Dynamics of the spherical Sherrington-Kirkpatrick model and average case complexity for top eigenvectors

by

Tingzhou Yu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Statistics

Waterloo, Ontario, Canada, 2023

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this Master's thesis, we investigate the Langevin dynamics on the spherical Sherrington-Kirkpatrick (SSK) model, a classical mean-field spin glass model. The first contribution of this thesis is the asymptotic limit of energy function of the SSK model, a critical property linked to the model's equilibrium state. The thermodynamic limit of energy of the system is characterized in terms of a system of integro-differential equations as the size of the system goes to infinity. Then we look at the behavior of the limiting dynamics as the time goes to infinity. This long time behavior of the energy has a phase transition. In the regime of below the critical inverse temperature, the limiting result is zero. In the regime of above the critical inverse temperature, the limiting result is a constant depending on the temperature.

The second contribution of this thesis is that we analyze the complexity of the zero-temperature Langevin dynamics (a.k.a. the gradient descent algorithm) on the SSK model. We establish lower and upper bound for the hitting time, defined as the first time required for the output of the algorithm to achieve a small overlap with the eigenvector corresponding to the smallest eigenvalue of the Wigner matrix.

Acknowledgements

Firstly, I would like to thank my advisor, Aukosh Jagannath, for his invaluable guidance and support during my time at the University of Waterloo.

I would like to thank Shoja'eddin Chenouri and Yi Shen for serving as a committee member of my thesis defense.

At the University of Waterloo, I would like to thank Adam Kolkiewicz, Pengfei Li, Paul Marriott, David Saunders, Leilei Zeng, for teaching me so much in the past two years and patiently answering my questions. I would like to thank the staff members of the Department of Statistics and Actuarial Sciences at UW for their friendly administration of the department. I would like to especially thank Mary Lou Dufton.

I want to thank friends during my studies: Wei Liang, Yiran Wang, Ziyue Shi, Jiayi Cheng, Dong Yao, Xiyue Han, Kecheng Li, and Yaolun Yin.

Dedication

This is dedicated to the one I love.

Table of Contents

List of Figures

Chapter 1

Introduction

The Sherrington-Kirkpatrick (SK) model is a widely studied mathematical model in statistical mechanics. It has drawn considerable interest from physicists, mathematicians, and computer scientists. This model was first introduced by Sherrington and Kirkpatrick in their seminal paper [SK75]. Since then, it has become a prominent tool for analyzing the behavior of spin glasses, which are disordered magnetic systems that exhibit complex behavior making them as a topic of great interest in physics [EA75].

This thesis focuses on several aspects of the SK model and related problems. We will first review the SK model and its basic properties in Section 1.1. In Section 1.2, we review the phenomenon of aging in the SK model. Finally, we discussed the complexity of gradient descent related to zero-temperature Langevin dynamics on SK model in linear algebra in Section 1.3. Specifically, we will focused on the problem of computing eigenvectors of symmetric Wigner matrices using gradient descent.

1.1 Sherrington-Kirkpatrick (SK) model

We begin by introducing the spin glass model. For dimension $N \geqslant 1$, we consider the state space $\{-1,+1\}^N$. In this model, there are N classical spins labeled as $[N] := \{1,2,\ldots,N\}$, which can take the values of either -1 or $+1$. The spins interact with each other through pairwise couplings J_{ij}, which are independent standard Gaussian random variables modeling the quality of interaction between agents i and j. Let $H_{\mathbf{J}} : \{-1,+1\}^N \to \mathbb{R}$ be the Hamiltonian. The interpretation of $H_{\mathbf{J}}(X)$ is the internal magnetic energy of each possible state $X \in \{-1,+1\}^N$. In order to understand the material in thermal equilibrium, one would like to describe the proportion of time spent in different states is given by the Gibbs measure on $\{-1,+1\}^N$ at inverse temperature $\beta \in \mathbb{R}$:

$$G_N(X) := \frac{\exp(\beta H_{\mathbf{J}}(X))}{Z_\beta}, \tag{1.1}$$

where

$$Z_\beta := \sum_{X \in \{-1,+1\}^N} e^{\beta H_{\mathbf{J}}(X)} \tag{1.2}$$

is called the partition function and μ is the probability measure defined on $\{-1,+1\}$.

A natural question in the spin glass model is the asymptotic behavior of the maximum of $H_{\mathbf{J}}$, for example the Dean's problem [Pan13b, Section 1.1]. In realistic models, the key feature of a spin glass is that different pairs of spins interact in very different ways which is known as the Edwards-Anderson spin glass model introduced by [SZ82].

The Sherrington-Kirkpatrick (SK) model is a mean-field simplification of the Edwards-Anderson spin glass model, in which one ignores the spatial locations of the spins. The SK model is motivated by the optimization problem called Dean's problem: given N agents labeled $[N] = \{1,\dots,N\}$ in which the like or dislike between individuals i and j is given by g_{ij}. Then Dean would like to maximize the "happiness" of the N-agent system defined by

$$H_g(X) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{ij} \mathbf{1}_{\{X_i = X_j\}}, \tag{1.3}$$

where $g_{ij} = g_{ji}$ for $1 \le i < j \le N$ are independent identically distributed (i.i.d.) random variables with mean 0 and variance 1, and g_{ii} for $i = 1,\dots,N$ are i.i.d. random variables with mean 0 and variance 2. Note that the system (1.3) is equivalently as

$$H_g(X) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{ij} X_i X_j, \tag{1.4}$$

which is the sum of the pairwise couplings multiplied by the product of the spins.

This is a non-trivial maximization problem, even in the case where $N = 3$ (see Figure 1.1 for a illustration). The presence of conflicting interactions between agents can lead to frustration and the inability to reconcile all intuitive assignments of the agents into two groups.

The SK model is associated with a free energy of the model defined via

$$F_N(\beta) = \frac{1}{N} \mathbb{E}\left(\log Z_N(\beta)\right), \tag{1.5}$$

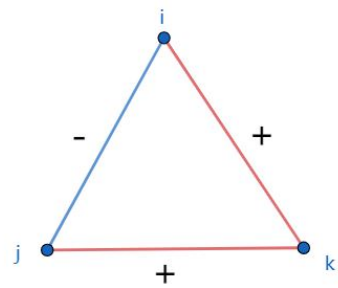

Figure 1.1: A simple example of the SK model used to illustrate the non-triviality of the problem, even in the case $N = 3$. Suppose there are three agents labeled i, j and k, where $J_{ij} < 0$, $J_{jk} > 0$ but $J_{ik} > 0$.

where $Z_N(\beta)$ is the partition function defined by

$$Z_N(\beta) := \sum_{X \in \{-1,+1\}^N} \exp\left(\frac{\beta}{\sqrt{N}} H_g(X)\right) \tag{1.6}$$

and inverse temperature parameter $\beta \in \mathbb{R}$ is tunable. The large N asymptotics of this free energy have been of interest to mathematicians studying the SK model. In 1979, [Par79] proposed a formula for this limit called the replica symmetric breaking solution. This formula was later confirmed by the work of [Tal06]. [Pan13a] generalized these results, providing a comprehensive understanding of the free energy and the thermodynamic properties of the SK model in the large N limit. Indeed, if we have a formula for the free energy, then this will give a formula for the leading order behavior of the Dean's problem [Pan13b, Section 1.1]. These mathematical advances have greatly enhanced our understanding of the behavior of spin glasses and have contributed to the development of new algorithms for solving optimization problems in statistics and machine learning (see e.g., [Cha07, DMM09, ZK16]).

Originally, the SK model considered the case where the spins are Ising spins and the probability measure follows the Bernoulli law $\mu(dx) = \frac{1}{2}(\delta_{-1} + \delta_{+1})$ defined on $\{-1,+1\}$. One natural dynamics associated with such dynamics are Glauber dynamics (see e.g., [Gru96, KMP01]).

One generalization of the Ising spins to a continuous configuration space M (e.g., a sphere instead of a hypercube) called spherical spin glasses. Let $\mathbb{S}^{N-1}(\sqrt{N})$ be $N-1$ dimensional sphere of radius $\sqrt{N}$ in $\mathbb{R}^N$:

$$\mathbb{S}^{N-1}(\sqrt{N}) := \{\boldsymbol{X} \in \mathbb{R}^N : \|\boldsymbol{X}\|^2 = N\}, \tag{1.7}$$

where $\|\cdot\|$ is the ℓ^2 norm.

The spherical Sherrington-Kirkpatrick (SSK) model [KTJ76] is defined by the Hamiltonian

$$H_g^{\mathrm{SSK}}(X) = -\frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{ij} X_i X_j, \tag{1.8}$$

where the spin variables $X = (X_1, \ldots, X_N) \in \mathbb{R}^N$ lie on the sphere $\mathbb{S}^{N-1}(\sqrt{N})$. The Hamiltonian of the SSK model is same as the SK model but with the constraint of spin variables on the sphere.

The associated Langevin dynamics of the SSK model (see (3.2) in Chapter 3) were considered by [SZ81, AG97, ADG01, BA03].

1.2 The phenomenon of aging in the SK model

In Section 1.1, we introduced the basic definition of the SK model, which is one of the most widely studied models of spin glasses. One of the main problems in the study of spin glasses is understanding the equilibrium phase transition (see e.g., [Geo11, GHM01]), which is a fundamental change in the system's properties that occurs as temperature is lowered and it can reveal information about the natural of the ground state of the system.

However, in some cases, the system may relax to equilibrium so slowly that it never actually reaches it. This phenomenon is known as "aging", and it is integral to the study of spin glasses dynamics, both experimentally and theoretically. Aging is a phenomenon that affects the system's decorrelation properties over time. According to [BA03], this means that the longer the system exists, the longer it takes to forget its past. This phenomenon has been studied extensively by various authors (see e.g., [CK93, CK94, VHO+07]). The study of aging and its effect on spin glass dynamics is an active area of research in physics, with important implications for the understanding of the low-temperature behavior of these systems.

In [ADG01], the authors focus on such systems and study the phenomenon of "aging" in spherical spin glasses to characterize the low-temperature behavior of the dynamics. They studied the Langevin dynamics for the SK model using correlation functions that satisfy a system of integro-differential equations known as the Crisanti-Horner-Sommers-Cugliandolo-Kurchan (CHSCK) equations [CS92, CK93]. This was the first mathematical literature on aging in spin glasses, and it opened the door for further research on this topic.

Our work in Chapter 3 aims to understand the asymptotic limit of the energy and focus on the long-time behavior of the energy of the system. The energy of the spherical spin glasses model is described by the Hamiltonian function, which describes the interactions between the spin variables on the surface of a high-dimensional sphere. The energy function is a key quantity in the spherical spin glasses model, as it determines the equilibrium properties of the system and governs its dynamical behavior (see e.g., [SN13, AC18]). By studying the long-time behavior of the energy function, we can gain insight into the way the energy of the system evolves over time and how it is affected by the dynamics of the spin variables. In particular, the long-time behavior of the energy function can reveal information about the aging behavior of the system.

1.3 The complexity of the gradient descent

In this section, we discuss the potential connection of our results on aging in spin glasses to linear algebra problems, and we review some background and previous results on the complexity of power method.

The study of aging in spin glasses model as discussed in Section 1.2 provides important insights into the dynamics of the energy function in the SK model, which is the focus of our work in Chapter 3. This insight can be used to develop more efficient algorithms for solving difficult linear algebra problems, such as computing the eigenvectors of Wigner random matrices. However, to achieve this goal, we need to understand the complexity of iterative methods, which have received less attention in complexity theory [Sma97].

The study of the complexity of algorithms in linear algebra has been a topic of interest for many years by [Sma97]. While direct algorithms that solve problems in a finite number of steps have been extensively studied, iterative methods such as those required for the matrix eigenvalue problem have received less attention in complexity theory in [Sma97].

The power method is a well-known iterative algorithm that approximates the eigenvector corresponding to the dominant eigenvalue, which is the largest in absolute value. However, for Hermitian random matrices, the complexity of the power method for obtaining a dominant eigenvector is infinite by [Kos88]. [Kos88] showed that the upper bound of the complexity is $O(N^2 \log N)$, conditioned on all the eigenvalues being positive. Another algorithm for calculating the dominant vector was investigated by [Kos91], who showed that, under certain conditions, the average number of iterations required is $O(\log N + \log|\log\epsilon|)$. [DT17] studied the performance of three algorithms for computing the eigenvalues of sample covariance matrices and showed that the complexity is about $O\left(\left(\frac{\log\epsilon^{-2}}{\log N} - \frac{3}{2}\right) N^{2/3} \log N\right)$, regardless of the specific distribution of the entries.

In our work presented in Chapter 4, we investigate the complexity of an algorithm that uses spherical gradient descent inside aging to study the equilibrium of the spherical SK model based on the zero-temperature dynamics, i.e., taking $\beta = \infty$ in the Langevin dynamics defined in (3.2). The spherical gradient descent is an optimization method that updates the spin variables in the direction of the negative gradient of the energy function on a unit sphere. This algorithm can be understood as the continuous analogue version of the power method. To ensure that our algorithm achieves its goal, we consider the hitting time when the overlap between the output and the eigenvector is positive, where overlap is a measure of similarity between two spin configurations. Our work represents an improvement over previous studies, as we provide a lower bound of $O(N^{2/3})$ and upper bound of $O(N^{2/3} \log N)$ for computing eigenvectors of Wigner random matrices whose entries are independent and identically distributed Gaussian variables. This insight may prove useful in developing more efficient algorithms for solving difficult linear algebra problems in the future.

1.4 Organization of the thesis

In Chapter 2, we review a variety of classical mathematical results that we will use in this thesis. The main contribution of the thesis begins in Chapter 3. The main results include the integro-differential equation of the energy of the SK model and an explicit formula for the limiting of the energy. In Chapter 4, we analyze the complexity of the zero-temperature dynamics for finding the eigenvectors corresponding to extreme eigenvalues of Wigner matrices.

