이학박사 학위논문

Metastability of Langevin dynamics

(랑주뱅 동역학의 메타안정성)

2023년 2월

서울대학교 대학원

수리과학부

이중경

Metastability of Langevin dynamics

(랑주뱅 동역학의 메타안정성)

지도교수 서 인 석

이 논문을 이학박사 학위논문으로 제출함

2022년 10월

서울대학교 대학원

수리과학부

이중경

이중경의 이학박사 학위논문을 인준함

2022년 11월

위 원 장 김 판 기 (인)

부위원장 서 인 석 (인)

위 원 강 남 규 (인)

위 원 남 경 식 (인)

위 원 박 형 빈 (인)

Metastability of Langevin dynamics

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

to the faculty of the Graduate School of

Seoul National University

by

Jung-Kyoung Lee

Dissertation Director : Professor Insuk Seo

Department of Mathematical Sciences
Seoul National University

February 2023

Abstract

Metastability of Langevin dynamics

Jung-Kyoung Lee

Department of Mathematical Sciences
The Graduate School
Seoul National University

In this thesis, we investigate metastability of non-reversible Langevin dynamics. We prove the Eyring–Kramers formula, which is a precise estimation of the expectation of transition time, for non-reversible metastable diffusion processes that have Gibbs invariant measures. In addition, we further develop the Eyring–Kramers formula by proving that a suitably time-rescaled non-reversible metastable diffusion process converges to a Markov chain on the deepest metastable valleys.

Finally, we introduce the Curie–Weiss–Potts model as an example of a metastable dynamics on complex potential function so that complex metastability occurs. We analyze the energy landscape of the Curie–Weiss–Potts model and the metastable behavior of the heat-bath Glauber dynamics associated with the Curie–Weiss–Potts model.

Key words: Metastability, statistical physics, Langevin dynamics, Eyring–Kramers formula, Markov chain model reduction, Curie–Weiss–Potts model
Student Number: 2017-29414

Contents

Chapter 1

Introduction

Metastability is a wide-spread phenomenon that occurs in various stochastic systems at the low temperature regime. This phenomenon exhibits transitions between (meta)stable states. The mathematical study of metastability dates back to the work of H. Eyring [28] and H.A. Kramers [43] in the early 20th century. Later in the 1960s, its first successful mathematical treatment was carried out in a sequence of pioneering studies by Freidlin and Wentzell from a large-deviation theoretical perspective [29]. In this thesis, we introduce mathematical study of metastability of Langevin dynamics considering sharp asymptotics of the mean of the transition time and Markov chain description of successive transitions. Also, we introduce the Curie–Weiss–Potts model as an example of metastable dynamics on complex potential.

1.1 Mathematical study of metastability

Consider a physical system whose energy landscape is given as shown in Figure 1.1. In this system, $\boldsymbol{m}$ is a metastable state, $\boldsymbol{s}$ is a stable state, and Δ is an energy barrier from $\boldsymbol{m}$ to $\boldsymbol{s}$. Suppose that the system is at state $\boldsymbol{m}$. If a sufficient impact larger than Δ is applied to the system, the transition from $\boldsymbol{m}$ to $\boldsymbol{s}$ occurs. In many probabilistic models, this impact is an accumulation

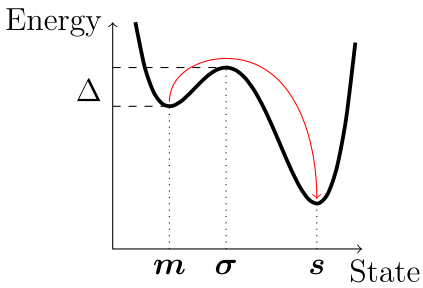


Figure 1.1: Single ground state

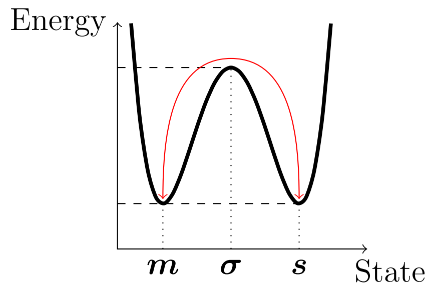


Figure 1.2: Multiple ground states

of randomness.

Let $\tau_{\boldsymbol{m}\to\boldsymbol{s}}$ be a transition time from $\boldsymbol{m}$ to $\boldsymbol{s}$. The first metastability result, called Freidlin–Wentzell theory [29], for this system is the following large-deviation type estimation of the expectation of $\tau_{\boldsymbol{m}\to\boldsymbol{s}}$:

$$\lim_{\epsilon\to 0} \epsilon \, \log \mathbb{E}[\tau_{\boldsymbol{m}\to\boldsymbol{s}}] \, = \, \Delta \ , \tag{1.1}$$

where $\epsilon > 0$ denotes the temperature.

Eyring–Kramers formula

The first more quantitative question in metastability is the derivation of the precise estimate of $\mathbb{E}[\tau_{\boldsymbol{m}\to\boldsymbol{s}}]$ as a refinement of the logarithmic estimate (1.1). Such a precise estimate is called the Eyring–Kramers formula (e.g., [8, 14, 28, 43, 51, 57, 58]) and takes the form

$$\mathbb{E}[\tau_{\boldsymbol{m}\to\boldsymbol{s}}] \simeq f(\epsilon) \exp \frac{\Delta}{\epsilon} \ . \tag{1.2}$$

Finding prefactor $f(\epsilon)$ is the main challenge in this problem.

Markov chain model reduction

If there are multiple ground states as shown Figure 1.2, we can expect that there will be successive transitions between those states. The canonical way to describe this kind of metastable behavior is the Markov chain model reduction (e.g., [2, 3, 49, 50, 52, 53, 54, 57, 59, 83]) which describes the successive inter-valley hopping dynamics as a Markov chain on much simpler state space. This also requires quantitatively precise information in the level of (1.2) regarding the metastable transition.

1.2 Langevin dynamics

In the study of the metastability of stochastic dynamical systems, one of the most important models is the overdamped Langevin dynamics given by a stochastic differential equation (SDE) of the form

$$d\boldsymbol{y}_\epsilon(t) \;=\; -\nabla U(\boldsymbol{y}_\epsilon(t))\,dt + \sqrt{2\epsilon}\,d\boldsymbol{w}_t\;, \tag{1.3}$$

where $(\boldsymbol{w}_t)_{t\ge 0}$ represents the standard d-dimensional Brownian motion, $\epsilon > 0$ is a small constant parameter corresponding to the magnitude of the noise, and $U:\mathbb{R}^d \to \mathbb{R}$ is a smooth Morse function[1] with finite critical points. In addition to its importance in large-deviation theory, mathematical physics, and engineering (cf. [29] and references therein), this process is also well-known for approximating the minibatch gradient descent algorithm widely used in deep learning (cf. [35] and references therin).

The metastable behaviors of the process $\boldsymbol{y}_\epsilon(\cdot)$, exhibited when U has multiple local minima, have attracted considerable attention in recent decades. Its first successful mathematical treatment was carried out in a sequence of pioneering studies by Freidlin and Wentzell in the 1960s from a large-deviation

[1] All the critical points of U are non-degenerate (i.e., the Hessian at each critical point is invertible) and isolated from others.

theoretical perspective, and these achievements have been summarized in [29]. Further, their accurate quantitative analysis has been thoroughly investigated in many studies. For instance, [14, 36] established the Eyring–Kramers formula, [13] provided the sharp asymptotics of low-lying spectra, [83, 86] described the metastable behavior as a limiting Markov chain under a suitable exponential time-rescaling, and [21, 22, 23, 60, 61, 65, 73] developed the quasi-stationary distribution approach for this process. The last approach is based on the theories from semi-classical analysis developed in [36, 70]. We note that these approaches are the most typical methods for quantitatively investigating the metastable behavior of a metastable process.

Metastable behavior of the dynamics

To heuristically explain the metastable behavior of the process $\boldsymbol{y}_\epsilon(\cdot)$, we regard this process as a small random perturbation of the dynamical system given by an ordinary differential equation (ODE) of the form

$$d\boldsymbol{y}(t) \;=\; -\nabla U(\boldsymbol{y}(t))\, dt\;. \tag{1.4}$$

Note that the stable equilibria of this dynamical system are given by the local minima of U. Hence, provided that $\epsilon \simeq 0$, the process $\boldsymbol{y}_\epsilon(\cdot)$ starting from a neighborhood of a local minimum of U will remain there for a sufficiently long time, as the noise is small compared to the drift term $-\nabla U(\boldsymbol{y}_\epsilon(t))dt$ that pushes the process toward the local minimum.

The metastability issue arises for the process $\boldsymbol{y}_\epsilon(\cdot)$ if U has multiple local minima. To illustrate the corresponding metastable behavior more clearly, we simply assume that U has two local minima $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ as shown in Figure 1.3, and we suppose that the process $\boldsymbol{y}_\epsilon(\cdot)$ starts at $\boldsymbol{m}_1$. If there is no noise, i.e., $\epsilon = 0$, the process always remains at $\boldsymbol{m}_1$. However, when ϵ is small but positive, random noise term $\sqrt{2\epsilon}\, d\boldsymbol{w}_t$ accumulates over a sufficiently long time and enables the process $\boldsymbol{y}_\epsilon(\cdot)$ to make a transition to a neighborhood of

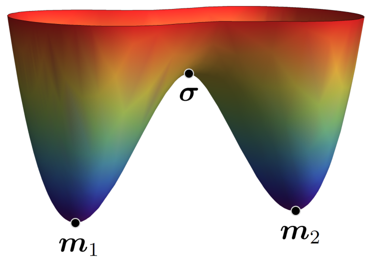


Figure 1.3: Double-well potential U with two minima $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ and a saddle point $\boldsymbol{\sigma}$ between them.

another minimum $\boldsymbol{m}_2$ after an exponentially long time, and this can be understood via the large-deviation principle (cf. [29]). This movement is called a *metastable transition*. Then, it remains for a long time in the neighborhood of $\boldsymbol{m}_2$ before making another transition. Such rare transitions between the neighborhoods of local minima constitute the dynamical metastable behavior of the process $\boldsymbol{y}_\epsilon(\cdot)$. We can expect richer behaviors when U has a more complex landscape.

Eyring–Kramers formula

The Eyring–Kramers formula is the sharp asymptotics, as $\epsilon \to 0$, of the mean of the time required to observe the transition described above. It was obtained for the one-dimensional case in classical studies [28, 43] conducted in the 1930s on the basis of explicit computation. The generalization of this result to arbitrary dimensions was finally accomplished in [14] a few decades later. We recall the double-well situation illustrated in Figure 1.3 to explain the Eyring–Kramers formula in a simple form. Let $\tau_{\mathcal{D}_\epsilon(\boldsymbol{m}_2)}$ denote the hitting time with respect to the process $\boldsymbol{y}_\epsilon(\cdot)$ of the set $\mathcal{D}_\epsilon(\boldsymbol{m}_2)$, which is a ball of radius ϵ centered at $\boldsymbol{m}_2$. Then, the Eyring–Kramers formula is the sharp estimate of the mean transition time $\mathbb{E}[\tau_{\mathcal{D}_\epsilon(\boldsymbol{m}_2)}|\boldsymbol{y}_\epsilon(0) = \boldsymbol{m}_1]$. The Freidlin–

Wentzell theory gives the large deviation estimate for this quantity as

$$\lim_{\epsilon\to 0}\epsilon\,\log\mathbb{E}[\,\tau_{\mathcal{D}_\epsilon(\boldsymbol{m}_2)}\,|\,\boldsymbol{y}_\epsilon(0)=\boldsymbol{m}_1\,]\;=\;U(\boldsymbol{\sigma})-U(\boldsymbol{m}_1)\;, \tag{1.5}$$

where $\boldsymbol{\sigma}$ is the saddle point between the two wells as shown in Figure 1.3. The Eyring–Kramers formula is a refinement of this result (cf. Corollary 3.1.5 of the thesis), and it gives the precise asymptotics of the expectation in (1.5).

The mean transition time is related to the quantification of the mixing property of the process $\boldsymbol{y}_\epsilon(\cdot)$. To explain it more precisely, we remark that the unique invariant measure for the process $\boldsymbol{y}_\epsilon(\cdot)$ is given by

$$\mu_\epsilon(d\boldsymbol{x})\;=\;\frac{1}{Z_\epsilon}\,e^{-U(\boldsymbol{x})/\epsilon}\,d\boldsymbol{x}\;, \tag{1.6}$$

where Z_ϵ is the constant given by

$$Z_\epsilon\;=\;\int_{\mathbb{R}^d}e^{-U(\boldsymbol{x})/\epsilon}\,d\boldsymbol{x}\;<\;\infty\;, \tag{1.7}$$

where we will impose suitable growth conditions for U in Section 2.1 to guarantee the finiteness of Z_ϵ. The measure $\mu_\epsilon(d\boldsymbol{x})$ corresponds to the *Gibbs measure* associated with the energy function U and inverse temperature ϵ and hence the constant Z_ϵ denotes the associated partition function. Therefore, we can regard the process $\boldsymbol{y}_\epsilon(\cdot)$ as a sampler of the Gibbs distribution $\mu_\epsilon(d\boldsymbol{x})$, which is exponentially concentrated on the global minima of U. There are two representative quantities for measuring this mixing property of the sampler $\boldsymbol{y}_\epsilon(\cdot)$: the spectral gap [13] and the mean transition time of the process from one local minimum to another [14]. Thus, by estimating the latter using the Eyring–Kramers formula, one can precisely measure the mixing property of $\boldsymbol{y}_\epsilon(\cdot)$.

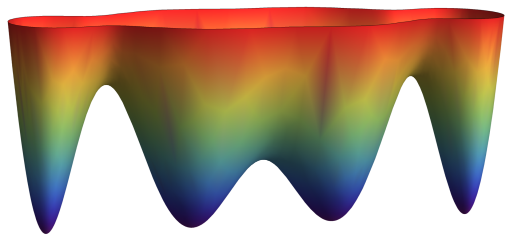


Figure 1.4: Example of potential U with multiple global minima.

Markov chain description of metastable behavior

The Eyring–Kramers formula focuses on a single metastable transition. Our next focus is not on such a single transition but on the full description of successive transitions via a suitable scaling limit. Now, suppose that U has multiple global minima as shown in Figure 1.4[2] and that the process $\boldsymbol{y}_\epsilon(\cdot)$ starts from a small fixed neighborhood of a minimum, which is called a (metastable) *valley*. More precisely, we can expect that once the process $\boldsymbol{y}_\epsilon(\cdot)$ makes a transition from one valley to another, then the next transition to another valley will take place after another exponentially long time. Hence, to comprehensively describe the metastable behavior, it is natural to prove that these successive metastable transitions converge in some sense to a continuous-time Markov process whose state space consists of the valleys of U. This limiting Markov process has a finite state space and is simpler than the original process so that this argument is called *Markov chain model reduction*. This proof of course requires highly accurate knowledge regarding the transition time in the level of the Eyring–Kramers formula, thereby providing a more detailed description of the metastable behavior.

Now, we review existing studies on the Markov chain description of metastable behavior. In [2, 3], a robust methodology based on potential theory has been introduced for the case in which the underlying dynamic is a Markov process on a *discrete* set. This method has been applied to many models

[2]This figure has been excerpted from [83, Figure 1.2].

such as the zero-range process [4, 45, 84], the inclusion process [7, 40, 41], the discrete version of the overdamped Langevin dynamics [52, 54], and the ferromagnetic systems [47, 53]. On the other hand, for metastable *diffusion* processes, a different methodology known as the PDE approach based on the analysis of a certain Poisson equation has been introduced in [55, 83]. In [83], a general methodology to deal with the solution to the corresponding Poisson equation when the underlying dynamics is reversible has been developed and successfully applied to the process $\boldsymbol{y}_\epsilon(\cdot)$.

1.3 Main contribution of the thesis

In this thesis, we consider a variant of the classical overdamped Langevin dynamics $\boldsymbol{y}_\epsilon(\cdot)$, which is obtained by adding a vector field to the drift term of the SDE (1.3). More precisely, we focus on the metastability of the diffusion process given by an SDE of the form

$$d\boldsymbol{x}_\epsilon(t) \;=\; -(\nabla U + \boldsymbol{\ell})(\boldsymbol{x}_\epsilon(t))\, dt + \sqrt{2\epsilon}\, d\boldsymbol{w}_t \;,$$

where $U : \mathbb{R}^d \to \mathbb{R}$ is the smooth potential function as described above. Further, $\boldsymbol{\ell} : \mathbb{R}^d \to \mathbb{R}^d$ is a smooth vector field that is orthogonal to the gradient field ∇U, i.e.,

$$\nabla U(\boldsymbol{x}) \cdot \boldsymbol{\ell}(\boldsymbol{x}) \;=\; 0 \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^d \;,$$

and it is incompressible:

$$(\nabla \cdot \boldsymbol{\ell})(\boldsymbol{x}) \;=\; 0 \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^d \;.$$

The first condition guarantees that the quasi-potential of the process $\boldsymbol{x}_\epsilon(\cdot)$ is U (cf. [29, Theorem 3.3.1]), and the second condition ensures that the invariant measure of the process $\boldsymbol{x}_\epsilon(\cdot)$ is the Gibbs measure $\mu_\epsilon(d\boldsymbol{x})$ (cf. Theorem

2.1.3). In this sense, the process $\boldsymbol{x}_\epsilon(\cdot)$ is another sampler of the Gibbs measure $\mu_\epsilon(d\boldsymbol{x})$. Indeed, we prove in Theorem 2.1.3 that both of the conditions are the necessary and sufficient conditions for the process $\boldsymbol{x}_\epsilon(\cdot)$ to have as an invariant measure the Gibbs distribution $\mu_\epsilon(d\boldsymbol{x})$ defined in (1.6) for all $\epsilon > 0$. For this reason, this generalized model has been investigated in many studies from different perspectives, e.g., [24, 37, 38, 62, 64, 81, 82].

Comparison to the process $\boldsymbol{y}_\epsilon(\cdot)$

One of the main features of the process $\boldsymbol{y}_\epsilon(\cdot)$ is the fact that it is *reversible* with respect to its Gibbs invariant measure $\mu_\epsilon(d\boldsymbol{x})$. Owing to this reversibility, many tools are available to investigate the process $\boldsymbol{y}_\epsilon(\cdot)$. However, *nearly none of these tools is applicable to non-reversible processes such as* $\boldsymbol{x}_\epsilon(\cdot)$. Hence, the quantitative analysis of the metastability of non-reversible processes has long been an open issue. To this end, many innovative studies such as [32, 45, 51, 54, 84] have been conducted in recent years, and several non-reversible metastable processes have been analyzed. In particular, low-lying spectra of $\boldsymbol{x}_\epsilon(\cdot)$ has been analyzed in [64]. In the thesis, we present the Eyring–Kramers formula and the Markov chain description of the metastable behavior of the process $\boldsymbol{x}_\epsilon(\cdot)$.

Moreover, it is widely believed that the non-reversible process $\boldsymbol{x}_\epsilon(\cdot)$ has a better mixing property than the reversible process $\boldsymbol{y}_\epsilon(\cdot)$. In fact, this belief has been quantitatively verified in Chapter 3 and [64] in view of the so-called Eyring–Kramers formula and low-lying spectra, respectively. We verify in Theorem 2.1.1 that the stable points of the process $\boldsymbol{x}_\epsilon(\cdot)$ are the local minima of U and hence identical to those of the process $\boldsymbol{y}_\epsilon(\cdot)$. Hence, we can compare the Eyring–Kramers formula of $\boldsymbol{x}_\epsilon(\cdot)$ with that of $\boldsymbol{y}_\epsilon(\cdot)$, and this comparison reveals that *the mean transition time of the dynamics* $\boldsymbol{x}_\epsilon(\cdot)$ *from one local minimum of U to another is always faster than that of the reversible dynamics* $\boldsymbol{y}_\epsilon(\cdot)$. This implies that we can accelerate the stochastic gradient descent algorithm by adding the incompressible field $\boldsymbol{\ell}$, which is orthogonal to ∇U. We remark that such an acceleration has been observed for the model when the diffusivity ϵ is kept constant (see [24, 37, 38, 62, 81, 82] and references therein). In particular, we refer to [31] for the explicit relation with the stochastic gradient descent algorithm.

General methodology of capacity estimation

Another main result of our study is the establishment of a straightforward and robust method for estimating a potential theoretic notion known as the capacity. In the proof of Eyring–Kramers formula based on the potential theoretic approach developed in [14], it is crucial to estimate the capacity between metastable valleys. In all the existing results based on this approach, such an estimation is carried out via variational principles such as the Dirichlet principle or the Thomson principle.

For the reversible case, this approach is less complex as the Dirichlet principle is an optimization problem over a space of functions. Hence, by taking a suitable test function that approximates the known optimizer of the variational principle, we can bound the capacity in a precise manner. This strategy is the essence of the potential theoretic approach to metastability. By contrast, for the non-reversible case, the variational expression of the capacity is destined to involve both the function and the so-called flow (cf. [54, Theorems 3.2 and 3.3]). Therefore, one must construct both the test function and the test flow to estimate the capacity precisely. Accordingly, when this approach is adopted for the non-reversible model, the major technical difficulty arises in the construction of the test flow. This problem has been resolved in existing studies such as [51, 54, 84] based on considerable computations.

In this thesis, we develop a robust methodology to estimate the capacity without relying on these variational principles. We use only a test function in the estimation of the capacity; ***no test flow is used even in the non-reversible case***. Hence, our methodology significantly reduces the complexity of the analysis of metastable non-reversible processes to the level of the reversible models. Therefore, our methodology is expected to present new possibilities for the analysis of non-reversible metastable random processes.

In summary, we develop a new methodology to estimate the capacity and use it to establish the Eyring–Kramers formula for the non-reversible and metastable diffusions $\boldsymbol{x}_\epsilon(\cdot)$.

Resolvent equation approach

The Markov model reduction has been presented for the reversible process $\boldsymbol{y}_\epsilon(\cdot)$ in [83] based on the partial differential equation (PDE) approach. In this thesis, we extend the PDE approach to the non-reversible setting in a robust manner and apply this method to the process $\boldsymbol{x}_\epsilon(\cdot)$. We note that the method of [83] relies on analysis of solutions of certain form of Poisson equations. It has been observed in [49, 50] that considering ***resolvent equation***, instead of Poisson equation, simplifies several argument and provide more robust and convenient methodology which in some sense provides a necessary and sufficient condition for the Markov chain description of metastable behavior. This method has been applied to a critical reversible zero-range process [49] to which the method of [2, 3] is not applicable because the metastable valley is too large. Hence, we will rely on this resolvent approach to analyze the metastability of the process $\boldsymbol{x}_\epsilon(\cdot)$ in the current thesis and this is the first application of this method to a non-reversible model.

Curie–Weiss–Potts model

The Potts model is a well-known mathematical model suitable for studying ferromagnetic spin system consisting of $q \geq 3$ spins. We refer to [89] a comprehensive review on the Potts model. In this thesis, we introduce the Potts model defined on large complete graphs without an external field to understand the associated energy landscape as well as the metastable behavior of the heat-bath Glauber dynamics to the highly precise level. This special case of the Potts model defined on complete graphs is called the *Curie–Weiss–Potts model*. First, we thoroughly analyze the energy landscape of the Curie–Weiss–Potts model. In particular, for the Curie–Weiss–Potts model with $q \ge 3$ spins and zero external field, we completely characterize all critical temperatures and phase transitions in view of the global structure of the energy landscape. We observe that there are three critical temperatures and four different regimes for $q < 5$, whereas there are four critical temperatures and five different regimes for $q \ge 5$. Our analysis extends the investigations performed in [18]; they provide the precise characterization of the second critical temperatures for all $q \ge 3$ and in [53], which provides a complete analysis of the energy landscape for $q = 3$. Based on our precise analysis of the energy landscape, we also perform a quantitative investigation of the metastable behavior of the heat-bath Glauber dynamics associated with the Curie–Weiss–Potts model.

1.4 General model

We conclude the introduction by explaining the importance of the process $\boldsymbol{x}_\epsilon(\cdot)$ in the study of metastability. For a vector field $\boldsymbol{b} : \mathbb{R}^d \to \mathbb{R}^d$, consider the dynamical system in $\mathbb{R}^d$ given by an ODE of the form

$$d\boldsymbol{z}(t) \;=\; -\boldsymbol{b}(\boldsymbol{z}(t))\, dt \quad ; \; t \ge 0 \; . \tag{1.8}$$

Suppose that this dynamics has several stable equilibria. An open problem in the study of metastability is to determine the metastable behavior for the following small random perturbation of (1.8):

$$d\boldsymbol{z}_\epsilon(t) \;=\; -\boldsymbol{b}(\boldsymbol{z}_\epsilon(t))\, dt + \sqrt{2\epsilon}\, d\boldsymbol{w}_t \quad ; \; t \ge 0 \; . \tag{1.9}$$

For this model, Freidlin and Wentzell [29] established the large-deviation-type analysis of the metastable behavior. In [55], metastability of the process $\boldsymbol{z}_\epsilon(\cdot)$ on a *one-dimensional torus* has been analyzed based on the explicit form of the solution to the Poisson equation. However, rigorous accurate quantitative analysis such as that based on the Eyring–Kramers formula or the Markov chain description is unknown for this general model and remains as a primary open question in this field. We refer to [8] for the Eyring–Kramers formula for $\boldsymbol{z}_\epsilon(\cdot)$ under a special set of assumptions.

The difficulty in the rigorous analysis of the process $\boldsymbol{z}_\epsilon(\cdot)$ is due to two factors: the non-reversibility and the lack of an explicit formula for the invariant measure. In Theorem 2.1.3 below, we prove that the process $\boldsymbol{z}_\epsilon(\cdot)$ defined in (1.9) has a Gibbs invariant measure (1.6) if and only if $\boldsymbol{b} = \nabla U + \boldsymbol{\ell}$ for some $\boldsymbol{\ell}$ such that $\nabla U \cdot \boldsymbol{\ell} \equiv 0$ and $\nabla \cdot \boldsymbol{\ell} \equiv 0$; hence, this is the model considered in this thesis. Thus, we completely overcome the difficulty arising from the non-reversibility in the study of the process $\boldsymbol{z}_\epsilon(\cdot)$ in this thesis. The problem arising from the lack of an understanding of the invariant measure of $\boldsymbol{z}_\epsilon(\cdot)$ is not addressed in our studies, as the model considered has an explicit Gibbs invariant measure; this problem should be investigated in future research.

We finally remark that an important model which is not discussed in this introduction is the underdamped Langevin dynamics. This dynamics is non-reversible and has the Gibbs measure as the invariant measure so that it seems at first glance that this model falls into our framework. However, the main challenge in this model is the fact that the diffusion coefficient is degenerate. Accordingly, rigorous quantitative study for this model is barely known (cf. [63]) and is an important future research problem.

Structure of the thesis

First, we present a precise definition of non-reversible Langevin dynamics in Chapter 2. Chapters 3 and 4 are devoted to the Eyring–Kramers formula and Markov chain model reduction for non-reversible Langevin dynamics, respectively. In Chapter 5, we introduce the Curie–Weiss–Potts model as an example of metastable dynamics on complex potential. We completely analyze the energy landscape of the model and prove the Eyring–Kramers formula and Markov chain model reduction for the heat-bath Glauber dynamics associated with the Curie–Weiss–Potts model.

Chapter 2

Model

2.1 Non-reversible Langevin dynamics

In this section, we introduce the fundamental features of the model. Recall the definition of $\boldsymbol{x}_\epsilon(\cdot)$:

$$d\boldsymbol{x}_\epsilon(t) \;=\; -(\nabla U + \boldsymbol{\ell})(\boldsymbol{x}_\epsilon(t))\, dt + \sqrt{2\epsilon}\, d\boldsymbol{w}_t \;, \tag{2.1}$$

with two conditions on $\boldsymbol{\ell}$:

$$\nabla U(\boldsymbol{x}) \cdot \boldsymbol{\ell}(\boldsymbol{x}) \;=\; 0 \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^d \;, \tag{2.2}$$

and

$$(\nabla \cdot \boldsymbol{\ell})(\boldsymbol{x}) \;=\; 0 \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^d \;. \tag{2.3}$$

The results stated in this chapter regarding the process $\boldsymbol{x}_\epsilon(\cdot)$ constitute the essence of this field.

Potential function U

To introduce the model rigorously, we must explain the potential function $U : \mathbb{R}^d \to \mathbb{R}$ in the SDE (2.1). We assume that the potential function $U \in$

$C^3(\mathbb{R}^d)$ is a Morse function that satisfies the growth conditions

$$\lim_{n\to\infty} \inf_{|\boldsymbol{x}|\geq n} \frac{U(\boldsymbol{x})}{|\boldsymbol{x}|} = \infty \,, \tag{2.4}$$

$$\lim_{|\boldsymbol{x}|\to\infty} \frac{\boldsymbol{x}}{|\boldsymbol{x}|} \cdot \nabla U(\boldsymbol{x}) = \infty \,, \text{ and} \tag{2.5}$$

$$\lim_{|\boldsymbol{x}|\to\infty} \{|\nabla U(\boldsymbol{x})| - 2\Delta U(\boldsymbol{x})\} = \infty \,, \tag{2.6}$$

where $|\boldsymbol{x}|$ denotes the Euclidean distance in $\mathbb{R}^d$. These conditions have been introduced in previous studies such as [14, 51, 83] to guarantee the positive recurrence of the diffusion process $\boldsymbol{y}_\epsilon(\cdot)$ given by (1.3) and the finiteness of Z_ϵ in (1.7). More precisely, it is well known (cf. [14]) that these conditions imply the tightness condition

$$\int_{\{\boldsymbol{x}:U(\boldsymbol{x})\geq a\}} e^{-U(\boldsymbol{x})/\epsilon} d\boldsymbol{x} \,\leq\, C_a\, e^{-a/\epsilon} \ \text{ for all } a \in \mathbb{R} \,, \tag{2.7}$$

where C_a is a constant that depends only on a, and hence imply the finiteness of the partition function Z_ϵ. Finally, we also assume that U is a Morse function, i.e., all the critical points of U are non-degenerate. We remark that the metastability of the reversible process $\boldsymbol{y}_\epsilon(\cdot)$ has been analyzed in [14] under the same set of assumptions.

Deterministic dynamical system $\boldsymbol{x}(\cdot)$

To explain the metastable behavior of the process $\boldsymbol{x}_\epsilon(\cdot)$, we first consider a deterministic dynamical system given by the ODE

$$d\boldsymbol{x}(t) \,=\, -(\nabla U + \boldsymbol{\ell})(\boldsymbol{x}(t))\, dt \,. \tag{2.8}$$

We can demonstrate that this dynamical system has essentially the same phase portrait as $\boldsymbol{y}(\cdot)$ defined in (1.4).

Theorem 2.1.1. *The following hold.*

1. *We have* $\boldsymbol{\ell}(\boldsymbol{c}) = 0$ *for all critical points* $\boldsymbol{c} \in \mathbb{R}^d$ *of* U*.*

2. *A point* $\boldsymbol{c} \in \mathbb{R}^d$ *is an equilibrium of the dynamical system* (2.8) *if and only if* $\boldsymbol{c} \in \mathbb{R}^d$ *is a critical point of* U*.*

3. *An equilibrium* $\boldsymbol{c} \in \mathbb{R}^d$ *of the dynamical system* (2.8) *is stable if and only if* $\boldsymbol{c}$ *is a local minimum of* U*.*

The proof is given in Section 2.2. We emphasize that the divergence-free condition (2.3) is not used in the proof of this theorem, whereas the orthogonality condition (2.2) plays a significant role. In view of part (3) of the previous theorem, we can observe that the process $\boldsymbol{x}_\epsilon(\cdot)$ is expected to exhibit metastable behavior when U has multiple local minima, and this is the situation that we are going to discuss in the current thesis.

Diffusion process $\boldsymbol{x}_\epsilon(\cdot)$

Now, we focus on the diffusion process $\boldsymbol{x}_\epsilon(\cdot)$. Under the condition (2.2) and conditions (2.4)–(2.6), we can prove the following property of the process $\boldsymbol{x}_\epsilon(\cdot)$. Note again that the condition (2.3) is not used.

Theorem 2.1.2. *The following hold.*

1. *There is no explosion for the diffusion process* $\boldsymbol{x}_\epsilon(\cdot)$*.*

2. *The diffusion process* $\boldsymbol{x}_\epsilon(\cdot)$ *is positive recurrent.*

The proof of this result is given in Section 2.3.

Invariant measure

Since the process $\boldsymbol{x}_\epsilon(\cdot)$ is positive recurrent, we know that this process has an invariant measure. Now, we prove that $\mu_\epsilon(d\boldsymbol{x})$ is the unique invariant measure for the process $\boldsymbol{x}_\epsilon(\cdot)$.

Before proceeding to the statement of this result, we first explain the role of the conditions (2.2) and (2.3). Recall the general model $\boldsymbol{z}_\epsilon(\cdot)$ given by the SDE (1.9). It is known from [29, Theorem 3.3.1] that if the quasi-potential V associated with (1.9) is of class C^1, we can write $\boldsymbol{b} = \nabla V + \boldsymbol{\ell}$ where $\nabla V \cdot \boldsymbol{\ell} \equiv 0$. Hence, the assumption (2.2) is nothing more than the regularity assumption on the quasi-potential. The special assumption regarding the field $\boldsymbol{\ell}$ is (2.3), and the role of this assumption is summarized below.

Theorem 2.1.3. *The following hold.*

1. *If $\boldsymbol{\ell}$ satisfies the conditions* (2.2) *and* (2.3)*, then the Gibbs measure $\mu_\epsilon(d\boldsymbol{x})$ is the unique invariant measure for the diffusion process $\boldsymbol{x}_\epsilon(\cdot)$.*

2. *On the other hand, suppose that the Gibbs measure $\mu_\epsilon(d\boldsymbol{x})$ is the invariant measure for the diffusion process $\boldsymbol{z}_\epsilon(\cdot)$ defined in* (1.9) *for all $\epsilon > 0$. Then, the vector field $\boldsymbol{b}$ can be written as $\boldsymbol{b} = \nabla U + \boldsymbol{\ell}$, where U and $\boldsymbol{\ell}$ satisfy* (2.2) *and* (2.3)*.*

The proof of this theorem is given in Section 2.3. Therefore, heuristically, the condition (2.3) can be regarded as a necessary and sufficient condition (up to the regularity of the quasi-potential) for the diffusion process $\boldsymbol{z}_\epsilon(\cdot)$ to have the Gibbs invariant measure.

Construction of $\boldsymbol{\ell}$

The result obtained in this thesis might be nearly useless if it is extremely difficult to find a non-trivial $\boldsymbol{\ell}$ satisfying the conditions (2.2) and (2.3) simultaneously. However, there is a simple way to generate a variety of $\boldsymbol{\ell}$ when the potential U is given. Let $\mathcal{M}_{d\times d}(\mathbb{R})$ be a space of $d \times d$ real matrices and let $J : \mathbb{R} \to \mathcal{M}_{d\times d}(\mathbb{R})$ be a smooth function such that the range of J consists of only skew-symmetric matrices. Then, a vector field of the form $\boldsymbol{\ell}(\boldsymbol{x}) = J(U(\boldsymbol{x}))\, \nabla U(\boldsymbol{x})$ satisfies the conditions (2.2) and (2.3) as observed in [47, Section 1] and [64, Section 1]. Moreover, unless J is a constant function, the model considered here is different from the one considered in [51].

Notations regarding $\boldsymbol{x}_\epsilon(\cdot)$

We conclude this section by defining some notations regarding the process $\boldsymbol{x}_\epsilon(\cdot)$. Let $\mathscr{L}_\epsilon$ denote the generator associated with the process $\boldsymbol{x}_\epsilon(\cdot)$. Then, $\mathscr{L}_\epsilon$ acts on $f \in C^2(\mathbb{R}^d)$ such that

$$\mathscr{L}_\epsilon f \;=\; -(\nabla U + \boldsymbol{\ell}) \cdot \nabla f + \epsilon \Delta f \;. \tag{2.9}$$

Under the conditions (2.2) and (2.3) on $\boldsymbol{\ell}$, we can rewrite this generator in the divergence form as

$$\mathscr{L}_\epsilon f \;=\; \epsilon e^{U/\epsilon} \nabla \cdot \left[e^{-U/\epsilon} \Big(\nabla f - \frac{1}{\epsilon} f \, \boldsymbol{\ell} \Big) \right] . \tag{2.10}$$

Finally, let $\mathbb{P}_{\boldsymbol{x}}^\epsilon$ denote the law of the process $\boldsymbol{x}_\epsilon(\cdot)$ starting from $\boldsymbol{x}$, and let $\mathbb{E}_{\boldsymbol{x}}^\epsilon$ denote the expectation with respect to $\mathbb{P}_{\boldsymbol{x}}^\epsilon$.

2.2 Deterministic dynamical system

In this section, we prove the properties of the dynamical systems $\boldsymbol{x}(\cdot)$ given by the ODE (2.8).

2.2.1 Preliminary results on matrix computations

In this subsection, we present few technical lemmas. We remark that all the vectors and matrices in this subsection are real. The first lemma below will be used to investigate the stable equilibria of the dynamical system $\boldsymbol{x}(\cdot)$.

Lemma 2.2.1. *Let $\mathbb{A}$, $\mathbb{B}$ be square matrices of the same size and suppose that $\mathbb{A}$ is symmetric positive definite and $\mathbb{A}\mathbb{B}$ is skew-symmetric. Then, all the eigenvalues of matrix $\mathbb{A} + \mathbb{B}$ are either positive real or complex with a positive real part. In particular, the matrix $\mathbb{A} + \mathbb{B}$ is invertible.*

Proof. By a change of basis, we may assume that $\mathbb{A} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_d)$

for some $\lambda_1, \ldots, \lambda_d > 0$. Let α be a real eigenvalue of $\mathbb{A}+\mathbb{B}$ and let $\boldsymbol{u}$ be the corresponding non-zero eigenvector. Then, we have

$$0 < |\mathbb{A}\boldsymbol{u}|^2 = \mathbb{A}\boldsymbol{u}\cdot(\mathbb{A}+\mathbb{B})\boldsymbol{u} = \alpha(\mathbb{A}\boldsymbol{u}\cdot\boldsymbol{u}) ,$$

where the first identity holds since $\mathbb{A}\mathbb{B}$ is skew-symmetric. This proves that $\alpha > 0$ since $\mathbb{A}$ is positive definite.

Next, let $z = a + ib$ be a complex eigenvalue of $\mathbb{A}+\mathbb{B}$ and let $\boldsymbol{u}+i\boldsymbol{w}$ be the corresponding non-zero eigenvector, where $\boldsymbol{u}$ and $\boldsymbol{w}$ are real vectors. Since $\mathbb{A}$ and $\mathbb{B}$ are real, we have

$$(\mathbb{A}+\mathbb{B})\boldsymbol{u} = a\boldsymbol{u} - b\boldsymbol{w} \quad \text{and} \quad (\mathbb{A}+\mathbb{B})\boldsymbol{w} = b\boldsymbol{u} + a\boldsymbol{w} .$$

Since $\mathbb{A}\mathbb{B}$ is skew-symmetric, we get

$$\begin{aligned} |\mathbb{A}\boldsymbol{u}|^2 &= \mathbb{A}\boldsymbol{u}\cdot(\mathbb{A}+\mathbb{B})\boldsymbol{u} = \mathbb{A}\boldsymbol{u}\cdot(a\boldsymbol{u} - b\boldsymbol{w}) , \\ |\mathbb{A}\boldsymbol{w}|^2 &= \mathbb{A}\boldsymbol{w}\cdot(\mathbb{A}+\mathbb{B})\boldsymbol{w} = \mathbb{A}\boldsymbol{w}\cdot(b\boldsymbol{u} + a\boldsymbol{w}) . \end{aligned}$$

By adding these two identities, we get

$$|\mathbb{A}\boldsymbol{u}|^2 + |\mathbb{A}\boldsymbol{w}|^2 = a(\mathbb{A}\boldsymbol{u}\cdot\boldsymbol{u} + \mathbb{A}\boldsymbol{w}\cdot\boldsymbol{w}) .$$

Therefore, we get $a > 0$ since $\mathbb{A}$ is positive definite. □

The next lemma is used to analyze the saddle points of the dynamical system (2.8). For a square matrix $\mathbb{M}$, let $\mathbb{M}^\dagger$ denote its transpose, and we write $\mathbb{M}^s = \frac{1}{2}(\mathbb{M}+\mathbb{M}^\dagger)$.

Lemma 2.2.2. *Let $\mathbb{A}$, $\mathbb{B}$ be square matrices of the same size and suppose that $\mathbb{A}^s$ is positive definite and $\mathbb{B}$ is a non-singular, symmetric matrix that has only one negative eigenvalue. Then, $\mathbb{A}\mathbb{B}$ is invertible and has only one negative eigenvalue with geometric multiplicity* 1.

Proof. By a change of basis, we may assume that $\mathbb{B} = \mathrm{diag}(-\lambda_1,\, \lambda_2,\, \ldots,\, \lambda_d)$ for some $\lambda_1,\, \ldots,\, \lambda_d > 0$. It is well known that a matrix $\mathbb{A}$ such that $\mathbb{A}^s$ is positive definite does not have a negative eigenvalue and $\det \mathbb{A} > 0$. Hence, we have $\det \mathbb{A}\mathbb{B} < 0$ so that $\mathbb{A}\mathbb{B}$ is invertible and has at least one negative eigenvalue.

First, assume that $\mathbb{A}\mathbb{B}$ has two different negative eigenvalues, $-a$, $-b$, and let $\boldsymbol{u} = (u_1,\, \ldots,\, u_d)$, $\boldsymbol{w} = (w_1,\, \ldots,\, w_d)$ be the corresponding eigenvectors. We claim that $u_1,\, w_1 \neq 0$. By contrast, suppose that $u_1 = 0$. Then, we have

$$\mathbb{B}\boldsymbol{u} \cdot \mathbb{A}^s \mathbb{B}\boldsymbol{u} \;=\; \mathbb{B}\boldsymbol{u} \cdot \mathbb{A}\mathbb{B}\boldsymbol{u} \;=\; -a\mathbb{B}\boldsymbol{u} \cdot \boldsymbol{u} \;=\; -a\sum_{j=2}^{d} \lambda_j u_j^2 \;<\; 0\;, \tag{2.11}$$

which is a contradiction since $\mathbb{A}^s$ is positive definite. By the same argument, we get $w_1 \neq 0$.

By the definition of a, b and by the positive definiteness of $\mathbb{A}^s$, for any $t \in \mathbb{R}$,

$$(\boldsymbol{u} + t\boldsymbol{w})^\dagger\, \mathbb{B}\, (a\boldsymbol{u} + bt\boldsymbol{w}) \;=\; -(\boldsymbol{u} + t\boldsymbol{w})^\dagger\, \mathbb{B}\mathbb{A}\mathbb{B}\, (\boldsymbol{u} + t\boldsymbol{w}) \;<\; 0\;.$$

Let $p = -u_1/(bw_1)$. By substituting t with ap in the previous equation, the first coordinate of $a\boldsymbol{u} + bt\boldsymbol{w} = a(\boldsymbol{u} + bp\boldsymbol{w})$ is zero; thus, we have

$$0 \;>\; (\boldsymbol{u} + ap\boldsymbol{w})^\dagger\, \mathbb{B}\, (a\boldsymbol{u} + abp\boldsymbol{w}) \;=\; a\sum_{j=2}^{d} \lambda_j\, (u_j + apw_j)\, (u_j + bpw_j)\;. \tag{2.12}$$

Similarly, substituting t with bp makes the first coordinate of $\boldsymbol{u} + bp\boldsymbol{w}$ zero, and we get

$$0 \;>\; (\boldsymbol{u} + bp\boldsymbol{w})^\dagger\, \mathbb{B}\, (a\boldsymbol{u} + b^2p\boldsymbol{w}) \;=\; \sum_{j=2}^{d} \lambda_j\, (au_j + b^2pw_j)\, (u_j + bpw_j)\;. \tag{2.13}$$

Computing $(b/a \times (2.12) + (2.13))$ gives

$$0 > \sum_{j=2}^{d} \lambda_j\,(u_j+bpw_j)(bu_j+abpw_j+au_j+b^2pw_j) \;=\; (a+b)\sum_{j=2}^{d} \lambda_j(u_j+bpw_j)^2\,,$$

which is a contradiction since we have assumed that $\lambda_2, \ldots, \lambda_d > 0$. Therefore, $\mathbb{AB}$ has only one negative eigenvalue $-a$.

Now, let us assume that there are two eigenvectors $\boldsymbol{u}$ and $\boldsymbol{w}$ corresponding to $-a$, which are linearly independent. Then, we can repeat the same computation as that presented above to get a contradiction, as we did not use the fact that $a \neq b$ in the computation. Hence, the dimension of the eigenspace corresponding to the eigenvalue $-a$ is 1. □

Remark 2.2.3. *Indeed, we can show that the algebraic multiplicity of the unique negative eigenvalue is also* 1 *by considering the Jordan decomposition.*

The following lemma is a direct consequence of the previous one.

Lemma 2.2.4. *Let $\mathbb{A}$, $\mathbb{B}$ be square matrices of the same size and suppose that $\mathbb{A}$ is a symmetric non-singular matrix with exactly one negative eigenvalue and $\mathbb{AB}$ is a skew-symmetric matrix. Then, the matrix $\mathbb{A} + \mathbb{B}$ is invertible and has only one negative eigenvalue, and its geometric multiplicity is* 1*.*

Proof. Since $\mathbb{A}$ is symmetric and $\mathbb{AB}$ is skew-symmetric, we have $-\mathbb{AB} = (\mathbb{AB})^\dagger = \mathbb{B}^\dagger\mathbb{A}$. Therefore, we get $\mathbb{BA}^{-1} = -\mathbb{A}^{-1}\mathbb{B}^\dagger = -(\mathbb{BA}^{-1})^\dagger$; thus, the matrix $\mathbb{BA}^{-1}$ is skew-symmetric. Let $\mathbb{I}$ be the identity matrix with the same size as $\mathbb{A}$. Then, by substituting $\mathbb{I} + \mathbb{BA}^{-1}$ and $\mathbb{A}$ for $\mathbb{A}$ and $\mathbb{B}$, respectively, in Lemma 2.2.2, we conclude the proof since $\mathbb{A} + \mathbb{B} = (\mathbb{I} + \mathbb{BA}^{-1})\mathbb{A}$. □

2.2.2 Equilibria of the dynamical system $\boldsymbol{x}(\cdot)$

In the remainder of the thesis, we use the following notations.

Notation 2.2.5. *For each critical point $\boldsymbol{c}$ of U, let $\mathbb{H}^{\boldsymbol{c}} = (\nabla^2 U)(\boldsymbol{c})$ denote the Hessian of U at $\boldsymbol{c}$ and let $\mathbb{L}^{\boldsymbol{c}} = (D\boldsymbol{\ell})(\boldsymbol{c})$ denote the Jacobian of $\boldsymbol{\ell}$ at $\boldsymbol{c}$.*

In this subsection, we analyze the equilibria of the dynamical system (2.8) by proving Theorem 2.1.1. First, we prove part (1) of the theorem.

Proof of part (1) of Theorem 2.1.1. Let $\boldsymbol{c} \in \mathbb{R}^d$ be a critical point of U. Since $\nabla U \cdot \boldsymbol{\ell} \equiv 0$ by (2.2), we have

$$\mathbf{0} \equiv \nabla [\, \nabla U \cdot \boldsymbol{\ell} \,] = (\nabla^2 U)\, \boldsymbol{\ell} + (D\boldsymbol{\ell})\, \nabla U \ .$$

Thus, we have $\mathbb{H}^{\boldsymbol{c}}\mathbb{L}^{\boldsymbol{c}} = \mathbf{0}$ as $\nabla U(\boldsymbol{c}) = \mathbf{0}$. Since $\mathbb{H}^{\boldsymbol{c}}$ is invertible as U is a Morse function, we get $\boldsymbol{\ell}(\boldsymbol{c}) = \mathbf{0}$. □

Now, we present a lemma that is a consequence of the condition (2.2) and part (1) of Theorem 2.1.1 that we have just proved.

Lemma 2.2.6. *For any critical point* $\boldsymbol{c}$ *of* U*, the matrix* $\mathbb{H}^{\boldsymbol{c}}\mathbb{L}^{\boldsymbol{c}}$ *is skew-symmetric.*

Proof. For small $\varepsilon > 0$ and $\boldsymbol{x} \in \mathbb{R}^d$, the Taylor expansion implies that

$$\nabla U(\boldsymbol{c} + \varepsilon \boldsymbol{x}) = \varepsilon\, \mathbb{H}^{\boldsymbol{c}}\, \boldsymbol{x} + O(\varepsilon^2) \quad \text{and} \quad \boldsymbol{\ell}(\boldsymbol{c} + \varepsilon \boldsymbol{x}) = \varepsilon\, \mathbb{L}^{\boldsymbol{c}}\, \boldsymbol{x} + O(\varepsilon^2) \ ,$$

since we have $\nabla U(\boldsymbol{c}) = \boldsymbol{\ell}(\boldsymbol{c}) = 0$ by part (1) of Theorem 2.1.1. By (2.2), we have

$$[\, \varepsilon \mathbb{H}^{\boldsymbol{c}}\, \boldsymbol{x} + O(\varepsilon^2) \,] \cdot [\, \varepsilon\, \mathbb{L}^{\boldsymbol{c}}\, \boldsymbol{x} + O(\varepsilon^2) \,] = 0 \ .$$

Dividing both sides by ε^2 and letting $\varepsilon \to 0$, we get $(\mathbb{H}^{\boldsymbol{c}}\, \boldsymbol{x}) \cdot (\mathbb{L}^{\boldsymbol{c}}\, \boldsymbol{x}) = 0$. Since the Hessian $\mathbb{H}^{\boldsymbol{c}}$ is symmetric, we can deduce that $\boldsymbol{x} \cdot \mathbb{H}^{\boldsymbol{c}}\, \mathbb{L}^{\boldsymbol{c}}\, \boldsymbol{x} = 0$ for all $\boldsymbol{x} \in \mathbb{R}^d$. This completes the proof. □

Now, we focus on parts (2) and (3) of Theorem 2.1.1.

Proof of parts (2) and (3) of Theorem 2.1.1. First, we focus on part (2). If $\boldsymbol{c}$ is a critical point of U, we have $(\nabla U + \boldsymbol{\ell})(\boldsymbol{c}) = \mathbf{0}$ by part (1); thus, $\boldsymbol{c}$ is an equilibrium of the dynamical system (2.8). On the other hand, suppose that $\boldsymbol{c}$ is an equilibrium, i.e., $(\nabla U + \boldsymbol{\ell})(\boldsymbol{c}) = \mathbf{0}$. Then, by (2.2), we have $0 = (\nabla U \cdot \boldsymbol{\ell})(\boldsymbol{c}) = -|\nabla U(\boldsymbol{c})|^2$; thus, $\nabla U(\boldsymbol{c}) = 0$.

For part (3), suppose that $\boldsymbol{c}$ is a local minimum of U such that the Hessian $\mathbb{H}^{\boldsymbol{c}}$ is positive definite. Since $\mathbb{H}^{\boldsymbol{c}}\mathbb{L}^{\boldsymbol{c}}$ is skew-symmetric by Lemma 2.2.6, we can insert $\mathbb{A} := \mathbb{H}^{\boldsymbol{c}}$ and $\mathbb{B} = \mathbb{L}^{\boldsymbol{c}}$ into Lemma 2.2.1 to conclude that all the eigenvalues of the matrix $\mathbb{H}^{\boldsymbol{c}} + \mathbb{L}^{\boldsymbol{c}}$ are either positive real or complex with a positive real part; hence, $\boldsymbol{c}$ a is stable equilibrium of the dynamical system $\boldsymbol{x}(\cdot)$ since $\mathbb{H}^{\boldsymbol{c}} + \mathbb{L}^{\boldsymbol{c}}$ is the Jacobian of the vector field $\nabla U + \boldsymbol{\ell}$ at $\boldsymbol{c}$.

For the other direction, suppose that $\boldsymbol{c}$ is a stable equilibrium of the dynamical system (2.8), i.e., the matrix $\mathbb{H}^{\boldsymbol{c}} + \mathbb{L}^{\boldsymbol{c}}$ is positive definite in the sense that

$$\boldsymbol{x} \cdot [\mathbb{H}^{\boldsymbol{c}} + \mathbb{L}^{\boldsymbol{c}}]\boldsymbol{x} > 0 \quad \text{for all } \boldsymbol{x} \neq \mathbf{0} \, . \tag{2.14}$$

Suppose now that the symmetric matrix $\mathbb{H}^{\boldsymbol{c}}$ is not positive definite so that there is a negative eigenvalue $-\lambda < 0$. Let $\boldsymbol{v}$ be the corresponding unit eigenvector. Since $\mathbb{H}^{\boldsymbol{c}}\mathbb{L}^{\boldsymbol{c}}$ is skew-symmetric by Lemma 2.2.6 and $\mathbb{H}^{\boldsymbol{c}}$ is symmetric, we have

$$2(\mathbb{H}^{\boldsymbol{c}})^2 = \mathbb{H}^{\boldsymbol{c}}[\mathbb{H}^{\boldsymbol{c}} + \mathbb{L}^{\boldsymbol{c}}] + [\mathbb{H}^{\boldsymbol{c}} + (\mathbb{L}^{\boldsymbol{c}})^\dagger]\mathbb{H}^{\boldsymbol{c}} \, ,$$

and thus we get

$$\begin{aligned} 2\lambda^2 &= \boldsymbol{v} \cdot 2(\mathbb{H}^{\boldsymbol{c}})^2\boldsymbol{v} = \mathbb{H}^{\boldsymbol{c}}\boldsymbol{v} \cdot [\mathbb{H}^{\boldsymbol{c}} + \mathbb{L}^{\boldsymbol{c}}]\boldsymbol{v} + \boldsymbol{v} \cdot [\mathbb{H}^{\boldsymbol{c}} + (\mathbb{L}^{\boldsymbol{c}})^\dagger]\mathbb{H}^{\boldsymbol{c}}\boldsymbol{v} \\ &= -\lambda\boldsymbol{v} \cdot [\mathbb{H}^{\boldsymbol{c}} + \mathbb{L}^{\boldsymbol{c}} + \mathbb{H}^{\boldsymbol{c}} + (\mathbb{L}^{\boldsymbol{c}})^\dagger]\boldsymbol{v} = -2\lambda\boldsymbol{v} \cdot [\mathbb{H}^{\boldsymbol{c}} + \mathbb{L}^{\boldsymbol{c}}]\boldsymbol{v} \, . \end{aligned}$$

This contradicts with (2.14) and therefore $\mathbb{H}^{\boldsymbol{c}} + \mathbb{L}^{\boldsymbol{c}}$ must be positive definite. This completes the proof. □

2.2.3 Saddle points of dynamical system (2.8)

Now, we focus on the saddle points. First, we prove that, for a saddle point $\boldsymbol{\sigma} \in \mathbb{R}^d$, the matrix $\mathbb{H}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}}$ has only one negative eigenvalue as the matrix $\mathbb{H}^{\boldsymbol{\sigma}}$ has only one negative eigenvalue.

Lemma 2.2.7. *Let $\boldsymbol{\sigma} \in \mathbb{R}^d$ be a saddle point such that $\mathbb{H}^{\boldsymbol{\sigma}}$ has only one negative eigenvalue. Then, the matrix $\mathbb{H}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}}$ has only one negative eigenvalue and is invertible.*

Proof. Let $\boldsymbol{\sigma} \in \mathbb{R}^d$ be a saddle point so that $\mathbb{H}^{\boldsymbol{\sigma}}$ has exactly one negative eigenvalue by the Morse lemma. Then, we can insert $\mathbb{A} := \mathbb{H}^{\boldsymbol{\sigma}}$ and $\mathbb{B} := \mathbb{L}^{\boldsymbol{\sigma}}$ into Lemma 2.2.4 owing to Lemma 2.2.6, and we can conclude that the matrix $\mathbb{H}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}}$ has only one negative eigenvalue and is invertible. □

Next, we compares the unique negative eigenvalues of $\mathbb{H}^{\boldsymbol{\sigma}}$ and $\mathbb{H}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}}$ for saddle point $\boldsymbol{\sigma} \in \mathbb{R}^d$.

Lemma 2.2.8. *Let $\lambda^{\boldsymbol{\sigma}}$ and $\mu^{\boldsymbol{\sigma}}$ be unique negative eigenvalues of $\mathbb{H}^{\boldsymbol{\sigma}}$ and $\mathbb{H}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}}$, respectively. Then, we have $\mu^{\boldsymbol{\sigma}} \geq \lambda^{\boldsymbol{\sigma}}$*

Proof. Denote by $-\lambda_1, \lambda_2, \dots, \lambda_d$ the eigenvalues of the symmetric matrix $\mathbb{H}^{\boldsymbol{\sigma}}$, where $\lambda_1, \dots, \lambda_d > 0$. Thus, $\lambda^{\boldsymbol{\sigma}} = \lambda_1$. Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_d$ denote the normal eigenvectors of $\mathbb{H}^{\boldsymbol{\sigma}}$ corresponding to the eigenvalues $-\lambda_1, \dots, \lambda_d$, respectively. Let $\boldsymbol{v}$ denote the unit eigenvector of $\mathbb{H}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}}$ corresponding to the unique negative eigenvalue $-\mu^{\boldsymbol{\sigma}}$ and write $\boldsymbol{v} = \sum_{i=1}^{d} a_i \boldsymbol{u}_i$. Since $\mathbb{H}^{\boldsymbol{\sigma}}\mathbb{L}^{\boldsymbol{\sigma}}$ is skew-symmetric by Lemma 2.2.6, we have

$$|\mathbb{H}^{\boldsymbol{\sigma}}\boldsymbol{v}|^2 \;=\; \boldsymbol{v} \cdot \mathbb{H}^{\boldsymbol{\sigma}}(\mathbb{H}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}})\boldsymbol{v} \;=\; -\mu^{\boldsymbol{\sigma}}\boldsymbol{v} \cdot \mathbb{H}^{\boldsymbol{\sigma}}\boldsymbol{v} \;.$$

Using the above-mentioned notations, we can rewrite this identity as

$$\sum_{i=1}^{d} a_i^2 \lambda_i^2 \;=\; -\mu^{\boldsymbol{\sigma}} \Big[-a_1^2 \lambda_1 + \sum_{i=2}^{d} a_i^2 \lambda_i \Big] \;. \tag{2.15}$$

First, suppose that $a_1 = 0$. Then, we have $\sum_{i=2}^{d} a_i^2 \lambda_i^2 = -\mu^{\boldsymbol{\sigma}} \sum_{i=2}^{d} a_i^2 \lambda_i$ and hence we get $a_2 = \cdots = a_d = 0$. This implies that $\boldsymbol{v} = 0$, which is a contradiction.

Thus, $a_1 \neq 0$. By (2.15), we have

$$a_1^2\lambda_1^2 \le \sum_{i=1}^{d} a_i^2\lambda_i^2 = \mu^{\boldsymbol{\sigma}} a_1^2\lambda_1 - \mu^{\boldsymbol{\sigma}} \sum_{i=2}^{d} a_i^2\lambda_i \le \mu^{\boldsymbol{\sigma}} a_1^2\lambda_1 \,.$$

Since $a_1 \neq 0$, we get $\mu^{\boldsymbol{\sigma}} \ge \lambda_1 = \lambda^{\boldsymbol{\sigma}}$. □

2.3 Properties of diffusion process

In this section, we prove the basic properties of the diffusion process $\boldsymbol{x}_\epsilon(\cdot)$.

2.3.1 Positive recurrence and non-explosion

First, we establish a technical lemma.

Lemma 2.3.1. *For all $\epsilon > 0$, there exists $r_0 = r_0(\epsilon) > 0$ such that $(\mathscr{L}_\epsilon U)(\boldsymbol{x}) \le -3$ for all $\boldsymbol{x} \notin \mathcal{D}_{r_0}(\mathbf{0})$.*

Proof. By (2.5) and (2.6), we can take r_0 to be sufficiently large such that

$$|\,\nabla U(\boldsymbol{x})\,| - 2\,\Delta U(\boldsymbol{x}) > \frac{\epsilon}{2} \quad \text{and} \quad |\,\nabla U(\boldsymbol{x})\,| > 2 \tag{2.16}$$

for all $\boldsymbol{x} \notin \mathcal{D}_{r_0}(\mathbf{0})$. Then, for $\boldsymbol{x} \notin \mathcal{D}_{r_0}(\mathbf{0})$, we have

$$\Delta U(\boldsymbol{x}) \le -\frac{\epsilon}{4} + \frac{1}{2}\,|\,\nabla U(\boldsymbol{x})\,| \le \frac{1}{4\epsilon}\,|\,\nabla U(\boldsymbol{x})\,|^2 \,.$$

Therefore,

$$(\mathscr{L}_\epsilon U)(\boldsymbol{x}) = -|\,\nabla U(\boldsymbol{x})\,|^2 + \epsilon\,\Delta U(\boldsymbol{x}) \le -\frac{3}{4}\,|\,\nabla U(\boldsymbol{x})\,|^2 \le -3 \,.$$

The last inequality follows from the second condition of (2.16). □

Now, we prove Theorem 2.1.2

Proof of Theorem 2.1.2. First, we prove part (1), i.e., the non-explosion property. By [87, Theorem at page 197], it suffices to check that there exists a smooth function $u : \mathbb{R}^d \to (0, \infty)$ such that

$$u(\boldsymbol{x}) \to \infty \text{ as } \boldsymbol{x} \to \infty \quad \text{and} \quad (\mathscr{L}_\epsilon u)(\boldsymbol{x}) \le u(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in \mathbb{R}^d \ . \tag{2.17}$$

We claim that $u = U + k_\epsilon$ with a sufficiently large constant k_ϵ satisfies all these conditions. First, we take k_ϵ to be sufficiently large such that $u > 0$. The former condition of (2.17) is immediate from (2.4). Now, it suffices to check the second condition. By Lemma 2.3.1, the function $\mathscr{L}_\epsilon u = \mathscr{L}_\epsilon U$ is bounded from above. Denote this bound by M_ϵ and then take k_ϵ to be sufficiently large such that $u(\boldsymbol{x}) > M_\epsilon$ for all $\boldsymbol{x} \in \mathbb{R}^d$. Then, the second condition of (2.17) follows.

The positive recurrence of $\boldsymbol{x}_\epsilon(\cdot)$ follows from Lemma 2.3.1 and [79, Theorem 6.1.3]. □

2.3.2 Invariant measure

By a slight abuse of notation, we write $\mu_\epsilon(\boldsymbol{x}) = Z_\epsilon^{-1} e^{-U(\boldsymbol{x})/\epsilon}$ (cf. (1.6)). Now, we prove Theorem 2.1.3. We can observe from the expression (2.10) of the generator $\mathscr{L}_\epsilon$ that the adjoint generator $\mathscr{L}_\epsilon^{\mathrm{a}}$ of $\mathscr{L}_\epsilon$ with respect to the Lebesgue measure $d\boldsymbol{x}$ can be written as

$$\mathscr{L}_\epsilon^{\mathrm{a}} f = \epsilon \nabla \cdot [\, e^{-U/\epsilon} \nabla (e^{U/\epsilon} f)\,] + \boldsymbol{\ell} \cdot \nabla (e^{U/\epsilon} f) \ . \tag{2.18}$$

Proof of Theorem 2.1.3. First, we prove part (1). With the expression (2.18) and the explicit form of $\mu_\epsilon(\boldsymbol{x})$, we can check that $\mathscr{L}_\epsilon^{\mathrm{a}} \mu_\epsilon = 0$. Therefore, by [87, Theorem at page 254] and part (1) of Theorem 2.1.2, the measure $\mu_\epsilon(d\boldsymbol{x})$ is the invariant measure for the process $\boldsymbol{x}_\epsilon(\cdot)$. The uniqueness follows from [87, Theorem at page 259] and [87, Theorem at page 260].

For part (2), let us assume that $\mu_\epsilon(d\boldsymbol{x})$ is the invariant measure for the dynamics $\boldsymbol{z}_\epsilon(\cdot)$ given in (1.9) for all $\epsilon > 0$. Note that the generator associated with the process $\boldsymbol{z}_\epsilon(\cdot)$ acts on $f \in C^2(\mathbb{R}^d)$ as

$$\widetilde{\mathscr{L}}_\epsilon f \;=\; -\boldsymbol{b} \cdot \nabla f + \epsilon \Delta f \;.$$

Hence, its adjoint generator with respect to the Lebesgue measure is given by

$$\widetilde{\mathscr{L}}_\epsilon^{\mathrm{a}} f \;=\; \nabla \cdot [\, f \boldsymbol{b}\,] + \epsilon \Delta f \;.$$

By [87, Theorem at page 259], we must have $\widetilde{\mathscr{L}}_\epsilon^{\mathrm{a}} \mu_\epsilon = 0$. By writing $\boldsymbol{\ell} = \boldsymbol{b} - \nabla U$, this equation can be expressed as $e^{-U/\epsilon} \left[\dfrac{1}{\epsilon} \nabla U \cdot \boldsymbol{\ell} + \nabla \cdot \boldsymbol{\ell} \right] = 0$. Since this holds for all $\epsilon > 0$, the vector field $\boldsymbol{\ell}$ must satisfy both (2.2) and (2.3). $\square$

Chapter 3

Eyring–Kramers formula

This chapter is devoted to prove the Eyring–Kramers formula for the process $\boldsymbol{x}_\epsilon(\cdot)$ (Theorem 3.1.3). We also remark that in a recent study [64], the model considered in this thesis was investigated in view of the low-lying spectra. Sharp estimates were established for the exponentially small eigenvalues of the generator associated with the process $\boldsymbol{x}_\epsilon(\cdot)$. See Corollary 3.1.6 to understand how our discovery is related to the result presented in [64].

3.1 Main result

In this section, we explain the Eyring–Kramers formula for the diffusion process $\boldsymbol{x}_\epsilon(\cdot)$. The main result is stated in Theorem 3.1.3 (and Corollary 3.1.5 for the simple double-well case).

3.1.1 Structure of metastable valleys

Let $\mathcal{M}$ denote the set of local minima of U. The starting point $\boldsymbol{m}_0 \in \mathcal{M}$ of the process $\boldsymbol{x}_\epsilon(\cdot)$ is fixed throughout the chapter. Note that $\boldsymbol{m}_0$ is a stable equilibrium of $\boldsymbol{x}(\cdot)$ by Theorem 2.1.1.

Let us fix $H \in \mathbb{R}$ such that $U(\boldsymbol{m}_0) < H$ and define $\Sigma = \Sigma^H$ as the set of saddle points of level H:

$$\Sigma := \{ \boldsymbol{\sigma} : U(\boldsymbol{\sigma}) = H \text{ and } \boldsymbol{\sigma} \text{ is a saddle point of } U \} .$$

We take H such that $\Sigma \neq \emptyset$. We define

$$\mathcal{H} = \{ \boldsymbol{x} \in \mathbb{R}^d : U(\boldsymbol{x}) < H \} , \tag{3.1}$$

and we assume that $\mathcal{H}$ has multiple connected components; hence, metastability occurs.

We decompose $\mathcal{H} = \mathcal{H}_0 \cup \mathcal{H}_1$, where $\mathcal{H}_0$ is the connected component of $\mathcal{H}$ containing $\boldsymbol{m}_0$ and $\mathcal{H}_1 = \mathcal{H} \setminus \mathcal{H}_0$. Note that $\mathcal{H}_1$ may not be connected. Let $\mathcal{M}_0$ and $\mathcal{M}_1$ denote the sets of local minima belonging to $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Let $\mathcal{D}_r(\boldsymbol{x})$ denote an open ball in $\mathbb{R}^d$ centered at $\boldsymbol{x}$ with radius r, and define

$$\mathcal{U}_\epsilon := \bigcup_{\boldsymbol{m} \in \mathcal{M}_1} \mathcal{D}_\epsilon(\boldsymbol{m}) .$$

In this chapter, we focus on the sharp asymptotics of the mean of the transition time from $\boldsymbol{m}_0$ to $\mathcal{U}_\epsilon$. Figure 3.1 illustrates the notations introduced above.

Notation 3.1.1. *Since the sets such as* Σ *and* $\mathcal{U}_\epsilon$ *depend on* H*, we add the superscript* H *to these notations, e.g.,* Σ^H*, when we want to emphasize the dependency on* H*.*

3.1.2 Eyring–Kramers constant for $\boldsymbol{x}_\epsilon(\cdot)$

In this subsection, we fix $\boldsymbol{\sigma} \in \Sigma$ and suppose that $\mathbb{H}^{\boldsymbol{\sigma}}$ has only one negative eigenvalue $-\lambda^{\boldsymbol{\sigma}}$. In the Eyring–Kramers formula for the reversible process $\boldsymbol{y}_\epsilon(\cdot)$ obtained in [14], an important constant is the so-called *Eyring–Kramers constant* defined by

$$\omega_{\text{rev}}^{\boldsymbol{\sigma}} = \frac{\lambda^{\boldsymbol{\sigma}}}{2\pi\sqrt{-\det \mathbb{H}^{\boldsymbol{\sigma}}}} . \tag{3.2}$$

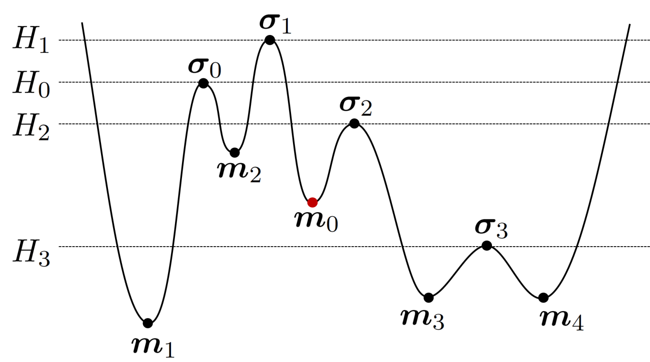


Figure 3.1: Example of landscape of the potential function U with five local minima $\{\boldsymbol{m}_i : 0 \le i \le 4\}$ and four saddle points $\{\boldsymbol{\sigma}_i : 0 \le i \le 3\}$. We assume that $U(\boldsymbol{m}_3) = U(\boldsymbol{m}_4)$ and write $H_i = U(\boldsymbol{\sigma_i})$, $0 \le i \le 3$. Our objective is to compute the transition time from the local minimum $\boldsymbol{m}_0$ to other local minima. We can select the level H according to our detailed objective. By taking $H = H_1$, we have $\mathcal{M}_1 = \{\boldsymbol{m}_1, \boldsymbol{m}_2\}$; hence, we focus on the transition time from $\boldsymbol{m}_0$ to $\mathcal{D}_\epsilon(\boldsymbol{m}_1) \cup \mathcal{D}_\epsilon(\boldsymbol{m}_2)$. This occurs at the level of H_1 since the process must pass through $\boldsymbol{\sigma}_1$ to make such a transition. For this case, we have $\mathcal{M}_0 = \{\boldsymbol{m}_0, \boldsymbol{m}_3, \boldsymbol{m}_4\}$ and $\mathcal{M}_0^\star = \{\boldsymbol{m}_3, \boldsymbol{m}_4\}$. On the other hand, by taking $H = H_2$, we have $\mathcal{M}_1 = \{\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4\}$. For this case, we compute the escape time from the metastable valley around $\boldsymbol{m}_0$. The selection $H = H_3$ is not available since the condition $U(\boldsymbol{m}_0) < H$ is violated; hence $\mathcal{H}$ does not contain $\boldsymbol{m}_0$. This level is meaningful when we start from, e.g., $\boldsymbol{m}_3$. Finally, the selection $H = H_0$ is not appropriate as Σ_0 becomes an empty set. For this case, we refer to Remark 3.1.4 (4) for further details.

Now, we introduce the corresponding constant for the process $\boldsymbol{x}_\epsilon(\cdot)$. By Lemma 2.2.7, $\mathbb{H}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}}$ has only one negative eigenvalue and is invertible. Let $-\mu^{\boldsymbol{\sigma}}$ denote the unique negative eigenvalue obtained in this lemma and define the Eyring–Kramers constant at $\boldsymbol{\sigma}$ by

$$\omega^{\boldsymbol{\sigma}} = \frac{\mu^{\boldsymbol{\sigma}}}{2\pi\sqrt{-\det \mathbb{H}^{\boldsymbol{\sigma}}}} . \tag{3.3}$$

Then, by Lemma 2.2.8, we can prove the following comparison result for the Eyring–Kramers constant.

