The Free Energy of Spherical Vector Spin Glasses

by

Justin Ko

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto

Abstract

The Free Energy of Spherical Vector Spin Glasses

Justin Ko
Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto
2020

We study a class of vector spin models with configurations restricted to subsets of the sphere. We will prove a constrained free energy formula for these models. This formula defines a large deviations principle for the limiting distribution of the overlaps under the asymptotic Gibbs measure. The thesis builds on the mathematical results used to prove free energy formulas for the classical Sherrington–Kirkpatrick spin glass [99], spherical spin models [30], and vector spin glass models [80]. The free energy formula proved in this thesis are true generalizations of the classical results, in the sense that these vector spin formulas restricted to one dimension coincide with the known results for classical models.

The first contribution of this thesis is a variational formula for contrained copies of classical spherical spin glasses sampled at different temperatures. The free energy for multiple systems of spherical spin glasses with constrained overlaps was studied in [83] where the authors proved an upper bound of the constrained free energy using Guerra's interpolation. In this thesis, we prove this upper bound is sharp. Our approach combines the ideas of the Aizenman–Sims–Starr scheme in [30] and the synchronization mechanism used in the vector spin models in [80] and [81]. We derive a vector version of the Aizenman–Sims–Starr scheme for spherical spin glass and use the synchronization property of arrays obeying the overlap-matrix form of the Ghirlanda–Guerra identities to prove the matching lower bound.

The second contribution of this thesis is the simplification of this variational formula to the form originally discovered for the classical spherical spin glass model by Crisanti and Sommers [40]. In particular, we prove the analogue of the Crisanti–Sommers variational formula for spherical spin glasses with vector spins. This formula is derived from the discrete Parisi variational formula for the limit of the free energy of constrained copies of spherical spin glasses. In vector spin models, the variations of the functional order parameters must preserve the monotonicity of matrix paths which introduces a new challenge in contrast to the derivation of the classical Crisanti–Sommers formula.

In memory of my grandmother.

Acknowledgements

I owe my sincere gratitude to my advisor, Dmitry Panchenko, for his support throughout my graduate studies. His door was always open if I ever needed guidance and he always found a way to point me in the right direction. It was his encouragement, attention to detail, and insight that made the completion of this thesis possible. His work as a mathematician and as a mentor is truly inspiring.

I want to thank Aukosh Jagannath for his immense help during the start of my graduate studies and for his continued support until now. He was like a second advisor and I am grateful for his mentorship. In addition, I want to thank Wei-Kuo Chen for his useful comments on several of my results and for giving me plenty of problems to think about. Antonio Auffinger also deserves a special mention for his thoughtful suggestions and careful reading of this thesis. Lastly, I also want to thank David Belius, Alice Guionnet, Subhabrata Sen, and Lenka Zdeborová for their helpful discussions and for introducing me to the multiple fascinating areas of spin glasses.

At the University of Toronto, I want to thank Dietrich Burbulla, Almut Burchard, Bob Haslhofer, Jeremy Quastel, and the rest of the professors I was fortunate enough to learn from. I would also like to thank Ashley Armogan, Marija Ignjatovic, Sonja Injac, Jemima Merisca, Patrina Seepersaud, and all the staff at the University of Toronto for their hard work.

I am grateful to have made so many good friends during my studies. Special thanks to Afiny Akdemir, Aghil Alaee, Robert Chang, Yichao Chen, Charlene Chu, Andrew Colinet, Keegan Dasilva Barbosa, Duncan Dauvergne, Wenkui Du, Leon Fröber, Zhichao Du, Vivian He, Fabian Henneke, Tyler Holden, Xiao Jie, Christopher Kennedy, Tomas Kojar, Kuni Ito, Bill Li, Jinhui Li, Qiaoyu Liang, David Miyamoto, Jeff Negrea, Adriano Pacifico, Mario Palasciano, Daniel Park, David Pechersky, Virginia Pedreira, Krishan Rajaratnam, Florian Schweiger, Russell Sharpe, Shuyang Shen, Matt Sourisseau, Dani Spivak, Motasem Suleiman, Kaidi Ye, Sina Zabanfahm, Shuangjian Zhang, Yuxin Zhou, and Farshid Zoghalchi for making my studies memorable.

Contents

Chapter 1

Introduction

The field of spin glasses originated in statistical physics. In 1975, physicists Sherrington and Kirkpatrick introduced a model of a spin glass with Ising spins [93]. Several years later, Parisi proposed a formula for the free energy of this model in the thermodynamic limit [84]. In 2006, Talagrand proved that the thermodynamic limit of the free energy in the Sherrington–Kirkpatrick model is given by Parisi's formula [99] resulting in numerous mathematical advances in this field.

The key feature exhibited in these models is the ultrametric support of the asymptotic Gibbs measure proven by Panchenko [75]. Perturbations of the Hamiltonians in a wide family of spin glass models can regularize the Gibbs measure forcing it to satisfy a stability property known as the generalized Ghirlanda–Guerra identities [52] in the thermodynamic limit. These identities imply that the limiting Gibbs measure in more complicated spin glass models exhibit the same key properties exhibited in the classical spin glass model such as: ultrametricity [75], positivity [101], and synchronization [78, 80]. These properties can be generalized to prove analogues of the Parisi formula formula in a wide range of more complicated spin glass models.

This thesis builds on these results to prove free energy formulas for high dimensional spin glasses. In particular, we will prove a free energy formula for constrained copies of spherical spin glasses, also called the spherical vector spin glasses. We will review the tools and techniques used to study the traditional spin glass models, and adapt these techniques to prove a free energy formula for general constrained vector spin models. The formulas proved in this thesis are true generalizations of the classical results, in the sense that the results proved for the vector spin models restricted to one dimension coincide with the known results for classical spin glass models.

1.1 Organization of the Thesis

In Chapter 1, we motivate the study of spin glasses and discuss the various mathematical contributions in the field leading to the derivation of the free energy formulas for several classical spin glass models. We introduce some traditional spin glass models such as the Sherrington–Kirkpatrick model [Section 1.2.1], spherical model [Section 1.2.2], and the vector spin model [Section 1.2.3]. Next we summarize the main contributions of this thesis and motivate where these new results fit in the wider field of spin glasses [Section 1.3, 1.4]. In Chapter 2, we review a variety of classical mathematical results used in the derivation of the Parisi variational formula for the Ising spin models. In Chapter 3, we include some

technical results that will be used to extend the results of Chapter 2 to the setting of spherical models and vector spin models.

The main contribution of the thesis begins in Chapter 4. We will motivate and prove a free energy formula for constrained copies of spherical spin glasses sampled independently at various temperatures. One can view this problem as the high dimensional analogue of the classical spherical spin glass models introduced in Chapter 1. The main result is a variational formula for the limit of the constrained free energy. In Chapter 5, we partially solve this optimization problem and express it as in the form of the original free energy formula predicted by physicists.

1.2 Overview of Related Spin Glass Models

In this section, we introduce and motivate several models of spin glasses and discuss the historical contributions in the study of these models. The techniques used to prove results for these classical models motivate the approaches we use to study the spherical vector spin models in this thesis [Section 1.3]. We will explain the significance of our results and summarize some areas of future research in Section 1.4.

We begin by providing a broad introduction to spin glasses. A common motivation for spin glasses is the optimization of a random function in high dimensions. For dimension $N \geq 1$, we let $\Sigma_N \subset \mathbb{R}^N$ denote the *configuration space* of our model. This is the domain of our optimization problem. Let $H_N : \Sigma_N \to \mathbb{R}$ be a random function called the *Hamiltonian*. A central question in spin glasses is the asymptotic behavior of the maximum of these random function. Some examples of the Hamiltonians that fall under the broad classification of spin glass models include the comfort function in the Dean's problem, the log-likelihood functions in high dimensional Bayesian inference, the number of failed clauses in the random k-SAT problem, and the volume of interesting half spaces in the perceptron model. An overview of such models can be found in various textbooks on spin glasses [68, 71, 76, 101, 102, 108].

In this thesis, we will focus on one type of Hamiltonian, namely the high dimensional random polynomial called the *mixed p-spin Hamiltonians* in spin glasses. We will give precise definitions of these random functions in (1.7), (1.26), (1.37). These Hamiltonians are normalized so that its variance grows linearly in N. Furthermore, the random functions are Gaussian, so by classical Gaussian concentration inequalities [Theorem 2.1.5], it suffices to study the asymptotics of the expected value of its maximum

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E} \max_{\boldsymbol{\sigma}\in\Sigma_N} H_N(\boldsymbol{\sigma}). \tag{1.1}$$

In our examples, the logarithm of the volume of Σ_N is of order N and H_N has variance of order N, so it is reasonable to suspect that the expected value of the maximum of the Hamiltonian is roughly of order N [101, Lemma A.3.1]. The precise asymptotics of (1.1) on the other hand is a very challenging problem. One objective moving forward is to prove a formula for (1.1).

To approach these optimization problems (1.1), we will use methods from statistical mechanics. Instead of solving the optimization problem directly, we introduce a quantity called the *free energy* and solve this relaxed problem instead. This statistical mechanics approach to study the distribution of maximum values is widely used in other areas of probability as well, and is not specific to the field of spin glasses. Let $\beta \geq 0$ denote the *inverse temperature* parameter, and consider the random measure

G_N on Σ_N called the *Gibbs measure*,

$$dG_N(\boldsymbol{\sigma}) = \frac{1}{Z_N} e^{\beta H_N(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma}) \qquad \text{and} \qquad G_N(A) = \int_A \frac{1}{Z_N} e^{\beta H_N(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma}) \tag{1.2}$$

where μ_N is a fixed finite measure on Σ_N called the *reference measure*. These Gibbs measures are probability measures that assign more weight to configurations near the maximums of $H_N(\boldsymbol{\sigma})$ when β is large. The normalization term Z_N of this probability measure is called the *partition function*

$$Z_N = \int_{\Sigma_N} e^{\beta H_N(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma}). \tag{1.3}$$

We let $\langle \cdot \rangle$ denote the expected value with respect to G_N,

$$\big\langle f(\boldsymbol{\sigma}) \big\rangle = \int_{\Sigma_N} \frac{f(\boldsymbol{\sigma}) e^{\beta H_N(\boldsymbol{\sigma})}}{Z_N}\, d\mu_N(\boldsymbol{\sigma}). \tag{1.4}$$

The maximum (1.1) is encoded by the *free energy* function $F_N : [0, \infty) \to \mathbb{R}$ given by

$$F_N(\beta) = \frac{1}{N} \mathbb{E} \log Z_N = \frac{1}{N} \mathbb{E} \log \int_{\Sigma_N} e^{\beta H_N(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma}). \tag{1.5}$$

To see the relation between the free energy and the maximum value, we consider the simple *Ising spin* case on the finite configuration space $\Sigma_N = \{-1, +1\}^N$ where μ_N uniform on the Σ_N. Since

$$e^{\beta \max_{\sigma \in \Sigma_N} H_N(\boldsymbol{\sigma})} \le \sum_{\sigma \in \{-1,+1\}^N} e^{\beta H_N(\boldsymbol{\sigma})} \le 2^N e^{\beta \max_{\sigma \in \Sigma_N} H_N(\boldsymbol{\sigma})}$$

we have

$$\frac{1}{N} \mathbb{E} \max_{\sigma \in \Sigma_N} H_N(\sigma) - \frac{\log 2}{\beta} \le \frac{1}{N\beta} \mathbb{E} \log \sum_{\sigma \in \{-1,+1\}^N} e^{\beta H_N(\boldsymbol{\sigma})} \cdot \frac{1}{2^N} \le \frac{1}{N} \mathbb{E} \max_{\sigma \in \Sigma_N} H_N(\sigma).$$

The maximum value of the Hamiltonian is given by the *zero-temperature* limit of the free energy,

$$\lim_{\beta \to \infty} \lim_{N \to \infty} \frac{F_N(\beta)}{\beta} = \lim_{N \to \infty} \frac{1}{N} \mathbb{E} \max_{\sigma \in \Sigma_N} H_N(\sigma). \tag{1.6}$$

The existence of this limit for the Sherrington–Kirkpatrick Hamiltonians we introduce in Section 1.2.1 was proved by Guerra and Toninelli [55].

