...MEAN-FIELD THEORY OF SPIN GLASSES AND BOOLEAN SATISFIABILITY...

JAESUNG SON

THESIS ADVISOR: ROBIN PEMANTLE

APRIL 16, 2018

1 Introduction

1.1 The Ising model

Suppose the Ising spins are localized the vertices of some region $\mathbb{L} = \{1, \cdots, L\}^d$ of a d-dimensional cubic lattice of side L. On each site $i \in \mathbb{L}$, there is an **Ising spin** $\sigma_i \in \{+1, -1\}$, which can be thought of as a spin that points either up $(+1)$ or down (-1).

A **configuration** of the system $\underline{\sigma} = (\sigma_1, \cdots, \sigma_N)$ is an assignment of values of all the spins in the system. So, the space of configurations $\mathcal{X}_N = \underbrace{\mathcal{X} \times \cdots \times \mathcal{X}}_{N \text{ times}} = \{+1, -1\}^{\mathbb{L}}$, with $\mathcal{X} = \{+1, -1\}$ and $N = L^d$.

Definition 1.1. We equip a configuration $\underline{\sigma}$ with an **energy function**, or **Hamiltonian**, as such:

$$E(\underline{\sigma}) = -J \sum_{i,j} \sigma_i \sigma_j - h \sum_{i \in \mathbb{L}} \sigma_i, \tag{1.2}$$

where the sum over i, j is over all unordered pairs of sites $i, j \in \mathbb{L}$ that are nearest neighbors and h measures the applied external magnetic field. The variables J's are interactions of a bond (ij) and are assigned a value $J > 0$ when the neighboring spins are aligned $\sigma_i = \sigma_j$ ($\uparrow\uparrow$ or $\downarrow\downarrow$) or a value $J < 0$ when the neighboring spins are anti-aligned $\sigma_i \neq \sigma_j$ ($\uparrow\downarrow$). The positive interaction $J > 0$ can lead to **ferromagnetism** (macroscopic magnetism) as all the pairs of spins in the system have the tendency to align in the same direction and thus is called a **ferromagnetic** interaction [N08]. The negative interaction $J < 0$ has the opposite effect in which spins tend to align in opposite directions and thus is called a **anti-ferromagnetic** interaction.

Ising solved the one-dimensional Ising model in his 1924 thesis, a problem given to him by Lenz, and showed the absence of any phase transitions [Is25]. Physical phase transitions can be understood as when a small temperature in some parameter such as temperature or pressure causes a drastic qualitative change in the state of the system. In the one-dimensional Ising model, Ising showed that for any positive, finite temperature, the system is always disordered.

Although Ising erroneously concluded the absence of any phase transitions in higher dimensions, Peierls, in 1936, showed that phase transitions in fact do occur in higher dimensions using an argument based on the free energy of domain walls, a boundary separating magnetic domains [PB36]. In 1944, Lars Onsager solved the two-dimensional Ising model by using the transfer-matrix method and showed that the model undergoes a phase transition between an ordered phase and a disordered phase [On44]. The attempt to extend the transfer-matrix method to the three-dimensional case was unsuccessful and the problem remains unsolved in the higher dimensional cases.

Definition 1.3. In general, the probability for the system to be found in a specific configuration x from a configuration space $\mathcal{X}$, $\mu_\beta(x)$, is given by the **Boltzmann distribution:**

$$\mu_\beta(x) = \frac{e^{-\beta E(x)}}{Z(\beta)}, \tag{1.4}$$

$$Z(\beta) = \sum_{x \in \mathcal{X}} e^{-\beta E(x)}. \tag{1.5}$$

Here, the parameter $\beta = 1/T$ is the inverse of temperature, with Boltzmann's constant $k_B = 1$. $Z(\beta)$ is a *normalization constant* and is called the **partition function**.

Note that at infinite temperature, $\beta = 0$, all configurations have equal weights according to the Boltzmann distribution and the energy function is irrelevant. In the case of the Ising model, all configurations have weight 2^{-N} and the Ising spins are completely independent. At zero temperature, $\beta = +\infty$, the Boltzmann distribution has only weights at the ground state(s), where, in general, the configuration $x_0 \in \mathcal{X}$ is a ground state if $E(x) \geq E(x_0)$ for any $x \in \mathcal{X}$. The minimum value of the energy $E_0 = E(x_0)$ is the ground state energy. If there is no magnetic field, $h = 0$, then there are two degenerate ground states: the configuration $\underline{\sigma}^{(+)}$ with all spins $\sigma_i = +1$ and the configuration $\underline{\sigma}^{(-)}$ with all spins $\sigma_i = -1$.

1.2 Thermodynamic potentials

Some of the properties of the Boltzmann distribution can be summarized through the thermodynamic potentials, which are functions of the temperature $1/\beta$ and the parameters that define the energy function $E(x)$.

Definition 1.6. **Free energy** is defined as

$$f(\beta) = -\frac{1}{\beta} \ln Z(\beta). \tag{1.7}$$

Definition 1.8. It may be more convenient to work with **free entropy**

$$\Phi(\beta) = -\beta f(\beta) = \ln Z(\beta). \tag{1.9}$$

Definition 1.10. From free entropy, we can derive the **internal energy** $U(\beta)$

$$U(\beta) = \frac{\partial}{\partial \beta}(\beta f(\beta)). \tag{1.11}$$

The thermodynamic potentials defined above are consolidated here for easier navigation and will appear later when we discuss the replica trick in detail.

1.3 Mean-field theory of spin glasses

The interaction of localized magnetic moments can be *ferromagnetic* in which all neighboring spins align in the same direction or *anti-ferromagnetic* in which all neighboring spins point in opposite directions. Spin glasses are *disordered* magnets which have a random magnetic spin structure, with some ferromagnetic and anti-ferromagnetic interactions, while ferromagnetic solids have an *ordered* spin structure in which all magnetic spins align in the same direction. The simplest model for ferromagnetism is the Ising model, which was discussed above.

1.3.1 Sherrington-Kirkpatrick model

In 1975, David Sherrington and Scott Kirkpatrick introduced an exactly solvable model of the spin glass called the **Sherrington Kirkpatrick model** (or SK model) [SK75]. Here, we will introduce the SK model and the idea of the replica calculation.

Consider the space of 2^N configurations of N Ising spins with values in $\{\pm 1\}$.

Definition 1.12. The energy function (or Hamiltonian) of the SK model is given by

$$E(\underline{\sigma}) = -\sum_{i<j} J_{ij}\sigma_i\sigma_j - h\sum_i \sigma_i, \tag{1.13}$$

where the first sum is over all $N(N-1)/2$ distinct pairs, the coupling random variables J_{ij}, $1 \le i < j \le N$ are independent and identically distributed (i.i.d) Gaussian $\mathcal{N}(J_0/N, J^2/N)$, and h is an external magnetic field that interacts with the spins.

1.3.2 The replica trick

A **sample** or instance of the SK model is given by one of the values of its 2^N energy levels. In order to describe typical samples, we have to compute the average log-partition function $\mathbb{E}\ln Z$. However, this is a rather difficult task and it is much easier to compute the integer moments of the partition function $\mathbb{E}Z^n$, with $n \in \mathbb{N}$, since Z is the sum of a large number of simple terms [MM09].

Lemma 1.14. *The* **replica method** *utilizes the following identity*

$$\mathbb{E}\ln Z = \lim_{n\to 0} \frac{\ln(\mathbb{E}Z^n)}{n} = \lim_{n\to 0} \frac{\lim \mathbb{E}Z^n - 1}{n} \tag{1.15}$$

by preparing $\mathbb{E}Z^n$ as if n were an integer and taking the limit as $n \to 0$ to apply the identity (1.15).

If we were to calculate $\mathbb{E}Z^n$ with n as a non-integer real number, the computation would be not much easier than that of $\mathbb{E}\ln Z$. The heart of the replica trick is to extrapolate $\mathbb{E}\ln Z$ by first assuming n to be an integer and later taking the limit as $n \to 0$. This bold justification has not yet been resolved; however, the replica method has been become standard practice as, when compared with other *exact* solutions, the method leads to the same results.

We will explore the derivation of the replica calculation for the Sherrington-Kirkpatrick model, including the replica symmetric solution and the 1-step replica symmetry breaking solution in Section 2. We only state the result of the full replica symmetry breaking solution and suggest Appendix B for the Parisi equation [N08]. We will also discuss the complications regarding the assumptions of the replica method.

1.4 Computational complexity

The theory of computational complexity deals with classifying computational problems by their difficulty, where an algorithm inputs some number N of variables and answers either yes or no. There are two classes of algorithms: polynomial and super-polynomial. Let $T(N)$ be the number of operations required to solve, in the worst-case, an instance of size N.

Definition 1.16. The algorithm is **polynomial** if there exists a constant k such that $T(N) = O(N^k)$.

Definition 1.17. In the big O notation, we say $f(x) = O(g(x))$ if and only if there exists some positive number M and real number x_0 such that $|f(x)| \leq M|g(x)|$ for $x \geq x_0$.

If the algorithm is not polynomial, it is **super-polynomial**.

Definition 1.18. To compare two decision problems $\mathcal{A}$ and $\mathcal{B}$, we say that $\mathcal{B}$ is **polynomially reducible** to $\mathcal{A}$ if the following conditions are satisfied [MM09]:

- There exists a mapping R which takes any instance I of the decision problem $\mathcal{B}$ to the instance $R(I)$ of $\mathcal{A}$ such that the solution of the instance $R(I)$ of $\mathcal{A}$ is the same solution as that of I of $\mathcal{B}$.
- The mapping R can be computed in polynomial time.

Such a mapping R is a **polynomial reduction**.

Some of the complexity classes of decision problems are as follows [MM09]:

- **P:** Polynomial problems can be solved by an algorithm running in polynomial time.
- **NP:** Non-deterministic polynomial problems can be solved by a non-deterministic algorithm running in polynomial time.
- **NP-complete:** A problem is NP-complete if it is in NP and if any other problem in NP can be polynomially reduced to it.
- **NP-hard:** Non-deterministic polynomial hard problems are not in NP but are as hard as NP-complete problems.

The following theorem by Stephen Cook and Leonid Levin is stated without proof:

Theorem 1.19. *The satisfiability problem is NP-complete.*

The implications of the theorems are quite remarkable in that any problem in NP is polynomially reducible to an instance of the satisfiability problem. Thus, any problem in NP can be solved in polynomial time.

Definition 1.20. We introduce the complexity class **#P**, which is the set of counting problems associated with decision problems in the complexity class NP.

Definition 1.21. Parallel to the decision problems, a counting problem is **#P-complete** if and only if it is in #P.

So, while an NP decision problem may ask whether there exists an assignment of variables that satisfies a given general Boolean formula, a #P counting problem would ask how many different variable assignments will satisfy that given general Boolean formula. Such a counting problem of counting the number of satisfying assignments of a given general Boolean formula is **#SAT**.

1.4.1 The satisfiability problem

Definition 1.22. A **clause** is a logical OR of some variables or their negations.

A variable x_i takes values 1 (`TRUE`) or 0 (`FALSE`) and its negation $\overline{x}_i := 1 - x_i$.

Definition 1.23. A variable or its negation is a **literal**, denoted by z_i.

A clause a involving K_a variables is a constraint that forbids exactly one of the 2^{K_a} possible assignments of these K_a variables. When a clause is not satisfied, it is said to be **violated**.

Definition 1.24. An instance of the satisfiability problem can be expressed by the **conjunctive normal form** (CNF):

$$F = C_1 \wedge \cdots \wedge C_M, \tag{1.25}$$

where $C_a = z_{i_1^a} \vee \cdots \vee z_{i_{K_a}^a}$, and $\{i_1^a, \cdots i_{K_a}^a\} \subseteq \{1, \cdots, N\}$.

Given F, is there an assignment of the x_i's in $\{0,1\}^N$ such that F is true? If so, then F is **SAT** and otherwise, F is **UNSAT**. This question of satisfiability is known as the **Boolean satisfiability problem**. We have the K**-satisfiability** (K-SAT) problem if we require all the clauses to have the same length $K_a = K$. The **MAX-SAT** problem is one in which we find the configuration that violates the fewest clauses. Note that if $K_a \leq 2$, then the problem is polynomial; however, if $K_a \leq K$ with $K \geq 3$, then the problem is NP-complete.

1.4.2 Random K-SAT threshold

An instance of a **random** K**-SAT** has only clauses of length K. $\text{SAT}_N(K, M)$ denotes the random K-SAT ensemble with N as the number of variables and M as the number of clauses. A formula in $\text{SAT}_N(K, M)$ with M clauses of size K is selected randomly from the $\binom{N}{K} 2^K$.

Definition 1.26. A parameter for the random K-SAT ensemble is the **clause density** $\alpha := M/N$.

An instance of the ensemble $\text{SAT}_N(K, \alpha)$ is generated by selecting each of the $\binom{N}{K} 2^K$ possible clauses independently with probability $\alpha N 2^{-K} / \binom{N}{K}$. $P_N(K, \alpha)$ is the probability that a randomly generated formula is satisfiable. As $\alpha \to 0$, the probability goes to 1 and as $\alpha \to \infty$, the probability goes to 0. Simulations indicate an existence of a phase transition at some finite value $\alpha_{\text{sat}}(K)$. So, for $\alpha < \alpha_{\text{sat}}(K)$, a random K-SAT formula is SAT with $P_N(K, \alpha) \to 1$ as $N \to \infty$, and for $\alpha > \alpha_{\text{sat}}(K)$, the formula is UNSAT with $P_N(K, \alpha) \to 1$ as $N \to \infty$. The existence of such a phase transition has been shown for the random 2-SAT at the critical clause density $\alpha_{\text{sat}}(2) = 1$ and the proof by bicycles, found in [MM09], is a result of [Go96] and is presented below.

Theorem 1.27. *Let $P_N(K=2,\alpha)$ be the probability for a $SAT_N(K=2,M)$ random formula to be SAT. Then,*

$$\lim_{N\to\infty} P_N(K=2,\alpha) = \begin{cases} 1 & \text{if } \alpha < 1, \\ 0 & \text{if } \alpha > 1. \end{cases} \tag{1.28}$$

Proof. We first show that a random formula is SAT with high probability for $\alpha < 1$. We define a directed graph $\mathcal{D}(F)$ of some formula F by associating each of the $2N$ literals with some vertex. Whenever there is a clause such as $\overline{x}_1 \vee x_2$, if $x_1 = 1$, then $x_2 = 1$, and if $x_2 = 0$, then $x_1 = 0$. We represent these cases graphically by drawing a directed edge from x_1 to x_2 and an undirected edge from $\overline{x}_2$ to $\overline{x}_1$. Then, F is UNSAT *iff* there exists some index $i \in \{1, \cdots, N\}$ such that $\mathcal{D}(F)$ contains a directed path from x_i to $\overline{x}_i$ and from $\overline{x}_i$ to x_i.