Chapter 2

Preliminaries

In this chapter, we collect some results that were used in the proof of the main results presented in Chapter 3 and Chapter 4 in random matrix theory and probability theory. In Section 2.1, we provide a brief review of the Wigner matrix, semicircle law, and Tracy-Widom law. In Section 2.2, we review some basic tools from stochastic analysis, including stochastic differential equations and Itô's formula. Finally, in Section 2.3, we review the definition and basic properties of the Bessel function. These results are key components in the proof of our main results.

2.1 Random matrix theory results

Random matrix theory is a branch of mathematics that studies the behavior of large matrices with random entries. While the idea and motivation of studying random matrices dates back to 1950s with the pioneering works of Wigner, Dyson, and others [Wig55, Dys70]. There are already several textbooks that describe in detail the theory and work related to random matrices (see e.g., [AGZ10, Tao12, TV14]). We will review some facts about semicircle law and edge universality in this section.

Let us start with the definition of the (symmetric) Winger matrix ensemble.

Definition 2.1.1. *[Tao12, Section 2.3] Let $N \geqslant 1$ be an integer. Consider a symmetric $N \times N$ matrix $Y = \{Y_{ij}\}_{1 \leq i,j \leq N}$, which hence has N eigenvalues. Assume that the following conditions hold:*

- $\{Y_{ij}\}_{1 \leq i \leq j \leq N}$ *are independent random variables;*

- *The diagonal entries $\{Y_{ii}\}_{1\leq i\leq N}$ are identically distributed with finite variance, and the off-diagonal entries $\{Y_{ij}\}_{1\leq i\leq j}$ are identically distributed with mean zero and unit variance.*

The matrix Y is known as a ***symmetric Wigner matrix****.*

When the random variables Y_{ij} and Y_{ii} are real Gaussian with $\mathbb{E}|Y_{ii}|^2 = 2$, the Wigner matrix Y will be called **Gaussian Orthogonal Ensemble (GOE)**. Similarly, when the strictly upper triangular entries Y_{ij} are complex and Y_{ii} are real Gaussian with $\mathbb{E}|Y_{ii}|^2 = 1$, the Wigner matrix Y will be called Gaussian Unitary Ensemble (GUE).

By [Tao12, Corollary 2.3.6], we have $\|Y\|_{\text{op}} = O(\sqrt{N})$. We always consider the normalized Wigner matrix $\mathbf{J} = \frac{1}{\sqrt{N}}Y$ in the following results.

Let (X,d) be a metric space. Let μ_n $(n \geqslant 1)$ and μ be Borel probability measures on X. Let $C_b(X)$ be a collection of all bounded and continuous functions defined on X. We say that μ_n converges in weakly to μ if

$$\int f d\mu_n \to \int f d\mu, \text{ as } n \to \infty, \text{ for all } f \in C_b(X). \tag{2.1}$$

One of the fundamental universality results in random matrix theory is the Wigner semicircle law. This describes the limiting distribution of eigenvalues of a Wigner matrix converges weakly to the semicircle law as the size of the matrix goes to infinity. The precise distribution of entries does not affect the conclusion of the law as long as the matrix satisfies conditions of the Wigner matrix as in Definition 2.1.1.

Given any $N \times N$ normalized Wigner matrix $\mathbf{J}$, we consider the empirical spectral measure of $\mathbf{J}$:

$$\mu_{\mathbf{J}} := \frac{1}{N}\sum_{i=1}^{N} \delta_{\sigma^i}, \tag{2.2}$$

where $\sigma^1 \leq \cdots \leq \sigma^N$ are eigenvalues of J and $\delta_\cdot$ is the Dirac measure.

Theorem 2.1.2. *[Tao12, Theorem 2.4.2] Let J be the $N \times N$ normalized symmetric Wigner matrix. Then the empirical spectral distribution $\mu_{\mathbf{J}}$ converges weakly to the (Wigner) semicircle law almost surely (hence also in probability)*

$$\mu_{sc} := \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{|x|\leq 2}dx. \tag{2.3}$$

There are several methods to prove the semicircle law. One of the methods to prove this result is the moment method. The idea of this method is to compare the moments of the empirical spectral measure of the random matrix with the moments of the semicircular law, and show that the convergence as the matrix size tends to infinity. Another method to prove the semicircle law is the Stieltjes transform method, which is a similar technique to Fourier method. The Stieltjes transform of a probability measure is a complex-valued function. More precisely, given any probability measure μ on the real line, we can form its **Stieltjes transform**: for any $z \in \mathbb{R} \setminus \mathrm{supp}(\mu)$

$$m(z) := \int_{\mathbb{R}} \frac{1}{x-z} d\mu(x). \tag{2.4}$$

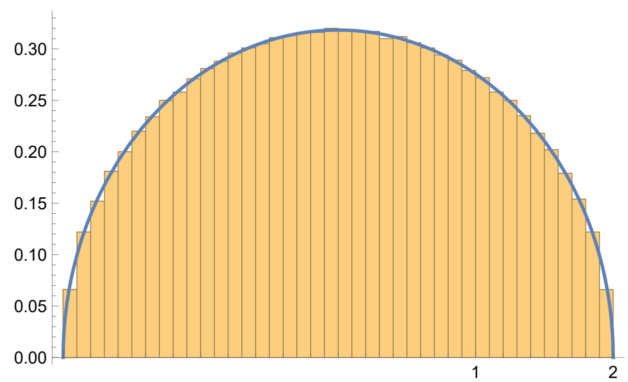

Figure 2.1: Semicircle law of GOE: plot of empirical spectral distribution of 10000×10000 matrix from GOE converging weakly to semicircle law.

Except for the semicircle law, there are two universal phenomenon about the local statics of eigenvalues of random matrices called the bulk universality and the edge universality (see e.g. [DKM+99, ESY11, TV11, LY14]).

This thesis will focus on the edge universality. The two endpoints -2 and $+2$ of the semicircle law are called the edge of spectrum. Note that the typical or average spacing between eigenvalues is of order $O(N^{-1})$ (i.e., they are roughly equally spaced with $[-2,2]$). However, for the Gaussian ensembles, Tracy and Widom [TW94, TW96] give more precise information about the rescaling of the largest eigenvalue around the edge, which is described as follows.

Lemma 2.1.3. *[AGZ10, Theorem 3.1.5] Let $\sigma^1 \le \cdots \le \sigma^N$ be the eigenvalues of a normalized Wigner matrix $\mathbf{J}$. Then we have*

$$\lim_{N\to\infty} \mathbb{P}\left(N^{2/3}(\sigma^N - 2) \le x\right) = F_\beta(x) \tag{2.5}$$

where $F_\beta(x)$ is the Tracy-Widom distribution functions described by the Painlevé equations for $\beta = 1,2,4$ corresponding to orthogonal, unitary, symplectic ensemble, respectively. This result also holds for the smallest eigenvalues σ^1.

See [BBD08] for the definition of the Painlevé equations.

[For93] proved the joint distribution of the k largest eigenvalues can be expressed in terms of the Airy kernel.

Lemma 2.1.4. *[For93] Let $\sigma^1 \le \cdots \le \sigma^N$ be the eigenvalues of a Wigner matrix J. Then the limiting joint distribution of the k largest eigenvalue for Gaussian ensembles*

$$\lim_{N\to\infty} \mathbb{P}(N^{2/3}(\sigma^N - 2) \le x_1, \dots, N^{2/3}(\sigma^{N-k+1} - 2) \le x_k) = F_{\beta,k}(x_1, \dots, x_k), \tag{2.6}$$

where $F_{\beta,k}(x_1,\dots,x_k)$ is still called the Tracy-Widom distribution. This result also holds for the smallest eigenvalues $\sigma^1,\dots,\sigma^k$.

Definition 2.1.5. *[TV14] We say that the Wigner matrix ensemble $\mathbf{J} = \{Y_{ij}\}_{1\le i,j\le N}$ **obeys condition C1** with constant C_0 [TV14] if one has*

$$\mathbb{E}|Y_{ij}|^{C_0} \le C, 1 \le i,j \le N \tag{2.7}$$

for some constant C (independent of N).

It was conjectured that the Tracy-Widom law holds for general Wigner matrices if the Wigner matrix $\mathbf{J} = \{Y_{ij}\}_{1\le i,j\le N}$ obeying condition **C1** with $C_0 = 4$ by the numerical results in [BBP07]. This fourth moment is the optimal value of the moment [TV14]. There has been a lot of partial progress on this conjecture [Sos99, Joh01, LY14].

[LY14] proved a simple necessary and sufficient criterion for Tracy-Widom law. Here we just list the sufficient condition as follows.

Theorem 2.1.6. *[LY14, Theorem 1.2] Let $\mathbf{J}$ be the normalized Wigner matrix defined in Definition 2.1.1. We denote by $\sigma_1 \le \sigma_2 \le \dots \sigma_N$ the eigenvalues of $\mathbf{J}$. If the off-diagonal entry of the Wigner matrix satisfies*

$$\lim_{s\to\infty} s^4 \mathbb{P}(|Y_{12}| \geqslant s) = 0, \tag{2.8}$$

then the joint distribution function of k rescaled the largest eigenvalues

$$\mathbb{P}(N^{2/3}(\sigma^N - 2) \le s_1, N^{2/3}(\sigma^{N-1} - 2) \le s_2, \dots, N^{2/3}(\sigma^{N-k+1} - 2) \le s_k) \tag{2.9}$$

has a limit as $N \to \infty$, which coincides with that in the GUE and GOE cases, i.e., it weakly converges to the Tracy-Widom distribution. This result also holds for the smallest eigenvalues $\sigma^1,\dots,\sigma^k$.

Note that any distribution with a finite fourth moment satisfies the criterion (2.8), however, the converse statement does not hold. See [LY14] for a counterexample.

Next, we will discuss the distribution of eigenvectors of Wigner matrices. For a given matrix $\mathbf{J}$ sampled from the GOE (a special case of Wigner matrices), the eigenvalues of $\mathbf{J}$ are almost surely distinct by [AGZ10, Theorem 2.5.2]. Due to the GOE ensemble's invariance under orthogonal transformations, we have the following Lemma.

Lemma 2.1.7. *[AGZ10, Corollary 2.5.4] Let $\{v_1, v_2, \dots, v_N\}$ be the eigenvectors corresponding to the eigenvalues $\sigma^1, \dots, \sigma^N$ of a matrix $\mathbf{J}$ drawn from GOE. Each of the eigenvectors $v_1, \dots, v_N$ is distributed uniformly on the unit sphere:*

$$\mathbb{S}_+^{N-1} := \{\mathbf{x} = (x_1, \dots, x_N) : x_i \in \mathbb{R}, \|\mathbf{x}\|_2 = 1, x_1 > 0\}. \tag{2.10}$$

Note that there would be the sign ambiguity for unit eigenvectors. To resolve this ambiguity, we randomly and independently choose each eigenvector from the two available options.

Recall the following fact about the Gaussian random vector in high dimensions.

Lemma 2.1.8. *[Ver18, Exercise 3.3.7] For a random vector $g \sim N(0, I_N)$, $g/\|g\|_2$ is uniformly distributed on the unit sphere $\mathbb{S}^{N-1}$.*

Since the direction $g/\|g\|_2$ is uniformly distributed on the unit sphere $\mathbb{S}^{N-1}$. Thus, we can represent a random vector $X \sim \mathrm{Unif}(\sqrt{N}\mathbb{S}^{N-1})$ as $X = \sqrt{N}\frac{g}{\|g\|_2}$.

Combing the above facts, we can prove the following Lemma that for a unit vector $q \in \mathbb{S}^{N-1}$, $\sqrt{N} v_i \cdot q$ are asymptotically normally distributed for $i = 1, \dots, N$.

Lemma 2.1.9. *Let $\mathbf{J} = \{Y_{ij}\}_{1 \le i,j \le N}$ be a matrix drawn from GOE. Let $q \in \mathbb{S}^{N-1}$ be a unit vector in $\mathbb{R}^N$. Let $v_i \in \mathbb{S}^{N-1}$ be chosen randomly among all unit eigenvectors corresponding to eigenvalue σ^i of $\mathbf{J}$. Then $\sqrt{N} v_i \cdot q$ converges to $N(0,1)$ in distribution as $N \to \infty$.*

Proof. By Lemma 2.1.7 and Lemma 2.1.8, we write $v_i = g/\|g\|_2$ and $q = h/\|h\|_2$ with $g = (g_1, \dots, g_N) \sim N(0, I_N)$ and $h = (h_1, \dots, h_N) \sim N(0, I_N)$, respectively. By Slutsky's theorem [CB21, Theorem 5.5.17], we have

$$\sqrt{N} v_i \cdot q = \frac{\frac{1}{\sqrt{N}} \sum_{i=1}^N g_i h_i}{\sqrt{\left(\frac{1}{N}\sum_i g_i^2\right)\left(\frac{1}{N}\sum_i h_i^2\right)}} \tag{2.11}$$

converges to $N(0,1)$ in distribution, where the numerator converges to $N(0,1)$ in distribution by central limit theorem, and the denominator converges to 1 in probability by the weak law of large number. □

One can extend the above Lemma to the general Wigner matrix ensembles:

Lemma 2.1.10. *[TV12] Let* $\mathbf{J} = \{Y_{ij}\}_{1\le i,j\le N}$ *be a random real symmetric matrix obeying the following condition for a sufficiently large constant* C_0*:*

$$\mathbb{E}|Y_{ij}|^{C_0} \le C \tag{2.12}$$

for some constants $C > 0$*. Assume that* $Y_{ij} = -Y_{ij}$ *for* $1 \le i, j \le N$*. Let* $q \in \mathbb{S}^{N-1}$ *be a unit vector in* $\mathbb{R}^N$*. For each* $i = 1, \dots, N$*, let* $v_i \in \mathbb{S}^{N-1}$ *be chosen randomly among all unit eigenvectors with eigenvalue* σ^i*. Then* $\sqrt{N} v_i \cdot q$ *tends to* $N(0,1)$ *in distribution as* $N \to \infty$*.*

2.2 Stochastic differential equations

Most of the concepts and proofs featured in this section were covered in the course STAT 902 by Prof. Yi Shen in Winter 2022. Other notable sources include [Øks03, LG16, Dur19].