In addition to the maximum, the free energy also encodes information about the central moments of the Hamiltonian. We can think of the free energy as a "moment generating" function of the distribution of the Hamiltonian under the Gibbs measure (1.2). We can differentiate $F_N(\beta)$ with respect to β to see that the free energy encodes several statistics about the random function H_N:

1. Expected value: Taking the first derivative of $F_N(\beta)$ gives

$$\frac{d}{d\beta} F_N(\beta) = \frac{1}{N} \mathbb{E} \big\langle H_N(\sigma) \big\rangle.$$

Conditionally on the randomness in H_N, the quantity on the right can be interpreted as the average of $H_N(\sigma)$ with respect to the Gibbs measure G_N defined in (1.2).

2. Variance: Taking the second derivative of $F_N(\beta)$ gives

$$\frac{d^2}{d\beta^2}F_N(\beta) = \frac{1}{N}\Big(\mathbb{E}\big\langle H_N^2(\sigma)\big\rangle - \mathbb{E}\big\langle H_N(\sigma)\big\rangle^2\Big).$$

Conditionally on the randomness in H_N, the quantity on the right can be interpreted as the variance of $H_N(\sigma)$ with respect to the Gibbs measure G_N defined in (1.2).

The derivatives of the free energy coincide with the moments of the random function under the averaged Gibbs measure. This relationship between the free energy and the behavior of H_N under the Gibbs measure is a key relationship will be explored in later chapters. The primary objective of this thesis is to prove some variational formulas for $F_N(\beta)$ at all parameters $\beta \geq 0$. These formulas will give us a tool to study the properties of the random function H_N.

The rest of this section is dedicated to a broad overview of results for classical spin glass models: the Sherrington–Kirkpatrick model [Section 1.2.1], the spherical spin glass models [Section 1.2.2], and the vector spin glass models [Section 1.2.3]. We will provide a historical overview of the main contributions in the analysis of these models leading to the derivation of the free energy formulas. A detailed overview of the early development of spin glasses in the perspective of one of the field's founders can be found in the series of articles [4, 5, 6, 7, 8, 9, 10].

The main contribution of this thesis is for multi-system versions of the one-dimension classical models with special constraints among the individual systems. We will build on the results for classical models to prove a formula for the free energy of the *spherical vector spin glass models*. The motivation for these multi-system models and a precise description of the constraints will be explained in Section 1.3.

1.2.1 The Sherrington–Kirkpatrick Model

We begin by introducing the classical Sherrington–Kirkpatrick model. This model was introduced by Sherrington and Kirkpatrick in 1975 to describe the magnetic behavior of disordered magnetic alloys called spin glasses [93]. It is also commonly motivated as a frustrated optimization problem called the *Dean's problem* [76, Section 1.1]. Mathematically, this model is equivalent to the optimization of a random a homogeneous quadratic polynomial on the hypercube $\{-1,+1\}^N$, which we now describe.

Let $N \geq 1$. Consider independent Gaussian random variables $g_{ij} \sim N(0,1)$ for $1 \leq i,j \leq N$ called the *disorder* of the model. We define the *configuration space* $\Sigma_N = \{-1,+1\}^N$ and consider the random Hamiltonian $H_N : \Sigma_N \to \mathbb{R}$ given by

$$H_N(\boldsymbol{\sigma}) = \frac{1}{\sqrt{N}}\sum_{i,j=1}^{N} g_{ij}\sigma_i\sigma_j, \tag{1.7}$$

where $\boldsymbol{\sigma} = (\sigma_1,\ldots,\sigma_N) \in \Sigma_N$ are called *spin configurations*. The coordinates $\sigma_i \in \{-1,+1\}$ are called *Ising spins*. In this setting, the Hamiltonian can be interpreted as a centered Gaussian process indexed by points on the hypercube $\{-1,+1\}^N$. Its covariance is given by

$$\mathbb{E}H_N(\boldsymbol{\sigma}^1)H_N(\boldsymbol{\sigma}^2) = N\bigg(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^1\sigma_i^2\bigg)^2 = NR_{1,2}^2, \tag{1.8}$$

where $R_{1,2}$ denotes the normalized inner product,

$$R_{1,2} = \frac{\boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2}{N} = \frac{1}{N}\sum_{i=1}^{N} \sigma_i^1 \sigma_i^2 \in [-1, 1]. \tag{1.9}$$

We call the normalized inner product $R_{1,2}$ the *overlap*, because it measures the proportion of the coordinates such that $\boldsymbol{\sigma}^1$ and $\boldsymbol{\sigma}^2$ are equal. We will see shortly that covariances defined as convex functions of the overlaps is one of the key features in more general spin glass models. The limiting distribution of the overlap under the Gibbs measure (1.2) is the parameter of the free energy formulas we will introduce later in (1.15).

As seen in (1.6), the precise asymptotics of the Hamiltonian is described by a quantity known as the free energy for the Sherrington–Kirkpatrick Hamiltonian with Ising spins,

$$F_N(\beta) = \frac{1}{N}\mathbb{E}\log Z_N = \frac{1}{N}\mathbb{E}\log \sum_{\boldsymbol{\sigma}\in\{-1,+1\}^N} e^{\beta H_N(\boldsymbol{\sigma})}. \tag{1.10}$$

A formula for the limit of (1.10) valid for small β called the *replica symmetric solution* was proposed by Sherrington and Kirkpatrick [93]. Several years later Parisi [84, 85] proposed a solution called the *replica symmetric breaking solution* that was valid for all β. His original method was based on a non-rigorous computation known as the *replica method* combined with the *ultrametric parametrization* of the replica matrices that appear in this calculation. A detailed description of the replica trick can be found in [106] or [101, Chapter 1.13].

In Parisi's original solution, the ultrametricity property was simply a constraint on the algebraic parametrization of the replica matrices appearing in the computation. Parisi's ansatz generalized the original replica symmetric ansantz proposed by Sherrington and Kirkpatrick, but the physical meaning behind this generalization was still a mystery. Only later in the works of Mezard, Parisi, Toulouse, and Virasoro [86, 69, 70] was ultrametricity reinterpreted more physically in terms of the properties of the Gibbs measure

$$G_N(\boldsymbol{\sigma}) = \frac{e^{\beta H_N(\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}\in\Sigma_N} e^{\beta H_N(\boldsymbol{\sigma})}}. \tag{1.11}$$

Geometrically, the ultrametric behavior of the replica matrices correspond to the ultrametric support of the Gibbs measure in the $N \to \infty$ limit. More precisely, if we consider an independent sample of three replica $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2, \boldsymbol{\sigma}^3$ from (1.11) normalized to have length 1, then with high probability the normalized samples satisfy an approximate ultrametric inequality,

$$\|\boldsymbol{\sigma}^2 - \boldsymbol{\sigma}^3\| \lesssim \max(\|\boldsymbol{\sigma}^1 - \boldsymbol{\sigma}^2\|, \|\boldsymbol{\sigma}^1 - \boldsymbol{\sigma}^3\|). \tag{1.12}$$

Another related concept to the ultrametric parametrization is the *order parameter* in Parisi's solution [86]. The order parameter implies that the free energy is determined by the distribution of the overlap $R_{1,2}$ defined in (1.9) of two *replica* $\boldsymbol{\sigma}^1$ and $\boldsymbol{\sigma}^2$ sampled independently from the same Gibbs measure (1.11). The concept of ultrametricity and the order parameters are closely connected, and it turns out that the ultrametric support of the Gibbs measure is the reason why the distribution of the entire replica matrix is determined by the functional order parameter [Theorem 2.2.12]. This geometric interpretation opened a path to a rigorous mathematical theory for these models.

Parisi's formula for the free energy was proven rigorously by Talagrand [99] following the replica

symmetry breaking interpolation discovered by Guerra [54]. In this thesis, we will be following the strategy of a later proof by Panchenko [77] using the ultrametric structure of the Gibbs measure, which was unproven at the time of Talagrand's proof. The approach outlined in this proof is very general and can be extended to prove results in a large class of more complicated spin glass models that we study in this thesis.

The key mathematical result in this proof of the free energy is Parisi's ultrametricity conjecture [Theorem 2.2.11] proved rigorously by Panchenko in [75] following a partial result by Arguin and Aizenman [11]. The proof is a cumulation of earlier results for a simplified model for spin glass with ultrametric structure known as the generalized energy models [43, 44, 45, 46] and the stability properties of the Gibbs measure known as the Ghirlanda–Guerra identities [52] and the Aizenman–Contucci stochastic stability [1]. This ultrametricity property allows us to approximate the limiting Gibbs measure in an appropriate sense by a family of random measures supported on ultrametric sets of a Hilbert space known as the Ruelle probability cascades [90, 23]. This connection is explained in [76, Chapter 2] and summarized in Section 2.2 of this thesis.

We will see later in Section 1.2.3 that the ultrametric structure of the Gibbs measure has far reaching consequences that will allow us to generalize the methodology used to study the Sherrington–Kirkpatrick model to more complex models. These key consequences will also be adapted to prove the formulas for the spherical vector spin models studied in this thesis.

The Parisi Formula

We now state the form of the variational formula for the limit of the free energy. For a probability measure μ on $[0,1]$, let

$$x(t) = \mu([0,t]) \quad \text{such that} \quad x(0) = 0 \quad \text{and} \quad x(1) = 1 \tag{1.13}$$

denote a right continuous non-decreasing function representing its cumulative density function on $[0,1]$. The *Parisi functional* on these cdfs is defined by

$$\mathscr{P}(x) = \mathcal{P}(x) - \beta^2 \int_0^1 x(t) t \, dt, \tag{1.14}$$

where $\mathcal{P}(x)$ has an explicit form that will be defined shortly in (1.16). The parameter $x(t)$ in the Parisi functional (1.14) is called the *functional order parameter*. The limit of the free energy is given by minimizing (1.14) over the all cdfs on $[0,1]$.