Lemma 3.1.2. *We have* $\omega^{\boldsymbol{\sigma}} \ge \omega^{\boldsymbol{\sigma}}_{\rm rev}$.

In Corollary 3.1.7, we prove that the process $\boldsymbol{x}_\epsilon(\cdot)$ is faster than $\boldsymbol{y}_\epsilon(\cdot)$ on the basis of this comparison result.

3.1.3 Eyring–Kramers formula for $\boldsymbol{x}_\epsilon(\cdot)$

For $\mathcal{A} \subset \mathbb{R}^d$, let $\overline{\mathcal{A}}$ denote the closure of $\mathcal{A}$. Define

$$\Sigma_0 = \overline{\mathcal{H}_0} \cap \overline{\mathcal{H}_1} \subset \Sigma . \tag{3.4}$$

We assume that $\Sigma_0 \neq \emptyset$[1]. For each $\boldsymbol{\sigma} \in \Sigma_0$, the Hessian $\mathbb{H}^{\boldsymbol{\sigma}}$ has only one negative eigenvalue as a consequence of the Morse lemma (cf. [71, Lemma 2.2]); hence, the Eyring–Kramers constant $\omega^{\boldsymbol{\sigma}}$ at $\boldsymbol{\sigma} \in \Sigma_0$ can be defined as in the previous subsection. Then, define

$$\omega_0 = \sum_{\boldsymbol{\sigma} \in \Sigma_0} \omega^{\boldsymbol{\sigma}} . \tag{3.5}$$

[1]The case $\Sigma_0 = \emptyset$ may occur, for instance, if we take $H = H_0$ in Figure 3.1. We can deal with this situation using our result by modifying H; see Remark 3.1.4(4).

Let h_0 denote the minimum of U on $\mathcal{H}_0$ and let $\mathcal{M}_0^\star$ denote the set of the deepest minima of U on $\mathcal{H}_0$:

$$\mathcal{M}_0^\star \;=\; \{\boldsymbol{m} \in \mathcal{M}_0 : U(\boldsymbol{m}) = h_0\}\ . \tag{3.6}$$

Define

$$\nu_0 \;=\; \sum_{\boldsymbol{m}\in\mathcal{M}_0^\star} \frac{1}{\sqrt{\det \mathbb{H}^{\boldsymbol{m}}}}\ . \tag{3.7}$$

Now, we are ready to state the Eyring–Kramers formula for the non-reversible process $\boldsymbol{x}_\epsilon(\cdot)$, which is the main result of the current chapter. For a sequence $(a_\epsilon)_{\epsilon>0}$ of real numbers, we write $a_\epsilon = o_\epsilon(1)$ if $\lim_{\epsilon\to 0} a_\epsilon = 0$.

Theorem 3.1.3. *We have*

$$\mathbb{E}^\epsilon_{\boldsymbol{m}_0}[\,\tau_{\mathcal{U}_\epsilon}\,] \;=\; [\,1 + o_\epsilon(1)\,]\,\frac{\nu_0}{\omega_0}\,\exp\frac{H - h_0}{\epsilon}\ . \tag{3.8}$$

Remark 3.1.4. *We state the following with regard to Theorem 3.1.3.*

1. *Heuristically, the process $\boldsymbol{x}_\epsilon(\cdot)$ starting at $\boldsymbol{m}_0$ first mixes among the neighborhoods of minima of $\mathcal{M}_0^\star$, and then makes a transition to $\mathcal{U}_\epsilon$ by passing through a neighborhood of the saddle in Σ_0 according to the Freidlin–Wentzell theory. This is the reason that the formula* (3.8) *depends on the local properties of the potential U at $\mathcal{M}_0^\star$ and Σ_0. A remarkable fact regarding the formula* (3.8) *is that the sub-exponential prefactor is dominated only by these local properties. This is mainly because the invariant measure is the Gibbs measure $\mu_\epsilon(d\boldsymbol{x})$. It is observed in [8] that an additional factor called "non-Gibbsianness" of the process should be introduced in the general case (i.e., in the analysis of the metastable behavior of the process $\boldsymbol{z}_\epsilon(\cdot)$).*

2. *Theorem 3.1.3 is a generalization of [14, Theorem 3.2], as the reversible case is the special $\boldsymbol{\ell} = \mathbf{0}$ case of our model. Moreover, a careful reading*

of our arguments reveals that the error term $o_\epsilon(1)$ *is indeed* $O(\epsilon^{1/2}\log\frac{1}{\epsilon})$ *which is the one that appeared in [14, Theorems 3.1 and 3.2].*

3. *The constants* ω_0, ν_0, *and* h_0 *and the set* $\mathcal{U}_\epsilon$ *are not changed if we take a different starting point* $\boldsymbol{m}_0' \in \mathcal{M}_0$. *In view of Theorem 3.1.3, this implies that all the transition times from a point in* $\mathcal{M}_0$ *to* $\mathcal{U}_\epsilon$ *are asymptotically the same. For instance, if we take* $H = H_1$ *in Figure 3.1, the expectation of the hitting time* $\tau_{\mathcal{D}_\epsilon(\boldsymbol{m}_1)\cup\mathcal{D}_\epsilon(\boldsymbol{m}_2)}$ *is asymptotically the same for the starting points* $\boldsymbol{m}_0$, $\boldsymbol{m}_3$ *and* $\boldsymbol{m}_4$. *This is because the process* $\boldsymbol{x}_\epsilon(\cdot)$ *sufficiently mixes in the valley* $\mathcal{H}_0$ *before moving to another valley.*

4. *Consider the case* $H = H_2$, *where the potential* U *is given as Figure 3.1 so that we have* $\mathcal{U}_\epsilon^{H_2} = \{\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3, \boldsymbol{m}_4\}$. *However, in time scale* $\exp\left\{\frac{H_2 - h_0}{\epsilon}\right\}$, *the diffusion process cannot move to the neighborhoods of* $\boldsymbol{m}_1$ *and* $\boldsymbol{m}_2$, *since* $\boldsymbol{\sigma}_2$ *is the only saddle point in* $\Sigma_0^{H_2}$ *and* $\boldsymbol{m}_3$ *and* $\boldsymbol{m}_4$ *are the only minima in the connected components of* $\mathcal{H}_1$ *whose boundary contains* $\boldsymbol{\sigma}_2$. *Our proof verifies this as well.*

5. *We can tune* H *such that* $\boldsymbol{m}_0$ *is the unique local minimum of* $\mathcal{H}_0$. *For example, in Figure 3.1, we can achieve this by selecting* $H = H_2$. *Then, the formula* (3.8) *becomes the asymptotics of the transition time from* $\boldsymbol{m}_0$ *to one of the other local minima, and this is the classic form of the Eyring–Kramers formula. We remark that all the existing studies [14, 51] on the Eyring–Kramers formula for metastable diffusion processes have dealt with only this situation. On the other hand, our result is more comprehensive in that we analyzed all the possible levels by carefully investigating the equilibrium potential in Section 3.5. Such a comprehensive result for a diffusion setting was barely known previously, see [70] where a similar setting along with the possibility of degenerate critical points has been discussed.*

6. *We use Notation 3.1.1 and suppose that* $\Sigma_0^{H_0} = \emptyset$*. Then, we have* $\mathbb{E}^{\epsilon}_{\boldsymbol{m}_0}[\,\tau_{\mathcal{U}_\epsilon}\,] \gg \exp\left\{\dfrac{H_0 - h_0}{\epsilon}\right\}$ *and the level* H_0 *is not appropriate to investigate this mean transition time. Instead, we define*

$$H^* \;=\; \sup\{\, H : \mathcal{U}_\epsilon^H = \mathcal{U}_\epsilon^{H_0}\,\}$$

so that at level H^* *the gate path from* $\boldsymbol{m}_0$ *to* $\mathcal{U}_\epsilon^{H_0}$ *firstly appears and hence* $\Sigma_0^{H^*} \neq \emptyset$*. Thus, we can estimate* $\mathbb{E}^{\epsilon}_{\boldsymbol{m}_0}[\,\tau_{\mathcal{U}_\epsilon^{H_0}}\,]$ *by taking* $H = H^*$*. For instance, in Figure 3.1, we have* $\Sigma_0^{H_0} = \emptyset$ *and* $H^* = H_1$*.*

7. *By selecting* $\boldsymbol{\ell}$ *appropriately, we can make* ω_0 *arbitrarily large.*

The proof of Theorem 3.1.3 is given in Section 3.3.

Double-well case

The Eyring–Kramers formula stated above has a simple form in the double-well case. Recall the double-well situation illustrated in Figure 1.3. For this case, the only meaningful selection of H is $U(\boldsymbol{\sigma})$, and $\Sigma_0 = \{\boldsymbol{\sigma}\}$ for this choice. With this H, we can interpret Theorem 3.1.3 as following corollary.

Corollary 3.1.5. *We have*

$$\mathbb{E}^{\epsilon}_{\boldsymbol{m}_1}[\,\tau_{\mathcal{D}_\epsilon(\boldsymbol{m}_2)}\,] \;=\; [\,1 + o_\epsilon(1)\,]\,\frac{2\pi}{\mu^{\boldsymbol{\sigma}}}\sqrt{\frac{-\det \mathbb{H}^{\boldsymbol{\sigma}}}{\det \mathbb{H}^{\boldsymbol{m}_1}}}\,\exp\frac{U(\boldsymbol{\sigma}) - U(\boldsymbol{m}_1)}{\epsilon}\;. \tag{3.9}$$

This is the classical form of the Eyring–Kramers formula. With this simple case, we explain why this result is a refinement of the Freidlin–Wentzell theory. By [29, Theorem 3.3.1], the quasi-potential $V(\cdot\,;\boldsymbol{m}_1)$ of the process $\boldsymbol{x}_\epsilon(\cdot)$ with respect to the local minimum $\boldsymbol{m}_1$ is given by $V(\boldsymbol{x};\boldsymbol{m}_1) = U(\boldsymbol{x}) - U(\boldsymbol{m}_1)$ on the domain of attraction of $\boldsymbol{m}_1$ with respect to the process $\boldsymbol{x}(\cdot)$. Hence, we can deduce the following large-deviation type result from the Freidlin–

Wentzell theory:

$$\lim_{\epsilon \to 0} \epsilon \, \log \mathbb{E}^{\epsilon}_{\boldsymbol{m}_1} [\, \tau_{\mathcal{D}_\epsilon(\boldsymbol{m}_2)} \,] \;=\; U(\boldsymbol{\sigma}) - U(\boldsymbol{m}_1) \;.$$

In the formula (3.9), we find the precise sub-exponential pre-factor associated with this large-deviation estimate.

We can also deduce from Corollary 3.1.5 a precise relation between the mean transition time and a low-lying spectrum of the generator $\mathscr{L}_\epsilon$ for the double-well case. In [64], the sharp asymptotics for the eigenvalue λ_ϵ of $\mathscr{L}_\epsilon$ with the smallest real part was obtained. Note that the generator $\mathscr{L}_\epsilon$ is not self-adjoint; hence, the eigenvalue might be a complex number.

Corollary 3.1.6. *For the double-well situation, we suppose that* $U(\boldsymbol{m}_1) \ge U(\boldsymbol{m}_2)$*. Let* λ_ϵ *denote the one with smallest real part among the non-zero eigenvalues of* $\mathscr{L}_\epsilon$*. Then, the following holds:*

$$\mathbb{E}^{\epsilon}_{\boldsymbol{m}_1} [\, \tau_{\mathcal{D}_\epsilon(\boldsymbol{m}_2)} \,] \;=\; \frac{1+o_\epsilon(1)}{\lambda_\epsilon} \;. \tag{3.10}$$

Note that λ_ϵ as well as the error term $o_\epsilon(1)$ in (3.10) is in general a non-real complex number. Suprisingly, it is verified in [64, Remark 1.10] that λ_ϵ is indeed a real number if U is a double-well potential and ϵ is sufficiently small. We remark that the inverse relationship between the low-lying spectrum and the mean transition time as in (3.10) has been rigorously verified in [13, 14] for a wide class of reversible models including $\boldsymbol{y}_\epsilon(\cdot)$.

Comparison with reversible case

The Eyring–Kramers formula for the reversible process $\boldsymbol{y}_\epsilon(\cdot)$ has been shown in [14, Theorem 3.2]. We can also recover[2] this result by inserting $\boldsymbol{\ell} = \boldsymbol{0}$. We now explain this result using our terminology and we provide a comparison

[2] Indeed, our result with $\boldsymbol{\ell} = \boldsymbol{0}$ strictly contains what has been established in [14]. See Remark 3.1.4-(3).

between reversible and non-reversible cases. Write

$$\omega_{0,\,\mathrm{rev}} \;=\; \sum_{\boldsymbol{\sigma}\in\Sigma_0} \omega_{\mathrm{rev}}^{\boldsymbol{\sigma}}\;,$$

and let $\mathbb{E}_{\boldsymbol{x},\,\mathrm{rev}}^{\epsilon}$ denote the expectation with respect to the reversible process $\boldsymbol{y}_\epsilon(\cdot)$ starting from $\boldsymbol{x}\in\mathbb{R}^d$. Then, as a consequence of Theorem 3.1.3 with $\boldsymbol{\ell}=\boldsymbol{0}$, we get the following corollary.

Corollary 3.1.7. *The following holds:*

$$\mathbb{E}_{\boldsymbol{m}_0,\,\mathrm{rev}}^{\epsilon}[\,\tau_{\mathcal{U}_\epsilon}\,] \;=\; [\,1+o_\epsilon(1)\,]\,\frac{\nu_0}{\omega_{0,\,\mathrm{rev}}}\;\exp\frac{H-h_0}{\epsilon}\;.$$

Therefore, we have $\mathbb{E}_{\boldsymbol{m}_0}^{\epsilon}[\,\tau_{\mathcal{U}_\epsilon}\,]\le\mathbb{E}_{\boldsymbol{m}_0,\,\mathrm{rev}}^{\epsilon}[\,\tau_{\mathcal{U}_\epsilon}\,]$ *for all small enough* ϵ*.*

Proof. The first assertion follows immediately from the fact that $\omega_{\mathrm{rev}}^{\boldsymbol{\sigma}}$ defined in (3.2) corresponds to $\omega^{\boldsymbol{\sigma}}$ with $\boldsymbol{\ell}=\boldsymbol{0}$. The second assertion follows from Lemma 3.1.2 which implies that $\omega_0\ge\omega_{0,\,\mathrm{rev}}$. □

In view of the fact that the dynamics $\boldsymbol{y}_\epsilon(\cdot)$ plays a crucial role in the stochastic gradient descent algorithm, we might be able to accelerate this algorithm by adding a suitable orthogonal, incompressible vector field to the drift part.

3.2 Potential theory

In this section, we introduce the potential theory related to the process $\boldsymbol{x}_\epsilon(\cdot)$. As in the previous studies, we prove the Eyring–Kramers formula based on the relation between the mean transition time and the potential theoretic notions, and this relation is recalled in Proposition 3.3.1. The difficulty, especially for the non-reversible process, in using this formula arises from the estimation of the capacity term appearing in the formula. In this chapter, as explained in the Introduction section, we develop a novel and simple way to estimate the capacity. In this section, we explain a formula given by Proposition 3.2.2 for the capacity which plays a crucial role in our method. We remark that this formula itself is not new; the method for handling this formula is the innovation of the current study, and will be explained in the remainder of this chapter. To explain this formula, we start by introducing the adjoint process, equilibrium potential, and capacity.

3.2.1 Adjoint process

The adjoint operator $\mathscr{L}_\epsilon^*$ of $\mathscr{L}_\epsilon$ with respect to the invariant measure $\mu_\epsilon(d\boldsymbol{x})$ can be written as

$$\mathscr{L}_\epsilon^* f \;=\; \epsilon e^{U/\epsilon} \nabla \cdot \left[e^{-U/\epsilon} \left(\nabla f + \frac{1}{\epsilon} f \boldsymbol{\ell} \right) \right] \;=\; -(\nabla U - \boldsymbol{\ell}) \cdot \nabla f + \epsilon \, \Delta f \;. \tag{3.11}$$

Note that the generator $\mathscr{L}_\epsilon^{\mathrm{a}}$ defined in (2.18) is an adjoint with respect to the Lebesgue measure, instead of $\mu_\epsilon(d\boldsymbol{x})$. The adjoint process $\boldsymbol{x}_\epsilon^*(\cdot)$ is the diffusion process associated with the generator $\mathscr{L}_\epsilon^*$; hence, it is given by the SDE

$$d\boldsymbol{x}_\epsilon^*(t) \;=\; -(\nabla U - \boldsymbol{\ell})(\boldsymbol{x}_\epsilon^*(t)) \, dt + \sqrt{2\epsilon} \, d\boldsymbol{w}_t \;.$$

Let $\mathbb{P}_{\boldsymbol{x}}^{\epsilon,*}$ denote the law of the process $\boldsymbol{x}_\epsilon^*(\cdot)$. We can prove that the process $\boldsymbol{x}_\epsilon^*(\cdot)$ is positive recurrent and has the unique invariant measure $\mu_\epsilon(d\boldsymbol{x})$ by an argument that is identical to that for $\boldsymbol{x}_\epsilon(\cdot)$.

3.2.2 Equilibrium potentials and capacities

In the remainder of this section, we fix two disjoint non-empty bounded domains $\mathcal{A}$ and $\mathcal{B}$ of $\mathbb{R}^d$ with $C^{2,\alpha}$-boundaries for some $\alpha \in (0, 1)$ such that the perimeters $\sigma(\mathcal{A})$ and $\sigma(\mathcal{B})$ are finite, and $d(\mathcal{A}, \mathcal{B}) > 0$. Now, we introduce the equilibrium potential and capacity between the two sets $\mathcal{A}$ and $\mathcal{B}$. Write $\Omega = (\overline{\mathcal{A}} \cup \overline{\mathcal{B}})^c$ so that $\partial\Omega = \partial\mathcal{A} \cup \partial\mathcal{B}$.

The equilibrium potentials $h_{\mathcal{A},\mathcal{B}}^\epsilon$, $h_{\mathcal{A},\mathcal{B}}^{\epsilon,*} : \mathbb{R}^d \to \mathbb{R}$ between $\mathcal{A}$ and $\mathcal{B}$ with respect to the processes $\boldsymbol{x}_\epsilon(\cdot)$ and $\boldsymbol{x}_\epsilon^*(\cdot)$ are given by

$$h^\epsilon_{\mathcal{A},\,\mathcal{B}}(\boldsymbol{x}) \;=\; \mathbb{P}^\epsilon_{\boldsymbol{x}}\,[\,\tau_{\mathcal{A}} < \tau_{\mathcal{B}}\,] \quad\text{and}\quad h^{\epsilon,\,*}_{\mathcal{A},\,\mathcal{B}}(\boldsymbol{x}) \;=\; \mathbb{P}^{\epsilon,\,*}_{\boldsymbol{x}}\,[\,\tau_{\mathcal{A}} < \tau_{\mathcal{B}}\,]$$

for $\boldsymbol{x} \in \mathbb{R}^d$, respectively.

The capacity between $\mathcal{A}$ and $\mathcal{B}$ with respect to the processes $\boldsymbol{x}_\epsilon(\cdot)$ and $\boldsymbol{x}_\epsilon^*(\cdot)$ are respectively defined by

$$\begin{aligned}\mathrm{cap}_\epsilon(\mathcal{A},\,\mathcal{B}) \;&=\; \epsilon\int_{\partial\mathcal{A}} (\nabla h^\epsilon_{\mathcal{A},\,\mathcal{B}}\cdot \boldsymbol{n}_\Omega)\,\sigma(d\mu_\epsilon) \quad\text{and} \\ \mathrm{cap}^*_\epsilon(\mathcal{A},\,\mathcal{B}) \;&=\; \epsilon\int_{\partial\mathcal{A}} (\nabla h^{\epsilon,\,*}_{\mathcal{A},\,\mathcal{B}}\cdot \boldsymbol{n}_\Omega)\,\sigma(d\mu_\epsilon)\;,\end{aligned} \tag{3.12}$$

where $\boldsymbol{n}_\Omega(\boldsymbol{x})$ is the outward normal vector to Ω at $\boldsymbol{x}$; hence, $\boldsymbol{n}_\Omega(\boldsymbol{x}) = -\boldsymbol{n}_{\mathcal{A}}(\boldsymbol{x})$ for $\boldsymbol{x} \in \partial\mathcal{A}$. Here, $\int_{\partial\mathcal{A}} f\,\sigma(d\mu_\epsilon)$ is a shorthand of $\int_{\partial\mathcal{A}} f(\boldsymbol{x})\,\mu_\epsilon(\boldsymbol{x})\,\sigma(d\boldsymbol{x})$. These capacities exhibit the following well-known properties.

Lemma 3.2.1. *The following properties hold.*

1. *We have*

$$\mathrm{cap}_\epsilon(\mathcal{A},\,\mathcal{B}) \;=\; \mathrm{cap}^*_\epsilon(\mathcal{A},\,\mathcal{B}) \;=\; \mathrm{cap}^*_\epsilon(\mathcal{B},\,\mathcal{A}) \;=\; \mathrm{cap}_\epsilon(\mathcal{B},\,\mathcal{A})\;.$$

2. *We have*

$$\mathrm{cap}_\epsilon(\mathcal{A},\,\mathcal{B}) \;=\; \epsilon\int_\Omega |\nabla h^\epsilon_{\mathcal{A},\,\mathcal{B}}|^2\,d\mu_\epsilon \;=\; \epsilon\int_\Omega |\nabla h^{\epsilon,\,*}_{\mathcal{A},\,\mathcal{B}}|^2\,d\mu_\epsilon\;.$$

Proof. We refer to [51, Lemmas 3.2 and 3.1] for the proof of parts (1) and (2), respectively. □

3.2.3 Representation of capacity

We keep the sets $\mathcal{A}$, $\mathcal{B}$, and Ω from the previous subsection. Then, for a function $f : \mathbb{R}^d \to \mathbb{R}$ that is differentiable at $\boldsymbol{x} \in \mathbb{R}^d$, we define a vector field

Φ_f at $\boldsymbol{x}$ as

$$\Phi_f(\boldsymbol{x}) \;=\; \nabla f(\boldsymbol{x}) + \frac{1}{\epsilon} f(\boldsymbol{x})\, \boldsymbol{\ell}(\boldsymbol{x})\;. \tag{3.13}$$

Let $C_0^\infty(\mathbb{R}^d)$ denote the class of smooth and compactly supported functions on $\mathbb{R}^d$. Let

$$\mathscr{C}_{\mathcal{A},\,\mathcal{B}} \;=\; \{\, f \in C_0^\infty(\mathbb{R}^d) : f \equiv 1 \text{ on } \mathcal{A}\;, \quad f \equiv 0 \text{ on } \mathcal{B}\,\}. \tag{3.14}$$

Hence, for $f \in \mathscr{C}_{\mathcal{A},\,\mathcal{B}}$, the vector field Φ_f is defined on $\mathbb{R}^d$. The following expression plays a crucial role in the estimation of the capacity.

Proposition 3.2.2. *For all $f \in \mathscr{C}_{\mathcal{A},\,\mathcal{B}}$, we have*

$$\epsilon \int_\Omega [\, \Phi_f \cdot \nabla h^\epsilon_{\mathcal{A},\,\mathcal{B}}\,]\, d\mu_\epsilon \;=\; \mathrm{cap}_\epsilon(\mathcal{A},\, \mathcal{B})\;. \tag{3.15}$$

Proof. Since f is compactly supported, we can apply the divergence theorem to rewrite the left-hand side of (3.15) as

$$\epsilon \int_{\partial\Omega} f\, [\, \nabla h^\epsilon_{\mathcal{A},\,\mathcal{B}} \cdot \boldsymbol{n}_\Omega\,]\, \sigma(d\mu_\epsilon) \;-\; \int_\Omega f\, (\mathscr{L}_\epsilon h^\epsilon_{\mathcal{A},\,\mathcal{B}})\, d\mu_\epsilon\;.$$

Since $f = \mathbf{1}_{\partial\mathcal{A}}$ on $\partial\Omega$ by the condition $f \in \mathscr{C}_{\mathcal{A},\,\mathcal{B}}$, the first term of the above-mentioned expression is equal to $\mathrm{cap}_\epsilon(\mathcal{A},\, \mathcal{B})$ by (3.12). On the other hand, the second integral is 0 since $\mathscr{L}_\epsilon h^\epsilon_{\mathcal{A},\,\mathcal{B}} \equiv 0$ on Ω by the property of the equilibrium potential. □

3.3 Proof of Eyring–Kramers formula

In this section, we prove the Eyring–Kramers formula stated in Theorem 3.1.3 up to the construction of a test function and analysis of the equilibrium potential.

3.3.1 Proof of Theorem 3.1.3

For convenience of notation, we will use the following abbreviations for the capacity and equilibrium potential between a small ball around the minimum $\boldsymbol{m}_0$ and $\mathcal{U}_\epsilon$:

$$\begin{aligned} \mathrm{cap}_\epsilon \;&=\; \mathrm{cap}_\epsilon(\,\mathcal{D}_\epsilon(\boldsymbol{m}_0), \mathcal{U}_\epsilon\,)\;, \\ h_\epsilon(\cdot) \;&=\; h^\epsilon_{\mathcal{D}_\epsilon(\boldsymbol{m}_0),\mathcal{U}_\epsilon}(\cdot) \quad \text{and} \quad h^*_\epsilon(\cdot) \;=\; h^{\epsilon,*}_{\mathcal{D}_\epsilon(\boldsymbol{m}_0),\mathcal{U}_\epsilon}(\cdot)\;. \end{aligned} \tag{3.16}$$

The proof of the Eyring–Kramers formula relies on the following formula regarding the mean transition time.

Proposition 3.3.1. *We have*

$$\mathbb{E}^\epsilon_{\boldsymbol{m}_0}[\,\tau_{\mathcal{U}_\epsilon}\,] \;=\; [\,1+o_\epsilon(1)\,]\,\frac{1}{\mathrm{cap}_\epsilon}\int_{\mathbb{R}^d} h^*_\epsilon\,d\mu_\epsilon\;. \tag{3.17}$$

This remarkable relation between the mean transition time and the potential theoretic notions was first observed in [14, Proposition 6.1] for the reversible case. Then, it was extended to the general non-reversible case in [51, Lemma 9.2]. Our proof is identical to that of the latter case; hence, we omit the details. Now, the proof of Theorem 3.1.3 is reduced to computing the right-hand side of (3.17). We shall estimate the capacity and integral terms separately. We emphasize here that, even if we rely on the general formula (3.17), the estimation of these two terms is carried out in a novel manner. For simplicity of notation, hereafter, we write

$$\alpha_\epsilon \;=\; Z_\epsilon^{-1}\,e^{-H/\epsilon}\,(2\pi\epsilon)^{d/2}\;. \tag{3.18}$$

Our main innovation in the proof of the Eyring–Kramers formula is the new strategy to prove the following proposition.

Proposition 3.3.2. *For* ω_0 *defined in* (3.5)*, we have*

$$\mathrm{cap}_\epsilon \,=\, [\,1+o_\epsilon(1)\,]\,\alpha_\epsilon\,\omega_0\ . \tag{3.19}$$

We present our proof, up to the construction of a test function, in the next subsection. Further, we need to estimate the integral term in (3.17).

Proposition 3.3.3. *For* ν_0 *defined in* (3.7)*, we have*

$$\int_{\mathbb{R}^d} h_\epsilon^*\,d\mu_\epsilon \,=\, [\,1+o_\epsilon(1)\,]\,Z_\epsilon^{-1}\,(2\pi\epsilon)^{d/2}\,e^{-h_0/\epsilon}\,\nu_0\ . \tag{3.20}$$

We heuristically explain that the last proposition holds. Define $\mathcal{G}=\{\boldsymbol{x}: U(\boldsymbol{x})<H-\beta\}$ for small $\beta>0$ and let $\mathcal{G}_i=\mathcal{H}_i\cap\mathcal{G}$ for $i=0,\,1$. Since the process starting from a point in $\mathcal{G}_0$ may touch the set $\mathcal{D}_\epsilon(\boldsymbol{m}_0)$ before climbing to the saddle point at level H, we can expect that $h_\epsilon^*\simeq 1$ on $\mathcal{G}_0$. By a similar logic, we have $h_\epsilon^*\simeq 0$ on $\mathcal{G}_1$. Since $\mu_\epsilon(\mathcal{G}^c)$ is negligible by (2.7), we can conclude that the left-hand side of (3.20) is approximately equal to $\mu_\epsilon(\mathcal{G}_0)$, whose asymptotics is given by the right-hand side of (3.20). We turn this into a rigorous argument in Section 3.5.4 on the basis of a delicate analysis of the equilibrium potential.

Now, we formally conclude the proof of Eyring–Kramers formula.

Proof of Theorem 3.1.3. The proof is completed by combining Propositions 3.3.1, 3.3.2, and 3.3.3. □

3.3.2 Strategy to prove Proposition 3.3.2

Instead of relying on the traditional approach, which uses the variational expression of the capacity given by the Dirichlet principle or the Thomson principle to estimate the capacity, we develop an alternative strategy in this subsection. This strategy is suitable for non-reversible cases in that neither the flow structure nor the test flow is used.

In Section 3.6, we construct a smooth test function $g_\epsilon \in \mathscr{C}_{\mathcal{D}_\epsilon(\boldsymbol{m}_0),\mathcal{U}_\epsilon}$ (cf. (3.14)) satisfying the following property.

Theorem 3.3.4. *We have*

$$\epsilon \int_{\Omega_\epsilon} [\, \Phi_{g_\epsilon} \cdot \nabla h_\epsilon \,] \, d\mu_\epsilon \;=\; [\, 1 + o_\epsilon(1) \,] \, \alpha_\epsilon \, \omega_0 + o_\epsilon(1) \, [\, \alpha_\epsilon \, \mathrm{cap}_\epsilon \,]^{1/2} \;, \tag{3.21}$$

where $\Omega_\epsilon = (\, \overline{\mathcal{D}_\epsilon(\boldsymbol{m}_0)} \cup \overline{\mathcal{U}_\epsilon} \,)^c$.

The left-hand side of (3.21) corresponding to cap_ϵ by Proposition 3.2.2 is believed to be equal to the first term at the right-hand side. Thus, the second error term is somewhat unwanted and appears just because of a technical reason explained in more detail at Remark 3.3.5. We can however absorb this second error term to the first error term at the right-hand side of (3.21) as illustrated in the proof below of Proposition 3.3.2. Note that we assume Theorem 3.3.4 at this moment.

Proof of Proposition 3.3.2. By Proposition 3.2.2 and Theorem 3.3.4, we get

$$\mathrm{cap}_\epsilon \;=\; [\, 1 + o_\epsilon(1) \,] \, \alpha_\epsilon \, \omega_0 + o_\epsilon(1) \, [\, \alpha_\epsilon \, \mathrm{cap}_\epsilon \,]^{1/2} \;.$$

By dividing both sides by α_ϵ and substituting $r_\epsilon = [\, \mathrm{cap}_\epsilon / \alpha_\epsilon \,]^{1/2}$, we can rewrite the previous identity as

$$r_\epsilon^2 \;=\; [\, 1 + o_\epsilon(1) \,] \, \omega_0 + o_\epsilon(1) \, r_\epsilon \;.$$

By solving this quadratic equation in r_ϵ, we get $r_\epsilon = [\, 1 + o_\epsilon(1) \,] \, (\omega_0)^{1/2}$. Squaring this completes the proof. □

Now we turn to Theorem 3.3.4. The core of our strategy is to find a suitable test function g_ϵ and to compute the left-hand side of (3.21). Indeed, we construct g_ϵ as an approximation of the equilibrium potential $h_\epsilon^*(\cdot)$ for the adjoint process (cf. (3.16)). The reason is that, by the divergence theorem, we can write the left-hand side of (3.21) as

$$\epsilon \int_{\Omega_\epsilon} [\, \Phi_{g_\epsilon} \cdot \nabla h_\epsilon \,] \, d\mu_\epsilon \;=\; -\int_{\Omega_\epsilon} h_\epsilon \, \mathscr{L}_\epsilon^* g_\epsilon \, d\mu_\epsilon + (\text{boundary terms}) \; . \tag{3.22}$$

To control the integration on the right-hand side, we try to make $\mathscr{L}_\epsilon^* g_\epsilon$ as small as possible (cf. Proposition 3.4.5); hence, in view of the fact that $\mathcal{L}_\epsilon^* h_\epsilon^* \equiv 0$ on Ω_ϵ by the property of the equilibrium potential, the test function g_ϵ should be an approximation of h_ϵ^*. The main contribution for the computation of the left-hand side of (3.22) comes from the boundary terms, and relevant computations are carried out in Proposition 3.4.6.

The construction of g_ϵ particularly focuses on the neighborhoods of the saddle points of Σ_0 as the equilibrium potential (and hence g_ϵ, which is an approximation of the equilibrium potential) drastically falls from 1 to 0 there. We carry out this construction around the saddle point in Section 3.4 on the basis of a linearization procedure that is now routine in this field, e.g., [14, 51]. Then, we extend these functions around the saddle points of Σ_0 to a continuous function on $\mathbb{R}^d$ belonging to $\mathscr{C}_{\mathcal{D}_\epsilon(\boldsymbol{m}_0),\mathcal{U}_\epsilon}$. This process will be performed in Section 3.6, and we finally obtain g_ϵ in (3.81). Then, we prove (3.21) on the basis of our analysis of the equilibrium potential carried out in Section 3.5.

Remark 3.3.5 ((Comparison with reversible case))**.** *Our strategy is relatively simple when the underlying process is reversible. In order to get a continuous test function g_ϵ, we need a mollification procedure (cf. Proposition 3.6.2), and we must include an additional term $o_\epsilon(1)\,[\alpha_\epsilon\, \mathrm{cap}_\epsilon]^{1/2}$ in* (3.21) *to compensate for this additional procedure. However, for the reversible case, we can get a continuous test function without this mollification procedure (cf. Remark 3.6.1) and we can prove that*

$$\epsilon \int_{\Omega_\epsilon} [\, \Phi_{g_\epsilon} \cdot \nabla h_\epsilon \,] \, d\mu_\epsilon \;=\; [\, 1 + o_\epsilon(1) \,] \, \alpha_\epsilon \, \omega_0 \; ,$$

instead of (3.21)*; hence, the proof of the Eyring–Kramers formula is more straightforward. This is the only technical difference between the reversible and non-reversible models in our methodology.*

The remainder of this chapter is devoted to proving Theorem 3.3.4, and in the course of the proof, Proposition 3.3.3 will also be demonstrated in Section 3.5.

3.4 Construction of test function around saddle point

We explain how we can construct the test function around a saddle point $\boldsymbol{\sigma} \in \Sigma_0$. Section 3.4.1 presents a preliminary analysis of the geometry around the saddle point. We acknowledge that several statements and proofs given in these sections are similar to those given in [51]; however, we try not to omit the proofs of these results, as the details of the computations are slightly different owing to the differences between the models. Then, we construct the test function $p_\epsilon^{\boldsymbol{\sigma}}$ on a neighborhood of $\boldsymbol{\sigma}$ in Section 3.4.2. Finally, we explain several computational properties of this test function in Sections 3.4.3–3.4.5. These properties play crucial role in the proof of Theorem 3.3.4.

Setting

In this section, we fix a saddle point $\boldsymbol{\sigma} \in \Sigma_0$ and simply write $\mathbb{H} = \mathbb{H}^{\boldsymbol{\sigma}} = (\nabla^2 U)(\boldsymbol{\sigma})$ and $\mathbb{L} = \mathbb{L}^{\boldsymbol{\sigma}} = (D\boldsymbol{\ell})(\boldsymbol{\sigma})$. Recall that $\mathbb{H}$ has only one negative eigenvalue because of the Morse lemma. Let $-\lambda_1, \lambda_2, \cdots, \lambda_d$ denote the eigenvalues of $\mathbb{H}$, where $-\lambda_1 = -\lambda_1^{\boldsymbol{\sigma}}$ denotes the unique negative eigenvalue. Let $\boldsymbol{e}_k = \boldsymbol{e}_k^{\boldsymbol{\sigma}}$ denote the eigenvector associated with the eigenvalue λ_k ($-\lambda_k$ if $k = 1$). *In addition, we assume the direction of* $\boldsymbol{e}_1$ *to be toward* $\mathcal{H}_0$, *i.e., for all sufficiently small* $r > 0$, $\boldsymbol{\sigma} + r\boldsymbol{e}_1 \in \mathcal{H}_0$.

By Lemma 2.2.7, the matrix $\mathbb{H}+\mathbb{L}$ has a unique negative eigenvalue $-\mu = -\mu^{\boldsymbol{\sigma}}$. We can readily observe that the matrix $\mathbb{H}-\mathbb{L}^{\dagger}$ is similar to $\mathbb{H}+\mathbb{L}$. To see this, first note that, since $\mathbb{H}\mathbb{L}$ is skew-symmetric by Lemma 2.2.6, we have $\mathbb{H}\mathbb{L} = -(\mathbb{H}\mathbb{L})^{\dagger} = -\mathbb{L}^{\dagger}\mathbb{H}$. Therefore, we can check the similarity as

$$\mathbb{H}^{-1}\,(\mathbb{H}-\mathbb{L}^{\dagger})\,\mathbb{H} \;=\; \mathbb{H}^{-1}\,(\mathbb{H}^{2}+\mathbb{H}\mathbb{L}) \;=\; \mathbb{H}+\mathbb{L}\,. \tag{3.23}$$

Hence, the matrix $\mathbb{H}-\mathbb{L}^{\dagger}$ also has a unique negative eigenvalue $-\mu$, and let $\boldsymbol{v} = \boldsymbol{v}^{\boldsymbol{\sigma}}$ denote the unit eigenvector of this matrix associated with the eigenvalue $-\mu$. *Finally, we assume without loss of generality that* $\boldsymbol{v}\cdot\boldsymbol{e}_1 \ge 0$. Indeed, this cannot be 0 because of the following lemma, which implies that $(\boldsymbol{v}\cdot\boldsymbol{e}_1)^2 > 0$.

Lemma 3.4.1. *We have*

$$\boldsymbol{v}\cdot\mathbb{H}^{-1}\boldsymbol{v} \;=\; -\frac{(\boldsymbol{v}\cdot\boldsymbol{e}_1)^2}{\lambda_1} + \sum_{k=2}^{d}\frac{(\boldsymbol{v}\cdot\boldsymbol{e}_k)^2}{\lambda_k} \;=\; -\frac{1}{\mu} \;<\; 0\,.$$

Proof. The first equality is obvious if we write $\boldsymbol{v} = \sum_{i=1}^{d} a_i\boldsymbol{e}_i$. Now, we focus on the second equality. Note that $\mathbb{H}-\mathbb{L}^{\dagger}$ is invertible by Lemma 2.2.1 and (3.23). Hence, we can compute

$$\begin{aligned}\boldsymbol{v}\cdot\mathbb{H}^{-1}\boldsymbol{v} \;&=\; \boldsymbol{v}\cdot\mathbb{H}^{-1}(\mathbb{H}-\mathbb{L}^{\dagger})(\mathbb{H}-\mathbb{L}^{\dagger})^{-1}\boldsymbol{v} \;=\; -\frac{1}{\mu}\boldsymbol{v}\cdot\mathbb{H}^{-1}(\mathbb{H}-\mathbb{L}^{\dagger})\boldsymbol{v}\\ &=\; -\frac{1}{\mu}\boldsymbol{v}\cdot\boldsymbol{v} + \frac{1}{\mu}\boldsymbol{v}\cdot\mathbb{H}^{-1}\mathbb{L}^{\dagger}\mathbb{H}\mathbb{H}^{-1}\boldsymbol{v}\,.\end{aligned}$$

Since $|\boldsymbol{v}|^2 = 1$, the first term in the last line is $-\frac{1}{\mu}$. On the other hand, since $\mathbb{L}^{\dagger}\mathbb{H} = -(\mathbb{H}\mathbb{L})^{\dagger}$ is skew-symmetric and $\mathbb{H}^{-1}$ is symmetric, the second term in the last line is 0. This completes the proof. □

For two vectors $\boldsymbol{u}$, $\boldsymbol{w}\in\mathbb{R}^d$, let $\boldsymbol{u}\otimes\boldsymbol{w}\in\mathbb{R}^{d\times d}$ denote their tensor product, i.e., $(\boldsymbol{u}\otimes\boldsymbol{w})_{ij} = u_iw_j$, where u_i and w_j are the ith and jth elements of $\boldsymbol{u}$

and $\boldsymbol{w}$, respectively. The following Lemma is a consequence of the previous lemma and is similar to [54, Lemmas 4.1 and 4.2].

Lemma 3.4.2. *The following hold.*

1. *The matrix* $\mathbb{H} + 2\mu\, \boldsymbol{v} \otimes \boldsymbol{v}$ *is symmetric positive definite and* $\det(\mathbb{H} + 2\mu\, \boldsymbol{v} \otimes \boldsymbol{v}) = -\det \mathbb{H}$.

2. *The matrix* $\mathbb{H}+\mu\, \boldsymbol{v}\otimes\boldsymbol{v}$ *is symmetric non-negative definite and* $\det(\mathbb{H}+\mu\, \boldsymbol{v}\otimes\boldsymbol{v}) = 0$. *The null space of the matrix* $\mathbb{H}+\mu\, \boldsymbol{v}\otimes\boldsymbol{v}$ *is one-dimensional and spanned by the vector* $\mathbb{H}^{-1}\boldsymbol{v}$.

Proof. By a change of coordinate, we can assume that $\boldsymbol{e}_i$ is the ith standard unit vector of $\mathbb{R}^d$ such that $\mathbb{H} = \mathrm{diag}(-\lambda_1,\, \lambda_2,\, \dots,\, \lambda_d)$. First, we show that $\mathbb{H} + \mu\, \boldsymbol{v} \otimes \boldsymbol{v}$ is non-negative definite. If $v_2 = \cdots = v_d = 0$, then, we have $v_1^2 = \mu/\lambda_1$ by Lemma 3.4.1; thus, $\mathbb{H} + \mu\, \boldsymbol{v} \otimes \boldsymbol{v} = \mathrm{diag}(0, \lambda_2, \dots, \lambda_d)$ is non-negative definite. Otherwise, for $\boldsymbol{x} = \sum_{i=1}^{d} x_i \boldsymbol{e}_i \in \mathbb{R}^d$, we can compute

$$\boldsymbol{x} \cdot [\,\mathbb{H} + \mu\, \boldsymbol{v} \otimes \boldsymbol{v}\,]\, \boldsymbol{x} \;=\; -\lambda_1 x_1^2 + \sum_{k=2}^{d} \lambda_k x_k^2 + \mu \Big(\sum_{i=1}^{d} x_i v_i \Big)^2 .$$

By minimizing the right-hand side over x_1 and using Lemma 3.4.1, we get

$$\sum_{k=2}^{d} \lambda_k x_k^2 - \frac{(\,\sum_{k=2}^{d} x_k v_k\,)^2}{\sum_{k=2}^{d} v_k^2/\lambda_k}\ ,$$

which is non-negative by Cauchy–Schwarz inequality. This proves that $\mathbb{H} + \mu\, \boldsymbol{v}\otimes\boldsymbol{v}$ is non-negative definite. Then, the matrix $\mathbb{H}+2\mu\, \boldsymbol{v}\otimes\boldsymbol{v}$ is non-negative definite as well. By the well-known formula

$$\det(\mathbb{A} + \boldsymbol{x} \otimes \boldsymbol{y}) \;=\; (1 + \boldsymbol{y}^\dagger \mathbb{A}^{-1} \boldsymbol{x}) \det \mathbb{A}\ , \tag{3.24}$$

along with Lemma 3.4.1, we can check that $\det(\mathbb{H}+2\mu\, \boldsymbol{v}\otimes\boldsymbol{v}) = -\det\mathbb{H} > 0$, and thus, $\mathbb{H} + 2\mu\, \boldsymbol{v} \otimes \boldsymbol{v}$ is indeed positive definite. Finally, we investigate the

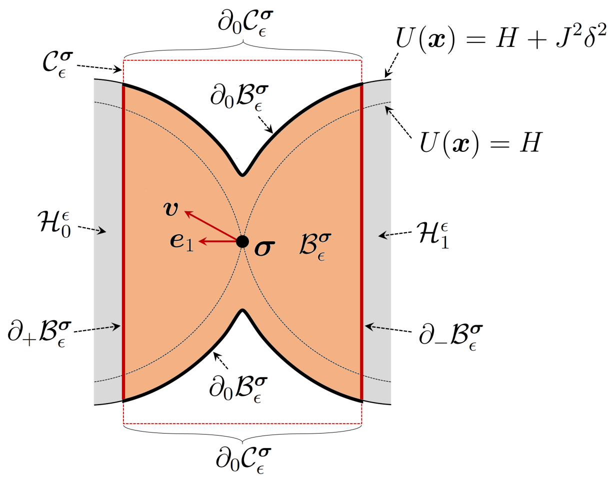


Figure 3.2: Illustration of the neighborhood structure around a saddle point $\boldsymbol{\sigma}$.

null space of $\mathbb{H} + \mu\boldsymbol{v} \otimes \boldsymbol{v}$. Suppose that $\boldsymbol{w} \in \mathbb{R}^d$ satisfies $(\mathbb{H} + \mu\boldsymbol{v} \otimes \boldsymbol{v})\boldsymbol{w} = 0$. Since $\mathbb{H}$ is invertible, we can rewrite this equation as $\boldsymbol{w} = -\mu(\boldsymbol{v} \cdot \boldsymbol{w})\mathbb{H}^{-1}\boldsymbol{v}$. Hence, the null space is a subspace of $\langle\, \mathbb{H}^{-1}\boldsymbol{v}\,\rangle$. On the other hand, if $\boldsymbol{w} = a\mathbb{H}^{-1}\boldsymbol{v}$ for some $a \in \mathbb{R}$, we can readily check that $(\mathbb{H} + \mu\boldsymbol{v} \otimes \boldsymbol{v})\boldsymbol{w} = \boldsymbol{0}$, and hence, $\langle\, \mathbb{H}^{-1}\boldsymbol{v}\,\rangle$ is indeed the null space. □

3.4.1 Neighborhood of saddle points

In this subsection, we specify the geometry around each saddle point $\boldsymbol{\sigma}$. Figure 3.2 illustrates the sets appearing in this section.

We focus on a neighborhood of $\boldsymbol{\sigma}$ with size of order δ, which is defined by

$$\delta \;=\; \delta(\epsilon) \;:=\; \Big(\epsilon \log \frac{1}{\epsilon}\Big)^{1/2}\,. \tag{3.25}$$

Let J be a sufficiently large constant that is independent of ϵ. There will be several class, e.g., Lemma 3.6.4, that require J to be sufficiently large; we suppose that J satisfies all such requirements. Define a box $\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ centered at

$\boldsymbol{\sigma}$ as

$$\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} = \Big\{ \boldsymbol{\sigma} + \sum_{i=1}^{d} \alpha_i \boldsymbol{e}_i^{\boldsymbol{\sigma}} \in \mathbb{R}^d : -\frac{J\delta}{\lambda_1^{1/2}} \le \alpha_1 \le \frac{J\delta}{\lambda_1^{1/2}} \text{ and } -\frac{2J\delta}{\lambda_j^{1/2}} \le \alpha_j \le \frac{2J\delta}{\lambda_j^{1/2}} \text{ for } 2 \le j \le d \Big\} .$$

Now, decompose the boundary $\partial\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ into $\partial_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$, $\partial_-\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$, and $\partial_0\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ such that

$$\begin{aligned} \partial_\pm\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} &= \Big\{ \boldsymbol{\sigma} + \sum_{i=1}^{d} \alpha_i \boldsymbol{e}_i^{\boldsymbol{\sigma}} \in \mathbb{R}^d : \alpha_1 = \pm\frac{J\delta}{\lambda_1^{1/2}} \Big\} , \qquad (3.26) \\ \partial_0\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} &= \partial\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \setminus (\partial_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \cup \partial_-\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}) . \end{aligned}$$

Lemma 3.4.3. *For $\boldsymbol{x} \in \partial_0\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$, we have $U(\boldsymbol{x}) \ge H + \frac{5}{4}J^2\delta^2$ for all sufficiently small $\epsilon > 0$.*

Proof. For $\boldsymbol{x} \in \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$, by the Taylor expansion of U at $\boldsymbol{\sigma}$,

$$U(\boldsymbol{x}) = H + \frac{1}{2}\Big[-\lambda_1 x_1^2 + \sum_{j=2}^{d} \lambda_j x_j^2 \Big] + O(\delta^3) . \qquad (3.27)$$

For $\boldsymbol{x} \in \partial_0\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$, $x_i = \pm 2J\delta/\sqrt{\lambda_i}$ for some $2 \le i \le d$. Therefore,

$$-\lambda_1 x_1^2 + \sum_{j=2}^{d} \lambda_j x_j^2 \ge -J^2\delta^2 + \lambda_i \Big(\frac{2J\delta}{\lambda_i^{1/2}} \Big)^2 = 3J^2\delta^2 .$$

Inserting this to (3.27) completes the proof. □

Hereafter, we assume that $\epsilon > 0$ is sufficiently small such that Lemma 3.4.3 holds. Define, for $\epsilon > 0$,

$$\begin{aligned} \mathcal{K}_\epsilon &= \{ \boldsymbol{x} \in \mathbb{R}^d : U(\boldsymbol{x}) < H + J^2\delta^2 \} \text{ and} \\ \mathcal{K} &= \{ \boldsymbol{x} \in \mathbb{R}^d : U(\boldsymbol{x}) < H + J^2 \} \qquad (3.28) \end{aligned}$$

so that $\mathcal{H} \subset \mathcal{K}_\epsilon \subset \mathcal{K}$ holds.

By Lemma 3.4.3, the boundary $\partial_0 \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ does not belong to $\mathcal{K}_\epsilon$. The neighborhood of $\boldsymbol{\sigma}$ in which we focus on the construction is the set $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} = \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \cap \mathcal{K}_\epsilon$. Now, we decompose the boundary $\partial \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ into $\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, $\partial_- \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, and $\partial_0 \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ such that

$$\partial_\pm \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \;=\; \partial_\pm \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \cap \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \quad \text{and} \quad \partial_0 \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \;=\; \partial \mathcal{B}_\epsilon \setminus (\, \partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \cup \partial_- \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \,)$$

so that we have $U(\boldsymbol{x}) = H + J^2\delta^2$ for all $\boldsymbol{x} \in \partial_0 \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ by Lemma 3.4.3.

Now, the set $\mathcal{K}_\epsilon \setminus \cup_{\boldsymbol{\sigma} \in \Sigma_0} \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ consists of several connected components. Let $\mathcal{H}_0^\epsilon$ denote one such component containing $\mathcal{M}_0$ and let $\mathcal{H}_1^\epsilon$ denote the union of the other components such that $\mathcal{M}_1 \subset \mathcal{H}_1^\epsilon$. By our convention on the direction of the vector $\boldsymbol{e}_1 = \boldsymbol{e}_1^{\boldsymbol{\sigma}}$ mentioned earlier in the current section, we have

$$\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \,\subset\, \partial \mathcal{H}_0^\epsilon \quad \text{and} \quad \partial_- \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \,\subset\, \partial \mathcal{H}_1^\epsilon \,. \tag{3.29}$$

This is illustrated in Figure 3.2.

3.4.2 Construction of test function around $\boldsymbol{\sigma}$ via linearization procedure

We construct a function $p_\epsilon^{\boldsymbol{\sigma}} : \mathbb{R}^d \to \mathbb{R}$ on $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, which acts as a building block for the global construction carried out in the following sections. As mentioned in Section 3.3.2, we would like to build a function approximating the equilibrium potential h_ϵ^* between $\mathcal{D}_\epsilon(\boldsymbol{m}_0)$ and $\mathcal{U}_\epsilon$. Thus, we expect $p_\epsilon^{\boldsymbol{\sigma}}$ to satisfy $\mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}} \simeq 0$, where $\mathscr{L}_\epsilon^*$ is defined in (3.11). To find this function, we linearize the generator $\mathscr{L}_\epsilon^*$ around $\boldsymbol{\sigma}$ by the first-order Taylor expansion such that, for smooth f,

$$\widetilde{\mathscr{L}_\epsilon^*} f \;=\; \epsilon\, \Delta f(\boldsymbol{x}) - \nabla f(\boldsymbol{x}) \cdot (\mathbb{H} - \mathbb{L})(\boldsymbol{x}) \;,$$

and we solve the linearized equation $\widetilde{\mathscr{L}_\epsilon^*} p_\epsilon^{\boldsymbol{\sigma}} \,=\, 0$. This equation can be explicitly solved using the separation of variables method. Note that in view of

(3.29), we would like to impose boundary conditions of the form $p_\epsilon^{\boldsymbol{\sigma}} \simeq 1$ on $\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ and $p_\epsilon^{\boldsymbol{\sigma}} \simeq 0$ on $\partial_-\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$. A test function satisfying all these requirements is given by

$$p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) \;=\; \frac{1}{c_\epsilon^{\boldsymbol{\sigma}}}\int_{-\infty}^{(\boldsymbol{x}-\boldsymbol{\sigma})\cdot\boldsymbol{v}} e^{-\frac{\mu}{2\epsilon}t^2}\,dt \quad ;\; \boldsymbol{x}\in\overline{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}}\;, \tag{3.30}$$

where

$$c_\epsilon^{\boldsymbol{\sigma}} \;=\; \int_{-\infty}^{\infty} e^{-\frac{\mu}{2\epsilon}t^2}\,dt \;=\; \sqrt{\frac{2\pi\epsilon}{\mu}}\;. \tag{3.31}$$

Note that $\boldsymbol{v}$ and μ are defined at the beginning of the current section. The crucial technical difficulty arises from the fact that the function $p_\epsilon^{\boldsymbol{\sigma}}$ is not constant along the boundary $\partial_\pm\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ unless the dynamics is reversible since $\boldsymbol{e}_1^{\boldsymbol{\sigma}}$ and $\boldsymbol{v}$ are linearly independent if $\boldsymbol{\ell}\neq 0$. This makes it difficult to patch these functions together. This issue will be thoroughly investigated in Section 3.6.

Since $p_\epsilon^{\boldsymbol{\sigma}}$ is smooth on $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, we can define $\Phi_{p_\epsilon^{\boldsymbol{\sigma}}}$ on $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$. Next, we must investigate the properties of $p_\epsilon^{\boldsymbol{\sigma}}$ and $\Phi_{p_\epsilon^{\boldsymbol{\sigma}}}$. *For the simplicity of notation, we assume that $\boldsymbol{\sigma}=\mathbf{0}$ in the remainder of the current section.*

3.4.3 Negligibility of $\mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}}$ on $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$

Our construction of $p_\epsilon^{\boldsymbol{\sigma}}$ suggests that $\mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}}$ is small on $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$. The next lemma precisely quantifies this heuristic observation.

Notation 3.4.4. *Let $C>0$ denote a positive constant independent of ϵ and $\boldsymbol{x}$. Different appearances of C may express different values.*

Proposition 3.4.5. *We have* $\displaystyle\int_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} |\mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}}|\,d\mu_\epsilon \;=\; o_\epsilon(1)\,\alpha_\epsilon.$

Proof. By inserting the explicit formula (3.30), we get

$$(\mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}})(\boldsymbol{x}) \;=\; (c_\epsilon^{\boldsymbol{\sigma}})^{-1}\,e^{-\frac{\mu}{2\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v})^2}\Big[\,-(\nabla U-\boldsymbol{\ell})(\boldsymbol{x})\cdot\boldsymbol{v}-\mu(\boldsymbol{x}\cdot\boldsymbol{v})\Big]\;.$$

Now, by applying the Taylor expansion of ∇U and ℓ around $\boldsymbol{\sigma}$, for $\boldsymbol{x} \in \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$,

$$\begin{aligned}(\mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}})(\boldsymbol{x}) &= -(c_\epsilon^{\boldsymbol{\sigma}})^{-1}\, e^{-\frac{\mu}{2\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v})^2} \Big[\{ (\mathbb{H}-\mathbb{L})\boldsymbol{x} + O(\delta^2) \} \cdot \boldsymbol{v} + \mu(\boldsymbol{x}\cdot\boldsymbol{v}) \Big] \\ &= -(c_\epsilon^{\boldsymbol{\sigma}})^{-1}\, e^{-\frac{\mu}{2\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v})^2} \Big[\boldsymbol{x}\cdot(-\mu\boldsymbol{v}) + \mu(\boldsymbol{x}\cdot\boldsymbol{v}) + O(\delta^2) \Big] \,,\end{aligned}$$

where the last line follows from the fact that $\boldsymbol{v}$ is an eigenvector of $(\mathbb{H}-\mathbb{L})^\dagger = \mathbb{H}-\mathbb{L}^\dagger$ associated with the eigenvalue $-\mu$. Now, recall $c_\epsilon^{\boldsymbol{\sigma}}$ from (3.31) to deduce that, for some constant $C>0$,

$$|\, (\mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}})(\boldsymbol{x}) \,| \le \frac{C\,\delta^2}{\epsilon^{1/2}}\, e^{-\frac{\mu}{2\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v})^2} \,.$$

By the second-order Taylor expansion, we can write

$$U(\boldsymbol{x}) = H + \frac{1}{2}\boldsymbol{x}\cdot\mathbb{H}\boldsymbol{x} + O(\delta^3) \quad \text{for } \boldsymbol{x}\in\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \,.$$

This expansion will be repeatedly used in the subsequent computation. Since $e^{-O(\delta^3)/\epsilon} = 1 + o_\epsilon(1)$ by the definition (3.25) of δ, we can conclude that

$$\int_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} |\, \mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}} \,|\, d\mu_\epsilon \le C\frac{\delta^2}{Z_\epsilon \epsilon^{1/2}} e^{-H/\epsilon} \int_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} e^{-\frac{1}{2\epsilon}\boldsymbol{x}\cdot(\mathbb{H}+\mu\boldsymbol{v}\otimes\boldsymbol{v})\boldsymbol{x}}\, d\boldsymbol{x} \,. \tag{3.32}$$

Now, the estimation of the last integral remains. This part is similar to [51, Lemma 8.7]; however, we repeat the argument here for the completeness of the proof. By part (2) of Lemma 3.4.2, let $\rho_1 = 0$ and $\rho_2, \dots, \rho_d > 0$ denote the eigenvalues of $\mathbb{H} + \mu\boldsymbol{v}\otimes\boldsymbol{v}$ and let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_d$ denote the corresponding unit eigenvectors. Let $\langle \boldsymbol{u}_2, \cdots, \boldsymbol{u}_d \rangle$ denote the subspace of $\mathbb{R}^d$ spanned by vectors $\boldsymbol{u}_2, \cdots, \boldsymbol{u}_d$. Since $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \subset \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$, there exists $M>0$ such that

$$\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \subset \bigcup_{a:|a|\le M\delta} (\, a\boldsymbol{u}_1 + \langle\, \boldsymbol{u}_2, \dots, \boldsymbol{u}_d \,\rangle\,) \,.$$

Hence, along with the change of variables $\boldsymbol{x} = \sum y_i \boldsymbol{u}_i$, we can bound the last integral in (3.32) by

$$\int_{-M\delta}^{M\delta} \Big[\int_{\mathbb{R}^{d-1}} \exp\Big\{ -\frac{1}{2\epsilon} \sum_{k=2}^{d} \rho_k y_k^2 \Big\} dy_2 \cdots dy_d \Big] dy_1 \;=\; C\,\delta\,\epsilon^{(d-1)/2}\;.$$

By inserting this into (3.32), we get $\int_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} |\,\mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}}\,|\,d\mu_\epsilon \;\le\; C\,\delta^3\,\epsilon^{-1}\,\alpha_\epsilon$. Since $\delta^3\,\epsilon^{-1} = o_\epsilon(1)$, the proof is completed. □

3.4.4 Property of $\Phi_{p_\epsilon^{\boldsymbol{\sigma}}}$ at the boundary of $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$

Next, we prove the following property of the vector field $\Phi_{p_\epsilon^{\boldsymbol{\sigma}}}$. Recall $\omega^{\boldsymbol{\sigma}}$ from (3.3).

Proposition 3.4.6. *We have*

$$\epsilon \int_{\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \Big[\Big(\Phi_{p_\epsilon^{\boldsymbol{\sigma}}} - \frac{1}{\epsilon}\boldsymbol{\ell} \Big) \cdot \boldsymbol{e}_1 \Big]\, \sigma(d\mu_\epsilon) \;=\; [\,1 + o_\epsilon(1)\,]\,\alpha_\epsilon\,\omega^{\boldsymbol{\sigma}}\;. \tag{3.33}$$

This estimate is indeed the key estimate in the proof of Theorem 3.3.4. The left-hand side of (3.33) corresponds to the boundary term in (3.22). The proof of this proposition is slightly complicated. Hence, we first establish some technical lemmas. For simplicity of notation, we assume in this subsection that $\boldsymbol{e}_i$ is the ith standard normal vector of $\mathbb{R}^d$; hence, we can write

$$\mathbb{H} \;=\; \mathrm{diag}(\,-\lambda_1,\,\lambda_2,\,\dots,\,\lambda_d\,) \quad \text{and} \quad \boldsymbol{v} \;=\; (\,v_1,\,\dots,\,v_d\,)\;.$$

Change of coordinate on $\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$

First, we introduce a change of coordinate that maps $\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ to a subset of $\mathbb{R}^{d-1}$ to simplify the integration in (3.33)

For $\mathbb{A} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{u} = (\,u_1,\,\dots,\,u_d\,) \in \mathbb{R}^d$, define $\widetilde{\mathbb{A}} \in \mathbb{R}^{(d-1)\times(d-1)}$ and

$\widetilde{\boldsymbol{u}} \in \mathbb{R}^{d-1}$ as

$$\widetilde{\mathbb{A}} = (\mathbb{A}_{i,j})_{2\le i,j\le d} \quad \text{and} \quad \widetilde{\boldsymbol{u}} = (u_2, \dots, u_d), \tag{3.34}$$

respectively. It is important to select a point of $\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ corresponding to the origin of $\mathbb{R}^{d-1}$ to simplify our computation. To this end, define $\boldsymbol{\gamma} = (\gamma_2, \dots, \gamma_d) \in \mathbb{R}^{d-1}$ as

$$\gamma_k = \frac{\lambda_1^{1/2}}{v_1} \cdot \frac{v_k}{\lambda_k} J\delta \quad ; \; k = 2, \dots, d. \tag{3.35}$$

Note that $v_1 \neq 0$ by Lemma 3.4.1. Define a map $\Pi_\epsilon : \partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \to \mathbb{R}^{d-1}$ that represents the change of coordinate as

$$\Pi_\epsilon(\boldsymbol{x}) = \widetilde{\boldsymbol{x}} + \boldsymbol{\gamma}. \tag{3.36}$$

Our careful selection of $\boldsymbol{\gamma}$ ensures that this map simplifies the computation of the crucial quadratic form.

Lemma 3.4.7. *For all $\boldsymbol{x} \in \partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, we have*

$$\boldsymbol{x} \cdot (\mathbb{H} + \mu\, \boldsymbol{v} \otimes \boldsymbol{v})\boldsymbol{x} = \Pi_\epsilon(\boldsymbol{x}) \cdot (\widetilde{\mathbb{H}} + \mu\, \widetilde{\boldsymbol{v}} \otimes \widetilde{\boldsymbol{v}})\, \Pi_\epsilon(\boldsymbol{x}).$$

Proof. Fix $\boldsymbol{x} = \big(\frac{J\delta}{\lambda_1^{1/2}}, x_2, \dots, x_d\big) \in \partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ and write $\Pi_\epsilon(\boldsymbol{x}) = \boldsymbol{y} = (y_2, \dots, y_d)$. Then, by Lemma 3.4.1, we can write

$$\boldsymbol{x} \cdot \boldsymbol{v} = \frac{J\,\delta}{\lambda_1^{1/2}} v_1 + \sum_{k=2}^{d} (y_k - \gamma_k) v_k = \boldsymbol{y} \cdot \widetilde{\boldsymbol{v}} + \frac{J\,\delta\,\lambda_1^{1/2}}{\mu\, v_1}.$$

Thus, we can write $\boldsymbol{x} \cdot (\mathbb{H} + \mu \boldsymbol{v} \otimes \boldsymbol{v})\boldsymbol{x}$ as

$$-\lambda_1 x_1^2 + \sum_{k=2}^{d} \lambda_k x_k^2 + \mu \Big(\boldsymbol{y} \cdot \widetilde{\boldsymbol{v}} + \frac{J\delta\lambda_1^{1/2}}{\mu v_1} \Big)^2 = \boldsymbol{y} \cdot (\widetilde{\mathbb{H}} + \mu \widetilde{\boldsymbol{v}} \otimes \widetilde{\boldsymbol{v}})\boldsymbol{y}.$$

The correction vector $\boldsymbol{\gamma}$ is designed to clear the linear terms and constant term here. $\square$

We can now show that the image of $\Pi_\epsilon(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})$ is comparable with a ball centered at the origin with a radius of order δ.

Lemma 3.4.8. *There exist constants r, $R > 0$ such that*

$$\mathcal{D}_{r\delta}^{(d-1)}(\mathbf{0}) \,\subset\, \Pi_\epsilon(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}) \,\subset\, \mathcal{D}_{R\delta}^{(d-1)}(\mathbf{0}) \;, \tag{3.37}$$

where $\mathcal{D}_a^{(d-1)}(\mathbf{0})$ denotes a sphere on $\mathbb{R}^{d-1}$ centered at the origin with radius a.

Proof. Since $\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \subset \mathcal{D}_{C\delta}^{(d-1)}(\mathbf{0})$ for sufficiently large $C > 0$ and $|\boldsymbol{\gamma}| = O(\delta)$, the existence of R is immediate from the definition of Π_ϵ.

Now we focus on the first inclusion of (3.37). For $\boldsymbol{\gamma} \in \mathbb{R}^{d-1}$ defined in (3.35), we write

$$\begin{aligned} \mathcal{P}_\delta \;&=\; \Big\{\, \boldsymbol{x} \in \mathbb{R}^d : x_1 = \frac{J\delta}{\lambda_1^{1/2}} \,\Big\} \subset \mathbb{R}^d \;\text{ and} \\ \overline{\boldsymbol{\gamma}} \;&=\; \Big(\, \frac{J\delta}{\lambda_1^{1/2}},\, -\gamma_2,\, \dots,\, -\gamma_d \Big) \in \mathcal{P}_\delta \;. \end{aligned}$$

Then, by the Taylor expansion and Lemma 3.4.1, we can check that

$$U(\overline{\boldsymbol{\gamma}}) \;=\; H - \frac{\lambda_1}{2\,\mu\,v_1^2}\, J^2\,\delta^2 + O(\delta^3) \;<\; H - c_0 J^2 \delta^2 \tag{3.38}$$

for all sufficiently small $\epsilon > 0$, provided that we take c_0 to be sufficiently small. Therefore, there exists $r > 0$ such that $\mathcal{D}_{r\delta}(\overline{\boldsymbol{\gamma}}) \cap \mathcal{P}_\delta \subset \partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$. Since $\Pi_\epsilon(\overline{\boldsymbol{\gamma}}) = \mathbf{0}$, we have $\mathcal{D}_{r\delta}^{(d-1)}(\mathbf{0}) = \Pi_\epsilon(\mathcal{D}_{r\delta}^{(d-1)}(\overline{\boldsymbol{\gamma}}) \cap \mathcal{P}_\delta)$. This completes the proof. $\square$

Now, we present three auxiliary lemmas (Lemmas 3.4.9, 3.4.10, and 3.4.11) that will be used in several instances including the proof of Proposition 3.4.6. The proofs of these technical results are deferred to the next subsection.

Lemma 3.4.9. *The matrix* $\widetilde{\mathbb{H}} + \mu\, \widetilde{\boldsymbol{v}} \otimes \widetilde{\boldsymbol{v}}$ *is positive definite and*

$$\det\big(\, \widetilde{\mathbb{H}} + \mu \widetilde{\boldsymbol{v}} \otimes \widetilde{\boldsymbol{v}}\,\big) \;=\; \mu\, \frac{v_1^2}{\lambda_1} \prod_{k=2}^{d} \lambda_k \;.$$

Proof. By (3.24) and Lemma 3.4.1,

$$\det\big(\, \widetilde{\mathbb{H}} + \mu \widetilde{\boldsymbol{v}} \otimes \widetilde{\boldsymbol{v}}\,\big) \;=\; (1 + \mu \widetilde{\boldsymbol{v}}^{\dagger}\, \widetilde{\mathbb{H}}^{-1}\, \widetilde{\boldsymbol{v}}) \det \widetilde{\mathbb{H}} \;=\; \frac{\mu v_1^2}{\lambda_1} \det \widetilde{\mathbb{H}} \;.$$

□

Recall $\partial_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ from (3.26) and define, for $a > 0$,

$$\partial_+^{1,\,a} \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \;=\; \{\boldsymbol{x} \in \partial_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} : \boldsymbol{x} \cdot \boldsymbol{v} \ge aJ\delta\} \;, \tag{3.39}$$

$$\partial_+^{2,\,a} \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \;=\; \big\{\boldsymbol{x} \in \partial_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} : U(\boldsymbol{x}) \ge H + aJ^2\delta^2\big\} \;. \tag{3.40}$$

Lemma 3.4.10. *There exists* $a_0 > 0$ *such that, for all* $a \in (0,\, a_0)$,

$$\partial_+^{1,\,a} \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \cup \partial_+^{2,\,a} \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \;=\; \partial_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \;.$$

Hereafter, the constant a_0 always refers to the one in the previous lemma. For $a > 0$, we write

$$\partial_+^{1,\,a} \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \;=\; \partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \cap \partial_+^{1,\,a} \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \;=\; \{\boldsymbol{x} \in \partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} : \boldsymbol{x} \cdot \boldsymbol{v} \ge aJ\delta\} \;, \tag{3.41}$$

$$\partial_+^{2,\,a} \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \;=\; \partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \cap \partial_+^{2,\,a} \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \;=\; \big\{\boldsymbol{x} \in \partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} : U(\boldsymbol{x}) \ge H + aJ^2\delta^2\big\} \,; \tag{3.42}$$

hence, we have

$$\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \;=\; \partial_+^{1,\,a} \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \cup \partial_+^{2,\,a} \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \tag{3.43}$$

for all $a \in (0,\, a_0)$ by the previous lemma. Now, we introduce the last lemma.

Lemma 3.4.11. *Let* $\mathbb{D}$ *be a positive-definite* $(d-1) \times (d-1)$ *matrix, Then,* *for all* $\boldsymbol{u}_1,\, \boldsymbol{u}_2 \in \mathbb{R}^{d-1}$ *and* $c \in (0,\, 1)$*, we have*

$$\int_{\Pi_\epsilon(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})\cap\{\boldsymbol{y}\in\mathbb{R}^{d-1}:\boldsymbol{y}\cdot\boldsymbol{u}_1\geq -c\delta\}} \frac{\boldsymbol{y}\cdot\boldsymbol{u}_2+\delta}{\boldsymbol{y}\cdot\boldsymbol{u}_1+\delta}\, e^{-1/(2\epsilon)\,\boldsymbol{y}\cdot\mathbb{D}\boldsymbol{y}}\, d\boldsymbol{y}$$
$$= \left[\, 1+o_\epsilon(1)\,\right] \frac{(2\pi\epsilon)^{(d-1)/2}}{\sqrt{\det(\mathbb{D})}}\ .$$

Now, we are ready to prove Proposition 3.4.6.