Stochastic differential equations (SDEs) are a powerful too for modeling systems that evolve randomly over time. They are widely used in many fields such as finance and machine learning (see e.g., [HØS00, LSS22]). Let us start with the definition of Brownian motion.

Definition 2.2.1. *[Dur19, Section 7.1] We say a real-valued process* B_t *for* $t \geqslant 0$ *is a one-dimensional Brownian motion if it satisfies the following properties:*

- $B_0 = 0$.
- B_t *has continuous paths.*
- *(independent increments) if* $t_0 < t_1 < \cdots < t_n$*, then* $B_{t_0}, B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}}$ *are independent.*
- *(stationary increments) For* $t > s \geqslant 0$*,* $B_t - B_s \sim N(0, t-s)$*.*

The one-dimensional Brownian motion defined above serves as the basic for the development of the theory of the stochastic integrals.

For $n > 1$, a n-dimensional Brownian motion $B_t = (B_t^1, \dots, B_t^n)$ starting at $x \in \mathbb{R}^n$ is defined as a collection of independent one-dimensional $B_t^1, \dots, B_n^t$, where $B_0^i = x_i$ for $i = 1, \dots, n$.

The Itô's formula is a central tool in stochastic calculus. In the following we give the simplest version of Itô's formula for one-dimensional Brownian motion from [Dur19, Theorem 7.6.1]. This formula can be extended to n-dimensional Brownian motion [Dur19, Theorem 7.6.7] and even to the semimartingale [LG16, Chapter 5].

Theorem 2.2.2 (Itô's formula)**.** *Suppose that $f:\mathbb{R}\to\mathbb{R}$ has two continuous derivatives. The with probability one, for all $t \geqslant 0$,*

$$f(B_t)-f(B_0)=\int_0^t f'(B_s)dB_s+\frac{1}{2}\int_0^t f''(B_s)ds. \tag{2.13}$$

Lemma 2.2.3 provides an integration by parts formula for stochastic integrals as follows.

Lemma 2.2.3. *[Øks03, Theorem 4.1.5] Let $f(t)$ be a continuous and of bounded variation with respect to $t\in[0,T]$. Then*

$$\int_0^t f(s)dB_s = f(t)B_t-\int_0^t B_s df(s). \tag{2.14}$$

Recall that a filtration on the probability space $(\Omega,\mathscr{F},\mathbb{P})$ is a non-decreasing family $\{\mathscr{F}_t\}_{t\geqslant 0}$ of sub-σ-algebra of $\mathscr{F}$ such that $\mathscr{F}_s\subset\mathscr{F}_t$ for $s\leq t\leq\infty$, where we define $\mathscr{F}_\infty := \sigma(\cup_{t\geqslant 0}\mathscr{F}_t)$. Recall that a stochastic process $\{X_t\}_{t\geqslant 0}$ is called adapted to a filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$ if X_t is $\mathscr{F}_t$-measurable for every $t\geqslant 0$.

One important result called Itô isometry from [Øks03, Corollary 3.1.7] is as follows.

Lemma 2.2.4 (Itô isometry)**.** *Let $f(t,\omega):[0,\infty)\times\Omega\to\mathbb{R}$ be a $\mathscr{B}\times\mathscr{F}$-measurable function, where $\mathscr{B}$ is the Borel $\sigma-$algebra on $[0,\infty)$. Assume that $f(t,\omega)$ is $\mathscr{F}_t$-adapted and $\mathbb{E}\left[\int_a^b f(t,\omega)^2dt\right]<\infty$. Let B_t be a Brownian motion on $[0,\infty)$. Then*

$$\mathbb{E}\left[\left(\int_a^b f(t,\omega)dB_t\right)^2\right]=\mathbb{E}\left[\int_a^b f(t,\omega)^2dt\right]. \tag{2.15}$$

One of main applications of Itô isometry is to compute the variance for random variables that are given as Itô integrals.

A stochastic differential equations (SDEs) has the following form [Øks03, Chapter 7]:

$$dX_t = a(t,X_t)dt+b(t,X_t)dB_t, \tag{2.16}$$

where $X_t \in \mathbb{R}^n$ is the the solution to the SDEs, $a(t,x) \in \mathbb{R}^n$ is the drift term that describes the deterministic evolution of the system, $b(t,x) \in \mathbb{R}^{n\times m}$ is the diffusion term that describes the random fluctuations, and B_t the is a m-dimensional Brownian motion.

The existence and uniqueness theorem of SDEs provides a theoretical framework for studying the solutions to SDEs (See e.g., [Øks03, Theorem 5.2.1]). Roughly speaking, it requires a linear growth upper bound and a Lipschitz condition for functions $a(t,x)$ and $b(t,x)$. It guarantees the existence of a unique solution to the SDEs under suitable conditions on the coefficients $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$. The solution X_t is called a strong solution if it satisfies the SDEs (2.16) almost surely for all time, given the Brownian motion B_t in advance.

A classic example of a SDE is the Ornstein-Uhlenbeck (OU) process (see e.g., [Øks03, Chapter 5]), which is widely used in finance and physics [MMS09]. The OU process is a mean-reverting process that models the behavior of systems that tend to return to a fixed equilibrium state. It is described by the following SDE: for two positive parameters $\theta, \sigma > 0$,

$$dX_t = -\theta X_t dt + \sigma dB_t. \tag{2.17}$$

By applying Itô's formula in Theorem 2.2.2 to $e^{\theta t} X_t$, we get

$$X_t = X_0 e^{-\theta t} + \sigma \int_0^t e^{-(\theta - s)} dB_s. \tag{2.18}$$

Another example is called the Langevin dynamics [XCZG18], which is a stochastic process that models the behavior of a particle in a viscous medium subject to a potential force and thermal fluctuations. It is described as a SDE as follows.

$$dX_t = -\nabla U(X_t)dt + \sqrt{2\beta^{-1}} dB_t, \tag{2.19}$$

where X_t represents the position of the particle at time t, $U(\cdot)$ is the potential energy of the particle, β is the inverse temperature, and B_t is the standard Brownian motion on $\mathbb{R}^d$.

2.3 Bessel function

In this section, we will introduce Bessel function and modfied Bessel function along with their integral representations, relations, and asymptotic behavior. These results will play a key role in the proof of our main results.

Bessel functions, first defined by the mathematician Daniel Bernoulli and then generalized by Friedrich Bessel, are canonical solutions of Bessel's differential equation in [AS48, Section 9.1].

An alternative definition of the (modified) Bessel function, for integer values of n, is possible using integral representation:

Definition 2.3.1. *[AS48, Section 9.1] The Bessel function is given by*

$$B_n(x) := \frac{i^{-n}}{\pi} \int_0^\pi e^{ix\cos\theta} \cos(n\theta) d\theta. \tag{2.20}$$

for $n \in \mathbb{N}$ and $x \in \mathbb{R}$.

The modified Bessel function is given by

$$I_n(x) := \frac{1}{\pi} \int_0^\pi e^{x\cos\theta} \cos(n\theta) d\theta. \tag{2.21}$$

for $n \in \mathbb{N}$ and $x \in \mathbb{R}$.

The relation between the Bessel function and the modified Bessel function is given by

$$I_n(x) = e^{-in\pi/2} B_n(xe^{i\pi/2}) \tag{2.22}$$

as shown in [AS48, Section 9.6].

We will use the following lemma about the recurrence relation and derivatives of modified Bessel functions.

Lemma 2.3.2. *[AS48, Section 9.6] Let I_n be the modified Bessel function defined as in Definition 2.3.1. For $n \in \mathbb{N}$,*

$$I_n'(x) = I_{n+1}(x) + \frac{n}{x} I_n(x), \tag{2.23}$$

and $I_0'(x) = I_1(x)$.

By [AS48, Section 9.6], we have the following asymptotic results of modified Bessel functions.

Lemma 2.3.3. *Let I_n be the modified Bessel function defined as in Definition 2.3.1. For $n \in \mathbb{N}$, as $x \to \infty$ we have*

$$\lim_{x\to\infty} \frac{I_n(x)}{x^{-1/2} e^x} = \frac{1}{\sqrt{2\pi}}. \tag{2.24}$$

Chapter 3

The long time behavior of the energy of the spherical SK model

In this chapter, we will consider the Langevin dynamics for the SSK model. Recall that the definition of the spherical Sherrington-Kirkpatrick (SSK) model in Section 1.1. The SSK model is described by the Hamiltonian

$$H_{\mathbf{J}}(X) = -\sum_{1\le i,j\le N} J_{ij}X_iX_j = -\boldsymbol{X}\cdot\mathbf{J}\boldsymbol{X} \tag{3.1}$$

where the spin variables $\boldsymbol{X} = (X_1,\dots,X_N) \in \mathbb{R}^N$ lie on the sphere constraint $\|\boldsymbol{X}\|^2 = \sum_{i=1}^N X_i^2 = N$, and $\mathbf{J} = \{J_{ij}\}_{1\le i,j\le N}$ is the normalized symmetric Wigner matrix with mean zero and $\mathbb{E}\left[J_{ij}^2\right] = \frac{1}{N}, \mathbb{E}\left[J_{ii}^2\right] = \frac{2}{N}$ for $1 \le i < j \le N$.

3.1 Main results

We consider the Langevin dynamics for the Sherrington-Kirkpatrick (SK) model defined by the following system of stochastic differential equations (SDEs) as in [ADG01]:

$$dX_t^i = \sum_{j=1}^N J_{ij}X_t^j dt - f'\left(\frac{1}{N}\sum_{j=1}^N (X_t^j)^2\right)X_t^j dt + \beta^{-1/2}dW_t^i, \tag{3.2}$$

where $\mathbf{J} = \{J_{ij}\}_{1\le i,j\le N}$ is a symmetric matrix of centered Gaussian random variables such that $\mathbb{E}[J_{ij}^2] = \frac{1}{N}$ and $\mathbb{E}[J_{ii}^2] = \frac{2}{N}$ for $1 \le i < j \le N$, $f : [0,\infty) \to \mathbb{R}$ satisfies f' to be non-negative and Lipschitz, β is a positive constant, and $\{W_t^i\}_{1\le i\le N}$ is an $N-$dimensional Brownian motion, independent of $\{J_{ij}\}_{1\le i,j\le N}$ and of the initial data $\{X_0^i\}_{1\le i\le N}$.

For any $N \geqslant 1$ and $T \geqslant 0$, the SDE (3.2) has a unique strong solution $\boldsymbol{X}_t = \{X_t^i : 1 \le i \le N, t \in [0,T]\}$ on $\mathscr{C}([0,T],\mathbb{R}^N)$. See [ADG01, Lemma 6.7] for a proof.

The second term in (3.2) is a Lagrange multiplier in order to implement a smooth spherical constraint [ADG01]. The simplification caused by the SSK model is the invariance under rotation for the SDE (3.2).

We write $\mathbf{J} = G^T D G$, where G is an orthogonal matrix with the uniform law on the sphere and $D = \mathrm{diag}(\sigma^1,\dots,\sigma^N)$ is the diagonal matrix of the eigenvalues $\{\sigma^i\}_{1\le i\le N}$ of $\mathbf{J}$. As $N\to\infty$, we have $\frac{1}{N}\sum_{i=1}^N \delta_{\sigma^i}$ converges weakly to the semicircle law μ_D with compact support $[-2,2]$ by Theorem 2.1.2.

To simplify the SDE (3.2), we let both sides of (3.2) be multiplied by the rotation matrix G which is invariant under rotation. We take $\boldsymbol{Y}_t := G\boldsymbol{X}_t$ and $B_t := GW_t$. Then the SDE under the rotation is given by

$$dY_t^i = \left(\sigma^i - f'\left(\|Y_t\|^2/N\right)\right) Y_t^i dt + \beta^{-1/2} dB_t^i, \tag{3.3}$$

where $\|\cdot\|$ is the ℓ^2 norm.

Denote by

$$K_N(t,s) := \frac{1}{N}\sum_{i=1}^N X_t^i X_s^i \tag{3.4}$$

the empirical correlation function. We use abbreviated notation $K_N(t) := K_N(t,t)$ for convenience. [ADG01] studied the dynamics of the empirical correlation K_N and the limiting point as $N\to\infty$ (N is the size of the system), which is the unique solution to a CHSCK equation as follows.

Theorem 3.1.1. *[ADG01, Theorem 2.3] Assume that the initial data $\{X_0^i\}_{1\le i\le N}$ are i.i.d with law μ_0 so that $\mathbb{E}_{X\sim\mu_0}[e^{\alpha X}] < \infty$ for some $\alpha > 0$. Fix $T \geqslant 0$. As $N\to\infty$, K_N converges almost surely to deterministic limits K. Moreover, the limit K is the unique solution to the following integro-differential equation:*

$$\begin{aligned} K(t,s) &= e^{-\int_0^t f'(K(w))dw - \int_0^s f'(K(w))dw} \mathop{\mathbb{E}}_{(\sigma,X_0)\sim\pi^\infty}[e^{\sigma(t+s)}X_0^2] \\ &\quad + \beta^{-1}\int_0^{t\wedge s} e^{-\int_r^t f'(K(w))dw - \int_r^s f'(K(w))dw} \mathop{\mathbb{E}}_{(\sigma,X_0)\sim\pi^\infty}[e^{\sigma(t+s-2r)}]dr, \end{aligned}$$

where $\pi^\infty = \mu_D \otimes \mu_0$ and here we write $K(s) := K(s.s)$.