Define a bicycle of length s as a path $(u, w_1, w_2, \cdots, w_s, v)$, where the w_i's are literals on s distinct variables, and $u, v \in \{w_1, \cdots, w_s, \overline{w}_1, \cdots, \overline{w}_s\}$. From above, if a formula is UNSAT, then there exists a cycle containing two literals x_i and $\overline{x}_i$ for some $i \in \{1, \cdots, N\}$. Since the probability that there exists a bicycle in $\mathcal{D}(F)$, denoted $\mathbb{P}(\mathcal{A})$, is bounded above by the expected number of bicycles, we have

$$\mathbb{P}(F \text{ is UNSAT}) \leq \mathbb{P}(\mathcal{A}) \leq \sum_{s=2}^{N} N^s 2^s (2s)^2 M^{s+1} \left(\frac{1}{4\binom{N}{2}}\right)^{s+1}, \tag{1.29}$$

where s is the size of the bicycle. The sum (1.29) is $O(1/N)$ when $\alpha < 1$.

The proof for that the random formula is UNSAT with high probability for $\alpha > 1$ follows from Theorem (1.34). $\square$

Upper bounds on the satisfiability threshold can be obtained by using the first moment method: let $U(F)$ be a random variable such that

$$U(F) = \begin{cases} 0 & \text{if } F \text{ is UNSAT}, \\ \geq 1 & \text{otherwise}. \end{cases} \tag{1.30}$$

Thus, for any formula F, we have

$$\mathbb{P}(F \text{ is SAT}) \leq \mathbb{E}U(F). \tag{1.31}$$

The choice of $U(F) = Z(F)$, where $Z(F)$ is the number of SAT assignments is ideal since $\mathbb{E}U(F)$ vanishes as $N \to \infty$ for α large enough. Since the probability that an assignment is SAT is uniform on all the possible assignments,

$$\mathbb{E}Z(F) = 2^N(1-2^{-K})^M = \exp\left[N\left(\ln 2 + \alpha \ln(1-2^{-K})\right)\right]. \tag{1.32}$$

So, if $\alpha > \alpha^*(K)$, where

$$\alpha^*(K) := -\frac{\ln 2}{\ln(1-2^{-K})}, \tag{1.33}$$

$\mathbb{E}Z(F)$ is exponentially small at large N and so $\mathbb{P}(F \text{ is SAT})$ vanishes as $N \to \infty$. We restate the result from [FP83], but found in [MM09], below as the following theorem.

Theorem 1.34 (FP83)**.** *If $\alpha > \alpha^*(K)$, then $\lim_{N\to\infty} \mathbb{P}(F \textit{ is } SAT) = 0$, and so, $\alpha_{sat}^{(N)}(K) \leq \alpha^*(K)$.*

As simulations suggest that there exists a phase transition in the random K-SAT between the SAT and UNSAT phases for any $K \geq 2$, we have the following conjecture.

Conjecture 1.35 (Satisfiability threshold conjecture)**.** *For any $K \geq 2$, there exists a threshold $\alpha_{\mathrm{sat}}(K)$ with*

$$\lim_{N\to\infty} P_N(K,\alpha) = \begin{cases} 1 & \text{if } \alpha < \alpha_{sat}(K), \\ 0 & \text{if } \alpha > \alpha_{sat}(K). \end{cases} \tag{1.36}$$

As mentioned previously, the conjecture has been proven for $K = 2$. The theorem below from Friedgut [Fr99] strongly supports the case for the conjecture and all that remains to prove the satisfiability threshold conjecture is that $\alpha_{\mathrm{sat}}^{(N)}(K) \to \alpha_{\mathrm{sat}}(K)$ as $N \to \infty$.

Theorem 1.37 (Friedgut's Theorem)**.** *There exists a sequence of $\alpha_{\mathrm{sat}}^{(N)}(K)$ such that, for any $\varepsilon > 0$,*

$$\lim_{N\to\infty} P_N(K,\alpha_N) = \begin{cases} 1 & \text{if } \alpha_N < \alpha_{\mathrm{sat}}^{(N)}(K) - \varepsilon, \\ 0 & \text{if } \alpha_N > \alpha_{\mathrm{sat}}^{(N)}(K) + \varepsilon. \end{cases} \tag{1.38}$$

In section 3, we clarify the satisfiability threshold conjecture explore some of the recent developments on the bounds for the conjecture.

2 Mean-field Theory of Spin Glasses

2.1 Sherrington-Kirkpatrick model

Suppose we have the space of 2^N configurations of N Ising spins that take values $\{\pm 1\}$.

Definition 2.1. The energy function, or the Hamiltonian, of the SK model is given by

$$E(\underline{\sigma}) = -\sum_{i<j} J_{ij}\sigma_i\sigma_j - h\sum_{i=1}^{N}\sigma_i, \tag{2.2}$$

where the first sum is over all $\binom{N}{2} = N(N-1)/2$ distinct pairs with $1 \le i < j \le N$, and h is an external magnetic field that interacts with the spins.

Definition 2.3. The interactions J_{ij}'s are i.i.d. coupling random variable $\sim \mathcal{N}(J_0/N, J^2/N)$ with density function

$$P(J_{ij}) = \sqrt{\frac{N}{2\pi J^2}}\exp\left\{-\frac{N}{2J^2}\left(J_{ij} - \frac{J_0}{N}\right)^2\right\}. \tag{2.4}$$

Some others let $J_{ij} \sim \mathcal{N}(0, 1/N)$ but this is merely a technicality and the result is not affected. The SK model is an **infinite-range interaction** (**mean-field**) model as there is no notion of Euclidean distance between the positions of the spins and it is a **fully connected** model since each spin interacts directly with all the other spins [MM09]. The SK model is the closest to the original spin glass problem.

2.2 The replica derivation

Note that a **sample** or instance of the SK model is given by one of the values of its 2^N energy levels, $E_1, \cdots, E_{2^N}$, and from (1.5), the partition function is

$$Z = \sum_{i=1}^{2^N}\exp\left(-\beta E_i\right). \tag{2.5}$$

Assuming that n is an integer, the first step of the replica trick involves finding an expression for Z^n

$$Z^n = \prod_{a=1}^{n} \left(\sum_{i_a=1}^{2^N} \exp\left(-\beta E_{i_a}\right) \right) = \sum_{i_1=1}^{2^N} \cdots \sum_{i_n=1}^{2^N} \exp\left(-\beta (E_{i_1} + \cdots + E_{i_n})\right) \tag{2.6}$$

$$= \sum_{\{\sigma_i^a\}} \exp\left(-\beta \left(\sum_{a=1}^{n} E_{i_a} \right) \right), \tag{2.7}$$

which is the partition function of a new system with configuration given by the n-tuple $(i_1, \cdots, i_n)$ with $i_a \in \{1, \cdots, 2^N\}$ and the sum $\sum_{\{\sigma_i^a\}}$ is over the 2^{Nn} different configurations. This sum is will be shortened henceforth by the trace representation Tr. Using our expression for the energy function (2.2) gives us

$$Z^n = \operatorname{Tr} \exp\left\{ -\beta \left[-\sum_{a=1}^{n} \left(\sum_{i<j} J_{ij} \sigma_i^a \sigma_j^a + h \sum_{i=1}^{N} \sigma_i^a \right) \right] \right\} \tag{2.8}$$

$$= \operatorname{Tr} \left(\left[\prod_{i<j} \exp\left\{ \beta J_{ij} \sum_{a=1}^{n} \sigma_i^a \sigma_j^a \right\} \right] \left[\prod_{i=1}^{N} \exp\left\{ \beta h \sum_{a=1}^{n} \sigma_i^a \right\} \right] \right). \tag{2.9}$$

2.2.1 Replica average of the partition function

To proceed with our replica calculation, we take the expectation of the new partition function (2.9):

$$\mathbb{E} Z^n \overset{(i)}{=} \operatorname{Tr} \left(\prod_{i<j} \mathbb{E} \left[\exp\left\{ \beta J_{ij} \sum_{a=1}^{n} \sigma_i^a \sigma_j^a \right\} \right] \prod_{i=1}^{N} \exp\left\{ \beta h \sum_{a=1}^{n} \sigma_i^a \right\} \right) \tag{2.10}$$

$$\overset{(ii)}{=} \operatorname{Tr} \left(\prod_{i<j} \exp\left\{ \frac{\beta J_0}{N} \sum_{a=1}^{n} \sigma_i^a \sigma_j^a + \frac{\left[\beta J \sum_{a=1}^{n} \sigma_i^a \sigma_j^a \right]^2}{2N} \right\} \prod_{i=1}^{N} \exp\left\{ \beta h \sum_{a=1}^{n} \sigma_i^a \right\} \right) \tag{2.11}$$

$$\overset{(iii)}{=} \operatorname{Tr} \exp\left\{ \frac{\beta J_0}{N} \sum_{i<j} \sum_{a=1}^{n} \sigma_i^a \sigma_j^a + \frac{\beta^2 J^2}{2N} \sum_{i<j} \left(\sum_{a=1}^{n} \sigma_i^a \sigma_j^a \right)^2 + \beta h \sum_{i=1}^{N} \sum_{a=1}^{n} \sigma_i^a \right\}, \tag{2.12}$$

where (i) is from the linearity of expectation and the independence of the J_{ij}'s, (ii) is from the moment generating function of the normal distribution, and (iii) is from the property of exponentiation.

With a little algebra, we have the following expression for the sum

$$\left(\sum_{a=1}^{n} \sigma_i^a \sigma_j^a \right)^2 = \sum_{a=1}^{n} \sum_{b=1}^{n} \sigma_i^a \sigma_j^a \sigma_i^b \sigma_j^b = n + 2 \sum_{a<b} \sigma_i^a \sigma_j^a \sigma_i^b \sigma_j^b, \tag{2.13}$$

with the n from that $(\sigma_i^a)^2 (\sigma_j^b)^2 = 1$ and the sum over ordered pairs from the symmetry of the expression.

Using (2.13) and some manipulation gives us the following expression:

$$\frac{\beta^2 J^2}{2N}\sum_{i<j}\left(\sum_{a=1}^{n}\sigma_i^a\sigma_j^a\right)^2 = \frac{\beta^2J^2}{2N}\cdot\frac{N(N-1)n}{2} + \frac{\beta^2J^2}{2N}2\sum_{a<b}\sum_{i<j}\sigma_i^a\sigma_j^a\sigma_i^b\sigma_j^b \tag{2.14}$$

$$= \frac{\beta^2J^2(N-1)n}{4} + \frac{\beta^2J^2}{2N}\left(\sum_{a<b}\sum_{i,j=1}^{N}\sigma_i^a\sigma_j^a\sigma_i^b\sigma_j^b - \sum_{a<b}\sum_{i=1}^{N}\sigma_i^a\sigma_i^a\sigma_i^b\sigma_i^b\right) \tag{2.15}$$

$$= \frac{\beta^2J^2Nn}{4} - \frac{\beta^2J^2n}{4} + \frac{\beta^2J^2}{2N}\sum_{a<b}\left(\sum_{i=1}^{N}\sigma_i^a\sigma_i^b\right)^2 - \frac{\beta^2J^2n(n-1)}{4}. \tag{2.16}$$

Similarly, we have

$$\frac{\beta J_0}{N}\sum_{i<j}\sum_{a=1}^{n}\sigma_i^a\sigma_j^a = \frac{\beta J_0}{2N}\left(\sum_{a=1}^{n}\sum_{i,j=1}^{N}\sigma_i^a\sigma_j^a - \sum_{a=1}^{n}\sum_{i=1}^{N}\sigma_i^a\sigma_i^a\right) \tag{2.17}$$

$$= \frac{\beta J_0}{2N}\sum_{a=1}^{n}\left(\sum_{i=1}^{N}\sigma_i^a\right)^2 - \frac{\beta J_0 n}{2}. \tag{2.18}$$

Since the constant terms (those without N) in (2.16) and (2.18) are irrelevant to the leading exponential order in the large N-limit, we discard them henceforth in our calculations, replacing $=$, the equal sign, with $\doteq$, the notation for the leading exponential order. Thus, we have

$$\mathbb{E}Z^n \doteq \operatorname{Tr}\exp\left\{\frac{\beta J_0}{2N}\sum_{a=1}^{n}\left(\sum_{i=1}^{N}\sigma_i^a\right)^2 + \frac{\beta^2J^2Nn}{4} + \frac{\beta^2J^2}{2N}\sum_{a<b}\left(\sum_{i=1}^{N}\sigma_i^a\sigma_i^b\right)^2 + \beta h\sum_{i=1}^{N}\sum_{a=1}^{n}\sigma_i^a\right\} \tag{2.19}$$

$$\doteq \exp\left(\frac{\beta^2J^2Nn}{4}\right)\operatorname{Tr}\exp\left\{\frac{\beta J_0}{2N}\sum_{a=1}^{n}\left(\sum_{i=1}^{N}\sigma_i^a\right)^2 + \frac{\beta^2J^2}{2N}\sum_{a<b}\left(\sum_{i=1}^{N}\sigma_i^a\sigma_i^b\right)^2 + \beta h\sum_{i=1}^{N}\sum_{a=1}^{n}\sigma_i^a\right\}. \tag{2.20}$$

Using the property of exponentiation, we have

(2.21)

$$\mathbb{E}Z^n \doteq \exp\left(\frac{\beta^2J^2Nn}{4}\right)\operatorname{Tr}\exp\left\{\beta h\sum_{i=1}^{N}\sum_{a=1}^{n}\sigma_i^a\right\}\prod_{a=1}^{n}\exp\left\{\frac{\beta J_0}{2N}\left(\sum_{i=1}^{N}\sigma_i^a\right)^2\right\}\prod_{a<b}\exp\left\{\frac{\beta^2J^2}{2N}\left(\sum_{i=1}^{N}\sigma_i^a\sigma_i^b\right)^2\right\}$$

2.2.2 Reduction by Gaussian integral

Lemma 2.22. *From the distribution of the Gaussian $\mathcal{N}(\mu,\sigma^2)$ random variable,*

$$\int_{-\infty}^{\infty}\frac{1}{2\pi\sqrt{\sigma^2}}e^{-(x-\mu)/(2\sigma^2)}dx, \tag{2.23}$$

we can derive the following identity:

$$\sqrt{\frac{b}{2\pi}}\int_{-\infty}^{\infty} e^{-bx^2/2+abx}dx = \exp\left(\frac{ba^2}{2}\right). \tag{2.24}$$

From lemma 2.22, letting $a = \sum_{i=1}^{N} \sigma_i^a/N$, $b = \beta J_0 N$, and the integration variables be m_a, $a \in \{1, \cdots, n\}$,

$$\begin{aligned}\prod_{a=1}^{n} \exp\left\{\frac{\beta J_0 N}{2}\left(\sum_{i=1}^{N}\sigma_i^a/N\right)^2\right\} &= \prod_{a=1}^{n}\left[\sqrt{\frac{\beta J_0 N}{2\pi}}\int_{-\infty}^{\infty}\exp\left\{-\frac{\beta J_0 N m_{ab}^2}{2} + \beta J_0 m_a \sum_{i=1}^{N}\sigma_i^a\right\} dm_a\right] \\ &= \int\left(\prod_{a=1}^{n} dm_a\sqrt{\frac{\beta J_0 N}{2\pi}}\right)\exp\left\{-\frac{\beta J_0 N}{2}\sum_{a=1}^{n} m_a^2\right\}\exp\left\{\beta J_0 \sum_{a=1}^{n}\left(\sum_{i=1}^{N}\sigma_i^a\right)m_a\right\},\end{aligned} \tag{2.25}$$

where we instead use the integral notation $\int dx f(x)$ instead of $\int f(x)dx$ for convenience. In this unusual integral notation, we take $\int\left(\prod_i dx_i c\right)\prod_i f(x_i)$ to be $\int\prod_i dx_i c f(x_i)$, where c is some constant.