Theorem 1.2.1 (*The Parisi Formula*)

The limit of the free energy (1.10) *equals*

$$\lim_{N\to\infty} F_N(\beta) = \inf_x \mathscr{P}(x), \tag{1.15}$$

where the infimum is taken over cumulative distribution functions on $[0,1]$.

The quantity $\mathcal{P}(x)$ appearing in (1.14) is given by the solution to a special non-linear parabolic

equation with a specific initial condition

$$\begin{cases} \Phi_t = -\beta^2(\Phi_{yy} + x(t)\Phi_y^2) & (t,y) \in (0,1) \times \mathbb{R} \\ \Phi(1,y) = \log(2\cosh(y)). \end{cases} \tag{1.16}$$

If $\Phi_\mu(t,y)$ denotes the solution to (1.16), then $\mathcal{P}(x) = \Phi_\mu(0,0)$ in (1.14). Auffinger and Chen proved that the functional $\mathcal{P}(x)$ is strictly convex, so $\mathscr{P}(x)$ has a unique minimizer [16]. Another proof of strict convexity for non-discrete μ using stochastic dynamic programming was discovered by Jagannath and Tobasco [59]. The probability measure μ associated with the minimizing $x(t)$ is called the *Parisi measure* and can be interpreted as the candidate for the limiting distribution of the overlap $R_{1,2}$ (1.9) under the Gibbs measure. This interpretation is correct for a class of models called *generic mixed p-spin models*.

The functional (1.14) is Lipschitz continuous [54, 100], so it suffices to restrict the infimum to discrete cdfs. In this setting, the functional $\mathcal{P}(x)$ can be computed recursively along the jumps of the cdf. We will state the discrete version of the Parisi variational formula because analogues of this discrete formula will show up when we describe the main results of this thesis in Section 1.3.

Every discrete cdf is encoded by sequences of parameters

$$\begin{aligned} 0 = x_0 < x_1 < \ldots < x_{r-2} < x_{r-1} \leq 1 \\ 0 = q_0 \leq q_1 \leq \ldots \leq q_{r-2} \leq q_{r-1} \leq 1 \end{aligned},$$

with the convention that $q_r = 1$ and $x_r = 1$. By continuity, we may take these inequalities to be strict. These sequences encode the jumps of the cdf of a probability measure μ on $[0,1]$, i.e. for $0 \leq k \leq r-1$

$$x(q) = \mu([0,q]) = x_k \qquad \text{for} \qquad q_k \leq q < q_{k+1}. \tag{1.17}$$

We now define a recursive formulation of $\mathcal{P}(x)$ that appears in the variational formula. Consider the increasing convex function $\xi : [0,1] \to \mathbb{R}^+$,

$$\xi(q) = \beta^2 q^2, \tag{1.18}$$

and the related covariance functions

$$\xi'(q) = 2\beta^2 q \qquad \text{and} \qquad \theta(q) = q\xi'(q) - \xi(q) = \beta^2 q^2. \tag{1.19}$$

The covariance function ξ appears as the covariance of the Sherrington–Kirkpatrick Hamiltonian, and the functions ξ' and θ appear as the covariances of the *cavity fields* [76, Chapter 3] that appear when we compare the Hamiltonians of different system sizes. Consider a sequence of centered Gaussians $(z_k)_{1 \leq k \leq r}$ with variance

$$\mathbb{E} z_k^2 = \xi'(q_k) - \xi'(q_{k-1}). \tag{1.20}$$

We define

$$X_r = \log \sum_{\sigma \in \{-1,+1\}} \exp\left(\sigma \sum_{1 \leq j \leq r} z_k\right) = \log 2\cosh\left(\sum_{1 \leq j \leq r} z_k\right) \tag{1.21}$$

and, recursively for $1 \le k \le r-1$, we define

$$X_k = \frac{1}{x_k} \log \mathbb{E}_k \exp(x_k X_{k+1}) \iff \exp(x_k X_k) = \mathbb{E}_k \exp(x_k X_{k+1}), \tag{1.22}$$

where $\mathbb{E}_k$ denotes the expectation with respect to z_{k+1}. If $x_k = 0$ then we interpret the formula to mean

$$X_k = \mathbb{E}_k X_{k+1}, \tag{1.23}$$

which is precisely what we get if we take the limit as $x_k \to 0$ in the recursive definition. The quantity X_0 is nonrandom, and we will denote it by $\mathcal{P}(x)$. Consider the functional

$$\mathscr{P}(x) = \mathcal{P}(x) - \frac{1}{2} \sum_{1 \le k \le r-1} x_k(\theta(q_{k+1}) - \theta(q_k)). \tag{1.24}$$

By continuity, the Parisi formula defined in Theorem 1.2.1 states that the limit of (1.10) is equal to the minimization of (1.24) over discrete cdfs. By a direct computation, it is also easy to see that the derivation of $\mathcal{P}(x)$ for discrete cdfs is consistent with the earlier PDE definition $\mathcal{P}(x) = \Phi_\mu(0,0)$.

Mixed p-spin Models

The Parisi variational formula (1.24) holds for a more general class of *mixed p-spin* models on the hypercube $\{-1,+1\}^N$. The Sherrington–Kirkpatrick Hamiltonian (1.7) discussed earlier in Section 1.2.1 is a special case of this model called the pure 2-spin model. Given a sequence of parameters $(\beta_p)_{p \ge 1}$, we define the *mixed p-spin* Hamiltonian as

$$H_N(\boldsymbol{\sigma}) = \sum_{p \ge 1} \beta_p H_{N,p}(\boldsymbol{\sigma}), \tag{1.25}$$

where $H_{N,p}(\boldsymbol{\sigma})$ denotes pure p-spin Hamiltonians

$$H_{N,p}(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1,\dots,i_p=1}^{N} g_{i_1 \dots i_p} \sigma_{i_1} \cdots \sigma_{i_p}. \tag{1.26}$$

It is easy to check that the covariance of these Gaussian processes are

$$\mathbb{E} H_{N,p}(\boldsymbol{\sigma}^1) H_{N,p}(\boldsymbol{\sigma}^2) = N \xi(R_{1,2}), \tag{1.27}$$

where

$$\xi(q) = \sum_{p \ge 1} \beta_p^2 x^p. \tag{1.28}$$

We assume that the inverse temperature parameters (β_p) are chosen so that these sums are well defined. The approaches used to prove the Parisi formula for the pure 2-spin model, also apply to these mixed p-spin models. If $\mathcal{P}(x)$ is the functional defined in (1.24) with respect to the mixed p-spin $\xi(q)$ defined in (1.28), then Parisi's formula [Theorem 1.2.1] extended to mixed p-spin models states

$$\lim_{N \to \infty} F_N(\beta) = \inf_x \Big(\mathcal{P}(x) - \frac{1}{2} \int_0^1 t \xi''(t) x(t)\, dt \Big). \tag{1.29}$$

The discrete form of the Parisi functional (1.24) for mixed p-spin models can be extended to all cdfs on $[0,1]$ by Lipschitz continuity. For continuous cdfs we can express $\mathcal{P}(x)$ in terms of the solution of a non-linear heat equation called the *Parisi PDE* [76, Lemma 4.3],

$$\begin{cases} \Phi_t = -\frac{\xi''(t)}{2}(\Phi_{yy} + x(t)\Phi_y^2) & (t,y) \in (0,1) \times \mathbb{R} \\ \Phi(1,y) = \log(2\cosh(y)). \end{cases} \tag{1.30}$$

In our main results, we will also prove a similar continuous extension of the discrete Parisi formula for spherical vector spins in Chapter 5 in Section 5.4.

The proof of the upper bound of the free energy uses a replica symmetry breaking interpolation bound [54] to interpolate between the mixed p-spin Hamiltonian (1.26) and the limiting Hamiltonian described by the Reulle probability cascades. The proof of the matching lower bound uses a cavity computation approach to compare systems of different sizes [2] and the limiting Gibbs measure which can be approximated by the same family of Reulle probability cascades appearing in the upper bound.

Several other variants of this free energy formulas (1.29) have been proved. Most noteworthy is the stochastic calculus reformulation of the formula discovered by Auffinger and Chen in [16]. The Parisi functional in this form is convex and therefore has a unique minimizer called the *Parisi measure*. There is also a formulation of the formula in terms of the TAP free energy [104] by Chen, Panchenko, and Subag [34, 35, 36]. Recently, there was a way to express the Parisi form of a modified *enriched free energy* as the solution of a class of Hamilton–Jacobi equations by Mourrat and Panchenko [73].

The properties of the *Parisi measure*, the unique minimizer of $\mathscr{P}(x)$ defined in (1.29), is another important question in the spin glasses. In the pure 2-spin Sherrington–Kirkpatrick model, the Parisi measure is concentrated on one point for $\beta \leq \frac{1}{\sqrt{2}}$. This is the notion of *replica symmetry* described by Sherrington and Kirkpatrick. For $\beta > \frac{1}{\sqrt{2}}$ is was proved by Toninelli [105] that the Parisi measure cannot concentrate on one point, so we have *replica symmetry breaking* at low temperatures. This phase transition between is called the de Almeida–Thouless line [41]. The phase diagrams and properties of the minimizers for mixed p-spin models were studied in [15, 16, 31, 61, 100]. It is conjectured that for large β, the minimizer has a continuous part, so we have a phenomenon called *full RSB* in Ising models. Recently, efficient optimization algorithms were developed for full RSB models in [3, 72].

Remark 1.2.1. In the one-dimensional setting, the free energy formula applies even when the covariance structure ξ defined in (1.28) is not convex on $\mathbb{R}$. This follows because a perturbation of the Hamiltonian forces the overlaps to take positive values [Theorem 2.3.2], and $\xi(q)$ is convex for $q \geq 0$. This is not true in the vector spin models, so we will have to restrict our analysis to even mixed p-spin models in higher dimensional models to ensure the covariances are convex.

1.2.2 The Spherical Spin Glass Models

We now introduce a generalization of the Ising spin models to a continuous configuration space. We call this model the spherical spin glasses, which uses configurations on a sphere instead of a hypercube. Surprisingly, the spherical model is much harder to solve mathematically, even though the variational formula has a simpler closed form. The main difficulty is the uniform measure on the sphere cannot be expressed as a product measure of its coordinates. This lack of symmetry between the spin coordinates changes the cavity computations to prove the lower bound.