Remark 3.1.2. *As emphasized in [ADG01], the aging is very dependent on initial conditions. In addition to considering i.i.d. initial condition, the author also considers other three types of initial conditions: the rotated independent initial conditions, the top eigenvector initial conditions, and the stationary initial conditions.*

Remark 3.1.3. *Based on the thermodynamic limit of $K_N(t,s)$ as $N \to \infty$, [ADG01] study the long time evaluations of $K(t,s)$ and established a dynamical phase transition in terms of the asymptotic of $K(t,s)$ in [ADG01, Proposition 3.2]. This is a first mathematical proof of the aging phenomenon.*

Next, we similarly consider how to describe how the energy of the system evolves over time. Recall that the quadratic Hamiltonian of SSK model is defined by $H_{\mathbf{J}}(\boldsymbol{X}_t) = \boldsymbol{X}_t^T \mathbf{J} \boldsymbol{X}_t$. Note that we have $H_{\mathbf{J}}(\boldsymbol{Y}_t) = \boldsymbol{Y}_t^T D \boldsymbol{Y}_t$, where $\boldsymbol{Y}_t = G X_t$ and $\mathbf{J} = G^T D G$.

Let

$$H_N(t) := \frac{1}{N} H_{\mathbf{J}}(\boldsymbol{Y}_t) = \frac{1}{N} \sum_{i=1}^{N} \sigma^i (Y_t^i)^2 \tag{3.5}$$

be the **energy** of the system.

Our first result characterizes the limiting behavior of the energy $H_N(t)$ of the SSK model as $N \to \infty$ for $t \in [0,T]$ as follows.

Theorem 3.1.4. *Assume that the initial data $\{X_0^i\}_{1 \le i \le N}$ are i.i.d with law μ_0 so that $\mathbb{E}_{X \sim \mu_0}[e^{\alpha X}] < \infty$ for some $\alpha > 0$. Fix $T \geqslant 0$. Let K be the solution defined as in Theorem 3.1.1. As $N \to \infty$, H_N converges almost surely to deterministic limits H. Moreover, the limit H is the unique solution to the following integro-differential equation:*

$$H(t) = e^{-2\int_0^t f'(K(w))dw} \mathop{\mathbb{E}}_{(\sigma, X_0) \sim \pi^\infty} [\sigma e^{2\sigma t} X_0^2] \tag{3.6}$$

$$+ \beta^{-1} \int_0^t e^{-2\int_s^t f'(K(w))dw} \mathop{\mathbb{E}}_{(\sigma, X_0) \sim \pi^\infty} [\sigma e^{2\sigma(t-s)}] ds \tag{3.7}$$

where $\pi^\infty = \mu_D \otimes \mu_0$ and here we write $K(s) := K(s.s)$.

Theorem 3.1.4 provides a precise characterization of $H(t)$. However, the expression of $H(t)$ is unclear because it involves the fixed point equation of $K(t)$ in Theorem 3.1.1. The key point of this model is that we can exactly study the long time behavior of the energy $H(t)$ as $t \to \infty$.

In order to precisely determine the limit of the energy, we will define $K(0,0)$ to be 1 and take function

$$f(x) = \frac{cx^2}{2}, \tag{3.8}$$

where c is a positive constant. By [Tao12, Theorem 2.4.2], recall that the density function of semicircle law μ_D is given by

$$d\mu_D = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{\{-2\le x\le 2\}}dx. \tag{3.9}$$

Let $m : \mathbb{R}\setminus \mathrm{supp}(\mu_D) \to \mathbb{R}$ be the Stieljes transform of the probability measure μ_D given by

$$m(s) = \mathbb{E}_{\sigma\sim\mu_D}\left[\frac{1}{s-\sigma}\right] = \frac{2}{s+\sqrt{s^2-4}}. \tag{3.10}$$

Let β_c be the critical temperature such that

$$\beta_c = \frac{c}{4}m(2) = \frac{c}{4}. \tag{3.11}$$

Our second result is as follows.

Theorem 3.1.5. *Assume that $K(0,0) = 1$ and $f(x) = \frac{cx^2}{2}$ for some positive constants $c > 0$. Let H be the unique solution given in Theorem 3.1.1. Then, for $\beta \leqslant \beta_c$, we have*

$$\lim_{t\to\infty} H(t) = 0. \tag{3.12}$$

For $\beta > \beta_c$, we have

$$\lim_{t\to\infty} H(t) = \frac{4\beta - c}{2^{\frac{5}{2}}\sqrt{\pi}c\beta} + \frac{1}{2}\beta^{-1}. \tag{3.13}$$

Theorem 3.1.5 describes the limit of energy function $H(t)$ as $t \to \infty$. This concludes a dynamical phase transition phenomenon. See Figure 3.1 for the existence of a jump discontinuity in the asymptotic limit of the function $H(t)$, where we set $c = 1$.

The proof of Theorem 3.1.5 utilizes tools and techniques from the paper [ADG01], with some modifications made to their results. We borrow the notation from [ADG01] and write

$$R(t) := \exp\left(2\int_0^t f'(K(w))dw\right) \tag{3.14}$$

with $K(w) = K(w,w)$.

Then the expression of $H(t)$ in Theorem 3.1.4 becomes

$$H(t) = R(t)^{-1}\left(\mathbb{E}[\sigma e^{2\sigma t}] + \beta^{-1}\int_0^t R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr\right) \tag{3.15}$$

Note that the limit of $H(t)$ is governed by the asymptotic of the derivative of the moment generating function of σ. So it suffices to characteristic the limit of $R(t)$ and $\mathbb{E}\left[\sigma e^{2\sigma t}\right]$.

Similarly, we consider the asymptotic limit of $H(t)/K(t)$ as $t \to \infty$.

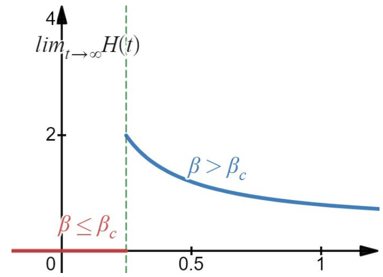

Figure 3.1: In this figure, we set $c = 1$ and plot the limiting behavior of the function $H(t)$ as t approaches infinity. $H(t)$ exhibits a jump discontinuity in the phase transition at the critical inverse temperature $\beta_c = 0.25$.

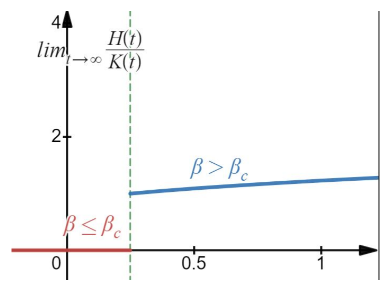

Figure 3.2: In this figure, we set $c = 1$ and plot the limiting behavior of the function $H(t)/K(t)$ as t approaches infinity. There is a jump discontinuity in the phase transition at the critical inverse temperature $\beta_c = 0.25$.

Corollary 3.1.6. *Assume the same setting as in Theorem 3.1.5. Then for $\beta \leqslant \beta_c$, we have*

$$\lim_{t\to\infty} \frac{H(t)}{K(t)} = 0. \tag{3.16}$$

For $\beta > \beta_c$, we have

$$\lim_{t\to\infty} \frac{H(t)}{K(t)} = \frac{2^{-3/2}(4\beta - c) + \sqrt{\pi}c}{2^{-5/2}(4\beta - c) + \sqrt{\pi}c} \tag{3.17}$$

The proof of Corollary 3.1.6 as $\beta > \beta_c$ is the same as the proof in Theorem 3.1.5. For the proof of the case $\beta \leq \beta_c$, we apply the same idea of proving [LSS22, Theorem 1.2]. See Figure 3.2 for the plot of the asymptotic limit of $H(t)/K(t)$, where we set $c = 1$.

Proof. Consider first the regime $\beta > \beta_c$. Recall that $R(t)$ is defined as in (3.14) and $K(t)$ is given in Theorem 3.1.1. We rewrite $K(t)$ as follows:

$$K(t) = R^{-1}(t)\left(\mathbb{E}[e^{2t\sigma}] + \beta^{-1}\int_0^t R(r)\mathbb{E}[e^{2(t-r)\sigma}]dr\right). \tag{3.18}$$

We apply Lemma 3.2.4 and Lemma 3.2.3, and plug in the asymptotic limit of $R(t)$ and $E[e^{2t\sigma}]$, we get

$$\lim_{t\to\infty} K(t) = \frac{2^{-\frac{7}{2}}\pi^{-\frac{1}{2}}(4\beta+1)}{C_\beta(4\beta-c)} + \frac{1}{2}\beta^{-1}. \tag{3.19}$$

Combining this result and Theorem 3.1.5, we obtained the desired result.

For $\beta < \beta_c$, we will show that $\lim_{t\to\infty} K(t) = C$ for some non-zero constants C. Take $h(t) = R(t)K(t)$. Note that

$$R'(t) = 2cK(t)R(t) = 2ch(t). \tag{3.20}$$

Thus, we have $2c\mathscr{L}_h(z) = z\mathscr{L}_R(z) - 1$.

By Lemma 3.2.5, we have

$$\mathscr{L}_R(z) = \frac{1 + cm(z)}{2z - c\beta^{-1}m(z)} \tag{3.21}$$

Hence, we get

$$\mathscr{L}_g(z) = \frac{1}{2c}\left(\frac{czm(z)(1+\beta^{-1}) - z}{2z - c\beta^{-1}m(z)}\right). \tag{3.22}$$

Note that $\mathscr{L}_g(z)$ has a simple pole at s_β, which is a solution to $2z = c\beta^{-1}m(z)$. Thus there exists a constant $C > 0$ such that

$$\lim_{z\to 0} z\mathscr{L}_g(z + s_\beta) = C \tag{3.23}$$

By [BA03, Lemma 7.2], we have

$$\lim_{t\to\infty} e^{-2s_\beta t}h(t) = C \tag{3.24}$$

Hence, $\lim_{t\to\infty} K(t) = C$ for some non-zero constants C. Hence, the desired follows from Theorem 3.1.5.

For $\beta = \beta_c$, we apply the same proof as $\beta < \beta_c$. Note that $\lim_{t\to\infty} K(t) = C_1$ for some non-zero constants C_1. Hence, we still obtain the desired result by Theorem 3.1.5. □

3.2 Proof of Theorem 3.1.5

We will first give the representation of the characteristic function of the semicircle law by the Bessel function.

Lemma 3.2.1. *Let $B_1(t)$ be the Bessel function defined as in Definition 2.3.1 for $n = 1$. Recall that the eigenvalues σ of $N \times N$ normalized symmetric Wigner matrix $\mathbf{J}$ follow the semicircle law with distribution μ_D as in (3.9). Then we have*

$$\mathbb{E}[e^{it\sigma}] = \frac{B_1(2t)}{t} \tag{3.25}$$

Proof. By [Chu01, Theorem 6.2.3], it is enough to calculate the inverse of the characteristic function of $B_1(2t)/t$ is the density of the semicircle law.

Note that by the inversion formula we have

$$\begin{aligned}\frac{1}{2\pi}\int_{\mathbb{R}} \frac{1}{t} B_1(2t) e^{-itx} dt &= \frac{1}{2i\pi^2}\int_{\mathbb{R}}\int_0^{\pi} \frac{1}{t} e^{2it\cos\theta}\cos\theta e^{-itx} d\theta dt \\ &= \frac{1}{2i\pi^2}\int_0^{\pi}\cos\theta \underbrace{\int_{\mathbb{R}} \frac{1}{t} e^{it(2\cos\theta - x)} dt}_{=-i\pi\cdot\mathrm{Sign}(2\cos\theta - x)} d\theta \\ &= -\frac{1}{2\pi}\int_0^{\pi}\cos\theta\cdot\mathrm{Sign}(2\cos\theta - x) d\theta,\end{aligned}$$

where $\mathrm{Sign}(\cdot)$ is the Sign function.

Clearly, we have $\mathrm{Sign}(2\cos\theta - x) = -1$ for $x > 2$, and $\mathrm{sgn}(2\cos\theta - x) = 1$ for $x < -2$. For both cases, the above integral is zero due to $\int_0^{\pi}\cos\theta d\theta = 0$.

Consider the case that $-2 \le x \le 2$. Set $u = 2\cos\theta - x$. The above integral becomes

$$\begin{aligned}-\frac{1}{2\pi}\int_{-2-x}^{2-x} \frac{u+x}{2}\frac{1}{\sqrt{1-(\frac{u+x}{2})^2}}\mathrm{Sign}(u) du &= \frac{1}{2\pi}\left\{\int_{-2-x}^{0} \frac{u+x}{2}\frac{1}{\sqrt{1-(\frac{u+x}{2})^2}} du - \int_0^{2-x} \frac{u+x}{2}\frac{1}{\sqrt{1-(\frac{u+x}{2})^2}} du\right\} \\ &= \frac{1}{2\pi}\left(2\int_{-1}^{\frac{x}{2}} \frac{y}{\sqrt{1-y^2}} dy - 2\int_{\frac{x}{2}}^{1} \frac{y}{\sqrt{1-y^2}} dy\right) \\ &= \frac{1}{2\pi}\sqrt{4-x^2},\end{aligned}$$

where we take variable $y = (u+x)/2$ in the second line for $-2 \le x \le 2$. □

The following Lemma derives the asymptotic limit of the derivative of the moment generating function of σ.

Lemma 3.2.2. *Recall that the eigenvalues σ of $N \times N$ normalized symmetric Wigner matrix $\mathbf{J}$ follow the semicircle law with distribution μ_D as in (3.9). Then we have:*

$$\lim_{t\to\infty} \frac{\mathbb{E}_{\sigma\sim\mu_D}[\sigma e^{t\sigma}]}{t^{-3/2}e^{2t}} = \frac{1}{2\sqrt{\pi}}. \tag{3.26}$$

Proof. Substitute t by it in Lemma 3.2.1, then we have

$$\mathbb{E}[e^{t\sigma}] = \frac{B_1(-2it)}{-2it} \tag{3.27}$$

where $B_1(\cdot)$ is the Bessel function.