Proof of Proposition 3.4.6. In view of the definition of $\Phi_{p_\epsilon^{\boldsymbol{\sigma}}}$ given in (3.13), we can write

$$\epsilon \int_{\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \Big[\, \Phi_{p_\epsilon^{\boldsymbol{\sigma}}} - \frac{1}{\epsilon}\boldsymbol{\ell} \,\Big] \cdot \boldsymbol{e}_1\, \sigma(d\mu_\epsilon) \;=\; I_1 - I_2\ , \tag{3.44}$$

where

$$I_1 \;=\; \epsilon \int_{\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \nabla p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x})\cdot\boldsymbol{e}_1\,\sigma(d\mu_\epsilon) \quad\text{and}\quad I_2 \;=\; \int_{\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} (1-p_\epsilon^{\boldsymbol{\sigma}})\,(\boldsymbol{\ell}\cdot\boldsymbol{e}_1)\,\sigma(d\mu_\epsilon)\ .$$

First, we compute I_1. By the explicit form of $p_\epsilon^{\boldsymbol{\sigma}}$ and the Taylor expansion of U, we can write

$$I_1 \;=\; [\, 1+o_\epsilon(1)\,]\, v_1\, \frac{\epsilon}{Z_\epsilon}\sqrt{\frac{\mu}{2\pi\epsilon}}e^{-\frac{H}{\epsilon}} \int_{\partial_+\mathcal{B}_\epsilon} e^{-\frac{1}{2\epsilon}\boldsymbol{x}\cdot(\mathbb{H}+\mu\boldsymbol{v}\otimes\boldsymbol{v})\boldsymbol{x}}\sigma(d\boldsymbol{x})\ . \tag{3.45}$$

By the change of variables $\boldsymbol{y} = \Pi_\epsilon(\boldsymbol{x})$, the last integral can be expressed as

$$\int_{\Pi_\epsilon(\partial_+\mathcal{B}_\epsilon)} e^{-\frac{1}{2\epsilon}\boldsymbol{y}\cdot(\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}})\boldsymbol{y}}\, d\boldsymbol{y} \;=\; [\, 1+o_\epsilon(1)\,]\, \frac{(2\pi\epsilon)^{(d-1)/2}}{\sqrt{\det(\,\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}}\,)}}\ ,$$

where the equality follows from the change of variables $\boldsymbol{z} = \epsilon^{-1/2}\boldsymbol{y}$ and

Lemma 3.4.8. Summing up, we get

$$I_1 \;=\; [\,1+o_\epsilon(1)\,]\,\frac{v_1\,\mu^{1/2}\,\alpha_\epsilon}{2\pi\sqrt{\det\,(\,\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}}\,)}}\;. \tag{3.46}$$

Next, we consider I_2. Let us take $a\in(0,\,a_0)$, where a_0 is the constant in Lemma 3.4.10, and decompose

$$I_2 \;=\; I_{2,\,1}+I_{2,\,2}\;, \tag{3.47}$$

where

$$\begin{aligned} I_{2,\,1} \;&=\; \int_{\partial_+^{1,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} (1-p_\epsilon^{\boldsymbol{\sigma}})\,(\boldsymbol{\ell}\cdot\boldsymbol{e}_1)\,\sigma(d\mu_\epsilon)\;,\\ I_{2,\,2} \;&=\; \int_{\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}\setminus\partial_+^{1,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} (1-p_\epsilon^{\boldsymbol{\sigma}})\,(\boldsymbol{\ell}\cdot\boldsymbol{e}_1)\,\sigma(d\mu_\epsilon)\;. \end{aligned}$$

First, we compute $I_{2,\,1}$. Recall the elementary inequality

$$\frac{b}{b^2+1}e^{-b^2/2} \;\le\; \int_b^\infty e^{-t^2/2}\,dt \;\le\; \frac{1}{b}e^{-b^2/2} \quad \text{for } b>0\;. \tag{3.48}$$

Now, for $\boldsymbol{x}\in\partial_+^{1,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, since we have $\sqrt{\dfrac{\mu}{\epsilon}}(\boldsymbol{x}\cdot\boldsymbol{v})\to\infty$ as $\epsilon\to 0$, we obtain from the definition of $p_\epsilon^{\boldsymbol{\sigma}}$ and (3.48) that

$$1-p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) \;=\; [\,1+o_\epsilon(1)\,]\,\frac{\epsilon^{1/2}}{(2\pi\mu)^{1/2}\,(\boldsymbol{x}\cdot\boldsymbol{v})}\;\exp\Big\{-\frac{\mu}{2\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v})^2\Big\}\,. \tag{3.49}$$

By the Taylor expansion of $\boldsymbol{\ell}$, we have

$$\boldsymbol{\ell}(\boldsymbol{x})\cdot\boldsymbol{e}_1 \;=\; \mathbb{L}\boldsymbol{x}\cdot\boldsymbol{e}_1+O(\delta^2)\;. \tag{3.50}$$

Our plan is to insert (3.49) and (3.50) into $I_{2,\,1}$ to complete the proof. To this end, we first explain that we can ignore the $O(\delta^2)$ term in (3.50). By (3.49), the Taylor expansion of U, and Lemma 3.4.2, we have

$$\begin{aligned}
&\Big|\,\delta^2 \int_{\partial_+^{1,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} (1-p_\epsilon^{\boldsymbol{\sigma}})\,\sigma(d\mu_\epsilon)\,\Big| \\
&\le C\,\frac{\delta\,\epsilon^{1/2}}{Z_\epsilon}\,e^{-H/\epsilon}\int_{\partial_+^{1,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \exp\Big\{-\frac{\mu}{2\epsilon}\boldsymbol{x}\cdot(\mathbb{H}+\mu\boldsymbol{v}\otimes\boldsymbol{v})\boldsymbol{x}\Big\}\,\sigma(d\boldsymbol{x}) \\
&\le C\,\frac{\delta\,\epsilon^{1/2}}{Z_\epsilon}\,e^{-H/\epsilon}\,\sigma(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}) \;=\; C\frac{\delta^d\,\epsilon^{1/2}}{Z_\epsilon}\,e^{-H/\epsilon} \;=\; o_\epsilon(1)\,\alpha_\epsilon\,. \qquad (3.51)
\end{aligned}$$

Hence, by combining (3.49), (3.50), and (3.51), we can write

$$I_{2,1} \;=\; o_\epsilon(1)\,\alpha_\epsilon + \frac{[\,1+o_\epsilon(1)\,]\,\alpha_\epsilon\,\epsilon}{(2\pi\epsilon)^{(d+1)/2}\,\mu^{1/2}}\int_{\partial_+^{1,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} e^{-\frac{1}{2\epsilon}\boldsymbol{x}\cdot[\,\mathbb{H}+\mu\boldsymbol{v}\otimes\boldsymbol{v}\,]\boldsymbol{x}}\,\frac{\mathbb{L}\boldsymbol{x}\cdot\boldsymbol{e}_1}{\boldsymbol{x}\cdot\boldsymbol{v}}\,\sigma(d\boldsymbol{x})\,. \qquad (3.52)$$

By the change of variables $\boldsymbol{y} = \Pi_\epsilon(\boldsymbol{x})$ and Lemma 3.4.7, we can write the last integral as

$$\begin{aligned}
&\int_{\Pi(\partial_+\mathcal{B}_\epsilon)\cap\{\boldsymbol{y}:\boldsymbol{y}\cdot\widetilde{\boldsymbol{v}}\ge c'J\delta\}} e^{-\frac{1}{2\epsilon}\boldsymbol{y}\cdot[\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}}]\boldsymbol{y}}\,\frac{\boldsymbol{y}\cdot\widetilde{\mathbb{L}}^\dagger\widetilde{\boldsymbol{v}} - \frac{J\delta\lambda_1}{v_1}\sum_{k=2}^d \mathbb{L}_{1k}\frac{v_k}{\lambda_k}}{\boldsymbol{y}\cdot\widetilde{\boldsymbol{v}}+\frac{J\delta\lambda_1}{\mu v_1}}d\boldsymbol{y} \\
&= (-\mu\mathbb{L}\mathbb{H}^{-1}\boldsymbol{v})\int_{\Pi(\partial_+\mathcal{B}_\epsilon)\cap\{\boldsymbol{y}:\boldsymbol{y}\cdot\widetilde{\boldsymbol{v}}\ge c'J\delta\}} e^{-\frac{1}{2\epsilon}\boldsymbol{y}\cdot[\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}}]\boldsymbol{y}}\,\frac{\boldsymbol{y}\cdot\boldsymbol{w}+\frac{J\delta\lambda_1}{\mu v_1}}{\boldsymbol{y}\cdot\widetilde{\boldsymbol{v}}+\frac{J\delta\lambda_1}{\mu v_1}}d\boldsymbol{y}
\end{aligned}$$

for some $\boldsymbol{w}\in\mathbb{R}^{d-1}$ and $c' = a - \dfrac{\lambda_1}{\mu v_1}$. Take $a\in(0,\,a_0)$ to be sufficiently small such that $c'<0$ (which is possible by the statement of Lemma 3.4.10). Evaluating the last integral via Lemmas 3.4.8 and 3.4.11 and inserting the result into (3.52), we conclude that

$$I_{2,1} \;=\; o_\epsilon(1)\,\alpha_\epsilon + [\,1+o_\epsilon(1)\,]\,\alpha_\epsilon\,\frac{\mu^{1/2}\,(-\mathbb{L}\mathbb{H}^{-1}\boldsymbol{v})\cdot\boldsymbol{e}_1}{2\pi\sqrt{\det(\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}})}}\,. \qquad (3.53)$$

Next, we consider $I_{2,2}$. By Lemma 3.4.10, we have $\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}\setminus\partial_+^{1,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \subset$

$\partial_+^{2,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$; hence,

$$|\, I_{2,2}\,| \;\le\; \frac{C}{Z_\epsilon}\int_{\partial_+^{1,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} e^{-U(\boldsymbol{x})/\epsilon}\,\sigma(d\boldsymbol{x}) \;\le\; \frac{C}{Z_\epsilon}\,e^{-H/\epsilon}\,e^{-cJ^2\delta^2/\epsilon}\,\sigma(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})\,, \quad (3.54)$$

where we applied trivial bounds[3] for $|1-p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x})|$ and $\boldsymbol{\ell}$ in the first inequality, while we used the condition $U(\boldsymbol{x})\ge H+aJ^2\delta^2$ for $\boldsymbol{x}\in\partial_+^{2,a}\mathcal{B}_\epsilon$ in the second one. Since $\sigma(\partial_+\mathcal{B}_\epsilon)=O(\delta^{d-1})$, we get

$$|\, I_{2,2}\,| \;\le\; \frac{C\,\delta^{d-1}}{Z_\epsilon}\,\epsilon^{cJ^2/2} = o_\epsilon(1)\,\alpha_\epsilon \quad (3.55)$$

for sufficiently large J. Hence, $I_{2,2}$ is negligible. By combining (3.47), (3.53), and (3.55), we get

$$I_2 \;=\; o_\epsilon(1)\,\alpha_\epsilon \;+\big[\,1+o_\epsilon(1)\,\big]\,\alpha_\epsilon\,\frac{\mu^{1/2}\,(-\mathbb{L}\mathbb{H}^{-1}\boldsymbol{v})\cdot\boldsymbol{e}_1}{2\pi\sqrt{\det\big(\,\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}}\,\big)}}\;. \quad (3.56)$$

By (3.46) and (3.56), we obtain

$$I_1 - I_2 \;=\; \big[\,1+o_\epsilon(1)\,\big]\,\alpha_\epsilon\,\frac{\mu^{1/2}\,(\boldsymbol{v}+\mathbb{L}\mathbb{H}^{-1}\boldsymbol{v})\cdot\boldsymbol{e}_1}{2\pi\sqrt{\det\big(\,\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}}\,\big)}}\;. \quad (3.57)$$

Since $\mathbb{H}\mathbb{L}=-\mathbb{L}^\dagger\mathbb{H}$ by the skew-symmetry of $\mathbb{H}\mathbb{L}$, we have $\mathbb{L}\mathbb{H}^{-1}=-\mathbb{H}^{-1}\mathbb{L}^\dagger$. Hence,

$$\begin{aligned}(\boldsymbol{v}+\mathbb{L}\mathbb{H}^{-1}\boldsymbol{v})\cdot\boldsymbol{e}_1 &= (\mathbb{I}-\mathbb{H}^{-1}\mathbb{L}^\dagger)\boldsymbol{v}\cdot\boldsymbol{e}_1 \;=\; \mathbb{H}^{-1}(\mathbb{H}-\mathbb{L}^\dagger)\boldsymbol{v}\cdot\boldsymbol{e}_1\\ &= -\mu\,\mathbb{H}^{-1}\boldsymbol{v}\cdot\boldsymbol{e}_1 \;=\; \frac{\mu}{\lambda_1}\,\boldsymbol{v}\cdot\boldsymbol{e}_1 \;=\; \frac{\mu\,v_1}{\lambda_1}\end{aligned} \quad (3.58)$$

since $-\mu$ is an eigenvalue of $\mathbb{H}-\mathbb{L}^\dagger$ associated with the eigenvector $\boldsymbol{v}$ and

[3] Since $\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}\subset\mathcal{K}$ where $\mathcal{K}$ is defined in (3.28) we can bound $\boldsymbol{\ell}$ by the $L^\infty(\mathcal{K})$ norm of $\boldsymbol{\ell}$. This argument will be used repeatedly in the remainder of the chapter without further mention.

$\mathbb{H}^{-1} = \mathrm{diag}(-1/\lambda_1, 1/\lambda_2, \cdots, 1/\lambda_d)$. Inserting this computation and Lemma 3.4.9 into (3.57), we get

$$I_1 - I_2 \;=\; [\,1 + o_\epsilon(1)\,]\,\alpha_\epsilon\,\frac{\mu}{2\pi\sqrt{\prod_{k=1}^d \lambda_k}} \;=\; [1 + o_\epsilon(1)]\,\alpha_\epsilon\,\omega^{\boldsymbol{\sigma}}\;.$$

This completes the proof. □

3.4.5 Proof of Lemmas 3.4.10 and 3.4.11

Proof of Lemma 3.4.10. By Lemma 3.4.1, we have $\lambda_1 \sum_{k=2}^d v_k^2/\lambda_k < v_1^2$. Thus, there exists $\varepsilon_0 \in (0,\, v_1)$ such that

$$(\lambda_1 + \varepsilon_0)\sum_{k=2}^d \frac{v_k^2}{\lambda_k} \;<\; (v_1 - \varepsilon_0)^2\;. \tag{3.59}$$

Let $a_0 = \varepsilon_0 \min\{1,\, \lambda_1^{-1/2},\, \lambda_1^{-1}\}$, and we claim that this constant a_0 satisfies the requirement of the lemma.

Fix $a \in (0,\, a_0)$, $\boldsymbol{x} \in \partial_+\mathcal{C}_\epsilon$ and suppose, on the other hand, that

$$\boldsymbol{x}\cdot\boldsymbol{v} \;<\; aJ\delta \;\le\; \varepsilon_0\frac{J\delta}{\lambda_1^{1/2}} \quad\text{and}\quad U(\boldsymbol{x}) - H \;<\; aJ^2\delta^2 \;\le\; \varepsilon_0\frac{J^2\delta^2}{\lambda_1}\;. \tag{3.60}$$

Since $U(\boldsymbol{x}) - H \;=\; \frac{1}{2}\boldsymbol{x}\cdot\mathbb{H}\boldsymbol{x} + O(\delta^3)$ by the Taylor expansion, the latter condition implies that $\boldsymbol{x}\cdot\mathbb{H}\boldsymbol{x} < \varepsilon_0\frac{J^2\delta^2}{\lambda_1}$ for all sufficiently small $\epsilon > 0$.

Write $\boldsymbol{x} \in \partial_+\mathcal{C}_\epsilon$ as $\boldsymbol{x} \;=\; \frac{J\delta}{\lambda_1^{1/2}}\big(\boldsymbol{e}_1 + \sum_{k=2}^d x_k\boldsymbol{e}_k\big)$ such that we can rewrite the two conditions of (3.60) respectively as

$$0 \;<\; v_1 - \varepsilon_0 \;<\; -\sum_{k=2}^d v_kx_k \quad\text{and}\quad \sum_{k=2}^d \lambda_k x_k^2 \;<\; \lambda_1 + \varepsilon_0\;.$$

By these two inequalities and (3.59), we have

$$\sum_{j=2}^{d} \lambda_j x_j^2 \sum_{k=2}^{d} \frac{v_k^2}{\lambda_k} \;<\; (\lambda_1+\varepsilon_0) \sum_{k=2}^{d} \frac{v_k^2}{\lambda_k} \;<\; (v_1-\varepsilon_0)^2 \;<\; \Big(\sum_{k=2}^{d} x_k v_k \Big)^2 \;,$$

which contradicts the Cauchy–Schwarz inequality; hence, the claim is proven.

□

Proof of Lemma 3.4.11. Write $\zeta = \zeta(\epsilon) = \sqrt{\log \frac{1}{\epsilon}}$ and let $\mathcal{Q}_\epsilon = \Pi_\epsilon(\partial_+ \mathcal{B}_\epsilon)$. Then, by the change of variables $\boldsymbol{z} = \epsilon^{-1/2} \boldsymbol{y}$, we can write the integral in the statement of the lemma as

$$\epsilon^{(d-1)/2} \int_{\epsilon^{-1/2}\mathcal{Q}_\epsilon \cap \{\boldsymbol{z}\in\mathbb{R}^{d-1}:\boldsymbol{z}\cdot\boldsymbol{u}_1 \ge -c\zeta\}} \frac{\boldsymbol{z}\cdot\boldsymbol{u}_2+\zeta}{\boldsymbol{z}\cdot\boldsymbol{u}_1+\zeta}\, e^{-(1/2)\boldsymbol{z}\cdot\mathbb{D}\boldsymbol{z}} d\boldsymbol{z} \;.$$

Fix $0<\alpha<1$. Then, since $\zeta \to \infty$ as $\epsilon \to 0$, by Lemma 3.4.8,

$$\mathcal{D}^{(d-1)}_{r\zeta^\alpha}(\boldsymbol{0}) \;\subset\; \epsilon^{-1/2}\mathcal{Q}_\epsilon \cap \{\boldsymbol{z}\in\mathbb{R}^{d-1} : \boldsymbol{z}\cdot\boldsymbol{u}_1 \ge -c\zeta\}$$

for all sufficiently small $\epsilon>0$. Now we decompose the integral into

$$\Big[\int_{\mathcal{D}^{(d-1)}_{r\zeta^\alpha}(\boldsymbol{0})} + \int_{\{\epsilon^{-1/2}\mathcal{Q}_\epsilon \setminus \mathcal{D}^{(d-1)}_{r\zeta^\alpha}(\boldsymbol{0})\}\cap\{\boldsymbol{z}\in\mathbb{R}^{d-1}:\boldsymbol{z}\cdot\boldsymbol{u}_1\ge -c\zeta\}} \Big] \frac{\boldsymbol{z}\cdot\boldsymbol{u}_2+\zeta}{\boldsymbol{z}\cdot\boldsymbol{u}_1+\zeta}\, e^{-(1/2)\boldsymbol{z}\cdot\mathbb{D}\boldsymbol{z}} d\boldsymbol{z} \;. \tag{3.61}$$

Let us consider the first integral. Note that

$$\sup_{\boldsymbol{z}\in\mathcal{D}^{(d-1)}_{r\zeta^\alpha}(\boldsymbol{0})} \Big| \frac{\boldsymbol{z}\cdot\boldsymbol{u}_2+\zeta}{\boldsymbol{z}\cdot\boldsymbol{u}_1+\zeta} - 1 \Big| \;=\; o_\epsilon(1) \;.$$

Thus, the first integral is

$$[\,1+o_\epsilon(1)\,] \int_{\mathcal{D}^{(d-1)}_{r\zeta^\alpha}(\boldsymbol{0})} e^{-(1/2)\boldsymbol{z}\cdot\mathbb{D}\boldsymbol{z}} d\boldsymbol{z} \;=\; [\,1+o_\epsilon(1)\,] \frac{(2\pi)^{(d-1)/2}}{\sqrt{\det(\mathbb{D})}} \;, \tag{3.62}$$

since $\mathcal{D}^{(d-1)}_{r\zeta^\alpha}(\mathbf{0}) \uparrow \mathbb{R}^{d-1}$ as $\epsilon \to 0$.

Now, we focus on the second integral. Since $\epsilon^{-1/2}\mathcal{Q}_\epsilon \subset \mathcal{D}^{(d-1)}_{R\zeta}(\mathbf{0})$ by Lemma 3.4.8, and since $\boldsymbol{z} \cdot \boldsymbol{u}_1 \ge -c\zeta$ for $c \in (0,\,1)$ by the statement of the lemma, there exists $C > 0$ such that

$$\sup_{\boldsymbol{z} \in \epsilon^{-1/2}\mathcal{Q}_\epsilon} \Big| \frac{\boldsymbol{z} \cdot \boldsymbol{u}_2 + \zeta}{\boldsymbol{z} \cdot \boldsymbol{u}_1 + \zeta} \Big| \le C \;.$$

Hence, the absolute value of the second integral in (3.61) is bounded from above by

$$C \int_{\mathcal{D}^{(d-1)}_{R\zeta}(\mathbf{0}) \setminus \mathcal{D}^{(d-1)}_{r\zeta^\alpha}(\mathbf{0})} e^{-(1/2)\boldsymbol{z}\cdot\mathbb{D}\boldsymbol{z}}\, d\boldsymbol{z} \;=\; o_\epsilon(1) \;. \tag{3.63}$$

By combining (3.61), (3.62), and (3.63), we complete the proof. □

3.5 Analysis of equilibrium potential

In this section, we establish a bound on the equilibrium potential h_ϵ and h^*_ϵ in Proposition 3.5.1. On the basis of this bound, we prove Proposition 3.3.3 in Section 3.5.4. Further, we remark that this bound plays an important role in the proof of Theorem 3.3.4 (cf. Section 3.6.4).

For two disjoint non-empty sets $\mathcal{A}$, $\mathcal{B} \subset \mathbb{R}^d$, let $\Gamma_{\mathcal{A},\mathcal{B}}$ be a set of all C^1-paths $\boldsymbol{\gamma} : [0,1] \to \mathbb{R}^d$ such that $\boldsymbol{\gamma}(0) \in \mathcal{A}$ and $\boldsymbol{\gamma}(1) \in \mathcal{B}$. Then, let $\mathfrak{H}_{\mathcal{A},\mathcal{B}}$ denote the height of the saddle points between $\mathcal{A}$ and $\mathcal{B}$:

$$\mathfrak{H}_{\mathcal{A},\mathcal{B}} := \inf_{\gamma \in \Gamma_{\mathcal{A},\mathcal{B}}} \sup_{t \in [0,\,1]} U(\,\boldsymbol{\gamma}(t)\,) \;.$$

3.5.1 Estimates of equilibrium potentials h_ϵ and h^*_ϵ

In this subsection, we prove the following proposition regarding the so-called leveling property of the equilibrium potential.

Proposition 3.5.1. *We can find a constant $C > 0$ satisfying the following bounds.*

1. *For all $\boldsymbol{y} \in \mathcal{H}_0$, the following holds:*

$$h_\epsilon(\boldsymbol{y}),\, h_\epsilon^*(\boldsymbol{y}) \,\geq\, 1 - C\,\epsilon^{-d}\, \exp\frac{\mathfrak{H}_{\{\boldsymbol{y}\},\, \mathcal{D}_\epsilon(\boldsymbol{m}_0)} - H}{\epsilon}\ .$$

2. *For all $\boldsymbol{y} \in \mathcal{H}_1$, the following holds:*

$$h_\epsilon(\boldsymbol{y}),\, h_\epsilon^*(\boldsymbol{y}) \,\leq\, C\,\epsilon^{-d}\, \exp\frac{U(\boldsymbol{y}) - H}{\epsilon}\ .$$

The proof of Proposition 3.5.1 relies on the following two bounds on the capacity.

Lemma 3.5.2. *There exists $C > 0$ such that for all $\boldsymbol{y} \in \mathcal{W}_0$ and $\boldsymbol{m} \in \mathcal{M}_0$,*

$$\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}),\, \mathcal{D}_\epsilon(\boldsymbol{m})) \,\geq\, C\,\epsilon^d\, Z_\epsilon^{-1}\, e^{-\mathfrak{H}_{\{\boldsymbol{y}\},\mathcal{D}_\epsilon(\boldsymbol{m})}/\epsilon}\ .$$

Lemma 3.5.3. *There exists $C > 0$ such that for all $\boldsymbol{y} \in \mathcal{H}_0$,*

$$\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}),\, \mathcal{U}_\epsilon) \,\leq\, C\, Z_\epsilon^{-1}\, e^{-H/\epsilon}\ .$$

We prove Lemmas 3.5.2 and 3.5.3 in Sections 3.5.2 and 3.5.3, respectively. Now, we prove Proposition 3.5.1

Proof of Proposition 3.5.1. Since the proofs for h_ϵ and h_ϵ^* are identical, we consider only h_ϵ. In [51, Proposition 7.9], it has been shown that there exists $C > 0$ such that

$$h_{\mathcal{A},\mathcal{B}}(\boldsymbol{x}) \,\leq\, C\, \frac{\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{x}),\, \mathcal{A})}{\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{x}),\, \mathcal{B})}\ , \tag{3.64}$$

provided that $\mathcal{A}$ and $\mathcal{B}$ are disjoint domains of sufficiently smooth bounds.

For part (1), we can use this bound to get

$$1-h_\epsilon(\boldsymbol{y}) \;=\; h_{\mathcal{U}_\epsilon,\,\mathcal{D}_\epsilon(\boldsymbol{m}_0)}(\boldsymbol{y}) \;\le\; C\,\frac{\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}),\,\mathcal{U}_\epsilon)}{\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}),\,\mathcal{D}_\epsilon(\boldsymbol{m}_0))}\;.$$

Now, by applying Lemmas 3.5.2 and 3.5.3, we complete the proof of part (1).

For part (2), we fix $\boldsymbol{y} \in \mathcal{H}_1$. Then, again by (3.64),

$$h_\epsilon(\boldsymbol{y}) \;=\; h_{\mathcal{D}_\epsilon(\boldsymbol{m}_0),\mathcal{U}_\epsilon}(\boldsymbol{y}) \;\le\; C\,\frac{\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}),\,\mathcal{D}_\epsilon(\boldsymbol{m}_0))}{\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}),\,\mathcal{U}_\epsilon)}\;.$$

By the same logic with the proofs of Lemmas 3.5.2 and 3.5.3, we get

$$\begin{aligned}\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}),\,\mathcal{D}_\epsilon(\boldsymbol{m}_0)) \;&\le\; \frac{Ce^{-H/\epsilon}e^{-H/\epsilon}}{Z_\epsilon} \quad \text{and}\\ \mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}),\,\mathcal{U}_\epsilon) \;&\ge\; \frac{C\epsilon^d}{Z_\epsilon}\,e^{-\mathfrak{H}_{\{\boldsymbol{y}\},\mathcal{U}_\epsilon}/\epsilon}\;.\end{aligned}$$

Since $\mathcal{U}_\epsilon$ contains all the local minima of $\mathcal{M}_1$ and $\mathcal{H}_1$ is a subset of the domain of attraction of $\mathcal{M}_1$, we have $\mathfrak{H}_{\{\boldsymbol{y}\},\mathcal{U}_\epsilon} = U(\boldsymbol{y})$ and the proof is completed. □

3.5.2 Proof of Lemma 3.5.2

For the lower bound case, the proof is a consequence of the existing estimate for the reversible case. Let $\mathrm{cap}^{\mathrm{s}}_\epsilon(\cdot,\,\cdot)$ denote the capacity with respect to the reversible process $\boldsymbol{y}_\epsilon(\cdot)$ given in (1.3), whose generator is $(1/2)(\mathscr{L}_\epsilon + \mathscr{L}^*_\epsilon)$. Then, it is well known that (cf. [32, Lemma 2.5]) for any two disjoint non-empty domains $\mathcal{A},\,\mathcal{B} \subset \mathbb{R}^d$ with smooth boundaries, we have the following equation:

$$\mathrm{cap}_\epsilon(\mathcal{A},\,\mathcal{B}) \;\ge\; \mathrm{cap}^{\mathrm{s}}_\epsilon(\mathcal{A},\,\mathcal{B})\;. \tag{3.65}$$

Therefore, it suffices to show the inequality for $\mathrm{cap}^{\mathrm{s}}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}),\,\mathcal{D}_\epsilon(\boldsymbol{m}))$, instead. The lower bound for this capacity can be obtained by optimizing the integration on the tube connecting $\mathcal{D}_\epsilon(\boldsymbol{y})$ and $\mathcal{D}_\epsilon(\boldsymbol{m})$. This is rigorously achieved by a parametrization of this tube. When we parametrize the tube successfully, we can use the idea of [14, Proposition 4.7] to complete the proof.

Let $\boldsymbol{\omega} : [0, L] \to \mathbb{R}^d$ be a smooth path such that $|\dot{\boldsymbol{\omega}}(t)| = 1$ for all $t \in [0,\, L]$. For $r > 0$, define $A_r(0)$, $A_r(L)$ by

$$\begin{aligned} A_r(0) \;&=\; \{\, \boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \cdot \dot{\boldsymbol{\omega}}(0) < 0,\, |\boldsymbol{x} - \boldsymbol{\omega}(0)| < r \,\} \\ A_r(L) \;&=\; \{\, \boldsymbol{x} \in \mathbb{R}^d : \boldsymbol{x} \cdot \dot{\boldsymbol{\omega}}(L) > 0,\, |\boldsymbol{x} - \boldsymbol{\omega}(L)| < r \,\} \end{aligned}$$

and define the *tubular neighborhood of* $\boldsymbol{\omega}$ *of radius* r by

$$\boldsymbol{\omega}_r \;=\; \{\boldsymbol{x} \in \mathbb{R}^d : |\boldsymbol{x} - \boldsymbol{\omega}(t)| < r \text{ for some } t \in [0, L]\} \setminus (\, A_r(0) \cup A_r(L) \,).$$

For $\rho > 0$, let $\mathcal{D}_\rho^{(d-1)}$ be a $(d-1)$-dimensional sphere of radius ρ centered at the origin.

Lemma 3.5.4. *There exists* $r_0 > 0$ *such that* $[0, L] \times \mathcal{D}_{r_0}^{(d-1)}$ *is diffeomorphic to* $\boldsymbol{\omega}_{r_0}$*. Furthermore, we can find a diffeomorphism* $\varphi : [0, L] \times \mathcal{D}_{r_0}^{(d-1)} \to \boldsymbol{\omega}_{r_0}$ *of the form*

$$\varphi(t,\, \boldsymbol{z}) \;=\; \boldsymbol{\omega}(t) + \mathbb{A}(t)\boldsymbol{z} \tag{3.66}$$

for some smooth $d \times (d-1)$ *matrix-valued function* $\mathbb{A}(\cdot)$ *of rank* $d-1$*, and it satisfies*

$$\Big|\det \frac{\partial \varphi}{\partial (t,\, \boldsymbol{z})}\Big| \;\geq\; \frac{1}{2} \quad \text{on } [0,\, L] \times \mathcal{D}_{r_0}^{(d-1)}\,. \tag{3.67}$$

Proof. The proof needs to recall several notions and results from differential geometry. We refer to [56] for a reference. We regard $\boldsymbol{\omega} = \boldsymbol{\omega}(\,[\,0,\, L\,]\,)$ as a one-dimensional compact manifold. Let $N\boldsymbol{\omega} \subset \mathbb{R}^d \times \mathbb{R}^d$ denote the normal bundle of $\boldsymbol{\omega}$. By the tubular neighborhood theorem (cf. [56, Theorem 6.24]), there exists $r_0 > 0$ such that $\boldsymbol{\omega}_{r_0}$ is diffeomorphic to $N\boldsymbol{\omega}_{r_0} = \{\, (\boldsymbol{p},\, \boldsymbol{v}) \in N\boldsymbol{\omega} : |\boldsymbol{v}| < r_0 \,\}$. The diffeomorphism $E : N\boldsymbol{\omega}_{r_0} \to \boldsymbol{\omega}_{r_0}$ is given by $E(\boldsymbol{p},\, \boldsymbol{v}) = \boldsymbol{p} + \boldsymbol{v}$. Since $\boldsymbol{\omega}$ is contractible, the vector bundle of $\boldsymbol{\omega}$ is trivial; thus, $N\boldsymbol{\omega}$ is diffeomorphic to $\boldsymbol{\omega} \times \mathbb{R}^{d-1}$. Let $\phi : \boldsymbol{\omega} \times \mathbb{R}^{d-1} \to N\boldsymbol{\omega}$ denote the corresponding diffeomorphism. Since this diffeomorphism preserves the vector space structure, the function $\phi(\boldsymbol{p},\, \boldsymbol{z})$ is linear in $\boldsymbol{z}$ and satisfies $|\pi_2(\phi(\boldsymbol{p},\, \boldsymbol{z}))| = |\boldsymbol{z}|$ where $\pi_2 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is the projection function for the second coordinate.

Since $\boldsymbol{\omega} \times \mathbb{R}^{d-1}$ is a trivial bundle of rank $d-1$, there are $d-1$ smooth sections $\sigma_j : \boldsymbol{\omega} \to \mathbb{R}^{d-1}$ which are linearly independent. By the Gram–Schmidt operation, we may assume that they are pointwise orthonormal, i.e., $\sigma_i(\boldsymbol{p}) \cdot \sigma_j(\boldsymbol{p}) = \delta_{i,\,j}$ for all $i,\, j$ and $\boldsymbol{p} \in \boldsymbol{\omega}$. Define a $d \times (d-1)$ matrix $\mathbb{B}(\boldsymbol{p}) = [\, \mathbb{B}_1(\boldsymbol{p}),\, \dots,\, \mathbb{B}_{d-1}(\boldsymbol{p})\,]$ by $\mathbb{B}_i(\boldsymbol{p}) = \pi_2(\phi(\boldsymbol{p},\, \sigma_i(\boldsymbol{p})))$ for $j = 1,\, \dots,\, d-1$. By the smoothness of ϕ and σ_j, we can observe that all the elements of $\mathbb{B}(\cdot)$ are smooth. Then, the diffeomorphism $\varphi : [0, L] \times \mathcal{D}_{r_0}^{(d-1)} \to \boldsymbol{\omega}_{r_0}$ can be written as

$$\varphi(t,\, \boldsymbol{z}) \;=\; \phi(\boldsymbol{\omega}(t),\, \boldsymbol{z}) \;=\; \boldsymbol{\omega}(t) + \mathbb{B}(\boldsymbol{\omega}(t))\boldsymbol{z}\ .$$

We can now take $\mathbb{A} = \mathbb{B} \circ \omega$ to get (3.66). Now we consider (3.67). We can write

$$\frac{\partial \varphi}{\partial (t,\, \boldsymbol{z})}(t,\, \mathbf{0}) \;=\; [\, \dot{\boldsymbol{\omega}}(t),\, \mathbb{A}(t)\,]\ .$$

Since all the column vectors in the matrix on the right-hand sides are normal and orthogonal to each other, we have $\big|\det \dfrac{\partial \varphi}{\partial (t,\, \boldsymbol{z})}(t,\, \mathbf{0})\big| = 1$. Hence, by taking r_0 to be sufficiently small, we get (3.67). □

Proposition 3.5.5. *Let $\boldsymbol{\omega} : [0,\, L] \to \mathbb{R}^d$ be a C^1-path connecting $\boldsymbol{y}$ and $\boldsymbol{m}$ such that $U(\boldsymbol{\omega}(t)) \le M$ and $|\dot{\boldsymbol{\omega}}(t)| = 1$ for all t. Moreover, let f be a smooth function such that $f \equiv 1$ on $\mathcal{D}_\epsilon(\boldsymbol{y})$ and $f \equiv 0$ on $\mathcal{D}_\epsilon(\boldsymbol{m})$. Then, there exists a constant $C > 0$ such that*

$$\epsilon \int_{\boldsymbol{\omega}_{r_0}} |\, \nabla f\, |^2\, d\mu_\epsilon \;\ge\; C\, L^{-1}\, \epsilon^d\, Z_\epsilon^{-1}\, e^{-M/\epsilon}\ ,$$

where r_0 is the constant obtained in Lemma 3.5.4 for the path $\boldsymbol{\omega}$.

Proof. By Lemma 3.5.4, we have

$$\begin{aligned}&\epsilon\int_{\boldsymbol{\omega}_{r_0}}|\nabla f|^2 d\mu_\epsilon\\&\ge\frac{\epsilon}{2Z_\epsilon}\int_{\mathcal{D}_\epsilon^{(d-1)}}\int_0^L|\nabla f(\boldsymbol{\omega}(t)+\mathbb{A}(t)\boldsymbol{z})|^2\,e^{-U(\boldsymbol{\omega}(t)+\mathbb{A}(t)\boldsymbol{z})/\epsilon}\,dt\,d\boldsymbol{z}\end{aligned}$$

for $\epsilon\in(0,r_0)$, where the factor of 2 appears because (3.67) is used for bounding the Jacobian of the change of variables from below. For $(t,\boldsymbol{z})\in[0,\,L]\times\mathcal{D}_\epsilon^{(d-1)}$, we have

$$\begin{aligned}&\frac{d}{dt}f(\boldsymbol{\omega}(t)+\mathbb{A}(t)\boldsymbol{z})\\&=\nabla f(\boldsymbol{\omega}(t)+\mathbb{A}(t)\boldsymbol{z})\cdot(\dot{\boldsymbol{\omega}}(t)+\dot{\mathbb{A}}(t)\boldsymbol{z})\;\le\;2|\nabla f(\boldsymbol{\omega}(t)+\mathbb{A}(t)\boldsymbol{z})|\,,\end{aligned}$$

where the last inequality holds for sufficiently small ϵ since $|\dot{\omega}(t)|=1$ and $|\boldsymbol{z}|\le\epsilon$. Summing up, we can write

$$\begin{aligned}&\epsilon\int_{\boldsymbol{\omega}_{r_0}}|\nabla f|^2\,d\mu_\epsilon\\&\ge\frac{\epsilon}{4Z_\epsilon}\int_{\mathcal{D}_\epsilon^{(d-1)}}\int_0^L\Big|\frac{d}{dt}f(\boldsymbol{\omega}(t)+\mathbb{A}(t)\boldsymbol{z})\Big|^2 e^{-U(\boldsymbol{w}(t)+\mathbb{A}(t)\boldsymbol{z})/\epsilon}\,dt\,d\boldsymbol{z}\,.\end{aligned}\tag{3.68}$$

Now, we can apply the idea of [14, Proposition 4.7]. Indeed, we can fix $\boldsymbol{z}\in\mathcal{D}_\epsilon^{(d-1)}$ and write $f_{\boldsymbol{z}}(t)=f(\boldsymbol{\omega}(t)+\mathbb{A}(t)\boldsymbol{z})$. Then, we can obtain the minimizer of the integral $\displaystyle\int_0^L\left|\frac{d}{dt}f_{\boldsymbol{z}}(t)\right|^2 e^{-U(\boldsymbol{w}(t)+\mathbb{A}(t)\boldsymbol{z})/\epsilon}dt$ explicitly as

$$f_{\boldsymbol{z}}(t)\;=\;\frac{\int_t^L\;e^{U(\boldsymbol{\omega}(s)+\mathbb{A}(s)\boldsymbol{z})/\epsilon}\,ds}{\int_0^L\;e^{U(\boldsymbol{\omega}(s)+\mathbb{A}(s)\boldsymbol{z})/\epsilon}\,ds}\,.$$

Inserting this solution into (3.68) gives

$$\epsilon \int_{\boldsymbol{\omega}_{r_0}} |\nabla f|^2 d\mu_\epsilon \geq \frac{\epsilon}{4Z_\epsilon} \int_{\mathcal{D}_\epsilon^{(d-1)}} \Big[\int_0^L e^{U(\omega(t)+\mathbb{A}(t)\boldsymbol{z})/\epsilon} dt \Big]^{-1} d\boldsymbol{z} .$$

Since $|\boldsymbol{z}| \leq \epsilon$, we have $U(\omega(t) + \mathbb{A}(t)\boldsymbol{z}) \leq M + C\,\epsilon$ for some constant $C > 0$, and the proof is completed. □

Now, we are ready to prove Lemma 3.5.2.

Proof of Lemma 3.5.2. Fix $\boldsymbol{y} \in \mathcal{H}_0$ and for some $L = L(\boldsymbol{y})$, let $\boldsymbol{\omega} : [0,\, L] \to \mathbb{R}^d$ be a C^1-path connecting $\boldsymbol{y}$ to $\mathcal{D}_\epsilon(\boldsymbol{m})$ such that $U(\boldsymbol{\omega}(t)) \leq \mathfrak{H}_{\{\boldsymbol{y}\},\,\mathcal{D}_\epsilon(\boldsymbol{m})}$ and $|\dot{\boldsymbol{\omega}}(t)| = 1$ for all $t \in [0,\, L]$. Since $\mathcal{H}_0$ is bounded, we can find L_0 such that $L(\boldsymbol{y}) < L_0$ for all $\boldsymbol{y} \in \mathcal{H}_0$. Then, recall the diffeomorphism $\varphi : [0,\, L] \times \mathcal{D}_{r_0}^{(d-1)} \to \boldsymbol{\omega}_{r_0}$ constructed in Lemma 3.5.4. Then, e

$$\mathrm{cap}_\epsilon^{\mathrm{s}}(\,\mathcal{D}_\epsilon(\boldsymbol{y}),\, \mathcal{D}_\epsilon(\boldsymbol{m})\,) \geq \epsilon \int_{\boldsymbol{\omega}_{r_0}} |\nabla h^{\epsilon,\,\mathrm{s}}_{\mathcal{D}_\epsilon(\boldsymbol{y}),\,\mathcal{D}_\epsilon(\boldsymbol{m})}|^2 \, d\mu_\epsilon \ ,$$

where $h^{\epsilon,\,\mathrm{s}}_{\mathcal{D}_\epsilon(\boldsymbol{y}),\,\mathcal{D}_\epsilon(\boldsymbol{m})}(\cdot)$ is the equilibrium potential between $\mathcal{D}_\epsilon(\boldsymbol{y})$ and $\mathcal{D}_\epsilon(\boldsymbol{m})$ with respect to the reversible process $\boldsymbol{y}_\epsilon(\cdot)$. Hence, by Proposition 3.5.5 and the fact that we can take $L(\boldsymbol{y})$ to be uniformly bounded by L_0, the proof is completed. □

3.5.3 Proof of Lemma 3.5.3

The upper bound cannot be proven by a comparison with reversible dynamics as in the lower bound case unless the dynamics satisfies the so-called sector condition, and that is exactly what has been used in [51]. However, the dynamics $\boldsymbol{x}_\epsilon(\cdot)$ does not necessarily satisfy the sector condition; hence, we must develop a new argument. We believe that our argument presented below is sufficiently robust to treat a wide class of models.

Proof of Lemma 3.5.3. For each set $\mathcal{A} \subset \mathbb{R}^d$ and $r > 0$, define

$$\mathcal{A}^{[r]} = \{ \boldsymbol{x} \in \mathbb{R}^d : |\boldsymbol{x} - \boldsymbol{y}| \le r \text{ for some } \boldsymbol{y} \in \mathcal{A} \} . \tag{3.69}$$

Suppose that ϵ is sufficiently small such that $\mathcal{H}_0^{[2\epsilon]}$ is disjoint from $\mathcal{U}_\epsilon$ and $\mathcal{H}_0^{[2\epsilon]} \subset \mathcal{K}$ (cf. (3.28)). Take a smooth function $q_\epsilon : \mathbb{R}^d \to \mathbb{R}$ such that, for some constant $C > 0$,

$$q_\epsilon \equiv 1 \text{ on } \mathcal{H}_0^{[\epsilon]} \ , \quad q_\epsilon \equiv 0 \text{ on } \mathbb{R}^d \setminus \mathcal{H}_0^{[2\epsilon]} \ , \text{ and } \ |\nabla q_\epsilon| \le \frac{C}{\epsilon} \mathbf{1}_{\mathcal{H}_0^{[2\epsilon]} \setminus \mathcal{H}_0^{[\epsilon]}} \ . \tag{3.70}$$

Since $q_\epsilon \in \mathscr{C}_{\mathcal{D}_\epsilon(\boldsymbol{y}), \mathcal{U}_\epsilon}$ (cf. (3.14)), we can deduce from Proposition 3.2.2 that

$$\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}), \mathcal{U}_\epsilon) = \epsilon \int_{\Omega_\epsilon} \Big[\nabla q_\epsilon \cdot \nabla h_\epsilon + \frac{1}{\epsilon} q_\epsilon \boldsymbol{\ell} \cdot \nabla h_\epsilon \Big] \, d\mu_\epsilon \ . \tag{3.71}$$

By the divergence theorem and (2.3), the second term on the right-hand side can be rewritten as

$$\int_{\partial \Omega_\epsilon} h_\epsilon \, q_\epsilon \, [\, \boldsymbol{\ell} \cdot \boldsymbol{n}_{\Omega_\epsilon}] \, \sigma(d\mu_\epsilon) - \int_{\Omega_\epsilon} h_\epsilon \, [\, \nabla q_\epsilon \cdot \boldsymbol{\ell} \,] \, d\mu_\epsilon \ . \tag{3.72}$$

Since $h_\epsilon = \mathbf{1}_{\partial \mathcal{D}_\epsilon(\boldsymbol{y})}$ on $\partial \Omega_\epsilon = \partial \mathcal{U}_\epsilon \cup \partial \mathcal{D}_\epsilon(\boldsymbol{y})$, $q_\epsilon \equiv 1$ on $\partial \mathcal{D}_\epsilon(\boldsymbol{y})$, and $\boldsymbol{n}_{\Omega_\epsilon} = -\boldsymbol{n}_{\mathcal{D}_\epsilon(\boldsymbol{y})}$, the first integral of (3.72) becomes

$$-\int_{\partial \mathcal{D}_\epsilon(\boldsymbol{y})} [\, \boldsymbol{\ell} \cdot \boldsymbol{n}_{\mathcal{D}_\epsilon(\boldsymbol{y})}] \, \sigma(d\mu_\epsilon) = \int_{\mathcal{D}_\epsilon(\boldsymbol{y})} (\, \nabla \cdot \boldsymbol{\ell} \,) \, d\mu_\epsilon + \int_{\mathcal{D}_\epsilon(\boldsymbol{y})} [\, \boldsymbol{\ell} \cdot \nabla \mu_\epsilon \,](\boldsymbol{x}) \, d\boldsymbol{x} \tag{3.73}$$

by the divergence theorem again. Note that the last two integrals are 0 by (2.3) and (2.2), respectively. Hence the first integral of (3.72) vanishes. For the second integral of (3.72), by the trivial bound $|h_\epsilon| \le 1$ and the last condition of (3.70), we have

$$\Big| \int_{\Omega_\epsilon} h_\epsilon \, [\nabla q_\epsilon \cdot \boldsymbol{\ell}] \, d\mu_\epsilon \Big| \le \frac{C}{\epsilon Z_\epsilon} \int_{\mathcal{H}_0^{[2\epsilon]} \setminus \mathcal{H}_0^{[\epsilon]}} e^{-U(\boldsymbol{x})/\epsilon} \, d\boldsymbol{x} \le \frac{C}{Z_\epsilon} e^{-H/\epsilon} \ , \tag{3.74}$$

where the second inequality follows from the fact that $U(\boldsymbol{x}) = H + O(\epsilon)$ on $\mathcal{H}_0^{[2\epsilon]} \setminus \mathcal{H}_0^{[\epsilon]}$ and that $\mathrm{vol}\,(\mathcal{H}_0^{[2\epsilon]} \setminus \mathcal{H}_0^{[\epsilon]}) = O(\epsilon)$. Summing up, we obtain from (3.71) that

$$\mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}), \mathcal{U}_\epsilon) \;\le\; \epsilon \int_{\Omega_\epsilon} [\nabla q_\epsilon \cdot \nabla h^\epsilon]\, d\mu_\epsilon + \frac{C}{Z_\epsilon}\, e^{-H/\epsilon}\,. \tag{3.75}$$

By the Cauchy–Schwarz inequality and part (2) of Lemma 3.2.1, the integral on the right-hand side is bounded from above by the square root of

$$\epsilon \int_{\Omega_\epsilon} |\, \nabla q_\epsilon\,|^2\, d\mu_\epsilon \times \mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}), \mathcal{U}_\epsilon)\;.$$

By a computation similar to (3.74), we get

$$\epsilon \int_{\Omega_\epsilon} |\nabla q_\epsilon|^2\, d\mu_\epsilon \;\le\; \frac{C}{\epsilon Z_\epsilon} \int_{\mathcal{H}_0^{[2\epsilon]} \setminus \mathcal{H}_0^{[\epsilon]}} e^{-U(\boldsymbol{x})/\epsilon}\, d\boldsymbol{x} \;\le\; \frac{C}{Z_\epsilon}\, e^{-H/\epsilon}\;.$$

Therefore, we can bound the integral on the right-hand side of (3.75) by

$$\Big[\, \frac{C}{Z_\epsilon}\, e^{-H/\epsilon}\, \mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}), \mathcal{U}_\epsilon) \Big]^{1/2} \;\le\; \frac{1}{2} \Big[\, \frac{C}{Z_\epsilon}\, e^{-H/\epsilon} + \mathrm{cap}_\epsilon(\mathcal{D}_\epsilon(\boldsymbol{y}), \mathcal{U}_\epsilon) \Big]\;.$$

Inserting this into (3.75) completes the proof. □

3.5.4 Proof of Proposition 3.3.3

Now, we are ready to prove Proposition 3.3.3, which is a crucial step in the proof of the Eyring–Kramers formula.

Proof of Proposition 3.3.3. Take $\beta > 0$ to be sufficiently small such that there is no critical point $\boldsymbol{c}$ of U such that $U(\boldsymbol{c}) \in [H - \beta,\, H)$. Then, we can decompose $\mathcal{G} = \{\boldsymbol{x} : U(\boldsymbol{x}) < H - \beta\}$ into $\mathcal{G}_0$, $\mathcal{G}_1$, where $\mathcal{G}_0 \subset \mathcal{H}_0$ and

$\mathcal{G}_1 \subset \mathcal{H}_1$. Write

$$\int_{\mathbb{R}^d} h_\epsilon^* \, d\mu_\epsilon \,=\, \Big[\int_{\mathcal{G}_0} + \int_{\mathcal{G}_1} + \int_{\mathcal{G}^c} \Big] \, h_\epsilon^* \, d\mu_\epsilon \tag{3.76}$$

and consider the three integrals separately. First, for $\boldsymbol{y} \in \mathcal{G}_0$, we have $\mathfrak{H}_{\{\boldsymbol{y}\},\mathcal{D}_\epsilon(\boldsymbol{m}_0)} < H - \beta$; thus, by part (1) of Proposition 3.5.1, we have $|\, h_\epsilon^*(\boldsymbol{y}) - 1\,| \le C\,\epsilon^{-d}\, e^{-\beta/\epsilon} = o_\epsilon(1)$. This bound ensures that

$$\int_{\mathcal{G}_0} h_\epsilon^* \, d\mu_\epsilon \,=\, [\,1 + o_\epsilon(1)\,]\, \mu_\epsilon(\mathcal{G}_0) \,=\, [\,1 + o_\epsilon(1)\,]\, Z_\epsilon^{-1}\, (2\pi\epsilon)^{d/2}\, e^{-h_0/\epsilon}\, \nu_0 \;, \tag{3.77}$$

where the second identity follows from the Laplace asymptotics for the function $e^{-U/\epsilon}$.

For the second integral, by part (2) of Proposition 3.5.1,

$$\int_{\mathcal{G}_1} h_\epsilon^* \, d\mu_\epsilon \,\le\, \frac{C}{Z_\epsilon\, \epsilon^d} \int_{\mathcal{G}_1} e^{[U(\boldsymbol{x}) - H]/\epsilon}\, e^{-U(\boldsymbol{x})/\epsilon}\, d\boldsymbol{x} \,=\, o_\epsilon(1)\; Z_\epsilon^{-1}\, (2\pi\epsilon)^{d/2}\, e^{-h_0/\epsilon}\, \nu_0 \;, \tag{3.78}$$

where the last line follows from $H > h_0$. Finally, for the last integral, by the bound $|h_\epsilon^*| \le 1$ and (2.7),

$$\int_{\mathcal{G}^c} h_\epsilon^* \, d\mu_\epsilon \,\le\, \mu_\epsilon(\mathcal{G}^c) \,\le\, Z_\epsilon^{-1}\, e^{-(H-\beta)/\epsilon} \,=\, o_\epsilon(1)\, Z_\epsilon^{-1}\, (2\pi\epsilon)^{d/2}\, e^{-h_0/\epsilon}\, \nu_0 \;. \tag{3.79}$$

By inserting (3.77), (3.78), and (3.79) into (3.76), the proof is completed. □

3.6 Construction of test function and proof of Theorem 3.3.4

In this section, we finally construct the test function $g_\epsilon \in \mathscr{C}_{\mathcal{D}_\epsilon(\boldsymbol{m}_0),\mathcal{U}_\epsilon}$ satisfying Theorem 3.3.4.

3.6.1 Construction of g_ϵ and proof of Theorem 3.3.4

Recall $\mathcal{H}_0^\epsilon$ and $p_\epsilon^{\boldsymbol{\sigma}}$ from Section 3.4.1 and (3.30), respectively, and define $f_\epsilon : \mathbb{R}^d \to \mathbb{R}$ as

$$f_\epsilon(\boldsymbol{x}) \;=\; \begin{cases} p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) & \boldsymbol{x} \in \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \text{ for some } \boldsymbol{\sigma} \in \Sigma_0 \;, \\ \mathbf{1}_{\mathcal{H}_0^\epsilon}(\boldsymbol{x}) & \text{otherwise} \;. \end{cases}$$

The function f_ϵ is not continuous on $\mathcal{K}_\epsilon$ in general; instead, it is discontinuous along the boundaries $\partial_\pm \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ and $\partial \mathcal{K}_\epsilon$.

Remark 3.6.1. *It can be readily checked that the function f_ϵ is continuous on $\mathcal{K}_\epsilon$ if we consider the reversible case, i.e., $\boldsymbol{\ell} \equiv \mathbf{0}$.*

For convenience, we formally define $\nabla f_\epsilon(\boldsymbol{x})$ as

$$\nabla f_\epsilon(\boldsymbol{x}) \;=\; \begin{cases} \nabla p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) & \boldsymbol{x} \in \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \text{ for some } \boldsymbol{\sigma} \in \Sigma_0 \;, \\ 0 & \text{otherwise} \;. \end{cases} \tag{3.80}$$

Note that this is not a weak derivative of f_ϵ; hence, elementary theorems such as the divergence theorem cannot be applied to this gradient. With this formal gradient, we can define Φ_{f_ϵ} formally as

$$\Phi_{f_\epsilon}(\boldsymbol{x}) \;=\; \nabla f_\epsilon(\boldsymbol{x}) + \frac{1}{\epsilon}\, f_\epsilon(\boldsymbol{x})\, \boldsymbol{\ell}(\boldsymbol{x}) \;=\; \begin{cases} \epsilon^{-1}\, \boldsymbol{\ell}(\boldsymbol{x}) & \boldsymbol{x} \in \mathcal{H}_0^\epsilon \;, \\ \Phi_{p_\epsilon^{\boldsymbol{\sigma}}}(\boldsymbol{x}) & \boldsymbol{x} \in \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \text{ for some } \boldsymbol{\sigma} \in \Sigma_0 \;, \\ \mathbf{0} & \text{otherwise} \;. \end{cases}$$

Note that this is a formal definition, and Proposition 3.2.2 is not applicable to Φ_{f_ϵ}.

Now, we mollify the function f_ϵ as in [51] to get the genuine test function g_ϵ. To this end, consider a smooth, positive, and symmetric function $\phi : \mathbb{R}^d \to \mathbb{R}$ that is supported on the unit sphere of $\mathbb{R}^d$ and satisfies $\displaystyle\int_{\mathbb{R}^d} \phi(\boldsymbol{x})\, d\boldsymbol{x} = 1$.

Then, for $r > 0$, define $\phi_r(\boldsymbol{x}) = r^{-d}\phi(r^{-1}\boldsymbol{x})$. For the function $f : \mathbb{R}^d \to \mathbb{R}$ and vector field $\boldsymbol{V} : \mathbb{R}^d \to \mathbb{R}^d$, we write

$$f^{(r)} \,=\, f * \phi_r \quad \text{and} \quad \boldsymbol{V}^{(r)} \,=\, \boldsymbol{V} * \phi_r \ ,$$

where $*$ represents the usual convolution. In the remaining subsections, we prove the following two propositions. Hereafter, we write $\eta = \epsilon^2$. The first one asserts that we can approximate $\Phi_{f_\epsilon^{(\eta)}}$ by Φ_{f_ϵ}.

Proposition 3.6.2. *We have*

$$\epsilon \int_{\mathbb{R}^d} |\, \Phi_{f_\epsilon^{(\eta)}} - \Phi_{f_\epsilon} \,|^2 \, d\mu_\epsilon \,=\, o_\epsilon(1)\, \alpha_\epsilon \ .$$

Next, we prove the following estimate.

Proposition 3.6.3. *We have*

$$\epsilon \int_{\mathbb{R}^d} [\, \Phi_{f_\epsilon} \cdot \nabla h_\epsilon \,] \, d\mu_\epsilon \,=\, [\, 1 + o_\epsilon(1) \,]\, \alpha_\epsilon \, \omega_0 \ .$$

Before proving these propositions, we explain why Theorem 3.3.4 is a consequence of these propositions. We define the test function g_ϵ explicitly as

$$g_\epsilon \,=\, f_\epsilon^{(\eta)} \quad \text{where } \eta \,=\, \epsilon^2 \ . \tag{3.81}$$

Proof of Theorem 3.3.4. By Proposition 3.6.3, it suffices to prove that

$$\epsilon \int_{\mathbb{R}^d} [\, (\Phi_{g_\epsilon} - \Phi_{f_\epsilon}) \cdot \nabla h_\epsilon \,] \, d\mu_\epsilon \,=\, o_\epsilon(1) \, [\, \alpha_\epsilon \, \mathrm{cap}_\epsilon \,]^{1/2} \ .$$

With the selection (3.81), this is immediate from the Cauchy–Schwarz inequality, Lemma 3.2.1, and Proposition 3.6.2. □

In Sections 3.6.2 and 3.6.3, we shall prove Propositions 3.6.2 and 3.6.3, respectively. We remark that the proof of Proposition 3.6.2 is nearly model-independent and is similar to the proof of [51, Lemma 6.4]. Hence, we explain the structure of the proof and refer to [51] for most of the details. Of course, there are several differences in the proofs, and we present the full details for such parts.

3.6.2 Proof of Proposition 3.6.2

By the Cauchy–Schwarz inequality, we can write

$$\epsilon \int_{\mathbb{R}^d} |\, \Phi_{f_\epsilon^{(\eta)}} - \Phi_{f_\epsilon} \,|^2 \, d\mu_\epsilon \,\le\, 3\,(I_1 + I_2 + I_3)\;,$$

where

$$\begin{aligned} I_1 &= \epsilon \int_{\mathbb{R}^d} |\, \nabla(f_\epsilon^{(\eta)}) - (\nabla f_\epsilon)^{(\eta)} \,|^2 \, d\mu_\epsilon\;, \\ I_2 &= \epsilon \int_{\mathbb{R}^d} |\, (\nabla f_\epsilon)^{(\eta)} - \nabla f_\epsilon \,|^2 \, d\mu_\epsilon\;, \;\text{ and} \\ I_3 &= \frac{1}{\epsilon} \int_{\mathbb{R}^d} (f_\epsilon^{(\eta)} - f_\epsilon)^2 \, |\boldsymbol{\ell}|^2 \, d\mu_\epsilon\;. \end{aligned}$$

To conclude the proof of Proposition 3.6.2, it suffices to prove that $I_1,\ I_2,\ I_3 = o_\epsilon(1)\,\alpha_\epsilon$. The proofs of $I_1 = o_\epsilon(1)\,\alpha_\epsilon$ and $I_2 = o_\epsilon(1)\,\alpha_\epsilon$ are identical to those of [51, Lemma 8.5] and [51, Assertions 8.C and 8.D], respectively. The term I_3 has not been investigated previously. We present the proof of $I_3 = o_\epsilon(1)\,\alpha_\epsilon$. Note that the functions $f_\epsilon^{(\eta)}$ and f_ϵ are supported on $\mathcal{K}$ for sufficiently small $\epsilon > 0$, and since $|\boldsymbol{\ell}|$ is bounded on $\mathcal{K}$, it suffices to prove the following lemma.

Lemma 3.6.4. *We have*

$$\frac{1}{\epsilon} \int_{\mathbb{R}^d} (f_\epsilon^{(\eta)} - f_\epsilon)^2 \, d\mu_\epsilon \,=\, o_\epsilon(1)\,\alpha_\epsilon\;. \tag{3.82}$$

Proof. Recall the notation $\mathcal{A}^{[r]}$ from (3.69) and define

$$\begin{aligned}\widetilde{\mathcal{B}}_\epsilon^{\boldsymbol{\sigma}} &= \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \setminus (\partial \mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]} \ , \text{ and} \\ \widetilde{\mathcal{H}}_i^\epsilon &= \mathcal{H}_i^\epsilon \setminus \Big[(\partial \mathcal{K}_\epsilon)^{[\eta]} \cup \Big({}_{\boldsymbol{\sigma}\in\Sigma_0} (\partial \mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]} \Big) \Big] \quad ; \ i = 1,\, 2 \ .\end{aligned}$$

By the Cauchy–Schwarz inequality, we have

$$\begin{aligned}[\, (f_\epsilon^{(\eta)} - f_\epsilon)(\boldsymbol{x})\,]^2 &= \Big(\int_{\mathbb{R}^d} (\, f_\epsilon(\boldsymbol{x}) - f_\epsilon(\boldsymbol{x}-\boldsymbol{y})\,)\, \phi_\eta(\boldsymbol{y})\, d\boldsymbol{y} \Big)^2 \\ &\le \int_{\mathbb{R}^d} (\, f_\epsilon(\boldsymbol{x}) - f_\epsilon(\boldsymbol{x}-\boldsymbol{y})\,)^2\, \phi_\eta(\boldsymbol{y})\, d\boldsymbol{y} \ .\end{aligned}$$

Since

$$f_\epsilon(\boldsymbol{x}) \;=\; f_\epsilon(\boldsymbol{x}-\boldsymbol{y}) \quad \text{if } \boldsymbol{x} \notin \mathcal{K}_\epsilon^{[\eta]} \text{ and } |\boldsymbol{y}| \le \eta \ , \tag{3.83}$$

the left-hand side of (3.82) is bounded from above by

$$\int_{\mathcal{K}_\epsilon^{[\eta]}} \int_{\mathbb{R}^d} \frac{1}{\epsilon}\, |f_\epsilon(\boldsymbol{x}) - f_\epsilon(\boldsymbol{x}-\boldsymbol{y})|^2\, \phi_\eta(\boldsymbol{y})\, d\boldsymbol{y}\, \mu_\epsilon(d\boldsymbol{x}) \ .$$

Now, we divide the integral $\displaystyle\int_{\mathcal{K}_\epsilon^{[\eta]}}$ in the previous case into

$$\int_{\widetilde{\mathcal{H}}_0^\epsilon} + \int_{\widetilde{\mathcal{H}}_1^\epsilon} + \int_{(\partial\mathcal{K}_\epsilon)^{[\eta]}} + \sum_{\boldsymbol{\sigma}\in\Sigma_0} \int_{\widetilde{\mathcal{B}}_\epsilon^{\boldsymbol{\sigma}}} + \sum_{\boldsymbol{\sigma}\in\Sigma_0} \int_{(\partial\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]} \setminus (\partial\mathcal{K}_\epsilon)^{[\eta]}} \tag{3.84}$$

and consider the five integrals separately.

The first two integrals are 0 for the same reason with regard to (3.83). Now, we consider the third one. Since $|\, f_\epsilon(\boldsymbol{x}) - f_\epsilon(\boldsymbol{x}-\boldsymbol{y})\, | \le 1$ for all $\boldsymbol{x},\, \boldsymbol{y} \in \mathbb{R}^d$, the integral is bounded from above by

$$\int_{(\partial\mathcal{K}_\epsilon)^{[\eta]}} \int_{\mathbb{R}^d} \frac{1}{\epsilon}\, \phi_\eta(\boldsymbol{y})\, d\boldsymbol{y}\, \mu_\epsilon(d\boldsymbol{x}) \;=\; \frac{1}{\epsilon}\, \mu_\epsilon(\, (\partial\mathcal{K}_\epsilon)^{[\eta]}\,) \ . \tag{3.85}$$

Since $U(\boldsymbol{y}) = H + J^2\,\delta^2$ for $\boldsymbol{y} \in \partial\mathcal{K}_\epsilon$, there exists $C > 0$ such that

$$U(\boldsymbol{x}) \,\geq\, H + J^2\,\delta^2 - C\,\eta \qquad \text{for all } \boldsymbol{x} \in (\partial\mathcal{K}_\epsilon)^{[\eta]}\;.$$

Hence, the right-hand side of (3.85) is bounded by

$$\frac{C}{\epsilon\,Z_\epsilon}\,e^{-H/\epsilon}\int_{(\partial\mathcal{K}_\epsilon)^{[\eta]}} \epsilon^{J^2}\,e^{C\eta/\epsilon}\,d\boldsymbol{x} \,\leq\, C\,\epsilon^{J^2-d/2-1}\,\alpha_\epsilon\,\mathrm{vol}((\partial\mathcal{K}_\epsilon)^{[\eta]}) \,=\, o_\epsilon(1)\,\alpha_\epsilon$$

for sufficiently large J, since $\mathrm{vol}\,(\,(\partial\mathcal{K}_\epsilon)^{[\eta]}\,) = O(1)$.

Next, we consider the fourth term in (3.84). Fix $\boldsymbol{\sigma} \in \Sigma_0$ and assume, for simplicity of notation, that $\boldsymbol{\sigma} = \mathbf{0}$. By the mean value theorem, for $\boldsymbol{x} \in \widetilde{\mathcal{B}}_\epsilon^{\boldsymbol{\sigma}}$ and $\boldsymbol{y} \in \mathcal{D}_\eta(\mathbf{0})$,

$$|\,f_\epsilon(\boldsymbol{x}) - f_\epsilon(\boldsymbol{x}-\boldsymbol{y})\,| \,\leq\, |\boldsymbol{y}|\sum_{k=1}^{d}\sup_{\boldsymbol{z}\in\mathcal{D}_\eta(\boldsymbol{x})}|\,\nabla_k f_\epsilon(\boldsymbol{z})\,|\;. \tag{3.86}$$

First, we remark from the expression (3.80) that, for $\boldsymbol{u} \in \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$,

$$\nabla_k f_\epsilon(\boldsymbol{u}) \,=\, \frac{1}{c_\epsilon^{\boldsymbol{\sigma}}}\exp\Big\{-\frac{\mu}{2\epsilon}(\boldsymbol{u}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2\Big\}\,v_k\;. \tag{3.87}$$

Since $\eta \ll \delta$ and $|\boldsymbol{x}| = O(\delta)$, we have

$$(\boldsymbol{z}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2 \,\geq\, (\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2 - C\eta\delta \quad \text{for } \boldsymbol{x} \in \widetilde{\mathcal{B}}_\epsilon^{\boldsymbol{\sigma}} \text{ and } \boldsymbol{z} \in \mathcal{D}_\eta(\boldsymbol{x})\;. \tag{3.88}$$

By combining (3.87) and (3.88), we get

$$|\nabla_k f_\epsilon(\boldsymbol{z})|^2 \,\leq\, \frac{C}{\epsilon}\exp\Big\{-\frac{\mu}{\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2\Big\}\;.$$

Inserting this into (3.86), we obtain, for $\boldsymbol{x} \in \widetilde{\mathcal{B}}_\epsilon^{\boldsymbol{\sigma}}$,

$$\int_{\mathbb{R}^d}|\,f_\epsilon(\boldsymbol{x}) - f_\epsilon(\boldsymbol{x}-\boldsymbol{y})\,|^2\,\phi_\eta(\boldsymbol{y})d\boldsymbol{y} \,\leq\, \frac{C\eta^2}{\epsilon}\exp\Big\{-\frac{\mu}{\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2\Big\}\;.$$

Therefore, the integral in the fourth term of (3.84) is bounded by

$$\frac{1}{\epsilon Z_\epsilon}\frac{C\eta^2}{\epsilon}e^{-H/\epsilon}\int_{\widetilde{\mathcal{B}}_\epsilon^{\boldsymbol{\sigma}}}\exp\Big\{-\frac{1}{2\epsilon}\boldsymbol{x}\cdot(\mathbb{H}^{\boldsymbol{\sigma}}+2\mu\boldsymbol{v}^{\boldsymbol{\sigma}}\otimes\boldsymbol{v}^{\boldsymbol{\sigma}})\boldsymbol{x}\Big\}\,d\boldsymbol{x}$$

by the Taylor expansion of U around $\boldsymbol{\sigma}$. By Lemma 3.4.2, the last integral is $O(\epsilon^{d/2})$; hence, the whole expression is $o_\epsilon(1)\,\alpha_\epsilon$.

Now, we consider the last integral of (3.84). We also fix $\boldsymbol{\sigma}$ and assume that $\boldsymbol{\sigma}=\boldsymbol{0}$. Since

$$(\partial\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}\setminus(\partial\mathcal{K}_\epsilon)^{[\eta]}\subset(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}\cup(\partial_-\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}\,,$$

it suffices to prove that the integral over $(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}$ is small, as the argument for $(\partial_-\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}$ is identical. Since $\eta\ll\delta$, by Lemma 3.4.10, there exists a constant $a>0$ such that

$$U(\boldsymbol{x})\,\ge\,aJ^2\delta^2\quad\text{or}\quad\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}}\,\ge\,aJ\delta \tag{3.89}$$

holds for all $\boldsymbol{x}\in(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}$. Let us first assume that the former holds. Then, since $|f_\epsilon|\le 1$ and $\mathrm{vol}\big((\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}\big)=O(1)$, by the first condition of (3.89), the integral over $\boldsymbol{x}\in(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}$ satisfying the former condition of (3.89) is bounded from above by

$$\frac{C}{\epsilon Z_\epsilon}\int_{(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}}e^{-U(\boldsymbol{x})/\epsilon}\,d\boldsymbol{x}\,\le\,\frac{C}{\epsilon Z_\epsilon}e^{-H/\epsilon}\,\epsilon^{aJ^2}\,\mathrm{vol}\big((\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}\big)\,=\,o_\epsilon(1)\,\alpha_\epsilon \tag{3.90}$$

for sufficiently large J.

Now, assume that the second condition of (3.89) holds for $\boldsymbol{x}\in(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}$.

As in the proof of Lemma 3.4.6, we can rewrite $1 - f_\epsilon(\boldsymbol{x})$ as

$$[\,1 + o_\epsilon(1)\,]\,\frac{\epsilon^{1/2}}{(2\pi\mu)^{1/2}\,(\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})}\,\exp\Big\{-\frac{\mu}{2\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2\Big\}$$
$$\le \frac{C\,\epsilon^{1/2}}{\delta}\,\exp\Big\{-\frac{\mu}{2\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2\Big\}\,.$$

Similarly, we can check that, for $\boldsymbol{y}\in\mathcal{D}_\eta(\boldsymbol{0})$,

$$|\,1 - f_\epsilon(\boldsymbol{x}-\boldsymbol{y})\,| \le \frac{C\,\epsilon^{1/2}}{\delta}\,\exp\Big\{-\frac{\mu}{2\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2\Big\}\,.$$

By the two bounds above, we can bound $|f_\epsilon(\boldsymbol{x}) - f_\epsilon(\boldsymbol{x}-\boldsymbol{y})|^2$ from above by

$$2\,[\,|1-f_\epsilon(\boldsymbol{x})|^2 + |1-f_\epsilon(\boldsymbol{x}-\boldsymbol{y})|^2\,] \;\le\; \frac{C\,\epsilon}{\delta^2}\exp\Big\{-\frac{\mu}{\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2\Big\}\,.$$

Hence, we can bound the last integral of (3.84) and restrict it to $\boldsymbol{x}\in(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}$, satisfying the second condition of (3.89), from above by

$$\frac{C}{\delta^2}\int_{(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}}\exp\Big\{-\frac{\mu}{\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2\Big\}\,\mu_\epsilon(d\boldsymbol{x})\,.$$

By applying the Taylor expansion of U around $\boldsymbol{\sigma}$, this is bounded by

$$\frac{1}{\delta^2\,Z_\epsilon}\,e^{-H/\epsilon}\int_{(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}}\exp\Big\{-\frac{\mu}{2\epsilon}\boldsymbol{x}\cdot[\,\mathbb{H}^{\boldsymbol{\sigma}} + 2\mu\boldsymbol{v}^{\boldsymbol{\sigma}}\otimes\boldsymbol{v}^{\boldsymbol{\sigma}}\,]\,\boldsymbol{x}\Big\}\,d\boldsymbol{x}\,.$$

By Lemma 3.4.2, there exists $c>0$ such that $\boldsymbol{x}\cdot[\,\mathbb{H}^{\boldsymbol{\sigma}} + 2\mu\,\boldsymbol{v}^{\boldsymbol{\sigma}}\otimes\boldsymbol{v}^{\boldsymbol{\sigma}}\,]\boldsymbol{x}\ge c\,|\boldsymbol{x}|^2$. Furthermore, there exists $C>0$ such that $|\boldsymbol{x}|\ge C\delta$ for all $\boldsymbol{x}\in(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}$. Therefore, we can bound the last centered display from above by

$$\frac{1}{Z_\epsilon}\,e^{-H/\epsilon}\,\epsilon^{cJ^2}\,\mathrm{vol}((\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}) \;=\; o_\epsilon(1)\,\alpha_\epsilon \tag{3.91}$$

for sufficiently large J since $\mathrm{vol}((\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}})^{[\eta]}) = O(1)$. By (3.90) and (3.91), we can verify that the last integral of (3.84) is $o_\epsilon(1)\,\alpha_\epsilon$, and this completes the proof. □

3.6.3 Proof of Proposition 3.6.3

First, note that we can write

$$\epsilon \int_{\mathbb{R}^d} [\, \Phi_{f_\epsilon} \cdot \nabla h_\epsilon \,] \, d\mu_\epsilon \;=\; A_1 + \sum_{\boldsymbol{\sigma} \in \Sigma_0} A_2(\boldsymbol{\sigma}) \;, \tag{3.92}$$

where

$$A_1 \;=\; \int_{\mathcal{H}_0^\epsilon} [\, \boldsymbol{\ell} \cdot \nabla h_\epsilon \,] \, d\mu_\epsilon \quad \text{and} \quad A_2(\boldsymbol{\sigma}) \;=\; \epsilon \int_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} [\, \Phi_{p_\epsilon^{\boldsymbol{\sigma}}} \cdot \nabla h_\epsilon \,] \, d\mu_\epsilon \;.$$

To estimate these integrals, we first mention a technical result.

