E}\big[\gamma^{-\tau_1(\mathscr{C})}\big|\mathbf{X}^{\tau_{r-1}(\mathscr{C})}\big]\Big] \\ &= A(r-1)\,\mathbb{E}\big[\gamma^{-\tau_1(\mathscr{C})}\big|\mathbf{X}^{\tau_{r-1}(\mathscr{C})}\big]\,, \end{aligned}$$

and the result follows from Corollary 1.9, (ii). □

Corollary 1.11. *Assume that* $\{\mathbf{X}^j\}_{j=0}^{\infty}$ *is a Markov process with Lyapunov function* $V : \mathbb{X} \to [1, \infty)$. *Then there exists an invariant probability measure.*

Proof. We apply Proposition 1.8 with a constant stopping time $J = j$ to conclude that

$$\sup_{j \geq 1} \mathbb{E}\big[V(\mathbf{X}^j)\big] < \infty\,.$$

Now the existence of invariant measure follows from a standard argument. Indeed, by the Chebyshev inequality, and compactness of the level sets of function V, the sequence of measures

$$\mu_n := \frac{1}{n}\sum_{l=1}^{n}\mathbb{P}\Big[\{\mathbf{X}^l \in \cdot\}\Big] \qquad (n \in \mathbb{N})\,,$$

is tight (Chapter 2, Definition 2.16). Therefore, by the Prohorov Theorem (Chapter 2, Theorem 2.2) there exists a convergent subsequence to the measure μ. Consequently, the measure μ is finite and invariant. Normalizing it if necessary, we obtain an invariant probability measure. □

Theorem 1.1. *Let* $\mathbf{X}$ *be a Markov process (or chain) with transition kernel* P. *Fix* $T > 0$. *Let* $\{\mathbf{X}^j\}_{j=0}^{\infty}$, *with* $\mathbf{X}^j := \mathbf{X}(jT)$ *be an embedded Markov chain with transition kernel* $P(T)$. *Assume that the Markov chain* $\{\mathbf{X}^j\}_{j=0}^{\infty}$ *has a Lyapunov function* $V : \mathbb{X} \to [1, \infty)$ *(with parameters* α *and* β*), and satisfies the minorization condition with the set*

$$\mathscr{C} := \mathscr{C}(2, \gamma) = \Big\{\mathbf{x};\, V(\mathbf{x}) \leq \frac{2\beta}{\gamma - \alpha}\Big\}$$

for some $\gamma \in (\sqrt{\alpha}, 1)$ *and parameter* η. *Then there exist a unique invariant measure* μ, *and constants* $r := r(\gamma) \in (0, 1)$, $\kappa := \kappa(\gamma) \in (0, \infty)$, *such that*

$$\Big|\mathbb{E}^{\mathbf{X}^0}\left[f(\mathbf{X}^j)\right] - \int_{\mathbb{X}} f\,\mathrm{d}\mu\Big| \leq \kappa r^j \left(\mathbb{E}\big[V(\mathbf{X}^0)\big] + 1\right) \qquad \forall \text{ measurable } f : |f| \leq V\,.$$

Proof. Step 1: Construction of an equivalent Markov chain with atomic structure. In this step we will construct a Markov chain $\{\mathbf{Z}^j\}_{j=0}^{\infty}$ with the same transition kernel as $\{\mathbf{X}^j\}_{j=0}^{\infty}$, which has an atomic structure, i.e. there exists a subset of the configuration space of non-zero probability such that the transition kernel of the Markov chain is the same for all points of the subset.

The minorization condition implies that we can define a new transition kernel as follows:

$$\widetilde{P}(\mathbf{x},\mathscr{A}) := \begin{cases} P(\mathbf{x},\mathscr{A}) & (\mathbf{x}\notin\mathscr{C})\,, \\ \frac{P(\mathbf{x},\mathscr{A})-\eta\nu[\mathscr{A}]}{1-\eta} & (\mathbf{x}\in\mathscr{C})\,. \end{cases}$$