By equation (2.22), we get

$$\mathbb{E}[e^{t\sigma}] = \frac{I_1(2t)}{2t}, \tag{3.28}$$

where $I_1(2t)$ is the modified Bessel function.

Thus, the derivative of the moment generating function can be expressed as

$$\mathbb{E}[\sigma e^{t\sigma}] = -t^{-2}I_1(2t) + 2t^{-1}I_1'(2t) \tag{3.29}$$

Combine (3.29) and Lemma 2.3.2 we have

$$\mathbb{E}[\sigma e^{t\sigma}] = \frac{I_2(2t)}{t}. \tag{3.30}$$

By Lemma 2.3.3, it yields the desired result.

□

Similarly, we have the following result.

Lemma 3.2.3. *Assume the same setting holds as in Lemma 3.2.2, we have*

$$\lim_{t\to\infty} \frac{\mathbb{E}_{\sigma\sim\mu_D}[e^{t\sigma}]}{t^{-3/2}e^{2t}} = \frac{1}{4\sqrt{\pi}}. \tag{3.31}$$

Let $m(s)$ be defined as in (3.10). We define

$$p(s,\beta) := \frac{2\beta s}{c} - m(s). \tag{3.32}$$

Recall that β_c is the critical temperature defined in (3.11). For any $\beta \in (0,\beta_c)$, there exists a unique solution of $p(s,\beta)=0$ on the interval $(2,\infty)$, denoted by s_β. We can solve for s_β as $s_\beta = 2(1-(1-\beta/\beta_c)^2)^{-1/2}$. For $\beta > \beta_c$, we simply define $s_\beta = 2$.

We can get the asymptotic limit of $R(t)$ as follows.

Lemma 3.2.4. *[ADG01, Lemma 3.3] Recall that $s_\beta = 2(1-(1-\beta/\beta_c)^2)^{-1/2}$ for any $\beta \in (0,\beta_c)$, and $s_\beta = 2$ for $\beta \geqslant \beta_c$. Let $\Psi = 0$ for $\beta < \beta_c$, $\Psi = \frac{1}{2}$ for $\beta = \beta_c$, and $\Psi = \frac{3}{2}$ for $\beta > \beta_c$. Then there exists a constant $C_\beta > 0$ such that*

$$\lim_{x\to\infty} \frac{R(x)}{x^{-\Psi} e^{2xs_\beta}} = C_\beta. \tag{3.33}$$

Moreover, we have

$$C_\beta = \begin{cases} \frac{\beta(cm(s_\beta)+1)}{2\beta - cm'(s_\beta)}, & \beta < \beta_c \\ \frac{\beta(c+1)}{c}, & \beta = \beta_c \\ \frac{c\beta(4\beta+1)}{(4\beta-c)^2}, & \beta > \beta_c \end{cases}$$

where c is the coefficient constant defined in (3.8) and $m(s)$ is defined as in (3.10).

Combining Lemma 3.2.2 and Lemma 3.2.4 we can characterize the limit of $H(t)$ as $t\to\infty$. In order to give a precise result of the asymptotic limit, we also need to do Laplace transformation on both sides of the equation (3.15) to get an identity as follows.

Define the Laplace transform of the function $R(t)$ for $z > 2$ by

$$\mathcal{L}_R(z) := \int_0^\infty e^{-2zt} R(t)dt. \tag{3.34}$$

Lemma 3.2.5. *The Laplace transform $\mathcal{L}_R(z)$ satisfies the equation*

$$2z\mathcal{L}_R(z) - 1 = cm(z)(1+\beta^{-1}\mathcal{L}_R(z)), \tag{3.35}$$

where c is the coefficient constant defined in (3.8) and $m(s)$ is defined as in (3.10).

Proof. Note that

$$K(t)R(t) = K(t)e^{2c\int_0^t K(w)dw} = \frac{1}{2c}\partial_t R(t).$$

Then we have the linear Volterra integro-differential equation

$$R'(t) = 2cK(t)R(t) = 2c\left(\mathbb{E}[e^{2\sigma t}] + \beta^{-1}\int_0^t R(r)\mathbb{E}[e^{2\sigma(t-r)}]dr\right). \tag{3.36}$$

The Laplace transform of the LHS in (3.36) is

$$\mathscr{L}_{R'}(z) = -R(0) + 2z\mathscr{L}_R(z) = -1 + 2z\mathscr{L}_R(z).$$

Note that the term inside the integral on RHS in (3.36) can be expressed as the convolution of $R(t)$ and $e^{2\sigma t}$. We write it as $(e^{2\sigma\cdot} * R)(t)$ and use the fact that the Laplace transform of this one is equal to product of the Laplace transform of each function.

Thus, the RHS becomes

$$\begin{aligned}\int_0^\infty e^{-2zt}R'(t)dt &= \int_0^\infty e^{-2zt}\left(2c\left(\mathbb{E}[e^{2\sigma t}] + \beta^{-1}\int_0^t R(r)\mathbb{E}[e^{2\sigma(t-r)}]dr\right)\right)dt\\ &= c\mathbb{E}\left[\frac{1}{z-\sigma}\right] + 2c\beta^{-1}\mathbb{E}\left[\int_0^\infty e^{-2t(z-\sigma)}dt\right]\mathscr{L}_g(z)\\ &= c\mathbb{E}\left[\frac{1}{z-\sigma}\right] + c\beta^{-1}\mathbb{E}\left[\frac{1}{z-\sigma}\right]\mathscr{L}_g(z)\end{aligned}$$

Hence, combining the Laplace transforms of the left and right sides, we obtain

$$2z\mathscr{L}_R(z) = 1 + cm(z) + c\beta^{-1}m(z)\mathscr{L}_R(z).$$

□

We now turn to the proof of Theorem 3.1.5.

Proof of Theorem 3.1.5. (i) We start by considering the case where $\beta > \beta_c$.

Using Lemma 3.2.4, we obtain an asymptotic limit for $R(t)$ as:

$$R(t) \sim_{t\to\infty} C_\beta t^{-3/2}e^{4t},$$

where $C_\beta = \frac{c\beta(4\beta+1)}{(4\beta-c)^2}$.

Combining the asymptotic limit for $R(t)$ and Lemma 3.2.2, we notice that the limit of the first term of $H(t)$ defined as in (3.15) is:

$$\lim_{t\to\infty} R(t)^{-1}\mathbb{E}[\sigma e^{2\sigma t}] = \lim_{t\to\infty} \frac{\frac{\mathbb{E}[\sigma e^{2\sigma t}]}{(2t)^{-3/2}e^{4t}}(2t)^{-3/2}e^{4t}}{\frac{R(t)}{C_\beta t^{-3/2}e^{4t}}C_\beta t^{-3/2}e^{4t}} = \frac{2^{-5/2}}{\sqrt{\pi}C_\beta}. \tag{3.37}$$

Next, we multiply the integral in equation (3.15) by the asymptotic limit of $R(t)$ and split it into three parts: for $t\to\infty, x\to\infty$, and $x/t\to 0$,

$$\begin{aligned}
C_\beta^{-1}t^{\frac{3}{2}}e^{-4t}\int_0^t R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr &= C_\beta^{-1}t^{\frac{3}{2}}e^{-4t}\bigg(\int_0^x R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr \\
&\quad + \int_x^{t-x} R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr + \int_{t-x}^t R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr\bigg) \\
&= C_\beta^{-1}t^{\frac{3}{2}}e^{-4t}\bigg(\int_0^x R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr \\
&\quad + \int_x^{t-x} R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr + \int_0^x R(t-r)\mathbb{E}[\sigma e^{2\sigma(r)}]dr\bigg) \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

We now estimate each term separately. For I_1, we have:

$$\begin{aligned}
I_1 &= C_\beta^{-1}t^{\frac{3}{2}}e^{-4t}\int_0^x R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr \\
&= C_\beta^{-1}t^{\frac{3}{2}}e^{-4t}\int_0^x R(r)\frac{\mathbb{E}[\sigma e^{2\sigma(t-r)}]}{(2(t-r))^{-\frac{3}{2}}e^{4(t-r)}}(2(t-r))^{-\frac{3}{2}}e^{4(t-r)}dr \\
&= 2^{-\frac{3}{2}}C_\beta^{-1}\int_0^x R(r)e^{-4r}\left(\frac{1}{2\sqrt{\pi}}+o(1)\right)\left(\frac{t}{t-r}\right)^{\frac{3}{2}}dr \\
&= 2^{-\frac{3}{2}}C_\beta^{-1}\int_0^x R(r)e^{-4r}\left(\frac{1}{2\sqrt{\pi}}+o(1)\right)(1+o(1))dr \\
&= 2^{-\frac{5}{2}}C_\beta^{-1}\frac{1}{\sqrt{\pi}}\int_0^x R(r)e^{-4r}dr
\end{aligned}$$

For I_2, we can show that it is of smaller order than I_1 and I_3 and can be neglected.

Indeed, we have

$$\begin{aligned}
I_2 &= C_\beta^{-1} t^{\frac{3}{2}} e^{-4t} \int_x^{t-x} R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr \\
&= C_\beta^{-1} t^{\frac{3}{2}} e^{-4t} \int_x^{t-x} \frac{R(r)}{C_\beta r^{-\frac{3}{2}} e^{4r}} C_\beta r^{-\frac{3}{2}} e^{4r} \frac{\mathbb{E}[\sigma e^{2\sigma(t-r)}]}{(2(t-r))^{-\frac{3}{2}} e^{4(t-r)}} (2(t-r))^{-\frac{3}{2}} e^{4(t-r)} dr \\
&= C_\beta^{-1} t^{\frac{3}{2}} e^{-4t} \int_x^{t-x} C_\beta r^{-\frac{3}{2}} e^{4r} \left(\frac{1}{2\sqrt{\pi}} + o(1)\right) (2(t-r))^{-\frac{3}{2}} e^{4(t-r)} dr \\
&= 2^{-\frac{5}{2}} \frac{1}{\sqrt{\pi}} \int_x^{t-x} \left(\frac{t}{r(t-r)}\right)^{\frac{3}{2}} dr \\
&= o(1)
\end{aligned}$$

Finally, for I_3, we have:

$$I_3 = \int_0^x \mathbb{E}[\sigma e^{2\sigma r}] e^{-4r} dr.$$

Note that we have

$$\int_0^\infty |R(r)| e^{-4r} dr < \infty \text{ and } \int_0^\infty \mathbb{E}[\sigma e^{2\sigma r}] e^{-4r} dr < \infty.$$

Then we have

$$\lim_{t\to\infty} H(t) = \frac{2^{-\frac{5}{2}}}{C_\beta\sqrt{\pi}} + \frac{2^{-\frac{5}{2}}\beta^{-1}}{C_\beta\sqrt{\pi}} \int_0^\infty R(r) e^{-4r} dr + \beta^{-1} \int_0^\infty \mathbb{E}[\sigma e^{2\sigma r}] e^{-4r} dr. \tag{3.38}$$

By Lemma 3.2.5, we have

$$\int_0^\infty e^{-4r} R(r) dr = \mathscr{L}_R(2) = \frac{\beta(1+c)}{4\beta - c} \tag{3.39}$$

Also, note that

$$\begin{aligned}
\int_0^\infty \mathbb{E}[\sigma e^{2\sigma r}] e^{-4r} dr &= \mathbb{E}\left[\sigma \int_0^\infty e^{-2(2-\sigma)r} dr\right] \\
&= \mathbb{E}\left[\frac{\sigma}{2(2-\sigma)}\right] \\
&= \frac{1}{2}\left(-1 + \mathbb{E}\left[\frac{2}{2-\sigma}\right]\right) = \frac{1}{2}.
\end{aligned} \tag{3.40}$$

Hence, plug (3.39) and (3.40) into (3.38) we have

$$\lim_{t\to\infty} H(t) = \frac{2^{-\frac{5}{2}}}{C_\beta\sqrt{\pi}} + \left(\frac{2^{-\frac{5}{2}}}{C_\beta\sqrt{\pi}}\right)\left(\frac{1+c}{4\beta-c}\right) + \frac{1}{2}\beta^{-1} = \frac{2^{-\frac{5}{2}}(4\beta-c)}{\sqrt{\pi}c\beta} + \frac{1}{2}\beta^{-1}. \tag{3.41}$$

(ii) As $\beta = \beta_c$, by Lemma 3.2.4, we have

$$R(t) \sim_{t\uparrow\infty} C_\beta t^{-1/2} e^{4t}$$

where $C_\beta = \frac{\beta(c+1)}{c}$.

By Lemma 3.2.2 and Lemma 3.2.4, the first term $R^{-1}(t)\mathbb{E}[\sigma e^{2\sigma t}]$ in $H(t)$ converges to 0 as $t \to \infty$.

Note that we have for $x \to \infty, t \to \infty, x/t \to 0$

$$\begin{aligned} C_\beta^{-1} t^{\frac{1}{2}} e^{-4t} \int_0^t R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr = C_\beta^{-1} t^{\frac{1}{2}} e^{-4t} \Bigg(\int_0^x R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr \\ + \int_x^{t-x} R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr + \int_0^x R(t-r)\mathbb{E}[\sigma e^{2\sigma r}]dr \Bigg) \\ =: E_1 + E_2 + E_3 \end{aligned}$$

Then we have

$$E_1 = 2^{-\frac{3}{2}} C_\beta^{-1} \int_0^x R(r) e^{-4r} \frac{t^{\frac{1}{2}}}{(t-r)^{\frac{3}{2}}} \left(\frac{1}{2\sqrt{\pi}} + o(1)\right) dr = o(1)$$

where it follows from $\frac{t^{\frac{1}{2}}}{(t-r)^{\frac{3}{2}}} = \frac{1}{t}(1+o(1)) = o(1)$.

Similarly, we have $E_3 = o(1)$.