We will state the 1 dimensional spherical spin glass model for mixed p-spin Hamiltonians with external field. There are very general models, in the sense that admissible functions $\xi(q)$ defined (1.28) on the sphere are equivalent to these mixed p-spin models [91]. Consider the $N-1$ dimensional sphere of radius $\sqrt{N}$ in $\mathbb{R}^N$ denoted by

$$S_N = \{\boldsymbol{\sigma} \in \mathbb{R}^N \mid \|\boldsymbol{\sigma}\|_2 = \sqrt{N}\}. \tag{1.31}$$

Let λ_N denote the normalized uniform surface measure on S_N, such that $\lambda_N(S_N) = 1$. The main difficulty in the mathematical analysis of the spherical model is the absence of the product measure structure in the reference measure. In the Ising spin case, we observe that

$$\mu_N(\boldsymbol{\sigma}) = \prod_{i=1}^{N} \mu(\sigma_i) \tag{1.32}$$

where μ is the uniform measure on $\{-1, +1\}$, i.e. $\mu(\pm 1) = \frac{1}{2}$. This combined with the symmetry of the Hamiltonian implies that Ising spin models mean that the Gibbs measure is invariant under permutations of the coordinates called *symmetry between sites*. This property is not immediately satisfied by spherical models, so some extra work has to be taken to reduce this model to the setting of the Ising spin case.

Let $H_N(\boldsymbol{\sigma})$ denote the mixed p-spin Hamiltonian defined in (1.25). Given *external field* $h \in \mathbb{R}$, we want to compute free energies of the form

$$F_N(\beta) = \frac{1}{N}\mathbb{E}\log \int_{S_N} e^{H_N(\boldsymbol{\sigma}) + h\sum_{i \le N} \sigma_i}\, d\lambda_N(\boldsymbol{\sigma}). \tag{1.33}$$

The dependence on the inverse temperature parameters (β_p) are encoded into the covariance structure ξ defined in (1.28), so we will not write them to simplify notation. The variational formula for the free energy was originally discovered by Crisanti and Sommers in [40] in 1992 using non-rigorous methods. The Parisi form of this functional was proven mathematically by Talagrand for even p-spin models in [98] and later extended to all mixed p-spin models by Chen [30]. The version of the formula we will present arises after solving the discrete Parisi formula for the spherical spin glass models for its Lagrange multiplier term and an extension from discrete probability measures to all probability measures by continuity [98, Section 4].

The Crisanti–Sommers Formula

We now state formula for the free energy in the form discovered by Crisanti and Sommers. Analogously to the Sherrington–Kirkpatrick model, the functional order parameters of these models can be interpreted as the limiting distributions of the overlaps under the Gibbs measure. Given a probability measure μ on $[0, 1]$, we denote its cdf $x : [0, 1] \to [0, 1]$ by

$$x(t) = \mu([0, t]).$$

Let $x^{-1}(q)$ denote the quantile transform of $x(t)$,

$$x^{-1}(q) = \inf\{t \in [0, 1] \mid q \le x(t)\}$$

and define $q_x = x^{-1}(1)$ to denote the largest point in the support of μ. We denote the Crisanti–Sommers functional by

$$\mathscr{C}(x) = \frac{1}{2}\bigg(\int_0^n x(t)(\xi'(t) + h^2)\, dt + \log|1 - q_x| + \int_0^{q_x} \frac{1}{\hat{x}(t)}\, dt \bigg), \tag{1.34}$$

where

$$\hat{x}(t) = \int_t^1 x(s)\, ds.$$

This functional defined on the space of probability measures on $[0,1]$ is the rate functional of (1.33).

Theorem 1.2.2 (*The Crisanti–Sommers Formula*)

The limit of the free energy (1.33) *equals*

$$\lim_{N\to\infty} F_N(\beta) = \inf_x \mathscr{C}(x), \tag{1.35}$$

where the infimum is taken over cumulative distribution functions on $[0,1]$.

In contrast to the Parisi formula for the Ising spins, this form of the variational formula is obviously convex, so its minimizer is unique. There is also an analogue of the Parisi form of this variational formula [98, Proposition 3.1] and [37, Equation (4)],

$$\mathscr{P}(x,b) = \frac{1}{2}\bigg(\frac{h^2}{b - \tilde{x}(0)} + \int_0^1 \frac{\xi''(t)}{b - \tilde{x}(t)}\, dt + b - 1 - \log b - \int_0^1 t\xi''(t)x(t)\, dt \bigg), \tag{1.36}$$

where $\tilde{x}(t) = \int_t^1 \xi''(s)x(s)\, ds$ and $b > \max(1, \tilde{x}(0))$. If we consider the pure 2-spin Hamiltonian, there is also an equivalent variational formula using the TAP approach [104] by Belius and Kistler [21]. The simpler form of the formula (1.35) has made it possible to prove several properties about the spherical models.

The Parisi and Crisanti–Sommers variational problems were studied in [15, 61, 98]. The Crisanti–Sommers formula has been applied to derive variational principles for the ground state energy in [17, 38, 60]. These variational formulas were used to explore related problems such as phase diagrams [62, 95], temperature chaos [29, 32, 33], disorder chaos and the geometry of the Gibbs measure [94, 97] and optimization alogorithms [96]. The application to temperature chaos is closely related to the multi-system models we study in this thesis and will be explained in more detail in Section 1.3.1. The dynamics of spherical spin glasses was also studied in [22, 42, 51]. Lastly, we refer the reader to [14, 62, 60, 32, 19, 38] for other recent work where various aspects of the spherical models have been studied.

1.2.3 The Vector Spin Glass Models

We now state a class of high dimensional models called vector spin models. These formulas arise when we consider coupled copies of the Ising spin glass models introduced earlier. These multi-systems models are used to study problems involving coupled systems that appear naturally when studying related problems such as temperature chaos [79], spectral gap estimates [12], and maximum cut on inhomogeneous random graphs [57]. These vector spin models encompass a large class of models such as the Potts spin glass [47], Heisenberg spin glass [24], and the Ghatak–Sherrington spin glass [50].

We fix $n \geq 2$. The main difference in the class of vector spin models in contrast to the models discussed before is the spin coordinates are now vectors in $\mathbb{R}^n$. Consider the *configuration space* $\Sigma_N \subseteq (\mathbb{R}^n)^N$. We

denote a *configuration* from Σ_N by $\vec{\boldsymbol{\sigma}} = (\vec{\sigma}_1, \ldots, \vec{\sigma}_N)$ where

$$\vec{\sigma}_i = (\sigma_i(1), \ldots, \sigma_i(n)) \in \mathbb{R}^n$$

denotes the vector spin coordinates and for $j \leq n$,

$$\boldsymbol{\sigma}(j) = (\sigma_1(j), \ldots, \sigma_N(j)) \in \mathbb{R}^N$$

denotes the configuration of the jth coordinate or copy. For each coordinate and inverse temperature vectors $(\vec{\beta}_p)_{p\geq 1}$, we let

$$H_N^j(\vec{\boldsymbol{\sigma}}) = \sum_{p\geq 1} \vec{\beta}_p(j) H_{N,p}^j(\boldsymbol{\sigma}(j)) \quad \text{where} \quad H_{N,p}^j(\vec{\boldsymbol{\sigma}}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1\ldots i_p} \sigma_{i_1}(j) \cdots \sigma_{i_p}(j), \quad (1.37)$$

denote the corresponding p-spin Hamiltonian (1.26) of the jth coordinate and let

$$H_N(\vec{\boldsymbol{\sigma}}) = \sum_{j\leq n} H_N^j(\vec{\boldsymbol{\sigma}}) \quad (1.38)$$

denote the Hamiltonian of the mixed vector p-spin model. Again, we assume that the vector of inverse temperature parameters $(\vec{\beta}_p)_{p\geq 1}$ are small enough so that the covariances of H_N is well defined. We restrict the study to even mixed p-spin models, which means that we assume $\vec{\beta}_p = 0$ for odd p. This ensures the covariance of the Hamiltonian is a convex function. This Hamiltonian coincides with the Gibbs measure associated with n independent samples of the configurations $\boldsymbol{\sigma}(j)$ from the same disorder in the even mixed p-spin Hamiltonian.

For these models, we assume that the reference measure on Σ_N can be expressed as the product of measures μ on $\mathbb{R}^n$ with compact support. We define

$$\mathscr{D} = \text{conv}\{\vec{\sigma} \otimes \vec{\sigma} \mid \vec{\sigma} \in \text{supp}(\mu)\} \quad (1.39)$$

to be the closed convex hull of the possible self overlaps. For $\boldsymbol{Q} = (Q^{i,j})_{i,j\leq n} \in \mathscr{D}$, we define

$$Q_N^\varepsilon = \left\{ \vec{\boldsymbol{\sigma}} \in \Sigma_N \ : \ \frac{\boldsymbol{\sigma}(i) \cdot \boldsymbol{\sigma}(j)}{N} \in [Q^{i,j} - \varepsilon, Q^{i,j} + \varepsilon] \right\} \quad (1.40)$$

to be the epsilon ball around $\boldsymbol{Q}$. Our goal will be to study constrained vector spin free energies of the form

$$F_N^\varepsilon(\boldsymbol{Q}) = \frac{1}{N}\mathbb{E} \log \int_{Q_N^\varepsilon} e^{H_N(\vec{\boldsymbol{\sigma}})} \, d\mu^{\otimes N}(\vec{\boldsymbol{\sigma}}). \quad (1.41)$$

The results for unconstrained free energies are described in [80, Theorem 1]. The variational formula for these vector spin free energies was proved by Panchenko [80] in 2018. This work builds on the techniques used to study the multispecies spin glass [78] and the Potts spin glass [81].

The key to these proofs relied on the ultrametric structure of the limiting distribution in the Gibbs measure originally proved for the Sherrington–Kirkpatrick model in [75]. The ultrametric structure of the marginals in the joint overlap arrays in these multiple system models imply a sychronization property for the generalized joint overlap arrays that define the covariances of the vector spin Hamiltonians

[Section 3.2]. As a consequence the limiting distribution of the Gibbs measures can be approximated by the same Ruelle probability cascades as observed in the one-dimensional models via a perturbation of the Hamiltonian. This ultrametric description of the limiting distribution is consistent with the ultrametric functional order parameters originally proposed by Parisi for the Sherrington–Kirkpatrick models described earlier in Section 1.2.1.

The Parisi Formula for Even Mixed Vector p-Spin Models

We now state the variational formula for the vector spin free energy (1.41). The limit of the free energy is equal to the minimization of a functional on the space of matrix paths. We present the discrete representation of this formula, which can be thought of as the high dimensional version of the Ising spin models in Section 1.2.1.