Let

$$\widetilde{\mathbf{X}}^{j+1} = \widetilde{\mathbf{h}}(\widetilde{\mathbf{X}}^j,\widetilde{\omega}) \qquad (\widetilde{\omega}\in\mathbf{\Omega})$$

be the corresponding Markov chain with transition kernel $\widetilde{P}$. Define the new Markov chain

$$\mathbf{Z}^{j+1} := \mathbf{h}(\mathbf{Z}^j,\omega_j) \tag{1.10}$$

where $\omega_j := \big(\widetilde{\omega}_j,\phi_j,\boldsymbol{\xi}_j\big)$ are i.i.d. random variables,

$$\mathbf{h}(\mathbf{x},\omega) := \mathbb{1}_{\mathscr{C}}(\mathbf{x})\Big[\phi\widetilde{\mathbf{h}}(\mathbf{x},\widetilde{\omega}) + (1-\phi)\boldsymbol{\xi}\Big] + \big(1-\mathbb{1}_{\mathscr{C}}(\mathbf{x})\big)\widetilde{\mathbf{h}}\big(\mathbf{x},\widetilde{\omega}\big)\,,$$

and ω_1 is distributed as $\omega := \big(\widetilde{\omega},\phi,\boldsymbol{\xi}\big)$, where ϕ, $\boldsymbol{\xi}$ are random variables which are independent from $\widetilde{\omega}$, such that $\mathbb{P}\big[\phi=1\big] = 1-\eta$, $\mathbb{P}\big[\phi=0\big] = \eta$, and $\boldsymbol{\xi}$ is distributed according to ν. Elementary calculations then imply that the transition kernel of the chain $\{\mathbf{Z}^j\}_{j=0}^{\infty}$ is the same as of the initial Markov chain $\{\mathbf{X}^j\}_{j=0}^{\infty}$.

Step 2: A coupling argument. Let $\{\mathbf{Z}^j\}_{j=0}^{\infty}$ and $\{\widetilde{\mathbf{Z}}^j\}_{j=0}^{\infty}$ be two realizations of the Markov chain (1.10) with the same random variables $(\{\phi_j\}_{j=0}^{\infty}, \{\boldsymbol{\xi}_j\}_{j=0}^{\infty})$ and independent random variables $\{\widetilde{\omega}_j^1\}_{j=0}^{\infty}, \{\widetilde{\omega}_j^2\}_{j=0}^{\infty}$. Denote $\mathcal{F}_j := \sigma\{(\mathbf{Z}^l,\widetilde{\mathbf{Z}}^l);\, l\leq j\}$ for $j\geq 0$. Our aim is to estimate the difference

$$\Big|\mathbb{E}\big[f(\mathbf{Z}^j)\big] - \mathbb{E}\big[f(\widetilde{\mathbf{Z}}^j)\big]\Big|\,,$$

for measurable f such that $|f|\leq V$. Without loss of generality we can assume that f is a non-negative function; otherwise, we may decompose f as a difference of non-negative functions. We define the coupling time by

$$\psi := \inf_{j\geq 0}\big\{(\mathbf{Z}^j,\widetilde{\mathbf{Z}}^j)\in\mathscr{C}\times\mathscr{C};\, \phi_j = 0\big\}\,.$$

Notice that

$$\mathbb{E}\big[f(\mathbf{Z}^j)\big] = \mathbb{E}\big[f(\mathbf{Z}^j)\mathbb{1}_{j\geq\psi}\big] + \mathbb{E}\big[f(\mathbf{Z}^j)\mathbb{1}_{j<\psi}\big]\,,$$

and since

$$\mathbb{E}\big[f(\mathbf{Z}^j)\mathbb{1}_{j\geq\psi}\big] = \mathbb{E}\big[f(\widetilde{\mathbf{Z}}^j)\mathbb{1}_{j\geq\psi}\big] \leq \mathbb{E}\big[f(\widetilde{\mathbf{Z}}^j)\big]\,,$$

and $f \leq V$, we conclude that

$$\mathbb{E}\big[f(\mathbf{Z}^j)\big] - \mathbb{E}\big[f(\widetilde{\mathbf{Z}}^j)\big] \leq \mathbb{E}\big[V(\mathbf{Z}^j)\mathbb{1}_{j<\psi}\big]\,.$$

We reverse the roles of $\mathbf{Z}$ and $\widetilde{\mathbf{Z}}$ to deduce that

$$\mathbb{E}\big[f(\widetilde{\mathbf{Z}}^j)\big] - \mathbb{E}\big[f(\mathbf{Z}^j)\big] \leq \mathbb{E}\big[V(\widetilde{\mathbf{Z}}^j)\mathbb{1}_{j<\psi}\big]\,.$$