Also, we have

$$E_2 = \frac{2^{-\frac{3}{2}}}{\sqrt{\pi}} \int_x^{t-x} \left(\frac{t}{r}\right)^{\frac{1}{2}} \left(\frac{1}{t-r}\right)^{\frac{3}{2}} dr = O(x^{-\frac{1}{2}}).$$

Hence, we have

$$\lim_{t\to\infty} H(t) = 0. \tag{3.42}$$

(iii) In the case of $\beta < \beta_c$:

By Lemma 3.2.4, we have

$$R(t) \sim_{t\uparrow\infty} C_\beta e^{2s_\beta t},$$

where $C_\beta = \frac{\beta(cm(s_\beta)+1)}{2\beta - cm'(s_\beta)}$.

By Lemma 3.2.2 and Lemma 3.2.4, the first term $R^{-1}(t)\mathbb{E}[\sigma e^{2\sigma t}]$ in $H(t)$ converges to 0 as $t \to \infty$.

Note that we have for $x \to \infty, t \to \infty, x/t \to 0$

$$\begin{aligned} C_\beta^{-1} e^{-2s_\beta t} \int_0^t R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr = C_\beta^{-1} e^{-2s_\beta t} \Big(\int_0^x R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr \\ + \int_x^{t-x} R(r)\mathbb{E}[\sigma e^{2\sigma(t-r)}]dr + \int_0^x R(t-r)\mathbb{E}[\sigma e^{2\sigma r}]dr \Big) \\ =: F_1 + F_2 + F_3 \end{aligned}$$

Then we have

$$F_1 = 2^{-\frac{3}{2}} C_\beta^{-1} \int_0^x R(r) e^{-4r} \frac{1}{(t-r)^{\frac{3}{2}}} e^{-2(s_\beta - 2)t} \left(\frac{1}{2\sqrt{\pi}} + o(1) \right) dr = o(1)$$

where it follows from $\frac{t^{\frac{1}{2}}}{(t-r)^{\frac{3}{2}}} = \frac{1}{t}(1+o(1)) = o(1)$ and $s_\beta > 2$.

Similarly, we have $F_3 = o(1)$.

Also, we have

$$F_2 = \frac{2^{-\frac{3}{2}}}{\sqrt{\pi}} \int_x^{t-x} e^{-2(s_\beta-2)(t-r)} \left(\frac{1}{t-r} \right)^{\frac{3}{2}} dr = O(x^{-\frac{1}{2}}).$$

Hence, we have

$$\lim_{t\to\infty} H(t) = 0. \tag{3.43}$$

□

3.3 Proof of Theorem 3.1.4

Recall that f' is non-negative and Lipschitz as defined in (3.2). Recall that $K = K(t,t)$ is defined in (3.4). Define $R_\tau^\theta(K) := e^{-\int_\tau^\theta f'(K(s))ds}$ and

$$DR_\tau^\theta(K) = \frac{d}{d\tau} R_\tau^\theta(K) = f'(K(\tau,\tau)) e^{-\int_\tau^\theta f'(K(s))ds} \tag{3.44}$$

We have the following bound on $R_\tau^\theta(K)$ and $DR_\tau^\theta(K)$.

Lemma 3.3.1. *[ADG01, Theorem 5.3] Recall that we define f', $R_\tau^\theta(K)$, and $DR_\tau^\theta(K)$ as above. Then we have*

1. *for any $0 \le \tau \le \theta \le T$ and $K \in \mathbf{C}([0,T]^2)$, we have*

$$0 \le R_\tau^\theta(K) \le 1, \text{ and } \int_0^t |R_\tau^\theta(K)| d\tau \le 1. \tag{3.45}$$

2. *for every $\theta \le T$, we have*

$$\sup_{\tau \le \theta} |R_\tau^\theta(K) - R_\tau^\theta(\widetilde{K})| \le \|f'\|_{\mathscr{L}} \int_0^\theta |K(s,s) - \widetilde{K}(s,s)| ds, \tag{3.46}$$

   where $\|f'\|_{\mathscr{L}}$ is the Lipschitz norm of f'.

3. *for any $0 \le \tau \le \theta \le T$,*

$$|DR_\tau^\theta(K) - DR_\tau^\theta(\widetilde{K})| \le \|f'\|_{\mathscr{L}} \left\{ |K(\tau,\tau) - \widetilde{K}(\tau,\tau)| + \left(DR_\tau^\theta(K) + DR_\tau^\theta(\widetilde{K}) \right) \int_0^\theta |K(s,s) - \widetilde{K}(s,s)| ds \right\}.$$

Consider the following collections of functions with domain space $\mathbb{R}^2 \times \mathbf{C}([0,T])$ for $T > 0$ and range space one of $\mathbf{C}([0,T]^j)$ for $j = 1,2,3$:

$$\mathscr{G}_1 := \{g_j, j = 1,2,3 : g_1(Y_0, \sigma, B_\cdot)(t) = \sigma e^{\sigma t}(Y_0)^2, g_2(\cdot)(t) = \sigma B_t^2, g_3(\cdot)(t) = \sigma Y_0 B_t\}.$$

$$\mathscr{G}_2 := \{g_j, j = 4,5 : g_4(Y_0, \sigma, B_\cdot)(s,t) = \sigma Y_0 B_s e^{\sigma t}, g_5(\cdot)(s,t) = \sigma^2 Y_0 B_s e^{\sigma t}\}.$$

$$\mathscr{G}_3 := \{g_j, j = 6,7 : g_6(Y_0, \sigma, B_\cdot)(u,v,t) = \sigma B_u B_v e^{\sigma t}, g_7(\cdot)(u,v,t) = \sigma^2 B_u B_v e^{\sigma t}\}.$$

Then our collection of functions is

$$\mathscr{G} = \mathscr{G}_1 \cup \mathscr{G}_2 \cup \mathscr{G}_3. \tag{3.47}$$

Define the empirical measure

$$\nu_T^N := \frac{1}{N}\sum_{i=1}^N \delta_{Y_0^i, \sigma^i, B^i_{[0,T]}}. \tag{3.48}$$

Define for $g \in \mathscr{G}$

$$\mathscr{C}_N := \int g(Y_0, \sigma, B_\cdot) d\nu_T^N(Y_0, \sigma, B_\cdot) = \frac{1}{N}\sum_{i=1}^N g(Y_0^i, \sigma^i, B_\cdot^i), \tag{3.49}$$

where note that for $g \in \mathscr{G}_j$, $\int g d\nu_T^N \in \mathbf{C}([0,T]^j)$ for $j = 1,2,3$.

Proof of Theorem 3.1.4. 1. Existence and uniqueness of the limit

Apply Ito's formula described as in Theorem 2.2.2, we have

$$Y_t^i = e^{\int_0^t (\sigma^i - f'(K_N(r)))dr} Y_0^i + \beta^{-1/2} \int_0^t e^{\int_s^t (\sigma^i - f'(K_N(r)))dr} dB_s^i. \tag{3.50}$$

Define $F_t(K,\sigma) = f'((K(t,t)) - \sigma$. By Lemma 2.2.3, we have

$$Y_t^i = e^{-\int_0^t F_r(K_N,\sigma^i)dr} Y_0^i + \beta^{-1/2} B_t^i + \left(-\beta^{-1/2} \int_0^t B_s^i F_s(K_N,\sigma^i) e^{-\int_s^t F_r(K_N,\sigma^i)dr} ds \right). \tag{3.51}$$

Then we have

$$Y_t^i = \underbrace{R_0^t(K_N) e^{\sigma^i t} Y_0^i}_{=:T_1^i(t)} + \underbrace{\beta^{-1/2} B_t^i}_{=:T_2^i(t)} + \underbrace{\left(-\beta^{-1/2} \int_0^t B_s^i e^{\sigma^i (t-s)} (DR_s^t(K_N) - \sigma^i R_s^t(K_N)) ds \right)}_{=:T_3^i(t)}. \tag{3.52}$$

We denote the sum of three terms as $T_1^i(t)$, $T_2^i(t)$, and $T_3^i(t)$, respectively. In this case,

$$Y_t^i = T_1^i(t) + T_2^i(t) + T_3^i(t). \tag{3.53}$$

Then the energy $H_N(t)$ becomes

$$H_N(t) = \frac{1}{N}\sum_{i=1}^N \sigma^i \left(\sum_{j=1}^3 (T_j^i(t))^2 + 2\sum_{j \neq k} T_j^i(t) T_k^i(t) \right), \tag{3.54}$$

Plug into the expression of $T_1^i(t), T_2^i(t), T_3^i(t)$

$$
\begin{aligned}
H_N(t) &= \frac{1}{N}\sum_{i=1}^N \sigma^i (R_0^t(K_N))^2 e^{2\sigma^i t}(Y_0^i)^2 + \frac{1}{N}\sum_{i=1}^N \sigma^i \beta^{-1}(B_t^i)^2 \\
&+ \frac{\beta^{-1}}{N}\sum_{i=1}^N \sigma^i \int_0^t\int_0^t B_u^i B_v^i e^{\sigma^i(2t-u-v)}\left(DR_u^t(K_N) - \sigma^i R_v^t(K_N)\right)\left(DR_v^t(K_N) - \sigma^i R_v^t(K_N)\right) du dv \\
&+ \frac{2}{N}\sum_{i=1}^N \sigma^i \Big\{\beta^{-1/2} Y_0^i B_t^i R_0^t(K_N) - \beta^{-1/2}\int_0^t Y_0^i B_s^i e^{\sigma^i(2t-s)} R_0^t(K_N)\left(DR_s^t(K_N) - \sigma^i R_s^t(K_N)\right) ds \\
&- \beta^{-1}\int_0^t B_t^i B_s^i e^{\sigma^i(t-s)}\left(DR_s^t(K_N) - \sigma^i R_s^t(K_N)\right) ds\Big\}
\end{aligned}
$$

Hence, the above equation of $H_N(t)$ specifies the function

$$
\Phi : \mathbf{C}([0,T]^2) \times \mathbf{C}([0,T])^{|\mathscr{G}_1|} \times \mathbf{C}([0,T]^2)^{|\mathscr{G}_2|} \times \mathbf{C}([0,T]^3)^{|\mathscr{G}_3|} \to \mathbf{C}([0,T])
$$

such that

$$
H_N = \Phi(K_N, \mathscr{C}_N).
$$

By Lemma 3.3.1, for any $\mathscr{C}_N$ defined as in (3.49) and $K_N, \widetilde{K}_N \in \mathbf{C}([0,T]^2)$, there exists a constant $C_1 > 0$ so that

$$
\sup_{0\le t\le T} |\Phi(K_N, \mathscr{C}_N) - \Phi(\widetilde{K}_N, \mathscr{C}_N)| \le C_1 \int_0^t |K_N(s,s) - \widetilde{K}_N(s,s)| ds \tag{3.55}
$$

Similarly, we apply Lemma 3.3.1, then for any $\mathscr{C}_N, \widetilde{\mathscr{C}}_N$ defined as in (3.49) and $\widetilde{K}_N \in \mathbf{C}([0,T]^2)$, there exists a constant $C_2 > 0$ so that

$$
\sup_{0\le t\le T} |\Phi(\widetilde{K}_N, \mathscr{C}_N) - \Phi(\widetilde{K}_N, \widetilde{\mathscr{C}}_N)| \le C_2 \|\mathscr{C}_N - \widetilde{\mathscr{C}}_N\|_\infty \tag{3.56}
$$

By Theorem 3.1.1, we know that K_N converges to the deterministic limit K almost surely and K is unique. By [LSS22, Lemma 3.7], each element in $\mathscr{C}_N$ converges to deterministic limits $\mathscr{C}$ almost surely under the i.i.d. initial conditions.

Combining (3.55) and (3.56), then we have

$$
\begin{aligned}
\|\Phi(K_N, \mathscr{C}_N) - \Phi(K, \mathscr{C})\|_\infty &= \|\Phi(K_N, \mathscr{C}_N) - \Phi(K_N, \mathscr{C}) + \Phi(K_N, \mathscr{C}) - \Phi(K, \mathscr{C})\|_\infty \\
&\le \|\Phi(K_N, \mathscr{C}_N) - \Phi(K_N, \mathscr{C})\|_\infty + \|\Phi(K_N, \mathscr{C}) - \Phi(K, \mathscr{C})\|_\infty \\
&\le C_1\|K_N - K\|_\infty + C_2\|\mathscr{C}_N - \mathscr{C}\|_\infty \to 0, \text{ as } N \to \infty.
\end{aligned}
$$

Hence, H_N converges to deterministic function H.

Also, we have $H = \Phi(K, \mathscr{C})$. Indeed,

$$|H - \Phi(K,\mathscr{C})| \le |H - H_N| + |H_N - \Phi(K,\mathscr{C})| \to 0, \text{ as } N \to \infty.$$

2. Equations for the limit points

   Next, we characterize the limit H as follows. Recall that $R_\tau^\theta(K) = e^{-\int_\tau^\theta f'(K(s))ds}$. Then Y_t can be expressed as

$$Y_t^i = R_0^t(K_N)e^{\sigma^i t}Y_0^i + \beta^{-1/2}\int_0^t R_s^t(K_N)e^{\sigma^i(t-s)}dB_s^i.$$

   Substitute the above expression of Y_t^i to $H_N(t)$ defined as in (3.5), then we have

$$H_N(t) = \frac{1}{N}\sum_{i=1}^N \sigma^i e^{2\sigma^i t}(Y_0^i)^2(R_0^t(K_N))^2 + \frac{\beta^{-1}}{N}\sum_{i=1}^N \sigma^i\left(\int_0^t R_s^t(K_N)e^{\sigma^i(t-s)}dB_s^i\right)^2 + \frac{2\beta^{-1/2}}{N}\sum_{i=1}^N \sigma^i\left(R_0^t(K_N)e^{\sigma^i t}Y_0^i\right)\int_0^t R_s^t(K_N)e^{\sigma^i(t-s)}dB_s^i.$$

   As $N \to \infty$, note that the first term is convergence to its expectation as in (3.6) by strong law of large number (SLLN) [Dur19, Theorem 2.4.1], the limit of the second term given in (3.7) is obtained by SLLN and Itô isometry [Øks03, Lemma 3.1.5], and the last term is convergence to zero by SLLN. Note that we take the limit that requires the empirical measure $\nu_T^N := \frac{1}{N}\sum_{i=1}^N \delta_{Y_0^i,\sigma^i,B^i_{[0,T]}}$ converges to the desired limits under the i.i.d. initial conditions. Hence, we obtain the desired integro-differential equation as in Theorem 3.1.4.