The *functional order parameter* of this model are sequences of parameters encoded by sequences

$$\begin{aligned} 0 = \ x_0 < \ x_1 < \ldots < \ x_{r-2} < \ x_{r-1} \leq \ 1 \\ \mathbf{0} = \boldsymbol{Q}_0 \leq \boldsymbol{Q}_1 \leq \ldots \leq \boldsymbol{Q}_{r-2} \leq \boldsymbol{Q}_{r-1} \leq \boldsymbol{Q} \end{aligned}, \tag{1.2.3}$$

with the convention that $\boldsymbol{Q}_r = \boldsymbol{Q}$ and $x_r = 1$. By continuity, we may take these inequalities to be strict. Given a monotone matrix path of positive semidefinite $n \times n$ matrices $\boldsymbol{\Phi} : \mathbb{R}^+ \to \mathbb{S}_n^+$ and the cdf of a discrete random variable $x(t) = \mu([0,t])$ on $\mathbb{R}^+$, these sequences (1.2.3) encode the quantile function $\boldsymbol{\pi}$ of a particular matrix valued random variable, i.e. for $1 \leq k \leq r$

$$\boldsymbol{\pi}(t) = \boldsymbol{\Phi}(x^{-1}(t)) = \boldsymbol{Q}_k \qquad \text{for} \qquad x_{k-1} < t \leq x_k.$$

Similarly to the order parameters in the Sherrington–Kirkpatrick model [Section 1.2.1] and spherical spin glass models [Section 1.2.2], the functional order parameter $\boldsymbol{\pi}$ describes the distribution of a matrix valued overlap in vector spin models.

We now describe the Parisi functional for vector spin models in terms of a recursive quantity. For $\boldsymbol{A} \in \mathbb{S}_+^n$, we define the functions

$$\boldsymbol{\xi}(\boldsymbol{A}) = \sum_{p \geq 2} (\vec{\beta}_p \otimes \vec{\beta}_p) \odot \boldsymbol{A}^{\odot p}, \tag{1.42}$$

and

$$\boldsymbol{\xi}'(\boldsymbol{A}) = \sum_{p \geq 2} p(\vec{\beta}_p \otimes \vec{\beta}_p) \odot \boldsymbol{A}^{\odot (p-1)} \quad \text{and} \quad \boldsymbol{\theta}(\boldsymbol{A}) = \sum_{p \geq 2} (p-1)(\vec{\beta}_p \otimes \vec{\beta}_p) \odot \boldsymbol{A}^{\odot p}, \tag{1.43}$$

where $\odot$ is the Hadamard product on $n \times n$ matrices and $\boldsymbol{A}^{\odot p}$ is the pth Hadamard power of $\boldsymbol{A}$. Since $\vec{\beta}_p = 0$ for odd p, these matrix valued functionals are convex in its coordinates.

Consider a sequence of centered Gaussian vectors $(\vec{z}_k)_{1 \leq k \leq r}$ with covariance

$$\mathrm{Cov}(\vec{z}_k) = \boldsymbol{\xi}'(\boldsymbol{Q}_k) - \boldsymbol{\xi}'(\boldsymbol{Q}_{k-1}).$$

Because the sequences of matrices are monotone, these Gaussian vectors are well defined by the Schur product theorem. Given symmetric $\boldsymbol{\Lambda} \in \mathbb{R}^{n \times n}$ called a *Lagrange multiplier*, we define

$$X_r = \log \int_{\mathbb{R}^n} \exp\left(\sum_{1 \leq k \leq r} (\vec{\omega}, \vec{z}_k) + (\boldsymbol{\Lambda}\vec{\omega}, \vec{\omega}) \right) d\mu(\vec{\omega})$$

where $(\cdot,\cdot)$ is the inner product on $\mathbb{R}^n$ and, recursively for $1 \le k \le r-1$, we define

$$X_k = \frac{1}{x_k} \log \mathbb{E}_k \exp(x_k X_{k+1}) \iff \exp(x_k X_k) = \mathbb{E}_k \exp(x_k X_{k+1}),$$

where $\mathbb{E}_k$ denotes the expectation with respect to $\vec{z}_{k+1}$. If $x_k = 0$ then we interpret the formula to mean

$$X_k = \mathbb{E}_k X_{k+1}.$$

The quantity X_0 is nonrandom, and we will denote it with $\mathcal{P}(\pi)$. Consider the functional

$$\mathscr{P}(\boldsymbol{\Lambda}, \boldsymbol{\pi}) = \mathcal{P}(\boldsymbol{\pi}) - \langle \boldsymbol{\Lambda}, \boldsymbol{Q} \rangle - \frac{1}{2} \sum_{1 \le k \le r-1} x_k \operatorname{Sum}(\boldsymbol{\theta}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\theta}(\boldsymbol{Q}_k)),$$

where $\operatorname{Sum}(\boldsymbol{A})$ denotes the sum of the entries of $\boldsymbol{A}$. This functional defined on the space of probability measures on monotone matrix paths $\boldsymbol{\pi} : \mathbb{R}^+ \to \mathbb{S}_n^+$ is the rate functional of (1.10).

Theorem 1.2.3

For any $n \ge 1$, the limit of the free energy is given by

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} F_N^{\varepsilon}(\boldsymbol{Q}) = \inf_{\pi, \boldsymbol{\Lambda}} \mathscr{P}(\boldsymbol{\Lambda}, \boldsymbol{\pi}) \tag{1.44}$$

where the infimum is taken over Lagrange multipliers and monotone matrix paths $\boldsymbol{\pi}$ encoded by (1.2.3).

The synchronization techniques used to prove the free energy formulas for vector spin models was used recently in other contexts in [57, 39]. In this thesis we give another application. The main contribution of this thesis is a similar vector spin formula for spherical spin glasses. As with the jump from the Ising model to the spherical model, the jump from the vector spin models to the spherical vector spin models is non-trivial.

Remark 1.2.2. In this thesis, bold face capitals such as $\boldsymbol{\Phi}$ and $\boldsymbol{A}$ usually denote matrices in $\mathbb{R}^{n \times n}$. Symbols with accent arrows such as $\vec{x}$ and $\vec{\sigma}$ denote vectors in $\mathbb{R}^n$. Bold Greek letters such as $\boldsymbol{\sigma}$ and $\boldsymbol{\omega}$ denote configurations in $\mathbb{R}^N$. Non formatted numbers such as x and q denotes numbers in $\mathbb{R}$. Later on, $\underline{x}$ and $\underline{\boldsymbol{Q}}$ will denote sequences of real numbers and positive semidefinite matrices.

1.3 Overview of the Main Results

We now introduce the main contribution of the thesis. Our main objective is to prove a variational formula for the vector spin analogue of the spherical spin glass model. This model shows up naturally when one wants to study questions about the support of the array of inner products generated from replicas sampled according to the Gibbs measure of spherical models. These contributions can be found in the author's recent papers [64] and [63] and are expanded on in Chapters 4 and Chapters 5 of this thesis. We begin with some motivation for the spherical vector spin models in the context of the spherical spin glass models in Section 1.2.2.

1.3.1 The Spherical Vector Spin Model

We consider n independent samples of spherical configurations sampled according to the Gibbs measure associated with a mixed p-spin Hamiltonian H_N^j with external field $\vec{h}$ defined in (1.37). The notation in this section is identical to the vector spin notation introduced in Section 1.2.3. In the spherical models, these configurations take values in S_N, the sphere of radius $\sqrt{N}$. In particular, the replica $\boldsymbol{\sigma}^j = \boldsymbol{\sigma}(j)$ are sampled according to the Gibbs measure G_N^j on S_N defined by

$$G_N^j(\boldsymbol{\sigma}(j)) \propto e^{H_N^j(\vec{\boldsymbol{\sigma}}) + \vec{h}(j)\sum_{i=1}^N \sigma_i(j)}. \tag{1.45}$$

These samples have a common disorder, but can be sampled at different temperatures. That is, the $g_{i_1,\dots,i_p}$ appearing in H_N^j are the same for $j \le n$, but the sequence of parameters $(\vec{\beta}_p(j))$ and $\vec{h}(j)$ may be different for each sample $j \le n$. Our goal is to understand the distribution of the overlap array generated by these replica,

$$\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) = \left(\frac{\boldsymbol{\sigma}(j)\cdot\boldsymbol{\sigma}(j')}{N}\right)_{j,j'\le n}. \tag{1.46}$$

The matrix $\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}})$ is called the *self overlap* of the vector spin configuration. Let $\boldsymbol{Q}$ be a positive definite matrix such that $Q^{i,i} = 1$ for $i \le n$ and $Q^{i,j} \in [-1,1]$ for $i \ne j$. Given ε, we say $\boldsymbol{A} \approx \boldsymbol{B}$ if the entries of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are within ε, i.e. $|A^{i,j} - B^{i,j}| \le \varepsilon$ for all $i, j \le n$. A natural question is what is the probability that the overlap array takes value $\boldsymbol{Q}$ under the average Gibbs measure

$$\mathbb{E}\big\langle\, \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\big\rangle, \tag{1.47}$$

where $\langle\cdot\rangle$ is the average with respect to the product Gibbs measure,

$$G_N(\vec{\boldsymbol{\sigma}}) = \prod_{i=1}^n G_N^j(\boldsymbol{\sigma}(j)) \propto e^{H_N(\vec{\boldsymbol{\sigma}}) + \sum_{i=1}^N (\vec{h}, \vec{\sigma}_i)} \tag{1.48}$$

since $H_N(\vec{\boldsymbol{\sigma}}) = \sum_{j\le n} H_N^j(\vec{\boldsymbol{\sigma}})$. The large deviations [Remark 1.3.1 and Section 4.1] are encoded by the constrained free energies of the form

$$F_N^\varepsilon(\boldsymbol{Q}) = \frac{1}{N}\mathbb{E}\log \int_{Q_N^\varepsilon} e^{H_N(\vec{\boldsymbol{\sigma}}) + \sum_{j=1}^n \vec{h}(j)\sum_{i=1}^N \sigma_i(j)}\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}) \tag{1.49}$$

where λ_N is the normalized uniform surface measure on the sphere S_N and $\lambda_N^n = \lambda_N^{\otimes n}$ is its product measure on $S_N^n = S_N^{\otimes n}$. The set $Q_N^\varepsilon \subseteq S_N^n$ is the set of spins with constrained self overlaps defined

$$Q_N^\varepsilon = \big\{\vec{\boldsymbol{\sigma}} \in S_N^n \mid \|\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) - \boldsymbol{Q}\|_\infty \le \varepsilon\big\}. \tag{1.50}$$

We call these the constrained free energies, and our goal will be to find a variational formula for the limit of (1.49) as $N \to \infty$ and $\varepsilon \to 0$. In physics, these are called *off equilibrium* free energies, because we constrain the overlaps to lie in a non-equilibrium state.