Similarly, letting $a = \sum_{i=1}^{N}\sigma_i^a\sigma_i^b/N$, $b = \beta^2 J^2 N$, and the integration variables be Q_{ab}, $1 \leq a < b <\leq n$,

$$\begin{aligned}\prod_{a<b} \exp\left\{\frac{\beta^2 J^2 N}{2}\left(\sum_{i=1}^{N}\sigma_i^a\sigma_i^b/N\right)^2\right\} &= \prod_{a<b}\left[\sqrt{\frac{\beta^2 J^2 N}{2\pi}}\int_{-\infty}^{\infty}\exp\left\{-\frac{\beta^2 J^2 N Q_{ab}^2}{2} + \beta^2 J^2 \sum_{i=1}^{N}\sigma_i^a\sigma_i^b Q_{ab}\right\} dQ_{ab}\right] \\ &= \int\left(\prod_{a<b} dQ_{ab}\sqrt{\frac{\beta^2 J^2 N}{2\pi}}\right)\exp\left\{-\frac{\beta^2 J^2 N}{2}\sum_{a<b} Q_{ab}^2\right\}\exp\left\{\beta^2 J^2 \sum_{a<b}\sum_{i=1}^{N}\sigma_i^a\sigma_i^b Q_{ab}\right\}.\end{aligned} \tag{2.26}$$

Substituting (2.25) and (2.26) into (2.21) gives us

$$\begin{aligned}\mathbb{E}Z^n \doteq \exp\left(\frac{\beta^2 J^2 N n}{4}\right) &\int\left(\prod_{a=1}^{n} dm_a\sqrt{\frac{\beta J_0 N}{2\pi}}\right)\int\left(\prod_{a<b} dQ_{ab}\sqrt{\frac{\beta^2 J^2 N}{2\pi}}\right) \\ &\cdot \exp\left\{-\left(\frac{\beta J_0 N}{2}\sum_{a=1}^{n} m_a^2 + \frac{\beta^2 J^2 N}{2}\sum_{a<b} Q_{ab}^2\right)\right\} \\ &\cdot \operatorname{Tr}\exp\left\{\beta\sum_{a=1}^{n}\sum_{i=1}^{N}(h + J_0 m_a)\sigma_i^a + \beta^2 J^2 \sum_{a<b}\sum_{i=1}^{N}\sigma_i^a\sigma_i^b Q_{ab}\right\}.\end{aligned} \tag{2.27}$$

Lemma 2.28. *If g is some arbitrary function, we have that*

$$\operatorname{Tr}\exp\left\{\sum_{i=1}^{N} g(\sigma_i^a)\right\} = \exp\left\{N \ln \operatorname{tr}\exp g(\sigma^a)\right\}, \tag{2.29}$$

where the trace $\operatorname{tr} = \sum_{\{\sigma^a\}}$ *is over all states of a single replicated spin* σ^a [FH93].

Using the relation (2.29), we have that

$$\mathbb{E}Z^n \doteq \int \left(\prod_{a=1}^{n} dm_a \sqrt{\frac{\beta J_0 N}{2\pi}} \right) \int \left(\prod_{a<b} dQ_{ab} \sqrt{\frac{\beta^2 J^2 N}{2\pi}} \right) \exp \left\{ NG(Q,m) \right\}, \tag{2.30}$$

where $G(Q,m)$ is a function of $n(n-1)/2+n$ variables Q_{ab}, $1 \le a < b \le n$, and m_a, $1 \le a \le n$, with

$$G(Q,m) := \frac{\beta^2 J^2 n}{4} - \frac{\beta J_0}{2} \sum_{a=1}^{n} m_a^2 - \frac{\beta^2 J^2}{2} \sum_{a<b} Q_{ab}^2 + \ln \left(\operatorname{tr} \exp \left\{ \beta \sum_{a=1}^{n} (h + J_0 m_a) \sigma^a + \beta^2 J^2 \sum_{a<b} \sigma^a \sigma^b Q_{ab} \right\} \right). \tag{2.31}$$

2.2.3 Method of steepest descent

The method of steepest descent (or the saddle-point method) approximates an integral in roughly the direction of the steepest descent. The integral to be approximated in (2.30) is in the form of

$$\int_a^b e^{\lambda g(z)} dz, \tag{2.32}$$

where a and b are some limits of integration and λ is large.

Because the integral will be dominated by the highest saddle point, $a < z_0 < b$, when the function $g(z)$ is near z_0, we have, from the Taylor series expansion,

$$g(z) \approx g(z_0) + \frac{1}{2} g''(z_0)(z - z_0)^2, \tag{2.33}$$

where the first derivative term is zero since we choose z_0 such that $g'(z_0) = 0$.

Lemma 2.34. *The integral in* (2.32) *can be approximated as such*

$$\int_a^b h(z) e^{\lambda g(z)} dz \approx \sqrt{\frac{2\pi}{\lambda \left| g''(z_0) \right|}} h(z_0) e^{\lambda g(z_0)}, \tag{2.35}$$

as $\lambda \to \infty$ if $g''(z_0) < 0$ and h is positive.

At the extremum of $G(Q,m)$, namely (Q^*, m^*) which we will determine later, we can approximate the integral (2.30) for large N using the saddle-point method (2.35)

$$\mathbb{E}Z^n|_{(Q^*,m^*)} \approx \exp \left\{ NG(Q^*, m^*) \right\} \tag{2.36}$$

$$\approx 1 + n \left\{ \frac{\beta^2 J^2 N}{4} - \frac{\beta J_0 N}{2n} \sum_{a=1}^{n} m_a^2 - \frac{\beta^2 J^2 N}{2n} \sum_{a<b} Q_{ab}^2 + \frac{1}{n} \ln \left(\operatorname{tr}\, e^{L(Q,m)} \right) \right\} \Bigg|_{(Q^*,m^*)}. \tag{2.37}$$

where the function $L(Q,m)$ of Q and m is defined as

$$L(Q,m) := \beta \sum_{a=1}^{n} (h + J_0 m_a)\sigma^a + \beta^2 J^2 \sum_{a<b} \sigma^a \sigma^b Q_{ab}, \tag{2.38}$$

and in taking the limit $n \to 0$ and letting $N < \infty$ be large, we obtain the last approximation (2.37) from the tangent line approximation of $e^{\lambda x}$ near $x = 0$:

$$e^{\lambda x} \approx 1 + \lambda x. \tag{2.39}$$

With the definition of free entropy (1.8) and the replica identity (1.15), we have the following expression for free energy

$$-\beta[f] = \lim_{n\to 0} \frac{\mathbb{E}Z^n - 1}{nN} \tag{2.40}$$

$$= \lim_{n\to 0} \left\{ \frac{\beta^2 J^2}{4} - \frac{\beta J_0}{2n} \sum_{a=1}^{n} m_a^2 - \frac{\beta^2 J^2}{2n} \sum_{a<b} Q_{ab}^2 + \frac{1}{n} \ln \left(\text{tr } e^{L(Q,m)} \right) \right\} \Bigg|_{(Q^*, m^*)}, \tag{2.41}$$

where $[f] = \dfrac{f}{N}$ is the configurational average of the free energy.

To estimate the integral (2.30) at large N via the saddle-point method, we differentiate the function $G(Q,m)$ (2.31) with respect to its arguments and set it equal to zero. For all $a < b$, we have that

$$\frac{\partial G}{\partial Q_{ab}} = -\beta^2 J^2 Q_{ab} + \frac{\text{tr } \beta^2 J^2 \left(\sigma^a \sigma^b\right) \exp\left\{ \beta \sum_{a=1}^{n} (h + J_0 m_a)\sigma^a + \beta^2 J^2 \sum_{a<b} \sigma^a \sigma^b Q_{ab} \right\}}{\text{tr } \exp\left\{ \beta \sum_{a=1}^{n} (h + J_0 m_a)\sigma^a + \beta^2 J^2 \sum_{a<b} \sigma^a \sigma^b Q_{ab} \right\}} = 0, \tag{2.42}$$

which implies

$$Q_{ab} = \langle \sigma^a \sigma^b \rangle_n. \tag{2.43}$$

Definition 2.44. We introduce here the notation

$$\langle f(\sigma) \rangle_n := \frac{1}{Z(Q_{ab}, m_a)} \text{tr } f(\sigma) \exp\left\{ \beta \sum_{a=1}^{n} (h + J_0 m_a)\sigma^a + \beta^2 J^2 \sum_{a<b} \sigma^a \sigma^b Q_{ab} \right\}, \tag{2.45}$$

$$Z(Q_{ab}, m_a) := \text{tr } \exp\left\{ \beta \sum_{a=1}^{n} (h + J_0 m_a)\sigma^a + \beta^2 J^2 \sum_{a<b} \sigma^a \sigma^b Q_{ab} \right\}, \tag{2.46}$$

where $f(\sigma) = f(\sigma^1, \cdots, \sigma^n)$ is any function.

Similarly, we have

$$\frac{\partial G}{\partial m_a} = -\beta J_0 m_a + \frac{\text{tr } (\beta J_0 \sigma^a) \exp\left\{\beta \sum_{a=1}^{n}(h + J_0 m_a)\sigma^a + \beta^2 J^2 \sum_{a<b} \sigma^a \sigma^b Q_{ab}\right\}}{\text{tr } \exp\left\{\beta \sum_{a=1}^{n}(h + J_0 m_a)\sigma^a + \beta^2 J^2 \sum_{a<b} \sigma^a \sigma^b Q_{ab}\right\}}, \tag{2.47}$$

which implies that

$$m_a = \langle \sigma^a \rangle_n. \tag{2.48}$$

We have the following second partial derivatives

$$\begin{aligned} \left.\frac{\partial G}{\partial Q_{ab}^2}\right|_{Q_{ab}=\langle\sigma^a\sigma^b\rangle_n} &= -\beta^2 J^2 + \left.\frac{\left\langle \left(\beta^2 J^2 \sigma^a \sigma^b\right)^2 \right\rangle_n Z(Q_{ab}, m_a) - \left\langle \beta^2 J^2 \sigma^a \sigma^b \right\rangle_n^2 \left(Z(Q_{ab}, m_a)\right)^2}{\left(Z(Q_{ab}, m_a)\right)^2}\right|_{Q_{ab}=\langle\sigma^a\sigma^b\rangle_n} \\ &= -\beta^2 J^2, \end{aligned} \tag{2.49}$$

$$\begin{aligned} \left.\frac{\partial G}{\partial m_a^2}\right|_{m_a=\langle\sigma^a\rangle_n} &= -\beta J_0 + \left.\frac{\left\langle \left(\beta J_0 \sigma^a\right)^2 \right\rangle_n Z(Q_{ab}, m_a) - \left\langle \beta J_0 \sigma^a \right\rangle_n^2 \left(Z(Q_{ab}, m_a)\right)^2}{\left(Z(Q_{ab}, m_a)\right)^2}\right|_{m_a=\langle\sigma^a\rangle_n} \\ &= -\beta J_0. \end{aligned} \tag{2.50}$$

We can also confirm from taking the partial derivatives of (2.41) our results in (2.43) and (2.48).

2.3 Replica symmetric solution

It is natural (but naïve) to assume **replica symmetry** (RS), $Q_{ab} = q$ for $a \neq b$, $Q_{ab} = q$ for $a = b$, and $m_a = m$, and derive a replica-symmetric solution. Changing the sums over $a < b$ to sums over $a \neq b$ by symmetry and substituting our values for Q_{ab} and m_a into the free energy expression (2.41) give us, before the limit $n \to 0$,

$$-\beta[f]_{\text{RS}} = \frac{\beta^2 J^2}{4} - \frac{\beta J_0}{2n} n m^2 - \frac{\beta^2 J^2}{4n}(n(n-1)q^2) + \frac{1}{n} \ln \text{tr } e^{L(Q_{\text{RS}}, m_{\text{RS}})}, \tag{2.51}$$

where $L(Q_{\text{RS}}, m_{\text{RS}})$ is simply

$$L(Q_{\text{RS}}, m_{\text{RS}}) := \beta \sum_{a=1}^{n}(h + J_0 m)\sigma^a + \frac{\beta^2 J^2 q}{2} \sum_{a \neq b} \sigma^a \sigma^b. \tag{2.52}$$

With a little algebra, we have that

$$\sum_{a\neq b}\sigma^a\sigma^b=\left(\sum_{a=1}^{n}\sigma^a\right)^2-\sum_{a=1}^{n}\sigma^a\sigma^a, \tag{2.53}$$

and so $L(Q_{\mathrm{RS}},m_{\mathrm{RS}})$ can be rewritten as

$$L(Q_{\mathrm{RS}},m_{\mathrm{RS}})=\beta\sum_{a=1}^{n}(h+J_0m)\sigma^a+\frac{\beta^2J^2q}{2}\left(\sum_{a=1}^{n}\sigma^a\right)^2-\frac{\beta^2J^2qn}{2} \tag{2.54}$$

The last term can be simplified using the Gaussian reduction identity (2.24) with the integration variable as z, $a=\sum_{a=1}^{n}\sigma^a$, and $b=\beta^2J^2q$

$$\ln\operatorname{tr}\,e^{L(Q_{\mathrm{RS}},m_{\mathrm{RS}})}=\ln\operatorname{tr}\sqrt{\frac{\beta^2J^2q}{2\pi}}\int_{-\infty}^{\infty}dz\exp\left\{-\frac{\beta^2J^2q}{2}z^2+\left(\sum_{a=1}^{n}\sigma^a\right)\beta^2J^2qz+\beta\sum_{a=1}^{n}(h+J_0m)\sigma^a-\frac{n}{2}\beta^2J^2q\right\}$$

$$=\ln\int_{-\infty}^{\infty}dz\left(\frac{e^{-z^2/2}}{\sqrt{2\pi}}\right)\exp\left\{-\frac{n}{2}\beta^2J^2q\right\}\operatorname{tr}\prod_{a=1}^{n}\exp\left\{\,(\beta J\sqrt{q}z+\beta h+J_0m)\,\sigma^a\right\} \tag{2.55}$$

$$=\ln\int_{-\infty}^{\infty}dz\left(\frac{e^{-z^2/2}}{\sqrt{2\pi}}\right)\exp\left\{-\frac{n}{2}\beta^2J^2q\right\}\prod_{a=1}^{n}\operatorname{tr}\exp\left\{\,(\beta J\sqrt{q}z+\beta h+J_0m)\,\sigma^a\right\}, \tag{2.56}$$

where in (2.55) we replaced β^2J^2qz with z, which does not change the expression as $\beta^2J^2q>0$ and the limits of integration are from $-\infty$ to ∞ and in (2.56) we used the property of exponentiation to interchange the order of tr $\prod_{a=1}^{n}$. Since the single replica spin can take one of two values $\sigma^a\in\{\pm1\}$, we have

$$=\ln\int_{-\infty}^{\infty}dz\left(\frac{e^{-z^2/2}}{\sqrt{2\pi}}\right)\exp\left\{-\frac{n}{2}\beta^2J^2q\right\}\prod_{a=1}^{n}2\cosh\left\{\,(\beta J\sqrt{q}z+\beta h+J_0m)\,\sigma^a\right\}$$

$$=\ln\int_{-\infty}^{\infty}\mathbf{D}z\exp\left\{n\ln\left[2\cosh\left(\beta J\sqrt{q}z+\beta h+\beta J_0m\right)\right]-\frac{n}{2}\beta^2J^2q\right\}$$

$$=\ln\left(1+n\int_{-\infty}^{\infty}\mathbf{D}z\ln\left(2\cosh\left(\beta\tilde{H}(z)\right)\right)-\frac{n}{2}\beta^2J^2q+O(n^2)\right), \tag{2.57}$$

where the hyperbolic function $\cosh(z)=(e^z+e^{-z})/2$, $\mathbf{D}z=dz\,e^{-z^2/2}/\sqrt{2\pi}$ is the Gaussian measure, $\tilde{H}(z)=J\sqrt{q}z+h+J_0m$, and $O(\cdots)$ is the big O notation. In the last equality (2.57), we used the Taylor expansion $e^z=1+z+O(z^2)$. Using our expression (2.57) in (2.51) and taking the limit $n\to0$ give us

$$-\beta[f]_{\mathrm{RS}}=\frac{\beta^2J^2}{4}(1-q)^2-\frac{\beta J_0m^2}{2}+\int\mathbf{D}z\ln\left(2\cosh\left(\beta\tilde{H}(z)\right)\right), \tag{2.58}$$

where in applying L'Hôpital's rule, we have the limit

$$\lim_{x\to 0}\frac{\ln(1+ax+bx^2)}{x}=\lim_{x\to 0}\frac{\left(\dfrac{a+2bx}{1+ax+bx^2}\right)}{1}=a. \tag{2.59}$$