□

Chapter 4

The complexity of the gradient descent

4.1 The hitting time of gradient descent

We consider the normalized $N \times N$ symmetric Wigner matrix $\mathbf{J} = \{J_{ij}\}_{1 \le i,j \le N}$ defined as in Definition 2.1.1 (e.g., GOE and GUE are two special cases of Wigner matrix). We denote N eigenvalues of $\mathbf{J}$ in increasing order as

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_N.$$

We define $v_1, v_2, \ldots, v_N$ be an orthonormal basis of eigenvectors of $\mathbf{J}$ so that $\mathbf{J}v_i = \lambda_i v_i$ for $i = 1, \ldots, N$.

It is clear that if the Wigner matrix $\mathbf{J}$ obeys condition **C1** with constant $C_0 = 4$ as in Definition 2.1.5, then the criterion 2.8 holds. Our main result needs to ensure that the Tracy-Widom law holds, so we only need to ensure that the Wigner matrix obeys condition **C1** with constant $C_0 = 4$ by Theorem 2.1.6.

Next, we consider the dynamics of the spin glasses model on the sphere. Recall that the SK model defined on the unit sphere is described by the Hamiltonian

$$H_{\mathbf{J}}(X) = X^T \mathbf{J} X,$$

where $X = (X_1, \ldots, X_N)$ with $\|X\|_2 = 1$ and $\mathbf{J}$ is the normalized symmetric Wigner matrix.

The zero-temperature dynamics (a.k.a., gradient descent algorithm) of the SK model on the unit sphere is defined as follows:

$$dX_t = -\nabla_{\mathbb{S}^{N-1}} H_{\mathbf{J}}(X_t)dt, \tag{4.1}$$

where the initial data $X_0 = \{X_0^i\}_{1\le i\le N}$ is uniformly distributed on the unit sphere $\mathbb{S}^{N-1}$, and $\nabla_{\mathbb{S}^{N-1}}$ is the gradient on the unit sphere $\mathbb{S}^{N-1}$ and

$$\nabla_{\mathbb{S}^{N-1}} f(x) := \nabla f(x) - (\nabla f(x)\cdot x)x,\ x \in \mathbb{R}^N$$

for smooth functions f.

Similarly, the Langevin dynamics of the SK model on the sphere are defined by the following SDE.

$$dX_t = -\nabla_{\mathbb{S}^{N-1}} H_{\mathbf{J}}(X_t)dt + \beta^{-1/2} dB_t, \tag{4.2}$$

where the initial condition X_0 is uniformly distributed on the unit sphere $\mathbb{S}^{N-1}$, B_t is the Brownian motion on the unit sphere $\mathbb{S}^{N-1}$, and β is the inverse temperature. Brownian motion on the unit sphere is a solution of the integral equation [Hsu02, Example 3.3.2].

In this chapter, we mainly study the algorithm complexity of using the gradient descent algorithm to find the extreme eigenvalues of the Wigner matrix. This is related to our previous main results about the asymptotic limit of energy in Chapter 3 for the description of the time to the equilibrium state of the SK model.

Recall that we randomly start with an initial condition X_0 uniformly distributed on the unit sphere $\mathbb{S}^{N-1}$, and consider the zero-temperature dynamics defined in (4.1).

Fix $\varepsilon \in (0,1)$. Denote by T_ε the hitting time of the overlap between the output X_t of the gradient descent and eigenvector v_1 corresponding to the smallest eigenvalue of $\mathbf{J}$ is greater than ε, that is

$$T_\varepsilon := \inf_{t>0}\{|v_1 \cdot X_t| \geqslant \varepsilon\}.$$

Our main result is the lower bound and upper bound of the hitting time T_ε as follows.

Theorem 4.1.1. *Assume that the normalized $N\times N$ symmetric Wigner matrix $\mathbf{J}$ obeys condition **C1** with constant $C_0 = 4$. Consider the gradient descent described in (4.1) with the same setting as before. Let the hitting time T_ε be defined as before. For every $\delta > 0$, there exist constants $C_1 = C_1(\delta) > 0$, $C_2 = C_2(\delta) > 0$, and $C_3 = C_3(\varepsilon,\delta) > 0$ such that*

$$\lim_{N\to\infty} \mathbb{P}\left(C_1 N^{2/3} < T_\varepsilon < C_2 N^{2/3}\log(C_3 N)\right) > 1-\delta.$$

The upper bound in Theorem 4.1.1 is also proved in [Bra22].

4.2 Proof of Theorem 4.1.1

Recall that $\lambda_1 \le \cdots \le \lambda_N$ are N eigenvalues of the Wigner matrix $\mathbf{J}$ in increasing order. Recall that v_i is the orthonormal basis of eigenvectors of $\mathbf{J}$ corresponding to the i-th smallest eigenvalue λ_i for $i = 1, \ldots, N$. Define the overlap of the output X_t and eigenvectors v_i of $\mathbf{J}$ by $h_i(t) := v_i \cdot X_t$ for $i = 1, \ldots, N$. Note that we can solve $h_i(t)$ for $i = 1, \ldots, N$ as follows.

Lemma 4.2.1. *Assume that the same setting holds as in Theorem 4.1.1. Then we have*

$$|h_j(t)| = \frac{|h_j(0)|e^{-2\lambda_j t}}{\sqrt{\sum_{i=1}^N h_i^2(0)e^{-4\lambda_i t}}}. \tag{4.3}$$

Proof. Note that the gradient descent (4.1) can be simplified as follows

$$dX_t = -(\nabla H_{\mathbf{J}}(X_t) - (\nabla H_{\mathbf{J}}(X_t) \cdot X_t)X_t)\,dt = -2\mathbf{J}X_t dt + 2H_{\mathbf{J}}(X_t)X_t dt. \tag{4.4}$$

By spectral decomposition we have

$$\mathbf{J} = \sum_i v_i v_i^T \lambda_i,$$

where $v_1, \ldots, v_N$ are eigenvectors corresponding to eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_N$ of $\mathbf{J}$.

Note that

$$H_{\mathbf{J}}(X_t) = X_t \cdot \mathbf{J}X_t = \sum_i \lambda_i (h_i(t))^2.$$

Consider the dot product v_1 on the both sides of (4.4) and substitute the above equation of $H_{\mathbf{J}}(X_t)$ to get

$$\begin{aligned}\frac{1}{2}h_1'(t) &= h_1(t)H_{\mathbf{J}}(X_t) - v_1 \cdot (\sum_i v_i v_i^T \lambda_i)X_t \\ &= h(t)H_{\mathbf{J}}(X_t) - \lambda_1 v_1 \cdot X_t \\ &= h_1(t)H_{\mathbf{J}}(X_t) - \lambda_1 h_1(t).\end{aligned}$$

By the fact that $\sum_i h_i(t)^2 = 1$ we have

$$\frac{1}{2}h_1'(t) = (H_{\mathbf{J}}(X_t) - \lambda_1)h(t) = \sum_{i=1}^N [(\lambda_i - \lambda_1)h_i^2]h_1(t). \tag{4.5}$$

where we write $h_i(t) = h_i$ for convenience, $i = 1, \ldots, N$.

Similarly, we have

$$\frac{1}{2}h'_j(t) = \sum_{i=1}^{N}[(\lambda_i - \lambda_j)h_i^2]h_j(t). \tag{4.6}$$

Multiply $h_j(t)$ on the both side of (4.6) yields

$$(h_j^2(t))' = 4\sum_{i=1}^{N}\left((\lambda_i - \lambda_j)h_i^2\right)h_j^2$$

Denote $f(t) = \sum_i \lambda_i h_i^2(t)$ and $F(t) = \int_0^t f(s)ds$.

Both sides of the equation are divided by h_j^2 yields for$j = 1, \ldots, N$

$$(\log h_j^2(t))' = 4f(t) - 4\lambda_j$$

Integrating the two sides with respect to time t yields

$$h_j^2(t) = h_j(0)^2 \exp\left(4F(t) - 4\lambda_j t\right) \tag{4.7}$$

Taking the derivative with respect to both sides of $F(t)$ and substitute (4.7) yields

$$F'(t) = \sum_i \lambda_i h_i^2(t) = e^{4F(t)}\left(\sum_i \lambda_i h_i^2(0)e^{-4\lambda_t t}\right).$$

Both sides are divided by the integration factor $e^{4F(t)}$ and integrating with respect to t to obtain

$$e^{-4F(t)} - 1 = \sum h_i^2(0)(e^{-4\lambda_i t} - 1). \tag{4.8}$$

So we get

$$F(t) = -\frac{1}{4}\log\left(\sum_i h_i^2(0)e^{-4\lambda_i t}\right). \tag{4.9}$$

Substituting (4.9) into equation (4.7) yields for $j = 1, 2, \ldots, N$

$$|h_j(t)| = \frac{|h_j(0)|e^{-2\lambda_j t}}{\sqrt{\sum_{i=1}^{N} h_i^2(0)e^{-4\lambda_i t}}}. \tag{4.10}$$

□

To prove Theorem 4.1.1, we need the following Lemma.

Lemma 4.2.2. *Consider a sequence of i.i.d. positive random variables* $X_1,\dots,X_k$. *For every constant* $C>0$, *we have*

$$\mathbb{P}\left(\frac{X_1+X_2+\cdots+X_k}{X_j}>C\right)>1-\frac{C}{k},\ \textit{for}\ j=1,\dots,k.$$

Proof. Note that

$$0<\mathbb{E}\left[\frac{X_1}{\sum_{i=1}^k X_i}\right]=\mathbb{E}\left[\frac{X_2}{\sum_{i=1}^k X_i}\right]=\cdots=\mathbb{E}\left[\frac{X_k}{\sum_{i=1}^k X_i}\right]<1.$$

Then we have

$$\mathbb{E}\left[\frac{X_1}{\sum_{i=1}^k X_i}\right]=\frac{1}{k}.$$

By Markov's inequality, we get for every constant $C>0$

$$\mathbb{P}\left(\frac{X_1+X_2+\cdots+X_k}{X_j}\le C\right)=\mathbb{P}\left(\frac{X_j}{X_1+X_2+\cdots+X_k}\geqslant 1/C\right)\le C\cdot\mathbb{E}\left[\frac{X_j}{\sum_i X_i}\right]=\frac{C}{k}. \tag{4.11}$$

□

Lemma 4.2.3. *[Lan22, Lemma 2.3] Let* **J** *be the normalized symmetric Wigner matrix. Assume that* **J** *obeys condition* ***C1*** *with constant* $C_0=4$. *Let* $\lambda_1<\lambda_2$ *be the two smallest eigenvalues of* **J**. *Then for every* $\varepsilon>0$, *there exists* $\delta>0$ *so that*

$$\mathbb{P}\left(N^{2/3}(\lambda_2-\lambda_1)\geqslant\delta\right)\geqslant 1-\varepsilon \tag{4.12}$$

for all N *large enough. This result also holds for the two largest eigenvalue* $\lambda_{N-1}<\lambda_N$.

This result follows from combining Theorem 2.1.6 and continuous mapping theorem in [Dur19, Theorem 3.2.10]. Moreover, we have $\lambda_2-\lambda_1=O_p(N^{-2/3})$.

We require the following Lemma.

Lemma 4.2.4. *Let* X *be a standard normal random variable. Then for every* $\epsilon>0$, *there exists a constant* $\delta>0$ *so that*

$$\mathbb{P}(|X|>\delta)\geqslant 1-\epsilon. \tag{4.13}$$

Proof. For every $\epsilon > 0$, there exists a sufficiently small constant $\delta \in \left(0, \sqrt{\frac{\pi}{2}}\epsilon\right)$ so that

$$\mathbb{P}(|X| \leq \delta) = \frac{1}{\sqrt{2\pi}} \int_{-\delta}^{\delta} e^{-x^2/2} dx \leq \frac{2}{\sqrt{2\pi}} \delta < \epsilon, \tag{4.14}$$

where the above inequality follows from the fact that $e^{-x^2/2} \leq 1$ for $x \in \mathbb{R}$. □

Armed with the previous results, we then can prove our main result.

The proof of Theorem 4.1.1. By Lemma 4.2.1, we have

$$|h_1(t)| = \frac{|h_1(0)| e^{-2\lambda_1 t}}{\sqrt{\sum_i h_i^2(0) e^{-4\lambda_i t}}} = \frac{|h_1(0)|}{\sqrt{h_1^2(0) + \sum_{i=2}^N h_i^2(0) e^{-4(\lambda_i - \lambda_1)t}}} \tag{4.15}$$

Next, we consider the upper and lower bound of the hitting time T_ε, respectively.