A variational formula for the constrained vector spin free energies (1.49) will give us a tool to determine likely values of the overlaps generated by independent samples from the spherical Gibbs

measures. To see how (1.49) is related to (1.47), notice that

$$\big\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\big\rangle = \int_{S_N^n} \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q}) \frac{e^{H_N(\vec{\boldsymbol{\sigma}})}}{\int_{S_N^n} e^{H_N(\vec{\boldsymbol{\sigma}})}\, d\lambda_N^n(\vec{\boldsymbol{\sigma}})}\, d\lambda_N^n(\sigma). \tag{1.51}$$

The normalization term decouples into the product of one dimensional spherical free energies (1.33), so

$$\begin{aligned} &\frac{1}{N}\log\big\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\big\rangle \\ &= \frac{1}{N}\log \int_{Q_N^\varepsilon} e^{H_N(\vec{\boldsymbol{\sigma}}) + \sum_{j=1}^n \vec{h}(j)\sum_{i=1}^N \sigma_i(j)}\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}) - \sum_{j=1}^n \frac{1}{N}\log \int_{S_N} e^{H_N^j(\boldsymbol{\sigma}) + \vec{h}(j)\sum_{i=1}^N \sigma_i}\, d\lambda_N^n(\boldsymbol{\sigma}) \end{aligned} \tag{1.52}$$

decomposes into the difference of the random constrained free energy (1.49) and a sum of random one dimensional free energies (1.33). If we can show that (1.52) is strictly less than $-\delta$ with high probability, then we can conclude with high probability that

$$\big\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\big\rangle \le e^{-\delta N}, \tag{1.53}$$

which means that the configuration $\boldsymbol{Q}$ is exponentially unlikely under the product Gibbs measure $\prod_{j=1}^n G_N^j$ on S_N^n. By Gaussian concentration [Proposition 4.1.1], it suffices to study the asymptotics of these quantities averaged over the disorder $g_{i_1,\dots,i_p}$,

$$\frac{1}{N}\mathbb{E}\log\big\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\big\rangle = F_N^\varepsilon(\boldsymbol{Q}) - \sum_{j=1}^n F_N^j \tag{1.54}$$

where $F_N^\varepsilon(\boldsymbol{Q})$ is the constrained free energy (1.49) and F_N^j are the one-dimensional free energies (1.33) with respect to temperature parameters $(\beta_p(j))$ and external field $\vec{h}(j)$. The primary objective of this thesis is to prove a free energy formula for $F_N^\varepsilon(\boldsymbol{Q})$.

Applications of the Free Energy Formulas

In the previous section, we concluded that the free energy for a given constraint (1.50) is strictly less than the sum of the one dimensional free energies (1.54) discussed in Section 1.2.2, then the configuration set Q_N^ε is exponentially unlikely [Proposition 4.1.1]. This gives us a tool to test valid configurations for the replica matrix in spherical spin glasses. The spherical vector spin Hamiltonians are general enough to include cases when the replicas may be taken at different temperatures with different external fields.

An explicit example where this formula can be directly applied is the temperature chaos problem mentioned in Section 1.2.2. Chaos in temperature means that the set of likely configurations looks quite different under small changes to the temperature. In particular, samples of two spin configurations $\boldsymbol{\tau}, \boldsymbol{\rho}$ from the Gibbs distributions at different temperatures result in an almost uncorrelated overlap,

$$\lim_{N\to\infty} \mathbb{E}\Big\langle \Big| \frac{\boldsymbol{\tau}\cdot\boldsymbol{\rho}}{N} \Big| \Big\rangle = 0 \tag{1.55}$$

where $\langle\cdot\rangle$ denotes the product of the Gibbs measures (1.45) at the different temperates. It was proved in [33] that chaos in temperature occurs in some models under an uncoupling condition on the supports of the individual Parisi measures [33, Theorem 2]. It remains an interesting problem to consider the

behavior in some full-RSB spherical models where this uncoupling condition fails. It was predicted that chaos in temperature for spherical models cannot be detected at the level of the free energy [89]. A recent geometric proof of this fact using the free energy landscapes was proved by Subag in [97]. It is still an interesting problem to verify this result as a direct consequence of the free energy and limiting behavior of the overlap arrays. The variational formulas for vector spin models are complicated and poses a difficult, but meaningful, analytical problem to study in future work.

Related work on the chaos in temperature for Ising spin models was studied in [79]. The disorder chaos problem for spin glasses was studied by Chen in [29] and Chatterjee [25, 26, 27]. In addition to the chaos problem, the free energies of coupled systems of spherical vector spin models have appeared in other contexts. In [12, 51], spectral gap estimates for generic spherical models were proved under various conditions on the Parisi measure. The free energy formulas can be used to prove large deviation principles [Remark 1.3.1] to extend these spectral gap estimates to the larger class of mixed even-p-spin spherical models. These formulas can also be used to study Franz–Parisi potential introduced in [92] and the related predictions on the landscapes of these free energies. Lastly, the coupled systems also appear in statistical analysis problems such as tensor principle component analysis [58].

Remark 1.3.1. The difference of free energies (1.54) in vector spin models defines an almost sure large deviations principle for the distribution of the overlap matrix [12, Theorem 1.13]. Consider a sequence of random probability measures on $\mathbb{R}^{n\times n}$

$$\mu_N(\cdot) = \langle R(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \in \cdot \rangle, \tag{1.56}$$

where $R(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}})$ is the overlap array generated from samples $\boldsymbol{\sigma}^1, \ldots, \boldsymbol{\sigma}^n$ samples independently from G_N^j defined in (1.45). The almost sure large deviations principle for μ_N follows as another consequence of the comparison of the free energies (1.54),

$$\lim_{\varepsilon\to 0}\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\log\langle R(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q}\rangle = \lim_{\varepsilon\to 0}\lim_{N\to\infty}\left(F_N^{\varepsilon}(\boldsymbol{Q}) - \sum_{j=1}^{n} F_N^j\right) = -I(\boldsymbol{Q}). \tag{1.57}$$

Gaussian concentration of the free energies [Theorem 2.1.3] implies that the same rate function $I(\boldsymbol{Q})$ holds almost surely for μ_N without averaging over the $g_{i_1,\ldots,i_p}$ in the Hamiltonians. The rate function $I(\boldsymbol{Q})$ is given precisely by the variational formulas for the limit of the free energies (1.49) and (1.33), which we define in Theorem 1.3.2 and Theorem 1.2.2 and formula for this rate function is stated precisely in equation (4.40).

1.3.2 Parisi Form of the Free Energy

The main contribution of this thesis is the free energy of multiple copies of spherical spin glasses with constrained overlaps introduced in [48, 49]. We now state the variational formula for the free energy defined in (1.49). For $\boldsymbol{A} \in \mathbb{S}_+^n$, we recall the functions (1.42) and (1.43)

$$\boldsymbol{\xi}(\boldsymbol{A}) = \sum_{p\geq 2}(\vec{\beta}_p \otimes \vec{\beta}_p) \odot \boldsymbol{A}^{\odot p}, \tag{1.58}$$

and

$$\boldsymbol{\xi}'(\boldsymbol{A}) = \sum_{p\geq 2} p(\vec{\beta}_p \odot \vec{\beta}_p) \odot \boldsymbol{A}^{\odot(p-1)} \quad \text{and} \quad \boldsymbol{\theta}(\boldsymbol{A}) = \sum_{p\geq 2}(p-1)(\vec{\beta}_p \otimes \vec{\beta}_p) \odot \boldsymbol{A}^{\odot p}, \tag{1.59}$$

where $\odot$ is the Hadamard product on $n \times n$ matrices and $\boldsymbol{A}^{\odot p}$ is the pth Hadamard power of $\boldsymbol{A}$. These functions are the matrix analogues of the covariance functions ξ and θ defined in (1.19). Consider the sequences of real numbers and positive semidefinite matrices of the form

$$\begin{aligned} 0 = x_0 \leq x_1 \leq \ldots \leq x_{r-2} \leq x_{r-1} \leq x_r = 1 \\ \boldsymbol{0} = \boldsymbol{Q}_0 \leq \boldsymbol{Q}_1 \leq \ldots \leq \boldsymbol{Q}_{r-2} \leq \boldsymbol{Q}_{r-1} \leq \boldsymbol{Q}_r = \boldsymbol{Q} \end{aligned} . \tag{1.60}$$

These sequences encode a monotone matrix path known as the *functional order parameter* for the vector spin models. The r-step discretization of the Parisi functional for spherical vector spin glasses is defined by

$$\begin{aligned} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) = \frac{1}{2}\Big[\langle \vec{h}\vec{h}^{\intercal}, \boldsymbol{\Lambda}_1^{-1}\rangle + \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - n - \log|\boldsymbol{\Lambda}| + \sum_{1 \leq k \leq r-1} \frac{1}{x_k} \log \frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|} + \langle \boldsymbol{\xi}'(\boldsymbol{Q}_1), \boldsymbol{\Lambda}_1^{-1}\rangle \\ - \sum_{1 \leq k \leq r-1} x_k \operatorname{Sum}\big(\boldsymbol{\theta}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\theta}(\boldsymbol{Q}_k)\big)\Big] \end{aligned} \tag{1.61}$$

where $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \operatorname{tr}(\boldsymbol{A}\boldsymbol{B})$ is the Frobenius inner product on symmetric matrices, $|\cdot|$ is the determinant, $\operatorname{Sum}(\boldsymbol{A})$ is the sum of the entries of the matrix and

$$\boldsymbol{\Lambda}_r = \boldsymbol{\Lambda}, \quad \boldsymbol{\Lambda}_p = \boldsymbol{\Lambda} - \sum_{p \leq k \leq r-1} x_k\big(\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_k)\big) \text{ for } 1 \leq p \leq r-1. \tag{1.62}$$

To simplify notation, we also set $\underline{x} = (x_k)_{k=0}^r$ and $\underline{\boldsymbol{Q}} = (\boldsymbol{Q}_k)_{k=0}^r$ denote the sequences in (1.60). The parameter $\boldsymbol{\Lambda}$ is a Lagrange multiplier term that appears when decoupling the constraint on the self overlaps. It was shown in [64] that the free energy (1.49) is given by the minimizer of the Parisi functional (1.61).

Theorem 1.3.1 (*Vector Spin Parisi Formula*)

The limit of the free energy with self overlaps constrained to $\boldsymbol{Q}$ equals

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} F_N(\boldsymbol{Q}_N^{\varepsilon}) = \inf_{r, \Lambda, x, Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}). \tag{1.63}$$

The infimum is over sequences of the form (1.60), $\boldsymbol{\Lambda}$ such that $|\boldsymbol{\Lambda}_1| > 0$, and all $r \geq 1$.

We now explain the mathematical and physics contributions cumulating to the variational formula for the free energy (1.63). Multiple copies of mixed even-p-spin spherical spin glasses with constrained self overlaps was first studied in [49, 48] using non-rigorous methods. A rigorous upper bound for the free energy of this model was proved in [83] by Panchenko and Talagrand using the Guerra replica symmetry breaking bound [54]. The sharp lower bound was proved in [64] by the author of this thesis using the synchronization mechanism described in [78, 80, 81] and the Aizenman–Sims–Starr scheme [2] for spherical spin glasses described in [30]. These results are a consequence of the ultrametric structure of generalized overlaps that satisfy the Ghirlanda–Guerra identities [52, 53] which was proved in [75]. Combining the upper and lower bound results in a discrete Parisi variational formula for the free energy of spherical spin glasses with vector spins. The Crisanti–Sommers representation of this free energy formula was proved by the author of the thesis in [63] using logarithmic barrier functions to partially solve the variation problem by removing the Lagrange multiplier constraint. Recently, Auffinger and Zhou proved several properties about the minimizers of the vector spin Crisanti–Sommers and Parisi

functionals as well as a variational formula for the ground state in this model in [20].