Solving the extremization condition for free energy (2.58) with respect to q gives us

$$0=\frac{\beta^2J^2}{2}(q-1)+\int \mathbf{D}z\left(\frac{2\sinh(\beta\tilde{H}(z))}{2\cosh(\beta\tilde{H}(z))}\right)\frac{\beta J z}{2\sqrt{q}} \tag{2.60}$$

$$=\frac{\beta^2J^2}{2}(q-1)-\tanh(\beta\tilde{H}(z))e^{-z^2/2}\beta J/\sqrt{8\pi q}\Big|_{-\infty}^{+\infty}+\int dz e^{-z^2/2}\beta J\ \mathrm{sech}^2(\beta\tilde{H}(z))\beta J\sqrt{q}/\sqrt{8\pi q} \tag{2.61}$$

$$=\frac{\beta^2J^2}{2}q-\frac{\beta^2J^2}{2}+\int \mathbf{D}z\beta^2J^2\ \mathrm{sech}^2(\beta\tilde{H}(z))/2, \tag{2.62}$$

by integration by parts using $u=\tanh(\beta\tilde{H}(z))$ and $dv=dze^{-z^2/2}\beta Jz/\sqrt{8\pi q}$ and from that $e^{-z^2/2}$ is an even function of z that decays to zero. This implies

$$q=1-\int \mathbf{D}z\,\mathrm{sech}^2\left(\beta\tilde{H}(z)\right)=\int \mathbf{D}z\tanh^2\left(\beta\tilde{H}(z)\right) \tag{2.63}$$

Solving the extremization condition for free energy (2.58) with respect to m gives us

$$0=-\beta J_0 m+\int \mathbf{D}z\left(\frac{2\sinh(\beta\tilde{H}(z))}{2\cosh(\beta\tilde{H}(z))}\right)\beta J_0, \tag{2.64}$$

which implies

$$m=\int \mathbf{D}z\tanh\left(\beta\tilde{H}(z)\right). \tag{2.65}$$

2.3.1 Phase diagram

The behavior of the solutions (2.63) and (2.65) depends on the parameters β and J_0. For simplicity, suppose that the external magnetic field $h=0$ for the rest of the section. If the distribution of the coupling random variable J_{ij} is symmetric, then $J_0=0$ and $\tilde{H}(z)=J\sqrt{q}z$. So, $\tanh(\beta\tilde{H}(Z))$ is an odd function of z. Thus, the magnetization m in (2.63) is 0 and there is no ferromagnetic phase. The free energy is now

$$-\beta[f]_{\mathrm{RS}}=\frac{1}{4}\beta^2J^2(1-q)^2+\int \mathbf{D}z\ln\left(2\cosh\left(\beta J\sqrt{q}z\right)\right) \tag{2.66}$$

$$=\frac{1}{4}\beta^2J^2-\frac{1}{2}\beta^2J^2q+\frac{1}{4}\beta^2J^2q^2+\ln 2+\frac{1}{2}\beta^2J^2q-\frac{1}{4}\beta^4J^4q^2+O(q^3) \tag{2.67}$$

$$=\frac{1}{4}\beta^2J^2+\ln 2+\frac{1}{4}\beta^2J^2q^2\left(1-q^2\right)+O(q^3). \tag{2.68}$$

Lemma 2.69. *We used the following asymptotic expansion below in* (2.67) *to obtain an expression*

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \ln\cosh(\beta J\sqrt{q}z)dz = -\sum_{n=1}^{N} \frac{B_{2n}}{n!n} 2^{n-1}\left(1-2^{2n}\right)(\beta J\sqrt{q})^{2n} + O\left(q^{2N+2}\right), \tag{2.70}$$

where $B_{2n} = \frac{(-1)^{n+1}2(2n)!}{(2\pi)^{2n}}\zeta(2n)$ *for* $n \geq 1$ *is the Bernoulli numbers and* $\zeta(2n) = \sum_{k=1}^{\infty}\frac{1}{k^{2n}}$ *is the Riemann zeta function.*

The Landau theory implies that the critical point is determined by the condition of a vanishing coefficient of the second order term q^2 and so the spin glass transition exists at $T = J =: T_f$ [N08].

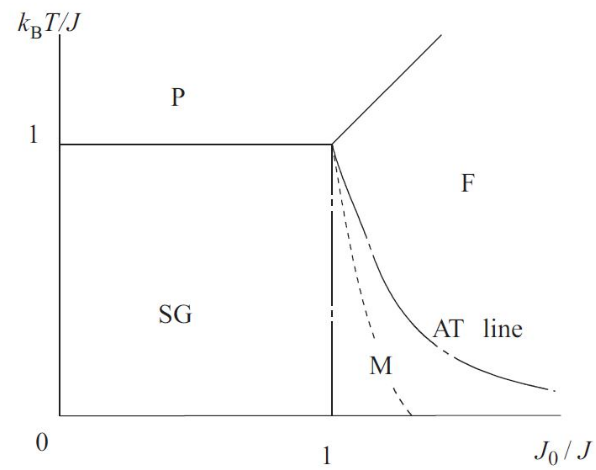

FIGURE 1. Phase diagram of the SK model. The dashed line is the boundary between the ferromagnetic (F) and spin glass (SG) phases and exists only under the ansatz of replica symmetry. The replica-symmetric solution is unstable below the AT line, and a mixed phase (M) emerges between the spin glass and ferromagnetic phases. The system is in the paramagnetic phase (P) in the high-temperature region [N08].

The boundary between the spin glass and ferromagnetic phases is given numerically by solving (2.63) and (2.65) and we can depict the phase diagram as shown above in figure 1 [N08].

2.3.2 Negative entropy

We see that the assumption for replica symmetry fails at low temperatures as the ground-state entropy for $J_0 = 0$ is the negative value $-1/2\pi$. Since $\lim_{x\to\infty} \tanh^2(x) = 1$, we have $q \to 1$ as $1/\beta = T \to 0$, and assume that $q = 1 - aT$, with $a > 0$ for small $T > 0$. To check this assumption, we have, for $\beta \to \infty$,

$$\int \mathbf{D}z \, \operatorname{sech}^2(\beta J z) = \frac{1}{\beta J} \int \mathbf{D}z \frac{d}{dz} \tanh(\beta J z) \tag{2.71}$$

$$\to \frac{1}{\beta J} \int \mathbf{D}z \, \{2\delta(z)\} \tag{2.72}$$

$$= \sqrt{\frac{2}{\pi}} \frac{T}{J}. \tag{2.73}$$

Thus, the assumption that $q = 1 - aT$ is justified with $a = \sqrt{\frac{2}{\pi}} \frac{1}{J}$. Substituting $q = 1 - aT$ into the expression for free energy (2.66) gives us, for large β,

$$-\beta [f]_{\mathrm{RS}} = \frac{1}{4} \beta^2 J^2 a^2 T^2 + 2 \int_0^\infty \mathbf{D}z \ln \left(e^{\beta J \sqrt{q} z} \left(1 + e^{-2\beta J \sqrt{q} z} \right) \right) \tag{2.74}$$

$$\approx \frac{J^2}{4} \frac{2}{J^2 \pi} + 2 \int_0^\infty \mathbf{D}z \left(\beta J \sqrt{q} z + \ln \left(1 + e^{-2\beta J \sqrt{q} z} \right) \right) \tag{2.75}$$

$$\approx \frac{1}{2\pi} + \frac{2\beta J (1 - aT/2)}{\sqrt{2\pi}} + 2 \int_0^\infty \mathbf{D}z \, e^{-2\beta J \sqrt{q} z}, \tag{2.76}$$

where we changed $\int_{-\infty}^{\infty} \mathbf{D}z$ to $2 \int_0^\infty \mathbf{D}z$ since $\cosh(z)$ is an event function of z and used the approximation $\ln(1 + z) \approx z$ for small z.

$$-\beta [f]_{\mathrm{RS}} \approx \frac{1}{2\pi} + \frac{2\beta J \left(1 - \sqrt{\frac{2}{\pi}} \frac{T}{2J} \right)}{\sqrt{2\pi}} + \exp \left\{ 2\beta^2 J^2 \left(1 - \sqrt{\frac{2}{\pi}} \frac{T}{J} \right) \right\} \tag{2.77}$$

$$= -\frac{1}{2\pi} + \sqrt{\frac{2}{\pi}} \beta J. \tag{2.78}$$

So, we have the following expression for the free energy in low temperature

$$[f]_{\mathrm{RS}} \approx \frac{T}{2\pi} - \sqrt{\frac{2}{\pi}} J, \tag{2.79}$$

with the ground-state entropy as $-\frac{1}{2\pi}$ and the ground-state energy $-\sqrt{\frac{2}{\pi}} J$.

Although the interchange of the limits $n \to 0$ and $N \to \infty$ to derive the free energy (2.41) using the method of steepest descent may have been the source of the negative entropy, we see in section 2.6 that the problem lies in the replica method itself.

2.4 Replica symmetry breaking

The free energy expression derived through the replica symmetry ansatz (RS ansatz) led to an erroneous negative entropy at low temperatures. So, it seems logical to "break" the replica symmetry and proceed with an approach in which the order parameter Q_{ab} possibly depends on the replica indices a and b and m_a on the replica index a.

2.4.1 The Parisi Ansatz

The $n \times n$ matrix $\{Q_{ab}\}$ for the replica-symmetric solution had $Q_{ab} = q$ for $a \neq b$ and 0 along the diagonals as such

$$\{Q_{ab}\} = \begin{pmatrix} 0 & q & \cdots & q \\ q & 0 & & q \\ \vdots & & \ddots & \vdots \\ q & q & \cdots & 0 \end{pmatrix}. \tag{2.80}$$

To start the first step of the replica symmetry breaking (1RSB), let $m_1 \leq n$ be a positive integer that divides the replicas into n/m_1 block(s). A "diagonal" block, which contains the main diagonal entries, has 0's along its diagonals and q_1 as its off-diagonal entries while an "off-diagonal" block, which does not contain the main diagonal entries, has q_0 as its entries as such

$$\begin{pmatrix} 0 & q_1 & q_1 & q_1 & q_0 & q_0 & q_0 & q_0 \\ q_1 & 0 & q_1 & q_1 & q_0 & q_0 & q_0 & q_0 \\ q_1 & q_1 & 0 & q_1 & q_0 & q_0 & q_0 & q_0 \\ q_1 & q_1 & q_1 & 0 & q_0 & q_0 & q_0 & q_0 \\ q_0 & q_0 & q_0 & q_0 & 0 & q_1 & q_1 & q_1 \\ q_0 & q_0 & q_0 & q_0 & q_1 & 0 & q_1 & q_1 \\ q_0 & q_0 & q_0 & q_0 & q_1 & q_1 & 0 & q_1 \\ q_0 & q_0 & q_0 & q_0 & q_1 & q_1 & q_1 & 0 \end{pmatrix} \tag{2.81}$$

FIGURE 2. 8×8 Parisi matrix with one step RSB ($n = 8, m_1 = 4$).

Then, in the second step of the replica symmetry breaking, the off-diagonal blocks remain the same, but the diagonal blocks are further divided into m_1/m_2 blocks. Within the subdivided blocks, the off-diagonal blocks remain the same while the diagonal blocks have q_2 instead of q_1 in their off-diagonal entries as such The numbers $n \geq m_1 \geq m_2 \geq \cdots \geq 1$ are integers and we define the function $q(x)$ as

$$q(x) = q_i \qquad \text{for } m_{i+1} \leq x \leq m_i. \tag{2.83}$$

(2.82)
$$\begin{pmatrix} 0 & q_2 & q_1 & q_1 & q_0 & q_0 & q_0 & q_0 \\ q_2 & 0 & q_1 & q_1 & q_0 & q_0 & q_0 & q_0 \\ q_1 & q_1 & 0 & q_2 & q_0 & q_0 & q_0 & q_0 \\ q_1 & q_1 & q_2 & 0 & q_0 & q_0 & q_0 & q_0 \\ q_0 & q_0 & q_0 & q_0 & 0 & q_2 & q_1 & q_1 \\ q_0 & q_0 & q_0 & q_0 & q_2 & 0 & q_1 & q_1 \\ q_0 & q_0 & q_0 & q_0 & q_1 & q_1 & 0 & q_2 \\ q_0 & q_0 & q_0 & q_0 & q_1 & q_1 & q_2 & 0 \end{pmatrix}$$

FIGURE 3. 8×8 Parisi matrix with two step RSB ($n = 8, m_1 = 4, m_2 = 2$).

Following the replica method, we take the limit $n \to 0$ and now have

(2.84)
$$0 \le m_1 \le m_2 \le \cdots \le 1 \qquad \text{for } 0 \le x \le 1,$$

treating $0 \le q(x) \le 1$ as a continuous function.