1. Lower bound of T_ε.

   For any $\delta > 0$, we fix the first k terms of the denominator of (4.15) and then we will find desired k below depending on ϵ and δ (independent of N):

$$|h_1(t)| \leq \frac{|h_1(0)|}{\sqrt{h_1^2(0) + \sum_{i=2}^k h_i^2(0) e^{-4t(\lambda_i - \lambda_1)}}}. \tag{4.16}$$

   Note that $(\lambda_i - \lambda_1) \leq (\lambda_k - \lambda_1)$ for $i = 1, \ldots, k-1$. Then we upper bound the above inequality:

$$|h_1(t)| \leq \frac{|h_1(0)|}{\sqrt{h_1^2(0) + e^{-4t(\lambda_k - \lambda_1)} \sum_{i=2}^k h_i^2(0)}} \tag{4.17}$$

   For $t \geqslant T_\epsilon$, we have

$$\epsilon \leq |h_1(t)| \leq \frac{|h_1(0)|}{\sqrt{h_1^2(0) + e^{-4t(\lambda_k - \lambda_1)} \sum_{i=2}^k h_i^2(0)}} \tag{4.18}$$

   and we get

$$T_\epsilon \geqslant \frac{1}{4(\lambda_k - \lambda_1)} \log\left(\frac{h_2^2(0) + \cdots + h_k^2(0)}{h_1^2(0)(\epsilon^{-2} - 1)}\right) \tag{4.19}$$

For any $\delta > 0$, we apply Lemma 4.2.2 (choose $C = 2\varepsilon^{-2} - 1$):

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{h_2^2(0) + \cdots + h_k^2(0)}{h_1^2(0)(\epsilon^{-2} - 1)} > 2\right) = \lim_{N\to\infty} \mathbb{P}\left(\frac{h_1^2(0) + h_2^2(0) + \cdots + h_k^2(0)}{h_1^2(0)} > \frac{2}{\epsilon^2} - 1\right) \quad (4.20)$$

$$\geqslant 1 - \delta/2, \quad (4.21)$$

where we take $k = [2(2\epsilon^{-2} - 1)/\delta] + 1$.

By the similar argument of Lemma 4.2.3, for any $\delta > 0$ there exists a constant $c_1 = c_1(\delta) > 0$ so that

$$\lim_{N\to\infty} \mathbb{P}(N^{2/3}(\lambda_k - \lambda_1) < c_1) \geqslant 1 - \delta/2 \quad (4.22)$$

Hence, for any $\delta > 0$ there exists $c_2 = \frac{\log 2}{4c_1}$ so that

$$\lim_{N\to\infty} \mathbb{P}(T_\varepsilon > c_2 N^{2/3}) \geqslant 1 - \delta. \quad (4.23)$$

2. Upper bound of T_ε.

   Note that $(\lambda_2 - \lambda_1) \leq (\lambda_i - \lambda_1)$ for $i = 3, \ldots, N$. We upper bound each term of the denominator in (4.15) by $e^{-4t(\lambda_i - \lambda_1)} \leq e^{-4t(\lambda_2 - \lambda_1)}$ for $i = 3, \ldots, N$. Then we have

$$|h_1(t)| \geqslant \frac{|h_1(0)|}{\sqrt{h_1^2(0) + e^{-4(\lambda_2 - \lambda_1)t} \sum_{i=2}^N h_i^2(0)}} = \frac{|h_1(0)|}{\sqrt{h_1^2(0) + e^{-4(\lambda_2 - \lambda_1)t}(1 - h_1^2(0))}} \quad (4.24)$$

   For $t \leq T_\varepsilon$, we get

$$\varepsilon \geqslant |h_1(t)| \geqslant \frac{|h_1(0)|}{\sqrt{h_1^2(0) + e^{-4(\lambda_2 - \lambda_1)t}(1 - h_1^2(0))}}. \quad (4.25)$$

   and this yields

$$T_\varepsilon \leq \frac{1}{4(\lambda_2 - \lambda_1)} \log\left(\frac{h_1^{-2}(0) - 1}{\varepsilon^{-2} - 1}\right). \quad (4.26)$$

   By Lemma 4.2.3, for any $\delta > 0$ there exists a constant $c_3 = c_3(\delta) > 0$ so that

$$\lim_{N\to\infty} \mathbb{P}\left(N^{2/3}(\lambda_2 - \lambda_1) \geqslant c_3\right) \geqslant 1 - \delta/2 \quad (4.27)$$

Note that $\sqrt{N}h_1(0)$ is asymptotic Gaussian by Lemma 2.1.10. By Lemma 4.2.4, for every $\delta > 0$, there exists a constant $c_4 > 0$ so that

$$\lim_{N\to\infty} \mathbb{P}\left(\sqrt{N}|h_1(0)| > c_4\right) \geqslant 1 - \delta/2. \tag{4.28}$$

For any $\delta > 0$, we take $c_5 = \frac{1}{c_4^2(\varepsilon^{-2}-1)} > 0$ and then get

$$\mathbb{P}\left(\frac{h_1^{-2}(0)-1}{\varepsilon^{-2}-1} < c_5 N\right) = \mathbb{P}\left(|h_1(0)| > \sqrt{\frac{c_4}{c_4^2+N}}\right) = \mathbb{P}\left(\sqrt{1+\frac{c_4^2}{N}}\sqrt{N}|h_1(0)| > c_4\right)$$

By inequality (4.28), for every $\delta > 0$ we have

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{h_1^{-2}(0)-1}{\varepsilon^{-2}-1} < c_5 N\right) \geqslant 1 - \delta/2. \tag{4.29}$$

Combining the two upper bounds (4.27) and (4.29), for every $\delta > 0$ there exist constants c_5 defined as above and $c_6 = \frac{1}{4c_3}$ so that

$$\lim_{N\to\infty} \mathbb{P}\left(T_\varepsilon < c_6 N^{2/3}\log(c_5 N)\right) \geqslant 1 - \delta. \tag{4.30}$$

□

References

[AC18] Antonio Auffinger and Wei-Kuo Chen. On the energy landscape of spherical spin glasses. *Advances in Mathematics*, 330:553–588, 2018.

[ADG01] G Ben Arous, Amir Dembo, and Alice Guionnet. Aging of spherical spin glasses. *Probability theory and related fields*, 120(1):1–67, 2001.

[AEK+14] Bloemendal Alex, Laszlo Erdos, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Isotropic local laws for sample covariance and generalized wigner matrices. *Electronic Journal of Probability*, 19:1–53, 2014.

[AG97] G Ben Arous and Alice Guionnet. Symmetric langevin spin glass dynamics. *The Annals of Probability*, 25(3):1367–1422, 1997.

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Number 118. Cambridge university press, 2010.

[AS48] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office, 1948.

[BA03] Gérard Ben-Arous. Aging and spin-glass dynamics. *arXiv preprint math/0304364*, 2003.

[BBD08] Jinho Baik, Robert Buckingham, and Jeffery DiFranco. Asymptotics of tracy-widom distributions and the total integral of a painlevé ii function. *Communications in Mathematical Physics*, 280:463–497, 2008.

[BBP07] Giulio Biroli, J-P Bouchaud, and Marc Potters. On the top eigenvalue of heavy-tailed random matrices. *Europhysics Letters*, 78(1):10001, 2007.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[Bra22] A. Brandenberger. Complexity and dynamics of a spherical spin glass system. *Masters Thesis*, 2022.

[BY88] Zhi-Dong Bai and Yong-Qua Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix. *The Annals of Probability*, pages 1729–1741, 1988.

[CB21] George Casella and Roger L Berger. *Statistical inference*. Cengage Learning, 2021.

[Cha07] Sourav Chatterjee. Estimation in spin glasses: A first step. 2007.

[Chu01] Kai Lai Chung. *A course in probability theory*. Academic press, 2001.

[CK93] Leticia F Cugliandolo and Jorge Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Physical Review Letters*, 71(1):173, 1993.

[CK94] Leticia F Cugliandolo and Jorge Kurchan. On the out-of-equilibrium relaxation of the sherrington-kirkpatrick model. *Journal of Physics A: Mathematical and General*, 27(17):5749, 1994.

[CS92] Andrea Crisanti and H-J Sommers. The spherical p-spin interaction spin glass model: the statics. *Zeitschrift für Physik B Condensed Matter*, 87(3):341–354, 1992.

[DKM+99] Percy Deift, Thomas Kriecherbauer, K T-R McLaughlin, Stephanos Venakides, and Xin Zhou. Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 52(11):1335–1425, 1999.

[DMM09] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[DT17] Percy Deift and Thomas Trogdon. Universality for eigenvalue algorithms on sample covariance matrices. *SIAM Journal on Numerical Analysis*, 55(6):2835–2862, 2017.

[Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[Dys70] Freeman J Dyson. Correlations between eigenvalues of a random matrix. *Communications in Mathematical Physics*, 19:235–250, 1970.

[EA75] Samuel Frederick Edwards and Phil W Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975.

[ESY11] Laszlo Erdos, Benjamin Schlein, and Horng-Tzer Yau. Universality of random matrices and local relaxation flow. *Inventiones mathematicae*, 185(1):75–119, 2011.

[For93] Peter J Forrester. The spectrum edge of random matrix ensembles. *Nuclear Physics B*, 402(3):709–728, 1993.

[Geo11] Hans-Otto Georgii. *Gibbs measures and phase transitions*, volume 9. Walter de Gruyter, 2011.

[GHM01] Hans-Otto Georgii, Olle Häggström, and Christian Maes. The random geometry of equilibrium phases. In *Phase transitions and critical phenomena*, volume 18, pages 1–142. Elsevier, 2001.

[Gru96] Malte Grunwald. Sanov results for glauber spin-glass dynamics. *Probability theory and related fields*, 106(2):187–232, 1996.

[Gui07] Alice Guionnet. *Dynamics for Spherical Models of Spin-Glass and Aging*, pages 117–144. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[HØS00] Yaozhong Hu, Bernt Øksendal, and Agnes Sulem. Optimal portfolio in a fractional black & scholes market. In *Mathematical Physics and Stochastic Analysis: Essays in Honour of Ludwig Streit*, pages 267–279. World Scientific, 2000.

[Hsu02] Elton P Hsu. *Stochastic analysis on manifolds*. Number 38. American Mathematical Soc., 2002.

[Joh01] Kurt Johansson Johansson. Universality of the local spacing distribution in certain ensembles of hermitian wigner matrices. *Communications in Mathematical Physics*, 215:683–705, 2001.

[KMP01] Claire Kenyon, Elchanan Mossel, and Yuval Peres. Glauber dynamics on trees and hyperbolic graphs. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 568–578. IEEE, 2001.

[Kos88] Eric Kostlan. Complexity theory of numerical linear algebra. *Journal of Computational and Applied Mathematics*, 22(2-3):219–230, 1988.

[Kos91] Eric Kostlan. Statistical complexity of dominant eigenvector calculation. *Journal of Complexity*, 7(4):371–379, 1991.

[KTJ76] John M Kosterlitz, David J Thouless, and Raymund C Jones. Spherical model of a spin-glass. *Physical Review Letters*, 36(20):1217, 1976.

[Lan22] Benjamin Landon. Free energy fluctuations of the two-spin spherical sk model at critical temperature. *Journal of Mathematical Physics*, 63(3):033301, 2022.

[LG16] Jean-Francois Le Gall. *Brownian motion, martingales, and stochastic calculus*. Springer, 2016.

[LSS22] Tengyuan Liang, Subhabrata Sen, and Pragya Sur. High-dimensional asymptotics of langevin dynamics in spiked matrix models. *arXiv preprint arXiv:2204.04476*, 2022.

[LY14] Ji Oon Lee and Jun Yin. A necessary and sufficient condition for edge universality of wigner matrices. *Duke Mathematical Journal*, 163(1):117–173, 2014.

[MMS09] Ross A Maller, Gernot Müller, and Alex Szimayer. Ornstein–uhlenbeck processes and extensions. *Handbook of financial time series*, pages 421–437, 2009.

[Øks03] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.

[Pan13a] Dmitry Panchenko. The parisi ultrametricity conjecture. *Annals of Mathematics*, pages 383–393, 2013.

[Pan13b] Dmitry Panchenko. *The sherrington-kirkpatrick model*. Springer Science & Business Media, 2013.

[Par79] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.

[SK75] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26):1792, 1975.

[Sma97] Steve Smale. Complexity theory and numerical analysis. *Acta numerica*, 6:523–551, 1997.

[SN13] Daniel L Stein and Charles M Newman. *Spin glasses and complexity*, volume 4. Princeton University Press, 2013.

[Sos99] Alexander Soshnikov. Universality at the edge of the spectrum in wigner random matrices. *Communications in mathematical physics*, 207:697–733, 1999.

[SZ81] Haim Sompolinsky and Annette Zippelius. Dynamic theory of the spin-glass phase. *Physical Review Letters*, 47(5):359, 1981.

[SZ82] Haim Sompolinsky and Annette Zippelius. Relaxational dynamics of the edwards-anderson model and the mean-field theory of spin-glasses. *Physical Review B*, 25(11):6860, 1982.

[Tal06] Michel Talagrand. The parisi formula. *Annals of mathematics*, pages 221–263, 2006.

[Tao12] Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

[TV11] Terence Tao and Van Vu. Random matrices: Universality of local eigenvalue statistics. *Acta Mathematica*, 206(1):127–204, 2011.

[TV12] Terence Tao and Van Vu. Random matrices: universal properties of eigenvectors. *Random Matrices: Theory and Applications*, 1(01):1150001, 2012.

[TV14] Terence Tao and Van Vu. Random matrices: the universality phenomenon for wigner ensembles. *Modern aspects of random matrix theory*, 72:121–172, 2014.

[TW94] Craig A Tracy and Harold Widom. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159:151–174, 1994.

[TW96] Craig A Tracy and Harold Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177:727–754, 1996.

[Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[VHO+07] Eric Vincent, Jacques Hammann, Miguel Ocio, Jean-Philippe Bouchaud, and Leticia F Cugliandolo. Slow dynamics and aging in spin glasses. In *Complex Behaviour of Glassy Systems: Proceedings of the XIV Sitges Conference Sitges, Barcelona, Spain, 10–14 June 1996*, pages 184–219. Springer, 2007.

[Wig55] Eugene Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955.

[XCZG18] Pan Xu, Jinghui Chen, Difan Zou, and Quanquan Gu. Global convergence of langevin dynamics based algorithms for nonconvex optimization. *Advances in Neural Information Processing Systems*, 31, 2018.

[ZK16] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.