As stated earlier, the upper bound of the free energy of this model was proved in [83]. The first objective of this thesis is to prove that this upper bound is sharp. The extension to the vector spin spherical models remains non-trivial despite the vast amount of work done for vector spin and spherical models. The first difficulty is the lack of the product space structure in the constrained spherical models in contrast to the vector spin models discussed in Section 1.2.3. To overcome this, we prove an approximate product measure structure for the uniform measure in the Poincaré limit [Theorem 3.1.1]. We use this fact to prove a vector version of the Aizenman–Sims-Starr scheme [Section 4.4] to reduce the spherical vector spin model to a form where the synchronization of vector spin glass [Section 3.2.1] can be applied directly. The spherical constraint on the coordinates adds some technical difficulties that must also must resolved when extending from the vector spin case to the spherical vector spin case. This proof is discussed in more detail in Chapter 4.

1.3.3 Crisanti–Sommers Form of the Free Energy

The equivalence of the one dimensional Parisi formula and the Crisanti–Sommers formula for spherical spin glasses was proved in [98] by showing that both functionals satisfy the same critical point conditions. In this thesis we prove the analogue of the Crisanti–Sommers functional for the spherical vector spin models and show that the limit of the free energy is obtained at the minimum of this functional. This variational formula for one dimensional vector spins is consistent with the classical Crisanti–Sommers formula. As we will see shortly, it is considerably simpler to analyze the properties in this form of the functional, which is a step in the direction to understanding the coupled systems discussed in Section 1.3.1.

A simplification of the vector spin Parisi functional (1.61) implies that the free energy in Theorem 1.3.1 can be obtained by minimizing a simpler higher dimensional functional resembling the Crisanti–Sommers functional for one dimensional spherical spin glasses. The parameters of the functional is the cdf of the trace of the overlap matrix, and the synchronized matrix path identifying the overlap matrix with its trace [80, Theorem 4]. Let

$$x(t) : [0, n] \to [0, 1] \quad \text{such that} \quad x(0) = 0 \quad \text{and} \quad x(n) = 1 \tag{1.64}$$

denote a right continuous non-decreasing function and

$$\boldsymbol{\Phi}(t) : [0, n] \to \mathbb{S}^n_+ \quad \text{such that} \quad \mathrm{tr}(\boldsymbol{\Phi}(t)) = t \quad \text{and} \quad \boldsymbol{\Phi}(0) = \mathbf{0} \quad \text{and} \quad \boldsymbol{\Phi}(n) = \boldsymbol{Q} \tag{1.65}$$

denote a 1-Lipschitz monotone matrix path in the space of $n \times n$ positive semidefinite matrices parametrized by its trace. A monotone matrix path is one with positive semidefinite increments, $\boldsymbol{\Phi}(t_2) - \boldsymbol{\Phi}(t_1) \in \mathbb{S}^n_+$ for $t_2 \geq t_1$. The largest point in the support of the measure associated with the c.d.f. $x(t)$ is denoted by

$$t_x := x^{-1}(1) = \inf\{t \in [0, n] \mid 1 \leq x(t)\}.$$

Assuming that $|\boldsymbol{Q} - \boldsymbol{\Phi}(t_x)| > 0$, we define the quantity

$$\mathscr{C}(x, \boldsymbol{\Phi}) = \frac{1}{2}\bigg(\int_0^n x(t) \langle \boldsymbol{\xi}'(\boldsymbol{\Phi}(t)) + \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{\Phi}'(t) \rangle \, dt + \log |\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(t_x)| + \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t) \rangle \, dt \bigg), \tag{1.66}$$

where $\hat{\boldsymbol{\Phi}}(t) : [0, n] \to \mathbb{R}^{n \times n}$ is a decreasing matrix path given by

$$\hat{\boldsymbol{\Phi}}(t) = \int_t^n x(s)\boldsymbol{\Phi}'(s)\, ds. \tag{1.67}$$

This variational formula is equivalent to the Parisi functional (1.61).

Theorem 1.3.2 (*Vector Spin Crisanti–Sommers Formula*)

The limit of the free energy with self overlaps constrained to $\boldsymbol{Q}$ *equals*

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} F_N^{\varepsilon}(\boldsymbol{Q}_N^{\varepsilon}) = \inf_{x, \boldsymbol{\Phi}} \mathscr{C}(x, \boldsymbol{\Phi}). \tag{1.68}$$

The infimum is over $x(t)$ *and* $\boldsymbol{\Phi}(t)$ *defined in* (1.64) *and* (1.65) *such that* $|\boldsymbol{Q} - \boldsymbol{\Phi}(t_x)| > 0$.

The proof of this fact follows the approach used to prove the equivalence of the Parisi form and Crisanti–Sommers form of the one-dimensional free energy formulas in Section 1.2.2. The proof for the one-dimensional models in [98, Section 4] computes the first variation of the discrete analogues of the Parisi form and Crisanti–Sommers form of the free energy to prove critical point conditions. The minimizers of both functionals satisfy identical critical point conditions, which will allow one to reduce from one form of the discrete formula to the other and vice versa. The key property of this computation is the strict monotonicity of the functional order parameters in the discrete models (1.2.1) dicussed in Section 1.2.1 and Section 1.2.2. By continuity, we can assume that

$$0 = q_0 < q_1 < \ldots < q_{r-2} < q_{r-1} = 1, \tag{1.69}$$

so the derivatives of the functionals with respect to the order parameters are equal 0 at the minimizer. This key fact allows us to prove critical point equations [98, Lemma 4.1] that relate the strictly increasing sequence (1.69) with the Lagrange multiplier to simplify the discrete Parisi form of the free energy. A similar reduction can be made for spherical vector spin models after we overcome a non-trivial obstacle in the derivation of the vector spin critical point equations.

In vector spin models [Section 1.2.3], the discrete functional order parameters are encoded by sequences of monotone matrices,

$$\mathbf{0} = \boldsymbol{Q}_0 \leq \boldsymbol{Q}_1 \leq \ldots \leq \boldsymbol{Q}_{r-2} \leq \boldsymbol{Q}_{r-1} \leq \boldsymbol{Q}_r = \boldsymbol{Q}. \tag{1.70}$$

In contrast to the one-dimensional models, this sequence may not be strictly increasing. In the vector spin models, the directional derivatives of the functionals at the minimizer preserving the monotonicity of the paths are non-negative. The critical point equations that one will get if we naively compute directional derivatives will result in a system of inequalities, which is insufficient to prove the equivalence of the Parisi and Crisanti–Sommers forms of the free energy.

To overcome this obstacle we use logarithmic barrier functions to force the minimizer to be strictly monotone by adding an infinitely large penalty when the matrix paths (1.70) are not strictly monotone. This will allow us to prove a system of approximate critical point equations for to prove the equivalence of some modified Parisi and Crisanti–Sommers forms. The removal of the logarithmic barrier is non-trivial because the penalties must be infinitely large, so simply taking the size of the barrier to zero is insufficient. Instead, we use a convexity argument to remove the barrier in the limit. The approximate

critical point equations become exact in the limit, so we can prove Crisanti–Sommers functional (1.66) and the Parisi functional (1.61) are equivalent. This proof is discussed in more detail in Chapter 5.

1.4 Future Areas of Research

In this section, we introduce some potential research areas using the vector spin free energies. These are problems are natural follow-ups to the work presented in the thesis. These problems are currently a work in progress.

TAP-Plefka Free Energy Formula

This thesis is focused on Parisi's replica symmetry breaking approach to solving spin glasses. There was another approach involving a skeleton expansion of the configuration space along maximal directions. This approach originally developed by physicists Thouless, Anderson, and Palmer [104] was adapted to the spherical Sherrington–Kirkpatrick model by Belius and Kistler [21]. It should be possible to prove a similar free energy formula for the vector spin model using a similar skeleton expansion without relying on any results involving the limiting distribution of the overlap array.

Hamilton Jacobi Free Energy Formula

Recently, there has been interest in studying the limit of the free energy using a techniques from partial differential equations. This approach expresses the limit of the free energy of the classical Sherrington–Kirkpatrick model as the solution of a Hamilton Jacobi equation [73]. Perhaps a similar method can be used to prove a formula for the vector spin models in terms of the solution of a Hamilton Jacobi partial differential equation.

Ground State Free Energy

The one dimensional Crisanti–Sommers formula was used as the starting point for deriving two equivalent variational formulas of the ground state free energy in [38] and [60]. A similar formula for the Ising spin models was studied in [18]. These formulas were used to describe the phase diagram of spherical spin glasses and study other problems such as disorder chaos. The methodology used to study these models was adapted to the vector spin case in a recent paper by Auffinger and Zhou [20]. In that paper, the Crisanti–Sommers formula for vector spin models was used to prove a similar variational formula for the ground state of coupled copies of spherical spin glasses [20, Theorem 6]. This can be used to further study questions about the vector spin models such as the phase diagram and algorithmic optimization of the vector spin Hamiltonians.

Statistical Interference

Spin glasses also has multiple applications in statistical inference [108]. Vector spin models show up naturally in constrained low-rank matrix estimation [67]. These symmetric vector spin models with *Boltzmann probability measures* are of the form

$$\mathbb{P}(X|Y) = \frac{1}{Z_X(Y)} \prod_{1\leq i\leq N} \mathbb{P}_X(\vec{x}_i) \prod_{1\leq i,j\leq N} e^{g(Y_{ij}, \frac{\vec{x}_i^{\intercal}\vec{x}_j}{\sqrt{N}})}$$

where $Y_{ij} \in \mathbb{R}$, $X \in \mathbb{R}^{N \times n}$, and $\vec{x} \in \mathbb{R}^n$. When $\mathbb{P}_X$ is a standard Gaussian on $\mathbb{R}^n$ and $Y_{ij} = g_{ij}$

$$g\Big(Y_{i,j}, \frac{\vec{x}_i^{\mathsf{T}} \vec{x}_j}{\sqrt{N}}\Big) = \beta Y_{i,j} \frac{\vec{x}_i^{\mathsf{T}} \vec{x}_j}{\sqrt{N}} = \frac{\beta}{\sqrt{N}} \sum_{k=1}^{n} \sum_{i,j=1}^{N} g_{ij} x_i(k) x_j(k)$$

then this model is equivalent to the spherical vector spin Sherrington–Kirkpatrick model.