Lemma 3.6.5. *There exists $C > 0$ such that*

$$\int_{\partial \mathcal{K}^\epsilon} \sigma(d\mu_\epsilon) \;\le\; C \, \epsilon^{J^2 - d/2} \, \alpha_\epsilon \;.$$

Proof. Since $U(\boldsymbol{x}) = H + J^2 \delta^2$ on $\partial \mathcal{K}^\epsilon$, we have

$$\int_{\partial \mathcal{K}^\epsilon} \sigma(d\mu_\epsilon) \;=\; \int_{\partial \mathcal{K}^\epsilon} \mu_\epsilon(\boldsymbol{x}) \, \sigma(d\boldsymbol{x}) \;=\; Z_\epsilon^{-1} \, e^{-H/\epsilon} \, \epsilon^{J^2} \, \sigma(\partial \mathcal{K}^\epsilon) \;.$$

Since $\sigma(\partial \mathcal{K}^\epsilon) = O(1)$, the proof is completed by the definition (3.18) of α_ϵ. □

We now consider A_1.

Lemma 3.6.6. *We can write*

$$A_1 \;=\; o_\epsilon(1) \, \alpha_\epsilon + \sum_{\boldsymbol{\sigma} \in \Sigma_0} A_{1,1}(\boldsymbol{\sigma}) \;,$$

where

$$A_{1,1}(\boldsymbol{\sigma}) \;=\; \int_{\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} [\, \boldsymbol{\ell} \cdot \boldsymbol{n}_{\mathcal{H}_0^\epsilon} \,] \, h_\epsilon \, \sigma(d\mu_\epsilon) \;. \tag{3.93}$$

Proof. By the divergence theorem, we have

$$\int_{\mathcal{H}_0^\epsilon} [\,\boldsymbol{\ell}\cdot\nabla h_\epsilon\,]\, d\mu_\epsilon \;=\; \int_{\partial\mathcal{H}_0^\epsilon} [\,\boldsymbol{\ell}\cdot\boldsymbol{n}_{\mathcal{H}_0^\epsilon}\,]\, h_\epsilon\, \sigma(d\mu_\epsilon)\;.$$

Write

$$\partial\widehat{\mathcal{H}}_0^\epsilon \;=\; \partial\mathcal{H}_0^\epsilon \setminus \Big[\bigcup_{\boldsymbol{\sigma}\in\Sigma_0} \partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}\Big] \;\subset\; \partial\mathcal{K}_\epsilon\;.$$

Then, it suffices to prove that

$$\int_{\partial\widehat{\mathcal{H}}_0^\epsilon} [\,\boldsymbol{\ell}\cdot\boldsymbol{n}_{\mathcal{H}_0^\epsilon}\,]\, h_\epsilon\, \sigma(d\mu_\epsilon) \;=\; o_\epsilon(1)\,\alpha_\epsilon\;.$$

Since $|h_\epsilon|$ and $|\boldsymbol{\ell}|$ are bounded on $\partial\widehat{\mathcal{H}}_0^\epsilon \subset \mathcal{K}$, and since $\partial\widehat{\mathcal{H}}_0^\epsilon \subset \partial\mathcal{K}_\epsilon$, the absolute value of the left-hand side of the previous case is bounded by $\int_{\partial\mathcal{K}_\epsilon} \sigma(d\mu_\epsilon)$, which is $o_\epsilon(1)\,\alpha_\epsilon$ for sufficiently large J by Lemma 3.6.5. This completes the proof. □

Now, we focus on $A_2(\boldsymbol{\sigma})$.

Lemma 3.6.7. *For $\boldsymbol{\sigma}\in\Sigma_0$, we can write*

$$A_2(\boldsymbol{\sigma}) \;=\; o_\epsilon(1)\,\alpha_\epsilon + A_{2,1}(\boldsymbol{\sigma})\;,$$

where

$$A_{2,1}(\boldsymbol{\sigma}) \;=\; \epsilon\int_{\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}\cup\partial_-\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} [\,\Phi_{p_\epsilon^{\boldsymbol{\sigma}}}\cdot\boldsymbol{n}_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}}\,]\, h_\epsilon\, \sigma(d\mu_\epsilon)\;. \tag{3.94}$$

Proof. By the divergence theorem, we can write

$$A_2(\boldsymbol{\sigma}) \;=\; -\int_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} (\,\mathscr{L}_\epsilon^*\, p_\epsilon^{\boldsymbol{\sigma}}\,)\, h_\epsilon\, d\mu_\epsilon + \epsilon\int_{\partial\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} [\,\Phi_{p_\epsilon^{\boldsymbol{\sigma}}}\cdot\boldsymbol{n}_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}}\,]\, h_\epsilon\, \sigma(d\mu_\epsilon)\;.$$

By Proposition 3.4.5, the first integral on the right-hand side is $o_\epsilon(1)\,\alpha_\epsilon$.

Hence, it suffices to prove that

$$\epsilon \int_{\partial_0 \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} [\, \Phi_{p_\epsilon^{\boldsymbol{\sigma}}} \cdot \boldsymbol{n}_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \,] \, h_\epsilon \, \sigma(d\mu_\epsilon) \;=\; o_\epsilon(1) \, \alpha_\epsilon \; . \tag{3.95}$$

By the explicit formula for $p_\epsilon^{\boldsymbol{\sigma}}$ and by the boundedness of $\boldsymbol{\ell}$ on $\mathcal{K}$, we can check that there exists $C > 0$ such that $|\, \Phi_{p_\epsilon^{\boldsymbol{\sigma}}} \,| \le C\epsilon^{-1}$ on $\partial_0 \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$. Therefore, the absolute value of the left-hand side of (3.95) is bounded from above by $C \int_{\partial_0 \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \sigma(d\mu_\epsilon)$. Since $\partial_0 \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \subset \partial \mathcal{K}_\epsilon$, the proof is completed by Lemma 3.6.5, provided that we take J to be sufficiently large. □

By (3.92) and Lemmas 3.6.6 and 3.6.7, it suffices to check the following Lemma to complete the proof of Proposition 3.6.3.

Lemma 3.6.8. *For $\boldsymbol{\sigma} \in \Sigma_0$, we have*

$$A_{1,1}(\boldsymbol{\sigma}) + A_{2,1}(\boldsymbol{\sigma}) \;=\; [\, 1 + o_\epsilon(1) \,] \, \alpha_\epsilon \, \omega^{\boldsymbol{\sigma}} \; .$$

We defer the proof of Lemma 3.6.8 to the next subsection and conclude the proof of Proposition 3.6.3 first.

Proof of Proposition 3.6.3. The proof is completed by combining 3.92 and Lemmas 3.6.6, 3.6.7, and 3.6.8. □

3.6.4 Proof of Lemma 3.6.8

As a consequence of Proposition 3.5.1, we can get the following estimate of the equilibrium potential at the boundaries $\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ and $\partial_- \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ for $\boldsymbol{\sigma} \in \Sigma_0$.

Lemma 3.6.9. *There exists a constant $C > 0$ such that, for all $\boldsymbol{\sigma} \in \Sigma_0$,*

$$h_\epsilon(\boldsymbol{x}) \;\ge\; 1 - C\,\epsilon^{-d} \exp \frac{U(\boldsymbol{x}) - H}{2\epsilon} \quad \forall \boldsymbol{x} \in \partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \quad \text{and}$$
$$h_\epsilon(\boldsymbol{x}) \;\le\; C\,\epsilon^{-d} \exp \frac{U(\boldsymbol{x}) - H}{2\epsilon} \quad \forall \boldsymbol{x} \in \partial_- \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \; .$$

Proof. Let us consider the first inequality. If $\boldsymbol{x} \in \partial_+ \mathcal{B}_\epsilon$ satisfies $U(\boldsymbol{x}) \geq H$, then the inequality is obvious for all sufficiently small ϵ. Otherwise, $\boldsymbol{x} \in \mathcal{H}_0$; hence, the bound follows from part (1) of Proposition 3.5.1 since we have $\mathfrak{H}_{\{\boldsymbol{x}\},\,\mathcal{D}_\epsilon(\boldsymbol{m}_0)} = U(\boldsymbol{x})$ for all sufficiently small ϵ. The proof of the second one is similar and left to the reader. $\square$

In the next lemma, we provide a consequence of the previous lemma.

Lemma 3.6.10. *For $\boldsymbol{\sigma} \in \Sigma_0$, we have*

$$\epsilon \int_{\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} |\, \nabla p_\epsilon^{\boldsymbol{\sigma}} \,| \, (\, 1 - h_\epsilon \,) \, \sigma(d\mu_\epsilon) \;=\; o_\epsilon(1) \, \alpha_\epsilon \;, \tag{3.96}$$

$$\int_{\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} (\, 1 - p_\epsilon^{\boldsymbol{\sigma}} \,) \, (\, 1 - h_\epsilon \,) \, \sigma(d\mu_\epsilon) \;=\; o_\epsilon(1) \, \alpha_\epsilon \;, \tag{3.97}$$

$$\epsilon \int_{\partial_- \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} |\, \nabla p_\epsilon^{\boldsymbol{\sigma}} \,| \, h_\epsilon \, \sigma(d\mu_\epsilon) \;=\; o_\epsilon(1) \, \alpha_\epsilon \;, \tag{3.98}$$

$$\int_{\partial_- \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} p_\epsilon^{\boldsymbol{\sigma}} \, h_\epsilon \, \sigma(d\mu_\epsilon) \;=\; o_\epsilon(1) \, \alpha_\epsilon \;. \tag{3.99}$$

Proof. Since the proofs of (3.98) and (3.99) are identical to those of (3.96) and (3.97), respectively, we focus only on (3.96) and (3.97).

Let us first consider (3.96). We use the explicit formula for $p_\epsilon^{\boldsymbol{\sigma}}$ and Lemma 3.6.9 to bound the left-hand side of (3.96) by

$$C \, \epsilon^{-1/2 - 3d/2} \, \alpha_\epsilon \int_{\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \exp \Big\{ - \frac{U(\boldsymbol{x}) - H}{2\epsilon} - \frac{\mu}{2\epsilon} (\boldsymbol{x} \cdot \boldsymbol{v}^{\boldsymbol{\sigma}})^2 \Big\} \, \sigma(d\boldsymbol{x}) \;. \tag{3.100}$$

By the Taylor expansion, the last line can be further bounded by

$$\begin{aligned} & C \, \epsilon^{-1/2 - 3d/2} \, \alpha_\epsilon \int_{\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \exp \Big\{ - \frac{1}{4\epsilon} \boldsymbol{x} \cdot [\mathbb{H}^{\boldsymbol{\sigma}} + 2\mu \boldsymbol{v}^{\boldsymbol{\sigma}} \otimes \boldsymbol{v}^{\boldsymbol{\sigma}}] \boldsymbol{x} \Big\} \, \sigma(d\boldsymbol{x}) \\ & \leq \; C \, \epsilon^{-1/2 - 3d/2} \, \alpha_\epsilon \int_{\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \exp \Big\{ - \frac{\gamma}{4\epsilon} |\boldsymbol{x}|^2 \Big\} \, \sigma(d\boldsymbol{x}) \;, \end{aligned} \tag{3.101}$$

where $\gamma > 0$ is the smallest eigenvalue of the positive-definite matrix $\mathbb{H} +$

$2\mu\boldsymbol{v} \otimes \boldsymbol{v}$ (cf. Lemma 3.4.2). Since there exists $C > 0$ such that $|\boldsymbol{x}| \ge CJ\delta$ for all $\boldsymbol{x} \in \partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, and since $\sigma(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}) = O(\delta^{d-1})$, we can bound (3.101) from above, for some c, $C > 0$, by

$$C\,\epsilon^{-1/2-3d/2}\,\delta^{d-1}\,\epsilon^{cJ^2}\,\alpha_\epsilon \;=\; C\,\Big(\log\frac{1}{\epsilon}\Big)^{\frac{d-1}{2}}\,\epsilon^{cJ^2-d-1} \;=\; o_\epsilon(1)\,\alpha_\epsilon$$

for sufficiently large J. This completes the proof of (3.96).

For (3.97), recall $\partial_+^{1,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ and $\partial_+^{2,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ from (3.41) and (3.42), respectively. By Lemma 3.4.10, it suffices to prove that, for $a \in (0,\, a_0)$,

$$\int_{\partial_+^{k,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} (\,1 - p_\epsilon^{\boldsymbol{\sigma}}\,)\,(\,1 - h_\epsilon\,)\,\sigma(d\mu_\epsilon) \;=\; o_\epsilon(1)\,\alpha_\epsilon \quad ;\; k = 1,\, 2\;. \tag{3.102}$$

For $k = 1$, by (3.49) and Lemma 3.6.9, we can bound the integral from above by

$$\frac{C\,e^{-H}\,\epsilon^{1/2}}{Z_\epsilon\,\epsilon^d\,\delta}\int_{\partial_+\mathcal{B}_\epsilon} \exp\Big\{\,-\,\frac{U(\boldsymbol{x})-H}{2\epsilon} - \frac{\mu}{2\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2\,\Big\}\,\sigma(d\boldsymbol{x})\;.$$

Hence, we can proceed as in the computation of (3.100) to prove that this is $o_\epsilon(1)\,\alpha_\epsilon$.

Now, we finally consider the $k = 2$ case of (3.102). Since $U(\boldsymbol{x}) \ge H + aJ^2\delta^2$ for $\boldsymbol{x} \in \partial_+^{2,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, the left-hand side of (3.102) with $k = 2$ is bounded from above by

$$\frac{1}{Z_\epsilon}\,e^{-H/\epsilon}\,\epsilon^{aJ^2}\,\sigma(\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}) \;\le\; \frac{C}{Z_\epsilon}\,e^{-H/\epsilon}\,\epsilon^{aJ^2}\,\delta^{d-1} \;=\; o_\epsilon(1)\,\alpha_\epsilon$$

for sufficiently large J. This completes the proof. □

Now, we are ready to prove Lemma 3.6.8.

Proof of Lemma 3.6.8. In view of the expressions (3.93) and (3.94) for $A_{1,1}(\boldsymbol{\sigma})$

and $A_{2,1}(\boldsymbol{\sigma})$, respectively, it suffices to prove the following estimates:

$$\epsilon \int_{\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \Big[\Big(\Phi - \frac{1}{\epsilon} \boldsymbol{\ell} \Big) \cdot \boldsymbol{n}_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \Big] \, h_\epsilon \, \sigma(d\mu_\epsilon) \; = \; [\, 1 + o_\epsilon(1) \,] \, \alpha_\epsilon \, \omega^{\boldsymbol{\sigma}} \; , \tag{3.103}$$

$$\epsilon \int_{\partial_- \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} [\, \Phi_{p_\epsilon^{\boldsymbol{\sigma}}} \cdot \boldsymbol{n}_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \,] \, h_\epsilon \, \sigma(d\mu_\epsilon) \; = \; o_\epsilon(1) \, \alpha_\epsilon \; . \tag{3.104}$$

Let us first consider (3.103). By (3.96) and (3.97) of Lemma 3.6.10, we can replace the $h_\epsilon(\boldsymbol{x})$ term with 1 with an error term of order $o_\epsilon(1)\, \alpha_\epsilon$. Then, we can apply Proposition 3.4.6 to prove (3.103). On the other hand, the estimate (3.104) is a direct consequence of (3.98) and (3.99) of Lemma 3.6.10. □

Chapter 4

Markov chain model reduction

This chapter is devoted to prove the Markov chain model reduction for the process $\boldsymbol{x}_\epsilon(\cdot)$ (Theorem 4.1.5). The proof of the Markov chain model reduction for the reversible process $\boldsymbol{y}_\epsilon(\cdot)$ in [83] is based on the Poisson equation approach. In this chapter, we extend this result to the non-reversible dynamics by considering resolvent equation instead of Poisson equation.

Remark. *Sets and constants including set of saddle points Σ and Eyring–Kramers constants $\omega^{\boldsymbol{\sigma}}$ are already defined in the previous chapter. However, our interest in this chapter is a different perspective of metastable behavior so that we need different saddle structure and metastable valleys (cf. Σ^* and $\mathcal{V}_i$ defined below). Hence, in spite of their similarities, for the completeness of the current chapter, we recall their definitions.*

4.1 Main result

In this section, we explain our main result regarding the Markov chain description of the metastable behavior of the process $\boldsymbol{x}_\epsilon(\cdot)$ when U has several local minima.

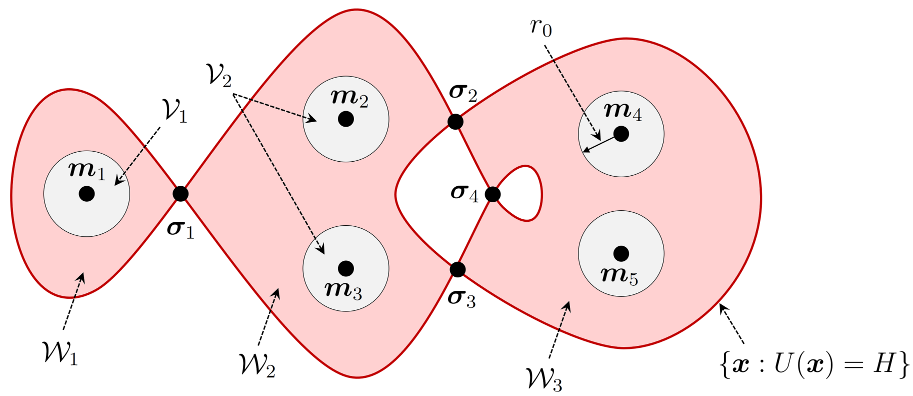


Figure 4.1: Example of landscape of U. In this example, we have $\Sigma = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3, \boldsymbol{\sigma}_4\}$ and the set $\{\boldsymbol{x} : U(\boldsymbol{x}) < H\}$ consists of three components $\mathcal{W}_1, \mathcal{W}_2, \mathcal{W}_3$. Hence, $S = \{1, 2, 3\}$. We have $\Sigma_{1,2} = \{\boldsymbol{\sigma}_1\}$, $\Sigma_{2,3} = \{\boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3\}$, and $\Sigma_{1,3} = \emptyset$. Therefore, $\Sigma^* = \{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_3\} \subsetneq \Sigma$. Suppose that $h_1 = h_2 = h < h_3$. Then, we have $S_\star = \{1, 2\}$. By assuming that $U(\boldsymbol{m}_2) = U(\boldsymbol{m}_3) = h$, two metastable valleys are defined by $\mathcal{V}_1 = \overline{\mathcal{D}_{r_0}(\boldsymbol{m}_1)}$ and $\mathcal{V}_2 = \overline{\mathcal{D}_{r_0}(\boldsymbol{m}_2)} \cup \overline{\mathcal{D}_{r_0}(\boldsymbol{m}_3)}$. Metastable valley is not defined for the shallow well $\mathcal{W}_3$.

4.1.1 Landscape of U and invariant measure

We first analyze the landscape of U. We refer to Figure 4.1 for an illustration of the notations introduced in this subsection.

For a concrete description, we fix a level H and define $\Sigma = \Sigma^H$ as the set of saddle points of level H:

$$\Sigma := \{\boldsymbol{\sigma} : U(\boldsymbol{\sigma}) = H \text{ and } \boldsymbol{\sigma} \text{ is a saddle point of } U\} .$$

By selecting H appropriately, we shall assume that Σ is a non-empty set. We now define

$$\mathcal{H} := \{\boldsymbol{x} \in \mathbb{R}^d : U(\boldsymbol{x}) < H\} ,$$

and denote by $\mathcal{W}_1, \ldots, \mathcal{W}_K$ the connected components of the set $\mathcal{H}$. These sets are called (metastable) wells for the potential function U corresponding to the level H. We focus on the transition of the process $\boldsymbol{x}_\epsilon(\cdot)$ among these wells. Various selections of H are possible; however, we focus on one fixed level to get a concrete result. The last paragraph of the current section explain how we can select various H to get a variety of results that provide a full description of the metastable behavior.

If $K = 1$, there is no interesting metastable behavior at level H, and we must take a smaller level to observe the metastable behavior. Therefore, we assume that $K \geq 2$. Now, we shall assume that the closure $\overline{\mathcal{H}}$ of $\mathcal{H}$ is a connected set. Otherwise, our analysis can be applied to each connected component of $\overline{\mathcal{H}}$, and this general situation is explained later. See the discussion after Theorem 4.1.5.

Write $S = \{1, \cdots, K\}$. For $i,\, j \in S$,[1] we write

$$\Sigma_{i,\,j} \;=\; \overline{\mathcal{W}}_i \cap \overline{\mathcal{W}}_j \;,$$

which denotes the set of saddle points between $\mathcal{W}_i$ and $\mathcal{W}_j$ of level H. Note that this set can be empty. Now, assume further that $\Sigma_{i,\,j} \cap \Sigma_{k,\,l} = \emptyset$ unless $\{i,\, j\} = \{k,\, l\}$; hence, there is no saddle point connecting three or more wells simultaneously. Write

$$\Sigma^* \;=\; \bigcup_{i,\,j \in S} \Sigma_{i,\,j} \;. \tag{4.1}$$

Then, we have $\Sigma^* \subseteq \Sigma$, and the equality may not hold (cf. Figure 4.1). By the Morse lemma, for each $\boldsymbol{\sigma} \in \Sigma^*$, the Hessian $(\nabla^2 U)(\boldsymbol{\sigma})$ has only one negative eigenvalue and $(d-1)$ positive eigenvalues, as we have assumed that U is a Morse function. We remark that this may not be true for $\boldsymbol{\sigma} \in \Sigma \setminus \Sigma^*$.

[1] In this thesis, writing "$a,\, b \in S$" implies that $a \in S$, $b \in S$, and $a \neq b$.

Metastable valleys

Now, we define the metastable valleys. We fix $i \in S$ and denote by h_i the minimum value of the potential U on the well $\mathcal{W}_i$, i.e.,

$$h_i := \min\{U(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{W}_i\} . \tag{4.2}$$

Define $\mathcal{M}_i$ as the set of the deepest minima of U on $\mathcal{W}_i$:

$$\mathcal{M}_i := \{\boldsymbol{m} \in \mathcal{W}_i : U(\boldsymbol{m}) = h_i\} .$$

Then, we can regard $H - h_i$ as the depth of the well $\mathcal{W}_i$. We write the ball in $\mathbb{R}^d$ centered at $\boldsymbol{x}$ with radius r as

$$\mathcal{D}_r(\boldsymbol{x}) := \{\boldsymbol{y} \in \mathbb{R}^d : |\boldsymbol{y} - \boldsymbol{x}| < r\} .$$

We take $r_0 > 0$ to be sufficiently small so that, for all $i \in S$ and for all $\boldsymbol{m} \in \mathcal{M}_i$,

$$\mathcal{D}_{2r_0}(\boldsymbol{m}) \subset \mathcal{W}_i \text{ and } \overline{\mathcal{D}_{2r_0}(\boldsymbol{m})} \setminus \{\boldsymbol{m}\} \text{ does not contain a critical point of } U. \tag{4.3}$$

Finally, the *metastable valley* corresponding to the well $\mathcal{W}_i$ is defined as

$$\mathcal{V}_i := \bigcup_{\boldsymbol{m} \in \mathcal{M}_i} \overline{\mathcal{D}_{r_0}\,(\boldsymbol{m})} , \tag{4.4}$$

where $\overline{\mathcal{D}_{r_0}\,(\boldsymbol{m})} = \{\boldsymbol{y} \in \mathbb{R}^d : |\boldsymbol{y} - \boldsymbol{x}| \le r_0\}$ denotes the closed ball. Our primary focus is the inter-valley dynamics among these sets $\mathcal{V}_i$.

Deepest valleys

We now characterize the deepest valleys of U, which will be the state space of the limiting Markov chain describing the metastable behavior. Recall h_i

from (4.2) and define

$$h := \min_{i\in S} h_i \quad \text{and} \quad S_\star := \{i \in S : h_i = h\} ,$$

so that $\{\mathcal{W}_i : i \in S_\star\}$ denotes the collection of the deepest wells. We assume that $|S_\star| \geq 2$ since the Markov chain description is trivial when $|S_\star| = 1$. Let

$$\mathcal{M}_\star := \bigcup_{i\in S_\star} \mathcal{M}_i = \{\boldsymbol{x} \in \mathbb{R}^d : U(\boldsymbol{x}) = h\} ,$$

so that the set $\mathcal{M}_\star$ denotes the set of global minima of U. Write $\mathcal{V}_\star = \bigcup_{i\in S_\star} \mathcal{V}_i$ so that $\mathcal{V}_\star$ denotes the set of deepest valleys. Finally, we write $\Delta = \mathbb{R}^d \setminus \mathcal{V}_\star$.

Invariant measure

With the construction of the metastable valleys, we can conclude that the invariant measure $\mu_\epsilon(d\boldsymbol{x})$ is concentrated on the set $\mathcal{V}_\star$. Moreover, we can compute the precise asymptotics for $\mu_\epsilon(\mathcal{V}_i)$ for each $i \in S_\star$. To this end, we recall Notation 2.2.5.

Notation 4.1.1. *For each $\boldsymbol{x} \in \mathbb{R}^d$, we write $\mathbb{H}^{\boldsymbol{x}} = (\nabla^2 U)(\boldsymbol{x})$ as the Hessian of U at $\boldsymbol{x}$ and $\mathbb{L}^{\boldsymbol{x}} = D\boldsymbol{\ell}(\boldsymbol{x})$ as the Jacobian of $\boldsymbol{\ell}$ at $\boldsymbol{x}$.*

For each $i \in S$, we define

$$\nu_i := \sum_{\boldsymbol{m}\in\mathcal{M}_i} \frac{1}{\sqrt{\det \mathbb{H}^{\boldsymbol{m}}}} ,$$

and write $\nu_\star = \sum_{i\in S_\star} \nu_i$. For a sequence $(a_\epsilon)_{\epsilon>0}$ of real numbers, we write $a_\epsilon = o_\epsilon(1)$ if $\lim_{\epsilon\to 0} a_\epsilon = 0$. The following asymptotics are useful in our discussion.

Proposition 4.1.2. *We have*

$$Z_\epsilon = [\,1+o_\epsilon(1)\,]\,(2\pi\epsilon)^{d/2}\,e^{-h/\epsilon}\,\nu_\star\;, \tag{4.5}$$
$$\mu_\epsilon(\mathcal{V}_i) = [\,1+o_\epsilon(1)\,]\,\frac{\nu_i}{\nu_\star}\quad ;\, i\in S_\star \quad \textit{and} \quad \mu_\epsilon(\Delta)\,=\,o_\epsilon(1)\;.$$

Proof. The proof is a consequence of an elementary computation based on the Laplace asymptotics. For further detail, we refer to [83, Proposition 2.2]. □

Eyring–Kramers constants

For $\boldsymbol{\sigma}\in\Sigma^*$, we previously mentioned that the Hessian $\mathbb{H}^{\boldsymbol{\sigma}}$ has only one negative eigenvalue by the Morse lemma. Further, the matrix $\mathbb{H}^{\boldsymbol{\sigma}}+\mathbb{L}^{\boldsymbol{\sigma}}$ also has only one negative eigenvalue by Lemma 2.2.7. Denote by $-\mu^{\boldsymbol{\sigma}}$ the unique negative eigenvalue of $\mathbb{H}^{\boldsymbol{\sigma}}+\mathbb{L}^{\boldsymbol{\sigma}}$. Recall the *Eyring–Kramers constant* at $\boldsymbol{\sigma}\in\Sigma^*$ defined by

$$\omega^{\boldsymbol{\sigma}}\;:=\;\frac{\mu^{\boldsymbol{\sigma}}}{2\pi\sqrt{-\det\mathbb{H}^{\boldsymbol{\sigma}}}}\;. \tag{4.6}$$

For i, $j\in S$, we define

$$\omega_{i,\,j}\;:=\;\sum_{\boldsymbol{\sigma}\in\Sigma_{i,\,j}}\omega^{\boldsymbol{\sigma}}\quad\text{and}\quad\omega_i\;:=\;\sum_{j\in S}\omega_{i,\,j}\;,$$

where we set $\omega_{i,\,i}=0$ for $i\in S$ for convenience of notation. Note that the connectedness of $\overline{\mathcal{H}}$ implies that $\omega_i>0$ for all $i\in S$.

4.1.2 Two Markov chains

Now, we construct two continuous-time Markov chains: $(\mathbf{x}(t))_{t\ge0}$ and $(\mathbf{y}(t))_{t\ge0}$. The Markov chain $\mathbf{y}(\cdot)$ describes the limiting metastable behavior of the diffusion process $\boldsymbol{x}_\epsilon(\cdot)$. The auxiliary Markov chain $\mathbf{x}(\cdot)$ is used in the construction of this limiting chain $\mathbf{y}(\cdot)$; moreover, it plays a crucial role in the proof. We refer to Remark 4.1.3 for the meaning of these Markov chains.

The construction of the limiting chain $\mathbf{y}(\cdot)$ is simple when all the wells have the same depth, i.e., $S_\star = S$ (cf. Remark 4.1.3). However, if $S_\star \subsetneq S$, the behavior of the process $\boldsymbol{x}_\epsilon(\cdot)$ on each shallow valley $\mathcal{V}_i$, $i \in S \setminus S_\star$, should be properly reflected in the construction; hence, the definition of $\mathbf{y}(\cdot)$ becomes more complex and should be done via the auxiliary chain $\mathbf{x}(\cdot)$ defined from now on.

Auxiliary Markov chain $\mathbf{x}(\cdot)$ on S

We define a probability measure $m(\cdot)$ on S by

$$m(i) := \omega_i / \sum_{j\in S} \omega_j \quad ; \; i \in S \; .$$

Let $(\mathbf{x}(t))_{t\ge 0}$ be the continuous-time Markov chain on S whose jump rate from $i \in S$ to $j \in S$ is given by $r_{\mathbf{x}}(i,\, j) = \omega_{i,\,j}/m(i)$. It is clear that the invariant measure for the Markov chain $\mathbf{x}(\cdot)$ is $m(\cdot)$, and moreover the process $\mathbf{x}(\cdot)$ is reversible with respect to $m(\cdot)$. We now introduce several potential theoretic notions regarding the process $\mathbf{x}(\cdot)$. These notions are used in the definition of the limiting Markov chain $\mathbf{y}(\cdot)$.

Denote by $L_{\mathbf{x}}$ the generator associated with the Markov chain $\mathbf{x}(\cdot)$ acting on $\mathbf{f} : S \to \mathbb{R}$ such that

$$(L_{\mathbf{x}}\mathbf{f})(i) = \sum_{j\in S} r_{\mathbf{x}}(i,\, j)\, [\, \mathbf{f}(j) - \mathbf{f}(i)\,] \quad ; \; i \in S \; .$$

Denote by $\mathbf{P}_i$ the law of process $\mathbf{x}(\cdot)$ starting at $i \in S$. For two disjoint non-empty subsets A, B of S, the equilibrium potential between A and B with respect to the process $\mathbf{x}(\cdot)$ is a function $\mathbf{h}_{A,\,B} : S \to \mathbb{R}$ defined by

$$\mathbf{h}_{A,\,B}(i) := \mathbf{P}_i\, [\, \tau_A < \tau_B\,] \quad ; \; i \in S \; ,$$

where τ_A, $A \subset S$, denotes the hitting time of the set A, i.e., $\tau_A = \inf\{t \ge 0 \, :$

$\mathbf{x}(t) \in A\}$. Define a bi-linear form $D_{\mathbf{x}}(\cdot, \cdot)$ by, for all $\mathbf{f}$, $\mathbf{g} : S \to \mathbb{R}$,

$$D_{\mathbf{x}}(\mathbf{f},\, \mathbf{g}) \;:=\; \sum_{i\in S} \mu(i)\, \mathbf{f}(i)\, [\, -(L_{\mathbf{x}}\mathbf{g})(i)\,] = \frac{1}{2} \sum_{i,\, j\in S} \omega_{i,\, j}\, [\, \mathbf{f}(i) - \mathbf{f}(j)\,]\, [\, \mathbf{g}(i) - \mathbf{g}(j)\,]\; . \tag{4.7}$$

Note that $D_{\mathbf{x}}(\mathbf{f},\, \mathbf{f})$ represents the Dirichlet form associated with the Markov chain $\mathbf{x}(\cdot)$. Finally, the capacity between two disjoint non-empty subsets A and B of S with respect to the process $\mathbf{x}(\cdot)$ is defined by

$$\mathrm{cap}_{\mathbf{x}}(A,\, B) \;:=\; D_{\mathbf{x}}(\mathbf{h}_{A,\, B},\, \mathbf{h}_{A,\, B})\; . \tag{4.8}$$

Limiting Markov chain $\mathbf{y}(\cdot)$ on $S_\star$

Recall that we assumed $|S_\star| \geq 2$. For $i,\, j \in S_\star$, define

$$\beta_{i,\, j} \;:=\; \frac{1}{2}[\mathrm{cap}_{\mathbf{x}}(\{i\},\, S_\star \setminus \{i\}) + \mathrm{cap}_{\mathbf{x}}(\{j\},\, S_\star \setminus \{j\}) - \mathrm{cap}_{\mathbf{x}}(\{i, j\},\, S_\star \setminus \{i, j\})]\; .$$

We set $\beta_{i,\, i} = 0$, $i \in S_\star$, for convenience and note that we have $\beta_{i,\, j} = \beta_{j,\, i}$ for all $i,\, j \in S_\star$. Then, we define $(\mathbf{y}(t))_{t\geq 0}$ as a continuous-time Markov chain on $S_\star$ with jump rate $r_{\mathbf{y}}(i,\, j)$ from $i \in S_\star$ to $j \in S_\star$ given by $r_{\mathbf{y}}(i,\, j) = \beta_{i,j}/\nu_i$. The process $\mathbf{y}(\cdot)$ defined in this manner is indeed the so-called trace process of $\mathbf{x}(\cdot)$ (cf. [2, Appendix])

Remark 4.1.3 (Comments on the processes $\mathbf{x}(\cdot)$ and $\mathbf{y}(\cdot)$)**.** *The auxiliary process $\mathbf{x}(\cdot)$ represents the inter-valley dynamics of the process $\boldsymbol{x}_\epsilon(\cdot)$ by assuming that it spends the same time scale at all valleys (which is not true in general). Since the process $\boldsymbol{x}_\epsilon(\cdot)$ spends a negligible time scale on shallow valleys, we can take the suitable trace of the process $\mathbf{x}(\cdot)$ on the (indices corresponding to) deepest valleys to get the correct process representing the inter-(deepest) valley dynamics of the process $\boldsymbol{x}_\epsilon(\cdot)$. This trace process is $\mathbf{y}(\cdot)$.*

4.1.3 Markov chain description via convergence of order process

Recall that $H-h$ represents the depth of the deepest wells. We can expect from Eyring–Kramers formula for $\boldsymbol{x}_\epsilon(\cdot)$ obtained in Theorem 3.1.3 that the order of the time scale for a metastable transition is

$$\theta_\epsilon \;:=\; \exp\frac{H-h}{\epsilon}\;.$$

Hence, we speed up the process $\boldsymbol{x}_\epsilon(\cdot)$ by a factor of θ_ϵ and then observe the index of the valley in which the speeded-up process is staying. To that end, we write

$$\widetilde{\boldsymbol{x}}_\epsilon(t)=\boldsymbol{x}_\epsilon(\theta_\epsilon t)\quad ;\; t\geq 0$$

the speeded-up process. In view of the fact that $\mu_\epsilon(\mathcal{V}_\star)=1-o_\epsilon(1)$ (cf. Proposition 4.1.2), this index belongs to the set $S_\star$ with dominating probability. We wish to prove that this index process converges to the process $\mathbf{y}(\cdot)$ defined in the previous subsection. The major technical issue in this heuristic explanation is the fact that the speeded-up process $\widetilde{\boldsymbol{x}}_\epsilon(\cdot)$ may stay in the set $\Delta=\mathbb{R}\setminus\mathcal{V}_\star$ with small probability, and for this case, the index process is not defined. Thus, to formulate this convergent result in a rigorous manner, we recall the notion of the *order process* introduced in [2, 3]. To define the order process, define

$$T_\epsilon(t)\;:=\;\int_0^t \mathbf{1}\{\widetilde{\boldsymbol{x}}_\epsilon(s)\in\mathcal{V}_\star\}\,ds\quad ;\; t\geq 0\;,$$

which measures the amount of time for which the speeded-up process $\widetilde{\boldsymbol{x}}_\epsilon(\cdot)$ stayed in $\mathcal{V}_\star$ until time t. Then, define $S_\epsilon(t)$ as the generalized inverse of the random increasing function $T_\epsilon(\cdot)$:

$$S_\epsilon(t)\;:=\;\sup\{s\geq 0: T_\epsilon(s)\leq t\}\quad ;\; t\geq 0\;. \tag{4.9}$$

Define the trace process $\boldsymbol{\xi}_\epsilon(\cdot)$ as

$$\boldsymbol{\xi}_\epsilon(t) \;:=\; \widetilde{\boldsymbol{x}}_\epsilon(S_\epsilon(t)) \quad ; \; t \ge 0 \; .$$

This process the one is obtained from the process $\widetilde{\boldsymbol{x}}_\epsilon(\cdot) = \boldsymbol{x}_\epsilon(\theta_\epsilon \cdot)$ by turning off the clock when the process $\widetilde{\boldsymbol{x}}_\epsilon(\cdot)$ does not belong to $\mathcal{V}_\star$. In other words, the trajectory of $\boldsymbol{\xi}_\epsilon(\cdot)$ is obtained by removing the excursions of $\widetilde{\boldsymbol{x}}_\epsilon(\cdot)$ at Δ. Hence, we have $\boldsymbol{\xi}_\epsilon(t) \in \mathcal{V}_\star$ for all $t \ge 0$; furthermore, the process $\boldsymbol{\xi}_\epsilon(\cdot)$ is a Markov process (with jump) on $\mathcal{V}_\star$.

First, we show that the process $\boldsymbol{\xi}_\epsilon(\cdot)$ is a relevant approximation of the process $\widetilde{\boldsymbol{x}}_\epsilon(\cdot)$ in the sense that the excursion of $\widetilde{\boldsymbol{x}}_\epsilon(\cdot)$ at Δ is negligible. Denote by $\mathbb{P}^\epsilon_{\boldsymbol{x}}$ the law of the original process $\boldsymbol{x}_\epsilon(\cdot)$ starting from $\boldsymbol{x} \in \mathbb{R}^d$ and by $\mathbb{E}^\epsilon_{\boldsymbol{x}}$ the expectation with respect to it.

Theorem 4.1.4. *For all $t \ge 0$, it holds that*

$$\lim_{\epsilon \to 0} \sup_{\boldsymbol{x} \in \mathcal{V}_\star} \mathbb{E}^\epsilon_{\boldsymbol{x}} \Big[\int_0^t \mathbf{1}_\Delta(\widetilde{\boldsymbol{x}}_\epsilon(s))\, ds \Big] \;=\; 0 \; .$$

The proof of this result is a direct consequence of the analysis of resolvent equation explained in Section 4.2 and will be explained therein.

By assuming this theorem, it now suffices to analyze the inter-valley behavior of the trace process $\boldsymbol{\xi}_\epsilon(\cdot)$. To this end, we define a projection $\Psi : \mathcal{V}_\star \to S_\star$ simply by

$$\Psi(\boldsymbol{x}) \;=\; i \quad \text{if } \boldsymbol{x} \in \mathcal{V}_i \quad ; \; i \in S_\star \; , \tag{4.10}$$

which maps a point belonging to a deepest valley to the index of that valley. Finally, define a process on $S_\star$ as

$$\mathbf{y}_\epsilon(t) \;:=\; \Psi(\boldsymbol{\xi}_\epsilon(t)) \quad ; \; t \ge 0 \; ,$$

which represents the valley where the trace process $\boldsymbol{\xi}_\epsilon(t)$ is staying. This process $\mathbf{y}_\epsilon(\cdot)$ is called the *order process*. Denote by $\mathbf{Q}^\epsilon_{\pi_\epsilon}$ the law of the order process $\mathbf{y}_\epsilon(\cdot)$ when the underlying process $\boldsymbol{x}_\epsilon(\cdot)$ starts from a distribution π_ϵ on $\mathbb{R}^d$, and denote by $\mathbf{Q}_i$ the law of the limiting Markov chain $\mathbf{y}(\cdot)$ starting from $i \in S_\star$. The following convergence theorem is the main result of the current chapter.

Theorem 4.1.5. *For every $i \in S_\star$ and for any sequence of Borel probability measures $(\pi_\epsilon)_{\epsilon>0}$ concentrated on $\mathcal{V}_i$, the law $\mathbf{Q}^\epsilon_{\pi_\epsilon}$ of the order process converges to $\mathbf{Q}_i$ as $\epsilon \to 0$.*

The proof of the theorem based on the resolvent approach developed in [50] is given in the next subsection. We remark that this is a generalization of [83, Theorem 2.3], as the reversible case is the special $\boldsymbol{\ell} = \mathbf{0}$ case of our model. Moreover, a careful reading of our arguments reveals that, the speed of the convergence of the finite dimensional marginals is given by

$$\begin{aligned}&\mathbf{Q}^\epsilon_{\pi_\epsilon}[\mathbf{y}_\epsilon(t_i) \in A_i \text{ for } i = 1, \ldots, k] \\ &= \big(1 + O\big(\epsilon^{1/2} \log \frac{1}{\epsilon}\big)\big)\, \mathbf{Q}_i[\mathbf{y}(t_i) \in A_i \text{ for } i = 1, \ldots, k]\end{aligned}$$

under the conditions of Theorem 4.1.5, where the error term $O(\epsilon^{1/2} \log \frac{1}{\epsilon})$ is identical to the one appeared in [14, Theorems 3.1 and 3.2] and depends on $t_1, \ldots, t_k$.

Discussion on general case

Thus far, we have assumed that $\overline{\mathcal{H}} = \{\boldsymbol{x} \in \mathbb{R}^d : U(\boldsymbol{x}) \le H\}$ is connected. However, our argument can be readily applied to the general situation without this assumption as follows. If $\overline{\mathcal{H}}$ is not connected, we take a connected component $\mathcal{X}$ and denote by $\mathcal{W}_1, \ldots, \mathcal{W}_K$ the connected component of $\mathcal{H}$ contained in $\mathcal{X}$. Let $S = \{1, \ldots, K\}$. Then, we can define all the notations as before, and Theorem 4.1.5 holds unchanged. This can be readily verified by coupling with the same dynamical systems reflected at the boundary of the connected component of the domain $\{\boldsymbol{x} \in \mathbb{R}^d : U(\boldsymbol{x}) < H + a\}$ for

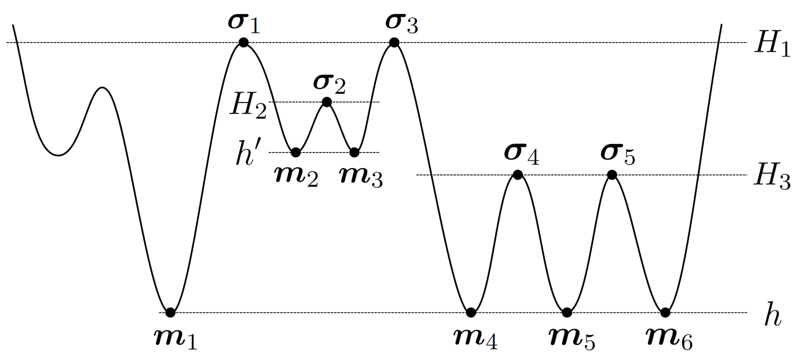


Figure 4.2: In this example of U, we have three possible choices of H: H_1, H_2, and H_3. By selecting $H = H_1$, we analyze the transitions between two deepest valleys $\mathcal{D}_{r_0}(\boldsymbol{m}_1)$ and $\mathcal{D}_{r_0}(\boldsymbol{m}_4) \cup \mathcal{D}_{r_0}(\boldsymbol{m}_5) \cup \mathcal{D}_{r_0}(\boldsymbol{m}_6)$. The time scale for these transitions is $e^{(H_1-h)/\epsilon}$, and Σ^* with this choice of H is $\{\boldsymbol{\sigma_1}, \boldsymbol{\sigma}_3\}$. Note that these two valleys are not directly connected, and all the transitions must pass through shallow valleys around $\boldsymbol{m}_2$ and $\boldsymbol{m}_3$. Hence, to get a precise Markov chain convergence, we must understand the behavior of the process in these shallow valleys. If we take $H = H_2$, we analyze the transitions between two shallow valleys $\mathcal{D}_{r_0}(\boldsymbol{m}_2)$ and $\mathcal{D}_{r_0}(\boldsymbol{m}_3)$. The time scale is now $e^{(H_2-h')/\epsilon}$. Finally, if we choose $H = H_3$, the successive transitions among three valleys $\mathcal{D}_{r_0}(\boldsymbol{m}_4)$, $\mathcal{D}_{r_0}(\boldsymbol{m}_5)$, and $\mathcal{D}_{r_0}(\boldsymbol{m}_6)$ are investigated in the time scale $e^{(H_3-h)/\epsilon}$. Note that these valleys are not distinguished at the level $H = H_1$; hence, we can analyze the metastable behavior with a higher resolution by taking this smaller H.

small enough a containing $\mathcal{X}$. This will be more precisely explained in [48] at which all the inter-valley (not restricted to the deepest valleys) dynamics are completely analyzed. Therefore, we can vary H to get different convergence results, and an example is given in Figure 4.2.

Connection to Eyring–Kramers formula

Now, we explain the connection between our result and the Eyring–Kramers formula obtained in Theorem 3.1.3. For simplicity, we suppose that $S = S_\star$ (i.e., $h_i = h$ for all $i \in S$) and that all the local minima of U are global minima. The explanation below is slightly more complicated without these assumptions, and we leave the details to the interested readers. Write $\tau_{\mathcal{A}}$, $\mathcal{A} \subset \mathbb{R}^d$, as the hitting time of the set $\mathcal{A}$. Then, by the Eyring–Kramers formula obtained in Theorem 3.1.3, we have, for $i \in S$ and $\boldsymbol{x} \in \mathcal{V}_i$,

$$\mathbb{E}_{\boldsymbol{x}}^{\epsilon}[\, \tau_{\mathcal{V}_\star \setminus \mathcal{V}_i} \,] \;=\; [\, 1 + o_\epsilon(1) \,] \, \frac{\nu_i}{\omega_i} \, \theta_\epsilon \; .$$

In other words, for the speeded-up process $\boldsymbol{x}_\epsilon(\theta_\epsilon \cdot)$ starting from a valley $\mathcal{V}_i$, the average of the transition time to other valleys is approximately ν_i/ω_i. This is in accordance with our result in that the limiting chain $\mathbf{y}(\cdot)$ starting from i jumps to one of the other sites at an average time of $[\, \sum_{j \in S} r_{\mathbf{y}}(i,\, j) \,]^{-1} = \nu_i/\omega_i$. On the other hand, our result provides more comprehensive information regarding the metastable behavior compared to the Eyring–Kramers formula, especially when $S \neq S_\star$.

4.2 Proof based on resolvent approach

In this section we review the resolvent approach developed in [50] and then prove Theorems 4.1.4 and 4.1.5 based on it.

4.2.1 Review of resolvent approach to metastability

Denote by $L_{\mathbf{y}}$ the generator associated with the limiting Markov chain $\mathbf{y}(\cdot)$ (defined in the previous section) that acts on $\mathbf{f} : S_\star \to \mathbb{R}$ such that

$$(L_{\mathbf{y}} \mathbf{f})(i) \;=\; \sum_{j \in S_\star} \frac{\beta_{i,j}}{\nu_i} \, [\, \mathbf{f}(j) - \mathbf{f}(i) \,] \; . \tag{4.11}$$

Recall (4.3) and define, for $i \in S$,

$$\widehat{\mathcal{V}}_i = \bigcup_{\boldsymbol{m} \in \mathcal{M}_i} \overline{\mathcal{D}_{2r_0}(\boldsymbol{m})} \; .$$

The following analysis of a solution to the resolvent equation is the main component of the resolvent approach.

Theorem 4.2.1. *Let* $\mathbf{f} : S_\star \to \mathbb{R}$ *be a given function and let* $\lambda > 0$ *where both* $\mathbf{f}$ *and* λ *are independent of* ϵ*. Then, the unique strong solution* $\phi_\epsilon^{\mathbf{f}}$ *to the resolvent equation (on* u*) on* $\mathbb{R}^d$

$$(\lambda - \theta_\epsilon \mathscr{L}_\epsilon)\, u \;=\; \sum_{i \in S_\star} [(\lambda - L_{\mathbf{y}})\mathbf{f}](i)\, \mathbf{1}_{\mathcal{V}_i} \tag{4.12}$$

satisfies

$$\lim_{\epsilon \to \infty} \sup_{\boldsymbol{x} \in \widehat{\mathcal{V}}_i} |\, \phi_\epsilon^{\mathbf{f}}(\boldsymbol{x}) - \mathbf{f}(i)\,| \;=\; 0 \quad \text{for all } i \in S_\star\,. \tag{4.13}$$

In [50, Theorem 2.3], it has been proven that this theorem implies Theorems 4.1.4 and 4.1.5 provided that the underlying metastable process $\boldsymbol{x}_\epsilon(\cdot)$ is a Markov process on discrete set. On the other hand, the proof of Theorem 4.1.5 based on Theorem 4.2.1 requires a slight technical modification since the solution $\phi_\epsilon^{\mathbf{f}}$ obtained in Theorem 4.2.1 does not belong to the core of the generator $\mathscr{L}_\epsilon$ associated with the process $\boldsymbol{x}_\epsilon(\cdot)$. We provide the proof here with emphasis on the modification. We remark that, we took supremum on $\widehat{\mathcal{V}}_i$ (instead of $\mathcal{V}_i$ as in [50]) in order to reserve enough space to carry out this modification.

The main idea is to replace the indicators in the right-hand side of (4.12) with smooth functions approximating the indicators so that we can recall the resolvent theory. Then, by using comparison argument, we shall solve all the technical problems. To that end, let us take $r_1 > r_0$ such that r_1 also satisfies the requirement (4.3) and define (cf. (4.4))

$$\mathcal{V}_{i,-} = \bigcup_{\boldsymbol{m} \in \mathcal{M}_i} \overline{\mathcal{D}_{r_0/2}(\boldsymbol{m})} \quad \text{and} \quad \mathcal{V}_{i,+} = \bigcup_{\boldsymbol{m} \in \mathcal{M}_i} \overline{\mathcal{D}_{r_1}(\boldsymbol{m})}$$

so that $\mathcal{V}_{i,-} \subset \mathcal{V}_i \subset \mathcal{V}_{i,+}$. Then, for each $i \in S_\star$, find smooth functions

$\zeta_{i,-},\, \zeta_{i,+} : \mathbb{R}^d \to [0,\,1]$ such that

$$\mathbf{1}_{\mathcal{V}_{i,-}} \le \zeta_{i,-} \le \mathbf{1}_{\mathcal{V}_i} \le \zeta_{i,+} \le \mathbf{1}_{\mathcal{V}_{i,+}} \;. \tag{4.14}$$

The key idea is to consider the functions $\psi^{\mathbf{f}}_{\epsilon,-}$ and $\psi^{\mathbf{f}}_{\epsilon,+}$ of the equations

$$(\lambda - \theta_\epsilon \mathscr{L}_\epsilon) u \;=\; \sum_{i \in S_\star} [(\lambda - L_{\mathbf{y}})\mathbf{f}](i)\, \zeta_{i,\pm} \;, \tag{4.15}$$

respectively. Then, $\psi^{\mathbf{f}}_{\epsilon,\pm}$ is now a smooth function that also satisfies (4.13) in the following sense.

Proposition 4.2.2. *We have that*

$$\lim_{\epsilon \to \infty} \sup_{\boldsymbol{x} \in \mathcal{V}_i} |\, \psi^{\mathbf{f}}_{\epsilon,\pm}(\boldsymbol{x}) - \mathbf{f}(i)\,| \;=\; 0 \quad \textit{for all } i \in S_\star \;. \tag{4.16}$$

Proof. Denote by $\phi^{\mathbf{f}}_{\epsilon,\pm}$ the solution to the equations

$$(\lambda - \theta_\epsilon \mathscr{L}_\epsilon) u \;=\; \sum_{i \in S_\star} [(\lambda - L_{\mathbf{y}})\mathbf{f}](i)\, \mathbf{1}_{\mathcal{V}_{i,\pm}} \;. \tag{4.17}$$

Since Theorem 4.2.1 holds for all $r_0 > 0$ satisfying (4.3), we can conclude that $\phi^{\mathbf{f}}_{\epsilon,\pm}$ also satisfies (4.13) in the sense that

$$\lim_{\epsilon \to \infty} \sup_{\boldsymbol{x} \in \mathcal{V}_i} |\, \phi^{\mathbf{f}}_{\epsilon,\pm}(\boldsymbol{x}) - \mathbf{f}(i)\,| \;=\; 0 \quad \text{for all } i \in S_\star \;. \tag{4.18}$$

Note that the appearance of $\widehat{\mathcal{V}}_i$ at (4.13) guarantees $\sup_{\boldsymbol{x} \in \mathcal{V}_i}$ in the previous estimate for $\phi^{\mathbf{f}}_{\epsilon,-}$. Therefore, the statement of proposition follows from (4.14) and the strong positivity of the operator $\lambda - \theta_\epsilon \mathscr{L}_\epsilon$. □

Now we use two functions $\psi^{\mathbf{f}}_{\epsilon,-}$ and $\psi^{\mathbf{f}}_{\epsilon,+}$ to prove Theorems 4.1.4 and 4.1.5, respectively. The huge benefit with these functions is the well-known expressions

$$\psi^{\mathbf{f}}_{\epsilon,\pm}(\boldsymbol{x}) = \mathbb{E}^{\epsilon}_{\boldsymbol{x}}\left[\int_0^{\infty} e^{-\lambda t} G_{\pm}(\boldsymbol{x}_{\epsilon}(t))dt\right] \tag{4.19}$$

where $G_{\pm} = \sum_{i\in S_\star}[(\lambda - L_{\mathbf{y}})\mathbf{f}](i)\,\zeta_{i,\pm}$ are bounded functions.

4.2.2 Proof of Theorem 4.1.4

Proof of Theorem 4.1.4. For $t > 0$, we have

$$\int_0^t \mathbf{1}_{\boldsymbol{\Delta}}(\widetilde{\boldsymbol{x}}_{\epsilon}(s))\,ds \le \int_0^t e^{\lambda t-\lambda s}\,\mathbf{1}_{\boldsymbol{\Delta}}(\widetilde{\boldsymbol{x}}_{\epsilon}(s))\,ds \le e^{\lambda t}\int_0^{\infty} e^{-\lambda s}\,\mathbf{1}_{\boldsymbol{\Delta}}(\widetilde{\boldsymbol{x}}_{\epsilon}(s))\,ds\ . \tag{4.20}$$

Also, by (4.14) and by definition of Δ, we have

$$\mathbf{1}_{\boldsymbol{\Delta}} \le 1 - \sum_{i\in S_\star}\zeta_{i,-}\ . \tag{4.21}$$

By (4.20) and (4.21), the proof of Theorem 4.1.4 is reduced to show that, for all $i \in S_\star$

$$\lim_{\epsilon\to 0}\sup_{\boldsymbol{x}\in\mathcal{V}_i}\mathbb{E}^{\epsilon}_{\boldsymbol{x}}\Big[\int_0^{\infty} e^{-\lambda s}\Big(1-\sum_{i\in S_\star}\zeta_{i,-}(\widetilde{\boldsymbol{x}}_{\epsilon}(s))\Big)\,ds\Big] \;=\; 0$$

or equivalently

$$\lim_{\epsilon\to 0}\sup_{\boldsymbol{x}\in\mathcal{V}_i}\mathbb{E}^{\epsilon}_{\boldsymbol{x}}\Big[\int_0^{\infty} e^{-\lambda s}\sum_{i\in S_\star}\zeta_{i,-}(\widetilde{\boldsymbol{x}}_{\epsilon}(s))\,ds\Big] \;=\; \frac{1}{\lambda}\ . \tag{4.22}$$

Note that the constant function $\mathbf{c} : S_\star \to \mathbb{R}$ defined by $\mathbf{c} \equiv \dfrac{1}{\lambda}$ satisfies $(\lambda - L_{\mathbf{y}})\mathbf{c} \equiv 1$ and therefore by (4.19), we can write

$$\mathbb{E}^{\epsilon}_{\boldsymbol{x}}\Big[\int_0^{\infty} e^{-\lambda s}\sum_{i\in S_\star}\zeta_{i,-}(\widetilde{\boldsymbol{x}}_{\epsilon}(s))\,ds\Big] = \psi^{\mathbf{c}}_{\epsilon,-}(\boldsymbol{x})\ .$$

Therefore, (4.22) is a direct consequence of Proposition 4.2.2. □

4.2.3 Proof of Theorem 4.1.5

Next we turn to the proof of Theorem 4.1.5. Here we need to use $\psi_{\epsilon,+}^{\mathbf{f}}$ instead. We first recall the following technical lemma from [50, Lemma 4.3].

Lemma 4.2.3. *Theorem 4.1.4 implies that, for all $t > 0$, (cf. (4.9))*

$$\lim_{\epsilon\to 0} \sup_{\boldsymbol{x}\in\mathcal{V}_\star} \mathbb{E}_{\boldsymbol{x}}^{\epsilon}\left[e^{-\lambda t} - e^{-\lambda S_\epsilon(t)}\right] = 0 \quad \textit{and}$$

$$\lim_{\epsilon\to 0} \sup_{\boldsymbol{x}\in\mathcal{V}_\star} \mathbb{E}_{\boldsymbol{x}}^{\epsilon}\left[\int_0^t \left\{e^{-\lambda s} - e^{-\lambda S_\epsilon(s)}\right\} ds\right] = 0\ .$$

Now we turn to the proof of Theorem 4.1.5.

Proof of Theorem 4.1.5. The argument given in [50] uses the solution $\phi_\epsilon^{\mathbf{f}}$ to prove Theorem 4.1.5 when the underlying metastable Markov process (in our case, $\boldsymbol{x}_\epsilon(\cdot)$) is defined on a discrete set. However, our proof requires a slight modification since our underlying Markov process $\boldsymbol{x}_\epsilon(\cdot)$ is now defined on $\mathbb{R}^d$. Technically speaking, the problem is the fact that $\phi_\epsilon^{\mathbf{f}} \notin C^2(\mathbb{R}^d)$ which implies that $\phi_\epsilon^{\mathbf{f}}$ does not belong to the core of the generator $\mathscr{L}_\epsilon$. Thus, we cannot conclude that

$$M_\epsilon^{\phi}(t) = e^{-\lambda t}\phi_\epsilon^{\mathbf{f}}(\widetilde{\boldsymbol{x}}_\epsilon(t)) - \phi_\epsilon^{\mathbf{f}}(\widetilde{\boldsymbol{x}}_\epsilon(0)) + \int_0^t e^{-\lambda s}(\lambda - \theta_\epsilon\mathscr{L}_\epsilon)\phi_\epsilon^{\mathbf{f}}(\widetilde{\boldsymbol{x}}_\epsilon(s))ds$$

is a martingale. This is the only place at which we can not use this function as in the proof of [50, Proposition 4.4]. This is the reason that we introduced the solution $\psi_{\epsilon,+}^{\mathbf{f}}$ which belongs to the core of $\mathscr{L}_\epsilon$, and we instead consider

$$M_\epsilon^{\psi}(t) = e^{-\lambda t}\psi_{\epsilon,+}^{\mathbf{f}}(\widetilde{\boldsymbol{x}}_\epsilon(t)) - \psi_{\epsilon,+}^{\mathbf{f}}(\widetilde{\boldsymbol{x}}_\epsilon(0)) + \int_0^t e^{-\lambda s}(\lambda - \theta_\epsilon\mathscr{L}_\epsilon)\psi_{\epsilon,+}^{\mathbf{f}}(\widetilde{\boldsymbol{x}}_\epsilon(s))ds$$

which is now a martingale.

In the proof of [50, Proposition 4.4] at which $M_\epsilon^\phi(t)$ was a martingale, the crucial ingredient of the proof is to consider $M_\epsilon^\phi(S_\epsilon(t))$ which can be rewritten in a simple form thanks to (4.12) and (4.13). Therefore, if we can show that

$$\lim_{\epsilon\to 0}\sup_{x\in\mathcal{V}_\star}\mathbb{E}_{\boldsymbol{x}}^\epsilon\left[\left|M_\epsilon^\psi(S_\epsilon(t))-M_\epsilon^\phi(S_\epsilon(t))\right|\right]=0\ \text{ for all } t\geq 0\ , \tag{4.23}$$

we can argue that $M_\epsilon^\phi(t)$ is a negligible perturbation of a martingale and therefore the proof of [50, Proposition 4.4] can still be applied. To prove (4.23), we need to prove that

$$\lim_{\epsilon\to 0}\sup_{x\in\mathcal{V}_\star}\mathbb{E}_{\boldsymbol{x}}^\epsilon\left|\psi_{\epsilon,+}^{\mathbf{f}}(\widetilde{\boldsymbol{x}}_\epsilon(S_\epsilon(t)))-\phi_\epsilon^{\mathbf{f}}(\widetilde{\boldsymbol{x}}_\epsilon(S_\epsilon(t)))\right|=0\ \text{ for all } t\geq 0, \text{ and} \tag{4.24}$$

$$\lim_{\epsilon\to 0}\sup_{x\in\mathcal{V}_\star}\mathbb{E}_{\boldsymbol{x}}^\epsilon\left[\int_0^{S_\epsilon(t)}e^{-\lambda s}\mathbf{1}_\Delta(\widetilde{\boldsymbol{x}}_\epsilon(s))ds\right]=0 \tag{4.25}$$

where the second one follows from the observation that, for some $C>0$,

$$\left|(\lambda-\theta_\epsilon\mathscr{L}_\epsilon)\psi_{\epsilon,+}^{\mathbf{f}}-(\lambda-\theta_\epsilon\mathscr{L}_\epsilon)\phi_\epsilon^{\mathbf{f}}\right|\leq C\sum_{i\in S_\star}\mathbf{1}_{\mathcal{V}_{i,+}\setminus\mathcal{V}_i}\leq C\mathbf{1}_\Delta\ .$$

Since $\widetilde{\boldsymbol{x}}_\epsilon(S_\epsilon(t))\in\mathcal{V}_\star$ by the definition (4.9) of $S_\epsilon(t)$, the estimate (4.24) is a direct consequence of (4.13) and Proposition 4.2.2.

Now it remains to prove (4.25). By the change of variable $s\leftarrow S_\epsilon(u)$,

$$\int_0^{S_\epsilon(t)}e^{-\lambda s}\mathbf{1}_{\mathcal{V}_\star}(\widetilde{\boldsymbol{x}}_\epsilon(s))ds=\int_0^t e^{-\lambda S_\epsilon(u)}\mathbf{1}_{\mathcal{V}_\star}(\widetilde{\boldsymbol{x}}_\epsilon(S_\epsilon(u)))du=\int_0^t e^{-\lambda S_\epsilon(u)}du$$

where the first equality follows from the definition of $S_\epsilon(\cdot)$ and the second one follows from $\widetilde{\boldsymbol{x}}_\epsilon(S_\epsilon(u)) \in \mathcal{V}_\star$ for all $u \ge 0$ by definition. Therefore,

$$\begin{aligned}
&\int_0^{S_\epsilon(t)} e^{-\lambda s}\mathbf{1}_\Delta(\widetilde{\boldsymbol{x}}_\epsilon(s))ds \\
&= \int_0^{S_\epsilon(t)} e^{-\lambda s}\left\{1 - \mathbf{1}_{\mathcal{V}_\star}(\widetilde{\boldsymbol{x}}_\epsilon(s))\right\}ds \\
&= \int_0^{S_\epsilon(t)} e^{-\lambda s}ds - \int_0^t e^{-\lambda S_\epsilon(u)}du \\
&= \left[\int_0^{S_\epsilon(t)} e^{-\lambda s}ds - \int_0^t e^{-\lambda s}ds\right] + \int_0^t \left\{e^{-\lambda s} - e^{-\lambda S_\epsilon(s)}\right\}ds\ .
\end{aligned}$$

Thus, (4.25) is a direct consequence of Lemma 4.2.3. □

The remainder of the chapter is focused on the proof of Theorem 4.2.1. Hence, we shall assume in the remainder of the chapter that both $\mathbf{f} : S_\star \to \mathbb{R}$ and $\lambda > 0$ are fixed and independent of ϵ. Moreover, we simply write ϕ_ϵ the solution $\phi_\epsilon^{\mathbf{f}}$ of equation (4.12). Moreover, we shall always implicitly assume that $\epsilon > 0$ is sufficiently small, as we are focusing on the asymptotics as $\epsilon \to 0$.

4.3 Analysis of resolvent equation

In this section, we prove Theorem 4.2.1 up to the construction of a certain test function, which will be deferred to Sections 4.4 and 4.5.

4.3.1 Energy estimate

In this subsection, we present a crucial energy estimate for the solutions of resolvent equation. Before proceeding to this estimate, we first remark that ϕ_ϵ is a bounded function as a consequence of [50, display (4.2)]. A detailed statement is given as the following proposition.

Proposition 4.3.1. *There exists $C > 0$ so that $\|\phi_\epsilon\|_{L^\infty(\mathbb{R}^d)} < C$ for all $\epsilon > 0$.*

Notation 4.3.2. *Here and later, we write $C > 0$ as a constant independent of ϵ and $\boldsymbol{x}$ (of course, C can possibly depend on $\mathbf{f}$ and λ). Different appearances of C possibly express different values.*

For sufficiently smooth function f, let us define the Dirichlet form $\mathscr{D}_\epsilon(f)$ with respect to the process $\boldsymbol{x}_\epsilon(\cdot)$ as

$$\mathscr{D}_\epsilon(f) := \int_{\mathbb{R}^d} f\,(-\mathscr{L}_\epsilon f)\, d\mu_\epsilon = \epsilon \int_{\mathbb{R}^d} |\nabla f|^2\, d\mu_\epsilon\ , \tag{4.26}$$

where the latter equality follows from an application of divergence theorem. Then, the flatness of the solution of resolvent equation (4.12) on each valley essentially follows from the following energy estimate (cf. [49, 76, 83]).

Proposition 4.3.3. *There exists $C > 0$ such that, for the solution ϕ_ϵ of (4.12),*

$$\mathscr{D}_\epsilon(\phi_\epsilon) \le C\,\theta_\epsilon^{-1}\ . \tag{4.27}$$

Proof. By multiplying both sides of (4.12) by $\phi_\epsilon d\mu_\epsilon$ and by performing the integral over $\mathbb{R}^d$, we get

$$\int_{\mathbb{R}^d} \phi_\epsilon(\lambda\phi_\epsilon - \theta_\epsilon\mathscr{L}_\epsilon\phi_\epsilon)\, d\mu_\epsilon \le C$$

by Proposition 4.3.1, since the right-hand side of (4.12) is a compactly supported bounded function independent of ϵ. The proof is completed by definition (4.26) of the Dirichlet form. □

4.3.2 Flatness of solution on each well

We first define

$$\delta := \delta(\epsilon) = \Big(\epsilon\,\log\frac{1}{\epsilon}\Big)^{1/2}\ , \tag{4.28}$$

which is an important scale in the analyses around saddle points carried out in the next section. Let $J > 0$ be a sufficiently large constant, and let $c_0 > 0$

be a constant that will be specified later in (4.71). For $i \in S$, define

$$\widehat{\mathcal{W}}_i \;:=\; \widehat{\mathcal{W}}_{i,\,\epsilon} \;=\; \{\boldsymbol{x} \in \mathcal{W}_i : U(\boldsymbol{x}) \le H - c_0 J^2 \delta^2\}\ . \tag{4.29}$$

Note that this set is connected if ϵ is sufficiently small, and we have $\mathcal{V}_i \subset \widehat{\mathcal{V}}_i \subset \widehat{\mathcal{W}}_i \subset \mathcal{W}_i$. For $i \in S$, denote by $\mathbf{m}_\epsilon(i)$ the average of ϕ_ϵ on $\widehat{\mathcal{W}}_i$, i.e.,

$$\mathbf{m}_\epsilon(i) \;=\; \frac{1}{\mathrm{vol}(\widehat{\mathcal{W}}_i)} \int_{\widehat{\mathcal{W}}_i} \phi_\epsilon(\boldsymbol{x})\, d\boldsymbol{x}\ ,$$

where $\mathrm{vol}(\mathcal{A}) \;=\; \displaystyle\int_{\mathcal{A}} d\boldsymbol{x}$ denotes the volume of a Lebesgue measurable set $\mathcal{A} \subset \mathbb{R}^d$ with respect to the Lebesgue measure. Remark from Proposition 4.3.1 that there exists $C > 0$ such that

$$\max_{i\in S} |\mathbf{m}_\epsilon(i)| \le C \tag{4.30}$$

for all $\epsilon > 0$. Our next objective is to prove that the function ψ_ϵ is close to its average value $\mathbf{m}_\epsilon(i)$ in $\widehat{\mathcal{W}}_i$ in the L^∞-sense.

Proposition 4.3.4. *For all $i \in S$, we have*

$$\|\phi_\epsilon - \mathbf{m}_\epsilon(i)\|_{L^\infty(\widehat{\mathcal{W}}_i)} \;=\; o_\epsilon(1)\ .$$

In [83, Section 4], it has been generally proven that the energy estimate of the form (4.27) is sufficient to prove Proposition 4.3.4 for the solution ϕ_ϵ. The argument presented therein is quite robust, and the reversibility is used only when the energy estimate is obtained. Hence, the methodology developed in [83] can be applied to Proposition 4.3.4 without any modification.

Remark 4.3.5. *In fact, the L^∞-boundedness such as Proposition 4.3.1 was not available when [83] started to prove the flatness result similar to Proposition 4.3.4. They obtained this boundedness as a byproduct of the proof of Proposition 4.3.4. Since we know this boundedness a priori owing to Propo-*

sition 4.3.1, the proof can indeed be written in even more concise form.

4.3.3 Characterization of $\mathbf{m}_\epsilon$ on deepest valleys via a test function

Since $\widehat{\mathcal{V}}_i \subset \widehat{\mathcal{W}}_i$ for all $i \in S$ by (4.3), it remains to prove the following proposition.

Proposition 4.3.6. *We have that*

$$|\,\mathbf{m}_\epsilon(i) - \mathbf{f}(i)\,| \,=\, o_\epsilon(1) \quad \text{for all } i \in S_\star\,.$$

Before proving Proposition 4.3.6, let us formally conclude the proof of Theorem 4.2.1 .