Hence we can conclude that

$$\Big|\mathbb{E}\big[f(\widetilde{\mathbf{Z}}^j)\big] - \mathbb{E}\big[f(\mathbf{Z}^j)\big]\Big| \leq \mathbb{E}\left[\big(V(\mathbf{Z}^j) + V(\widetilde{\mathbf{Z}}^j)\big)\mathbb{1}_{j<\psi}\right]\,. \tag{1.11}$$

The last step of the proof is to show the following inequality: For any $\gamma \in (\sqrt{\alpha}, 1)$ there exists $r \in (0,1)$ such that

$$\mathbb{E}\left[\Big(V(\mathbf{Z}^j) + V(\widetilde{\mathbf{Z}}^j)\Big)\mathbb{1}_{j<\psi}\right] \leq C\Big(\mathbb{E}\big[V(\mathbf{Z}^0)\big] + \mathbb{E}\big[V(\widetilde{\mathbf{Z}}^0)\big] + 1\Big)r^j \qquad (j \geq 1)\,. \tag{1.12}$$

Then we have

$$\Big|\mathbb{E}\big[f(\widetilde{\mathbf{Z}}^j)\big] - \mathbb{E}\big[f(\mathbf{Z}^j)\big]\Big| \leq C\Big(\mathbb{E}\big[V(\mathbf{Z}^0)\big] + \mathbb{E}\big[V(\widetilde{\mathbf{Z}}^0)\big] + 1\Big)r^j \qquad (j \geq 1)\,,$$

and the result follows. In fact, we can take the second Markov chain $\{\widetilde{\mathbf{Z}}^j\}_{j=0}^{\infty}$ to be stationary, i.e. $\mathcal{L}(\widetilde{\mathbf{Z}}_j) = \mu$ for $j \geq 0$, where μ is the invariant measure.

Step 3: Proof of the tail estimate (1.12). We will consider the product Markov chain $\mathbf{M} \equiv \{\mathbf{M}_j\}_{j=0}^{\infty} := \{(\mathbf{Z}^j, \widetilde{\mathbf{Z}}^j)\}_{j=0}^{\infty}$. Define the function $\widetilde{V} : \mathbb{X} \times \mathbb{X} \to \mathbb{R}^+$, where $\widetilde{V}(\mathbf{x}, \mathbf{y}) := V(\mathbf{x}) + V(\mathbf{y})$. Then $\widetilde{V}$ is a Lyapunov function for the chain $\mathbf{M}$. Indeed, by the definition of the Lyapunov function for the chains $\mathbf{Z}$ and $\widetilde{\mathbf{Z}}$ we have

$$\mathbb{E}[\widetilde{V}(\mathbf{M}^{j+1})|\mathcal{F}_j] \leq \alpha\widetilde{V}(\mathbf{M}^j) + 2\beta\,.$$

Hence we can apply Proposition 1.7 with parameter $s = 1$ to conclude that, for any $\gamma \in (\alpha, 1)$, we have that

$$\mathbb{E}\big[\widetilde{V}(\mathbf{M}^{j+1})|\mathcal{F}_j\big] \leq \gamma\widetilde{V}(\mathbf{M}^j) + 2\beta\mathbb{1}_{\widetilde{\mathscr{C}}}(\mathbf{M}_j)\,,$$

where

$$\widetilde{\mathscr{C}} := \left\{(\mathbf{x}, \mathbf{y}) \in \mathbb{X} \times \mathbb{X};\, \widetilde{V}(\mathbf{x}, \mathbf{y}) \leq \frac{2\beta}{\gamma - \alpha}\right\}\,.$$

Evidently, if $\mathbf{M}_j = \big(\mathbf{Z}^j, \widetilde{\mathbf{Z}}^j\big) \in \widetilde{\mathscr{C}}$ then $\mathbf{Z}^j, \widetilde{\mathbf{Z}}^j \in \mathscr{C}$. Define

$$\widetilde{\psi} := \inf_{j\geq 0}\big\{\big(\mathbf{Z}^j, \widetilde{\mathbf{Z}}^j\big) \in \widetilde{\mathscr{C}};\, \phi_j = 0\big\}\,.$$