2.4.2 One-step replica symmetry breaking

With our construction of the first step replica symmetry breaking, we now have the following expression

(2.85)
$$\sum_{a<b} Q_{ab}\sigma^a\sigma^b = \frac{1}{2}\left\{ q_0 \left(\sum_{a=1}^{n} \sigma^a \right)^2 + (q_1 - q_0) \sum_{\text{block}}^{n/m_1} \left(\sum_{a\in\text{block}}^{m_1} \sigma_a \right)^2 - nq_1 \right\},$$

where the first sum fills all the entries of $\{Q_{ab}\}$ as q_0, the second sum replaces all the entries q_0 in the diagonal block with q_1, and the last term removes the entries q_2 from the main diagonal and makes the diagonal entries zero [N08]. With the same approach of adding and subtracting terms, the sum with the quadratic term of Q_{ab} from (2.41) is

(2.86)
$$\lim_{n\to 0} \frac{1}{n} \sum_{a\ne b} Q_{ab}^2 = \lim_{n\to} \frac{1}{n} \left(n^2 q_0^2 - (n/m_1) m_1^2 (q_1^2 - q_0^2) - nq_1^2 \right) = m_1(q_1^2 - q_0^2) - q_1^2.$$

Substituting our expression (2.85) in the function $L(Q,m)$ (2.38), we have

(2.87)
$$L(Q_{\text{RSB}}, m_{\text{RSB}}) = \beta \sum_{a=1}^{n} (h + J_0 m)\sigma^a + \frac{\beta^2 J^2}{2} \left\{ q_0 \left(\sum_{a=1}^{n} \sigma^a \right)^2 + (q_1 - q_0) \sum_{\text{block}}^{n/m_1} \left(\sum_{a\in\text{block}}^{m_1} \sigma_a \right)^2 - nq_1 \right\}$$

We can simplify the last term using the Gaussian reduction identity (2.24) with the integration variable u, $a_1 = \sum_{a=1}^{n} \sigma^a$, and $b_1 = \beta^2 J^2 q_0$ and the integration variable v, $a_2 = \sum_{a \in \text{block}}^{m_1} \sigma_a$, and $b_2 = \beta^2 J^2 (q_1 - q_0)$.

$$
\begin{aligned}
\text{(2.88)} \quad \ln \operatorname{tr} e^{L(Q_{\text{RSB}}, m_{\text{RSB}})} &= \ln \operatorname{tr} \int_{-\infty}^{\infty} \sqrt{\frac{\beta^2 J^2 q_0}{2\pi}} du \int_{-\infty}^{\infty} \left(\prod_{\text{block}}^{n/m_1} \sqrt{\frac{\beta^2 J^2 (q_1 - q_0)}{2\pi}} \right) dv \\
&\cdot \exp \left\{ -\beta^2 J^2 q_0 u^2/2 + \left(\sum_{a=1}^{n} \sigma^a \right) \beta^2 J^2 q_0 u + \beta \sum_{a=1}^{n} (h + J_0 m) \sigma^a - \frac{n}{2} \beta^2 J^2 q_1 \right\} \\
&\cdot \exp \left\{ \sum_{\text{block}}^{n/m_1} \left(-\beta^2 J^2 (q_1 - q_0) v^2/2 + \left(\sum_{a \in \text{block}}^{m_1} \sigma^a \right) \beta^2 J^2 (q_1 - q_0) v \right) \right\} \\
\text{(2.89)} \quad &= -\frac{n}{2} \beta^2 J^2 q_1 + \ln 2 + \frac{1}{m_1} \int_{-\infty}^{\infty} \mathbf{D}u \left\{ \ln \int_{-\infty}^{\infty} \mathbf{D}v \cosh^{m_1} \eta(u, v) \right\},
\end{aligned}
$$

where the derivation utilizes the same methods as was used in the RS ansatz.

Definition 2.90. We define the function

$$
\eta(u, v) = \beta \left(J \sqrt{q_0} u + J \sqrt{q_1 - q_0} v + J_0 m + h \right). \tag{2.91}
$$

Using our two expressions above (2.86) and (2.89) for free energy (2.41), we have

(2.92)

$$
-\beta f_{1\text{RSB}} = -\frac{\beta J_0}{2} m^2 - \frac{\beta^2 J^2}{4} \left(1 - m_1 (q_1^2 - q_0^2) + q_1^2 - 2q_1 \right) + \ln 2 + \frac{1}{m_1} \int \mathbf{D}u \ln \left(\int \mathbf{D}v \cosh^{m_1} \eta(u, v) \right),
$$

where we let $m_a = m$ be replica symmetric.

The parameters m, m_1, q_0, and q_1 are all between 0 and 1. Solving the extremization conditions of free energy (2.92) with respect to m, q_0, and q_1 gives the following [N08]

$$
m = \int \mathbf{D}u \frac{\int \mathbf{D}v \cosh^{m_1} \eta(u, v) \tanh \eta(u, v)}{\int \mathbf{D}v \cosh^{m_1} \eta(u, v)} \tag{2.93}
$$

$$
q_0 = \int \mathbf{D}u \left(\frac{\int \mathbf{D}v \cosh^{m_1} \eta(u, v) \tanh \eta(u, v)}{\int \mathbf{D}v \cosh^{m_1} \eta(u, v)} \right)^2 \tag{2.94}
$$

$$
q_1 = \int \mathbf{D}u \frac{\int \mathbf{D}v \cosh^{m_1} \eta(u, v) \tanh^2 \eta(u, v)}{\int \mathbf{D}v \cosh^{m_1} \eta(u, v)}. \tag{2.95}
$$

When $J_0 = h = 0$, $\nu(u, v) = \beta(J\sqrt{q_0}u + J\sqrt{q_1 - q_0}v)$ is an odd function of u and v, which implies $m = 0$ from (2.93). The order parameter q_1 can be positive when $T < T_f = J$ because the first term in the expansion (2.95) is $\beta^2 J^2 q$ [N08]. Thus, the RS and 1RSB ansatz result in the same transition point.

2.5 Full replica symmetry breaking solution

We have that the entropy per spin with $J_0 = 0$ and $T = 0$ changes from $-\frac{1}{2\pi} \approx -0.16$ in the RS solution to approximately -0.01 in the 1RSB solution [N08]. This seems to indicate that the symmetry breaking ansatz leads us towards a stable solution and that we should make further steps in our replica symmetry breaking. Thus, we present the expression free energy (2.41) using a full replica symmetry breaking solution (FRSB) found in [N08] and the interested reader should refer to Appendix B of [N08] for a complete derivation. For simplicity, we will only consider the case $J_0 = 0$. We have the following expression for the K-step RSB (K-RSB).

$$\sum_{a\neq b} q_{ab}^l = q_0^l n^2 + (q_1^l - q_0^l)m_1^2 \cdot \frac{n}{m_1} + (q_2^l - q_1^l)m_2^2 \cdot \frac{m_1}{m_2} \cdot \frac{n}{m_1} + \cdots - q_K^l \cdot n \tag{2.96}$$

$$= n\sum_{j=0}^{K}(m_j - m_{j+1})q_j^l, \tag{2.97}$$

where l is some integer, $m_0 = n$, and $m_{K+1} = 1$. As $n \to 0$, we can replace $m_j - m_{j+1} \to -dx$, giving us

$$\frac{1}{n}\sum_{a\neq b} q_{ab}^l \to \int_0^1 q^l(x)dx. \tag{2.98}$$

The internal energy for $J_0 = 0, h = 0$ can be evaluated by the partial derivative with respect to β of (2.41)

$$U = \frac{\partial}{\partial\beta}(\beta f) = -\frac{\beta J^2}{2}\left(1 - \frac{2}{n}\sum_{a<b} Q_{ab}^2\right) = -\frac{\beta J^2}{2}\left(1 + \int_0^1 q^2(x)dx\right). \tag{2.99}$$

The magnetic susceptibility can be evaluated by the second partial derivative with respect to h of (2.41) as

$$\chi = \beta\left(1 + \frac{1}{n}\sum_{a\neq b} Q_{ab}\right) \to \beta\left(1 - \int_0^1 q(x)dx\right). \tag{2.100}$$

The expression for free energy [N08] is given by

$$\beta f = -\frac{\beta^2 J^2}{4}\left\{1 + \int_0^1 q(x)dx - 2q(1)\right\} - \int \mathbf{D}u f_0(0, \sqrt{q(0)}u), \tag{2.101}$$

where f_0, with the initial condition $f_0(1, h) = \ln(2\cosh(\beta h))$ satisfies the Parisi equation

$$\frac{\partial f_0(x,h)}{\partial x} = -\frac{J^2}{2}\frac{dq}{dx}\left\{\frac{\partial^2 f_0}{\partial h^2} + x\left(\frac{\partial f_0}{\partial h}\right)^2\right\}. \tag{2.102}$$

2.6 Possible issues with the replica trick

As we have discussed previously, to evaluate $\mathbb{E}\ln Z$, we employ the replica method (1.15). We first take n to be integers to calculate the nth moment $\mathbb{E}Z^n$ and proceed to take the limit as $n \to 0$. A natural complication arises as the two steps are not compatible with each other.

For the replica trick to hold, a sufficient condition of uniqueness of the moment problem is given by Carleman (stated without proof).

Theorem 2.103. *Let Z be a real random variable with moments $m_n = \mathbb{E}Z^n$ such that*

$$\sum_{n=1}^{\infty} m_{2n}^{-1/(2n)} = \infty. \tag{2.104}$$

Then, any random variable with the same moments as Z is identically distributed to Z.

So, if the sum in (2.104) converges, then the distribution of Z is not necessarily determined by its integer moments.

However, the SK model does not satisfy Carleman's condition and there is yet to be a formal mathematical justification for the use of the replica trick in solving models in statistical mechanics like the SK model [Ta07].

In deriving the expression for free energy (2.41), we interchanged the limits $n \to 0$ and $N \to \infty$ as such

$$\lim_{N\to\infty} \frac{\mathbb{E}\ln Z}{N} = \lim_{N\to\infty}\lim_{n\to 0} \frac{\ln \mathbb{E}Z^n}{Nn} \tag{2.105}$$

$$= \lim_{n\to 0} \frac{1}{n}\left(\lim_{N\to\infty} \frac{\ln \mathbb{E}Z^n}{N}\right) \tag{2.106}$$

to apply the method of steepest descent.

Hemmen and Palmer prove the validity of interchanging the limits for the SK model and find that the complications with SK model lies only in taking the moments from integers to real n [HP78].

3 Satisfiability

From Section 1, we found that the threshold for the 2-SAT is $\alpha^* = 1$ using the first moment method [FP83]. It is known that the K-SAT problem is NP-complete for $K \geq 3$ [GJ79], which may explain the different behavior of the $K = 2$ and the $K \geq 3$ problems. Following the construction of Monasson and Zecchina [MZ97], we minimize the energy of the random K-SAT with respect to the number of unsatisfied clauses. Let the Ising spin σ_i take value $+1$ if the logical variable $x_i = \texttt{TRUE}$ and value -1 if the logical variable $x_i = \texttt{FALSE}$. Let the coupling random variable C_{li} be $+1$ if the clause C_l includes x_i, -1 if it includes $\overline{x}_i$, and 0 if the clause includes neither x_i nor $\overline{x}_i$. Since we have a clause length of K, we choose K non-zero C_{li} from $\{C_{l1}, \cdots, C_{lN}\}$ and assign ± 1 randomly for these K random variables. So, we have that

$$C_{li}\sigma_i = \begin{cases} +1 & \text{if } \sigma_i = C_{li}, \\ 0 & \text{if } C_{li} = 0, \\ -1 & \text{if } \sigma_i \neq C_{li}. \end{cases} \tag{3.1}$$

Thus, the K-SAT problem is satisfiable if there exists an assignment of x_i's (and thus a configuration $\underline{\sigma}$) such that there is at least one σ_i for which $C_{li}\sigma_i = 1$ for every $l \in \{1, \cdots, M\}$. Since a clause contains only K logical variables, the minimum value of $\sum_{i=1}^{N} C_{li}\sigma_i = -K$, which is coincidentally the only case that the clause C_l is unsatisfied. So, if $\sum_{i=1}^{N} C_{li}\sigma_i > -K$, then there exists at least one σ_i such that $C_{li}\sigma_i = 1$ and the clause C_l is satisfied.

Definition 3.2. The energy function (or Hamiltonian) of the K-SAT problem is given by

$$E(\underline{\sigma}) = \sum_{l=1}^{M} \delta\left(\sum_{i=1}^{N} C_{li}\sigma_i, -K\right), \tag{3.3}$$

where $\delta(i, j)$ is the Kronecker delta function with $\delta(i, j) = 1$ if $i = j$ and $\delta(i, j) = 0$ if $i \neq j$. From this construction, we see that the ground-state energy level $E(\underline{\sigma}) = 0$ implies that the problem is satisfiable.

3.1 The replica derivation

Note that an instance of the satisfiability problem is given by one of the values of its $\left(\binom{N}{K} 2^K\right)^M$ energy levels, E_i, $1 \leq i \leq \binom{N}{K}^M 2^{KM}$, and from (1.5), the partition function is

$$Z = \sum_i \exp\left(-\beta E_i\right). \tag{3.4}$$

Assuming that n is an integer, the first step of the replica trick involves finding an expression for Z^n

$$Z^n = \prod_{a=1}^{n} \left(\sum_{i_a} \exp\left(-\beta E_{i_a}\right) \right) = \sum_{i_1} \cdots \sum_{i_n} \exp\left(-\beta (E_{i_1} + \cdots + E_{i_n})\right) \tag{3.5}$$

$$= \sum_{\{\sigma_i^a\}} \exp\left(-\beta \left(\sum_{a=1}^{n} E_{i_a} \right) \right), \tag{3.6}$$

which is the partition function of a new system with configuration given by the n-tuple $(i_1, \cdots, i_n)$ with $i_a \in \left\{ 1, \cdots, \binom{N}{K}^M 2^{KM} \right\}$ and the sum $\sum_{\{\sigma_i^a\}}$ is over the $\binom{N}{K}^{Mn} 2^{KMn}$ different configurations. This sum will henceforth shortened by the trace representation Tr. Using our expression for the energy function (3.3) gives us

$$Z^n = \operatorname{Tr} \exp \left\{ -\beta \left(-\sum_{a=1}^{n} \left[\sum_{l=1}^{M} \delta \left(\sum_{i=1}^{N} C_{li} \sigma_i^a, -K \right) \right] \right) \right\} \tag{3.7}$$

$$= \operatorname{Tr} \prod_{l=1}^{M} \exp \left\{ -\beta \left[-\sum_{a=1}^{n} \delta \left(\sum_{i=1}^{N} C_{li} \sigma_i^a, -K \right) \right] \right\}. \tag{3.8}$$

3.1.1 Replica average of the partition function

To proceed with our replica calculation, we take the expectation of the new partition function (3.8):

$$\mathbb{E} Z^n = \operatorname{Tr} \left(\zeta_K(\underline{\sigma})^M \right) \tag{3.9}$$

where used the linearity of expectation and the independence of the l clauses and define $\zeta_K(\underline{\sigma})$ below.

Definition 3.10. As we will be working with the expectation in (3.9), we define the expression as such

$$\zeta_K(\underline{\sigma}) := \mathbb{E} \exp \left\{ -\beta \left[-\sum_{a=1}^{n} \delta \left(\sum_{i=1}^{N} C_i \sigma_i^a, -K \right) \right] \right\}. \tag{3.11}$$

From (3.1), we can express the Kronecker delta function in (3.13) by a product of Kronecker delta functions

$$\delta \left(\sum_{i=1}^{N} C_i \sigma_i^a, -K \right) = \prod_{i=1, C_i \neq 0}^{N} \delta\left(\sigma_i^a, -C_i\right), \tag{3.12}$$

where the product is only over i such that C_i takes a non-zero value.