Proof of Theorem 4.2.1. By Propositions 4.3.4 and 4.3.6, we have $\|\phi_\epsilon - \mathbf{f}(i)\|_{L^\infty(\widehat{\mathcal{W}}_i)} = o_\epsilon(1)$ for all $i \in S_\star$. Since $\widehat{\mathcal{V}}_i \subset \widehat{\mathcal{W}}_i$, the proof is completed. □

Now, we turn to Proposition 4.3.6. The following proposition is the key in the proof of Proposition 4.3.6.

Proposition 4.3.7. *Let $\mathbf{g} = \mathbf{g}_\epsilon : S \to \mathbb{R}$ be a function that might depend on ϵ which is uniformly bounded in the sense that*

$$\sup_{\epsilon>0} \max_{i\in S} |\mathbf{g}(i)| < \infty\,. \tag{4.31}$$

Then, there exists a uniformly (in ϵ) bounded continuous function $Q_\epsilon^{\mathbf{g}} : \mathbb{R}^d \to \mathbb{R}$ that satisfies, for all $i \in S$,

$$Q_\epsilon^{\mathbf{g}}(\boldsymbol{x}) \,\equiv\, \mathbf{g}(i) \quad \text{for all } \boldsymbol{x} \in \mathcal{V}_i \quad \text{and} \tag{4.32}$$

$$\theta_\epsilon \int_{\mathbb{R}^d} Q_\epsilon^{\mathbf{g}}\,(\mathscr{L}_\epsilon \phi_\epsilon)\, d\mu_\epsilon \,=\, -\frac{1}{\nu_\star}\, D_{\mathbf{x}}(\mathbf{g}, \mathbf{m}_\epsilon) + o_\epsilon(1)\,. \tag{4.33}$$

The construction of the test function $Q_\epsilon^{\mathbf{g}}$ stated in the proposition above is the most crucial part of the proof and hence its proof is postponed to the next sections. At this moment, we prove Proposition 4.3.6 by assuming Proposition 4.3.7. Recall the bi-linear form $D_{\mathbf{x}}(\cdot,\,\cdot)$ defined in (4.7) and define another bi-linear form $D_{\mathbf{y}}(\mathbf{f},\,\mathbf{g})$ for $\mathbf{f},\,\mathbf{g} : S_\star \to \mathbb{R}$ as

$$D_{\mathbf{y}}(\mathbf{f},\,\mathbf{g}) \;:=\; \sum_{i\in S_\star} \mathbf{f}(i)\,(-L_{\mathbf{y}}\mathbf{g})(i)\,\frac{\nu_i}{\nu_\star} \;=\; \frac{1}{\nu_\star}\sum_{i\in S} \beta_{i,\,j}\,(\mathbf{f}(j)-\mathbf{f}(i))\,(\mathbf{g}(j)-\mathbf{g}(i))\;. \tag{4.34}$$

We recall some relations between $D_{\mathbf{x}}(\cdot,\cdot)$ and $D_{\mathbf{y}}(\cdot,\cdot)$ proved in [83]. For $\mathbf{u} : S_\star \to \mathbb{R}$, we define the harmonic extension $\widetilde{\mathbf{u}} : S \to \mathbb{R}$ as the extension of $\mathbf{u}$ to S satisfying $(L_{\mathbf{x}}\widetilde{\mathbf{u}})(i) \;=\; 0$ for all $i \in S \setminus S_\star$.

Lemma 4.3.8. *Let* $\mathbf{u},\,\mathbf{v} : S_\star \to \mathbb{R}$ *and let* $\widetilde{\mathbf{u}}$ *and* $\widetilde{\mathbf{v}}$ *be the harmonic extensions of* $\mathbf{u}$ *and* $\mathbf{v}$*, respectively. Then, we have* $D_{\mathbf{x}}(\widetilde{\mathbf{u}},\,\widetilde{\mathbf{v}}) = \nu_\star\, D_{\mathbf{y}}(\mathbf{u},\,\mathbf{v})$*. Moreover, for any extensions* $\mathbf{v}_1, \mathbf{v}_2$ *of* $\mathbf{v}$*, we have* $D_{\mathbf{x}}(\widetilde{\mathbf{u}},\,\mathbf{v}_1) = D_{\mathbf{x}}(\widetilde{\mathbf{u}},\,\mathbf{v}_2)$*.*

Proof. See [83, Lemma 4.3]. □

Now, we prove Proposition 4.3.6.

Proof of Proposition 4.3.6. Let us define $\mathbf{h}_\epsilon : S_\star \to \mathbb{R}$ as

$$\mathbf{h}_\epsilon(i) \;:=\; \mathbf{m}_\epsilon(i) - \mathbf{f}(i) \quad \text{for all } i \in S_\star\;, \tag{4.35}$$

and let $\widetilde{\mathbf{h}}_\epsilon$ be the harmonic extension of $\mathbf{h}_\epsilon$. Then, by the maximum principle and (4.30), there exists $C > 0$ such that

$$\max_{i\in S}\left|\widetilde{\mathbf{h}}_\epsilon(i)\right| = \max_{i\in S_\star}|\mathbf{h}_\epsilon(i)| \le C\;. \tag{4.36}$$

Therefore, we can construct a test function $Q_\epsilon^{\widetilde{\mathbf{h}}_\epsilon}$ constructed in Proposition 4.3.7.

Now, by Proposition 4.1.2, (4.12) and (4.32), we have

$$\begin{aligned}\int_{\mathbb{R}^d} Q_\epsilon^{\widetilde{\mathbf{h}}_\epsilon}(\lambda\phi_\epsilon - \theta_\epsilon \mathscr{L}_\epsilon \phi_\epsilon)\, d\mu_\epsilon &= (1+o_\epsilon(1)) \sum_{i\in S_\star} \mathbf{h}_\epsilon(i)\,(\lambda\mathbf{f} - L_{\mathbf{y}}\mathbf{f})(i)\,\frac{\nu_i}{\nu_\star} \\ &= \lambda \sum_{i\in S_\star} \mathbf{h}_\epsilon(i)\,\mathbf{f}(i)\,\frac{\nu_i}{\nu_\star} + D_{\mathbf{y}}(\mathbf{h}_\epsilon,\, \mathbf{f}) + o_\epsilon(1)\ ,\end{aligned} \tag{4.37}$$

where the last line follows from the definition of $D_{\mathbf{y}}$ and (4.36). The crucial idea in the proof is to compute the left-hand side of (4.37) in a different way and to compare with the previous computation. To that end, we first observe from Propositions 4.1.2, 4.3.1, 4.3.4 and (4.32) that

$$\lambda \int_{\mathbb{R}^d} Q_\epsilon^{\widetilde{\mathbf{h}}_\epsilon}\, \phi_\epsilon\, d\mu_\epsilon = \lambda \sum_{i\in S_\star} \mathbf{h}_\epsilon(i)\, \mathbf{m}_\epsilon(i)\,\frac{\nu_i}{\nu_\star} + o_\epsilon(1)\ . \tag{4.38}$$

By Proposition 4.3.7 and uniform boundedness of $Q_\epsilon^{\widetilde{\mathbf{h}}_\epsilon}$, we have

$$\int_{\mathbb{R}^d} Q_\epsilon^{\widetilde{\mathbf{h}}_\epsilon}(-\theta_\epsilon \mathscr{L}_\epsilon \phi_\epsilon)\, d\mu_\epsilon = \frac{1}{\nu_\star} D_{\mathbf{x}}(\widetilde{\mathbf{h}}_\epsilon,\, \mathbf{m}_\epsilon) + o_\epsilon(1)\ . \tag{4.39}$$

Denote by $\mathbf{m}_\epsilon^\star : S_\star \to \mathbb{R}$ the restriction of $\mathbf{m}_\epsilon : S \to \mathbb{R}$ on $S_\star$, and denote by $\widetilde{\mathbf{m}}_\epsilon^\star$ the harmonic extension of $\mathbf{m}_\epsilon^\star$. Then, by Lemma 4.3.8, we have

$$D_{\mathbf{x}}(\widetilde{\mathbf{h}}_\epsilon,\, \mathbf{m}_\epsilon) = D_{\mathbf{x}}(\widetilde{\mathbf{h}}_\epsilon,\, \widetilde{\mathbf{m}}_\epsilon^\star) = \nu_\star\, D_{\mathbf{y}}(\mathbf{h}_\epsilon,\, \mathbf{m}_\epsilon^\star)\ .$$

Inserting this into (4.39) and combining with (4.38), we can conclude that

$$\int_{\mathbb{R}^d} Q_\epsilon^{\widetilde{\mathbf{h}}_\epsilon}(\lambda\phi_\epsilon - \theta_\epsilon \mathscr{L}_\epsilon \phi_\epsilon)\, d\mu_\epsilon = \lambda \sum_{i\in S_\star} \mathbf{h}_\epsilon(i)\, \mathbf{m}_\epsilon(i)\,\frac{\nu_i}{\nu_\star} + D_{\mathbf{y}}(\mathbf{h}_\epsilon,\, \mathbf{m}_\epsilon^\star) + o_\epsilon(1)\ .$$

Comparing this with (4.37) and inserting (4.35), we get

$$\lambda \sum_{i\in S_\star} \mathbf{h}_\epsilon(i)^2 \frac{\nu_i}{\nu_\star} + D_{\mathbf{y}}(\mathbf{h}_\epsilon,\, \mathbf{h}_\epsilon) = o_\epsilon(1)\ .$$

This implies that $\max_{i\in S_\star} |\mathbf{h}_\epsilon(i)| = o_\epsilon(1)$ and therefore by recalling the definition (4.35) of $\mathbf{h}_\epsilon$ completes the proof. □

4.4 Construction of test function

In this section, we explicitly define the test function $Q_\epsilon^{\mathbf{g}}$, which is an approximating solution to the following elliptic equation:

$$\begin{cases} \mathscr{L}_\epsilon^* u \;=\; 0 & \text{on } \mathbb{R}^d \setminus (\cup_{i\in S}\mathcal{V}_i) \quad \text{and} \\ u \;=\; \mathbf{g}(i) & \text{on } \mathcal{V}_i \text{ for each } i \in S\ . \end{cases} \tag{4.40}$$

Although we share the same philosophy is this construction with the reversible case [83], the detailed construction and entailed computations are more complicated compared to the ones therein because of the non-reversibility.

4.4.1 Neighborhoods of saddle points

To construct the approximating solution to (4.40), we mainly focus on a neighborhood of each saddle point $\boldsymbol{\sigma} \in \Sigma_{i,\,j}$ for some $i,\, j \in S$, as the function u suddenly changes its value from $\mathbf{g}(i)$ to $\mathbf{g}(j)$ around such a saddle point. Therefore, we carefully define several notations regarding this neighborhood. In this subsection, we fix $i,\, j \in S$ and consider a saddle point $\boldsymbol{\sigma} \in \Sigma_{i,\,j}$. ***In addition, we assume that*** $i < j$ ***in this subsection.***

Notation 4.4.1. *We use the following notations in this subsection.*

1. *We abbreviate* $\mathbb{H} = \mathbb{H}^{\boldsymbol{\sigma}}$ *and* $\mathbb{L} = \mathbb{L}^{\boldsymbol{\sigma}}$.

2. *Since the symmetric matrix* $\mathbb{H}$ *has only one negative eigenvalue, we denote by* $-\lambda_1,\, \lambda_2,\, \dots,\, \lambda_d\, (= -\lambda_1^{\boldsymbol{\sigma}},\, \lambda_2^{\boldsymbol{\sigma}},\, \dots,\, \lambda_d^{\boldsymbol{\sigma}})$ *the eigenvalues of* $\mathbb{H}$, *where* $-\lambda_1$ *denotes the unique negative eigenvalue.*

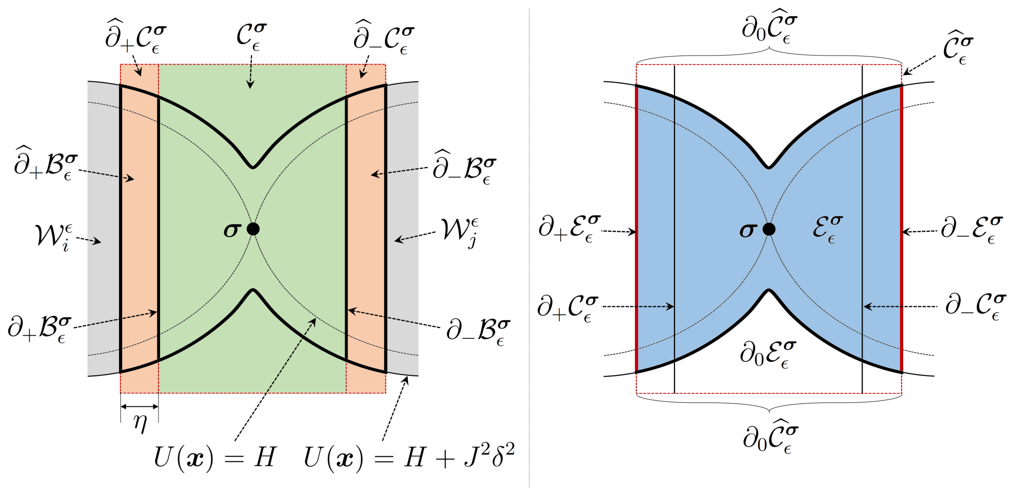


Figure 4.3: Illustration of various sets around a saddle point $\boldsymbol{\sigma}$ introduced in Section 4.4.1.

3. *Denote by* $\boldsymbol{e}_1^{\boldsymbol{\sigma}}$ *the unit eigenvector associated with the eigenvalue* $-\lambda_1$*, and by* $\boldsymbol{e}_k^{\boldsymbol{\sigma}}$*,* $k \geq 2$*, the unit eigenvector associated with the eigenvalue* λ_k*. In addition, we assume that the direction of* $\boldsymbol{e}_1^{\boldsymbol{\sigma}}$ *is toward* $\mathcal{W}_i$*, i.e., for all sufficiently small* $a > 0$*,* $\boldsymbol{\sigma} + a\boldsymbol{e}_1^{\boldsymbol{\sigma}} \in \mathcal{W}_i$*. Then, for* $\boldsymbol{x} \in \mathbb{R}^d$ *and* $k = 1, \dots, d$*, we write* $x_k = (\boldsymbol{x} - \boldsymbol{\sigma}) \cdot \boldsymbol{e}_k^{\boldsymbol{\sigma}}$*. In other words, we have* $\boldsymbol{x} = \boldsymbol{\sigma} + \sum_{m=1}^{d} x_m \boldsymbol{e}_m^{\boldsymbol{\sigma}}$*.*

Now, we define several sets around $\boldsymbol{\sigma}$. Figure 4.3 illustrates the sets appearing in this section. Recall δ from (4.28) and recall that $J > 0$ is a sufficiently large constant. Define an auxiliary set

$$\mathcal{T}_\epsilon^{\boldsymbol{\sigma}} := \Big\{ \boldsymbol{x} \in \mathbb{R}^d : x_k \in \big[-\frac{2J\delta}{\lambda_k^{1/2}}, \frac{2J\delta}{\lambda_k^{1/2}} \big] \ \text{ for } 2 \leq k \leq d \Big\} .$$

Then, define a box $\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ centered at $\boldsymbol{\sigma}$ as

$$\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} := \Big\{ \boldsymbol{x} \in \mathbb{R}^d : x_1 \in \big[-\frac{J\delta}{\lambda_1^{1/2}}, \frac{J\delta}{\lambda_1^{1/2}} \big] \Big\} \cap \mathcal{T}_\epsilon^{\boldsymbol{\sigma}} .$$

The boundary sets $\partial_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ and $\partial_-\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ defined below will be used later.

$$\partial_\pm\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \;=\; \Big\{\, \boldsymbol{x}\in\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} : x_1 = \pm\frac{J\delta}{\lambda_1^{1/2}} \,\Big\}\;. \tag{4.41}$$

We define another scale

$$\eta \;:=\; \eta(\epsilon) \;=\; \epsilon^2\;. \tag{4.42}$$

Then, define the enlargements of boundaries $\partial_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ and $\partial_-\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ as

$$\begin{aligned}
\widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} &= \Big\{\, \boldsymbol{x}\in\mathbb{R}^d : x_1 \,\in\, \Big[\,\frac{J\delta}{\lambda_1^{1/2}},\, \frac{J\delta}{\lambda_1^{1/2}}+\eta\,\Big] \Big\} \cap \mathcal{T}_\epsilon^{\boldsymbol{\sigma}}\;,\\
\widehat{\partial}_-\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} &= \Big\{\, \boldsymbol{x}\in\mathbb{R}^d : x_1 \,\in\, \Big[\,-\frac{J\delta}{\lambda_1^{1/2}}-\eta,\, -\frac{J\delta}{\lambda_1^{1/2}}\,\Big] \Big\} \cap \mathcal{T}_\epsilon^{\boldsymbol{\sigma}}\;.
\end{aligned}$$

With these enlarged boundaries, we can expand $\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ to

$$\widehat{\mathcal{C}}_\epsilon^{\boldsymbol{\sigma}} \;=\; \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \,\cup\, \widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \,\cup\, \widehat{\partial}_-\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}\;.$$

Let

$$\partial_0\widehat{\mathcal{C}}_\epsilon^{\boldsymbol{\sigma}} \;=\; \Big\{\, \boldsymbol{x}\in\widehat{\mathcal{C}}_\epsilon^{\boldsymbol{\sigma}} : x_k = \pm\frac{2J\delta}{\lambda_k^{1/2}} \;\text{ for some }\; 2\le k\le d \Big\}\;.$$

Then, by a Taylor expansion of U around $\boldsymbol{\sigma}$, we can readily verify that

$$U(\boldsymbol{x}) \;\ge\; H+\frac{3}{2}\,J^2\,\delta^2\,[\,1+o_\epsilon(1)\,] \quad \text{for all } \boldsymbol{x}\in\partial_0\widehat{\mathcal{C}}_\epsilon^{\boldsymbol{\sigma}}\;. \tag{4.43}$$

For the detailed proof, we refer to Lemma 3.4.3. Now, we define

$$\mathcal{K}_\epsilon \;=\; \{\boldsymbol{x}\in\mathbb{R}^d : U(\boldsymbol{x}) < H + J^2\delta^2\}\;,$$

so that, by (4.43), the boundary $\partial_0\widehat{\mathcal{C}}_\epsilon^{\boldsymbol{\sigma}}$ does not belong to $\mathcal{K}_\epsilon$ provided that ϵ is sufficiently small. Then, we define

$$\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \;=\; \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}\cap\mathcal{K}_\epsilon \quad,\; \widehat{\partial}_\pm\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \;=\; \widehat{\partial}_\pm\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}\cap\mathcal{K}_\epsilon \quad\text{and}\quad \mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \;=\; \widehat{\mathcal{C}}_\epsilon^{\boldsymbol{\sigma}}\cap\mathcal{K}_\epsilon$$

so that $\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} = \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \cup \widehat{\partial}_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \cup \widehat{\partial}_-\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$. Denote by $\partial\mathcal{E}_\epsilon^{\boldsymbol{\sigma}}$ the boundary of the set $\mathcal{E}_\epsilon^{\boldsymbol{\sigma}}$ and decompose it into

$$\partial\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \;=\; \partial_+\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \,\cup\, \partial_-\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \,\cup\, \partial_0\mathcal{E}_\epsilon^{\boldsymbol{\sigma}}$$

such that

$$\begin{aligned}\partial_\pm\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} &= \Big\{ \boldsymbol{x} \in \partial\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} : x_1 = \pm\Big(\frac{J\delta}{\lambda_1^{1/2}} + \eta\Big) \Big\} \;\; \text{and ,} \\ \partial_0\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} &= \Big\{ \boldsymbol{x} \in \partial\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} : x_1 \neq \pm\Big(\frac{J\delta}{\lambda_1^{1/2}} + \eta\Big) \Big\} \,.\end{aligned}$$

Then, by (4.43) (one can readily check from Figure 4.3), for sufficiently small $\epsilon > 0$,

$$U(\boldsymbol{x}) \;=\; H + J^2\delta^2 \quad \text{for all } \boldsymbol{x} \in \partial_0\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \,. \tag{4.44}$$

Furthermore, by our selection of the direction of vector $\boldsymbol{e}_1^{\boldsymbol{\sigma}}$ (cf. Notation 4.4.1-(3)), we have

$$\partial_+\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \;\subset\; \partial\mathcal{W}_i^\epsilon \;\text{ and }\; \partial_-\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \;\subset\; \partial\mathcal{W}_j^\epsilon \,. \tag{4.45}$$

Similarly, we decompose $\partial\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ into $\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, $\partial_-\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, and $\partial_0\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$ such that

$$\begin{aligned}\partial_\pm\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} &= \Big\{ \boldsymbol{x} \in \partial\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} : x_1 = \pm\frac{J\delta}{\lambda_1^{1/2}} \Big\} \,, \;\; \text{and} \\ \partial_0\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} &= \Big\{ \boldsymbol{x} \in \partial\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} : x_1 \neq \pm\frac{J\delta}{\lambda_1^{1/2}} \Big\} \,.\end{aligned} \tag{4.46}$$

4.4.2 Decomposition of $\mathcal{K}_\epsilon$

Now, we turn to the global picture. Recall Σ^* from (4.1). By (4.44), we can observe that $\mathcal{K}_\epsilon \setminus (\cup_{\boldsymbol{\sigma}\in\Sigma^*}\mathcal{E}_\epsilon^{\boldsymbol{\sigma}})$ consists of K connected components, and we denote by $\mathcal{W}_i^\epsilon$, $i \in S$, the component among them containing $\mathcal{V}_i$. Then, we can decompose $\mathcal{K}_\epsilon$ such that

$$\mathcal{K}_\epsilon = \Big[\bigcup_{i\in S} \mathcal{W}_i^\epsilon \Big] \cup \Big[\bigcup_{\boldsymbol{\sigma}\in\Sigma^*} \mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \Big] . \tag{4.47}$$

The test function $Q_\epsilon^{\mathbf{g}}$ is constructed on this global structure of $\mathcal{K}_\epsilon$.

4.4.3 Construction of function $Q_\epsilon^{\mathbf{g}}$

Construction around a saddle point

We start by introducing the building block for the construction of $Q_\epsilon^{\mathbf{g}}$, which is a function on $\mathcal{E}_\epsilon^{\boldsymbol{\sigma}} = \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \cup \widehat{\partial}_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \cup \widehat{\partial}_-\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$. First, let us focus on the set $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$. Recall that $\mathbb{H}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}}$ has a unique negative eigenvalue $-\mu^{\boldsymbol{\sigma}}$. Denote by $\mathbb{A}^\dagger$ the transpose of the matrix $\mathbb{A}$. Then, we can readily verify that (cf. (3.23)) the matrix $\mathbb{H}^{\boldsymbol{\sigma}} - (\mathbb{L}^{\boldsymbol{\sigma}})^\dagger$ is similar to $\mathbb{H}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}}$ and hence has the unique negative eigenvalue $-\mu^{\boldsymbol{\sigma}}$. We denote by $\boldsymbol{v}^{\boldsymbol{\sigma}}$ the unit eigenvector of $\mathbb{H}^{\boldsymbol{\sigma}} - (\mathbb{L}^{\boldsymbol{\sigma}})^\dagger$ associated with $-\mu^{\boldsymbol{\sigma}}$. We assume that $\boldsymbol{v}^{\boldsymbol{\sigma}} \cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}} > 0$, as we can take $-\boldsymbol{v}^{\boldsymbol{\sigma}}$ instead if this inner product is negative. We note that $\boldsymbol{v}^{\boldsymbol{\sigma}} \cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}} \neq 0$ by Lemma 3.4.1.

Recall the definition 3.30 of a function $p_\epsilon^{\boldsymbol{\sigma}} : \mathbb{R}^d \to \mathbb{R}$ as

$$p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) := \frac{1}{c_\epsilon^{\boldsymbol{\sigma}}} \int_{-\infty}^{(\boldsymbol{x}-\boldsymbol{\sigma})\cdot\boldsymbol{v}^{\boldsymbol{\sigma}}} e^{-\frac{\mu^{\boldsymbol{\sigma}}}{2\epsilon}t^2}\, dt \quad ; \ \boldsymbol{x} \in \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} , \tag{4.48}$$

where the normalizing constant $c_\epsilon^{\boldsymbol{\sigma}}$ is given by

$$c_\epsilon^{\boldsymbol{\sigma}} = \int_{-\infty}^{\infty} e^{-\frac{\mu^{\boldsymbol{\sigma}}}{2\epsilon}t^2}\, dt = \sqrt{\frac{2\pi\epsilon}{\mu^{\boldsymbol{\sigma}}}} . \tag{4.49}$$

Note that we defined the function on $\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ containing $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$. The function $p_\epsilon^{\boldsymbol{\sigma}}$ introduced here is identical to (3.30). It is remarkable that the test function for the Eyring–Kramers formula and that for the Markov chain convergence share the building block, while the global construction from this building block is carried out in a different manner.

The function $p_\epsilon^{\boldsymbol{\sigma}}$ is an approximating solution $\mathscr{L}_\epsilon^* f \simeq 0$ with approximating boundary conditions $f \simeq 1$ on $\partial_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ and $f \simeq 0$ on $\partial_- \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ by our assumption that $\boldsymbol{v}^{\boldsymbol{\sigma}} \cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}} > 0$. The approximating property $\mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}} \simeq 0$ can be quantified in the following proposition, which has been proven in Proposition 3.4.5.

Proposition 4.4.2. *For all $\boldsymbol{\sigma} \in \Sigma^*$, we have*

$$\theta_\epsilon \int_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} |\, \mathscr{L}_\epsilon^* p_\epsilon^{\boldsymbol{\sigma}} \,|\, d\mu_\epsilon \;=\; o_\epsilon(1) \;.$$

Now, we focus on the properties $p_\epsilon^{\boldsymbol{\sigma}} \simeq 1$ on $\partial_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ and $p_\epsilon^{\boldsymbol{\sigma}} \simeq 0$ on $\partial_- \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$. When suitably extending this function to get a continuous function on $\mathbb{R}^d$, these asymptotic equalities along the boundaries cause technical problems. They become the exact equality for the reversible case considered in [83] as $\boldsymbol{v}^{\boldsymbol{\sigma}} = \boldsymbol{e}_1^{\boldsymbol{\sigma}}$. For our case, the discontinuity is a natural consequence of the non-reversibility; hence, we need an additional continuation procedure. In Chapter 3, this continuation has been carried out by mollification via a smooth mollifier. For the current problem, such a procedure does not work, and we take a different path of construction. The enlarged set $\mathcal{E}_\epsilon^{\boldsymbol{\sigma}}$ is introduced for performing this continuation procedure.

Now, we continuously extend $p_\epsilon^{\boldsymbol{\sigma}}$ to $\widehat{\mathcal{C}}_\epsilon^{\boldsymbol{\sigma}}$. For each $\boldsymbol{x} = \boldsymbol{\sigma} + \sum_{k=1}^{d} x_k \boldsymbol{e}_k^{\boldsymbol{\sigma}} \in \widehat{\partial}_\pm \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$, we write

$$\overline{\boldsymbol{x}} \;=\; \boldsymbol{\sigma} \,\pm\, \frac{J\delta}{(\lambda_1^{\boldsymbol{\sigma}})^{1/2}}\, \boldsymbol{e}_1^{\boldsymbol{\sigma}} \,+\, \sum_{k=2}^{d} x_k\, \boldsymbol{e}_k^{\boldsymbol{\sigma}} \;\in\; \partial_\pm \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \;,$$

where the boundaries $\partial_\pm \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ are defined in (4.41). Then, define $p_\epsilon^{\boldsymbol{\sigma}}$ on the enlarged boundaries $\widehat{\partial}_\pm \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ as

$$p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) \;=\; \begin{cases} 1 + \dfrac{1}{\eta} \Big[(\boldsymbol{x} - \boldsymbol{\sigma}) \cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}} - \dfrac{J\delta}{(\lambda_1^{\boldsymbol{\sigma}})^{1/2}} - \eta \Big] (1 - p_\epsilon^{\boldsymbol{\sigma}}(\overline{\boldsymbol{x}})) & \text{for } \boldsymbol{x} \in \widehat{\partial}_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \;, \\ \dfrac{1}{\eta} \Big[(\boldsymbol{x} - \boldsymbol{\sigma}) \cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}} + \dfrac{J\delta}{(\lambda_1^{\boldsymbol{\sigma}})^{1/2}} + \eta \Big] p_\epsilon^{\boldsymbol{\sigma}}(\overline{\boldsymbol{x}}) & \text{for } \boldsymbol{x} \in \widehat{\partial}_- \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \;. \end{cases} \tag{4.50}$$

By such an extension, we can check that $p_\epsilon^{\boldsymbol{\sigma}}$ is continuous on $\widehat{\mathcal{C}}_\epsilon^{\boldsymbol{\sigma}}$. Now, we regard $p_\epsilon^{\boldsymbol{\sigma}}$ as a function on $\mathcal{E}_\epsilon^{\boldsymbol{\sigma}}$. Then, we can check that $p_\epsilon^{\boldsymbol{\sigma}}$ satisfies the exact boundary conditions

$$p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) \;=\; \begin{cases} 1 & \text{if } \boldsymbol{x} \in \partial_+ \mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \;, \\ 0 & \text{if } \boldsymbol{x} \in \partial_- \mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \;. \end{cases} \tag{4.51}$$

Now, we claim that the cost of this continuation procedure is tolerable.

Lemma 4.4.3. *For all* $\boldsymbol{\sigma} \in \Sigma^*$*, we have*

$$\theta_\epsilon \, \epsilon \int_{\widehat{\partial}_\pm \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}} |\nabla p_\epsilon^{\boldsymbol{\sigma}}|^2 \, d\mu_\epsilon \;=\; o_\epsilon(1) \;.$$

We defer the technical proof of this lemma to the next subsection.

Global construction

For $\mathbf{g} = \mathbf{g}_\epsilon : S \to \mathbb{R}$, we can now define the function $Q_\epsilon^{\mathbf{g}} : \mathbb{R}^d \to \mathbb{R}$. First, we define this function on $\mathcal{K}_\epsilon$ (cf. (4.47)) such that

$$Q_\epsilon^{\mathbf{g}}(\boldsymbol{x}) \;=\; \begin{cases} \mathbf{g}(i) & \text{for } \boldsymbol{x} \in \mathcal{W}_i^\epsilon,\; i \in S \;, \\ \mathbf{g}(j) + (\mathbf{g}(i) - \mathbf{g}(j))\, p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) & \text{for } \boldsymbol{x} \in \mathcal{E}_\epsilon^{\boldsymbol{\sigma}},\; \boldsymbol{\sigma} \in \Sigma_{i,j} \text{ for } i < j \;. \end{cases} \tag{4.52}$$

By (4.45) and (4.51), the function $Q_\epsilon^{\mathbf{g}}$ is continuous on $\mathcal{K}_\epsilon$. Since for all $\boldsymbol{\sigma} \in \Sigma^*$, it holds that

$$p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) \;\in\; [0,\, 1] \quad \text{and} \quad — \nabla p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) \,| \;\le\; C\, \eta^{-1} \quad \text{for all } \boldsymbol{x} \in \mathcal{E}_\epsilon^{\boldsymbol{\sigma}} \;,$$

we can check that

$$\|\, Q_\epsilon^{\mathbf{g}} \,\|_{L^\infty(\mathcal{K}_\epsilon)} \;=\; \|\, \mathbf{g} \,\|_\infty \quad \text{and} \quad \|\nabla Q_\epsilon^{\mathbf{g}}\|_{L^\infty(\mathcal{K}_\epsilon)} \;\le\; C\, \eta^{-1} \,\|\, \mathbf{g} \,\|_\infty \;,$$

where $\|\mathbf{g}\| = \max_{i\in S} |\mathbf{g}(i)|$. Note that $Q_\epsilon^{\mathbf{g}}$ is not differentiable along the boundary of $\mathcal{E}_\epsilon^{\boldsymbol{\sigma}}$ for each $\boldsymbol{\sigma} \in \Sigma^*$. In this computation and subsequent computations, we implicitly regard $\nabla Q_\epsilon^{\mathbf{g}}$ as an a.e. defined function except for these discontinuity surfaces. Then, we can continuously extend this function to $\mathbb{R}^d$ such that

$$\| Q_\epsilon^{\mathbf{g}} \|_{L^\infty(\mathbb{R}^d)} = \| \mathbf{g} \|_\infty \quad \text{and} \quad \| \nabla Q_\epsilon^{\mathbf{g}} \|_{L^\infty(\mathbb{R}^d)} \le C\, \eta^{-1} \| \mathbf{g} \|_\infty \, . \tag{4.53}$$

In particular, we have uniformly boundedness of $Q_\epsilon^{\mathbf{g}}$ since we assumed in Proposition 4.3.7 that $\|\mathbf{g}\|$ is uniformly bounded in ϵ. Note also that the condition (4.32) of Proposition 4.3.7 is satisfied by $Q_\epsilon^{\mathbf{g}}$ immediately from its definition in (4.52). The last and the most technical part is to check that $Q_\epsilon^{\mathbf{g}}$ satisfies (4.33). This will be carried out in the next section. Before doing that, we conclude the proof of Lemma 4.4.3.

4.4.4 Proof of Lemma 4.4.3

Before proving Lemma 4.4.3, we explain a decomposition of the extended boundary $\widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$, which will be used several times later. Define, for $a > 0$,

$$\widehat{\partial}_+^{1,a}\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} = \{ \boldsymbol{x} \in \widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} : \overline{\boldsymbol{x}} \cdot \boldsymbol{v} \ge aJ\delta \} \, , \tag{4.54}$$

$$\widehat{\partial}_+^{2,a}\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} = \{ \boldsymbol{x} \in \widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} : U(\boldsymbol{x}) \ge H + aJ^2\delta^2 \} \, . \tag{4.55}$$

Lemma 4.4.4. *There exists $a_0 > 0$ such that, for all $a \in (0, a_0)$,*

$$\widehat{\partial}_+^{1,a}\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \cup \widehat{\partial}_+^{2,a}\mathcal{C}_\epsilon^{\boldsymbol{\sigma}} = \widehat{\partial}_+\mathcal{C}_\epsilon \, .$$

The proof is a direct consequence of Lemma 3.4.10 as $\eta \ll \delta$ and is omitted.

Proof of Lemma 4.4.3. Fix $\boldsymbol{\sigma} \in \Sigma^*$, and for convenience of notation, we assume that $\boldsymbol{\sigma} = \mathbf{0}$. We only consider the integral on $\widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ since the proof for the case $\widehat{\partial}_- \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$ is essentially the same. Write $\boldsymbol{e}_1^{\boldsymbol{\sigma}} = (e_1, \dots, e_d)$ and $\boldsymbol{v}^{\boldsymbol{\sigma}} = (v_1, \dots, v_d)$. Then, by the explicit formula (4.50) for $p_\epsilon^{\boldsymbol{\sigma}}$, we have, for $\boldsymbol{x} \in \widehat{\partial}_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$,

$$\nabla_1 p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) = \frac{e_1}{\eta} [\, 1 - p_\epsilon^{\boldsymbol{\sigma}}(\overline{\boldsymbol{x}}) \,] \,, \tag{4.56}$$

$$\begin{aligned} \nabla_i p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x}) = & \frac{e_i}{\eta} [\, 1 - p_\epsilon^{\boldsymbol{\sigma}}(\overline{\boldsymbol{x}}) \,] \\ & + \frac{v_i^{\boldsymbol{\sigma}}}{\eta \, c_\epsilon^{\boldsymbol{\sigma}}} e^{-\frac{\mu}{2\epsilon}(\overline{\boldsymbol{x}} \cdot \boldsymbol{v}^{\boldsymbol{\sigma}})^2} \Big[\boldsymbol{x} \cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}} - \frac{J\delta}{(\lambda_1^{\boldsymbol{\sigma}})^{1/2}} - \eta \Big] \quad ; \; i \ge 2 \,. \end{aligned} \tag{4.57}$$

Since the absolute value of the term in the second pair of brackets in (4.57) is bounded by η for $\boldsymbol{x} \in \widehat{\partial}_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$, we can conclude that

$$|\nabla p_\epsilon^{\boldsymbol{\sigma}}(\boldsymbol{x})|^2 \,\le\, \frac{C}{\eta^2} [\, 1 - p_\epsilon^{\boldsymbol{\sigma}}(\overline{\boldsymbol{x}}) \,]^2 + \frac{C}{\epsilon} e^{-\frac{\mu}{\epsilon}(\overline{\boldsymbol{x}} \cdot \boldsymbol{v}^{\boldsymbol{\sigma}})^2} \,.$$

Now, for $a \in (0, a_0)$ where a_0 is the constant appearing in Lemma 4.4.4, it suffices to prove that

$$\theta_\epsilon \, \epsilon \int_{\widehat{\partial}_+^{k,\,a} \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}} \Big[\frac{1}{\eta^2} [\, 1 - p_\epsilon^{\boldsymbol{\sigma}}(\overline{\boldsymbol{x}}) \,]^2 + \frac{1}{\epsilon} e^{-\frac{\mu}{\epsilon}(\overline{\boldsymbol{x}} \cdot \boldsymbol{v}^{\boldsymbol{\sigma}})^2} \Big] \mu_\epsilon(\boldsymbol{x}) \, d\boldsymbol{x} \,=\, o_\epsilon(1) \;\text{ for } k = 1,\, 2 \,. \tag{4.58}$$

We first consider the case $k = 1$. By the elementary inequality $\int_b^\infty e^{-t^2/2} dt \le \frac{1}{b} e^{-b^2/2}$ for $b > 0$, we can deduce that

$$1 - p_\epsilon^{\boldsymbol{\sigma}}(\overline{\boldsymbol{x}}) \,\le\, \frac{C}{\overline{\boldsymbol{x}} \cdot \boldsymbol{v}^{\boldsymbol{\sigma}}} e^{-\frac{\mu}{2\epsilon}(\overline{\boldsymbol{x}} \cdot \boldsymbol{v}^{\boldsymbol{\sigma}})^2} \,\le\, \frac{C}{\delta} e^{-\frac{\mu}{2\epsilon}(\overline{\boldsymbol{x}} \cdot \boldsymbol{v}^{\boldsymbol{\sigma}})^2} \quad \text{for } \boldsymbol{x} \in \widehat{\partial}_+^{1,\,a} \mathcal{C}_\epsilon^{\boldsymbol{\sigma}} \,.$$

Hence, the left-hand side of (4.58) is bounded from above by

$$C \frac{\theta_\epsilon \epsilon}{\eta^2 \delta^2} \int_{\widehat{\partial}_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}} e^{-\frac{\mu}{\epsilon}(\overline{\boldsymbol{x}} \cdot \boldsymbol{v}^{\boldsymbol{\sigma}})^2} \mu_\epsilon(\boldsymbol{x}) d\boldsymbol{x} \,\le\, C \frac{1}{\epsilon^{d/2+3} \delta^2} \int_{\widehat{\partial}_+ \mathcal{C}_\epsilon^{\boldsymbol{\sigma}}} e^{-\frac{1}{2\epsilon} \boldsymbol{x} \cdot [\mathbb{H}^{\boldsymbol{\sigma}} + 2\mu^{\boldsymbol{\sigma}} \, \boldsymbol{v}^{\boldsymbol{\sigma}} \otimes \boldsymbol{v}^{\boldsymbol{\sigma}}] \boldsymbol{x}} d\boldsymbol{x} \,, \tag{4.59}$$

where we applied (4.5), the Taylor expansion of U around $\boldsymbol{\sigma} = \boldsymbol{0}$, and the fact that

$$e^{-\frac{\mu}{\epsilon}(\overline{\boldsymbol{x}}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2} \;=\; [\,1+o_\epsilon(1)\,]\, e^{-\frac{\mu}{\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v}^{\boldsymbol{\sigma}})^2} \;.$$

Since the matrix $\mathbb{H}^{\boldsymbol{\sigma}} + 2\mu^{\boldsymbol{\sigma}}\boldsymbol{v}^{\boldsymbol{\sigma}} \otimes \boldsymbol{v}^{\boldsymbol{\sigma}}$ is positive definite by Lemma 3.4.2, we can write

$$\boldsymbol{x}\cdot(\mathbb{H}^{\boldsymbol{\sigma}} + 2\,\mu^{\boldsymbol{\sigma}}\,\boldsymbol{v}^{\boldsymbol{\sigma}} \otimes \boldsymbol{v}^{\boldsymbol{\sigma}})\,\boldsymbol{x} \;\ge\; C\,|\boldsymbol{x}|^2 \;\ge\; C\,J^2\,\delta^2 \;, \tag{4.60}$$

as there exists $C>0$ such that $|\boldsymbol{x}-\boldsymbol{\sigma}| \ge CJ\delta$ for all $\boldsymbol{x} \in \widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$. By inserting (4.60) into (4.59), we can bound the right-hand side of (4.59) from above by

$$C\,\frac{1}{\epsilon^{d/2+3}\delta^2}\,\mathrm{vol}(\widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}})\,\epsilon^{CJ^2} \;=\; o_\epsilon(1)\;,$$

where the equality holds for sufficiently large J since $\mathrm{vol}(\widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}) = O(\eta\delta^{d-1})$.

Next, we consider the case $k=2$ of (4.58). For this case, by (4.5), the left-hand side of (4.58) is bounded from above by

$$C\,\epsilon^{1-d/2}\int_{\widehat{\partial}_+^{2,\,a}\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}} \frac{1}{\eta^2}\,e^{-\frac{U(\boldsymbol{x})-H}{\epsilon}}\,d\boldsymbol{x} \;\le\; \frac{C\,\epsilon^{1-d/2}}{\eta^2}\,\epsilon^{aJ^2}\,\mathrm{vol}(\widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}) \;=\; o_\epsilon(1)\;,$$

where the inequality holds from the definition of $\widehat{\partial}_+^{2,\,a}\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$, and the last equality holds for sufficiently large J since $\mathrm{vol}(\widehat{\partial}_+\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}) = O(\eta\delta^{d-1})$. This completes the proof. □

4.5 Proof of Proposition 4.3.7

We fix $\mathbf{g} = \mathbf{g}_\epsilon : S \to \mathbb{R}$ throughout this section which is *uniformly bounded in* ϵ in the sense of (4.31). The function $Q_\epsilon^{\mathbf{g}}$ appearing in this section is the one defined in (4.52).

4.5.1 Reduction to local computations around saddle points

In this subsection, we reduce the proof of Proposition 4.3.7 to two local estimates around saddle point $\boldsymbol{\sigma} \in \Sigma^*$. We perform this reduction via the following proposition.

Proposition 4.5.1. *It holds that*

$$\begin{aligned}&\theta_\epsilon \int_{\mathbb{R}^d} Q_\epsilon^{\mathbf{g}} \, (\mathscr{L}_\epsilon \phi_\epsilon) \, d\mu_\epsilon \\ &\quad = \, o_\epsilon(1) \, + \sum_{i,\, j \in S,\, i<j} \; \sum_{\boldsymbol{\sigma} \in \Sigma_{i,\,j}} (\, \mathbf{g}(i) - \mathbf{g}(j) \,) \, [\, A_1(\boldsymbol{\sigma}) + A_2^+(\boldsymbol{\sigma}) + A_2^-(\boldsymbol{\sigma}) \,] \, ,\end{aligned}$$

where

$$\begin{aligned}A_1(\boldsymbol{\sigma}) &= \, -\theta_\epsilon \, \epsilon \int_{\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \nabla p_\epsilon^{\boldsymbol{\sigma}} \cdot \Big[\nabla \phi_\epsilon - \frac{1}{\epsilon} \phi_\epsilon \boldsymbol{\ell} \Big] \, d\mu_\epsilon \quad \textit{and} \\ A_2^{\pm}(\boldsymbol{\sigma}) &= \, \theta_\epsilon \int_{\widehat{\partial}_\pm \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} \phi_\epsilon \, (\nabla p_\epsilon^{\boldsymbol{\sigma}} \cdot \boldsymbol{\ell}) \, d\mu_\epsilon \, .\end{aligned}$$

Proof. By the divergence theorem, we can write

$$\theta_\epsilon \int_{\mathbb{R}^d} Q_\epsilon^{\mathbf{g}} \, (\mathscr{L}_\epsilon \phi_\epsilon) \, d\mu_\epsilon \, = \, -\theta_\epsilon \, \epsilon \int_{\mathbb{R}^d} \nabla Q_\epsilon^{\mathbf{g}} \cdot \Big[\nabla \phi_\epsilon - \frac{1}{\epsilon} \phi_\epsilon \boldsymbol{\ell} \Big] \, d\mu_\epsilon \, .$$

Note that we can apply the divergence theorem since $Q_\epsilon^{\mathbf{g}}$ is continuous, while its gradient is not defined along $\partial \mathcal{E}_\epsilon^{\boldsymbol{\sigma}}$. Now, let us investigate the right-hand side. First, note that $\nabla Q_\epsilon^{\mathbf{g}} \equiv 0$ on $\mathcal{W}_i^\epsilon$ by definition. Hence, we can rewrite the right-hand side as

$$-\, \theta_\epsilon \, \epsilon \Big(\int_{\mathbb{R}^d \setminus \mathcal{K}_\epsilon} + \sum_{\boldsymbol{\sigma} \in \Sigma^*} \int_{\mathcal{E}_\epsilon^{\boldsymbol{\sigma}}} \Big) \, \nabla Q_\epsilon^{\mathbf{g}} \cdot \Big[\nabla \phi_\epsilon - \frac{1}{\epsilon} \phi_\epsilon \boldsymbol{\ell} \Big] \, d\mu_\epsilon \, . \tag{4.61}$$

Now we consider two integrals separately.

First integral of (4.61)**:** We will separately show that

$$\theta_\epsilon\, \epsilon \int_{\mathbb{R}^d \setminus \mathcal{K}_\epsilon} (\nabla Q_\epsilon^{\mathbf{g}} \cdot \nabla \phi_\epsilon)\, d\mu_\epsilon = \, o_\epsilon(1) \quad \text{and} \tag{4.62}$$

$$\theta_\epsilon \int_{\mathbb{R}^d \setminus \mathcal{K}_\epsilon} (\nabla Q_\epsilon^{\mathbf{g}} \cdot \boldsymbol{\ell})\, \phi_\epsilon\, d\mu_\epsilon = \, o_\epsilon(1)\ . \tag{4.63}$$

For (4.62), by (4.53) and the Cauchy–Schwarz inequality,

$$\Big| \theta_\epsilon\, \epsilon \int_{\mathbb{R}^d \setminus \mathcal{K}_\epsilon} \nabla Q_\epsilon^{\mathbf{g}} \cdot \nabla \phi_\epsilon\, d\mu_\epsilon \Big|^2 \, \le \, C\, \theta_\epsilon^2\, \epsilon\, \frac{1}{\eta^2}\, \mu_\epsilon(\mathbb{R}^d \setminus \mathcal{K}_\epsilon)\, \mathscr{D}_\epsilon(\phi_\epsilon)\ . \tag{4.64}$$

Note that the uniform boundedness of $\mathbf{g}$ is implicitly used here. We shall repeatedly use this boundedness in the later arguments as well. Since $U > H + \delta^2 J^2$ on $\mathbb{R}^d \setminus \mathcal{K}_\epsilon$, by (2.7) and Proposition 4.1.2,

$$\mu_\epsilon(\mathbb{R}^d \setminus \mathcal{K}_\epsilon) \, \le \, \frac{C}{Z_\epsilon}\, e^{-\frac{H+\delta^2 J^2}{\epsilon}} \, \le \, C\, \theta_\epsilon^{-1}\, \epsilon^{J^2 - \frac{d}{2}}\ . \tag{4.65}$$

Applying this and Proposition 4.3.3 to (4.64) yields

$$\Big| \theta_\epsilon\, \epsilon \int_{\mathbb{R}^d \setminus \mathcal{K}_\epsilon} (\nabla Q_\epsilon^{\mathbf{g}} \cdot \nabla \phi_\epsilon)\, d\mu_\epsilon \Big| \, \le \, \frac{C}{\eta}\, \epsilon^{\frac{1}{2}J^2 - \frac{d}{4} + \frac{1}{2}}\ .$$

Since $\eta = \epsilon^2$, we obtain (4.62) by taking J to be sufficiently large.

Now, we turn to (4.63). By (4.53), Proposition 4.3.1, (4.65) we can bound the absolute value of the left-hand side of (4.63) by $C\theta_\epsilon \eta^{-1} \mu_\epsilon(\mathbb{R}^d \setminus \mathcal{K}_\epsilon) = o_\epsilon(1)$.

Second integral of (4.61)**:** For each $\boldsymbol{\sigma} \in \Sigma_{i,j}$ with $i,\, j \in S$, we have $\nabla Q_\epsilon^{\mathbf{g}} = (\mathbf{g}(i) - \mathbf{g}(j)) \nabla p_\epsilon^{\boldsymbol{\sigma}}$ on $\mathcal{E}_\epsilon^{\boldsymbol{\sigma}}$. Therefore, we can prove that this integral is $A_1(\boldsymbol{\sigma}) + A_2^+(\boldsymbol{\sigma}) + A_2^-(\boldsymbol{\sigma})$ provided that we can prove

$$\theta_\epsilon\, \epsilon \int_{\widehat{\partial}_\pm \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}} (\nabla p_\epsilon^{\boldsymbol{\sigma}} \cdot \nabla \phi_\epsilon)\, d\mu_\epsilon \, = \, o_\epsilon(1)\ .$$

This follows from the Cauchy–Schwarz inequality, Proposition 4.3.3, and

Lemma 4.4.3. □

Based on the previous proposition, it suffices to estimate $A_1(\boldsymbol{\sigma})$ and $A_2^{\pm}(\boldsymbol{\sigma})$ for each $\boldsymbol{\sigma} \in \Sigma^*$. These estimates are carried out via the following proposition.

Proposition 4.5.2. *For $i, j \in S$ with $i < j$ and $\boldsymbol{\sigma} \in \Sigma_{i,j}$, we have*

$$A_1(\boldsymbol{\sigma}) \;=\; \frac{\lambda_1^{\boldsymbol{\sigma}}}{2\pi\nu_\star\sqrt{-\det \mathbb{H}^{\boldsymbol{\sigma}}}}\left(\mathbf{m}_\epsilon(j) - \mathbf{m}_\epsilon(i)\right) + o_\epsilon(1)\;, \tag{4.66}$$

and

$$A_2^+(\boldsymbol{\sigma}) \;=\; -\frac{\lambda_1^{\boldsymbol{\sigma}}}{2\pi\nu_\star\sqrt{-\det \mathbb{H}^{\boldsymbol{\sigma}}}}\,\frac{\left(\mathbb{L}^{\boldsymbol{\sigma}}\,(\mathbb{H}^{\boldsymbol{\sigma}})^{-1}\,\boldsymbol{v}^{\boldsymbol{\sigma}}\right)\cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}}}{\boldsymbol{v}^{\boldsymbol{\sigma}}\cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}}}\,\mathbf{m}_\epsilon(i) + o_\epsilon(1)\;, \tag{4.67}$$

$$A_2^-(\boldsymbol{\sigma}) \;=\; +\frac{\lambda_1^{\boldsymbol{\sigma}}}{2\pi\nu_\star\sqrt{-\det \mathbb{H}^{\boldsymbol{\sigma}}}}\,\frac{\left(\mathbb{L}^{\boldsymbol{\sigma}}\,(\mathbb{H}^{\boldsymbol{\sigma}})^{-1}\,\boldsymbol{v}^{\boldsymbol{\sigma}}\right)\cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}}}{\boldsymbol{v}^{\boldsymbol{\sigma}}\cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}}}\,\mathbf{m}_\epsilon(j) + o_\epsilon(1)\;. \tag{4.68}$$

Before proving this proposition, we conclude the demonstration of Proposition 4.3.7 by assuming this proposition.

Proof of Proposition 4.3.7. First, we check that

$$\begin{aligned}&\left(\boldsymbol{v}^{\boldsymbol{\sigma}} + \mathbb{L}^{\boldsymbol{\sigma}}\,(\mathbb{H}^{\boldsymbol{\sigma}})^{-1}\,\boldsymbol{v}^{\boldsymbol{\sigma}}\right)\cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}} \;=\; \left(\mathbb{I} - (\mathbb{H}^{\boldsymbol{\sigma}})^{-1}\,(\mathbb{L}^{\boldsymbol{\sigma}})^{\dagger}\right)\boldsymbol{v}^{\boldsymbol{\sigma}}\cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}}\\ &\;=\; (\mathbb{H}^{\boldsymbol{\sigma}})^{-1}\left(\mathbb{H}^{\boldsymbol{\sigma}} - (\mathbb{L}^{\boldsymbol{\sigma}})^{\dagger}\right)\boldsymbol{v}^{\boldsymbol{\sigma}}\cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}} \;=\; -\mu^{\boldsymbol{\sigma}}\,(\mathbb{H}^{\boldsymbol{\sigma}})^{-1}\,\boldsymbol{v}^{\boldsymbol{\sigma}}\cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}} \;=\; \frac{\mu^{\boldsymbol{\sigma}}}{\lambda_1^{\boldsymbol{\sigma}}}\left(\boldsymbol{v}^{\boldsymbol{\sigma}}\cdot \boldsymbol{e}_1^{\boldsymbol{\sigma}}\right)\;,\end{aligned}$$

where the first identity follows from the fact that the matrix $\mathbb{H}^{\boldsymbol{\sigma}}\mathbb{L}^{\boldsymbol{\sigma}}$ is a skew-symmetric matrix by Lemma 2.2.6, and the last identity follows from the fact that $\boldsymbol{e}_1^{\boldsymbol{\sigma}}$ is the eigenvector of $\mathbb{H}^{\boldsymbol{\sigma}}$ associated with the eigenvalue $-\lambda_1^{\boldsymbol{\sigma}}$. We can combine this computation with Proposition 4.5.2 to get

$$A_1(\boldsymbol{\sigma}) + A_2^+(\boldsymbol{\sigma}) + A_2^-(\boldsymbol{\sigma}) \;=\; \frac{\omega^{\boldsymbol{\sigma}}}{\nu_\star}\left(\mathbf{m}_\epsilon(j) - \mathbf{m}_\epsilon(i)\right) + o_\epsilon(1)\;, \tag{4.69}$$

where the Eyring–Kramers constant $\omega^{\boldsymbol{\sigma}}$ is defined in (4.6), and we implicitly used (4.30). Inserting (4.69) into Proposition 4.5.1 completes the proof. □

Now, it remains to prove Proposition 4.5.2. We provide the estimates of $A_1(\boldsymbol{\sigma})$ and $A_2^{\pm}(\boldsymbol{\sigma})$ in Sections 4.5.3 and 4.5.4, respectively.

4.5.2 Change of coordinates on $\partial_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$

Hereafter, it suffices to focus only on a single saddle point $\boldsymbol{\sigma}$; hence, in the remainder of the current section, we recall Notation 4.4.1 and use the following conventions: *we fix $\boldsymbol{\sigma} \in \Sigma_{i,j}$ for some $i, j \in S$ with $i < j$, assume that $\boldsymbol{\sigma} = \mathbf{0}$ for simplicity of notation, and drop the superscript $\boldsymbol{\sigma}$ from the notations, e.g., we write p_ϵ and $\mathcal{B}_\epsilon$ instead of $p_\epsilon^{\boldsymbol{\sigma}}$ and $\mathcal{B}_\epsilon^{\boldsymbol{\sigma}}$, respectively.*

Before proceeding to the proof of Proposition 4.5.2, we recall in this subsection a change of coordinate introduced in Section 3.4.4, which maps $\partial_+\mathcal{B}_\epsilon$ to a subset of $\mathbb{R}^{d-1}$. For $\mathbb{A} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{u} = (u_1, \dots, u_d) \in \mathbb{R}^d$, we define $\widetilde{\mathbb{A}} \in \mathbb{R}^{(d-1)\times(d-1)}$ and $\widetilde{\boldsymbol{u}} \in \mathbb{R}^{d-1}$ as

$$\widetilde{\mathbb{A}} := (\mathbb{A}_{i,j})_{2\le i,j\le d} \quad \text{and} \quad \widetilde{\boldsymbol{u}} := (u_2, \dots, u_d)\,, \tag{4.70}$$

respectively. Define a vector $\boldsymbol{\gamma} = (\gamma_2, \dots, \gamma_d) \in \mathbb{R}^{d-1}$ by $\gamma_k = \frac{v_k}{v_1} \cdot \frac{\lambda_1^{1/2}}{\lambda_k} J\delta$ for $2 \le k \le d$, where $\boldsymbol{v} = \boldsymbol{v}^{\boldsymbol{\sigma}} = (v_1, \dots, v_d)$ denotes the eigenvector introduced in Section 4.4.3, at which it has been mentioned that $v_1 = \boldsymbol{v} \cdot \boldsymbol{e} \neq 0$. Define

$$\mathcal{P}_\delta := \left\{ \boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d : x_1 = \frac{J\delta}{\lambda_1^{1/2}} \right\} \subset \mathbb{R}^d\,,$$

so that $\partial_+\mathcal{B}_\epsilon,\ \partial_+\mathcal{C}_\epsilon \subset \mathcal{P}_\delta$, and define a map $\Pi_\epsilon : \mathcal{P}_\delta \to \mathbb{R}^{d-1}$ by $\Pi_\epsilon(\boldsymbol{x}) = \widetilde{\boldsymbol{x}} + \boldsymbol{\gamma}$. This maps the change of coordinate from $\partial_+\mathcal{B}_\epsilon$ to $\mathbb{R}^{d-1}$, which simplifies computations significantly. For instance, we have the following result.

Lemma 4.5.3. *For all $\boldsymbol{x} \in \partial_+\mathcal{B}_\epsilon$, we have*

$$\boldsymbol{x} \cdot (\mathbb{H} + \mu\, \boldsymbol{v} \otimes \boldsymbol{v})\, \boldsymbol{x} \;=\; \Pi_\epsilon(\boldsymbol{x}) \cdot (\widetilde{\mathbb{H}} + \mu\, \widetilde{\boldsymbol{v}} \otimes \widetilde{\boldsymbol{v}})\, \Pi_\epsilon(\boldsymbol{x})\,.$$

In addition, the matrix $\widetilde{\mathbb{H}} + \mu\, \widetilde{\boldsymbol{v}} \otimes \widetilde{\boldsymbol{v}}$ *is positive definite and*

$$\det\big(\widetilde{\mathbb{H}} + \mu\, \widetilde{\boldsymbol{v}} \otimes \widetilde{\boldsymbol{v}}\big) \;=\; (\boldsymbol{v} \cdot \boldsymbol{e}_1)^2\, \frac{\mu}{\lambda_1} \prod_{k=2}^{d} \lambda_k\ .$$

Proof. We refer to Lemmas 3.4.7 and 3.4.9 for the proof. □

In Lemma 3.4.8, it has been verified that the image of $\Pi_\epsilon(\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}})$ is comparable with a ball in $\mathbb{R}^{d-1}$ centered at the origin with radius of order δ. In the next lemma, we slightly strengthen this result. Recall the definition of $\widehat{\mathcal{W}}_i$ from (4.29).

Lemma 4.5.4. *There exist constants* r, $R > 0$ *such that*

$$\mathcal{D}_{r\delta}^{(d-1)} \;\subset\; \Pi_\epsilon(\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \cap \widehat{\mathcal{W}}_i) \;\subset\; \Pi_\epsilon(\partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}}) \;\subset\; \mathcal{D}_{R\delta}^{(d-1)}\ ,$$

where $\mathcal{D}_a^{(d-1)}$ *represents a ball in* $\mathbb{R}^{d-1}$ *of radius* a *centered at the origin.*

Proof. In view of Lemma 3.4.8, it suffices to show the first inclusion. Define

$$\begin{aligned} \mathcal{P}_\delta \;&:=\; \Big\{\, \boldsymbol{x} = (x_1,\, \dots,\, x_d) \in \mathbb{R}^d : x_1 = \frac{J\delta}{\lambda_1^{1/2}} \,\Big\} \subset \mathbb{R}^d \quad \text{and} \\ \overline{\boldsymbol{\gamma}} \;&:=\; \Big(\, \frac{J\delta}{\lambda_1^{1/2}},\, -\gamma_2,\, \cdots,\, -\gamma_d \Big) \in \mathcal{P}_\delta\ . \end{aligned}$$

Then, it has been shown in (3.38) that

$$U(\overline{\boldsymbol{\gamma}}) \;=\; H - \frac{\lambda_1}{2\,\mu\, v_1^2}\, J^2\, \delta^2 + O(\delta^3) \;<\; H - 2\, c_0\, J^2\, \delta^2 \tag{4.71}$$

for all sufficiently small $\epsilon > 0$ if we take $c_0 > 0$ as a sufficiently small constant. By inserting this c_0 into the definition of $\widehat{\mathcal{W}}_i$ in (4.29), we can find sufficiently small $r > 0$ such that $\mathcal{D}_{r\delta}(\overline{\boldsymbol{\gamma}}) \cap \mathcal{P}_\delta \subset \partial_+ \mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \cap \widehat{\mathcal{W}}_i$. Since $\Pi_\epsilon(\overline{\boldsymbol{\gamma}}) = \mathbf{0}$, we have $\mathcal{D}_{r\delta}^{(d-1)} = \Pi_\epsilon(\mathcal{D}_{r\delta}(\overline{\boldsymbol{\gamma}}) \cap \mathcal{P}_\delta)$ and the proof is completed. □

4.5.3 Estimate of $A_1(\boldsymbol{\sigma})$

Now, we prove (4.66) of Proposition 4.5.2. By the divergence theorem, we can write

$$A_1(\boldsymbol{\sigma}) \;=\; -\theta_\epsilon\, \epsilon \int_{\partial\mathcal{B}_\epsilon} \phi_\epsilon\, [\, \nabla p_\epsilon \cdot \boldsymbol{n}_{\mathcal{B}_\epsilon}\,]\, \sigma(d\mu_\epsilon) + \theta_\epsilon \int_{\mathcal{B}_\epsilon} (\mathscr{L}_\epsilon^* p_\epsilon)\, \phi_\epsilon\, d\mu_\epsilon \;.$$

By Propositions 4.3.1 and 4.4.2, we can observe that the second term at the right-hand side is $o_\epsilon(1)$. Therefore, we can write

$$A_1(\boldsymbol{\sigma}) \;=\; K_0 + K_+ + K_- + o_\epsilon(1)\;, \tag{4.72}$$

where

$$\begin{aligned} K_0 \;&=\; -\theta_\epsilon\, \epsilon \int_{\partial_0\mathcal{B}_\epsilon} \phi_\epsilon\, [\, \nabla p_\epsilon \cdot \boldsymbol{n}_{\mathcal{B}_\epsilon}\,]\, \sigma(d\mu_\epsilon) \quad \text{and} \\ K_\pm \;&=\; -\theta_\epsilon\, \epsilon \int_{\partial_\pm\mathcal{B}_\epsilon} \phi_\epsilon\, [\, \nabla p_\epsilon \cdot \boldsymbol{n}_{\mathcal{B}_\epsilon}\,]\, \sigma(d\mu_\epsilon)\;. \end{aligned}$$

First, we show that $K_0 = o_\epsilon(1)$. For $\boldsymbol{x} \in \partial_0\mathcal{B}_\epsilon$, by (4.48) and (4.49),

$$|\, \nabla p_\epsilon(\boldsymbol{x}) \cdot \boldsymbol{n}_{\mathcal{B}_\epsilon}(\boldsymbol{x})\,| \;=\; \Big|\, \frac{1}{c_\epsilon^{\boldsymbol{\sigma}}}\, e^{-\frac{\mu}{2\epsilon}(\boldsymbol{x}\cdot\boldsymbol{v})^2}\, \boldsymbol{v} \cdot \boldsymbol{n}_{\mathcal{B}_\epsilon}(\boldsymbol{x}) \Big| \;\le\; \frac{C}{\epsilon^{1/2}}\;.$$

Therefore, by Proposition 4.3.1 and (4.44) along with the fact that $\partial_0\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \subset \partial_0\mathcal{E}_\epsilon^{\boldsymbol{\sigma}}$, we have

$$|\, K_0\,| \;\le\; C\,\theta_\epsilon\, \epsilon^{1/2}\, Z_\epsilon^{-1}\, e^{-(H+J^2\delta^2)/\epsilon}\, \sigma(\partial_0\mathcal{B}_\epsilon) \;\le\; C\, \epsilon^{J^2-(d+1)/2}\, \delta^{d-1} \;=\; o_\epsilon(1) \tag{4.73}$$

for sufficiently large J, where we used $\sigma(\partial_0\mathcal{B}_\epsilon) = O(\delta^{d-1})$ in the second inequality. Next, we estimate K_+ and K_-.

Lemma 4.5.5. *We have*

$$K_+ = -\frac{\lambda_1}{2\pi\nu_\star\sqrt{-\det\mathbb{H}}}\,\mathbf{m}_\epsilon(i) + o_\epsilon(1) \quad \textit{and}$$
$$K_- = \frac{\lambda_1}{2\pi\nu_\star\sqrt{-\det\mathbb{H}}}\,\mathbf{m}_\epsilon(j) + o_\epsilon(1)\ .$$

Proof. We only prove the estimate for K_+ since the proof for K_- is identical. Since $\boldsymbol{n}_{\mathcal{B}_\epsilon} = \boldsymbol{e}_1$ on $\partial_+\mathcal{B}_\epsilon$, by the Taylor expansion of U around $\boldsymbol{\sigma}$, explicit formula (4.48) for p_ϵ, and (4.5), we can write

$$K_+ \;=\; -[\,1+o_\epsilon(1)\,]\,\frac{\epsilon\,\mu^{1/2}\,(\boldsymbol{v}\cdot\boldsymbol{e}_1)}{(2\pi\epsilon)^{(d+1)/2}\,\nu_\star}\int_{\partial_+\mathcal{B}_\epsilon} e^{-\frac{1}{2\epsilon}\boldsymbol{x}\cdot(\mathbb{H}+\mu\,\boldsymbol{v}\otimes\boldsymbol{v})\,\boldsymbol{x}}\,\phi_\epsilon(\boldsymbol{x})\,\sigma(d\boldsymbol{x})\ .$$

With the notations introduced in Section 4.5.2, we perform the change of variable $\boldsymbol{y} = \Pi_\epsilon(\boldsymbol{x})$ in the previous integral to deduce that

$$K_+ \;=\; -[\,1+o_\epsilon(1)\,]\,\frac{\epsilon\,\mu^{1/2}\,(\boldsymbol{v}\cdot\boldsymbol{e}_1)}{(2\pi\epsilon)^{(d+1)/2}\,\nu_\star}\int_{\Pi_\epsilon(\partial_+\mathcal{B}_\epsilon)} e^{-\frac{1}{2\epsilon}\boldsymbol{y}\cdot(\widetilde{\mathbb{H}}+\mu\,\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}})\,\boldsymbol{y}}\,\phi_\epsilon(\Pi_\epsilon^{-1}(\boldsymbol{y}))\,d\boldsymbol{y}\ , \tag{4.74}$$

where we applied Lemma 4.5.3 to the exponential term. Let $r > 0$ be the constant appearing in Lemma 4.5.4. By Proposition 4.3.4 and Lemma 4.5.4, we have $\phi_\epsilon(\Pi_\epsilon^{-1}(\boldsymbol{y})) = \mathbf{m}_\epsilon(i) + o_\epsilon(1)$ for $\boldsymbol{y} \in \mathcal{D}_{r\delta}^{(d-1)}$. Thus, we have

$$\int_{\mathcal{D}_{r\delta}^{(d-1)}(\mathbf{0})} e^{-\frac{1}{2\epsilon}\boldsymbol{y}\cdot(\widetilde{\mathbb{H}}+\mu\,\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}})\,\boldsymbol{y}}\,\phi_\epsilon(\Pi_\epsilon^{-1}(\boldsymbol{y}))\,d\boldsymbol{y}$$
$$= \frac{(2\pi\epsilon)^{(d-1)/2}}{\sqrt{\det(\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}})}}\,[\,\mathbf{m}_\epsilon(i)+o_\epsilon(1)\,]\ .$$

Since the integral on $\Pi_\epsilon(\partial_+\mathcal{B}_\epsilon)\setminus\mathcal{D}_{r\delta}^{(d-1)}(\mathbf{0}) \subset \mathbb{R}^{d-1}\setminus\mathcal{D}_{r\delta}^{(d-1)}(\mathbf{0})$ is $o_\epsilon(1)$ by Proposition 4.3.1, we can conclude from (4.74) that

$$K_+ \;=\; -\,[\,1+o_\epsilon(1)\,]\,\frac{\mu^{1/2}\,(\boldsymbol{v}\cdot\boldsymbol{e}_1)}{2\pi\nu_\star\sqrt{\det(\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}})}}\,\mathbf{m}_\epsilon(i)+o_\epsilon(1)\ .$$

The proof is completed by the second part of Lemma 4.5.3. □

Now, (4.66) can be obtained by combining (4.72), (4.73), and Lemma 4.5.5.

4.5.4 Estimate of $A_2^{\pm}(\boldsymbol{\sigma})$

Now, we estimate $A_2^{+}(\boldsymbol{\sigma})$ and $A_2^{-}(\boldsymbol{\sigma})$. Since the proof is identical, it suffices to consider $A_2^{+}(\boldsymbol{\sigma})$, i.e., (4.67). Write $\boldsymbol{\ell} = (\ell_1, \dots, \ell_d)$ and $\boldsymbol{v} = (v_1, \dots, v_d)$. Then, by (4.56) and (4.57), we can write

$$A_2^{+}(\boldsymbol{\sigma}) \;=\; M_1 + M_2 \;, \tag{4.75}$$

where

$$\begin{aligned} M_1 \;&=\; \frac{\theta_\epsilon}{\eta C_\epsilon} \int_{\widehat{\partial}_+\mathcal{B}_\epsilon} \phi_\epsilon(\boldsymbol{x}) \Big[\, \boldsymbol{x}\cdot\boldsymbol{e}_1 - \frac{J\delta}{\lambda_1^{1/2}} - \eta \,\Big]\, e^{-\frac{\mu}{2\epsilon}(\overline{\boldsymbol{x}}\cdot\boldsymbol{v})^2} \sum_{k=2}^{d} v_k\, \ell_k(\boldsymbol{x})\, \mu_\epsilon(d\boldsymbol{x}) \;, \\ M_2 \;&=\; \theta_\epsilon \int_{\widehat{\partial}_+\mathcal{B}_\epsilon} \phi_\epsilon(\boldsymbol{x})\, \frac{1-p_\epsilon(\overline{\boldsymbol{x}})}{\eta}\, \ell_1(\boldsymbol{x})\, \mu_\epsilon(d\boldsymbol{x}) \;. \end{aligned}$$

First, we show that M_1 is negligible.

Lemma 4.5.6. *We have that* $M_1 = o_\epsilon(1)$.

Proof. Since $\Big|\boldsymbol{x}\cdot\boldsymbol{e}_1 - \dfrac{J\delta}{\lambda_1^{1/2}} - \eta\Big| \le \eta$ for $\boldsymbol{x} \in \widehat{\partial}_+\mathcal{B}_\epsilon$, by Proposition 4.3.1, it suffices to prove that

$$\frac{\theta_\epsilon}{C_\epsilon}\,\frac{1}{Z_\epsilon} \int_{\widehat{\partial}_+\mathcal{C}_\epsilon} e^{-\frac{\mu}{2\epsilon}(\overline{\boldsymbol{x}}\cdot\boldsymbol{v})^2}\, e^{-\frac{U(\boldsymbol{x})}{\epsilon}}\, d\boldsymbol{x} \;=\; o_\epsilon(1) \;. \tag{4.76}$$

By applying $U(\boldsymbol{x}) = U(\overline{\boldsymbol{x}}) + O(\epsilon^2)$ and then applying the Taylor expansion to $U(\overline{\boldsymbol{x}})$ (with respect to $\boldsymbol{\sigma} = \boldsymbol{0}$), the left-hand side of the above equality can be bounded from above by

$$\frac{C}{\epsilon^{(d+1)/2}} \int_{\widehat{\partial}_+\mathcal{C}_\epsilon} e^{-\frac{1}{2\epsilon}\overline{\boldsymbol{x}}\cdot(\mathbb{H}+\mu\,\boldsymbol{v}\otimes\boldsymbol{v})\overline{\boldsymbol{x}}}\, d\boldsymbol{x} \;=\; \frac{C\eta}{\epsilon^{(d+1)/2}} \int_{\partial_+\mathcal{C}_\epsilon} e^{-\frac{1}{2\epsilon}\boldsymbol{x}\cdot\,(\mathbb{H}+\mu\,\boldsymbol{v}\otimes\boldsymbol{v})\,\boldsymbol{x}}\, \sigma(d\boldsymbol{x})\ .$$

Using the change of variable $\boldsymbol{y} = \Pi_\epsilon(\boldsymbol{x})$ and applying Lemma 4.5.3, we can check that the last integral is bounded by $C\epsilon^{(d-1)/2}$. Hence, the left-hand side of (4.76) is bounded from above by $\dfrac{C\eta}{\epsilon^{(d+1)/2}} \times C\epsilon^{(d-1)/2} = o_\epsilon(1)$. This proves the lemma. □

Next, we estimate M_2.