So, substituting (3.12) into (3.13) and taking the configurational average over the C_i's give us

$$\zeta_K(\underline{\sigma}) = \frac{1}{2^K} \sum_{Ci_1 = \pm 1} \cdots \sum_{C_{i_K} = \pm 1} \frac{1}{N^K} \sum_{i_1=1}^{N} \cdots \sum_{i_K=1}^{N} \exp \left\{ -\beta \left[-\sum_{a=1}^{n} \prod_{k=1}^{K} \delta\left(\sigma_{i_k}^a, -C_{i_k}\right) \right] \right\}, \tag{3.13}$$

where we ignore the term $O(N^{-1})$ as it goes to zero in the large N limit.

Definition 3.14. We can rewrite the expression for $\zeta_K(\underline{\sigma})$ (3.13) in terms of $\{c(\underline{s})\}_{\underline{s}}$, which is the set of the number of sites with a specific pattern in the replica space $\underline{s} = \{s^1, \cdots s^n\}$:

$$Nc(\underline{s}) := \sum_{i=1}^{N} \prod_{a=1}^{n} \delta\left(\sigma_i^a, s^a\right). \tag{3.15}$$

$\zeta_K(\underline{\sigma})$ (3.13) only depends on $\underline{\sigma}$ only through $\{c(\underline{s})\}$ since if we choose $-C_{i_k}\sigma_{i_k}^a = s_k^a$.

Note from (3.1) that $s_k^a = -C_{i_k}\sigma_{i_k} = 1$ iff $C_{i_k} \neq \sigma_{i_k}$ and so $\delta\left(s_k^a, 1\right) = \delta\left(\sigma_k^a, -C_k\right)$ over nonzero C_k. So, (3.13) can be rewritten as

$$\zeta_K(\underline{\sigma}) = \zeta_k(\{c\}) = \frac{1}{2^K} \sum_{C_1=\pm1} \cdots \sum_{C_K=\pm1} \sum_{\underline{s}_1} \cdots \sum_{\underline{s}_K} c\left(-C_1\underline{s}_1\right) \cdots c\left(-C_K\underline{s}_K\right) \exp\left\{-\beta \sum_{a=1}^{n} \prod_{k=1}^{K} \delta\left(s_k^a, 1\right)\right\}. \tag{3.16}$$

Since $\sigma_i^a s_i^a = 1$ iff $\delta\left(\sigma_i^a, s_i^a\right) = 1$ and $\sigma_i^a s_i^a = -1$ iff $\delta\left(\sigma_i^a, s_i^a\right) = 0$, we have

$$c(\underline{s}) = \frac{1}{N} \sum_{i=1}^{N} \prod_{a=1}^{n} \left(\frac{1 + \sigma_i^a s_i^a}{2}\right) \tag{3.17}$$

$$= \frac{1}{2^n} \left(\frac{1}{N} \sum_{i=1}^{N} 1 + \frac{1}{N} \sum_{a=1}^{n} Q^a s^a + \frac{1}{N} \sum_{a<b} Q^{ab} s^a s^b + \cdots + \frac{1}{N} \sum_{a<b<\cdots<n} Q^{ab\cdots n} s^a \cdots s^n\right). \tag{3.18}$$

Definition 3.19. We define the multioverlaps $Q^{ab\cdots}$ as

$$Q^{ab\cdots} = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^a \sigma_i^b \cdots, \tag{3.20}$$

where we assume that the multioverlaps with an odd number of indices $Q^a = Q^{abc} = \cdots = 0$ [MZ97].

Remark 3.21. *We have the symmetry $c(\underline{s}) = c(-\underline{s})$ because the multioverlaps with odd number of indices are zero. Thus, since C_i takes only values ±1, we may remove C_i from $c(-C_i\underline{s}_i)$ in the expression* (3.16)*.*

Remark 3.22. *The average partition function* (3.9) *can now be simplified as*

$$\mathbb{E}Z^n = \int_0^1 \prod_{\underline{s}} dc(\underline{s}) e^{-NE_0(\{c\})} \mathrm{Tr} \prod_{\underline{s}} \delta\left\{c(\underline{s}) - \frac{1}{N} \sum_{i=1}^{N} \prod_{a=1}^{n} \delta\left(\sigma_i^a, s^a\right)\right\}, \tag{3.23}$$

$$E_0\left(\{c\}\right) := -\alpha \ln\left\{\sum_{\underline{s}_1,\cdots,\underline{s}_K} c(\underline{s}_1) \cdots c(\underline{s}_K) \cdot \prod_{a=1}^{n} \left(1 + \left(e^{-\beta} - 1\right) \prod_{k=1}^{K} \delta\left(s_k^a, 1\right)\right)\right\}, \tag{3.24}$$

where $\alpha = M/N$ is the clause density.

The first term in (3.23) is a straightforward result of algebraic manipulation

$$\exp\left(-NE_0(\{c\})\right) = \exp\left(N\alpha \ln\left\{\sum_{\underline{s}_1,\cdots,\underline{s}_K} c(\underline{s}_1)\cdots c(\underline{s}_K)\cdot\prod_{a=1}^{n}\left(1+\left(e^{-\beta}-1\right)\prod_{k=1}^{K}\delta\left(s_k^a,1\right)\right)\right\}\right) \tag{3.25}$$

$$= \left\{\sum_{\underline{s}_1,\cdots,\underline{s}_K} c(\underline{s}_1)\cdots c(\underline{s}_K)\cdot\prod_{a=1}^{n}\left(1+\left(e^{-\beta}-1\right)\prod_{k=1}^{K}\delta\left(s_k^a,1\right)\right)\right\}^M \tag{3.26}$$

$$\overset{(\dagger)}{=} \left\{\sum_{\underline{s}_1,\cdots,\underline{s}_K} c(\underline{s}_1)\cdots c(\underline{s}_K)\cdot\exp\left(-\beta\sum_{a=1}^{n}\prod_{k=1}^{K}\delta\left(s_k^a,1\right)\right)\right\}^M, \tag{3.27}$$

where in (†), we use the multiplicative property of exponentiation and that

$$\exp\left(-\beta\prod_{k=1}^{K}\delta\left(s_k^a,1\right)\right) = \begin{cases} e^{-\beta} & \text{when } \prod_{k=1}^{K}\delta\left(s_k^a,1\right)=1, \\ 1 & \text{when } \prod_{k=1}^{K}\delta\left(s_k^a,1\right)=0. \end{cases} \tag{3.28}$$

The second term in (3.23) applies the constraint

$$\sum_{\underline{s}} c(\underline{s}) = 1 \tag{3.29}$$

in the form of the Dirac delta, using the definition of $c(\underline{s})$ in 3.14.

Remark 3.30. *We will now simplify the second term in* (3.23). *Applying the trace operation over the spin variables* $\underline{s}$ *gives the entropy*

$$\mathrm{Tr}\prod_{\underline{s}}\delta\left\{c(\underline{s})-\frac{1}{N}\sum_{i=1}^{N}\prod_{a=1}^{n}\delta\left(\sigma_i^a,s^a\right)\right\} = \frac{N!}{\prod_{\underline{s}}(Nc(\underline{s}))!}. \tag{3.31}$$

Using the following approximation by Stirling

$$\ln(n!) \sim n \ln n - n, \tag{3.32}$$

yields

$$\ln\left(\frac{N!}{\prod\limits_{\underline{s}}(Nc(\underline{s}))!}\right) \sim \ln\left(N^N\right) - N - \sum_{\underline{s}} \ln\left[(Nc(\underline{s}))^{Nc(\underline{s})}\right] + \sum_{\underline{s}} Nc(\underline{s}) \tag{3.33}$$

$$\stackrel{(\dagger)}{=} \ln\left(N^N\right) - \left(\sum_{\underline{s}} Nc(\underline{s}) \ln N + \ln\left[c(\underline{s})^{Nc(\underline{s})}\right]\right) \tag{3.34}$$

$$\stackrel{(\ddagger)}{=} -\sum_{\underline{s}} \ln[c(\underline{s})^{Nc(\underline{s})}], \tag{3.35}$$

where in (†) and (‡) we used that $\sum_{\underline{s}} c(\underline{s}) = 1$ since $c(\underline{s})$ is the fraction of sites i, among the N possible indices, such that $\sigma_i^a = s^a$ for all $a \in \{1, \cdots, n\}$.

With a little algebraic manipulation, we have that

$$-\sum_{\underline{s}} \ln[c(\underline{s})^{Nc(\underline{s})}] = \ln\left(\prod_{\underline{s}} c(\underline{s})^{-Nc(\underline{s})}\right) \tag{3.36}$$

$$= -N \ln\left(\prod_{\underline{s}} c(\underline{s})^{c(\underline{s})}\right) \tag{3.37}$$

$$= -N \sum_{\underline{s}} c(\underline{s}) \ln c(\underline{s}). \tag{3.38}$$

Thus, our average partition function is now

$$\mathbb{E}Z^n = \int_0^1 \prod_{\underline{s}} dc(\underline{s}) \exp\left\{-N\left(E_0(\{c\}) + \sum_{\underline{s}} c(\underline{s}) \ln c(\underline{s})\right)\right\} \tag{3.39}$$

Applying the method of steepest descent (2.35) to the integral above (3.39)

$$\mathbb{E}Z^n \approx \exp\left\{-N\left(E_0(\{c\}) + \sum_{\underline{s}} c(\underline{s}) \ln c(\underline{s})\right)\right\} \tag{3.40}$$

$$\stackrel{(\dagger)}{\approx} 1 - NE_0(\{c\}) - N\sum_{\underline{s}} c(\underline{s}) \ln c(\underline{s}), \tag{3.41}$$

where in (†) we use the approximation $e^{\lambda x} \approx 1 + \lambda x$ for small x.

With the definition of free entropy (1.8) and the replica identity (1.15), we have the following expression for configurational average of free entropy in the limit $N, M \to \infty$

$$-\beta[f] = -\frac{\beta f}{N} = \frac{\mathbb{E}Z^n - 1}{N} \tag{3.42}$$

$$= -E_0\left(\{c\}\right) + \sum_{\underline{s}} c(\underline{s}) \ln c(\underline{s}). \tag{3.43}$$

3.2 Replica-symmetric solution

The free entropy in (3.43) is extremized with respect to $c(\underline{s})$ and similarly to the RS solution in the SK model, we assume $c(\underline{s})$ is symmetric under permutation of $s^1, \cdots, s^n$ to obtain the RS solution. We can express $c(\underline{s})$ in terms of the distribution function of the local magnetization $P(m)$ as such

$$c(\underline{s}) = \int_{-1}^{1} dm P(m) \prod_{a=1}^{n} \frac{1 + m s^a}{2}. \tag{3.44}$$

Such a $c(\underline{s})$ is replica-symmetric.

We show the following two remarks in the Appendix 4.1.

Remark 3.45. *The extremization condition of the free energy* (3.43) *gives the following equation for* $P(m)$

$$P(m) = \frac{1}{2\pi(1-m^2)} \int_{-\infty}^{\infty} du \cos\left(\frac{u}{2} \ln\left(\frac{1+m}{1-m}\right)\right) \cdot \exp\left\{-\alpha K + \alpha K \int_{-1}^{1} \prod_{k=1}^{K-1} dm_k P(m_k) \cos\left(\frac{u}{2} \ln A_{K-1}\right)\right\}, \tag{3.46}$$

$$A_{K-1} = 1 + \left(e^{-\beta} - 1\right) \prod_{k=1}^{K-1} \frac{1 + m_k}{2}. \tag{3.47}$$

Remark 3.48. *The free energy is expressed in terms of* $P(m)$ *as*

$$-\frac{\beta f}{Nn} = \ln 2 + \alpha(1-K) \int_{-1}^{1} \prod_{k=1}^{K} dm_k P(m_k) \ln A_k + \frac{\alpha K}{2} \int_{-1}^{1} \prod_{k=1}^{K-1} dm_k P(m_k) \ln(A_{K-1}) - \frac{1}{2} \int_{-1}^{1} dm P(m) \ln\left(1 - m^2\right). \tag{3.49}$$

3.2.1 The $K = 1$ case

When $K = 1$, $A_0 = e^{-\beta}$ and (3.46) gives

$$P(m) = \frac{1}{2\pi(1-m^2)} \int_{-\infty}^{\infty} du \cos\left(\frac{u}{2}\ln\left(\frac{1+m}{1-m}\right)\right) e^{-\alpha+\alpha\cos(u\beta/2)}. \tag{3.50}$$

Lemma 3.51. *We can express the exponential cosine in* (3.50) *using the modified Bessel functions*

$$e^{z\cos\theta} = \sum_{k=-\infty}^{\infty} I_k(z)\cos(k\theta), \tag{3.52}$$

where the kth modified Bessel function

$$I_k(z) = \left(\frac{z}{2}\right)^k \sum_{n=0}^{\infty} \frac{\left(\frac{z^2}{4}\right)^n}{n!\Gamma(k+n+1)}. \tag{3.53}$$

which gives us

$$P(m) = \frac{e^{-\alpha}}{2\pi(1-m^2)} \int_{-\infty}^{\infty} du \cos\left(\frac{u}{2}\ln\left(\frac{1+m}{1-m}\right)\right) \sum_{k=-\infty}^{\infty} I_k(\alpha)\cos\left(ku\beta/2\right). \tag{3.54}$$

Lemma 3.55. *Using the exponential relation of cosine*

$$\cos(\theta) = \frac{e^{i\theta}+e^{-i\theta}}{2}, \tag{3.56}$$

and the the Fourier transform of $e^{i\omega_0 u}$

$$\int_{-\infty}^{\infty} e^{iu(\omega-\omega_0)}du = 2\pi\delta(\omega-\omega_0), \tag{3.57}$$

along with the additivity of integrals, gives us

(3.58)

$$P(m) = \frac{e^{-\alpha}}{2\pi(1-m^2)} \sum_{k=-\infty}^{\infty} I_k(\alpha)/4 \int_{-\infty}^{\infty} du$$
$$\left(e^{iu/2\left[\ln\left(\frac{1+m}{1-m}\right)+k\beta\right]} + e^{-iu/2\left[\ln\left(\frac{1+m}{1-m}\right)-k\beta\right]} + e^{iu/2\left[\ln\left(\frac{1+m}{1-m}\right)-k\beta\right]} + e^{-iu/2\left[\ln\left(\frac{1+m}{1-m}\right)+k\beta\right]} \right)$$

(3.59)

$$= \frac{e^{-\alpha}}{2\pi(1-m^2)} \sum_{k=-\infty}^{\infty} I_k(\alpha)/2 \cdot 2\pi \left[\delta\left(\left[\ln\left(\frac{1+m}{1-m}\right)+k\beta\right]/2\right) + \delta\left(-\left[\ln\left(\frac{1+m}{1-m}\right)-k\beta\right]/2\right) + \delta\left(\left[\ln\left(\frac{1+m}{1-m}\right)-k\beta\right]/2\right) + \delta\left(-\left[\ln\left(\frac{1+m}{1-m}\right)+k\beta\right]/2\right) \right]$$

(3.60)

$$\overset{(\dagger)}{=} \frac{e^{-\alpha}}{(1-m^2)} \sum_{k=-\infty}^{\infty} I_k(\alpha) \left[\delta\left(\ln\left(\frac{1+m}{1-m}\right)+k\beta\right) + \delta\left(\ln\left(\frac{1+m}{1-m}\right)-k\beta\right) \right],$$

where, in (†), we used the scaling property of the Dirac delta for non-zero scalar c

$$\delta(cx) = \frac{\delta(x)}{|c|}. \tag{3.61}$$

Lemma 3.62. *Composing the Dirac delta with a smooth function g, with g' nowhere zero, gives the following identity*

$$\delta\left(g(x)\right) = \frac{\delta(x-x_0)}{|g'(x_0)|}, \tag{3.63}$$

where x_0 is the real root of the function g.