Lemma 4.5.7. *We have*

$$M_2 \;=\; -\frac{1}{2\pi\nu_\star}\sqrt{\frac{\mu}{\det(\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}})}}\,[\,(\mathbb{L}\,\mathbb{H}^{-1}\,\boldsymbol{v})\cdot\boldsymbol{e}_1\,]\,\mathbf{m}_\epsilon(i) + o_\epsilon(1)\ . \tag{4.77}$$

Proof. Let $a \in (0,\, a_0)$, where a_0 is the constant appearing in Lemma 4.4.4. Let us define

$$\widehat{\partial}_+^{2,\,a}\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \;:=\; \widehat{\partial}_+\mathcal{B}_\epsilon^{\boldsymbol{\sigma}} \,\cap\, \widehat{\partial}_+^{2,\,a}\mathcal{C}_\epsilon^{\boldsymbol{\sigma}}$$

and write

$$M_2 \;=\; M_{2,\,1} + M_{2,\,2}\ , \tag{4.78}$$

where

$$\begin{aligned} M_{2,\,1} \;&=\; \theta_\epsilon \int_{\widehat{\partial}_+^{2,\,a}\mathcal{B}_\epsilon} \phi_\epsilon(\boldsymbol{x})\frac{1-p_\epsilon(\overline{\boldsymbol{x}})}{\eta}\ell_1(\boldsymbol{x})\,\mu_\epsilon(\boldsymbol{x})\,d\boldsymbol{x}\ , \text{ and} \\ M_{2,\,2} \;&=\; \theta_\epsilon \int_{\widehat{\partial}_+\mathcal{B}_\epsilon\setminus\widehat{\partial}_+^{2,\,a}\mathcal{B}_\epsilon} \phi_\epsilon(\boldsymbol{x})\frac{1-p_\epsilon(\overline{\boldsymbol{x}})}{\eta}\ell_1(\boldsymbol{x})\,\mu_\epsilon(\boldsymbol{x})\,d\boldsymbol{x}\ . \end{aligned}$$

First, we check that $M_{2,\,1} = o_\epsilon(1)$. By Proposition 4.3.1, it suffices to show that $\dfrac{\theta_\epsilon}{\eta}\mu_\epsilon(\widehat{\partial}_+^{2,\,a}\mathcal{B}_\epsilon) = o_\epsilon(1)$. This is a consequence of the bound $U(\boldsymbol{x}) \geq H + aJ^2\delta^2$ on $\widehat{\partial}_+^{2,\,a}\mathcal{B}_\epsilon$, which holds by definition, the bound $\mathrm{vol}(\widehat{\partial}_+^{2,\,a}\mathcal{B}_\epsilon) \leq \mathrm{vol}(\widehat{\partial}_+\mathcal{B}_\epsilon) \leq C\eta\delta^{d-1}$, and (4.5).

Now, we turn to $M_{2,2}$. By Lemma 4.4.4, we have $\overline{\boldsymbol{x}} \cdot \boldsymbol{v} \geq cJ\delta$ for $\boldsymbol{x} \in \widehat{\partial}_+\mathcal{B}_\epsilon \setminus \widehat{\partial}_+^{2,a}\mathcal{B}_\epsilon$; hence, we can use the elementary inequality

$$\frac{b}{b^2+1}e^{-b^2/2} \leq \int_b^\infty e^{-t^2/2}dt \leq \frac{1}{b}e^{-b^2/2} \quad \text{for } b>0$$

to obtain

$$1-p_\epsilon(\overline{\boldsymbol{x}}) = [\,1+o_\epsilon(1)\,]\,\frac{\epsilon^{1/2}}{(2\pi\mu)^{1/2}\,(\overline{\boldsymbol{x}}\cdot\boldsymbol{v})}\,e^{-\frac{\mu}{2\epsilon}(\overline{\boldsymbol{x}}\cdot\boldsymbol{v})^2}\;.$$

Now, we apply this result along with the Taylor expansions of U and $\boldsymbol{\ell}$ around $\boldsymbol{\sigma}=\boldsymbol{0}$ to $M_{2,2}$ to get

$$\begin{aligned}&M_{2,2}\\&=\frac{1+o_\epsilon(1)}{(2\pi)^{(d+1)/2}\,\nu_\star\,\mu^{1/2}\,\epsilon^{(d+3)/2}}\int_{\widehat{\partial}_+\mathcal{B}_\epsilon\setminus\widehat{\partial}_+^{2,a}\mathcal{B}_\epsilon}\phi_\epsilon(\boldsymbol{x})\,\frac{\mathbb{L}\overline{\boldsymbol{x}}\cdot\boldsymbol{e}_1}{\overline{\boldsymbol{x}}\cdot\boldsymbol{v}}\,e^{-\frac{1}{2\epsilon}\overline{\boldsymbol{x}}\cdot(\mathbb{H}+\mu\,\boldsymbol{v}\otimes\boldsymbol{v})\,\overline{\boldsymbol{x}}}\,d\boldsymbol{x}\;.\end{aligned}$$

Note here that we have replaced several $\boldsymbol{x}$'s with $\overline{\boldsymbol{x}}$'s without changing the error term since $|\overline{\boldsymbol{x}}-\boldsymbol{x}| = O(\eta)$. Let $r>0$ be the constant appearing in Lemma 4.5.4. Then, we claim that, for all sufficiently small $\epsilon>0$,

$$\Big(\frac{J\delta}{\lambda_1^{1/2}},\,\frac{J\delta}{\lambda_1^{1/2}}+\eta\Big]\times\Pi_\epsilon^{-1}(\mathcal{D}_{r\delta/2}(\boldsymbol{0}))\subset\widehat{\partial}_+\mathcal{B}_\epsilon\cap\widehat{\mathcal{W}}_i\subset\widehat{\partial}_+\mathcal{B}_\epsilon\setminus\widehat{\partial}_+^{2,a}\mathcal{B}_\epsilon\;.$$

The second inclusion is immediate from the definitions of $\widehat{\mathcal{W}}_i$ and $\widehat{\partial}_+^{2,a}\mathcal{B}_\epsilon$. On the other hand, the first inclusion is a consequence of Lemma 4.5.4 and the fact that $\eta=\epsilon^2$. For convenience, we write

$$\mathcal{A}_\epsilon = \Big(\frac{J\delta}{\lambda_1^{1/2}},\,\frac{J\delta}{\lambda_1^{1/2}}+\eta\Big]\times\Pi_\epsilon^{-1}(\mathcal{D}_{r\delta/2}(\boldsymbol{0}))\;.$$

We further decompose $M_{2,2} = M_{2,2,1}+M_{2,2,2}$ where $M_{2,2,1}$ and $M_{2,2,2}$ are obtained from $M_{2,2}$ by replacing the integral $\int_{\widehat{\partial}_+\mathcal{B}_\epsilon\setminus\widehat{\partial}_+^{2,a}\mathcal{B}_\epsilon}$ with

$\displaystyle\int_{(\widehat{\partial}_+\mathcal{B}_\epsilon \setminus \widehat{\partial}_+^{2,\,a}\mathcal{B}_\epsilon)\setminus \mathcal{A}_\epsilon}$ and $\displaystyle\int_{\mathcal{A}_\epsilon}$, respectively. We argue that $M_{2,\,2,\,1} = o_\epsilon(1)$. By Proposition 4.3.1 and the fact that $\overline{\boldsymbol{x}} \cdot \boldsymbol{v} \geq aJ\delta$ on $\widehat{\partial}_+\mathcal{B}_\epsilon \setminus \widehat{\partial}_+^{2,\,a}\mathcal{B}_\epsilon$, it suffices to show that

$$\int_{\mathbb{R}^{d-1}\setminus \Pi_\epsilon^{-1}(\mathcal{D}_{r\delta/2}(\mathbf{0}))} e^{-\frac{1}{2\epsilon}\overline{\boldsymbol{x}}\cdot\,(\mathbb{H}+\mu\,\boldsymbol{v}\otimes\boldsymbol{v})\,\overline{\boldsymbol{x}}}\,d\widetilde{\boldsymbol{x}} \;=\; \epsilon^{(d-1)/2}\,o_\epsilon(1)\;,$$

where $\widetilde{\boldsymbol{x}}$ is defined in (4.70). The previous identity can be directly verified by the change of variable $\boldsymbol{y} = \Pi_\epsilon(\overline{\boldsymbol{x}})$.

Next, we turn to $M_{2,\,2,\,2}$. On $\mathcal{A}_\epsilon$, we have $\phi_\epsilon(\boldsymbol{x}) = \mathbf{m}_\epsilon(i) + o_\epsilon(1)$ by Proposition 4.3.4. Hence, we can write

$$\begin{aligned} M_{2,\,2,\,2} \;&=\; \frac{1+o_\epsilon(1)}{(2\pi)^{(d+1)/2}\,\nu_\star\,\mu^{1/2}\,\epsilon^{(d-1)/2}}\,[\,\mathbf{m}_\epsilon(i)+o_\epsilon(1)\,] \\ &;\times \int_{\Pi_\epsilon^{-1}(\mathcal{D}_{r\delta/2}(\mathbf{0}))} \frac{\mathbb{L}\overline{\boldsymbol{x}}\cdot\boldsymbol{e}_1}{\overline{\boldsymbol{x}}\cdot\boldsymbol{v}}\, e^{-\frac{1}{2\epsilon}\overline{\boldsymbol{x}}\cdot\,(\mathbb{H}+\mu\,\boldsymbol{v}\otimes\boldsymbol{v})\,\overline{\boldsymbol{x}}}\,d\widetilde{\boldsymbol{x}}\;. \end{aligned} \tag{4.79}$$

We can use Lemma 3.4.11 to show that the last integral can be written as

$$[\,1+o_\epsilon(1)\,]\,\frac{(2\pi\epsilon)^{(d-1)/2}\,(-\mu\,\mathbb{L}\mathbb{H}^{-1}\boldsymbol{v})\cdot\boldsymbol{e}_1}{\sqrt{\det{(\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}})}}}\;.$$

Inserting this into (4.79) along with the fact that $M_{2,\,2,\,1} = o_\epsilon(1)$ proves that

$$M_{2,\,2} \;=\; -\frac{1}{2\pi\nu_\star}\sqrt{\frac{\mu}{\det(\widetilde{\mathbb{H}}+\mu\widetilde{\boldsymbol{v}}\otimes\widetilde{\boldsymbol{v}})}}\,[\,(\mathbb{L}\mathbb{H}^{-1}\boldsymbol{v})\cdot\boldsymbol{e}_1\,]\,\mathbf{m}_\epsilon(i)+o_\epsilon(1)\;. \tag{4.80}$$

Combining this estimate with the fact that $M_{2,\,1} = o_\epsilon(1)$ completes the proof. □

Now, the proof of (4.67) follows immediately from (4.75) and Lemmas 4.5.3, 4.5.6, 4.5.7. This concludes the proof of Proposition 4.3.7.

Chapter 5

Curie–Weiss–Potts model

In this chapter, we completely analyze the energy landscape of the Curie–Weiss–Potts model for all $q \geq 4$. Based on this result, we prove metastable behavior of heat-bath Glauber dynamics associated with the Curie–Weiss–Potts model. The model exhibits phase transitions as an inverse temperature β varies so that different aspects of metastable behavior are observed. Our result reveals that different phase transitions are observed in the case of $q \leq 4$ and the case of $q \geq 5$ because the number of critical temperatures is different.

5.1 Studies on the Curie–Weiss–Potts model

The Curie–Weiss–Potts model is investigated in various studies; e.g., [6, 12, 18, 19, 25, 26, 44, 53, 85, 88] and references therein. We note that the rigorous mathematical definition of the Curie–Weiss–Potts model is presented in the next section.

The Curie–Weiss model

The Ising case of the Curie–Weiss–Potts model, i.e., the corresponding spin system consisting only of $q = 2$ spins, is the famous Curie–Weiss model. It is well-known that the Curie–Weiss model without an external field exhibits a phase transition at the critical (inverse) temperature $\beta_c > 0$. It is mainly because the number of global minima of the potential function associated with the empirical magnetization is one for the high temperature regime $\beta \leq \beta_c$ while it becomes two for the low temperature regime $\beta > \beta_c$, where $\beta > 0$ represents the inverse temperature (cf. [80, Chapter 9] for more detail). It is also well-known that such a phase transition for the structure of the energy landscape is closely related to the mixing property of the associated heat-bath Glauber dynamics. In [66], it has been shown that the Glauber dynamics exhibits the so-called cut-off phenomenon which is a signature of the fast mixing for the high-temperature regime (i.e., $\beta < \beta_c$) and the metastability for the low-temperature regime (i.e., $\beta > \beta_c$). The metastability for the low-temperature regime has been more deeply investigated in [17].

The Curie–Weiss–Potts model with $q = 3$

The picture for the Curie–Weiss model explained above has been fully extended to the Curie–Weiss–Potts model consisting of $q = 3$ spins. The complete description of the energy landscape has been obtained recently in [44, 53], where three critical temperatures

$$0 < \beta_1 < \beta_2 < \beta_3 = 3$$

are characterized. More precisely, it has been shown that the potential function associated with the empirical magnetization (which will be explained in detail in section 5.2.3) has

- the unique global minimum for $\beta \in (0, \beta_1)$,
- one global minimum and three local minima for $\beta \in (\beta_1, \beta_2)$,
- three global minima and one local minimum for $\beta \in (\beta_2, \beta_3)$, and
- three global minima for $\beta \in (\beta_3, \infty)$.

The articles [44, 53] also analyzed the associated saddle structure. Based on this analysis, [53] discussed the quantitative feature of the metastable behavior of the heat-bath Glauber dynamics in view of the Eyring–Kramers formula and Markov chain model reduction (cf. [2, 3, 46]) for all the low-temperature regime $\beta > \beta_1$. Because of the abrupt change in the structure of the potential function at $\beta = \beta_2$ and $\beta = \beta_3$, the metastable behaviors of the Glauber dynamics in three low-temperature regimes (β_1, β_2), (β_2, β_3), and (β_3, ∞) turned out to be both quantitatively and qualitatively different. For the high-temperature regime $(0, \beta_1)$, the cut-off phenomenon has been verified in [19] for all $q \geq 3$. Adjoining all these works completes the picture for the Curie–Weiss–Potts model with $q = 3$ spins.

The Curie–Weiss–Potts model with $q \geq 4$

Compared to the Curie–Weiss–Potts model with $q = 2$ or 3 spins, the analysis of the case with $q \geq 4$ spins is not completed so far. In many literature, two critical temperatures $\beta_1(q) < \beta_2(q)$ for the Curie–Weiss–Potts model with $q \geq 4$ spins are observed and the phase transitions near these critical temperatures have been analyzed. For instance, in [19], the phase transition from the fast mixing (the cut-off phenomenon) to the slow mixing (due to the appearance of new local minima) at $\beta = \beta_1(q)$ has been confirmed. In [26], it has been observed that the limiting distributions of the empirical magnetization exhibits the abrupt change at $\beta = \beta_2(q)$. In [18], the phase transition around $\beta_2(q)$ also has been studied in view of the equivalence and non-equivalence of ensembles.

These studies focus on the phase transitions involved with the local and the global minima of the potential function. However, in order to investigate the metastable behavior whose main objective is to analyze the transitions between neighborhoods of local minima (i.e., the metastable states), the precise understanding of the *saddle structure* is also required. To the best of our knowledge, the analysis of the saddle structure as well as the metastable behavior of the heat-bath Glauber dynamics for $q \geq 4$ has not been analyzed yet.

Main contribution of the chapter

The main result of the present work is to provide the complete description of the energy landscape including the saddle structure and to analyze dynamical features of the Glauber dynamics based on it for the Curie–Weiss–Potts models with $q \geq 4$ spins.

First, we observe that for $q = 4$, as in the case of $q = 3$, the potential function has three critical temperatures

$$0 < \beta_1(4) < \beta_2(4) < \beta_3(4) = 4 \ ,$$

and moreover the associated metastable behavior is quite similar to that of the case $q = 3$. On the other hand, for $q \geq 5$, we will deduce that there are four critical temperatures

$$0 < \beta_1(q) < \beta_2(q) < \beta_3(q) < \beta_4(q) = q \ ,$$

where two critical temperatures $\beta_1(q)$ and $\beta_2(q)$ play essentially the same role with $\beta_1(3)$ and $\beta_2(3)$ (and hence $\beta_1(4)$ and $\beta_2(4)$), respectively. Surprisingly, our work reveals that the role of the third critical temperature $\beta_3(q)$ for $q \leq 4$ is divided into the third and fourth critical temperatures $\beta_3(q)$ and $\beta_4(q)$ for $q \geq 5$. More precisely, for $q \leq 4$, the change in the saddle gates between global minima and the disappearance of the local minimum representing the chaotic configuration happen simultaneously at $\beta = \beta_3(q) = q$; however, for $q \geq 5$, the change of saddle gates happens at $\beta = \beta_3(q) < q$ and the disappearance of the chaotic local minimum occurs at $\beta = \beta_4(q) = q$. Hence, for $q \geq 5$, we observe another type of metastable behavior at $\beta \in [\beta_3(q), \beta_4(q))$ compared to the case $q \leq 4$.

Remark. *We can also consider the Curie–Weiss–Potts model under an external field. For such models with $q = 3$ spins, the energy landscape has been completely analyzed in [53, Sections 5, 6]. We expect similar results but rigorous demonstration seems to be very complicated for general $q \geq 4$; hence we leave it for future research. We also remark that the Curie–Weiss–Potts model with an random external field has been studied in [85, Section 5].*

Other studies on the Potts model

Although the present work focuses on the Potts model on complete graphs, we also note that the Ising and Potts models on the lattice are widely studied as well. For instance, we refer to [80] and the references therein for the phase transition, to [67, 68, 69] for the cut-off phenomenon in the high-temperature regime, and to [1, 5, 9, 10, 11, 15, 42, 72, 74, 75, 77] for the metastability in the low-temperature regime. In addition, we refer to [34, 39] for the Potts model in many spins or large dimensions and to [16, 20] for the study of metastability of the Ising model on random graphs.

5.2 Model

In this section, we introduce the formal definition of the Curie–Weiss–Potts model, which will be analyzed in the present work. Fix an integer $q \geq 3$ and let $S = \{1, \dots, q\}$ be the set of spins.

5.2.1 Curie–Weiss–Potts measure

For a positive integer N, let us denote by[1] $K_N = \{1, \dots, N\}$ the set of sites. Let $\Omega_N = S^{K_N}$ be the configuration space of spins on K_N. Each configuration is represented as $\sigma = (\sigma_1, \dots, \sigma_N) \in \Omega_N$ where $\sigma_v \in S$ denotes a spin at site $v \in K_N$. Let $\boldsymbol{h} = (h_1, \dots, h_q) \in \mathbb{R}^q$ be the external magnetic field. The

[1] We write K_N to emphasize that our model is on the complete graph

Hamiltonian associated with the Curie–Weiss–Potts model with the external field $\boldsymbol{h}$ is given by

$$\mathbb{H}_N(\sigma) \;=\; -\frac{1}{2N}\sum_{1\le u,v\le N}\mathbf{1}(\sigma_u=\sigma_v)\;-\;\sum_{v=1}^{N}\sum_{j=1}^{q}h_j\mathbf{1}(\sigma_v=j)\;\;;\;\sigma\in\Omega_N\;,$$

where $\mathbf{1}$ denotes the usual indicator function. Then, the Gibbs measure associated with the Hamiltonian at the (inverse) temperature $\beta>0$ is given by

$$\mu_N^\beta(\sigma)\;=\;\frac{1}{Z_N(\beta)}e^{-\beta\mathbb{H}_N(\sigma)}\;\;;\;\sigma\in\Omega_N\;,$$

where $Z_N(\beta)=\sum_{\sigma\in\Omega_N}e^{-\beta\mathbb{H}_N(\sigma)}$ is the partition function. The measure $\mu_N^\beta(\cdot)$ denotes the Curie–Weiss–Potts measure on Ω_N at the inverse temperature β.

5.2.2 Heat-bath Glauber dynamics

Now, we define a heat-bath Glauber dynamics associated with the Curie–Weiss–Potts measure $\mu_N^\beta(\cdot)$. For $\sigma\in\Omega_N$, $v\in K_N$, and $k\in S$, denote by $\sigma^{v,\,k}$ the configuration whose spin σ_v at site v is flipped to k, i.e.,

$$(\sigma^{v,\,k})_u\;=\;\begin{cases}\sigma_u & u\neq v\;,\\ k & u=v\;.\end{cases}$$

Then, we will consider a heat-bath Glauber dynamics associated with generator $\mathcal{L}_N$ which acts on $f:\Omega_N\to\mathbb{R}$ as

$$(\mathcal{L}_Nf)(\sigma)\;=\;\frac{1}{N}\sum_{v=1}^{N}\sum_{k=1}^{q}c_{v,\,k}(\sigma)[f(\sigma^{v,\,k})-f(\sigma)]\;,$$

where

$$c_{v,\,k}(\sigma)\;=\;\exp\left\{-\frac{\beta}{2}[\mathbb{H}_N(\sigma^{v,\,k})-\mathbb{H}_N(\sigma)]\right\}\;.$$

It can be observed that this dynamics is reversible with respect to the Curie–Weiss–Potts measure $\mu_N^\beta(\cdot)$. Henceforth, denote by $\sigma(t) = \sigma^{K_N}(t) = (\sigma_1(t), \dots, \sigma_N(t))$ the continuous time Markov process associated with the generator $\mathcal{L}_N$.

5.2.3 Empirical magnetization

For each spin $k \in S$, denote by $r_N^k(\sigma)$ the proportion of spin k of configuration $\sigma \in \Omega_N$, i.e.,

$$r_N^k(\sigma) := \frac{1}{N}\sum_{v=1}^{N} \mathbf{1}(\sigma_v = k) ,$$

and define the proportional vector $\boldsymbol{r}_N(\sigma)$ as

$$\boldsymbol{r}_N(\sigma) := (r_N^1(\sigma), \dots, r_N^{q-1}(\sigma)) ,$$

which represents the *empirical magnetization* of the configuration σ containing the macroscopic information of σ.

Define Ξ as

$$\Xi = \{\boldsymbol{x} = (x_1, \dots, x_{q-1}) \in (\mathbb{R}_{\geq 0})^{q-1} : x_1 + \cdots + x_{q-1} \leq 1\} , \tag{5.1}$$

and then define a discretization of Ξ as

$$\Xi_N = \Xi \cap (\mathbb{Z}/N)^{q-1} .$$

With this notation, we immediately have $\boldsymbol{r}_N(\sigma) \in \Xi_N$ for $\sigma \in \Omega_N$.

For the Markov process $\big(\sigma(t)\big)_{t\geq 0}$, we write $\boldsymbol{r}_N(\cdot) = \boldsymbol{r}_N(\sigma(\cdot))$ which is a stochastic process on Ξ_N expressing the evolution of the empirical magnetization. Since the model is defined on the complete graph K_N, we obtain the following proposition.

Proposition 5.2.1. *The process* $\big(\boldsymbol{r}_N(t)\big)_{t\geq 0}$ *is a continuous time Markov* *chain on* Ξ_N *whose invariant measure is given by*

$$\nu_N^\beta(\boldsymbol{x}) := \mu_N^\beta(\boldsymbol{r}_N^{-1}(\boldsymbol{x})) \ \ ; \ \boldsymbol{x} \in \Xi_N$$

where $\boldsymbol{r}_N^{-1}(\boldsymbol{x})$ *denotes the set* $\{\sigma \in \Omega_N \,:\, \boldsymbol{r}_N(\sigma) = \boldsymbol{x}\}$. *Furthermore,* $\boldsymbol{r}_N(\cdot)$ *is reversible with respect to* ν_N^β.

The proof of this proposition including jump rates is given in Section 5.5.1. Let $\mathbb{P}_{\boldsymbol{x}}^{N,\beta}$ be the law of the Markov chain $\boldsymbol{r}_N(\cdot)$ starting at $\boldsymbol{x} \in \Xi_N$ and let $\mathbb{E}_{\boldsymbol{x}}^{N,\beta}$ be the corresponding expectation.

More on the measure $\nu_N^\beta(\cdot)$

For $\boldsymbol{y} \in \Xi$, let $\widehat{\boldsymbol{y}} = (y,\, \dots,\, y_{q-1},\, y_q) \in \mathbb{R}^q$ where $y_q = 1 - (y_1 + \cdots + y_{q-1})$. Then, the Hamiltonian $\mathbb{H}_N$ can be written as

$$\mathbb{H}_N(\sigma) \ = \ NH(\boldsymbol{r}_N(\sigma)) \ \ ; \ \sigma \in \Omega_N$$

where

$$H(\boldsymbol{x}) \ = \ -\frac{1}{2}|\widehat{\boldsymbol{x}}|^2 - \boldsymbol{h} \cdot \widehat{\boldsymbol{x}} \ \ ; \ \boldsymbol{x} \in \Xi \, . \tag{5.2}$$

Therefore, by Proposition 5.2.1, the invariant measure $\nu_N^\beta(\cdot)$ of the process $\boldsymbol{r}_N(t)$ on Ξ_N can be written as

$$\begin{aligned} \nu_N^\beta(\boldsymbol{x}) \ &= \ \sum_{\sigma:\boldsymbol{r}_N(\sigma)=\boldsymbol{x}} \frac{1}{Z_N(\beta)} \exp\{-\beta \mathbb{H}_N(\sigma)\} \\ &= \binom{N}{(Nx_1)\cdots(Nx_q)} \frac{1}{Z_N(\beta)} \exp\{-\beta N H(\boldsymbol{x})\} \\ &=: \frac{1}{(2\pi N)^{(q-1)/2} Z_N(\beta)} \exp\{-\beta N F_{\beta,\,N}(\boldsymbol{x})\} \ , \end{aligned} \tag{5.3}$$

where, by Stirling's formula, we can write

$$F_{\beta,\,N}(\boldsymbol{x}) \ = \ F_\beta(\boldsymbol{x}) + \frac{1}{N} G_{\beta,\,N}(\boldsymbol{x}) \ ,$$

where

$$F_\beta(\boldsymbol{x}) \;=\; H(\boldsymbol{x}) + \frac{1}{\beta}S(\boldsymbol{x}) \quad \text{and} \quad G_{\beta,\,N}(\boldsymbol{x}) \;=\; \frac{\log(x_1 \cdots x_q)}{2\beta} + O(N^{-1}) \;. \quad (5.4)$$

In this equation, $H(\cdot)$ is the energy functional defined in (5.2) and $S(\cdot)$ is the entropy functional defined by

$$S(\boldsymbol{x}) = \sum_{i=1}^{q} x_i \log(x_i) \;,$$

and $G_{\beta,\,N}(\boldsymbol{x})$ converges to $\log(x_1 \cdots x_q)/(2\beta)$ uniformly on every compact subsets of $\operatorname{int}\Xi$.

Main objectives of the chapter

Now, we can express the main purpose of the current chapter in a more concrete manner. In this chapter, we consider the Curie–Weiss–Potts model when there is no external magnetic field; i.e., $\boldsymbol{h} = \boldsymbol{0}$. Therefore, from now on, we assume $\boldsymbol{h} = \boldsymbol{0}$. Under this assumption, the first objective is to analyze the function $F_\beta(\cdot)$ expressing the energy landscape of the empirical magnetization of the Curie–Weiss–Potts model. This result will be explained in Section 5.3. The second concern is to investigate the metastable behavior of the process $\boldsymbol{r}_N(\cdot)$ in the low-temperature regime. This will be explained in Section 5.4. Latter part of the chapter is devoted to proofs of these results.

5.3 Main result for energy landscape

In view of Proposition 5.2.1, (5.3), and (5.4), the structure of the invariant measure $\nu_N^\beta(\cdot)$ of the process $\boldsymbol{r}_N(\cdot)$ is essentially captured by the potential function $F_\beta(\cdot)$; hence, the investigation of $F_\beta(\cdot)$ is crucial in the analysis of the energy landscape and the metastable behavior of $\boldsymbol{r}_N(\cdot)$. In this section, we explain our detailed analysis of the function $F_\beta(\cdot)$.

Note that the function $F_\beta(\cdot) = H(\cdot) + \beta^{-1}S(\cdot)$ express the competition between the energy and the entropy represented by $H(\cdot)$ and $S(\cdot)$, respectively. Since there is a β^{-1} factor in front of the entropy functional, we can expect that the entropy dominates the competition when β is small (i.e., the temperature is high). Since entropy is uniquely minimized at the equally distributed configuration $(1/q, \dots, 1/q) \in \Xi$, we can expect that the potential $F_\beta(\cdot)$ also has the unique minimum when β is small. On the other hand, if β is large enough (i.e., the temperature is low), the energy $H(\cdot)$ with q minima dominates the system, and therefore, we can expect that the potential F_β also has q global minima. In this section we provide the complete characterization of the complicated pattern of transition from this high-temperature regime to low-temperature regime in a precise level.

In Section 5.3.1, we define several points that will be shown to be critical points. In Section 5.3.2, we introduce several critical values of (inverse) temperature β. In Section 5.3.3, we summarize the results on the energy landscape $F_\beta(\cdot)$. In Section 5.3.4, as a by-product of these results, we compute the mean-field free energy.

5.3.1 Critical points of $F_\beta(\cdot)$

Let us first investigate critical points of $F_\beta(\cdot)$. We recall that

$$F_\beta(\boldsymbol{x}) \;=\; -\frac{1}{2}\sum_{k=1}^{q} x_k^2 \;+\; \frac{1}{\beta}\sum_{k=1}^{q} x_k \log x_k \;\;;\; \boldsymbol{x} \in \Xi \;.$$

Notation 5.3.1. *We have following notations for convenience.*

1. *Since there is no risk of confusion, we will write the point $\boldsymbol{x} = (x_1, \dots, x_{q-1}) \in \Xi$ as $\boldsymbol{x} = (x_1, \dots, x_{q-1}, x_q) \in \mathbb{R}^q$ where $x_q = 1 - x_1 - \cdots - x_{q-1}$.*
2. *Let $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{q-1}\}$ be the orthonormal basis of $\mathbb{R}^{q-1}$ and $\boldsymbol{e}_q = \boldsymbol{0} \in \mathbb{R}^{q-1}$. According to the convention above, the vectors $\boldsymbol{e}_1, \dots, \boldsymbol{e}_q$ can be regarded as an orthonormal basis of $\mathbb{R}^q$.*

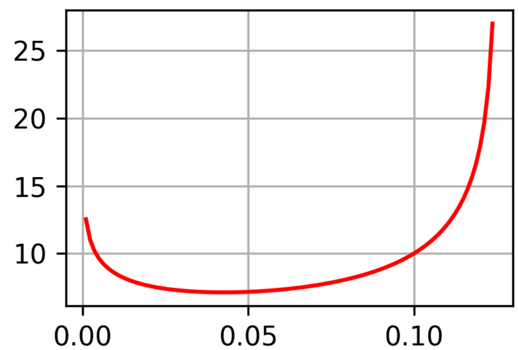


Figure 5.1: Graph of $g_2(t)$ for $q = 10$.

Now, we explain the candidates for the critical points of $F_\beta(\cdot)$ playing important role in the analysis of the energy landscape. The first candidate is

$$\mathbf{p} := (1/q,\ \dots,\ 1/q) \in \Xi\ ,$$

which represents the state where the spins are equally distributed.

In order to introduce the other candidates, we fix $i \in \mathbb{N} \cap [1, q/2]$ and let $j = q - i$. Define $g_i : (0, 1/j) \to \mathbb{R}$ as

$$g_i(t) \ := \ \frac{i}{1 - qt} \log\left(\frac{1 - jt}{it}\right)\ , \tag{5.5}$$

where we set $g_i(1/q) = q$ so that g_i becomes a continuous function on $(0, 1/j)$. We refer to Figure 5.1 for an illustration of graph of g_i. Then, it will be verified by Lemma 5.6.1 in Section 5.6.1 (and we can expect from the graph illustrated in Figure 5.1) that $g_i(t) = \beta$ has at most two solutions. We denote by $u_i(\beta) \le v_i(\beta)$ these solutions, provided that they exist. If there is only one solution, we let $u_i(\beta) = v_i(\beta)$ be this solution.

For $k \in S$, let

$$\mathbf{u}_1^k = \mathbf{u}_1^k(\beta) := \Big(u_1(\beta),\, \dots,\, 1-(q-1)u_1(\beta),\, \dots,\, u_1(\beta)\Big) \in \Xi\ , \tag{5.6}$$

$$\mathbf{v}_1^k = \mathbf{v}_1^k(\beta) := \Big(v_1(\beta),\, \dots,\, 1-(q-1)v_1(\beta),\, \dots,\, v_1(\beta)\Big) \in \Xi\ , \tag{5.7}$$

where $1-(q-1)u_1(\beta)$ and $1-(q-1)v_1(\beta)$ are located at the k-th component of $\mathbf{u}_1^k$ and $\mathbf{v}_1^k$, respectively. For $k,\, l \in S$, let

$$\begin{aligned}\mathbf{u}_2^{k,\,l} = \mathbf{u}_2^{k,\,l}(\beta) := \Big(u_2(\beta),\, \dots,\, &\frac{1-(q-2)u_2(\beta)}{2},\, \dots, \\ &\dots,\, \frac{1-(q-2)u_2(\beta)}{2},\, \dots,\, u_2(\beta)\Big) \in \Xi\ ,\end{aligned} \tag{5.8}$$

where $\dfrac{1-(q-2)u_2(\beta)}{2}$ is located at the k-th and l-th components. Of course, each of these points is well defined only when $u_1(\beta)$,$v_1(\beta)$, or $u_2(\beta)$ exists, respectively. Then, let

$$\mathcal{U}_1 := \{\mathbf{u}_1^k : k \in S\}\ ,\ \mathcal{U}_2 := \{\mathbf{u}_2^{k,\,l} : k,\, l \in S\}\ ,\ \text{and}\ \mathcal{V}_1 := \{\mathbf{v}_1^k : k \in S\}\ .$$

We remark that these sets depend on β although we omit β in the expressions for the simplicity of the notation.

Since we assumed that $\boldsymbol{h} = \mathbf{0}$, by symmetry, we can expect that the elements in $\mathcal{U}_1$ have the same properties; for instance, for all $k, l \in S$, we have $F_\beta(\mathbf{u}_1^k) = F_\beta(\mathbf{u}_1^l)$, and $\mathbf{u}_1^k$ is a critical point of $F_\beta(\cdot)$ if and only if $\mathbf{u}_1^l$ is. Of course the elements in $\mathcal{U}_2$ or $\mathcal{V}_1$ respectively have the same properties. Thus, it suffices to analyze their representatives, and hence select these representatives as

$$\mathbf{u}_1 = \mathbf{u}_1^q\ ,\ \mathbf{u}_2 = \mathbf{u}_2^{q-1,\,q}\ ,\ \text{and}\ \mathbf{v}_1 = \mathbf{v}_1^q\ . \tag{5.9}$$

Now, we have the following preliminary classification of critical points. We remark that a saddle point is a critical point at which the Hessian has only one negative eigenvalue.

Proposition 5.3.2. *The following hold.*

1. *If* $\mathbf{c} \in \Xi$ *is a local minimum of* F_β*, then* $\mathbf{c} \in \{\mathbf{p}\} \cup \mathcal{U}_1$*.*

2. *If* $\mathbf{s} \in \Xi$ *is a saddle point of* F_β*, then* $\mathbf{s} \in \mathcal{V}_1 \cup \mathcal{U}_2$ *for* $q \geq 4$ *and* $\mathbf{s} \in \mathcal{V}_1$ *for* $q = 3$*.*

Remark 5.3.3. *The set* $\mathcal{U}_2$ *is not defined for* $q = 3$ *since the set* $\mathcal{U}_i$ *is defined only when* $i \leq q/2$*. This will be explained in Section 5.6.1.*

The proof of this proposition is an immediate consequence of Proposition 5.6.3 in Section 5.6.1. The above proposition permits us to focus only on $\{\mathbf{p}\} \cup \mathcal{U}_1 \cup \mathcal{U}_2 \cup \mathcal{V}_1$ when we analyze the energy landscape in view of the metastable behavior, since the critical points of index greater than 1 cannot play any role, as the metastable transition always happens at the neighborhood of a saddle point (a critical point of index 1).

5.3.2 Critical temperatures

In this subsection, we introduce critical temperatures

$$0 < \beta_1(q) < \beta_2(q) < \beta_3(q) \leq q \ ,$$

at which the phase transitions in the energy landscape occur. The precise definition of these critical temperatures are given in (5.31) of Section 5.6.2. Henceforth, we write $\beta_i = \beta_i(q)$, $1 \leq i \leq 3$, since there is no risk of confusion.

To describe the role of these critical temperatures, we regard β as increasing from 0 to ∞. Figure 5.2 shows the role of $\mathbf{p}$, $\mathcal{U}_1$, $\mathcal{V}_1$, and $\mathcal{U}_2$ according to inverse temperature. Section 5.6 will prove this figure.

At $\beta = \beta_1$, the dynamics exhibits phase transition from fast mixing to slow mixing, and this is proven in [19]. Furthermore, the behavior of the dynamics changes from cutoff phenomenon to metastability. This phase transition is due to the appearance of new local minima $\mathcal{U}_1$ of $F_\beta(\cdot)$ other than $\mathbf{p}$ at $\beta = \beta_1$.

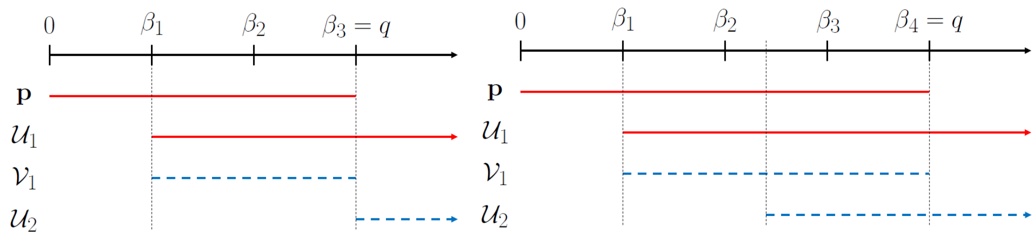


Figure 5.2: Role of each critical point according to temperature. Solid lines imply local minima and dashed lines imply saddle points.

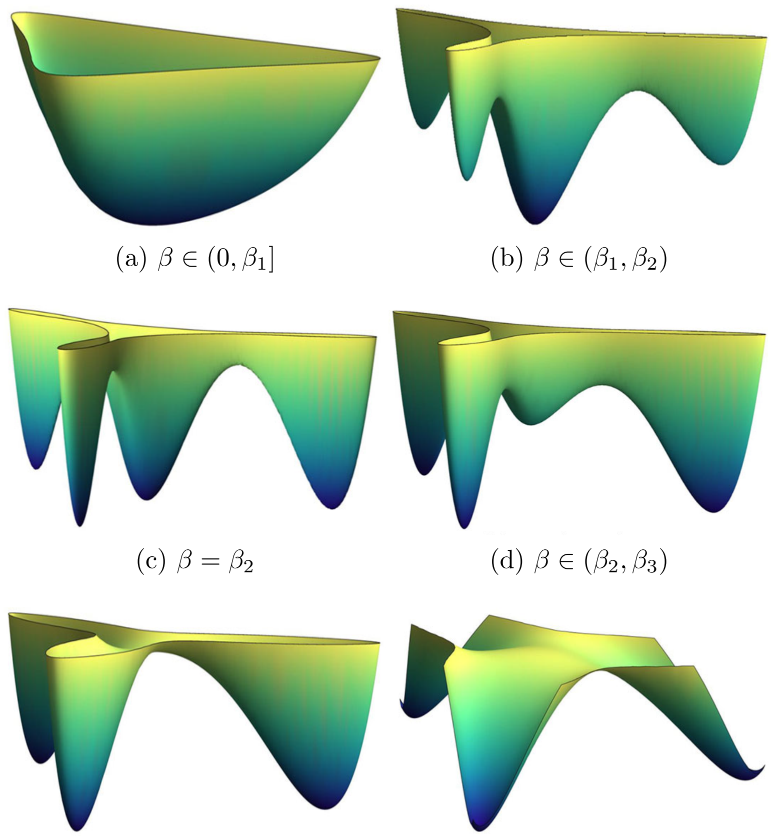
At $\beta = \beta_2$, the ground states of dynamics change from $\mathbf{p}$ to elements of $\mathcal{U}_1$, as observed in [18, Theorem 3.1(b)]. To explain the role of critical temperatures β_3 and q, we have to divide the explanation into several cases. Let us first assume that $q \geq 5$ so that $\beta_3 < q$. At $\beta = \beta_3$, the saddle gates among the ground states in $\mathcal{U}_1$ is changed from $\mathcal{V}_1$ to $\mathcal{U}_2$ (since the heights $F_\beta(\mathbf{v}_1)$ and $F_\beta(\mathbf{u}_2)$ are reversed at this point) and at $\beta = q$, the local minimum $\mathbf{p}$ becomes a local maximum. On the other hand, for $q \leq 4$, we have $\beta_3 = q$. At $\beta = \beta_3$, the change of the saddle gates and the disappearance of the local minimum $\mathbf{p}$ occur simultaneously. We refer to [53] for the detailed description when $q = 3$.

5.3.3 Stable and metastable sets

We define some metastable sets based on the results explained earlier. If $q \geq 4$, define H_β as (cf. (5.9))

$$H_\beta = \begin{cases} F_\beta(\mathbf{v}_1) \ , & \beta \in (\beta_1, \beta_3) \ , \\ F_\beta(\mathbf{u}_2) \ , & \beta \in [\beta_3, \infty) \ . \end{cases} \tag{5.10}$$

When $q = 3$, we set $H_\beta = F_\beta(\mathbf{v}_1)$ for all $\beta > \beta_1$ (cf. Remark 5.3.3). It will be verified in Lemma 5.6.7 and (5.31) that H_β is the height of the lowest saddle



(a) $\beta \in (0, \beta_1]$ (b) $\beta \in (\beta_1, \beta_2)$

(c) $\beta = \beta_2$ (d) $\beta \in (\beta_2, \beta_3)$

(e) $\beta = \beta_3$ (f) $\beta \in (\beta_3, \infty)$

Figure 5.3: Energy landscape of F_β when $q = 3$.

points.

Let $\widehat{S} := S \cup \{\mathfrak{o}\}$ and $\mathbf{u}_1^{\mathfrak{o}} := \mathbf{p}$. Let[2] $\mathcal{W}_k = \mathcal{W}_k(\beta)$, $k \in S$, be the connected component of $\{F_\beta < H_\beta\}$ containing $\mathbf{u}_1^k$ and let $\mathcal{W}_{\mathfrak{o}} = \mathcal{W}_{\mathfrak{o}}(\beta)$ be the connected component of $\{F_\beta < F_\beta(\mathbf{v}_1)\}$ containing $\mathbf{u}_1^{\mathfrak{o}}$. For $k,\, l \in \widehat{S}$, let $\Sigma_{k,\,l} = \Sigma_{k,\,l}(\beta) := \overline{\mathcal{W}_k} \cap \overline{\mathcal{W}_l}$ be a set of saddle gates of height H_β between $\mathbf{u}_1^k$ and $\mathbf{u}_1^l$.

Now, we can state the main result on energy landscape and the proofs of

[2]We define the set $\mathcal{W}_k$, $k \in S$, and $\mathcal{W}_{\mathfrak{o}}$ as the empty set if the set $\{F_\beta < H_\beta\}$ does not contain $\mathbf{u}_1^k$ and $\{F_\beta < F_\beta(\mathbf{v}_1)\}$ does not contain $\mathbf{u}_1^{\mathfrak{o}}$ respectively.

theorems in this section will presented in Section 5.9. The first result holds for all $q \geq 3$.

Theorem 5.3.4. *For $q \geq 3$, the following hold.*

1. *If $\beta \leq \beta_1$, there is no critical point other than $\mathbf{p}$, which is the global minimum.*

2. *For $\beta \in (\beta_1, q)$, we have $\mathcal{W}_{\mathfrak{o}} \neq \emptyset$ and for $\beta \in [q, \infty)$, we have $\mathcal{W}_{\mathfrak{o}} = \emptyset$.*

3. *Let $\mathcal{M}_\beta$ be a set of local minima of F_β. Then, we have*

$$\mathcal{M}_\beta = \begin{cases} \{\mathbf{p}\} & \beta \in (0, \beta_1] \ , \\ \{\mathbf{p}\} \cup \mathcal{U}_1 & \beta \in (\beta_1, q) \ , \\ \mathcal{U}_1 & \beta \in [q, \infty) \ . \end{cases}$$

4. *Let $\mathcal{M}^\star_\beta$ be a set of global minima of F_β. Then, we have*

$$\mathcal{M}^\star_\beta = \begin{cases} \{\mathbf{p}\} & \beta \in (0, \beta_2) \ , \\ \{\mathbf{p}\} \cup \mathcal{U}_1 & \beta = \beta_2 \ , \\ \mathcal{U}_1 & \beta \in (\beta_2, \infty) \ . \end{cases}$$

Since there is only one minimum if $\beta \leq \beta_1$, we now consider $\beta > \beta_1$. Before we write the main result on metastable sets, we would like to emphasize that [53, Proposition 4.4] proved the case when $q = 3$, while the proof for the case $q \geq 4$ is the main novel contents of the current chapter. We first consider the case $q \leq 4$. See Figure 5.3[3] for the visualization of the following and above theorem.

Theorem 5.3.5. *For $q \leq 4$, the following hold.*

1. *$\beta_3 = q$.*

[3] This figures are excerpt from [53, Fig 4]

2. *For* $\beta \in (\beta_1, q)$*, the sets* $\mathcal{W}_k$*,* $k \in \widehat{S}$*, are nonempty and disjoint. For* $k, l \in S$*,* $\Sigma_{k,l} = \emptyset$ *and for* $k \in S$*,* $\Sigma_{\mathfrak{o},k} = \{\mathbf{v}_1^k\}$*.*

3. *For* $\beta = q$*, we have* $\mathcal{W}_{\mathfrak{o}} = \emptyset$*. The sets* $\mathcal{W}_k$*,* $k \in S$*, are nonempty and disjoint. For* $k, l \in S$*,* $\Sigma_{k,l} = \{\mathbf{p}\}$*.*

4. *For* $\beta \in (q, \infty)$*, we have* $\mathcal{W}_{\mathfrak{o}} = \emptyset$*. The sets* $\mathcal{W}_k$*,* $k \in S$*, are nonempty and disjoint. For* $k, l \in S$*,*

$$\Sigma_{k,l} = \begin{cases} \{\mathbf{v}_1^m\} \ , \ \text{where } m \in S \setminus \{k, l\} \ , & \text{if } q = 3 \ , \\ \{\mathbf{u}_2^{k,l}\} \ , & \text{if } q = 4 \ . \end{cases}$$

Next, we consider the case $q \geq 5$. Note that the crucial difference compared to the previous theorem lies in the third and fifth statements. See Figures 5.4 and 5.5 for the visualization of the following theorem and Theorem 5.3.4.

Theorem 5.3.6. *For* $q \geq 5$*, the following hold.*

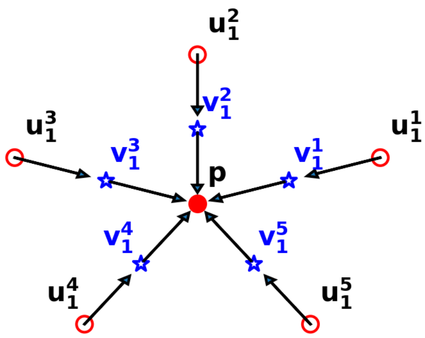
1. $\beta_3 < q$*.*

2. *For* $\beta \in (\beta_1, \beta_3)$*, the sets* $\mathcal{W}_k$*,* $k \in \widehat{S}$*, are nonempty and disjoint. For* $k, l \in S$*,* $\Sigma_{k,l} = \emptyset$ *and for* $k \in S$*,* $\Sigma_{\mathfrak{o},k} = \{\mathbf{v}_1^k\}$

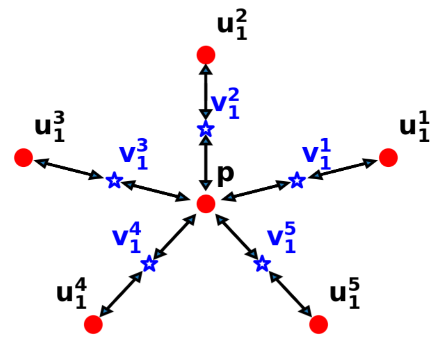
3. *For* $\beta = \beta_3$*, the sets* $\mathcal{W}_k$*,* $k \in \widehat{S}$*, are nonempty and disjoint. For* $k, l \in S$*,* $\Sigma_{k,l} = \{\mathbf{u}_2^{k,l}\}$ *and for* $k \in S$*,* $\Sigma_{\mathfrak{o},k} = \{\mathbf{v}_1^k\}$*.*

4. *For* $\beta \in (\beta_3, \infty)$*, the sets* $\mathcal{W}_k$*,* $k \in S$*, are nonempty and disjoint. For* $k, l \in S$*,* $\Sigma_{k,l} = \{\mathbf{u}_2^{k,l}\}$ *and for* $k \in S$*,* $\Sigma_{\mathfrak{o},k} = \emptyset$*.*

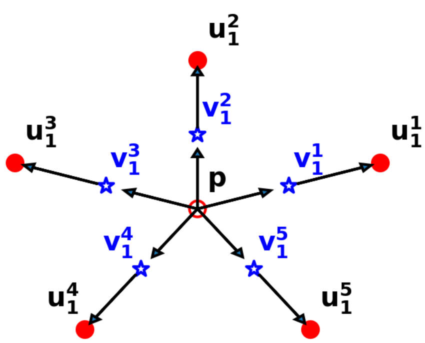
5. *For* $\beta \in (\beta_3, q)$*, we have* $F_\beta(\mathbf{v}_1) > H_\beta$*. Furthermore, the set* $\{F_\beta < F_\beta(\mathbf{v}_1)\}$ *has only two connected components, the well* $\mathcal{W}_{\mathfrak{o}}$ *and the other containing* $\mathcal{U}_1$*. The saddle points between them are* $\mathcal{V}_1$*.*



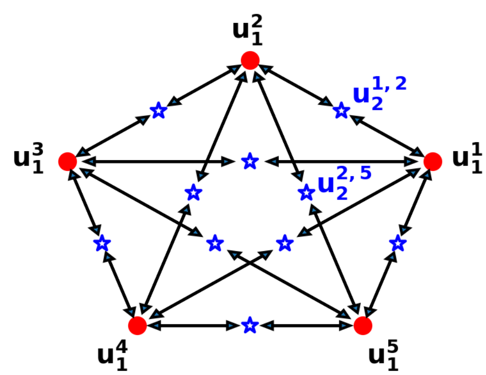
(a) (β_1, β_2)



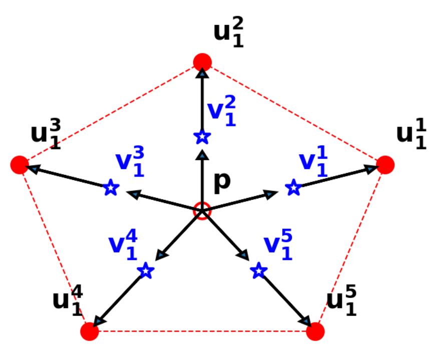
(b) $\beta = \beta_2$



(c) (β_2, β_3)

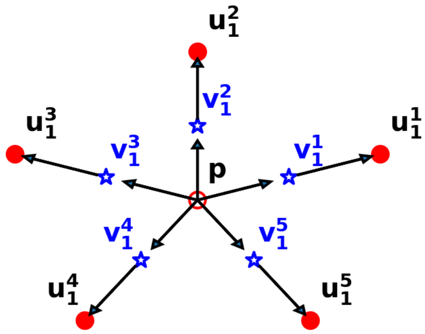


(d) (β_3, ∞)

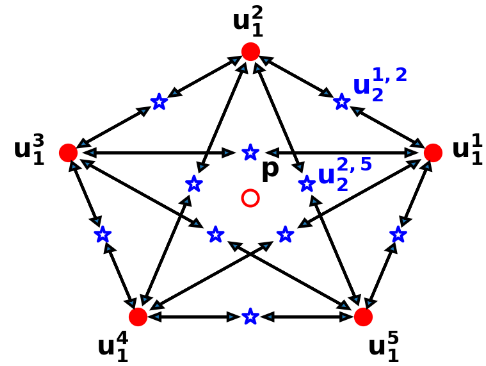


(e) (β_3, q)

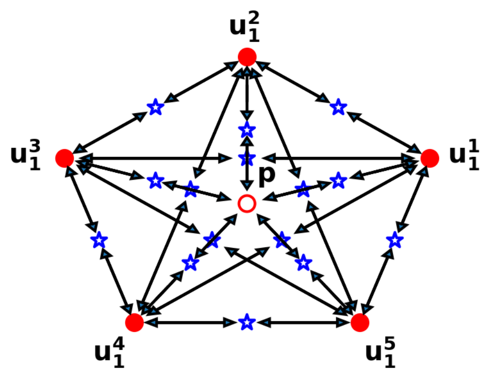
Figure 5.4: Illustration of energy landscape of F_β when $q = 5$. The first four figures are $\{F_\beta \le H_\beta\}$ and the last figure is $\{F_\beta \le F_\beta(\mathbf{v}_1)\}$. The star-shaped vertices and circles represent saddle points and local minima, respectively. The empty circles are shallower minima. Each arrow represents a path from one shallower minimum to another deeper minimum passing through a saddle point.



(a) $\mathcal{U}_1$ and $\mathcal{V}_1$



(b) $\mathcal{U}_1$ and $\mathcal{U}_2$



(c) $\mathcal{U}_1$, $\mathcal{V}_1$, and $\mathcal{U}_2$

Figure 5.5: Illustration of energy landscape of F_β when $q = 5$ and $\beta = \beta_3$. The figures are $\{F_\beta \le H_\beta\}$. The star-shaped vertices and circles represent saddle points and local minima, respectively. The empty circles are shallower minima. Each arrow represents a path from one shallower minimum to another deeper minimum passing through a saddle point. Note that when $\beta = \beta_3$, $\mathcal{U}_2$ and $\mathcal{V}_1$ exist simultaneously. The first two figures are illustrations of saddle structures of $\mathcal{V}_1$ and $\mathcal{U}_2$, respectively. The last figures is a combination of the previous two figures.

5.3.4 Mean-field free energy

In this subsection, we compute the mean-field free energy of the Curie–Weiss–Potts model defined by

$$\psi(\beta) := -\lim_{N\to\infty} \frac{1}{\beta N} \log Z_N(\beta) \ . \tag{5.11}$$

It is well known that the Curie–Weiss model with $q = 2$ spins exhibits the second-order phase transition at the unique critical temperature $\beta = \beta_c$, while the Curie–Weiss–Potts model with $q \geq 3$ spins exhibits the first-order phase transition at $\beta = \beta_2$ (cf. [18, 26, 78]). We now reconfirm this folklore by computing the free energy explicitly. This computation is based on the following observation (cf. [26, display (2.4)]):

$$\lim_{N\to\infty} \frac{1}{\beta N} \log Z_N(\beta) = \sup_{\boldsymbol{x}\in\Xi} \{-F_\beta(\boldsymbol{x})\} \ . \tag{5.12}$$

We give a rigorous proof in Appendix B.

Now, let us assume that $q \geq 3$ so that by (5.11), (5.12), and Theorems 5.3.4, we can deduce that

$$\psi(\beta) = \begin{cases} F_\beta(\mathbf{p}) & \text{if } \beta \leq \beta_2 \ , \\ F_\beta(\mathbf{u}_1) & \text{if } \beta > \beta_2 \ . \end{cases} \tag{5.13}$$

Corollary 5.3.7. *We have that*

$$\psi'(\beta) = \begin{cases} -\dfrac{1}{\beta^2} S(\mathbf{p}) & \textit{if } \beta < \beta_2 \ , \\ -\dfrac{1}{\beta^2} S(\mathbf{u}_1) & \textit{if } \beta > \beta_2 \ . \end{cases} \tag{5.14}$$

In particular, the Curie–Weiss–Potts model with $q \geq 3$ exhibits the first-order phase transition at $\beta = \beta_2$.

Proof. Let $\mathbf{c}(\beta) \in \Xi$ be a critical point of $F_\beta(\cdot)$. Then, since $F_\beta = H + \beta^{-1} S$, we have

$$\frac{d}{d\beta}F_{\beta}(\mathbf{c}(\beta)) = \nabla F_{\beta}(\mathbf{c}(\beta)) \cdot \dot{\mathbf{c}}(\beta) - \frac{1}{\beta^2}S(\mathbf{c}(\beta)) \ .$$

Since $\nabla F_{\beta}(\mathbf{c}(\beta)) = 0$, we get (5.14) from (5.13). Since S attains its unique local minimum at $\mathbf{p}$ and $\mathbf{u}_1 \neq \mathbf{p}$, $\psi'(\cdot)$ is discontinuous at $\beta = \beta_2$. □

5.4 Main result for metastability

In this section, we analyze the metastable behavior of $\boldsymbol{r}_N(\cdot)$ based on the analysis of the energy landscape carried out in the previous section and the general results obtained by [54]. As inverse temperature β varies, the behavior of this dynamics changes both qualitatively and quantitatively thanks to the structural phase transitions explained in the previous section.

Since the invariant measure ν_N^{β} is exponentially concentrated in neighborhoods of ground states, the corresponding Markov process $\boldsymbol{r}_N(\cdot)$ stays most of the time at these neighborhoods. The abrupt transitions between such stable states are the metastable behavior of the process $\boldsymbol{r}_N(\cdot)$ and one of the natural ways of describing these hopping dynamics among the neighborhoods of the ground states is the *Markov chain model reduction.* A comprehensive understanding of such approaches can be obtained from [2, 3, 46].

When the dynamics starts from a local minimum which is not a global minimum, we have to estimate the mean of the transition time to the global minimum in order to quantitatively understand the metastable behavior. This estimation is known as the *Eyring–Kramers formula.* In this section we provide the Markov chain model reduction and Eyring–Kramers formula for the metastable process $\boldsymbol{r}_N(\cdot)$.

Such a metastable behavior is observed only when there are multiple local minima; and hence we cannot expect metastable behavior at the high-temperature regime $\beta \leq \beta_1$ for which $\mathbf{p}$ is the unique local (and global) minimum. Hence, we assume $\beta > \beta_1$ in this section.

5.4.1 Some preliminaries

In this subsection, we introduce several notions crucial to the description of the metastable behavior.

Some constants

We first define the so-called Eyring–Kramers constants which play fundamental role in the quantitative analysis of metastability. Recall the definition of $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_q\}$ from Notation 5.3.1. Define $(q-1)\times(q-1)$ matrices $\mathbb{A}^{i,j}$, $i, j \in S$, and $\mathbb{A}(\boldsymbol{x})$ as

$$\mathbb{A}^{i,j} = (\boldsymbol{e}_j - \boldsymbol{e}_i)(\boldsymbol{e}_j - \boldsymbol{e}_i)^\dagger \quad \text{and} \quad \mathbb{A}(\boldsymbol{x}) = \sum_{1\le i<j\le q} w^{i,j}(\boldsymbol{x})\mathbb{A}^{i,j} \ .$$

As we will see in Section 5.5.3, these constants are related to the drift of empirical magnetization as a consequence of spin update from i to j or j to i. Since $\mathbb{A}^{i,j}$, $i, j \in S$, are positive definite, $\mathbb{A}(\boldsymbol{x})$ satisfies [54, display (A.1)] and hence, by [54, Lemma A.1], for all $k, l \in S$, the matrices $(\nabla^2 F_\beta)(\mathbf{u}_2^{k,l})\mathbb{A}(\mathbf{u}_2^{k,l})^\dagger$ and $(\nabla^2 F_\beta)(\mathbf{v}_1^k)\mathbb{A}(\mathbf{v}_1^k)^\dagger$ have the unique negative eigenvalue which will be denoted respectively by $-\mu_{k,l} = -\mu_{k,l}(\beta)$ and $-\mu_{\mathfrak{o},k} = -\mu_{\mathfrak{o},k}(\beta)$.

Now, let us define the *Eyring–Kramers constant*s corresponding to our model as

$$\omega_{k,l} = \omega_{k,l}(\beta) := \frac{\mu_{k,l}(\beta)}{\sqrt{-\det[(\nabla^2 F_\beta)(\mathbf{u}_2^{k,l})]}} e^{-\beta G_\beta(\mathbf{u}_2^{k,l})} \ , \quad k,\, l \in S \ , \tag{5.15}$$

$$\omega_{\mathfrak{o},k} = \omega_{\mathfrak{o},k}(\beta) := \frac{\mu_{\mathfrak{o},k}(\beta)}{\sqrt{-\det[(\nabla^2 F_\beta)(\mathbf{v}_1^k)]}} e^{-\beta G_\beta(\mathbf{v}_1^k)} \ , \quad k \in S \ . \tag{5.16}$$

By symmetry, we have $\omega_{k,l} = \omega_{k',l'}$ for all $k,\, l \in S$ and $k',\, l' \in S$ and $\omega_{o,k} = \omega_{\mathfrak{o},k'}$ for all $k,\, k' \in S$. Hence, let us write $\omega_{\mathfrak{o}} = \omega_{\mathfrak{o},1}$ and $\omega_1 = \omega_{1,2}$.

Next, define

$$\nu_k = \nu_k(\beta) := \frac{\exp(-\beta G_\beta(\mathbf{u}_1^k))}{\sqrt{\beta^2 \det[(\nabla^2 F_\beta)(\mathbf{u}_1^k)]}} \;, \quad k \in S \;, \tag{5.17}$$

$$\nu_{\mathfrak{o}} = \nu_{\mathfrak{o}}(\beta) := \frac{\exp(-\beta G_\beta(\mathbf{p}))}{\sqrt{\beta^2 \det[(\nabla^2 F_\beta)(\mathbf{p})]}} \;. \tag{5.18}$$

As explained in [54, display (2.8)], the constants ν_k, $k \in S$, and $\nu_{\mathfrak{o}}$ are the normalized asymptotic mass of the neighborhood of $\mathbf{u}_1^k$ and $\mathbf{p}$, respectively. By the symmetry, we also obtain $\nu_1 = \cdots = \nu_q$.

Time scales

The constant H_β defined in (5.10) denotes the height of the lowest saddle points. Let $\theta_k = \theta_k(\beta)$, $k \in \widehat{S}$, be the depth of well $\mathcal{W}_k(\beta)$, i.e.,

$$\begin{cases} \theta_1 = \cdots = \theta_q = \beta[H_\beta - F_\beta(\mathbf{u}_1)] \;, \\ \theta_{\mathfrak{o}} = \beta[F_\beta(\mathbf{v}_1) - F_\beta(\mathbf{p})] \;. \end{cases}$$

Then, $e^{N\theta_1}$ and $e^{N\theta_{\mathfrak{o}}}$ represent the time scales on which $\boldsymbol{r}_N(\cdot)$ exhibits metastability. For $\beta \geq q$, the constant $\theta_{\mathfrak{o}}$ is meaningless since $\mathcal{W}_{\mathfrak{o}} = \emptyset$.

Order process and Markov chain model reduction

Let $\delta = \delta(\beta) > 0$ be a small enough number such that $\delta < \min\{\theta_{\mathfrak{o}}, \theta_1\}$. If $\beta \geq q$, since $\theta_{\mathfrak{o}}$ is not defined, let $\delta = (1/2)\theta_1$. For $k \in S$, define

$$\mathcal{W}_k^\delta = \mathcal{W}_k \cap \{\boldsymbol{x} \in \Xi \,:\, F_\beta(\boldsymbol{x}) < H_\beta - \delta\} \;,$$
$$\mathcal{W}_{\mathfrak{o}}^\delta = \mathcal{W}_{\mathfrak{o}} \cap \{\boldsymbol{x} \in \Xi \,:\, F_\beta(\boldsymbol{x}) < F_\beta(\mathbf{v}_1) - \delta\} \;.$$

For $k \in \widehat{S}$, define $\mathcal{E}_N^k = \mathcal{E}_N^k(\beta)$ as

$$\mathcal{E}_N^k = \mathcal{W}_k^\delta \cap \Xi_N \;.$$

This set $\mathcal{E}_N^k$ is called the metastable set, provided that it is not an empty set. For $A \subset \widehat{S}$, we write

$$\mathcal{E}_N^A = \bigcup_{k \in A} \mathcal{E}_N^k \ .$$

Let T be S, $\widehat{S}$, or $\{\mathfrak{o},\ S\}$. Denote by $\Psi_N = \Psi_N^\beta : \Xi_N \to T \cup \{N\}$ the projection map given by

$$\Psi_N(\boldsymbol{x}) \ = \ \sum_{k \in T} k \mathbf{1}\{\boldsymbol{x} \in \mathcal{E}_N^k\} + N \mathbf{1}\{\boldsymbol{x} \in \Xi_N \setminus \mathcal{E}_N^T\} \ .$$

Let us define the so-called *order process* by $\mathbf{X}_N(t) = \Psi_N(\boldsymbol{r}_N(t))$ which represents the index of metastable set at which the process $\boldsymbol{r}_N(\cdot)$ is staying.

Definition 5.4.1 (Markov chain model reduction)**.** *Let* $\mathbf{X}(\cdot)$ *be a continuous time Markov chain on* T*. We say that the metastable behavior of the process* $\boldsymbol{r}_N(\cdot)$ *is described by a Markov Process* $\mathbf{X}(\cdot)$ *in the time scale* θ_N *if, for all* $k \in T$ *and for all sequence* $(\boldsymbol{x}_N)_{N \geq 1}$ *such that* $\boldsymbol{x}_N \in \mathcal{E}_N^k$ *for all* $N \geq 1$*, the finite dimensional marginals of the process* $\mathbf{X}_N(\theta_N \cdot)$ *under* $\mathbb{P}_{\boldsymbol{x}_N}^{N,\beta}$ *converges to that of the Markov chain* $\mathbf{X}(\cdot)$ *as* $N \to \infty$*.*

In the previous definition, it is clear that the Markov chain $\mathbf{X}(\cdot)$ describes the inter-valley dynamics of the process $\boldsymbol{r}_N(\cdot)$ accelerated by a factor of θ_N.

5.4.2 Metastability results for $q \leq 4$

We can now state the main result for the metastable behavior. First, we consider the case $q \leq 4$ whose result is essentially the same as that in [53, Section 4.3] where only the case $q = 3$ was considered.

We define limiting Markov chains when $q \leq 4$.

Definition. *Let* $q \leq 4$ *and*[4] $i \in \{\, (1, 2),\ (2),\ (2, 3),\ (3, \infty),\ (4)\, \}$*. Let* $\mathbf{Y}_q^i(\cdot)$ *be a Markov chain on* T *with jump rate* $r_q^i : T \times T \to \mathbb{R}$ *given by*

[4] Here, $(1, 2)$, $(2, 3)$, $(3, \infty)$ are single element of the given set.

1. $r_q^{(1,2)}(k,l) = \mathbf{1}\{l = \mathfrak{o}\}\dfrac{\omega_{\mathfrak{o}}}{\nu_1}\,,\ T = \widehat{S}$.

2. $r_q^{(2)}(k,l) = \mathbf{1}\{l = \mathfrak{o}\}\dfrac{\omega_{\mathfrak{o}}}{\nu_1} + \mathbf{1}\{k = \mathfrak{o}\}\dfrac{\omega_{\mathfrak{o}}}{\nu_{\mathfrak{o}}}\,,\ T = \widehat{S}$.

3. $r_q^{(2,3)}(k,l) = \dfrac{\omega_{\mathfrak{o}}}{q\nu_1}\,,\ T = S$.

4. $r_q^{(3,\infty)}(k,l) = \begin{cases} \dfrac{\omega_0}{\nu_1}\,, & q = 3 \\ \dfrac{\omega_1}{\nu_1}\,, & q = 4 \end{cases}\,,\ T = S$.

5. $r_q^{(4)}(k,l) = \mathbf{1}\{k = \mathfrak{o}\}\dfrac{\omega_{\mathfrak{o}}}{\nu_{\mathfrak{o}}}\,,\ T = \widehat{S}$.

The following theorem is the metastability result for $q \le 4$. We remark that the case when $q = 3$ is proven in [53, Section 4.3].

Theorem 5.4.2. *Let $q \le 4$. Then, the metastable behavior of the process $\boldsymbol{r}_N(\cdot)$ is described by (cf. Definition 5.4.1)*

1. *$\beta \in (\beta_1, \beta_2)$: the process $\mathbf{Y}_q^{(1,2)}(\cdot)$ in the time scale $2\pi N e^{\theta_1}$.*

2. *$\beta = \beta_2$: the process $\mathbf{Y}_q^{(2)}(\cdot)$ in the time scale $2\pi N e^{\theta_1}$.*

3. *$\beta \in (\beta_2, q)$: the process $\mathbf{Y}_q^{(2,3)}(\cdot)$ in the time scale $2\pi N e^{\theta_1}$ and by the process $\mathbf{Y}_q^{(4)}(\cdot)$ in the time scale $2\pi N e^{\theta_{\mathfrak{o}}}$.*

4. *$\beta \in (q, \infty)$: the process $\mathbf{Y}_q^{(3,\infty)}(\cdot)$ in the time scale $2\pi N e^{\theta_1}$.*

The proof follows from Theorems 5.3.4 and 5.3.5, Proposition 5.5.3, and [53, Theorem 2.2, Remark 2.10, 2.11].

Remark 5.4.3. *As mentioned in [53], we cannot investigate the case $\beta = \beta_3 = q$ with the current method since $\mathbf{p}$ is a degenerate saddle point.*

Remark 5.4.4. *The qualitative feature of the metastable behavior of $\boldsymbol{r}_N(\cdot)$ is essentially same for $q = 3$ and $q = 4$. The only difference is that the saddle* *points between metastable sets are defined in different ways; however, when $\beta > \beta_3$, the points in $\mathcal{V}_1$ for $q = 3$ and $\mathcal{U}_2$ for $q = 4$ play the same role since all the points belonging to these sets represent states in which most sites are aligned to two spins equally.*

5.4.3 Metastability results for $q \geq 5$

As in the previous subsection, we start by defining limiting Markov chains. Note that there are two different Markov chains.

Definition. *Let $q \geq 5$ and $i \in \{ (1,2),\, (2),\, (2,3),\, (3),\, (3,\infty),\, (4),\, (5) \}$. Let $\mathbf{Y}_q^i(\cdot)$ be a Markov chain on T with jump rate $r_q^i : T \times T \to \mathbb{R}$ with jump rate $r_q^i : T \times T \to \mathbb{R}$ given by*

1. $r_q^{(1,2)}(k,l) = \mathbf{1}\{l = \mathfrak{o}\} \dfrac{\omega_{\mathfrak{o}}}{\nu_1}\,,\ T = \widehat{S}.$

2. $r_q^{(2)}(k,l) = \mathbf{1}\{l = \mathfrak{o}\} \dfrac{\omega_{\mathfrak{o}}}{\nu_1} + \mathbf{1}\{k = \mathfrak{o}\} \dfrac{\omega_{\mathfrak{o}}}{\nu_{\mathfrak{o}}}\,,\ T = \widehat{S}.$

3. $r_q^{(2,3)}(k,l) = \dfrac{\omega_{\mathfrak{o}}}{q\nu_1}\,,\ T = S.$

4. $r_q^{(3)}(k,l) = \dfrac{1}{\nu_1}(\dfrac{\omega_{\mathfrak{o}}}{q} + \omega_1)\,,\ T = S.$

5. $r_q^{(3,\infty)}(k,l) = \dfrac{\omega_1}{\nu_1}\,,\ T = S.$

6. $r_q^{(4)}(k,l) = \mathbf{1}\{k = \mathfrak{o}\} \dfrac{\omega_{\mathfrak{o}}}{\nu_{\mathfrak{o}}}\,,\ T = \widehat{S}.$

7. $r_q^{(5)}(k,l) = \mathbf{1}\{k = \mathfrak{o}\} \dfrac{q\omega_{\mathfrak{o}}}{\nu_{\mathfrak{o}}}\,,\ T = \{\mathfrak{o},\, S\}.$

Now, we present the metastability result for $q \geq 5$. The new metastable behaviors are observed when $\beta = \beta_3$ and $\beta \in (\beta_3, q)$.