Corollary 3.64. *A corollary of the above lemma is the following*

$$\delta\left(x^2-a^2\right) = \frac{1}{2|a|}\left(\delta(x+a)+\delta(x-a)\right) \tag{3.65}$$

Since $\tanh\left(\frac{x}{2}\right) = \frac{e^x-1}{e^x+1}$, using lemma 3.62 and 3.64 in equation (3.60) yields

$$P(m) = e^{-\alpha} \sum_{k=-\infty}^{\infty} I_k(\alpha)\delta\left(m - \tanh\frac{\beta k}{2}\right). \tag{3.66}$$

Lemma 3.67. *The summation of the kth modified Bessel function has the following identity*

$$\sum_{k=1}^{\infty} I_k(z) = \frac{1}{2}\left(e^z - I_0(z)\right) \tag{3.68}$$

Since $\tanh(x) \to \pm 1$ as $x \to \pm\infty$, as temperature T goes to zero ($\beta = 1/T \to \infty$), using lemma 3.67, the local magnetization (3.66) becomes

$$P(m) = e^{-\alpha} I_0(\alpha)\delta(m) + \frac{1}{2}\left[1 - e^{-\alpha} I_0(\alpha)\right]\left[\delta(m-1) + \delta(m+1)\right]. \tag{3.69}$$

Substituting our expression for (3.66) into (3.49) for $K = 1$ gives us

$$-\frac{\beta F}{Nn} = \ln 2 + \frac{\alpha}{2}\int_{-1}^{1} dm P(m) \ln e^{-\beta} - \frac{1}{2}\int_{-1}^{1} dm P(m) \ln\left(1 - m^2\right) \tag{3.70}$$

$$= \ln 2 - \frac{\alpha\beta}{2} - \frac{1}{2}e^{-\alpha}\sum_{k=-\infty}^{\infty} I_k(\alpha)\int_{-1}^{1} dm\, \delta\left(m - \tanh\frac{\beta k}{2}\right)\ln\left(1 - m^2\right). \tag{3.71}$$

Lemma 3.72. *We use the identity*

$$\int_{-1}^{1} \delta(x-a)\ln\left(1-x^2\right)dx = H(1-a)H(a+1)\ln\left(1-a^2\right), \tag{3.73}$$

where $H(x) = \int_{-\infty}^{x} \delta(\xi)d\xi$ *is the Heaviside function, to derive,*

$$-\frac{\beta F}{Nn} = \ln 2 - \frac{\alpha\beta}{2} - \frac{1}{2}e^{-\alpha}\sum_{k=-\infty}^{\infty} I_k(\alpha)\ln\left(1 - \tanh^2\frac{\beta k}{2}\right) \tag{3.74}$$

$$= \ln 2 - \frac{\alpha\beta}{2} - \frac{1}{2}e^{-\alpha}\sum_{k=-\infty}^{\infty} I_k(\alpha)\ln\left(1 - \tanh^2\frac{\beta k}{2}\right) \tag{3.75}$$

$$= \ln 2 - \frac{\alpha\beta}{2} + e^{-\alpha}\sum_{k=-\infty}^{\infty} I_k(\alpha)\ln\cosh\frac{\beta k}{2}, \tag{3.76}$$

which, as $\beta \to \infty$, yields

$$\frac{E(\alpha)}{N} = \lim_{\beta\to\infty} -\frac{\ln 2}{\beta} + \frac{\alpha}{2} - e^{-\alpha}\sum_{k=-\infty}^{\infty} I_k(\alpha)\frac{\ln\cosh\frac{\beta k}{2}}{\beta} \tag{3.77}$$

$$= \frac{\alpha}{2} - e^{-\alpha}\sum_{k=-\infty}^{\infty} I_k(\alpha)\lim_{\beta\to\infty}\frac{\ln\cosh\frac{\beta k}{2}}{\beta} \tag{3.78}$$

$$= \frac{\alpha}{2} - e^{-\alpha}\left(\sum_{k=1}^{\infty}\frac{k}{2}I_k(\alpha) - \sum_{k=-1}^{-\infty}\frac{k}{2}I_k(\alpha)\right), \tag{3.79}$$

where the limit can be evaluated by L'Hôpital's rule

$$\lim_{\beta\to\infty} \frac{\ln\cosh\dfrac{\beta k}{2}}{\beta} = \lim_{\beta\to\infty} \frac{\dfrac{k}{2}\sinh\dfrac{\beta k}{2}}{\cosh\dfrac{\beta k}{2}} = \pm\frac{k}{2}, \tag{3.80}$$

where we have $k/2$ when $k \geq 0$ and $-k/2$ when $k < 0$.

Lemma 3.81. *Using the definition of the Bessel function* (3.53)*, we have for integer n that*

$$I_{-n}(z) = I_n(z). \tag{3.82}$$

Lemma 3.83. *We also have the following relationship involving the neighboring terms for any real ν*

$$\nu I_\nu(z) = \frac{z}{2}\left(I_{\nu-1}(z) - I_{\nu+1}(z)\right), \tag{3.84}$$

Using lemmas 3.81 and 3.83 in (3.79), we have

$$\frac{E(\alpha)}{N} = \frac{\alpha}{2} - e^{-\alpha}\left(\sum_{k=1}^{\infty}\frac{k}{2}I_k(\alpha) + \sum_{k=1}^{\infty}\frac{k}{2}I_k(\alpha)\right) \tag{3.85}$$

$$= \frac{\alpha}{2} - e^{-\alpha}\left(\sum_{k=1}^{\infty}\frac{\alpha}{2}\left(I_{k-1}(\alpha) - I_{k+1}(\alpha)\right)\right) \tag{3.86}$$

$$= \frac{\alpha}{2} - \frac{\alpha}{2}e^{-\alpha}\left(I_0(\alpha) + I_1(\alpha)\right), \tag{3.87}$$

where the telescoping sum

$$\sum_{k=1}^{n}\frac{z}{2}\left(I_{k-1}(z) - I_{k+1}(z)\right) = \frac{z}{2}\left(I_0(z) + I_1(z) - I_n(z) - I_{n+1}(z)\right) \longrightarrow \frac{z}{2}\left(I_0(z) + I_1(z)\right) \tag{3.88}$$

for z sufficiently small.

Since $\alpha > 0$, the ground-state energy is always positive. So, for $K = 1$ the SAT problem is unsatisfiable with probability 1.

3.2.2 The $K \geq 2$ case

As we found previously in Theorem 1.27, the critical point for $K = 2$ is $\alpha_{\text{sat}}^{(N)}(2) = 1$. Comparing the results of numerical simulations to the RS solutions for $K \geq 2$ indicate that the RS theory is correct for $\alpha < \alpha_{\text{sat}}^{(N)}(2)$ but not for $\alpha > \alpha_{\text{sat}}^{(N)}(2)$. A further discussion of the $K \geq 2$ case can be found in [MZ97].

3.3 Satisfiability threshold for random K-SAT

From above in Theorem 1.27, we have shown a proof by bicycles for the 2-SAT [Go96]. The 2-SAT problem is unsurprisingly much simpler to analyze than the K-SAT problem for $K \geq 3$. Indeed, in terms of computational complexity, the 2-SAT problem is polynomial while the K-SAT problem is NP-complete for $K \geq 3$. As discussed previously, numerical simulations suggest that the random K-SAT has a phase transition between the SAT and UNSAT phases for any $K \geq 2$.

Friedgut's theorem 3.91, restated below, states that there exists a sharp threshold *sequence* $\alpha_{\text{sat}}^{(N)}(K)$ for $K \geq 2$. An important note is that the theorem does not state whether $\alpha_{\text{sat}}^{(N)}(K)$ converges to a unique limit, which is what Conjecture 3.89, restated below, requires.

Conjecture 3.89 (Satisfiability threshold conjecture). *For any $K \geq 2$, there exists a threshold $\alpha_{\text{sat}}(K)$ with*

$$\lim_{N\to\infty} P_N(K,\alpha) = \begin{cases} 1 & \textit{if } \alpha < \alpha_{sat}(K), \\ 0 & \textit{if } \alpha > \alpha_{sat}(K). \end{cases} \tag{3.90}$$

As mentioned previously, the conjecture has been proven for $K = 2$. The theorem below from Friedgut [Fr99] strongly supports the case for the conjecture and all that remains to prove the satisfiability threshold conjecture is that $\alpha_{\text{sat}}^{(N)}(K) \to \alpha_{\text{sat}}(K)$ as $N \to \infty$.

Theorem 3.91 (Friedgut's Theorem). *There exists a sequence of $\alpha_{\text{sat}}^{(N)}(K)$ such that, for any $\varepsilon > 0$,*

$$\lim_{N\to\infty} P_N(K,\alpha_N) = \begin{cases} 1 & \textit{if } \alpha_N < \alpha_{\text{sat}}^{(N)}(K) - \varepsilon, \\ 0 & \textit{if } \alpha_N > \alpha_{\text{sat}}^{(N)}(K) + \varepsilon. \end{cases} \tag{3.92}$$

3.3.1 Upper bound

With the moment method (1.31), we have the upper bound (1.33) given by Theorem 1.34. The upper bound from [FP83] is not sharp, but is actually quite close to the more precise upper bound found by [KKKS98] presented below.

Theorem 3.93 (KKKS98). *Let ϵ_K denote an error term that decays to zero as $K \to \infty$. Then,*

$$\limsup_{N\to\infty} \alpha_{\text{sat}}^{(N)}(K) \leq 2^K \ln 2 - \frac{1}{2}(1 + \ln 2) + \epsilon_K. \tag{3.94}$$

The upper bound is achieved by truncating the first moment to locally maximal solution. The following is an outline of the result from [KKKS98].

Let $\mathscr{A}_n$ be the class of all truth assignments and $\mathscr{P}_n$ be the random class of truth assignments that satisfy a random formula ϕ. We define a class smaller than $\mathscr{P}_n$ as follows.

Definition 3.95. For a random formula ϕ, $\mathscr{P}_n^\sharp$ is defined the random class of truth assignments A such that: (i) $A \vDash \phi$ and (ii) any assignment obtained from A by changing exactly one `FALSE` value of A to `TRUE` does not satisfy ϕ. Such a construction is known as "single flips."

Remark 3.96. *Here, $\vDash$ is the symbol for entailment and we say that a formula ϕ is a semantic consequence within some system of a set of statements A ($A \vDash \phi$) if and only if every model which makes members of A true makes ϕ true.*

Consider the lexicographic ordering among truth assignments in which the value `FALSE` is considered smaller than the value `TRUE` and the values of variables with higher index are of lower priority in establishing the way two assignments compare. We see that $\mathscr{P}_n^\sharp$ is the set of elements of $\mathscr{P}_n$ that are local maxima in the lexicographic ordering of assignments, where the neighborhood of determination of local maximality is the set of assignments that differ from A in at most one position [KKKS98].

We thus have the following method of moments upper bound:

Lemma 3.97. *Let F be a random formula.*

$$\mathbb{P}(F \text{ is SAT}) \leq \mathbb{E}\left|\mathscr{P}_n^\sharp\right|. \tag{3.98}$$

Proof. From definition 3.95, we see that if an instantiation ϕ of random formula is satisfiable, the $\mathscr{P}_n^\sharp(\phi) \neq \emptyset$. We straightforwardly have

$$\mathbb{P}(F \text{ is SAT}) = \sum_\phi \mathbb{P}(\phi) \cdot \mathbb{1}_\phi, \tag{3.99}$$

where the indicator random variable

$$\mathbb{1}_\phi = \begin{cases} 1 & \text{if } \phi \text{ is satisfiable} \\ 0 & \text{otherwise.} \end{cases} \tag{3.100}$$

Also,

$$\mathbb{E}\left|\mathscr{P}_n^\sharp\right| = \sum_\phi \mathbb{P}(\phi)\cdot\left|\mathscr{P}_n^\sharp(\phi)\right|. \tag{3.101}$$

Since $\mathscr{P}_n^\sharp(\phi) \neq \emptyset$, $\left|\mathscr{P}_n^\sharp(\phi)\right| \geq 1$ and the lemma follows. □

The single flip method sharpens the previous lowest upper bound of 4.758 for the 3-SAT due to Kamath et al. [KMPS95] to 4.667 [KKKS98]. By defining an even smaller subset of $\mathscr{P}_n$ with a more general double flip method further improves the bound to a value between 4.601 and 4.60108 [KKKS98].

3.3.2 Lower bound

More recent advances in finding a lower bound for $\alpha_{\text{sat}}^{(N)}(K)$ have used the second moment method below.

Lemma 3.102 (Second moment method)**.** *Let $X \geq 0$ be a random variable with finite variance. Then,*

$$\mathbb{P}(X>0) \geq \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}. \tag{3.103}$$

Remark 3.104. *This sequence of development can be briefly summarized as such:*

$$\liminf_{N\to\infty} \alpha_{sat}^{(N)}(K) \geq \begin{cases} 2^{K-1}\ln 2 - d_k & \text{[AM02];} \\ 2^K \ln 2 - (K+1)\dfrac{\ln 2}{2} - 1 - \epsilon_K & \text{[AP04];} \\ 2^K \ln 2 - \dfrac{3}{2}\ln 2 + \epsilon_K & \text{[CP12],} \end{cases} \tag{3.105}$$

where $d_k \to (1+\ln 2)/2$ and, as before, $\epsilon_K \to 0$ as $K \to \infty$.

The second moment method brings two issues in the random K-SAT model: the first concerning the geometry of the solution space of possible assignments $\texttt{SOL} \subseteq \{\pm 1\}^N$, and the second concerning the geometry of the underlying bipartite graph [DSS16].

Remark 3.106. *Coja-Oghlan and Panagiotou* [CP16] *addressed both issues simultaneously and showed that*

$$\liminf_{N\to\infty} \alpha_{sat}^{(N)}(K) \geq 2^K \ln 2 - \frac{1}{2}(1+\ln 2) - \epsilon_K, \tag{3.107}$$

which matches the upper bound in Theorem 3.93 up to the error term ϵ_K.