Theorem 5.4.5. *Let $q \geq 5$. Then, the metastable behavior of the process $\boldsymbol{r}_N(\cdot)$ is described by*

1. *$\beta \in (\beta_1, \beta_2)$: the process $\mathbf{Y}_q^{(1,2)}(\cdot)$ in the time scale $2\pi N e^{\theta_1}$.*

2. *$\beta = \beta_2$: the process $\mathbf{Y}_q^{(2)}(\cdot)$ in the time scale $2\pi N e^{\theta_1}$.*

3. *$\beta \in (\beta_2, \beta_3)$: the process $\mathbf{Y}_q^{(2,3)}(\cdot)$ in the time scale $2\pi N e^{\theta_1}$ and by the process $\mathbf{Y}_q^{(4)}(\cdot)$ in the time scale $2\pi N e^{\theta_{\mathbf{o}}}$.*

4. *$\beta = \beta_3$: the process $\mathbf{Y}_q^{(3)}(\cdot)$ in the time scale $2\pi N e^{\theta_1}$ and by the process $\mathbf{Y}_q^{(4)}(\cdot)$ in the time scale $2\pi N e^{\theta_{\mathbf{o}}}$.*

5. *$\beta \in (\beta_3, q)$: the process $\mathbf{Y}_q^{(3,\infty)}(\cdot)$ in the time scale $2\pi N e^{\theta_1}$ and by the process $\mathbf{Y}_q^{(5)}(\cdot)$ in the time scale $2\pi N e^{\theta_{\mathbf{o}}}$.*

6. *$\beta \in [q, \infty)$: the process $\mathbf{Y}_q^{(3,\infty)}(\cdot)$ in the time scale $2\pi N e^{\theta_1}$.*

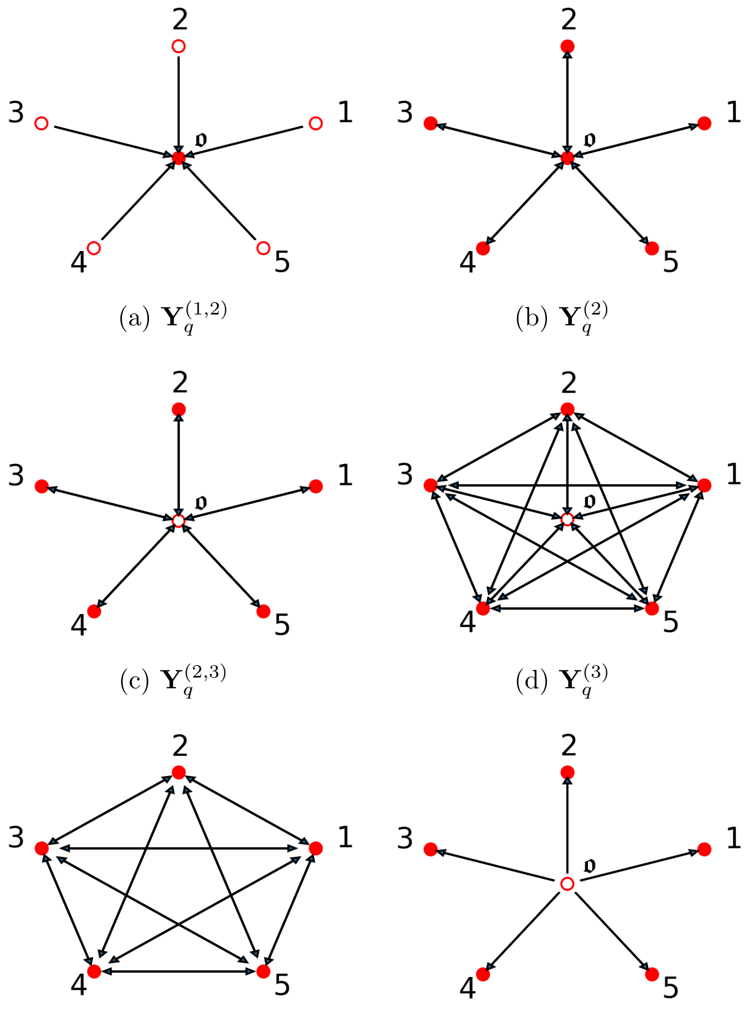
The proof follows from Theorems 5.3.4 and 5.3.6, Proposition 5.5.3, and [54, Theorem 2.2, Remarks 2.10, 2.11].

Remark 5.4.6. *Notably, in contrast to the case $q \leq 4$, we can describe the metastable behavior for all $\beta \in (\beta_1, \infty)$ since the saddle points are nondegenerate when $\beta = \beta_3$.*

We can now provide a more intuitive explanation of Theorem 5.4.5. See Figure 5.6 also for the description of metastable behavior. Note that if $\beta_2 < \beta < q$, there are two time scales.

- $\mathbf{Y}_q^{(1,2)}$: If $\beta_1 < \beta < \beta_2$, in the time scale $2\pi N e^{\theta_1}$, $\boldsymbol{r}_N(\cdot)$ starting from $\mathcal{E}_N^S$, reaches $\mathcal{E}_N^{\mathbf{o}}$ and stays there forever. When it goes from $\mathcal{E}_N^k$, $k \in S$, to $\mathcal{E}_N^{\mathbf{o}}$, it visits the neighborhood of $\mathbf{v}_1^k$.

- $\mathbf{Y}_q^{(2)}$: If $\beta = \beta_2$, in the time scale $2\pi N e^{\theta_1}$, the process $\boldsymbol{r}_N(\cdot)$ goes around each well in $\mathcal{E}_N^{\widehat{S}}$. However, when $\boldsymbol{r}_N(\cdot)$ starting from $\mathcal{E}_N^k$, $k \in S$, goes to $\mathcal{E}_N^l$, $l \in S \setminus \{k\}$, it must pass through $\mathcal{E}_N^{\mathbf{o}}$ and as in the case $\beta \in (\beta_1, \beta_2)$, it visits the neighborhood of $\mathbf{v}_1^k$ and then the neighborhood of $\mathbf{v}_1^l$.



(a) $\mathbf{Y}_q^{(1,2)}$ (b) $\mathbf{Y}_q^{(2)}$

(c) $\mathbf{Y}_q^{(2,3)}$ (d) $\mathbf{Y}_q^{(3)}$

(e) $\mathbf{Y}_q^{(3,\infty)}$ (f) $\mathbf{Y}_q^{(4)}$

Figure 5.6: Figure about metastability when $q = 5$

- $\mathbf{Y}_q^{(2,3)}$: If $\beta_2 < \beta < \beta_3$, in the time scale $2\pi N e^{\theta_1}$, the process $\boldsymbol{r}_N(\cdot)$ starting from $\mathcal{E}_N^S$ travels $\mathcal{E}_N^{\widehat{S}}$, however, it stays in $\mathcal{E}_N^{\mathbf{o}}$ in negligible time. Furthermore, when $\boldsymbol{r}_N(\cdot)$ goes from $\mathcal{E}_N^k$, $k \in S$, to $\mathcal{E}_N^l$, $l \in S \setminus \{k\}$, it must visit $\mathcal{E}_N^{\mathbf{o}}$.

- $\mathbf{Y}_q^{(3)}$: If $\beta = \beta_3$, in the time scale $2\pi N e^{\theta_1}$, the process $\boldsymbol{r}_N(\cdot)$ starting from $\mathcal{E}_N^S$ travels $\mathcal{E}_N^{\widehat{S}}$, however, it stays in $\mathcal{E}_N^{\mathbf{o}}$ in negligible time. Furthermore, there are two ways in which $\boldsymbol{r}_N(\cdot)$ goes from $\mathcal{E}_N^k$, $k \in S$, to $\mathcal{E}_N^l$, $l \in S \setminus \{k\}$. First, it goes to $\mathcal{E}_N^l$ directly and must pass through the neighborhood of $\mathbf{u}_2^{k,l}$. Second, it visits $\mathcal{E}_N^{\mathbf{o}}$ and stays there for a negligible period of time.

- $\mathbf{Y}_q^{(3,\infty)}$: If $\beta > \beta_3$, in the time scale $2\pi N e^{\theta_1}$, the process $\boldsymbol{r}_N(\cdot)$ starting from $\mathcal{E}_N^S$ travels $\mathcal{E}_N^S$ without visiting $\mathcal{E}_N^{\mathbf{o}}$. As in the case $\beta = \beta_3$, it must pass through the neighborhood of $\mathbf{u}_2^{k\,l}$, $k, l \in S$, when it goes from $\mathcal{E}_N^k$ to $\mathcal{E}_N^l$.

- $\mathbf{Y}_q^{(4)}$: If $\beta_2 < \beta \le \beta_3$, in the second time scale $2\pi N e^{\theta_{\mathbf{o}}}$, the process $\boldsymbol{r}_N(\cdot)$ starting from $\mathcal{E}_N^{\mathbf{o}}$, goes to $\mathcal{E}_N^k$, $k \in S$, and stays there forever. As $\mathbf{Y}_q^{(1,2)}$, $\mathbf{Y}_q^{(2)}$, and $\mathbf{Y}_q^{(2,3)}$, it passes through the neighborhood of $\mathbf{v}_1^k$ when it moves to $\mathcal{E}_N^k$ from $\mathcal{E}_N^{\mathbf{o}}$.

- $\mathbf{Y}_q^{(5)}$: If $\beta_3 < \beta < q$, in the second time scale $2\pi N e^{\theta_{\mathbf{o}}}$, the process $\boldsymbol{r}_N(\cdot)$ starting from $\mathcal{E}_N^{\mathbf{o}}$, goes to $\mathcal{E}_N^S$ and stays there forever. This dynamics is similar to $\mathbf{Y}_q^{(4)}$; however, $\mathcal{E}_N^k$, $k \in S$, are not distinguishable.

5.4.4 Eyring–Kramers formulae

In this subsection, we present the Eyring–Kramers formula with regard to metastable behavior.

Before we state the result, we introduce some notations. Let $[\boldsymbol{x}]_N$ be the nearest point in Ξ_N of $\boldsymbol{x} \in \Xi$. If there is more than one such point, one of them is chosen arbitrarily. For $\mathcal{A} \subset \Xi$, define $[\mathcal{A}]_N$ as

$$[\mathcal{A}]_N \;=\; \{[\boldsymbol{x}]_N \,:\, \boldsymbol{x} \in \mathcal{A}\}\ .$$

Denote by $H_{\mathcal{A}}$ the hitting time of the set $[\mathcal{A}]_N$ by the process $\boldsymbol{r}_N(\cdot)$:

$$H_{\mathcal{A}} \;=\; \inf\{t > 0 \,:\, \boldsymbol{r}_N(t) \in [\mathcal{A}]_N\}\ .$$

If $\mathcal{A} = \{\boldsymbol{x}\}$, we simply write $H_{\mathcal{A}} = H_{\boldsymbol{x}}$.

We have the following theorem.

Theorem 5.4.7. *Let $q \ge 3$. We have the following.*

1. *For $\beta_1 < \beta \le \beta_2$ and $k \in S$, we have*

$$\mathbb{E}^{N,\,\beta}_{\mathbf{u}_1^k}[H_{\mathbf{p}}] \;=\; [1 + o_N(1)]\frac{\nu_1}{\omega_{\mathbf{0}}}2\pi N\exp(N\theta_1)\ .$$

2. *For $\beta_2 \le \beta < q$, we have*

$$\mathbb{E}^{N,\,\beta}_{\mathbf{p}}[H_{\mathcal{U}_1}] \;=\; [1 + o_N(1)]\frac{\nu_{\mathbf{0}}}{q\omega_{\mathbf{0}}}2\pi N\exp(N\theta_0)\ .$$

3. *For $\beta > \beta_3$ and $k \in S$, we have*

$$\mathbb{E}^{N,\,\beta}_{\mathbf{u}_1^k}[H_{\mathcal{U}_1\setminus\{\mathbf{u}_1^k\}}] \;=\; [1 + o_N(1)]\frac{\nu_1}{(q-1)\omega_1}2\pi N\exp(N\theta_1)\ .$$

The proof follows from Theorems 5.3.4-5.3.6, Proposition 5.5.3, and [54, Theorem 2.5, Remarks 2.10, 2.11].

Because of the Eyring–Kramers formula, we can derive the large-deviation-type estimates on spectral gap and mixing time. To explain this more concretely, let λ_N be the spectral gap of the process $\boldsymbol{r}_N(\cdot)$ (which will be defined explicitly in (5.20)), and define the mixing time as

$$t_{\rm mix}^N = t_{\rm mix}^N(\delta) := \inf\left\{ t>0 \,:\, \sup_{\boldsymbol{x}\in\Xi_N} \left\| P^t(\boldsymbol{x},\cdot) - \nu_N^\beta \right\|_{\rm TV} < \delta \right\} \quad ; \; \delta \in (0,\,1) \;,$$

where $P^t(\boldsymbol{x},\cdot)$ is a distribution of $\boldsymbol{r}_N(t)$ with initial condition $\boldsymbol{r}_N(0) = \boldsymbol{x}$ and $\|\cdot\|_{\rm TV}$ denotes the total variation distance defined by

$$\|\mu-\nu\|_{TV} := \frac{1}{2}\sum_{\boldsymbol{x}\in\Xi_N} |\mu(\boldsymbol{x}) - \nu(\boldsymbol{x})|$$

for any two probability measures μ and ν on Ξ_N. Then, by the arguments in [13, 85], we can observe that

$$\lim_{N\to\infty} \frac{1}{N}\log t_{\rm mix}^N = \lim_{N\to\infty}\frac{1}{N}\log\frac{1}{\lambda_N} = \max\{\theta_{\mathfrak{o}},\,\theta_1\}\;.$$

Note that the Eyring–Kramers type estimate on the spectral gap cannot follow immediately from the results of [13, 85] since there are several valleys with same depth.

5.5 Preliminary analysis on potential and generator

In this section, we conduct several preliminary analyses. In Section 5.5.1, we prove Proposition 5.2.1. In particular, we compute the jump rates of Markov chain $\boldsymbol{r}_N(\cdot)$. In Section 5.5.2, we decompose the generator $\mathscr{L}_N$ into several simple generators $\mathscr{L}_{N,\boldsymbol{x}}^{i,j}$, $\boldsymbol{x}\in\Xi_N$, $i,j\in S$. Via this decomposition of $\mathscr{L}_N$, we can handle $\mathscr{L}_N$ using the method developed in [54] since our model is a special case of [54, Remarks 2.10, 2.11]; this correspondence will be explained in Section 5.5.3.

5.5.1 Dynamics of proportion vector.

We prove Proposition 5.2.1 in this section.

Proof of Proposition 5.2.1. Let $\boldsymbol{e}_j^N := \frac{1}{N}\boldsymbol{e}_j$, $j \in S$ (cf. Notation 5.3.1). Fix configurations $\sigma, \tau \in \Omega_N$ such that $\boldsymbol{r}_N(\sigma) = \boldsymbol{r}_N(\tau)$ and let

$$\boldsymbol{x} = (x_1,\, \dots,\, x_{q-1}) = \boldsymbol{r}_N(\sigma) = \boldsymbol{r}_N(\tau) \in \Xi_N \ .$$

For some sites $u_1, u_2, v_1, v_2 \in K_N$ such that $\sigma_{u_1} = \sigma_{v_1} = \tau_{u_2} = \tau_{v_2}$, let $i = \sigma_{v_1}$. Then, the Markovity of the process $\boldsymbol{r}_N(t)$ can be inferred from the identity

$$\begin{aligned} c_{u_1,\,j}(\sigma) = \ & c_{v_1,\,j}(\sigma) \ = \ c_{u_2,\,j}(\tau) \ = \ c_{v_2,\,j}(\tau) \\ & = \ \exp\left\{-\frac{N\beta}{2}[H(\boldsymbol{r}_N(\sigma^{v_1,\,j})) - H(\boldsymbol{r}_N(\sigma))]\right\} \\ & = \ \exp\left\{-\frac{N\beta}{2}[H(\boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N) - H(\boldsymbol{x})]\right\} \ , \end{aligned}$$

for $j \neq i$. Hence, $\boldsymbol{r}_N(\cdot)$ is a Markov chain.

Since there are Nx_i sites whose spins are i, the jump rate $R_N(\cdot,\cdot)$ of $\boldsymbol{r}_N(\cdot)$ can be written as

$$R_N(\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N) \ = \ \frac{Nx_i}{N} \exp\left\{-\frac{N\beta}{2}[H(\boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N) - H(\boldsymbol{x})]\right\} \ . \tag{5.19}$$

Hence, the generator $\mathscr{L}_N$ of $\boldsymbol{r}_N(\cdot)$ is given as

$$\mathscr{L}_N f(\boldsymbol{x}) \ = \ \sum_{i,\,j \in S} R_N(\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N)\,[f(\boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N) - f(\boldsymbol{x})] \ , \tag{5.20}$$

for $f : \Xi_N \to \mathbb{R}$.

Finally, this dynamics is reversible with respect to ν_β^N since we have the following detailed balance condition

$$\nu_\beta^N(\boldsymbol{x})\, R_N(\boldsymbol{x}, \boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N) \ = \ \nu_\beta^N(\boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N)\, R_N(\boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N, \boldsymbol{x}) \ ,$$

so that ν_β^N is the invariant measure. □

5.5.2 Cyclic decomposition

For $1 \le i < j \le q$, let $\gamma_N^{i,j}$ be the cycle $(\boldsymbol{e}_0^N, \boldsymbol{e}_j^N - \boldsymbol{e}_i^N, \boldsymbol{e}_0^N)$ of length 2 on $(\mathbb{Z}/N)^{q-1}$ and let $(\gamma_N^{i,j})_{\boldsymbol{x}} = \boldsymbol{x} + \gamma_N^{i,j}$. Define $\widehat{\Xi}_N^{i,j}$ as

$$\widehat{\Xi}_N^{i,j} \;=\; \{\boldsymbol{x} \in \Xi_N \,:\, (\gamma_N^{i,j})_{\boldsymbol{x}} \subset \Xi_N\} \;=\; \{\boldsymbol{x} \in \Xi_N \,:\, x_j \le 1 - N^{-1},\, x_i \ge N^{-1}\}\ .$$

Define a jump rate $\widetilde{R}_N^{i,j}$ associated with this cycle as

$$\begin{aligned}\widetilde{R}_{N,0}^{i,j}(\boldsymbol{x}) \;&=\; \exp\Big\{-N\beta[\overline{F}_{\beta,N}^{i,j}(\boldsymbol{x}) - F_{\beta,N}(\boldsymbol{x})]\Big\}\ ,\\ \widetilde{R}_{N,1}^{i,j}(\boldsymbol{x}) \;&=\; \exp\Big\{-N\beta[\overline{F}_{\beta,N}^{i,j}(\boldsymbol{x}) - F_{\beta,N}(\boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N)]\Big\}\ ,\end{aligned}$$

where

$$\overline{F}_{\beta,N}^{i,j}(\boldsymbol{x}) \;=\; \frac{1}{2}[F_{\beta,N}(\boldsymbol{x}) + F_{\beta,N}(\boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N)]\ .$$

Let $\mathscr{L}_{N,\boldsymbol{x}}^{i,j}$, $\boldsymbol{x} \in \widehat{\Xi}_N$, be a generator acting on $f : \Xi_N \to \mathbb{R}$ as

$$(\mathscr{L}_{N,\boldsymbol{x}}^{i,j} f)(\boldsymbol{y}) = \begin{cases} \widetilde{R}_{N,0}^{i,j}(\boldsymbol{x}) \Big[f(\boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N) - f(\boldsymbol{x})\Big] & \boldsymbol{y} = \boldsymbol{x}\ ,\\ \widetilde{R}_{N,1}^{i,j}(\boldsymbol{x}) \Big[f(\boldsymbol{x}) - f(\boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N)\Big] & \boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}_j^N - \boldsymbol{e}_i^N\ ,\\ 0 & \text{otherwise}\ . \end{cases} \tag{5.21}$$

Here, $\mathscr{L}_{N,\boldsymbol{x}}^{i,j}$ can be regarded as a generator of a Markov chain on the cycle $(\gamma_N^{i,j})_{\boldsymbol{x}}$.

Let

$$w^{i,j}(\boldsymbol{x}) \;:=\; \sqrt{x_i x_j}\ ,\ \text{and}\ \ w_N^{i,j}(\boldsymbol{x}) \;:=\; \sqrt{x_i(x_j + \frac{1}{N})}\ .$$

By (5.3), we can write

$$\exp\{-\beta N[F_{\beta,N}(\boldsymbol{x}) - H(\boldsymbol{x})]\} \;=\; (2\pi N)^{(q-1)/2}\binom{N}{(Nx_1)\cdots(Nx_q)}\ .$$

By elementary computations, we obtain

$$R_N(\boldsymbol{x}, \boldsymbol{x}+\boldsymbol{e}_j^N-\boldsymbol{e}_i^N)/\widetilde{R}_{N,0}^{i,j}(\boldsymbol{x}) \;=\; R_N(\boldsymbol{x}+\boldsymbol{e}_j^N-\boldsymbol{e}_i^N, \boldsymbol{x})/\widetilde{R}_{N,1}^{i,j}(\boldsymbol{x}) \;=\; w_N^{i,j}(\boldsymbol{x}) \;,$$

so that

$$\begin{aligned}
&R_N(\boldsymbol{x}, \boldsymbol{x}+\boldsymbol{e}_j^N-\boldsymbol{e}_i^N)[f(\boldsymbol{x}+\boldsymbol{e}_j^N-\boldsymbol{e}_i^N)-f(\boldsymbol{x})] \\
&\; = w_N^{i,j}(\boldsymbol{x})\, \mathscr{L}_{N,\boldsymbol{x}}^{i,j} f(\boldsymbol{x}) \quad \text{and} \\
&R_N(\boldsymbol{x}, \boldsymbol{x}+\boldsymbol{e}_i^N-\boldsymbol{e}_j^N)[f(\boldsymbol{x}+\boldsymbol{e}_i^N-\boldsymbol{e}_j^N)-f(\boldsymbol{x})] \\
&\; = w_N^{i,j}(\boldsymbol{x}+\boldsymbol{e}_i^N-\boldsymbol{e}_j^N)\, \mathscr{L}_{N,\boldsymbol{x}+\boldsymbol{e}_i^N-\boldsymbol{e}_j^N}^{i,j} f(\boldsymbol{x}) \;.
\end{aligned}$$

Hence, by (5.21), we can write

$$\begin{aligned}
\mathscr{L}_N f(\boldsymbol{x}) &= \sum_{1\le i<j\le q} [w_N^{i,j}(\boldsymbol{x})\, \mathscr{L}_{N,\boldsymbol{x}}^{i,j} + w_N^{i,j}(\boldsymbol{x}+\boldsymbol{e}_i^N-\boldsymbol{e}_j^N)\, \mathscr{L}_{N,\boldsymbol{x}+\boldsymbol{e}_i^N-\boldsymbol{e}_j^N}^{i,j}] f(\boldsymbol{x}) \\
&= \sum_{1\le i<j\le q} \sum_{\boldsymbol{y}\in\widehat{\Xi}_N^{i,j}} w_N^{i,j}(\boldsymbol{y}) \mathscr{L}_{N,\boldsymbol{y}}^{i,j} f(\boldsymbol{x}) \;. \qquad (5.22)
\end{aligned}$$

Since $w_N^{i,j}$ converges uniformly to $w^{i,j}$ and is uniformly Lipschitz on every compact subset of $\operatorname{int}\Xi$, our model is a special case of [54, Remarks 2.10, 2.11] provided that several technical requirements are verified. These requirements will be verified in the next subsection.

Remark 5.5.1. *[54, Section 2] assumes that for $\gamma_N^{i,j} = (\boldsymbol{z}_0, \boldsymbol{z}_1)$, $\boldsymbol{z}_1 - \boldsymbol{z}_0$ generates $\mathbb{Z}^{q-1}$; however, this requirement is needed to make $\boldsymbol{r}_N(\cdot)$ be irreducible. Since $(\gamma_N^{i,j})_{1\le i<j\le q}$ generate $\mathbb{Z}^{q-1}$, we do not need this assumption.*

5.5.3 Requirements for $F_{\beta,N}$ and $\mathscr{L}_N$

In this subsection, we verify that our model is a special case of [54, Remarks 2.10, 2.11].

First, we give some properties of $F_\beta(\cdot)$ and $G_{\beta,N}(\cdot)$. By the following proposition, $F_\beta(\cdot)$ and $G_{\beta,N}(\cdot)$ fulfill the requirements in the first paragraph of [54, Section 2].

Proposition 5.5.2. *The functions $F_\beta(\cdot)$ and $G_{\beta,N}(\cdot)$ satisfy the following properties.*

1. *F_β is twice-differentiable and there is no critical points at $\partial\,\Xi$. For all $\boldsymbol{x} \in \partial\Xi$, $\nabla F_\beta(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}) > 0$.*

2. *The second partial derivatives of $F_\beta(\cdot)$ are Lipschitz-continuous on every compact subset of int Ξ.*

3. *On each compact subset of int Ξ, $G_{\beta,N}(\cdot)$ is uniformly Lipschitz and converges uniformly to $G_\beta(\boldsymbol{x}) := (1/2\beta)\log(x_1 \cdots x_q)$ as $N \to \infty$.*

4. *There are finitely many critical points of $F_\beta(\cdot)$.*

Proof. (1)-(3) are straightforward. By Lemma 5.6.2 in Section 5.6.1, there are finitely many critical points of $F_\beta(\cdot)$. □

Next, fix one of saddle points $\mathbf{s}$. Note that $\nabla^2 F_\beta(\mathbf{s})$ has a unique negative eigenvalue. As in [54, Section 4.3], define $(\mathscr{L}_N^{i,j})^{\mathbf{s}}$ as

$$(\mathscr{L}_N^{i,j})^{\mathbf{s}} f(\boldsymbol{x}) \;=\; \frac{1}{N^2}(\boldsymbol{e}_j - \boldsymbol{e}_i)^\dagger \nabla^2 f(\boldsymbol{x})(\boldsymbol{e}_j - \boldsymbol{e}_i) - \frac{1}{N}\mathbb{A}^{i,j}\nabla^2 F(\mathbf{s})(\boldsymbol{x} - \mathbf{s}) \cdot \nabla f(\boldsymbol{x}) \,.$$

Denote by $-\lambda_1^{\mathbf{s}}, \lambda_2^{\mathbf{s}}, \dots, \lambda_{q-1}^{\mathbf{s}}$ the eigenvalues of $\nabla^2 F_\beta(\mathbf{s})$ where $\lambda_i^{\mathbf{s}} > 0$ for all $i = 1, \dots\, q-1$, and by $\boldsymbol{a}_1^{\mathbf{s}}, \boldsymbol{a}_2^{\mathbf{s}}, \dots, \boldsymbol{a}_{q-1}^{\mathbf{s}}$ the corresponding eigenvectors. Let $\epsilon_N := N^{-2/5} \ll N^{-1}$ so that ϵ_N satisfies [54, displays (4.7), (4.8)]. Define a neighborhood of $\mathbf{s}$ as

$$\mathcal{C}_N^{\mathbf{s}} \;:=\; \left\{ \mathbf{s} + \sum_{k=1}^{q-1} x_k \mathbf{a}_k^{\mathbf{s}} : |x_1| \le \epsilon_N,\, |x_k| \le \sqrt{\frac{2\lambda_1^{\mathbf{s}}}{\lambda_k}}\epsilon_N,\, 2 \le k \le q-1 \right\} \cap \Xi_N \;.$$

Then, by the next proposition, definitions (5.15)-(5.18) are consistent with [54, Remarks 2.10, 2.11].

Proposition 5.5.3. *For a smooth function* $f : \Xi \to \mathbb{R}$*, we have uniformly on* $\mathcal{C}_N^{\mathbf{s}}$*,*

$$\mathscr{L}_N f \;=\; [1+O(\epsilon_N)] \sum_{1\le i<j\le q} w^{i,j}(\mathbf{s})(\mathscr{L}_N^{i,j})^{\mathbf{s}} f \;.$$

Proof. Since $|\boldsymbol{x}-\mathbf{s}| = O(\epsilon_N)$, by (5.21) and the second order Taylor expansion on $\mathcal{C}_N^{\mathbf{s}}$, we have

$$\sum_{\boldsymbol{y}\in\widehat{\Xi}_N^{i,j}} \mathscr{L}_{N,\boldsymbol{y}}^{i,j} f(\boldsymbol{x}) \;=\; [1+O(\epsilon_N)](\mathscr{L}_N^{i,j})^{\mathbf{s}} f(\boldsymbol{x}) \;.$$

Hence, on $\mathcal{C}_N^{\mathbf{s}}$, since $w_N^{i,j}(\boldsymbol{x}) = [1+O(N^{-1})]w^{i,j}(\boldsymbol{x}) = [1+O(\epsilon_N)]w^{i,j}(\mathbf{s})$, we have

$$\begin{aligned}
\mathscr{L}_N f(\boldsymbol{x}) \;&=\; \sum_{1\le i<j\le q} \sum_{\boldsymbol{y}\in\widehat{\Xi}_N^{i,j}} w_N^{i,j}(\boldsymbol{y}) \mathscr{L}_{N,\boldsymbol{y}}^{i,j} f(\boldsymbol{x}) \\
&=\; [1+O(\epsilon_N)] \sum_{1\le i<j\le q} w^{i,j}(\mathbf{s}) \sum_{\boldsymbol{y}\in\widehat{\Xi}_N^{i,j}} \mathscr{L}_{N,\boldsymbol{y}}^{i,j} f(\boldsymbol{x}) \\
&=\; [1+O(\epsilon_N)] \sum_{1\le i<j\le q} w^{i,j}(\mathbf{s})(\mathscr{L}_N^{i,j})^{\mathbf{s}} f(\boldsymbol{x}) \;.
\end{aligned}$$

□

5.6 Investigation of critical points and temperatures

This section is devoted to the investigation of critical points and temperatures including their definitions. We will provide a preliminary analysis of the critical points in Section 5.6.1 and of the critical temperatures in section 5.6.2.

5.6.1 Classification of critical points

We recall that

$$F_\beta(\boldsymbol{x}) \;=\; -\frac{1}{2}\sum_{k=1}^{q} x_k^2 \;+\; \frac{1}{\beta}\sum_{k=1}^{q} x_k \log x_k \;,$$

and that $x_q = 1-(x_1+\cdots+x_{q-1})$. For $1 \le k \le q-1$,

$$\frac{\partial}{\partial x_k} F_\beta(\boldsymbol{x}) \;=\; -(x_k - x_q) \;+\; \frac{1}{\beta}(\log x_k - \log x_q) \;.$$

If $\dfrac{\partial}{\partial x_k} F_\beta(\boldsymbol{x}) = 0$, we must have $x_k - \dfrac{1}{\beta}\log x_k = x_q - \dfrac{1}{\beta}\log x_q$. Hence,

$$\nabla F_\beta(\boldsymbol{x}) = 0 \text{ if and only if } x_k - \frac{1}{\beta}\log x_k,\; 1 \le k \le q,\; \text{are the same} \;. \tag{5.23}$$

By (5.23), $\mathbf{p} = (1/q,\, \dots,\, 1/q)$ is a critical point.

By elementary computation, we can check that the equation $t - \dfrac{1}{\beta}\log t = c$ has at most two positive real solutions for fixed β, $c > 0$. Hence, if $(x_1, \dots, x_q)$ is a critical point[5], x_k's can have at most 2 values by (5.23). Hereafter, we assume $\mathbf{c}$ is a critical point and

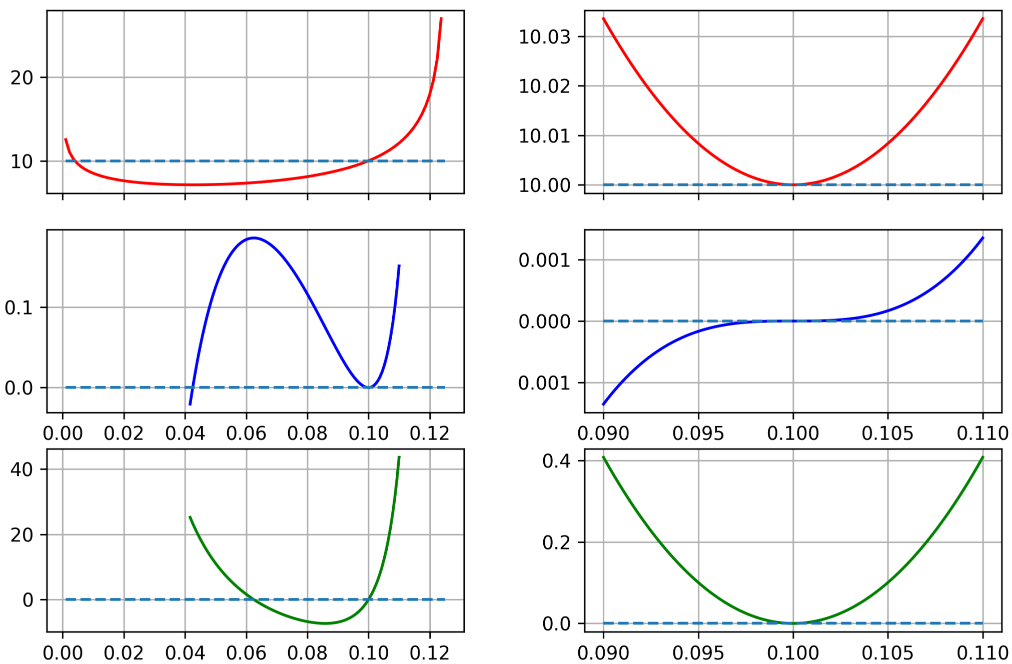
$$\mathbf{c} \;=\; (t,\, \dots,\, t, (1-jt)/i,\, \dots,\, (1-jt)/i) \;,$$

where j is the number of t's and $i = q - j$. Observe that by symmetry, each permutation of coordinates of $\mathbf{c}$ has the same properties. Without loss of generality, we may assume

$$1 \le i \le q/2 \le j \le q-1 \text{ and } t \ne 1/q \;.$$

The point $\mathbf{p}$ will be analyzed separately.

[5] Recall Notaion 5.3.1.



(a) $i = 2$.

(b) $i = q/2$.

Figure 5.7: Graphs of $g_i(t)$, $h_i(t)$, and $h_i'(t)$ when $q = 10$.

By (5.23), we obtain

$$t - \frac{1}{\beta}\log t \;=\; \frac{1-jt}{i} - \frac{1}{\beta}\log\left(\frac{1-jt}{i}\right) \;,$$

which implies

$$\beta \;=\; \frac{i}{1-qt}\log\left(\frac{1-jt}{it}\right) \;=\; g_i(t)\;. \tag{5.24}$$

Lemma 5.6.1. *Fix $q \geq 3$, $1 \leq i \leq q/2$ and $j = q - iP$. Then, the function $g_i : (0, 1/j) \to \mathbb{R}$ has the unique minimum, say m_i. Furthermore, if $\beta > g_i(m_i)$, $\beta = g_i(t)$ has two solutions.*

Proof. Define $h_i : (0, 1/j) \to \mathbb{R}$ as[6]

$$h_i(t) \;:=\; \log\frac{1-jt}{it} + \frac{qt-1}{qt(1-jt)}\;. \tag{5.25}$$

By elementary computation, we obtain

$$g_i'(t) \;=\; \frac{qi}{(1-qt)^2}h_i(t) \;\;\text{and}\;\; h_i'(t) \;=\; \frac{(qt-1)(2jt-1)}{q(1-jt)^2t^2}\;. \tag{5.26}$$

There are two cases, where $i < q/2$ and $i = q/2$. By elementary computation, we can show that the graphs of g_i, h_i, h_i' are given by Figure 5.7, which completes the proof. □

For $1 \leq i \leq q/2$, let

$$\beta_{s,\,i} = \beta_{s,\,i}(q) \;:=\; g_i(m_i)\;, \tag{5.27}$$

where m_i is the unique minimum of $g_i(\cdot)$ given in the above lemma.

If $\beta \geq \beta_{s,\,i}$, there are one or two solutions of $\beta = g_i(t)$ which will be

[6] As $g_i(\cdot)$; the function $h_i(\cdot)$ can be continuously extended to $(1, 1/j)$.

denoted by $u_i = u_i(\beta)$, $v_i = v_i(\beta)$ where $u_i \le v_i$. Let

$$\mathcal{U}_i = \mathcal{U}_i(\beta) = \{\text{permutations of } (u_i, \ldots, u_i, (1-ju_i)/i, \ldots, (1-ju_i)/i)\} ,$$
$$\mathcal{V}_i = \mathcal{V}_i(\beta) = \{\text{permutations of } (v_i, \ldots, v_i, (1-jv_i)/i, \ldots, (1-jv_i)/i)\} ,$$

for $\beta \ge \beta_{s,i}$. We have the following candidates of the critical points of F_β.

Lemma 5.6.2. *A critical point of F_β is exactly one of the following cases.*

1. $\mathbf{p} = (1/q, \ldots, 1/q)$.
2. *For $1 \le i \le q/2$ and $\beta \in (\beta_{s,i}, \infty)$, elements of $\mathcal{U}_i$.*
3. *For $1 \le i \le q/2$ and $\beta \in (\beta_{s,i}, \infty) \setminus \{q\}$, elements of $\mathcal{V}_i$.*
4. *For $1 \le i < q/2$ and $\beta = \beta_{s,i}$, elements of $\mathcal{U}_i = \mathcal{V}_i$.*

Proof. By part (1) of Proposition 5.5.2, points in $\partial\Xi$ cannot be critical points. Then, the proof follows from (5.23) and Lemma 5.6.1. □

Finally, we have the following results for critical points. The proof for $q = 3$ is given in [53, Proposition 4.2].

Proposition 5.6.3. *The minima and saddle points of F_β for $q = 3$, $q = 4$, and $q \ge 5$ are given by Tables 5.1, 5.2, and 5.3, respectively.*

	$\mathbf{p}$	$\mathcal{U}_1(\beta)$	$\mathcal{V}_1(\beta)$
$\beta \in (0, \beta_{s,1})$	minimum		
$\beta = \beta_{s,1}$	minimum	degenerate	degenerate
$\beta \in (\beta_{s,1}, q)$	minimum	minima	saddle points
$\beta = q$	degenerate	minima	degenerate
$\beta \in (q, \infty)$	maximum	minima	saddle points

Table 5.1: Classification of critical points when $q = 3$

	$\mathbf{p}$	$\mathcal{U}_1(\beta)$	$\mathcal{V}_1(\beta)$	$\mathcal{U}_2(\beta)=\mathcal{V}_2(\beta)$
$\beta \in (0, \beta_{s,1})$	minimum			
$\beta = \beta_{s,1}$	minimum	degenerate	degenerate	
$\beta \in (\beta_{s,1}, q)$	minimum	minima	saddle points	
$\beta = q$	degenerate	minima	degenerate	degenerate
$\beta \in (q, \infty)$	maximum	minima	index ≥ 2	saddle points

Table 5.2: Classification of critical points when $q = 4$

	$\mathbf{p}$	$\mathcal{U}_1(\beta)$	$\mathcal{V}_1(\beta)$	$\mathcal{U}_2(\beta)$
$\beta \in (0, \beta_{s,1})$	minimum			
$\beta = \beta_{s,1}$	minimum	degenerate	degenerate	
$\beta \in (\beta_{s,1}, \beta_{s,2})$	minimum	minima	saddle points	
$\beta = \beta_{s,2}$	minimum	minima	saddle points	degenerate
$\beta \in (\beta_{s,2}, q)$	minimum	minima	saddle points	saddle points
$\beta = q$	degenerate	minima	degenerate	degenerate
$\beta \in (q, \infty)$	maximum	minima	index ≥ 2	saddle points

Table 5.3: Classification of critical points when $q = 5$

Section 5.7 proves the above proposition. Until now, we classified all minima and saddle points for all $q \geq 3$.

5.6.2 Definition of critical temperatures

In the previous subsection, we defined several temperatures $\beta_{s,i}$, $1 \leq i \leq q/2$. In this subsection, we prove several properties of such temperatures and moreover introduce new temperatures. Then, we select the critical temperatures at which phase transitions occur.

The first lemma is about the order of $\beta_{s,i}$.

Lemma 5.6.4. *We have* $\beta_{s,1} < \beta_{s,2} < \cdots < \beta_{s,\lfloor q/2 \rfloor}$. *If* q *is even, we have* $\beta_{s,q/2} = q$ *and otherwise,* $\beta_{s,\lfloor q/2 \rfloor} < q$.

Proof. In this proof, we regard i as a real number and claim that $g_i(t)$ increases as $i \in [1, q]$ increases for fixed $t < 1/q$. By elementary computation, we obtain

$$\frac{d}{di} g_i(t) \; = \; \frac{1}{1-qt} \Big(\log \frac{1-jt}{it} + \frac{it}{1-jt} - 1 \Big) \; .$$

By the inequality $x - 1 > \log x$, we can conclude that $\frac{d}{di} g_i(t) > 0$. Hence, $g_i(t) < g_{i+1}(t)$ if $t < 1/q$.

Hereafter, let $i \in \mathbb{Z}$. Suppose $i < q/2 - 1$. Since m_i, $m_{i+1} < 1/q$, we obtain

$$\beta_{s,i} = g_i(m_i) \le g_i(m_{i+1}) < g_{i+1}(m_{i+1}) = \beta_{s,i+1} \; ,$$

by the above claim. The first inequality holds since m_i is a minimum of g_i. If $i = q/2 - 1$, since $m_i < 1/q = m_{i+1}$, we obtain

$$\beta_{s,i} = g_i(m_i) < g_i(m_{i+1}) = q = \beta_{s,q/2} \; .$$

If $i < q/2$, we have $m_i < 1/q$ so that $\beta_{s,i} < g_i(1/q) = q$. This with the above argument prove the second assertion. □

Remark. *In particular, by the above lemma, we have* $\beta_{s,1} < \beta_{s,2} = q$ *for* $q = 4$ *and* $\beta_{s,1} < \beta_{s,2} < q$ *for* $q \ge 5$.

The relative order of heights of critical points changes with changes in β, and the phase transition is owing to this fact. We will explain when and how this order is changed. Since the proofs are technical, they are postponed to Section 5.8.

Order of heights of p and $\mathcal{U}_1$

Define β_c as

$$\beta_c(q) \; := \; \frac{2(q-1)}{q-2} \log(q-1) \; , \tag{5.28}$$

which is introduced in [18, display (3.3)]. Then, we obtain the following.

Lemma 5.6.5. *For $q \geq 3$, we have $\beta_{s,1} < \beta_c$ and for $q \geq 4$, we have $\beta_{s,1} < \beta_c < \beta_{s,2}$.*

The proof of the lemma is given in Section 5.8.1. The following lemma is an important property of β_c.

Lemma 5.6.6. *Let $q \geq 3$. Then, we have*

$$\begin{cases} F_\beta(\mathbf{p}) < F_\beta(\mathbf{u}_1) & \text{if } \beta \in (\beta_{s,1}, \beta_c) \ , \\ F_\beta(\mathbf{p}) = F_\beta(\mathbf{u}_1) & \text{if } \beta = \beta_c \ , \\ F_\beta(\mathbf{p}) > F_\beta(\mathbf{u}_1) & \text{if } \beta \in (\beta_c, \infty) \ . \end{cases} \tag{5.29}$$

This result is the same as [18, Theorem 3.1(b)]. The proof is provided in [18, Appendices A, B] via convex-duality.

We may assume that β increases from a very small positive number. Observe that the elements of $\mathcal{U}_1$ and $\mathcal{V}_1$ simultaneously appear when $\beta = \beta_{s,1}$ and the elements of $\mathcal{U}_2$ appear when $\beta = \beta_{s,2}$. By the above two lemmas, before the appearance of critical points in $\mathcal{U}_2$, the heights of $\mathbf{p}$ and $\mathbf{u}_1$ are reversed.

Order of heights of $\mathcal{V}_1$ and $\mathcal{U}_2$

We have the following lemma about the heights of $\mathbf{u}_2$ and $\mathbf{v}_1$. The critical temperature β_m given in the following lemma is the crucial development of this chapter.

Lemma 5.6.7. *Let $q \geq 5$. We have a critical temperature $\beta_m \in (\beta_{s,2}, q)$ such that*

$$\begin{cases} F_\beta(\mathbf{v}_1) < F_\beta(\mathbf{u}_2) & \text{if } \beta_{s,2} \leq \beta < \beta_m \ , \\ F_\beta(\mathbf{v}_1) = F_\beta(\mathbf{u}_2) & \text{if } \beta = \beta_m \ , \\ F_\beta(\mathbf{v}_1) > F_\beta(\mathbf{u}_2) & \text{if } \beta_m < \beta \leq q \ . \end{cases} \tag{5.30}$$

The proof of the lemma is given in Section 5.8.2.

Up to this point, we have obtained four critical values

$$0 < \beta_{s,1} < \beta_c < \beta_{s,2} < \beta_m < q \ ,$$

when $q \geq 5$. If $q = 4$, we have $\beta_{s,2} = q$, else if $q = 3$, $\beta_{s,2}$ is not defined. Thus, if $q \leq 4$, define $\beta_m = q$ so that

$$0 < \beta_{s,1} < \beta_c < \beta_m = q \ .$$

We conclude this section with the definition of the critical temperatures at which the phase transitions occur. We can now define critical temperatures $\beta_1, \beta_2, \beta_3$ appearing in Section 5.3.2. The critical temperatures are given by

$$\beta_1(q) := \beta_{s,1}(q), \quad \beta_2(q) := \beta_c(q), \quad \beta_3(q) := \beta_m(q) \ . \tag{5.31}$$

5.7 Critical points of F_β

In this section, we prove Proposition 5.6.3 for $q \geq 4$. For the case $q = 3$, we refer to [53] and we will only highlight the difference.

5.7.1 Eigenvalues of Hessian of F_β at critical points

First, we investigate $\mathbf{p} = (1/q, \ldots, 1/q)$, which is always a critical point for all $\beta > 0$. The following lemma proves the property of $\mathbf{p}$.

Lemma 5.7.1. *The point* $\mathbf{p}$ *is a local minimum of* F_β *if* $\beta < q$*, a local maximum of* F_β *if* $\beta > q$*, and a degenerate critical point when* $\beta = q$*.*

Proof. Let $\mathbb{1} = (1, \ldots, 1)^\dagger$ be a $(q-1) \times 1$ matrix. By elementary computation, we obtain

$$\nabla^2 F_\beta(\mathbf{p}) \ = \ \frac{q-\beta}{\beta}\Big(\mathrm{diag}(1, \ldots, 1) + \mathbb{1}\mathbb{1}^\dagger\Big)$$

whose eigenvalues are $(q-\beta)/\beta$ with multiplicity $q-2$ and $q(q-\beta)/\beta$ with 1. This completes the proof. $\square$

Now, for $i \in [1, q/2] \cap \mathbb{N}$, $j = q - i$, and $\beta = g_i(t)$, define $a \in \mathbb{R}$ and $b \in \mathbb{R}$ as

$$a = a(i,t) = -1 + 1/\beta t \ , \quad b = b(i,t) = -1 + i/\beta(1-jt) \ . \tag{5.32}$$

We have the following lemma about eigenvalues of Hessian of F_β at critical points.

Lemma 5.7.2. *Let $i \in [1, q/2] \cap \mathbb{N}$ and $j = q - i$. Moreover, let $t \in (0, 1/j)$ and $\beta = g_i(t)$. Then, $\mathbf{c} = (t, \dots, t, (1-jt)/i, \dots, (1-jt)/i)$ is a critical point of F_β and eigenvalues of $\nabla^2 F_\beta(\mathbf{c})$ constitute one of the following cases.*

1. *If $i \geq 2$, all eigenvalues of $\nabla^2 F_\beta(\boldsymbol{c})$ are a, b with multiplicative $j-1$, $i-2$, respectively, and the roots of $\lambda^2 - (a+qb)\lambda + b(ia+jb)$.*

2. *If $i = 1$, all eigenvalues of $\nabla^2 F_\beta(\boldsymbol{c})$ are a with multiplicative $j-1$ and $a + (q-1)b$ with multiplicative 1.*

Proof. By Lemma 5.6.2, $\mathbf{c}$ is a critical point of F_β since $\beta = g_i(t)$. By elementary computation, we have

$$\begin{aligned} \frac{\partial^2}{\partial x_k^2} F_\beta(\boldsymbol{x}) &= -1 + \frac{1}{\beta x_k} + \Big(-1 + \frac{1}{\beta x_q} \Big) \ , \\ \frac{\partial^2}{\partial x_k \partial x_l} F_\beta(\boldsymbol{x}) &= -1 + \frac{1}{\beta x_q} \ , \end{aligned}$$

so that

$$\frac{\partial^2}{\partial x_k \partial x_l} F_\beta(\mathbf{c}) = \begin{cases} -1 + \dfrac{1}{\beta t} + (-1 + \dfrac{i}{\beta(1-jt)}) & \text{if } 1 \leq k = l \leq j \\ 2(-1 + \dfrac{i}{\beta(1-jt)}) & \text{if } j+1 \leq k = l \leq q-1 \\ -1 + \dfrac{i}{\beta(1-jt)} & \text{if } k \neq l \end{cases}$$

Then, we can write $\nabla^2 F_\beta(\mathbf{c})$ as

$$\nabla^2 F_\beta(\mathbf{c}) \;=\; \mathbb{D} + b\mathbb{1}\mathbb{1}^\dagger \ ,$$

where

$$\mathbb{D} \;=\; \mathrm{diag}(\underbrace{a, \dots, a}_{j}, \underbrace{b, \dots, b}_{q-1-j}) \ .$$

Let $\mathbb{I} = \mathbb{I}_{q-1}$ be a $(q-1)$-identity matrix. By the formula

$$\det(A + \boldsymbol{v}\boldsymbol{w}^\dagger) = \det A(1 + \boldsymbol{v}^\dagger A^{-1}\boldsymbol{w}) \ ,$$

we can write

$$\begin{aligned}\det(\nabla^2 F_\beta(\mathbf{c}) - \lambda\mathbb{I}) =&\; \det(\mathbb{D} - \lambda\mathbb{I} + b\mathbb{1}\mathbb{1}^\dagger) \\ =&\; (a-\lambda)^j (b-\lambda)^{i-1}\Big[1 + b\big(\frac{j}{a-\lambda} + \frac{i-1}{b-\lambda}\big)\Big] \ .\end{aligned}$$

Hence, we obtain

$$\begin{aligned}&\det(\nabla^2 F_\beta(\mathbf{c}) - \lambda\mathbb{I}) \\ &= \begin{cases} (a-\lambda)^{j-1}(b-\lambda)^{i-2}(\lambda^2 - (a+qb)\lambda + b(ia+jb)) & \text{if } i \ge 2 \ , \\ (a-\lambda)^{j-1}(a+jb-\lambda) = (a-\lambda)^{q-2}(a+(q-1)b-\lambda) & \text{if } i = 1 \ . \end{cases}\end{aligned}$$

The proof of the lemma arises directly from this explicit computation of characteristic polynomial of Hessian of $F_\beta(\mathbf{c})$. □

We have the following lemma about the sign of the eigenvalues of $\nabla^2 F_\beta(\mathbf{c})$. Recall the definition of m_i from Lemma 5.6.1.

Lemma 5.7.3. *Let $i \in [1, q/2] \cap \mathbb{N}$ and $j = q - i$. Moreover, let $t \in (0, 1/j)$ and $\beta = g_i(t)$. Then, we have the following table regarding the sign of each value. If $i = q/2$, we ignore $t = m_i$ and $t \in (m_i, 1/q)$.*

	$t \in (0, m_i)$	$t = m_i$	$t \in (m_i, 1/q)$	$t = 1/q$	$t \in (1/q, 1/j)$
a	$+$	$+$	$+$	0	$-$
b	$-$	$-$	$-$	0	$+$
$ia + jb$	$+$	0	$-$	0	$+$
$b(ia + jb)$	$-$	0	$+$	0	$+$

Proof. First, suppose that $t < 1/q$. Then,

$$a > 0 \iff \frac{1}{t} > \beta = \frac{i}{1-qt} \log\left(\frac{1-jt}{it}\right) \Leftrightarrow \frac{1-qt}{it} > \log\left(\frac{1-jt}{it}\right) .$$

By substituting $x = (1-jt)/(it)$, one can deduce that $a > 0$ is equivalent to $t \neq 1/q$ which implies $a > 0$. Moreover, by the same argument above, we have $b < 0$. In the same manner, if $t > 1/q$, we obtain $a < 0$ and $b > 0$.

Now, we investigate the sign of $ia + jb$. We write

$$ia + jb \;=\; -i + \frac{i}{\beta t} - j + \frac{ij}{\beta(1-jt)} \;=\; -q + \frac{i}{\beta t(1-jt)} .$$

By elementary computation, $ia+jb = 0$ if $t = 1/q$. Hence, $ia+jb > 0$ if and only if

$$\frac{i}{qt(1-jt)} \;>\; \beta \;=\; \frac{i}{1-qt} \log\left(\frac{1-jt}{it}\right) .$$

First, assume $t < 1/q$. Then, $ia + jb > 0$ if and only if

$$h_i(t) \;=\; \log\left(\frac{1-jt}{it}\right) + \frac{qt-1}{qt(1-jt)} \;<\; 0 .$$

By investigating the graph of h_i (cf. Figure 5.7), the above inequality holds if and only if $t < m_i$. Second, assume $t > 1/q$. Then, $ia + jb > 0$ if and only if $h_i(t) > 0$ if and only if $t > 1/q$. Hence, $ia + jb > 0$ if and only if $t < m_i$ or $t > 1/q$.

The case when $t = 1/q$ can be proven by the argument in the first paragraph of this proof. If $t = m_i$, then $ia + jb = 0$ since $h_i(m_i) = 0$. The above argument can prove the case when $i = q/2$ since $m_{q/2} = 1/q$. $\square$

Now, we study the critical points more deeply. We note that the Morse index of a critical point is the number of negative eigenvalues of the Hessian at that point.

5.7.2 Critical points of Morse index 0 or 1

When we consider critical points in $\mathcal{U}_i$ or $\mathcal{V}_i$, we assume that $\beta > \beta_{s,i}$ since when $\beta = \beta_{s,i}$, the elements of $\mathcal{U}_i = \mathcal{V}_i$ are degenerate. The case when $\beta = \beta_{s,i}$ is treated in Section 5.7.4.

By the Morse theory, critical points with more than 2 negative eigenvalues can be neither saddle points nor minima. Hence, the critical points with only positive eigenvalues or only one negative eigenvalue and $q-2$ positive eigenvalues are relevant to the landscape of F_β. We select these critical points in this subsection.

As in (5.32), for $i \in [1, q/2] \cap \mathbb{N}$, $j = q - i$, and $\beta > \beta_{s,i}$, when we consider $\mathbf{u}_i \in \mathcal{U}_i$, let

$$a = a(\mathbf{u}_i) := -1 + \frac{1}{\beta u_i} \ , \quad b = b(\mathbf{u}_i) := -1 + \frac{1}{\beta(1 - j u_i)} \ ,$$

and when we consider $\mathbf{v}_i \in \mathcal{V}_i$, let

$$a = a(\mathbf{v}_i) := -1 + \frac{1}{\beta v_i} \ , \quad b = b(\mathbf{v}_i) := -1 + \frac{1}{\beta(1 - j v_i)} \ .$$

Lemma 5.7.4. *Let $q \geq 4$. If $\beta > \beta_{s,1}$, $\mathcal{U}_1$ is a set of local minima. If $\beta > \beta_{s,2}$, $\mathcal{U}_2$ is a set of saddle points. If $\beta_{s,1} < \beta < q$, $\mathcal{V}_1$ is a set of saddle points else if $\beta > q$, each point in $\mathcal{V}_1$ has at least two negative eigenvalues.*

Proof. Consider $\mathbf{u}_1 \in \mathcal{U}_1$. Eigenvalues of $\nabla^2 F_\beta(\mathbf{u}_1)$ are a with multiplicative $q-2$ and $a + (q-1)b$ with multiplicative 1. By Lemma 5.7.3, if $\beta > \beta_{s,1}$, then since $u_1 < m_1 < 1/q$, we obtain a, $a + (q-1)b > 0$; hence, $\mathbf{u}_1$ is a local minimum.

Next, consider $\mathbf{v}_1 \in \mathcal{V}_1$. Eigenvalues of $\nabla^2 F_\beta(\mathbf{v}_1)$ are a with multiplicative $q-2$ and $a+(q-1)b$ with multiplicative 1. By Lemma 5.7.3, if $\beta_{s,1} < \beta < q$, then since $m_1 < v_1 < 1/q$, we obtain $a > 0$ and $a + (q-1)b < 0$; hence, it is a saddle point. If $\beta > q$, then since $v_1 > 1/q$, we obtain $a < 0$ and $a + (q-1)b > 0$ so that $\mathbf{v}_1$ has more than two negative eigenvalues.

Finally, let $i \geq 2$, $j = q - i$, and $\beta > \beta_{s,i}$. In this case, $\mathbf{u}_i$ has eigenvalues a, b with multiplicative $j-1$, $i-2$ and the roots of $\lambda^2 - (a+qb)\lambda + b(ia+jb)$. Since $u_i < m_i \leq 1/q$ for all i and $\beta > \beta_{s,i}$, by Lemma 5.7.3, $a > 0$, $b < 0$, and $b(ia + jb) < 0$ so that it has j positive eigenvalues and $i - 1$ negative eigenvalues. Hence, $\mathbf{u}_2$ is a saddle point. □

Remark 5.7.5. *For $q = 3$, by the same argument, $\nabla^2 F_\beta(\mathbf{v}_1)$ has only one negative eigenvalue and two positive eigenvalues for $\beta \in (\beta_{s,1}, \infty) \setminus \{q\}$.*

5.7.3 Critical points of Morse index larger than 1

In this subsection, we eliminate unneeded critical points.

Lemma 5.7.6. *Let $q \geq 5$. For $i \in [3, q/2] \cap \mathbb{N}$ and $\beta > \beta_{s,i}$, each point in $\mathcal{U}_i$ has at least two negative eigenvalues. And for $i \in [2, q/2] \cap \mathbb{N}$ and $\beta \in (\beta_{s,i}, \infty) \setminus \{q\}$, each point in $\mathcal{V}_i$ has at least two negative eigenvalues.*

Proof. By the proof of Lemma 5.7.4, $\mathbf{u}_i$ for $i \geq 3$ has at least two negative eigenvalues. Now, let $i \geq 2$, $j = q - i$, and $\beta \in (\beta_{s,i}, \infty) \setminus \{q\}$. In this case, each points in $\mathcal{V}_i$ has eigenvalues a, b with multiplicative $j-1$, $i-2$, and the roots of $\lambda^2 - (a + qb)\lambda + b(ia + jb)$. If $\beta_{s,i} < \beta < q$, then $v_i < 1/q$ so that $a > 0$, $b < 0$, and $b(ia + jb) > 0$. In this case,

$$a + qb \;=\; ia + jb + (1-i)a + (q-j)b \;<\; ia + jb \;<\; 0\;,$$

so that the two roots of $\lambda^2 - (a+qb)\lambda + b(ia+jb)$ are negative. Hence, it has

$j-1$ positive eigenvalues and i negative eigenvalues. If $\beta > q$, then $v_i > 1/q$ so that $a < 0$, and points in $\mathcal{V}_i$ have at least $j-1$ negative eigenvalues, where $j-1 \geq 2$ since $q \geq 5$. $\square$

Lemma 5.7.7. *Let $q = 4$ and $\beta \geq q$. Then, we have $\mathcal{V}_2 = \mathcal{U}_2$.*

Proof. Observe that $\beta_{s,2} = q$. If $\beta = q$, $\mathcal{V}_2 = \mathcal{U}_2$ since there is only one solution m_2 to $q = g_2(t)$. Suppose $\beta > q$. By elementary computation, we obtain

$$g_2\Big(\frac{1}{4} - t\Big) = g_2\Big(\frac{1}{4} + t\Big) \quad \text{for } t \in \Big[0, \frac{1}{4}\Big) \ ,$$

so that $v_2 = (1/2) - u_2$. Hence, $\mathbf{v}_2 = (u_2, u_2, v_2, v_2)$ is a permutation of $\mathbf{u}_2$, that is, each element of $\mathcal{V}_2$ is one of the elements of $\mathcal{U}_2$ so that $\mathcal{V}_2 = \mathcal{U}_2$. $\square$

By lemmas in this subsection, $\mathcal{U}_i$, $i \geq 3$, and $\mathcal{V}_i$, $i \geq 2$, are not of interest.

5.7.4 At critical temperature

In this subsection, we investigate the critical points at the critical temperatures, that is, at $\beta = \beta_{s,i}$ or $\beta = q$. The point $\mathbf{u}_i = \mathbf{v}_i$ is degenerate when $\beta = \beta_{s,i}$ and the point $\mathbf{p} = \mathbf{v}_i$ is degenerate when $\beta = q$ by Lemma 5.7.2 and 5.7.3.

Lemma 5.7.8. *If $i \leq q/2$ and $\beta = \beta_{s,i}$, the point $\mathbf{u}_i = \mathbf{v}_i$ is not a local minimum. If $\beta = q$, the point $\mathbf{p} = \mathbf{v}_i$ is not a local minimum.*

Proof. Fix $1 \leq i \leq j \leq q-1$ such that $i + j = q$ and define $\boldsymbol{\ell}_i : [0, 1/j] \to \Xi$ as

$$\boldsymbol{\ell}_i(s) \;=\; \Big(s, \dots, s, \frac{1-js}{i}, \dots, \frac{1-js}{i}\Big) \ .$$

We therefore obtain

$$\begin{aligned} F_\beta(\boldsymbol{\ell}_i(s)) \;&=\; -\frac{1}{2}\Big[js^2 + i\Big(\frac{1-js}{i}\Big)\Big] + \frac{1}{\beta}\Big[js\log s + (1-js)\log\Big(\frac{1-js}{i}\Big)\Big] \\ &=\; -\frac{1}{2i}(jqs^2 - 2js + 1) + \frac{1}{\beta}\Big[\frac{(1-js)(1-qs)}{i} g_i(s) + \log s\Big] \ . \end{aligned}$$

By (5.25) and (5.26), we have

$$\frac{d}{ds}F_\beta(\boldsymbol{\ell}_i(s)) \;=\; \frac{j}{i}(1-qs)+\frac{j}{\beta i}(qs-1)g_i(s) \;=\; \frac{j}{\beta i}(1-qs)(\beta-g_i(s))\ .$$

We claim that $F_{\beta_{s,\,i}}(\boldsymbol{\ell}_i(m_i))$ and $F_q(\boldsymbol{\ell}_i(1/q))$ are not the local minima of $F_{\beta_{s,\,i}}(\boldsymbol{\ell}_i(s))$ and $F_q(\boldsymbol{\ell}_i(s))$, respectively, and this completes the proof.

For the first claim, assume $i<j$, and note that $m_i<1/q$. Then, $1-qs>0$ and $\beta_{s,\,i}-g_i(s)<0$ if s is in a neighborhood of m_i and $s\neq m_i$. In this case, $\frac{d}{ds}F_{\beta_{s,\,i}}(\boldsymbol{\ell}_i(s))<0$ near m_i so that $\mathbf{u}_i=\mathbf{v}_i$ is not a local minimum. If $i=j$, $\beta_{s,\,i}=q$ so that it suffices to show the second assertion.

Next, note that $v_i(q)=1/q$ so that we have $g_i(s)<\beta=q$, $1-qs>0$ if $s<1/q$ and $g_i(s)>q$, $1-qs<0$ if $s>1/q$. Therefore, $\frac{d}{ds}F_q(\boldsymbol{\ell}_i(s))>0$ near $1/q$ so that $\mathbf{p}=\mathbf{v}_i$ is not a local minimum. □

Even though $\mathbf{u}_i$, $i\geq 3$, is not a saddle point if $\beta>\beta_{s,\,i}$, we cannot exclude the possibility that $\mathbf{u}_i$ is a saddle point when $\beta=\beta_{s,\,i}$; however, by the next two lemmas, $\mathcal{U}_i(\beta_{s,\,i})$, $i\geq 3$, are irrelevant to the landscape of F_β.

Lemma 5.7.9. *Let $q\geq 8$ and $i\geq 4$. Then, if $\beta=\beta_{s,\,i}$, $\mathbf{u}_i=\mathbf{v}_i$ is not a saddle point.*

Proof. By Lemma 5.7.2, $-1+1\,/\,[\beta_{s,\,i}\{-ju_i(\beta_{s,\,i})\}]$ is an eigenvalue of $\nabla F_{\beta_{s,\,i}}$ at $\mathbf{u}_i$ with a multiple of at least two. Hence, by Lemma 5.7.3, it has at least two negative eigenvalues. □

Lemma 5.7.10. *Let $q\geq 6$. We have $F_{\beta_{s,\,3}}(\mathbf{u}_3)>F_{\beta_{s,\,3}}(\mathbf{u}_2)$. Furthermore, if $q\geq 7$, we have $F_{\beta_{s,\,3}}(\mathbf{u}_3)>F_{\beta_{s,\,3}}(\mathbf{v}_1)$. Hence, $\mathbf{u}_3$ cannot be a saddle point lower than $\mathbf{u}_2$ or $\mathbf{v}_1$.*

The proof is presented in Section 5.8.3. We remark that if $q=6$, we have $\beta_{s,\,3}=q$ so that $\mathbf{v}_1(\beta_{s,\,3})=\mathbf{p}$ and the second assertion is not needed.

5.8 Analysis of energy landscape

In this section, we prove lemmas introduced in Section 5.6.2 and Lemma 5.7.10. To prove these lemmas, we need numerical computation given in Appendix 5.11.

5.8.1 Proof of Lemma 5.6.5

Lemma 5.8.1. *If $q \geq 4$, we have $v_1(\beta_{s,2}) > \dfrac{1}{2(q-1)}$.*

Proof. Fix $\beta = \beta_{s,2}$ and write $v_1 = v_1(\beta_{s,2})$ for convenience. Since $\beta_{s,2} = g_2(m_2) = g_1(v_1)$, we have

$$\frac{2}{1-qm_2}\log\frac{1-(q-2)m_2}{2m_2} = \frac{1}{1-qv_1}\log\frac{1-(q-1)v_1}{v_1} \tag{5.33}$$

Let

$$v_1^* = \frac{1}{2q}+\frac{m_2}{2}\ ,\ \text{so that}\ \frac{1}{1-qv_1^*} = \frac{2}{1-qm_2}\ . \tag{5.34}$$

We claim that $g_1(v_1^*) \leq g_1(v_1)$, that is, by (5.33),

$$\frac{1}{1-qv_1^*}\log\frac{1-(q-1)v_1^*}{v_1^*} \leq \frac{2}{1-qm_2}\log\frac{1-(q-2)m_2}{2m_2}\ .$$

By (5.34), the above inequality is equivalent to

$$\frac{1-(q-1)v_1^*}{v_1^*} \leq \frac{1-(q-2)m_2}{2m_2}\ .$$

By plugging v_1^* given in (5.34) into this inequality, it becomes $q^2m_2-2qm_2+1 \geq 0$. Hence, since g_1 is increasing at v_1, we obtain $v_1^* \leq v_1$.

Finally, we claim that

$$v_1^* = \frac{1+qm_2}{2q} > \frac{1}{2(q-1)}\ ,\ \text{i.e.,}\ m_2 > \frac{1}{q(q-1)}\ .$$

According to Figure 5.7, we can show this by

$$h_2\Big(\frac{1}{q(q-1)}\Big) = \log\frac{q^2-2q+2}{2} - \frac{q(q-1)(q-2)}{q^2-2q+2} < 0 \ .$$

This holds if $q = 4$ or $q = 5$ by elementary computation. Now, assume $q \geq 6$. Therefore, we obtain

$$\log\frac{q^2-2q+2}{2} < \log q^2 = 2\log q < q-2 < \frac{q(q-1)(q-2)}{q^2-2q+2} \ ,$$

which completes the proof. □

We can prove Lemma 5.6.5 by the aforementioned lemma.

Proof of Lemma 5.6.5. Since $\beta_c = g_1(\frac{1}{q(q-1)}) = g_1(\frac{1}{2(q-1)})$, we have $\beta_{s,1} < \beta_c$. By Lemma 5.8.1, since $g_1(t)$ is increasing on $(m_1, 1/(q-1)\,)$ and $m_1 < 1/(2q-2)$, we obtain

$$\beta_{s,2} = g_1(v_1) > g_1(\frac{1}{2(q-1)}) = \beta_c \ .$$

□

5.8.2 Proof of Lemma 5.6.7

We first introduce two lemmas.

Lemma 5.8.2. *Let $q \geq 5$. When $\beta = \beta_{s,2}$, we have $F_{\beta_{s,2}}(\mathbf{v}_1) < F_{\beta_{s,2}}(\mathbf{u}_2)$ and when $\beta = q$, we have $F_q(\mathbf{v}_1) = F_q(\mathbf{p}) > F_q(\mathbf{u}_2)$.*

The proof of the above lemma is given in Section 5.8.3.

Lemma 5.8.3. *Let $q \geq 5$. $\beta^2\frac{d}{d\beta}[F_\beta(\mathbf{u}_2) - F_\beta(\mathbf{v}_1)]$ decreases as β increases in $(\beta_{s,2}, q)$.*

Proof. For $t = t(\beta)$, which satisfies $\beta = g_i(t)$, let

$$\mathbf{c}_i \;=\; \mathbf{c}_i(\beta) \;=\; \Big(t, \ldots, t, \frac{1-jt}{it}, \ldots, \frac{1-jt}{it}\Big) \;. \tag{5.35}$$

Since $\mathbf{c}_i$ is a critical point, by the proof of Corollary 5.3.7, we have

$$\frac{d}{d\beta} F_\beta(\mathbf{c}_i) = -\frac{1}{\beta^2} S(\mathbf{c}_i) \;.$$

Define a function $k_i : (0, 1) \to \mathbb{R}$ as

$$k_i(t) := (1 - jt) \log \frac{1-jt}{it} + \log t \;. \tag{5.36}$$

By elementary computations, we obtain $S(\mathbf{c}_i) = k_i(t)$ so that we have

$$\frac{d}{d\beta} F_\beta(\mathbf{c}_i) \;=\; -\frac{1}{\beta^2} k_i(t) \;. \tag{5.37}$$

Now, by (5.37), we obtain

$$\beta^2 \frac{d}{d\beta}\Big[F_\beta(\mathbf{u}_2) - F_\beta(\mathbf{v}_1)\Big] \;=\; k_1(v_1(\beta)) - k_2(u_2(\beta)) \;. \tag{5.38}$$

Observe that the value $u_2(\beta)$ decreases and the value $v_1(\beta)$ increases as β increases. By elementary computation, for $t \in (0, 1/q)$, we obtain

$$k_i'(t) \;=\; -j \log \frac{1-jt}{it} + (1-jt)\Big(\frac{-j}{1-jt} - \frac{1}{t}\Big) + \frac{1}{t} \;=\; -j \log \frac{1-jt}{it} < 0 \;, \tag{5.39}$$

so that $k_i(t)$ decreasing on $(0, 1/q)$. Hence, (5.38) decreases as β increases in $(\beta_{s,2}, q)$. □

We can now prove Lemma 5.6.7.

Proof of Lemma 5.6.7. By Lemma 5.8.2, there is $\beta_0 \in (\beta_{s,2}, q)$, such that

$\frac{d}{d\beta}[F_\beta(\mathbf{u}_2) - F_\beta(\mathbf{v_1})] < 0$. Hence, by Lemma 5.8.3, we can deduce that there is only one critical value $\beta_m \in (\beta_{s,2}, q)$, such that

$$F_{\beta_m}(\mathbf{u}_2) = F_{\beta_m}(\mathbf{v}_1) \ . \tag{5.40}$$

□

5.8.3 Proofs of Lemmas 5.7.10 and 5.8.2

Before we go further, we conduct some computations. Recall the definition of m_2 from Lemma 5.6.1. Since $\beta_{s,i} = g_i(m_i) = \frac{i}{1-qm_i} \log \frac{1-jm_i}{im_i}$ and m_i is the minimum of g_i, we have

$$\begin{aligned} 0 = h_i(m_i) =& \log \frac{1-jm_i}{im_i} + \frac{qm_i - 1}{qm_i(1-jm_i)} \\ =& \frac{1-qm_i}{i}\beta_{s,i} - \frac{1-qm_i}{qm_i(1-jm_i)} \ , \end{aligned}$$

so that

$$qjm_i^2 - qm_i = qm_i(jm_i - 1) = -\frac{i}{\beta_{s,i}} \ . \tag{5.41}$$

For $\mathbf{c}_i$ defined in (5.35), since $S(\mathbf{c}_i) = k_i(t)$ and $\beta = g_i(t)$, we can write

$$F_\beta(\mathbf{c}_i) = \frac{1}{2i}\Big[qjt^2 - 2qt + 1\Big] + \frac{1}{\beta}\log t \ . \tag{5.42}$$

Hence, by (5.41) and $\beta_{s,i} = g_i(m_i)$, we have

$$\begin{aligned} F_{\beta_{s,i}}(\mathbf{u}_i) =& \frac{1-qm_i}{2i} + \frac{1}{\beta_{s,i}}\Big(\log m_i - \frac{1}{2}\Big) \qquad (5.43) \\ =& \frac{1}{2\beta_{s,i}} \log \frac{1-jm_i}{im_i} + \frac{1}{\beta_{s,i}} \log m_i - \frac{1}{2\beta_{s,i}} \ . \end{aligned}$$

By (5.41) again, we obtain

$$F_{\beta_{s,\,i}}(\mathbf{u}_i) = -\frac{1}{2\beta_{s,\,i}}\log(qe\beta_{s,\,i}) \ . \tag{5.44}$$

Now, we introduce two technical lemmas required in the proof of Lemmas 5.7.10 and 5.8.2.

Lemma 5.8.4. *For $q \ge 6500$, we have*

$$\frac{1}{\beta_{s,\,2}}\Big(\log qm_2 - \frac{1}{2}\Big) > \frac{(q-1)}{8q}(qm_2)^2 - \frac{1}{4}m_2 + \frac{-q^2+4q+1}{8q(q-1)} \ .$$

The proof is given in Section 5.10.

Lemma 5.8.5. *Let $q \ge 5$. Define $f_c(\beta) = -\dfrac{1}{2\beta}\log(qe\beta)$ and*

$$\Phi(\beta) = \frac{d}{d\beta}[f_c(\beta) - F_\beta(\mathbf{u}_2)] \ .$$

Then, we have $\Phi(\beta) > 0$ for $\beta > \beta_{s,\,2}$.

Proof. We have

$$\frac{d}{d\beta}f_c(\beta) = \ \frac{1}{2\beta^2}\log qe\beta - \frac{1}{2\beta}\frac{1}{\beta} \ = \ \frac{1}{2\beta^2}\log q\beta \ .$$

By (5.37), we obtain

$$\beta^2\frac{d}{d\beta}[f_c(\beta) - F_\beta(\mathbf{u}_2)] \ = \ \frac{1}{2}[\log q\beta + 2k_2(u_2)] \ .$$

By (5.39), the above expression is increasing function of β since u_2 decreases as β increases. Hence, it is sufficient to show $\Phi(\beta_{s,\,2}) > 0$. First, let $q \ge 55 >$

e^4. By (5.39)

$$\begin{aligned}\log q\beta + 2k_2(u_2) &> \log q\beta_{s,2} + 2k_2(\frac{1}{2j}) \\ &= \log q\beta_{s,2} + \log\frac{2j-j}{i} + 2\log\frac{1}{2j} = \log\frac{q\beta_{s,2}}{4ij} \ ,\end{aligned}$$

where we use $u_2 < 1/(2j)$ for the inequality. Since $\beta_{s,2} > \beta_c > 2\log q$, we obtain

$$\frac{q\beta_{s,2}}{4ij} > \frac{2q\log q}{8(q-2)} > \frac{q}{q-2} \ .$$

Finally, for $5 \le q \le 54$, by Proposition 5.11.1, we have $\Phi(\beta_{s,2}) > 0$. □

By the above lemmas, Lemma 5.8.2 can be proven.

Proof of Lemma 5.8.2. By Proposition 5.11.1 given in appendix, we can check that $F_{\beta_{s,2}}(\mathbf{u}_2) > F_{\beta_{s,2}}(\mathbf{v}_1)$ holds for $5 \le q \le 6500$. Now, suppose that $q > 6500$. By (5.42) and (5.43), we can write

$$\begin{aligned}F_{\beta_{s,2}}(\mathbf{u}_2) &= -\frac{1}{4}qm_2 + \frac{1}{4} + \frac{1}{\beta_{s,2}}\Big(\log m_2 - \frac{1}{2}\Big) \ , \\ F_{\beta_{s,2}}(\mathbf{v}_1) &= \frac{1}{2}\Big[q(q-1)\Big(v_1 - \frac{1}{q-1}\Big)^2 - \frac{1}{q-1}\Big] + \frac{1}{\beta_{s,2}}\log v_1 \ .\end{aligned}$$

By the proof of Lemma 5.8.1, we have

$$\frac{qm_2+1}{2q} = v_1^* \le v_1 < \frac{1}{q} \ ,$$

so that

$$F_{\beta_{s,2}}(\mathbf{v}_1) < \frac{1}{2}\left[q(q-1)\Big(\frac{qm_2+1}{2q} - \frac{1}{q-1}\Big)^2 - \frac{1}{q-1}\right] - \frac{1}{\beta_{s,2}}\log q \ .$$

Hence, the lemma can be proven if we can prove

$$
\begin{aligned}
&-\frac{1}{4}qm_2+\frac{1}{4}+\frac{1}{\beta_{s,2}}(\log m_2-\frac{1}{2})\\
&>\frac{1}{2}\Big[q(q-1)\Big(\frac{qm_2+1}{2q}-\frac{1}{q-1}\Big)^2-\frac{1}{q-1}\Big]-\frac{1}{\beta_{s,2}}\log q\\
&=\frac{1}{8}q(q-1)(m_2)^2-\frac{1}{4}(q+1)m_2+\frac{(q+1)^2}{8q(q-1)}-\frac{1}{\beta_{s,2}}\log q\ .
\end{aligned}
$$

This is the content of Lemma 5.8.4. Finally, by Lemma 5.8.5, we obtain $F_q(\mathbf{p})-F_q(\mathbf{u}_2)=f_c(q)-F_q(\mathbf{u}_2)>0$ since $f_c(\beta_{s,2})=F_{\beta_{x,2}}(\mathbf{u}_2)$. □

Now, we prove Lemma 5.7.10.

Proof of Lemma 5.7.10. Since the proof for $F_{\beta_{s,3}}(\mathbf{u}_3)>F_{\beta_{s,3}}(\mathbf{v}_1)$ is exactly the same as the proof of Lemma 5.8.2 including numerical verification, we omit it. By (5.44), we can write

$$F_{\beta_{s,3}}(\mathbf{u}_3)\ =\ f_c(\beta_{s,3})\ .$$

Hence, by Lemma 5.8.5 and by Proposition 5.11.1, we have

$$F_{\beta_{s,3}}(\mathbf{u}_3)\ =\ f_c(\beta_{s,3})\ >\ F_{\beta_{s,3}}(\mathbf{u}_2)\ .$$

□

5.9 Characterization of metastable sets

In this section, we prove Theorems 5.3.4-5.3.6. First, we prove Theorem 5.3.4.

Proof of Theorem 5.3.4. The first assertion is immediate from Lemmas 5.6.2 and 5.6.4. The third assertion is proven by Proposition 5.6.3 and Lemma 5.7.8. The fourth assertion is Lemma 5.6.6.

Now, it remains to show the second assertion. For $\beta \in (\beta_1, \beta_2]$, since $\mathbf{p}$ is the global minimum and $\mathbf{v}_1$ is a saddle point, we have $F_\beta(\mathbf{p}) < F_\beta(\mathbf{v}_1)$ so that $\mathcal{W}_{\mathfrak{o}} \neq \emptyset$. By the same argument in the proof of Lemma 5.8.3, we have

$$\frac{d}{d\beta}[F_\beta(\mathbf{v}_1) - F_\beta(\mathbf{p})] = -\frac{1}{\beta^2}[k_1(v_1(\beta)) + \log q] \ .$$

By 5.39, $k_1(\cdot)$ is decreasing on $(0,\, 1/q)$ and increasing on $(1/q,\, 1/(q-1)\,)$. Since $k_1(1/q) = -\log q$, we have $k_1(v_1(\beta)) + \log q > 0$ for $\beta \in (\beta_1, q)$ so that

$$\frac{d}{d\beta}[F_\beta(\mathbf{v}_1) - F_\beta(\mathbf{p})] < 0 \ .$$

Since $\mathbf{v}_1 = \mathbf{p}$ when $\beta = q$, we have $F_\beta(\mathbf{v}_1) > F_\beta(\mathbf{p})$ for $\beta < q$ and $F_\beta(\mathbf{v}_1) < F_\beta(\mathbf{p})$ for $\beta > q$. □

5.9.1 Proof of Theorem 5.3.6

Before we go further, we recall the height between two points. Let $\boldsymbol{a}$, $\boldsymbol{b} \in \operatorname{int}\Xi$, and let $\Gamma_{\boldsymbol{a},\boldsymbol{b}}$ be a set of all C^1-path $\gamma : [0,1] \to \operatorname{int}\Xi$, such that $\gamma(0) = \boldsymbol{a}$ and $\gamma(1) = \boldsymbol{b}$. Then, we can define the height $\mathfrak{H}(\boldsymbol{a},\boldsymbol{b})$ between $\boldsymbol{a}$ and $\boldsymbol{b}$ as $\mathfrak{H}(\boldsymbol{a},\boldsymbol{b}) = \inf_{\gamma\in\Gamma_{\boldsymbol{a},\boldsymbol{b}}} \sup_{0\le t\le 1} F_\beta(\gamma(t))$. We prove Theorem 5.3.6 in several steps.

Lemma 5.9.1. *Let $q \geq 4$. If $\beta > \beta_m$, the sets $\mathcal{W}_i(\beta)$, $i \in S$, are different. In particular, they are nonempty.*

Proof. Since the elements of $\mathcal{U}_1$ are the lowest minima, we have $F_\beta(\mathbf{u}_1) < H_\beta$ so that $\mathcal{W}_i$'s are nonempty. Without loss of generality, suppose $\mathcal{W}_1 = \mathcal{W}_2$. Since $\mathbf{u}_1^1$, $\mathbf{u}_1^2 \in \mathcal{W}_1$ and $\mathcal{W}_1$ is connected, there is a C^1-path $\gamma : [0,1] \to \mathcal{W}_1$, such that $\gamma(0) = \mathbf{u}_1^1$, $\gamma(1) = \mathbf{u}_1^2$. Therefore, we have $F_\beta(\gamma(t)) < H_\beta$ for $0 \leq t \leq 1$, so that

$$F_\beta(\mathbf{u}_1^1) < \mathfrak{H}(\mathbf{u}_1^1, \mathbf{u}_1^2) < H_\beta \ .$$

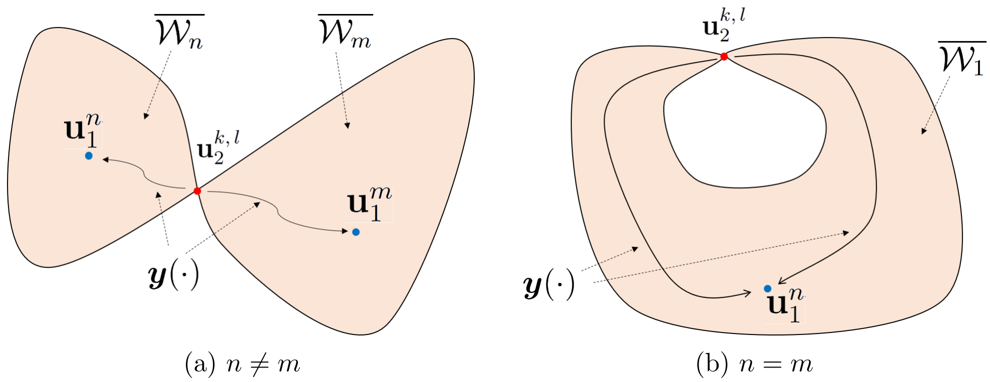


Figure 5.8: Paths from $\mathbf{u}_2^{k,l}$ to $\mathbf{u}_1^1$ and $\mathbf{u}_1^m$

Then, there is a saddle point $\boldsymbol{\sigma}(\mathbf{u}_1^1, \mathbf{u}_1^2)$, such that $F_\beta(\boldsymbol{\sigma}(\mathbf{u}_1^1, \mathbf{u}_1^2)) = \mathfrak{H}(\mathbf{u}_1^1, \mathbf{u}_1^2)$. However, by Proposition 5.6.3, the values of saddle points are greater than or equal to H_β. This is contradiction. Hence, $\mathcal{W}_i$'s are different. □

Lemma 5.9.2. *Let $q \geq 4$. If $\beta > q$, the set $\Sigma_{i,j}$ is singleton for all $i, j \in S$.*

Proof. First, we claim that $\Sigma_{i,j}$'s are not empty. Suppose one of $\Sigma_{i,j}$'s is empty. Then, by symmetry, all of them are empty. We will derive a contradiction from this.

Let us fix $1 \leq k < l \leq q$. Since $\mathbf{u}_2^{k,l}$ is a saddle point, there is a unit eigenvector $\boldsymbol{w}$ that corresponds to the unique negative eigenvalue of $\nabla^2 F_\beta(\mathbf{u}_2^{k,l})$. There exists $\eta > 0$, such that $F_\beta(\mathbf{u}_2^{k,l} + t\boldsymbol{w}) < H_\beta$ for all $0 < |t| < \eta$. Now, consider the path $\boldsymbol{y}(t)$ described by the ordinary differential equation

$$\dot{\boldsymbol{y}}(t) \;=\; -\nabla F_\beta(\boldsymbol{y}(t)), \quad \boldsymbol{y}(0) \;=\; \mathbf{u}_2^{k,l} + \eta\boldsymbol{w} \;. \tag{5.45}$$

Then, $\boldsymbol{y}(t)$ converges to a critical point whose height is less than H_β as $t \to \infty$. If this convergent point is not a local minimum, we can find an eigenvector $\boldsymbol{w}_1$ corresponding to a negative eigenvalue of the Hessian of F_β at that point. Then, by the same argument defining the path (5.45), the next path converges to another critical point whose height is lower than that of the previous critical point. Finally, this path converges to a local minimum. Since there is no local minimum other than $\mathcal{U}_1$, $\boldsymbol{y}(t)$ converges to some elements of $\mathcal{U}_1$, say $\mathbf{u}_1^n$. Since $\mathcal{W}_i$'s are different, $\boldsymbol{y}(\cdot)$ converges to only one minimum. By the same argument, the similar path starting at $\mathbf{u}_2^{k,\,l} - \eta \boldsymbol{w}$ converges to some $\mathbf{u}_1$, say $\mathbf{u}_1^m$. If $n \neq m$, $\mathbf{u}_2^{k,\,l} \in \Sigma_{n,\,m}$ so that $\Sigma_{n,\,m}$ is not empty. So, we have $m = n$. In this case, we obtain $\mathbf{u}_2^{k,\,l} \in \overline{\mathcal{W}_1}$ and $\mathbf{u}_2^{k,\,l} \notin \overline{\mathcal{W}_a}$ for all $a \neq n$. See Figure 5.8 for the visualization these paths.

By symmetry, since $\mathcal{U}_2$ has $q(q-1)/2$ elements and the number of $\mathcal{W}_i$ is q, there are $(q-1)/2$ elements in $\mathcal{U}_2$ corresponding to each $\mathcal{W}_i$, that is, $|\overline{\mathcal{W}_1} \cap \mathcal{U}_2| = (q-1)/2$, where $|A|$ is the number of elements of set A. If $\mathbf{u}_2^{1,\,a} \in \overline{\mathcal{W}_1}$, for some $2 \leq a \leq q$, we obtain $\mathbf{u}_2^{1,\,a} \in \overline{\mathcal{W}_a}$ by symmetry, and therefore $\Sigma_{1,\,a} = \overline{\mathcal{W}_1} \cap \overline{\mathcal{W}_a} \neq \emptyset$. Hence, we have $\mathbf{u}_2^{1,\,a} \notin \overline{\mathcal{W}_1}$. If $\mathbf{u}_2^{a,\,b} \in \overline{\mathcal{W}_1}$ for some $1 < a,\, b$, since $q \geq 4$ and by symmetry, $\mathbf{u}_2^{a,\,b} \in \overline{\mathcal{W}_m}$ for some $m \neq 2,\, a,\, b$, and this contradicts the assumption that $\Sigma_{1,\,m} = \overline{\mathcal{W}_1} \cap \overline{\mathcal{W}_m} = \emptyset$. Hence, all of $\Sigma_{i,\,j}$'s are nonempty.

Observe that the elements of $\Sigma_{i,\,j}$ are saddle points and $F_\beta(\boldsymbol{x}) = H_\beta$ for all $\boldsymbol{x} \in \Sigma_{i,\,j}$. Hence, by Proposition 5.6.3, $\Sigma_{i,\,j} \subset \mathcal{U}_2$. Since $\nabla^2 F_\beta(\mathbf{u}_2)$'s are nondegenerate and have only one negative eigenvalue, each element of $\mathcal{U}_2$ connects only two wells, i.e., $\Sigma_{i,\,j} \cap \Sigma_{k,\,l} = \emptyset$ if $\{i, j\} \neq \{k, l\}$. Therefore, since $\mathcal{U}_2$ has $q(q-1)/2$ elements, $\Sigma_{i,\,j}$ has at most one point so that we obtain $|\Sigma_{i,\,j}| = 1$. □

We can now prove Theorem 5.3.6.

Proof of Theorem 5.3.6. The first assertion follows from the definition of critical temperatures (5.31) and Lemma 5.6.7.

Let $\beta > q$. By Lemma 5.9.2, to prove $\Sigma_{i,\,j} = \{\mathbf{u}_2^{i,\,j}\}$, without loss of generality, it is sufficient to show that $\Sigma_{1,\,2} \neq \{\mathbf{u}_2^{1,\,4}\}$ and $\Sigma_{1,\,2} \neq \{\mathbf{u}_2^{3,\,4}\}$. First, suppose $\Sigma_{1,\,2} = \{\mathbf{u}_2^{1,\,4}\}$. Then, by symmetry, we obtain $\mathbf{u}_2^{1,\,4} \in \Sigma_{1,\,3}$, which contradicts to $\Sigma_{1,\,2} \cap \Sigma_{1,\,3} = \emptyset$. Second, suppose $\Sigma_{1,\,2} = \{\mathbf{u}_2^{3,\,4}\}$ so that by symmetry, we have $\Sigma_{1,\,5} = \{\mathbf{u}_2^{3,\,4}\}$ which is also contradiction. Hence, we obtain $\Sigma_{1,\,2} = \{\mathbf{u}_2^{1,\,2}\}$.

Since F_β is continuous in β, the values H_β and $\mathfrak{H}(\mathbf{u}_1^i(\beta), \mathbf{u}_1^j(\beta))$, $i, j \in S$, are continuous in β. Note that $\mathfrak{H}(\mathbf{u}_1^i(\beta), \mathbf{u}_1^j(\beta)) = F_\beta(\mathbf{u}_2) = H_\beta$ for $\beta \geq q$ since there is no saddle point other than $\mathcal{U}_2$. Since $F_\beta(\mathbf{v}_1) > H_\beta$ if $\beta > \beta_m = \beta_3$ and there is no saddle point other than the elements of $\mathcal{U}_2 \cup \mathcal{V}_1$, by continuity, we obtain

$$\mathfrak{H}(\mathbf{u}_1^i(\beta), \mathbf{u}_1^j(\beta)) = H_\beta \text{ if } \beta \geq \beta_3 .$$

Hence, $\mathbf{u}_2^{i,j}$ is a saddle point between $\mathbf{u}_1^i$ and $\mathbf{u}_1^j$ and $\Sigma_{i,j} = \{\mathbf{u}_2^{i,j}\}$ if $\beta > \beta_3$. Coupled with Lemma 5.9.1, the fourth assertion holds except that $\Sigma_{\mathfrak{o},i} = \emptyset$.

If $\beta \geq q$, $\mathcal{W}_{\mathfrak{o}} = \emptyset$. Let $\beta_3 \leq \beta < q$. Without loss of generality, suppose that $\Sigma_{\mathfrak{o},1} = \overline{\mathcal{W}_{\mathfrak{o}}} \cap \overline{\mathcal{W}_1} \neq \emptyset$. Let $\boldsymbol{a} \in \overline{\mathcal{W}_{\mathfrak{o}}} \cap \overline{\mathcal{W}_1}$. Note that $\boldsymbol{a} \in \mathcal{W}_{\mathfrak{o}}$ since $F_\beta(\boldsymbol{a}) \leq H_\beta < F_\beta(\mathbf{v}_1)$. Since $\boldsymbol{a} \in \overline{\mathcal{W}_1}$, $\boldsymbol{a}$ is connected to $\mathbf{u}_1^1$ in $\{\boldsymbol{x} : F_\beta(\boldsymbol{x}) \leq H_\beta\}$. In addition, since $H_\beta < F_\beta(\mathbf{v}_1)$ and $\boldsymbol{a} \in \mathcal{W}_{\mathfrak{o}}$, $\mathcal{W}_{\mathfrak{o}}$ must contain $\mathbf{u}_1^1$. We, therefore, obtain $\mathfrak{H}(\mathbf{p}, \mathbf{u}_1^1(\beta)) < F_\beta(\mathbf{v}_1)$ so that $\mathfrak{H}(\mathbf{p}, \mathbf{u}_1^1(\beta)) = H_\beta$. By continuity, we get

$$\mathfrak{H}(\mathbf{p}, \mathbf{u}_1^1(\beta)) = H_\beta \text{ for } \beta_3 \leq \beta < q ,$$

so that $F_\beta(\mathbf{p}) \leq H_\beta$. However, it is in contradiction to $F_q(\mathbf{p}) = F_q(\mathbf{v}_1) > H_q$. Hence, we obtain $\Sigma_{\mathfrak{o},i} = \emptyset$ for $i \in S$.

By the same argument and symmetry, the second assertion can be proven for $\beta \in (\beta_{s,1}, \beta_{s,2}) = (\beta_1, \beta_{s,2})$. By continuity argument, we can extend these to $\beta \in (\beta_1, \beta_3)$. The third assertion holds because of the first and fourth assertions, symmetry, and continuity. Finally, the fifth assertion can be proven by the same argument. □

5.9.2 Proof of Theorem 5.3.5

If $q = 4$, $\Sigma_{1,2} \neq \{\mathbf{u}_2^{3,4}\}$ cannot be proven by symmetry argument. Hence, we directly prove the Theorem 5.3.5.

Proof of Theorem 5.3.5. By Lemma 5.6.7 and (5.31), we obtain the first assertion.

Consider $\mathcal{K}_{i,j} = \{\, \boldsymbol{x} \in \Xi \,:\, x_i = x_j = \max\{x_1, \ldots, x_4\}\,\}$. It can be observed that these six planes divide Ξ into four pieces, and each plain contains one element of $\mathcal{U}_2$ and $\mathbf{u}_2^{i,j} \in \mathcal{K}_{i,j}$. We claim that $H_\beta = F_\beta(\mathbf{u}_2^{i,j}) < F_\beta(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{K}_{i,j}$ if $\beta > q$. Note that $\mathbf{p}$ is not local minimum if $\beta \geq q$.

Let $\widetilde{F}_\beta(\boldsymbol{x})$ be a restriction of F_β to $\mathcal{K}_{3,4}$ and let $\mathcal{K}_{3,4}^o = \{\boldsymbol{x} \in \mathcal{K}_{3,4} \,:\, x_3 = x_4 > x_1, x_2\}$. Since $x_3 = x_4 = \dfrac{1}{2}(1 - x_1 - x_2)$,

$$\frac{\partial}{\partial x_i}\widetilde{F}_\beta(\boldsymbol{x}) = -x_i + \frac{1}{\beta}\log x_i + x_3 - \frac{1}{\beta}\log x_3 \ ,$$

so that if $\boldsymbol{x} \in \mathcal{K}_{3,4}$ is a critical point, we have

$$-x_1 + \frac{1}{\beta}\log x_1 = -x_2 + \frac{1}{\beta}\log x_2 = -x_3 + \frac{1}{\beta}\log x_3 \ .$$

Since $x_3 = x_4 > x_1, x_2$, if $\beta \geq q$, the critical points in $\mathcal{K}_{3,4}^o$ are $\mathbf{u}_2^{3,4}$, $\mathbf{v}_2^{1,2}$. From the proof Lemma 5.7.7, we obtain $\mathbf{u}_2^{3,4} = \mathbf{v}_2^{1,2} = (u_2, u_2, v_2, v_2)$.

Let $a = -1 + \dfrac{1}{\beta u_2}$ and $b = -1 + \dfrac{1}{\beta v_2}$. We therefore obtain

$$\nabla^2 \widetilde{F}_\beta(\mathbf{u}_2^{3,4}) \ = \ \begin{pmatrix} a + \frac{1}{2}b & \frac{1}{2}b \\ \frac{1}{2}b & a + \frac{1}{2}b \end{pmatrix} \ .$$

The eigenvalues of $\nabla^2 \widetilde{F}_\beta(\mathbf{u}_2^{3,4})$ are a and $a + b$. By Lemma 5.7.3, a, $b > 0$ so that $\mathbf{u}_2^{3,4}$ is a local minimum in $\mathcal{K}_{3,4}^o$. Since this is the unique critical point, $\mathbf{u}_2^{3,4}$ is the unique minimum in $\mathcal{K}_{3,4}^o$. Since $\mathcal{K}_{3,4}$ is a closure of $\mathcal{K}_{3,4}^o$ and there is no critical point in $\mathcal{K}_{3,4}^o \setminus \{\mathbf{u}_2^{3,4}\}$, $\mathbf{u}_2^{3,4}$ is the unique minimum in $\mathcal{K}_{3,4}$. Hence, $\mathcal{W}_i$'s are different if $\beta > q$.

Let $\beta > q$. By the definition of $\mathcal{K}_{i,j}$, we obtain $\overline{\mathcal{W}_k} \cap \mathcal{K}_{i,j} = \emptyset$ if $k \neq i, j$ so that $\Sigma_{i,j} \subset \mathcal{K}_{i,j}$. By Lemma 5.9.2, $\Sigma_{i,j}$ are not empty. It can be observed $F_\beta(\boldsymbol{x}) = H_\beta$ and $\nabla F_\beta(\boldsymbol{x}) = 0$ if $\boldsymbol{x} \in \Sigma_{i,j}$. Since $\Sigma_{i,j} \subset \mathcal{K}_{i,j}$, we have

$\Sigma_{i,j} = \{\mathbf{u}_2^{i,j}\}$, thus the fourth assertion is proved.

For the third assertion, note that $F_q(\boldsymbol{x}) = H_q$ for all $\boldsymbol{x} \in \Sigma_{i,j}$ and $\mathbf{p}$ is the only point in $\mathcal{K}_{i,j}$, such that $F_q(\boldsymbol{x}) = H_q$. Moreover, we obtain $F_q(\boldsymbol{x}) > H_q = F_q(\mathbf{p})$ if $\boldsymbol{x} \in \mathcal{K}_{i,j}^o$, and finally we can deduce $F_q(\boldsymbol{x}) > H_q = F_q(\mathbf{p})$ if $\boldsymbol{x} \in \mathcal{K}_{i,j} \setminus \{\mathbf{p}\}$ using elementary calculus. Hence, $\mathcal{W}_i$'s are different if $\beta = q$.

For the second assertion, we can use the symmetry argument and the proofs are the same as the proof of Theorem 5.3.6. □

5.10 Proof of Lemma 5.8.4

This section is devoted to the proof of Lemma 5.8.4. In Section 5.10.1, we provide an auxiliary lemma to prove Lemma 5.10.1. In section 5.10.2, we prove this auxiliary lemma. So far, we have fixed an integer $q \geq 3$; however, in this section, we consider q as a real number and several variables as functions of q. For example, $m_2 = m_2(q)$, $j(q) = q - 2$, and $\beta_{s,2} = \beta_{s,2}(q)$.

5.10.1 Proof of Lemma 5.8.4

Lemma 5.10.1. *The function $f_\star$ of q is defined as*

$$f_\star(q) = \frac{1}{\beta_{s,2}}\Big(\log qm_2 - \frac{1}{2}\Big) - \frac{1}{8}(qm_2)^2 + \frac{1}{4}m_2 + \frac{251}{2002}\ . \tag{5.46}$$

Then, if $q > e^8$, $f_\star'(q) = \dfrac{d}{dq}f_\star(q) > 0$.

Proof of Lemma 5.8.4. By Proposition 5.11.1, we obtain $f_\star(6500) > 0$. We observe that $\dfrac{(q-1)}{8q} < \dfrac{1}{8}$ and $\dfrac{-q^2+4q+1}{8q(q-1)} < -\dfrac{251}{2002}$ if $q > 1000$. Hence, Lemma 5.10.1 proves Lemma 5.8.4. □

5.10.2 Proof of Lemma 5.10.1

Let $s_2 = s_2(q) = qm_2$. In the first lemma, we compute $m_2' = (d/dq)m_2$, $s_2' = (d/dq)s_2$, and $\beta_{s,2}' = (d/dq)\beta_{s,2}$.

Lemma 5.10.2. *We have*

$$m_2' = \frac{d}{dq}m_2 = -\frac{m_2(1-jm_2-qjm_2^2)}{q(1-2jm_2)} \ ,$$
$$s_2' = \frac{d}{dq}s_2 == -\frac{js_2^2(1-s_2)}{q(q-2js_2)} \ ,$$
$$\beta_{s,2}' = \frac{d}{dq}\beta_{s,2} = \frac{1}{1-s_2}\left(\beta_{s,2}s_2' - 2\frac{-s_2+s_2^2+qs_2'}{(q-js_2)s_2}\right) \ .$$

Proof. We observe that

$$\beta_{s,2} = g_2(m_2) = \frac{2}{1-qm_2}\log\frac{1-jm_2}{2m_2} = \frac{2}{qm_2(1-jm_2)} \ ,$$

so that

$$\log(1-jm_2) - \log 2m_2 = \frac{2}{q}\Big(\frac{1}{2m_2} - \frac{1}{1-jm_2}\Big) \ .$$

By differentiating this equation in q, we get

$$\frac{-m_2-jm_2'}{1-jm_2} - \frac{m_2'}{m_2} = -\frac{2}{q^2}\Big(\frac{1}{2m_2} - \frac{1}{1-jm_2}\Big) + \frac{2}{q}\Big(-\frac{m_2'}{2m_2^2} + \frac{-m_2-jm_2'}{(1-jm_2)^2}\Big) \ .$$

By elementary computation, we can write

$$m_2' = -\frac{m_2(1-jm_2-qjm_2^2)}{q(1-2jm_2)} \ . \tag{5.47}$$

Let $s_2 = qm_2$. Then,

$$s_2' = m_2 + qm_2' = -\frac{js_2^2(1-s_2)}{q(q-2js_2)} \ . \tag{5.48}$$

Next, we compute $\beta'_{s,2}$. Note that

$$\beta_{s,2} = \frac{2}{1-s_2}\log\frac{q-js_2}{2s_2} \ ,$$

so that

$$\begin{aligned}\beta'_{s,2} &= -\frac{2s'_2}{(1-s_2)^2}\log\frac{q-js_2}{2s_2} + \frac{2}{1-s_2}(\frac{1-s_2-js'_2}{q-js_2} - \frac{s'_2}{s_2}) \\ &= \frac{1}{1-s_2}\left(s'_2\frac{2}{1-s_2}\log\frac{q-js_2}{2s_2} + 2\frac{s_2-s_2^2-js_2s'_2-qs'_2+js_2s'_2}{(q-js_2)s_2}\right) \\ &= \frac{1}{1-s_2}\left(\beta_{s,2}s'_2 - 2\frac{-s_2+s_2^2+qs'_2}{(q-js_2)s_2}\right) \ . \end{aligned} \tag{5.49}$$

□

The next lemma provides the bound of $m_2(q)$.

Lemma 5.10.3. *Let $q > e^8$. We have*

$$\frac{1}{2q\log q} < m_2(q) < \frac{1}{q\log q} \ .$$

Proof. It can be observed that $h_2(m_2) = 0$ and $h_2(t) > 0$ if $m_2 < t < 1/q$. We claim that

$$h_2(a) \ = \ \log\frac{1-ja}{2a} + \frac{qa-1}{qa(1-ja)} \ > 0 \ ,$$

where $a = 1/q\log q$. The above inequality can be written as

$$\log\frac{q\log q - j}{2} > \left(\frac{q\log q - q}{q\log q - j}\right)\log q \ .$$

Since the right-hand side is smaller than $\log q$, it suffices to show that

$$\log q + \log\frac{\log q - 1 + 2/q}{2} > \log q \ ,$$

which is true if $q > e^3$. Hence, $m_2 < 1/q \log q$. Next, we have $m_2 > (1/2q)\log q$ since

$$\log\Big(q\log q - \frac{j}{2}\Big) - 2\Big\{\frac{q\log q - q/2}{(q\log q - j/2)}\Big\}\log q < 0\ ,$$

which is true if $q > e^8$. □

In the next two lemmas, we prove that some quantities are positive.

Lemma 5.10.4. *Let $q > e^8$. We have*

$$m_2' - s_2 s_2' > 0\ .$$

Proof. We have

$$\begin{aligned} m_2' - s_2 s_2' &= -\frac{m_2(1 - jm_2 - jqm_2^2)}{q(1-2jm_2)} + \frac{js_2^3(1-s_2)}{q(q-2js_2)} \\ &= \frac{s_2(-1 + jm_2 + jq(q+1)m_2^2 - jq^3m_2^3)}{q(q-2js_2)}\ . \end{aligned}$$

It suffices to show that

$$jq(q+1)m_2^2 - jq^3m_2^3 - 1 > 0\ .$$

Since $\dfrac{1}{2q\log q} < m_2 < \dfrac{1}{q\log q}$, we obtain

$$\begin{aligned} &jq(q+1)m_2^2 - jq^3m_2^3 - 1 \\ &\quad > \frac{jq(q+1)}{4q^2(\log q)^2} - \frac{jq^3}{q^3(\log q)^3} - 1 \\ &\quad = \frac{1}{q(\log q)^3}\Big[\frac{(q+1)(q-2)}{4}\log q - q(q-2) - q(\log q)^3\Big] \\ &\quad > \frac{1}{q(\log q)^3}[2(q+1)(q-2) - q(q-2) - q(\log q)^3] \\ &\quad = \frac{1}{q(\log q)^3}[q^2 - q(\log q)^3 - 4] > 0\ . \end{aligned}$$

In the second and third inequalities, we use $q > e^8$. Hence, $m_2' - s_2 s_2' > 0$. $\square$

Lemma 5.10.5. *Let $q > e^8$. We have*

$$\Big(\frac{1}{2} - \log s_2\Big)\beta_{s,2}' + \beta_{s,2}\frac{s_2'}{s_2} > 0 \ . \tag{5.50}$$

Proof. Let $A(q) = \dfrac{1}{2} - \log s_2$. From Lemma 5.10.3, we obtain

$$\frac{5}{2} < \frac{1}{2} + \log 8 < \frac{1}{2} + \log\log q < A(q) < \frac{1}{2} + \log(2\log q) \ ,$$

and

$$\begin{aligned} A(q)\beta_{s,2}' + \beta_{s,2}\frac{s_2'}{s_2} &= \frac{s_2'}{1-s_2}\Big[A(q)\beta_{s,2} - \frac{2q}{q-2}\frac{A(q)}{s_2^2}\Big] + \frac{s_2'}{1-s_2}\Big[\frac{1-s_2}{s_2}\beta_{s,2}\Big] \\ &= \frac{s_2'}{1-s_2}\Big[\beta_{s,2}\Big(\frac{1}{s_2} + A(q) - 1\Big) - \frac{2q}{q-2}\frac{A(q)}{s_2^2}\Big] \ . \end{aligned}$$

Hence, since $s_2' < 0$, it suffices to show that

$$\Big(\frac{2q}{q-2}\Big)\frac{A(q)}{s_2^2} > \beta_{s,2}\Big(\frac{1}{s_2} + A(q) - 1\Big) = \frac{\beta_{s,2}}{s_2}[1 + (A(q)-1)s_2] \ ,$$

i.e.,

$$\beta_{s,2} < \frac{1}{1 + (A(q)-1)s_2} \cdot \frac{2q}{q-2} \cdot \frac{A(q)}{s_2} \ .$$

Since, $s_2 < 1/\log q$, the right-hand side is greater than

$$\begin{aligned} \frac{1}{1 + (A(q)-1)s_2} \cdot \frac{2qA(q)}{q-2}\log q &> \frac{1}{1 + (A(q)-1)s_2}\Big(\frac{5q}{q-2}\Big)\log q \\ &> \frac{5}{1 + (A(q)-1)s_2}\log q \ , \end{aligned}$$

and

$$\begin{aligned}\beta_{s,2} < g_2(1/q\log q) &= \frac{2\log q}{\log q - 1}\log\frac{q\log q-(q-2)}{2}\\ &< \frac{5}{2}\log(q\log q) \;<\; \frac{15}{4}\log q\ ,\end{aligned}$$

where the last inequality is equivalent to $1/2 > \log(\log q)/\log q$ which is true for $q > e^8$.

Hence, it is enough to show that

$$\frac{1}{1+(A(q)-1)s_2} > \frac{3}{4}\ ,\ \text{i.e.,}\ \frac{1}{3} > (A(q)-1)s_2\ .$$

Since $0 < A(q) < 1/2 + \log(2\log q)$ and $s_2 < 1/\log q$, we obtain, for $q > e^8$,

$$(A(q)-1)s_2 < \frac{\log(2\log q)-1/2}{\log q} < \frac{1}{3}\ .$$

□

Now, we derive the proof of Lemma 5.10.1.

Proof of Lemma 5.10.1. By Lemma 5.10.2,

$$-s_2 + s_2^2 + qs_2' \;=\; -\frac{js_2^2(1-s_2)}{q-2js_2} - s_2(1-s_2) \;=\; (q-js_2)\frac{q}{js_2}s_2'\ ,$$

so that

$$\beta_{s,2}' = \frac{1}{1-s_2}\Big(\beta_{s,2}s_2' - 2\frac{q}{js_2^2}s_2'\Big) = \frac{1}{1-s_2}\Big(\beta_{s,2} - \frac{2q}{js_2^2}\Big)s_2'\ .$$

Now, we return to $f_\star(q)$. We have

$$\begin{aligned} f_\star(q) &= \frac{1}{\beta_{s,2}}(\log qm_2 - \frac{1}{2}) - \frac{1}{8}(qm_2)^2 + \frac{1}{4}m_2 + \frac{251}{2002} \\ &= \frac{1}{\beta_{s,2}}(\log s_2 - \frac{1}{2}) - \frac{1}{8}(s_2)^2 + \frac{1}{4}m_2 + \frac{251}{2002} \ , \end{aligned}$$

so that

$$\begin{aligned} f_\star'(q) &= -\frac{\beta_{s,2}'}{\beta_{s,2}^2}\Big(\log s_2 - \frac{1}{2}\Big) + \frac{1}{\beta_{s,2}}(\frac{s_2'}{s_2}) + \frac{1}{4}(m_2' - s_2 s_2') \\ &= \frac{1}{\beta_{s,2}^2}\Big[\beta_{s,2}'(\frac{1}{2} - \log s_2) + \beta_{s,2}(\frac{s_2'}{s_2})\Big] + \frac{1}{4}(m_2' - s_2 s_2') \ . \end{aligned}$$

Finally, Lemmas 5.10.4 and 5.10.5 prove Lemma 5.10.1. □

5.11 Numerical computations

Recall the definition (5.46) of $f_\star(\cdot)$. In this section, we verify several inequalities numerically. Our purpose is the following proposition. The proof is presented at the end of this section.

Proposition 5.11.1. *The following hold.*

1. *For $5 \le q \le 6500$, we have $F_{\beta_{s,2}}(\mathbf{u}_2) > F_{\beta_{s,2}}(\mathbf{v}_1)$.*
2. *For $6 \le q \le 54$, we have $\frac{d}{d\beta}[f_c(\beta) - F_\beta(\mathbf{u}_2)]\Big|_{\beta=\beta_{s,2}} > 0$.*
3. *$f_\star(6500) > 0$.*

Bounds of $\beta_{s,2}$, m_2 and v_1.

We will obtain the bounds of $\beta_{s,2}$, m_2, and v_1. Fix $q \ge 5$ and let $j = q - 2$. By gradient descent method, we obtain the following.

Algorithm 5.11.2. *We define $\beta_{s,2}^u$ and $\beta_{s,2}^l$ in the following way.*

1. *$t_0 \leftarrow 1\,/\,(2q-4)$.*

2. *While $g_2'(t_i) > 10^{-6}$, let $t_{i+1} \leftarrow t_i - g_2'(t_i)/(300q^2)$.*

3. *If $g_2'(t_i) \leq 10^{-6}$, let $m_2^* \leftarrow t_i$.*

Let $\beta_{s,2}^u := g_2(m_2^*) + (36/q)|g_2'(m_2^*)|$ and $\beta_{s,2}^l := g_2(m_2^*) - (36/q)|g_2'(m_2^*)|$.

We record m_2^* in the above algorithm and let

$$\rho_m := g_2'(m_2^*)/q \; .$$

Algorithm 5.11.3. *We define m_2^u and m_2^l in the following way.*

1. *If $h_2(m_2^*) \geq 0$, let $m_2^u := m_2^* + \rho_m$.*
   (a) *$t_0 \leftarrow m_2^*$.*
   (b) *While $h_2(t_i) \geq 0$, let $t_{i+1} \leftarrow t_i - \rho_m$.*
   (c) *If $h_2(t_i) < 0$, let $m_2^l := t_i - \rho_m$.*

2. *If $h_2(m_2^*) < 0$, let $m_2^l := m_2^* - \rho_m$.*
   (a) *$t_0 \leftarrow m_2^*$.*
   (b) *While $h_2(t_i) \leq 0$, let $t_{i+1} \leftarrow t_i + \rho_m$.*
   (c) *If $h_2(t_i) > 0$, let $m_2^u := t_i + \rho_m$.*

By Newton method, we approximate v_1 which satisfies $g_1(v_1) = \beta_{s,2}$.

Algorithm 5.11.4. *We define v_1^u and v_1^l in the following way.*

1. *Let $t_0 = 0.8/q$ and $t_{-1} = 0$.*
   (a) *While $|t_i - t_{i-1}| > 10^{-5}/q$, let $t_{i+1} \leftarrow t_i - (g_1(t_i) - \beta_{s,2}^u)/g_1'(t_i)$.*
   (b) *If $|t_i - t_{i-1}| \leq 10^{-5}/q$, let $v_1^* := t_i$ and $\rho_v := |t_i - t_{i-1}|$.*

2. *If $g_1(v_1^*) > \beta_{s,2}^u$, let $v_1^u := v_1^* + \rho_v$.*

3. *If* $g_1(v_1^*) \leq \beta_{s,2}^u$, *let*
   (a) $a_0 \leftarrow v_1^*$.
   (b) *While* $g_1(a_i) \leq \beta_{s,2}^u$, *let* $a_{i+1} \leftarrow a_i + \rho_v$.
   (c) *If* $g_1(a_i) > \beta_{s,2}^u$, *let* $v_1^u := a_i + \rho_v$.
4. *If* $g_1(v_1^*) < \beta_{s,2}^l$, *let* $v_1^l := v_1^* - \rho_v$.
5. *If* $g_1(v_1^*) \geq \beta_{s,2}^l$, *let*
   (a) $b_0 \leftarrow v_1^*$.
   (b) *While* $g_1(b_i) \geq \beta_{s,2}^l$, *let* $b_{i+1} \leftarrow b_i - \rho_v$.
   (c) *If* $g_1(b_i) < \beta_{s,2}^l$, *let* $v_1^l := b_i - \rho_v$.

Lemma 5.11.5. *We have* $\beta_{s,2}^l < \beta_{s,2} < \beta_{s,2}^u$, $m_2^l < m_2 < m_2^u$, *and* $v_1^l < v_1 < v_1^u$.

Proof. From the Taylor's theorem, we obtain

$$g_2(m_2 + t) = g_2(m_2) + g_2'(m_2 + t^*)t$$

for some $t^* \in (0, t)$ if $t > 0$ or $t^* \in (t, 0)$ if $t < 0$. Since h_2 is increasing in the neighborhood of m_2, we obtain

$$\begin{aligned} |g_2'(m_2 + t^*)| &= \left| \frac{2q}{[1 - q(m_2 + t^*)]^2} h_2(m_2 + t^*) \right| \\ &\leq \left| \frac{2q}{[1 - q(m_2 + t^*)]^2} h_2(m_2 + |t|) \right| \\ &= \left(\frac{1 - q(m_2 + |t|)}{1 - q(m_2 + t^*)} \right)^2 |g_2'(m_2 + |t|)| \; . \end{aligned}$$

Since $m_2 + t^*$, $m_2 < 1/(2j)$, we obtain

$$\frac{1 - q(m_2 + |t|)}{1 - q(m_2 + t^*)} \leq \frac{1}{1 - q(m_2 + t^*)} \leq \frac{1}{1 - q/(2j)} = \frac{2q - 4}{q - 4} \leq 6 \; ,$$

where the last inequality is from $q \ge 5$. Hence, we have

$$|g_2'(m_2 + t^*)| \le 36 |g_2'(m_2 + |t|)| \ ,$$

so that we have

$$\begin{aligned} |\beta_{s,2} - g_2(m_2 + t)| &= |g_2(m_2) - g_2(m_2 + t)| \\ &\le |g_2'(m_2 + t^*)||t| \ \le \frac{36}{q} |g_2'(m_2 + |t|)| \ , \end{aligned}$$

which proves the first claim. In the last inequality, we use the fact that $|t| < 1/q$.

Since $h_2(t) > 0$ if $t > m_2$ and $h_2(t) < 0$ if $t < m_2$, the second claim is true. Finally, since g_1 is increasing in the neighborhood of v_1, the third claim holds. □

We finally prove Proposition 5.11.1.

Proof of Proposition 5.11.1. From Lemma 5.11.5, we obtain

$$\beta_{s,2}^l < \beta_{s,2} < \beta_{s,2}^u \ , \ m_2^l < m_2 < m_2^u \ , \text{ and } v_1^l < v_1 < v_1^u \ .$$

By elementary computation, we have

$$\begin{aligned} F_{\beta_{s,2}}(\mathbf{u}_2) - F_{\beta_{s,2}}(\mathbf{v}_1) &\ge \frac{1}{4}[q(q-2)\Big(m_2^u - \frac{1}{q-2}\Big)^2 - \frac{2}{q-2}] + \frac{1}{\beta_{s,2}^l} \log m_2^l \\ &\quad - \frac{1}{2}\Big[q(q-1)\Big(v_1^l - \frac{1}{q-1}\Big)^2 - \frac{1}{q-1}\Big] - \frac{1}{\beta_{s,2}^u} \log v_1^u \ , \\ \log q\beta_{s,2} + 2k_2(m_2) &\ge \log(q\beta_{s,2}^l) + 2k_2(m_2^u) \ , \\ f_\star(6500) &\ge \frac{1}{\beta_{s,2}^l}\Big(\log q m_2^l - \frac{1}{2}\Big) - \frac{1}{8}(q m_2^u)^2 + \frac{1}{4} m_2^l + \frac{251}{2002} \ . \end{aligned}$$

The second inequality holds since $k_2(\cdot)$ is decreasing according to (5.39). From the numerical computations, we find that the right-hand sides of the displayed equations are positive for $5 \le q \le 6500$, and this completes the proof. □

5.12 Proof of (5.12)

Proof of (5.12). Since we have

$$Z_N(\beta) = \sum_{\boldsymbol{x}\in\Xi} \frac{N!}{(Nx_1)!\cdots(Nx_q)!} \exp\{-\beta N H(\boldsymbol{x})\} \ ,$$

we can use the elementary bound

$$k\log k - k \le \log k! \le (k+1)\log(k+1) - k \ ,$$

to obtain

$$\begin{aligned}&\sum_{\boldsymbol{x}\in\Xi} \exp\Big\{ -\beta N\Big[H(\boldsymbol{x}) + \frac{1}{\beta}\sum_{i=1}^{q}\big(x_i + \frac{1}{N}\big)\log\big(x_i+\frac{1}{N}\big)\Big] - q\log N\Big\} \\ &\le Z_N(\beta) \le \sum_{\boldsymbol{x}\in\Xi} \exp\Big\{ -\beta N F_\beta(\boldsymbol{x}) + \log(N+1) + N\log\big(1+\frac{1}{N}\big)\Big\} \ .\end{aligned}$$

Hence, by the definition of F_β (5.4), we can obtain

$$\sup_{\boldsymbol{x}\in\Xi}\{-F_\beta(\boldsymbol{x})\} + O\Big(\frac{\log N}{N}\Big) \le \frac{1}{\beta N}\log Z_N(\beta) \le \sup_{\boldsymbol{x}\in\Xi}\{-F_\beta(\boldsymbol{x})\} + O\Big(\frac{\log N}{N}\Big)$$

and the proof is completed. □

Bibliography

[1] L. Alonso, R. Cerf: The three dimensional polyominoes of minimal area. Electron. J. Comb. **3**, R27 (1996)

[2] J. Beltrán, C. Landim: Tunneling and metastability of continuous time Markov chains. J. Stat. Phys. **140**, 1065–1114 (2010)

[3] J. Beltrán, C. Landim: Tunneling and metastability of continuous time Markov chains II. J. Stat. Phys. **149**, 598–618 (2012)

[4] J. Beltrán, C. Landim: Metastability of reversible condensed zero range processes on a finite set. Probab. Theory Related Fields. **152**, 781–807 (2012)

[5] G. Ben Arous, R. Cerf : Metastability of the three dimensional Ising model on a torus at very low temperatures. Electron. J. Probab. **1**, 1–55 (1996)

[6] A. Bianchi, A. Bovier, D. Ioffe: Sharp asymptotics for metastability in the random field Curie–Weiss model. Electron. J. Probab. **14**, 1541–1603 (2009)

[7] A. Bianchi, S. Dommers, C. Giardina: Metastability in the reversible inclusion process. Electron. J. Probab. **22**, 1–34 (2017)

[8] F. Bouchet, J. Reygner: Generalisation of the Eyring–Kramers transition

rate formula to irreversible diffusion processes. Ann. Henri Poincar (B) Probability and Statistics. **17**, 3499–3532 (2016)

[9] A. Bovier, F. den Hollander: *Metastabillity: A Potential-theoretic approach.* Grundlehren der mathematischen Wissenschaften. Springer. (2015)

[10] A. Bovier, F. den Hollander, F.R. Nardi: Sharp asymptotics for Kawasaki dynamics on a finite box with open boundary. Probab. Theory Relat. Fields **135**, 265-310 (2006)

[11] A. Bovier, F. den Hollander, C. Spitoni: Homogeneous nucleation for Glauber and Kawasaki dynamics in large volumes at low temperatures. Ann. Probab. **38**, 661–713 (2010)

[12] A. Bovier, M. Eckhoff, V. Gayrard, M. Klein: Metastability in stochastic dynamics of disordered mean-field models. Probab. Theory Relat. Fields **119**, 99–161 (2001)

[13] A. Bovier, M. Eckhoff, V. Gayrard, M. Klein: Metastability and low lying spectra in reversible Markov chains. Commun. Math. Phys. **228**, 219–255 (2002)

[14] A. Bovier, M. Eckhoff, V. Gayrard, M. Klein: Metastability in reversible diffusion process I. Sharp asymptotics for capacities and exit times. J. Eur. Math. Soc. **6**, 399–424 (2004)

[15] A. Bovier, F. Manzo: Metastability in Glauber dynamics in the low-temperature limit: beyond exponential asymptotics. J. Stat. Phys. **107**, 757–779 (2002)

[16] A. Bovier, S. Marello, E. Pulvirenti: Metastability for the dilute Curie–Weiss model with Glauber dynamics. Electron. J. Probab. **26**, 1-38 (2021)

[17] M. Cassandro, A. Galves, E. Olivieri, M.E. Vares : Metastable behavior of stochastic dynamics: A pathwise approach. J. Stat. Phys. **35**, 603–634 (1984)

[18] M. Costeniuc, R. S. Ellis, H. Touchette: Complete analysis of phase transitions and ensemble equivalence for the Curie–Weiss–Potts model. J. Math. Phys. **46**, 063301 (2005)

[19] P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres, A. Sly: Glauber dynamics for the mean-field Potts model. J. Stat. Phys. **149**, 432–477 (2012)

[20] F. den Hollander, O. Jovanovski: Glauber dynamics on the Erdős-R nyi random graph. *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*, 519–589, Progress in Probabability, **77**, Springer. (2021)

[21] G. Di Ges, T. Leliévre, D. Le Peutrec, B. Nectoux: Jump Markov models and transition state theory: the quasi-stationary distribution approach. Faraday Discussions, **196**, 469–495 (2016)

[22] G. Di Ges, T. Lelivre, D. Le Peutrec, B. Nectoux: Sharp asymptotics of the first exit point density. Annals of PDE. **5**(1), (2019)

[23] G. Di Ges, T. Lelivre, D. Le Peutrec, B. Nectoux: The exit from a metastable state: Concentration of the exit point distribution on the low energy saddle points, part 1. J. Math. Pures Appl. **138**, 242–306 (2020)

[24] A. B. Duncan, T. Leliévre, G.A. Pavliotis: Variance reduction using nonreversible Langevin samplers. J. Stat. Phys. **163**, 457–491 (2016)

[25] P. Eichelsbacher, B. Martschink: On rates of convergence in the Curie–Weiss–Potts model with an external field. Ann. Henri Poincar: Probab. Statist. **51**, 252–282 (2015)

[26] R. S. Ellis, K. Wang: Limit theorems for the empirical vector of the Curie–Weiss–Potts model. Stoch. Process. Their Appl. **35**, 59–79 (1990)

[27] L. C. Evans: Partial differential equations. Graduate studies in Mathematics. **19**. American Mathematical Society. (1998)

[28] H. Eyring: The activated complex in chemical reactions. J. Chem. Phys. **3**, 107–115 (1935)

[29] M. I. Freidlin, A. D. Wentzell: On small random perturbation of dynamical systems. Usp. Math. Nauk **25** (1970) [English transl., Russ. Math. Surv. **25** (1970)]

[30] A. Friedman: *Stochastic differential equations and applications*. Academic Press, Cambridge, 1975.

[31] X. Gao, M. Gürbüzbalaban, L. Zhu: Breaking reversibility accelerates Langevin dynamics for global non-convex optimization. arXiv:1812.07725. (2019)

[32] A. Gaudillere, C. Landim: A Dirichlet principle for non reversible Markov chains and some recurrence theorems. Probab. Theory Related Fields. **158**, 55–89 (2014)

[33] D. Gilbarg, N.S. Trudinger: *Elliptic partial differential equations of second order*. Grundlehren der mathematischen Wissenschaften, **224**. Springer, Berlin, 2015.

[34] R. B. Griffiths, P. A. Pearce: Potts model in the many-component limit. J. Phys. A. **13,** 2143–2148 (1980)

[35] M. Gürbüzbalaban, L. Sagun, U. Simsekli: A tail-index analysis of stochastic gradient noise in deep neural networks. Proceedings of the 36th International Conference on Machine Learning, **97**, 5827-5837 (2019)

[36] B. Helffer, M. Klein, F. Nier: Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach. Mat. Contemp. **26**, 41–85 (2004)

[37] C.-R. Hwang, S.-Y. Hwang-Ma, S.-J. Sheu: Accelerating Gaussian diffusions. Ann. Appl. Probab. **3** , 897–913 (1993)

[38] C.-R. Hwang, S.-Y. Hwang-Ma, S.-J. Sheu: Accelerating diffusions. Ann. Appl. Probab. **15**, 1433–1444 (2005)

[39] H. Kesten, R. H. Schonmann: Behavior in large dimensions of the Potts and Heisenberg models. Rev. Math. Phys. **1,** 147–182 (1982)

[40] S. Kim: Second Time Scale of the Metastability of Reversible Inclusion Processes. arXiv:2007.05918 (2020)

[41] S. Kim, I. Seo: Condensation and metastable behavior of non-reversible inclusion processes. arXiv:2007.05202 (2020)

[42] S. Kim, I. Seo: Metastability of stochastic Ising and Potts models on lattices without external fields. arXiv:2102.05565 (2021)

[43] H. A. Kramers: Brownian motion in a field of force and the diffusion model of chemical reactions. Physica. **7**, 284–304 (1940)

[44] C. Klske, D. Meiner. Stable and metastable phases for the Curie–Weiss–Potts model in vector-valued fields via singularity theory. J. Stat. Phys. **181**, 968–989 (2020)

[45] C. Landim: Metastability for a non-reversible dynamics: The evolution of the condensate in totally asymmetric zero range processes. Commun. Math. Phys. **330**, 1–32 (2014)

[46] C. Landim: Metastable Markov chains. Probab. Surv. **16**, 143–227 (2019)

[47] C. Landim, P Lemire: Metastability of the two-dimensional Blume-Capel model with zero chemical potential and small magnetic field. J. Stat. Phys. **164**, 346-376 (2016)

[48] C. Landim, J. Lee, I. Seo: Metastability of small random perturbation of dynamical system. In preparation. (2022)

[49] C. Landim, D. Marcondes, I. Seo: Metastable behavior of reversible, critical zero-range processes. arXiv:2006.04214 (2020)

[50] C. Landim, D. Marcondes, I. Seo: A resolvent approach to metastability. arXiv:2102.00998 (2021)

[51] C. Landim, M. Mariani, I. Seo: A Dirichlet and a Thomson principle for non-selfadjoint elliptic operators with application to non-reversible ,etastable diffusion processes. Arch. Rational Mech. Anal. **231**, 887–938 (2017)

[52] C. Landim, R. Misturini, K. Tsunoda: Metastability of reversible random walks in potential fields. J. Stat. Phys. **160**, 1449–1482 (2015)

[53] C. Landim, I. Seo: Metastability of non-reversible, mean-field Potts model with three spins. J. Stat. Phys. **165**, 693–726 (2016)

[54] C. Landim, I. Seo: Metastability of non-reversible random walks in a potential field, the Eyring–Kramers transition rate formula. Commun. Pure Appl. Math. **71**, 203– 266 (2018)

[55] C. Landim, I. Seo: Metastability of one-dimensional, non-reversible diffusions with periodic boundary conditions. Ann. Henri Poincar (B) Probability and Statistics. **55**, 1850–1889 (2019)

[56] J. M. Lee: *Introduction to smooth manifolds*. Graduate Texts in Mathematics, **218**. Springer-Verlag, New York, 2012.

[57] J. Lee: Energy landscape and metastability of Curie–Weiss–Potts model. J. Stat. Phys. **187**(2) (2022)

[58] J. Lee, I. Seo: Non-reversible metastable diffusions with Gibbs invariant measure I: Eyring–Kramers formula. Probab. Theory Related Fields. **182**, 849–903 (2022)

[59] J. Lee, I. Seo: Non-reversible metastable diffusions with Gibbs invariant measure II: Markov chain convergence. arXiv:2008.08295. (2020)

[60] T. Leliévre, D. Le Peutrec, B. Nectoux: Exit event from a metastable state and Eyring–Kramers law for the overdamped Langevin dynamics. arXiv:1811.06786 (2018)

[61] T. Leliévre, D. Le Peutrec, B. Nectoux: The exit from a metastable state: concentration of the exit point distribution on the low energy saddle points, part 2. arXiv:2012.08311 (2020)

[62] T. Leliévre, F. Nier, G.A. Pavliotis: Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. J. Stat. Phys. **152**, 237–274 (2013)

[63] T. Leliévre, M. Ramil, J. Reygner. Quasi-stationary distribution for the Langevin process in cylindrical domains, part I: existence, uniqueness and long-time convergence. arXiv:2101.11999 (2021)

[64] D. Le Peutrec and L. Michel. Sharp spectral asymptotics for nonreversible metastable diffusion processes. Probability and Mathematical Physics. **1**, 3–53 (2019)

[65] D. Le Peutrec, B. Nectoux: Repartition of the quasi-stationary distribution and first exit point density for a double-well potential. SIAM J. Math. Anal. **52**, 581–604 (2020)

[66] D.A. Levin, M.J. Luczak, Y. Peres: Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. Probab. Theory Relat. Fields **146**, 223 (2010)

[67] E. Lubetzky, A. Sly: Cutoff for the Ising model on the lattice. Invent. Math. **191**, 719–755 (2013)

[68] E. Lubetzky, A. Sly: Information percolation and cutoff for the stochastic ising model. J. Am. Math. Soc. **29**, 729–774 (2016)

[69] E. Lubetzky, A. Sly: Universality of cutoff for the Ising model. Ann. Probab. **45**, 3664–3696 (2017)

[70] L. Michel: About small eigenvalues of the Witten Laplacian. Pure Appl. Anal. **1**, 149 - 206 (2019)

[71] J. Milnor: *Morse theory*. Annals of Mathematics Studies, **51**. Princeton University Press, New Jersey, 1969.

[72] F.R. Nardi, A. Zocca: Tunneling behavior of Ising and Potts models in the low-temperature regime. Stoch. Process. Their Appl. **129**, 4556–4575 (2019)

[73] B. Nectoux: Mean exit time for the overdamped Langevin process: the case with critical points on the boundary. Communications in Partial Differential Equations. (2021)

[74] E.J. Neves: A discrete variational problem related droplets at low temperatures. J. Stat. Phys. **80**, 103–123 (1995)

[75] E.J. Neves, R.H. Schonmann: Critical droplets and metastability for a glauber dynamics at very low temperatures. Commun. Math. Phys. **137**, 209–230 (1991)

[76] C. Oh, F. Rezakhanlou: Metastability of zero range processes via Poisson equations. Unpublished manuscript. (2019)

[77] E. Olivieri, M.E. Vares : *Large deviations and metastability.* Encyclopedia of Mathematics and Its Applications. **100**. Cambridge University Press, Cambridge. (2005)

[78] M. Ostilli, F. Mukhamedov: Continuous- and discrete-time Glauber dynamics. First- and second order phase transitions in mean-field Potts models. EPL. **101**(6) (2013)

[79] R. G. Pinsky: *Positive harmonic functions and diffusion.* Cambridge Studies in Advanced Mathematics, **45**. Cambridge University Press, Cambridge, 1995.

[80] F. Rassoul-Agha, T. Sepplinen: *A course on large deviations with an introduction to Gibbs measures.* Graduate Studies in Mathematics. **162**. American Mathematical Society. (2015)

[81] L. Rey-Bellet, K. : Irreversible Langevin samplers and variance reduction: a large deviations approach. Nonlinearity. **28**, (2015)

[82] L. Rey-Bellet, K. Spiliopoulos: Improving the convergence of reversible samplers. J. Stat. Phys. **164**, 472–494 (2016)

[83] F. Rezakhanlou, I. Seo: Scaling limit of small random perturbation of dynamical systems. arXiv:1812.02069 (2018)

[84] I. Seo: Condensation of non-reversible zero-range processes. Commun. Math. Phys. **366**, 781–839 (2019)

[85] M. Slowik: Contributions to the potential theoretic approach to metastability with applications to the random field Curie–Weiss–Potts model. Doctoral Thesis, Technische Universitt Berlin. (2012)

[86] M. Sugiura: Metastable behaviors of diffusion processes with small parameter. J. Math. Soc. Japan **47**, 755–788 (1995)

[87] S. R. S. Varadhan: *Lectures on diffusion problems and partial differential equations.* Tata Institute of Fundamental Research, Bombay, 1980.

[88] K. Wang: Solutions of the variational problem in the Curie–Weiss–Potts model. Stoch. Process. Their Appl. **50**, 245–252 (1994).

[89] F.Y. Wu: The Potts model. Rev. Mod. Phys. **54**, 235–268 (1982)

국문초록

본 학위논문에서는 비가역적 랑주뱅 동역학의 메타안정성을 연구했다. 그 결과로 Gibbs 불변분포를 갖는 비가역적 확산확률과정의 Eyring–Kramers 공식을 증명했는데, 이 공식은 전이 시간의 기댓값을 정확히 추산하는 것이다. 이에 더해 Eyring–Kramers 공식을 발전시켜, 적절한 시간 규모에서 비가역적 확산확률과정이 최솟값 사이의 마르코프 사슬로 수렴한다는 것을 증명했다.

마지막으로, 복잡한 메타안정성을 나타내는 복잡한 잠재함수 위의 메타안정적 동역학의 예시로써 Curie–Weiss–Potts 모형을 소개한다. 이 모형의 에너지 분포와 해당 모형 연관 된 heat-bath Glauber 동역학의 메타안정성을 연구했다.

주요어휘: 메타안정성, 통계물리, 랑주뱅 동역학, 마르코프사슬 모형 단순화, Curie–Weiss–Potts 모형
학번: 2017-29414

감사의 글

약 12년 전, 친구들과 스키장에 있을 때 부모님으로부터 서울대학교 수리과학부통계학과군에 합격했다는 전화를 받았었습니다. 그리고 다음 날, 학부 선배로부터 신입생 오리엔테이션에 관한 전화를 받았을 때의 기억이 아직도 생생합니다. 단순히 수학이 좋아서 큰 고민 없이 수학과 진학을 선택했는데, 그때의 저는 정말 무식하고 용감했던 것 같습니다. 하지만 지금 생각해보면 잘한 선택이었고, 과거의 용감한 저에게 정말 고마운 생각이 듭니다.

중간의 군 생활 2년을 제외하고, 학부 4년 반과 대학원 5년 반, 총 10년을 서울대에서 보내고, 이제는 12년 만에 정든 학교를 떠나게 됐습니다. 저 혼자만의 힘으로는 긴 학위과정을 무사히 마칠 수 없었기에, 저에게 도움을 주신 분들께 감사의 인사를 전하고자 합니다. 가장 먼저, 박사과정을 무사히 마칠 수 있도록 지도해주신 서인석 교수님께 무한한 감사의 말씀을 드립니다. 교수님의 아낌없는 지원과 가르침 덕분에, 박사과정 동안 수학 공부에 전념할 수 있었고 많은 것을 배울 수 있었습니다. 학부만 졸업하고 아무것도 몰랐던 저를, 믿고 지도해주시고 올바른 길로 인도해주셔서 정말 감사드립니다. 학위과정 동안 받은 모든 가르침을 잊지 않고, 앞으로 열심히 공부하는 지혜로운 수학자가 되겠습니다. 진심으로 감사드리고 존경합니다.

다음으로, 공동연구를 진행하고 있는 Claudio Landim 교수님께도 너무나 감사드립니다. 교수님께 정말 많은 것을 배울 수 있었고, 교수님 덕분에 국제학회도 참가하며 많은 사람을 만나고 많은 경험을 할 수 있었습니다. 앞으로도 교수님과 좋은 연구를 할 수 있도록 열심히 공부하겠습니다.

바쁘신 와중에, 학위논문 심사위원을 맡아주신 심사위원장 김판기 교수님을 비롯하여 박형빈 교수님, 먼 걸음 해주신 강남규 교수님과 남경식 교수님께 정말 감사드립니다. 또한 논문자격시험에서 재밌는 주제를 할 수 있게 지도해주신 이훈희 교수님께도 감사드립니다. 좋은 강의를 해주시고 무사히 졸업을 할 수 있도록 도움을 주신 모든 수리과학부 교수님들께도 감사드립니다.

연구실 동료들에게도 감사의 인사를 전합니다. 특히, 교수님의 첫 제자로 같이 공부를 시작한 선우에게 진심으로 고맙다고 하고 싶습니다. 동생이지만 때로는 친구처럼 지낼 수 있어서 외롭지 않은 대학원 생활을 보낼 수 있었고, 옆에서 성실히 공부하는 모습을 보면서 많은 점을 배우고 자극받을 수 있었습니다. 같이 즐겁게 스터디를 했던 연구실 후배 지호, 승우, 동준, 산하, 대철, 용우, 젊은 박사로서의 마음가짐과 앞으로의 연구 분야에 대해 많은 점을 배울 수 있었던 Mouad Ramil 박사님께도 진심으로 감사드립니다. 앞으로, 가까운 동료 연구자로서, 어려운 길을 다 같이 함께 할 수 있으면 좋겠습니다.

다음으로는 대학원 친구들에게 감사를 전하고 싶습니다. 학부 동기이자 선배 연구자로서 이정표가 되어준 성수, 공부하면서 지칠 때 조언을 아끼지 않았던 종민이, 학위논문과 논문심사를 준비하는 데 정말 많은 도움을 준 호식이 형, 화현회 친구이면서 논문자격시험에 큰 도움을 준 상준이에게 깊은 감사를 전합니다. 대학원 생활을 즐겁게 해준 경성이, 기우, 재무, 수경누나도 좋은 시간을 함께 보내주어서 고맙습니다.

대학원 밖에서 도움을 준 소중한 분들에게도 감사의 인사를 전합니다. 먼저, 철없던 시절 화현회에서 만나 같이 희망찬 미래를 꿈꾸었던 용석이형, 민규, 두기에게도 고마움을 전합니다. 이제는 결혼도 했고 각자의 인생을 살기 바쁘지만, 앞으로도 자주 만나 이야기를 나눌 수 있으면 좋겠습니다. 아울러, 이제 막 세상의 빛을 보게 된 두기의 딸 세이의 탄생을 축하합니다. 만나면 지금도 고등학생 때인 것 같아 너무 즐거운, 가장 오래 알고 지낸 수지고 친구 경진이, 병훈이, 왕산이, 동휘, 다빈이, 성현이에게 큰 고마움을 느낍니다. 같이 즐겁게 전공 공부도 하고 서로의 고민도 들어주었던 학부 친구 규태, 해강이, 상일이, 전역하고 화현회에 돌아올 수 있게 같은 팀으로 연주회에 올라주었던 규범이, 영규, 경화, 현화에게도 고마움을 전합니다. 덕분에 저의 대학 생활이 즐거움으로 가득했습니다. 입대 전부터 친하게 지냈던 화현회 친구 세진이, 종호형, 중현이, 자전거 타면서 친해져서 대학원 말년에 즐겁게 같이 운동한 슬하, 한아도 고맙습니다. 마지막으로, 5년이 넘는 대학원 생활에, 항상 함께하며, 가장 큰 힘과 위로가 되어준 소중한 주리에게 진심을 다해 고마움을 전합니다. 기쁠 때나 슬플 때, 항상 주리가 옆에 있어 주어서 정말 다행이었습니다.

사랑하는 가족에게도 감사의 인사를 전합니다. 부모님이 바쁘실 때 돌봐주시고 사랑을 주신 외할머니와 외할아버지, 제가 대학원을 진학하고 연구자의 길을 걷기로 결심하는 데 많은 영감을 주신 외삼촌께도 정말 감사드립니다. 특히, 초등학생 때부터 외할머니께서 저에게 대학원에 가라고 말씀하셨기 때문에 어렸을 때부터 공부하는 직업을 갖고 싶었던 것 같습니다. 이제는 하늘에서 지켜보시겠지만, 할머니 덕분에 저도 외삼촌처럼 어엿한 박사가 되었습니다. 정말 감사드립니다.

마지막으로, 태어났을 때부터 조건 없는 무한한 지원과 사랑을 주신 저희 부모님과, 함께 성장하면서 외로울 때 친구가 되어준 동생 중익이에게 말로 표현할 수 없는 감사를 드립니다. 제가 부모님의 아들인 것과 중익이의 형인 것이 자랑스럽습니다. 앞으로는 제가 받은 사랑을 돌려드리고, 우리 가족이 제가 아들이고 형인 것이 자랑스러울 수 있는 삶을 살겠습니다.

하고 싶은 말도, 감사한 분도 많아서 글이 길어졌습니다. 여기에 모두 담을 수는 없었지만, 지난 30년간 제가 잘 살아올 수 있도록 영향을 준 친구, 선후배, 선생님을 포함한 모든 분께 감사를 드리며 이만 마칩니다. 저도 앞으로 누군가에게 도움을 줄 수 있는 삶을 살도록 하겠습니다. 모두 진심으로 감사합니다!

2023년 2월 이중경올림