Remark 3.108. *Ding, Sly, and Sun* [DSS16] *resolve the satisfiability threshold conjecture 3.89 for large K: For $K \geq K_0$, with K_0 an absolute constant, the random K-SAT has a sharp satisfiability threshold α_{SAT}.*

4 Appendix

4.1 Local magnetization density and free energy of the K-SAT

In this appendix, taken largely from [N08], we derive the self-consistent equation (3.46) and the free energy (3.49) from the variational free energy (3.43) under the RS Ansatz. The function $c(\underline{s})$ (3.44) only depends on $\underline{s}$ from the number of down spins j in the set $\underline{s} = (s_1, \cdots, s_n)$ if we assume symmetry between replicas. So, we will also use the notation $c(j)$ interchangeably with $c(\underline{s})$. The free energy from (3.43) is now

$$-\frac{\beta f}{N} = -\sum_{j=0}^{n} \binom{n}{j} c(j) \ln c(j) + \alpha \ln \left\{ \sum_{j_1=0}^{n} \cdots \sum_{j_K=0}^{n} c(j_1) \cdots c(j_K) \right. \\ \left. \cdot \sum_{\underline{s}_1(j_1)} \cdots \sum_{\underline{s}_K(j_K)} \prod_{a=1}^{n} \left(1 + \left(e^{-\beta} - 1 \right) \prod_{k=1}^{K} \delta\left(s_k^a, 1\right) \right) \right\}, \tag{4.1}$$

where the sum over $\underline{s}_i(j_i)$ is for the $\underline{s}_i$ with j_i down spins.

Taking the partial derivative of (4.1) with respect to $c(j)$ gives us

$$\frac{\partial}{\partial c(j)} \left(-\frac{\beta f}{N} \right) = -\binom{n}{j} \left(\ln c(j) + 1 \right) + \frac{\alpha K g}{f}. \tag{4.2}$$

Definition 4.3. We define f and g above as

(4.4)

$$f := \sum_{j_1=0}^{n} \cdots \sum_{j_K=0}^{n} c(j_1) \cdots c(j_K) \cdot \sum_{\underline{s}_1(j_1)} \cdots \sum_{\underline{s}_K(j_K)} \prod_{a=1}^{n} \left(1 + \left(e^{-\beta} - 1 \right) \prod_{k=1}^{K} \delta\left(s_k^a, 1\right) \right),$$

(4.5)

$$g := \sum_{j_1=0}^{n} \cdots \sum_{j_{K-1}=0}^{n} c(j_1) \cdots c(j_{K-1}) \cdot \sum_{\underline{s}_1(j_1)} \cdots \sum_{\underline{s}_{K-1}(j_{K-1})} \sum_{\underline{s}(j)} \prod_{a=1}^{n} \left(1 + \left(e^{-\beta} - 1 \right) \delta\left(s^a, 1\right) \prod_{k=1}^{K-1} \delta\left(s_k^a, 1\right) \right).$$

From our expression for $c(\underline{s})$ (3.44) reproduced below

$$c(\underline{s}) = \int_{-1}^{1} dm P(m) \prod_{a=1}^{n} \frac{1 + m s^a}{2}, \tag{4.6}$$

we can express f and g in terms of the local magnetization $P(m)$ as such

$$f = \int_{-1}^{1} \prod_{k=1}^{K} dm_k P(m_k)(A_K)^n \tag{4.7}$$

$$g = \binom{n}{j} \int_{-1}^{1} \prod_{k=1}^{K-1} dm_k P(m_k)(A_{K-1})^{n-j}, \tag{4.8}$$

where, as before in (3.47), we have

$$A_K = 1 + \left(e^{-\beta} - 1\right) \prod_{k=1}^{K} \frac{1+m_k}{2}. \tag{4.9}$$

We derive (4.7) and (4.8) as follows.

Changing to $c(j_i)$'s to $c(\underline{s}_i)$'s, substituting (3.44) into (4.4), and summing over $\underline{s}_1$ to $\underline{s}_K$ give us

$$f = \sum_{\underline{s}_1(j_1)} \cdots \sum_{\underline{s}_K(j_K)} c(\underline{s}_1) \cdots c(\underline{s}_K) \prod_{a=1}^{n} \left(1 + \left(e^{-\beta-1}\right) \prod_{k=1}^{K} \delta\left(s_k^a, 1\right) \right) \tag{4.10}$$

$$= \int_{-1}^{1} \prod_{k=1}^{K} dm_k P(m_k)(A_K)^n. \tag{4.11}$$

Since the sum over $\underline{s}(j)$ in g in (4.5) is for $\underline{s}$ with j down spins, which implies $\delta\left(s^a, 1\right) = 0$, we have

$$g := \sum_{j_1=0}^{n} \cdots \sum_{j_{K-1}=0}^{n} c\left(j_1\right) \cdots c\left(j_{K-1}\right) \cdot \sum_{\underline{s}_1(j_1)} \cdots \sum_{\underline{s}_{K-1}(j_{K-1})} \sum_{\underline{s}(j)} \prod_{s^a=1,a=1}^{n} \left(1 + \left(e^{-\beta} - 1\right) \prod_{k=1}^{K-1} \delta\left(s_k^a, 1\right) \right) \tag{4.12}$$

$$= \sum_{\underline{s}_1(j_1)} \cdots \sum_{\underline{s}_{K-1}(j_{K-1})} \sum_{\underline{s}(j)} c(\underline{s}_1) \cdots c(\underline{s}_{K-1}) \prod_{s^a=1,a=1}^{n} \left(1 + \left(e^{-\beta} - 1\right) \prod_{k=1}^{K-1} \delta\left(s_k^a, 1\right) \right), \tag{4.13}$$

where the product $\prod_{s^a=1,a=1}^{n}$ is over all replicas with $s^a = 1$. Substituting (3.44) into (4.13) and summing over $\underline{s}_1$ to $\underline{s}_{K-1}$ yield the desired result

$$g = \sum_{\underline{s}(j)} \int_{-1}^{1} \prod_{k=1}^{K-1} dm_k P(m_k) \prod_{s^a=1,a=1}^{n} A_{K-1} \tag{4.14}$$

$$= \binom{n}{j} \int_{-1}^{1} \prod_{k=1}^{K-1} dm_k P(m_k)(A_{K-1})^{n-j}, \tag{4.15}$$

since there are precisely $n-j$ up spins and $\binom{n}{j}$ ways to arrange such assignments.

Using the symmetry $c(\underline{s}) = c(-\underline{s})$ from Remark 3.21, we can take into account the symmetry $c(j) = c(n-j)$ in the extremization condition (4.2).

Subject to the normalization condition

$$\sum_{j=0}^{n}\binom{n}{j}c(j)=1 \tag{4.16}$$

and the Lagrange multiplier method, the extremization condition is now

$$0=-2(\ln c(j)+1)+\alpha K\int_{-1}^{1}\prod_{k=1}^{K-1}dm_kP(m_k)f^{-1}\left[(A_{K-1})^j+(A_{K-1})^{n-j}\right]-2\lambda. \tag{4.17}$$

It is straightforward to see with algebra that

$$c(j)=\exp\left\{-\lambda-1+\frac{\alpha K}{2f}\int_{-1}^{1}dm_kP(m_k)\left[(A_{K-1})^j+(A_{K-1})^{n-j}\right]\right\}. \tag{4.18}$$

Now, taking the limit of the number of replicas $n\to 0$ produces the self-consistent equation for $P(m)$. The value of the Lagrange multiplier λ in the limit $n\to 0$ is obtained from that $c(0)=1$ and $\lambda=\alpha K-1$. The distribution $P(m)$ is derived from the inverse relation of

$$c(j)=\int_{-1}^{1}dmP(m)\left(\frac{1+m}{2}\right)^{n-j}\left(\frac{1-m}{2}\right)^{j} \tag{4.19}$$

in the limit $n\to 0$:

$$P(m)=\frac{1}{\pi(1-m^2)}\int_{-\infty}^{\infty}dyc(iy)\exp\left(-iy\ln\left(\frac{1-y}{1+y}\right)\right). \tag{4.20}$$

Substituting the expression for $c(j)$ from (4.18) and $\lambda=\alpha K-1$ above results in (3.46).

For the free energy (3.49), we must consider the $O(n)$ terms to derive the free energy in terms of $P(m)$. We can condense (4.1) using our notation as such

$$-\frac{\beta f}{N}=-\sum_{j=0}^{n}\binom{n}{j}\ln c(j)+\alpha\ln f. \tag{4.21}$$

Lemma 4.22. *Recall the series expansion at 0 for the exponential function a^x for $a>0$*

$$a^x=\sum_{k=0}^{\infty}\frac{(x\ln a)^k}{k!}. \tag{4.23}$$

Note that the term $(A_K)^n$ in f (4.7) does not depend on k and is thus a constant term. Letting $a=A_k$ and $x=n$ and applying Lemma 4.22 give the following series expansion for f

$$f=1+n\mu+O(n^2), \tag{4.24}$$

where we define

$$\mu := \int_{-1}^{1} \prod_{k=1}^{K} dm_k P(m_k) \ln A_k. \tag{4.25}$$

Substituting (4.18) and using the normalization condition to compute the first term on the right-hand side of (4.21) give

$$-\sum_{j=0}^{n} \binom{n}{j} c(j) \ln c(j) = \lambda + 1 - \frac{\alpha K}{2f} \sum_{j=0}^{n} \binom{n}{j} c(j) \int_{-1}^{1} \prod_{k=1}^{K-1} dm_k P(m_k) \left[(A_{K-1})^{n-j} + (A_{K-1})^{j} \right]. \tag{4.26}$$

To expand λ to $O(n)$, we use equate (4.18) and (4.19) to solve for $e^{\lambda+1}$

$$e^{\lambda+1} = \frac{\exp\left\{ \frac{\alpha K}{2f} \int_{-1}^{1} \prod_{k=1}^{K-1} dm_k P(m_k) \left[(A_{K-1})^{n-j} + (A_{K-1})^{j} \right] \right\}}{\int_{-1}^{1} dm P(m) \left(\frac{1+m}{2} \right)^{n-j} \left(\frac{1-m}{2} \right)^{j}} \tag{4.27}$$

Since the left hand side of (4.27) is independent of j, we let $j = 0$ in (4.27) and expand to $O(n)$ as such

$$\lambda + 1 = \alpha K + n \left(\alpha K(-\mu + \nu/2) + \ln 2 - \int_{-1}^{1} dm P(m) \ln \left(1 - m^2\right) \right) + O(n^2), \tag{4.28}$$

where we define ν similarly

$$\nu := \int_{-1}^{1} \prod_{k=1}^{K-1} dm_k P(m_k) \ln A_{K-1}. \tag{4.29}$$

We use our expression for $c(j)$ (4.19) to evaluate the sum in the right hand side of (4.26)

$$\sum_{j=0}^{n} \binom{n}{j} \int_{-1}^{1} dm_k P(m_k) \left(\frac{1+m_k}{2} \right)^{n-j} \left(\frac{1-m_k}{2} \right)^{j} \int_{-1}^{1} \prod_{k=1}^{K-1} dm_k P(m_k) \left[(A_{K-1})^{n-j} + (A_{K-1})^{j} \right] \tag{4.30}$$

$$\overset{(\dagger)}{=} 2 \int_{-1}^{1} \prod_{k=1}^{K} dm_k P(m_k) \left(\frac{1+m_K}{2} A_{K-1} + \frac{1-m_K}{2} \right)^{n} \tag{4.31}$$

$$= 2f, \tag{4.32}$$

where, in (†), a little algebraic manipulation reveals

$$\frac{1+m_K}{2} A_{K-1} + \frac{1-m_K}{2} = \frac{1+m_k+1-m_K}{2} + \left(e^{-\beta} - 1\right) \prod_{k=1}^{K} \frac{1+m_k}{2} \tag{4.33}$$

$$= A_K. \tag{4.34}$$

Using (4.21), (4.24), (4.26), (4.28), and (4.32), we have the desired result

$$-\frac{\beta f}{Nn} = \ln 2 + \alpha(1-K)\mu + \frac{\alpha K}{2}\nu - \frac{1}{2}\int_{-1}^{1} dm P(m) \ln\left(1-m^2\right) + O(n). \tag{4.35}$$

References

[AM02] Achlioptas, D., and Moore, C. The Asymptotic Order of the Random K-SAT Threshold, Proceedings of the 43 rd Annual IEEE Symposium on Foundations of Computer Science, 2002.

[AP04] Achlioptas, D., and Moore, C. The Threshold for Random k-SAT is $2^k \log 2 - O(k)$, Journal of the American Mathematical Society, Volume 17, Issue 4, 2002.

[1] Binder, K., and Young, P. Spin glasses: Experimental facts, theoretical concepts, and open questions. Reviews of Modern Physics, Volume 58, Issue 4, 801-976. 1986.

[CP12] Coja-Oghlan, A., and Panagiotou, K. Going after the k-SAT Threshold. 2012.

[CP16] Coja-Oghlan, A., and Panagiotou, K. The asymptotic k-SAT threshold. Advances in Mathematics, Volume 288, 985-1068.

[DSS16] Ding, J., Sly, A., and Sun, N. Proof of the Satisfiability Conjecture for Large k. 2016.

[FH93] Fischer, K., and Hertz, J. Spin glasses. Cambridge University Press. 1993.

[FP83] Franco, J., and Paull, M. Probabilistic Analysis of the Davis Putnam Procedure for Solving the Satisfiability Problem. Discrete Applied Mathematics, Volume 5, 77-87. 1983.

[Fr99] Friedgut, E. Sharp thresholds of graph proprties, and the k-sat problem. Journal of the American Mathematical Society, Volume 12, Issue 4, 10171054.

[GJ79] Garey, M., and Johnson, D. Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman, New York. 1979.

[Go96] Goerdt, A. A Threshold for Unsatisfiability. Journal of Computer and System Sciences, Volume 53, 469-486. 1996.

[HP78] Hemmen, J., and Palmer, R. The replica method and a solvable spin glass model. Journal of Physics A: A Mathematical and General, Volume 12, Issue 4, 563-580. 1978.

[Is25] Ising, E. Beitrag zur Theorie des Ferromagnetismus. Zeitschrift fr Physik, Volume 31, Issue 1, 253-258. 1925.

[KMPS95] Kamath, A., Motwani, R., Palem, K., and Spirakis, P. Tail bounds for occupancy and the satisfiability threshold conjecture, Random Structures & Algorithms, Volume 7, 59-80. 1995.

[KKKS98] Kirousis, L., Kranakis, E., Krizanc, D., and Stamatiou, Y. Approximating the Unsatisfiability Threshold of Random Formulas. Random Structures & Algorithms, Volume 12, Issue 3, 253-269. 1998.

[MM09] Mézard, M., and Montanari, A. Information, Physics, and Computation. Oxford University Press. 2009.

[MZ97] Monasson, R., and Zecchina, R. Statistical mechanics of the random K-satisfiability model. Physical Review, Volume 56, Issue 2, 1357-1370. 1997.

[N08] Nishimori, H. Statistical physics of spin glasses and information processing: an introduction. Oxford Univ. Press, 2008

[On44] Onsager, L. Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition. Physical Review, Volume 65, Issue 3-4, 117-49. 1944.

[PB36] Peierls, R., and M. Born. On Ising's model of ferromagnetism. Mathematical Proceedings of the Cambridge Philosophical Society, Volume 32, Issue 3 (1936): 477-481. 1936.

[SK75] Sherrington, D., and Kirkpatrick, S. Solvable Model of a Spin-Glass. Physical Review Letters, Volume 35, Issue 26, 17921796. 1975.

[Ta07] Tanaka, T. Moment Problem in Replica Method. Interdisciplinary Information Sciences, Volume 13, Issue 1, 17-23. 2007.