In a standard statistical inference problem, we want to estimate the ground truth X from some observed data noisy data Y. In this setting, if g is the log likelihood, of Y given the observation $x_i^{\mathsf{T}} x_j$ then the vector spin model is the posterior probability under Bayesian inference. A common question is to find some conditions on the models, such as the ratio of the size of the noise to the observations, to guarantee that the inner products generated of the observations only take one value. This is called a *replica symmetric* solution. Rigorous results applied to the low rank matrix estimation problem is discussed in [66]. Understanding the phase transitions of the vector spin models is one approach to finding such conditions for general vector spin inference problems.

Chapter 2

Preliminaries for Classical Models

In this chapter we summarize some results in probability theory and spin glass theory. These standard results are essential in the probabilistic derivation of the free energy formula using the asymptotic Gibbs measure developed by Panchenko and Talagrand. Readers familiar with the Sherrington–Kirkpatrick model can safely skip this chapter. The proofs of these results can be found in various textbooks on these subjects, so we will not recreate the proofs here unless the details of the proof is useful to understand the derivation of the spherical vector spin free energies in Chapter 4.

The results in this chapter apply to spin glass models with configurations in finite or countable set Σ and a measure $G(\boldsymbol{\sigma})$ on Σ. As seen in Chapter 1, when Σ is taken to be the N dimensional hypercube $\{-1,+1\}^N$ we get the classical Sherrington–Kirkpatrick model. We will introduce an extension of these results to spherical and vector spin models in the next chapter.

Outline of the Chapter

In Section 2.1, we introduce several elementary facts about Gaussian processes and functions of Gaussian processes. Next in Section 2.2, we review the properties of the Ruelle probability cascades and its connection with the limiting distribution of the overlaps in spin glass models. Finally in Section 2.3, we will explain the Ghirlanda–Guerra identities and its connections between the Ruelle probability cascades and the asymptotic Gibbs measure by means of a regularizing perturbation of the Gibbs measure. These are key techniques and results will be applied when we prove free energy formulas for spin glasses.

2.1 Gaussian Distributions

In this section, we prove several facts about Gaussian distributions that will be used throughout the proofs of the free energy formulas. These techniques will be applied throughout the thesis. These results will be stated for Gaussian processes indexed by finite or countable configuration spaces. We will these results to Gaussian processes indexed by points on the sphere in the next section.

2.1.1 Gaussian Integration by Parts

Gaussian integration by parts by is the main tool used to simplify averages with respect to Gibbs measure. It is also applied in the interpolation method used by Guerra [54]. It follows from the following

characterization of the standard normal.

Theorem 2.1.1 (*Stein's Lemma*)

If g has standard normal distribution, then

$$\mathbb{E}gf(g) = \mathbb{E}f'(g) \tag{2.1}$$

for all absolutely continuous functions such that $\mathbb{E}|f'(g)| < \infty$. Conversely, if (2.1) holds for all bounded, continuous and piecewise continuously differentiable functions f with $\mathbb{E}|f'(g)| < \infty$ then g is standard normal.

Proof. We present the proof adapted from [28, Lemma 2.1].
($\Longrightarrow$) Suppose that $\mathbb{E}|F'(g)| < \infty$. We start by splitting the region of integration,

$$\begin{aligned}\mathbb{E}F'(g) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} F'(x)e^{-x^2/2}\,dx + \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} F'(x)e^{-x^2/2}\,dx \\ &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\int_{-\infty}^{x} F'(x)te^{-t^2/2}\,dtdx + \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty}\int_{x}^{\infty} F'(x)te^{-t^2/2}\,dtdx.\end{aligned}$$

To simplify the integration terms, since the condition $\mathbb{E}|F'(g)| < \infty$ implies that

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty}\int_{x}^{\infty} |F'(x)te^{-t^2/2}|\,dtdx &= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty}\int_{x}^{\infty} |F'(x)|te^{-t^2/2}\,dtdx \\ &= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} |F'(x)|e^{-x^2/2}\,dx \le \mathbb{E}|F'(g)| < \infty\end{aligned}$$

and

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\int_{-\infty}^{x} |F'(x)te^{-t^2/2}|\,dtdx &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\int_{-\infty}^{x} |F'(x)|(-te^{-t^2/2})\,dtdx \\ &= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} |F'(x)|e^{-x^2/2}\,dx \le \mathbb{E}|F'(g)| < \infty.\end{aligned}$$

We can use Fubini's theorem to see that

$$\begin{aligned}-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\int_{-\infty}^{x} F'(x)te^{-t^2/2}\,dtdx &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\int_{t}^{0} F'(x)te^{-t^2/2}\,dxdt \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}(F(t)-F(0))e^{-t^2/2}dt\end{aligned}$$

and

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty}\int_{x}^{\infty} F'(x)te^{-t^2/2}\,dtdx &= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty}\int_{0}^{t} F'(x)te^{-t^2/2}\,dxdt \\ &= \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty}(F(t)-F(0))e^{-t^2/2}dt.\end{aligned}$$

These computations imply for $g \sim N(0,1)$,

$$\mathbb{E}F'(g) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(F(t)-F(0))te^{-t^2/2}dt = \mathbb{E}g(F(g)-F(0)) = \mathbb{E}gF(g),$$

since $\mathbb{E}gF(0) = 0$ by symmetry.

($\Longleftarrow$) For fixed $t \in \mathbb{R}$, consider Stein's equation

$$f'(w) - wf(w) = \mathbb{1}(w \leq t) - \mathbb{P}(g \leq t). \tag{2.2}$$

Intuitively, this equation measures how far a distribution is from being Gaussian by testing the integration by parts error. A solution to this ordinary differential equation can be solved using an integrating factor $e^{-\frac{w^2}{2}}$. Multiplying both sides by this integrating factor implies that

$$\frac{d}{dw}\big(e^{-\frac{w^2}{2}} f(w)\big) = e^{-\frac{w^2}{2}}\big(\mathbb{1}(w \leq t) - \mathbb{P}(g \leq t)\big).$$

Integrating both sides implies

$$f_t(w) = e^{\frac{w^2}{2}} \int_{-\infty}^{w} e^{-\frac{x^2}{2}}\big(\mathbb{1}(x \leq t) - \mathbb{P}(g \leq t)\big)\,dx = -e^{\frac{w^2}{2}} \int_{w}^{\infty} e^{-\frac{x^2}{2}}\big(\mathbb{1}(x \leq t) - \mathbb{P}(g \leq t)\big)\,dx,$$

which can be simplified to

$$f_t(w) = \begin{cases} \sqrt{2\pi} e^{\frac{w^2}{2}}\, \mathbb{P}(g \leq w)[1 - \mathbb{P}(g \leq t)] & \text{if } w \leq t \\ \sqrt{2\pi} e^{\frac{w^2}{2}}\, \mathbb{P}(g \leq t)[1 - \mathbb{P}(g \leq w)] & \text{if } w > t. \end{cases}$$

Furthermore, since $1 - \mathbb{P}(g \leq w) = \mathbb{P}(g > w) \sim \frac{1}{w\sqrt{2\pi}} e^{-\frac{w^2}{2}}$, this solution is bounded. It is also immediate that $f_t(w)$ is continuous and piecewise differentiable.

Suppose that the random variable W satisfies (2.1) holds for all bounded, continuous and piecewise continuously differentiable functions f with $\mathbb{E}|f'(g)| < \infty$. Let $f_t(w)$ is a solution to (2.2). Our integration by parts assumption applied to the function $f_t(w)$ implies that

$$0 = \mathbb{E}(f_t'(W) - W f_t(W)) = \mathbb{E}\,\mathbb{1}(W \leq t) - \mathbb{P}(g \leq t) \implies \mathbb{P}(W \leq t) = \mathbb{P}(g \leq t).$$

This holds for all t, so W is a standard Gaussian. □

The first implication of Stein's Lemma can be generalized to higher dimensions. Let $F : \mathbb{R}^n \to \mathbb{R}$.

Lemma 2.1.1

If $g = (g_1, \ldots, g_n)$ is a Gaussian vector in $\mathbb{R}^n$, then

$$\mathbb{E}g_1 F(g) = \sum_{i=1}^{n} \mathbb{E}(g_1 g_i)\mathbb{E}\frac{\partial}{\partial g_i} F(g),$$

provided that $\mathbb{E}|\frac{\partial}{\partial g_i} F(g)| < \infty$ or $\mathbb{E}(g_1 g_i) = 0$ for all $i \leq n$.

Proof. The case when $F : \mathbb{R}^n \to \mathbb{R}$ follows from the one dimensional case. We first state the modification of Stein's Lemma to centered Gaussian random variables with variance σ^2. If $g \sim N(0, \sigma^2)$, then $g \overset{d}{=} \sigma\tilde{g}$ where $\tilde{g} \sim N(0, 1)$. Therefore, (2.1) implies that

$$\mathbb{E}gF(g) = \mathbb{E}\sigma\tilde{g}F(\sigma\tilde{g}) = \sigma\mathbb{E}\frac{d}{d\tilde{g}} F(\sigma\tilde{g}) = \sigma^2 \mathbb{E}F'(\sigma\tilde{g}) = \mathbb{E}g^2 \cdot \mathbb{E}F'(g). \tag{2.3}$$

Suppose that $\vec{g} \sim N(0, \boldsymbol{C})$ where $\boldsymbol{C} = (C_{ij})_{i,j \le n}$ is the covariance matrix. The Gaussian vector $\tilde{g}$ defined by

$$\tilde{g} = (\tilde{g}_\ell)_{1 \le \ell \le n} = \Big(g_\ell - \frac{C_{1\ell}}{C_{11}} g_1 \Big)_{1 \le \ell \le n}$$

is independent from g_1 because $\mathbb{E} g_1 \tilde{g}_\ell = 0$. This implies that

$$g = \Big(\tilde{g} + \frac{C_{1\ell}}{C_{11}} g_1 \Big)_{1 \le \ell \le n}.$$

Since g_1 and $\tilde{g}$ are independent, we can use Fubini's theorem to integrate with respect to g_1 first conditionally on $\tilde{g}$ and apply our result in one-dimensions to conclude that

$$\begin{aligned}
\mathbb{E} g_1 F(g) = \mathbb{E}_{\tilde{g}} \mathbb{E}_{g_1} g_1 F(g) &= \mathbb{E}_{\tilde{g}} \mathbb{E}_{g_1} g_1 F\Big(\tilde{g}_1 + \frac{C_{1\ell}}{C_{11}} g_1, \dots, \tilde{g}_n + \frac{C_{1\ell}}{C_{11}} g_1 \Big) \\
&= \mathbb{E}_{\tilde{g}} C_{11} \mathbb{E}_{g_1} \frac{\partial}{\partial x} F\Big(\tilde{g}_1 + \frac{C_{1\ell}}{C_{11}} x, \dots, \tilde{g}_n + \frac{C_{1\ell}}{C_{11}} x \Big) \Big|_{x = g_1} \\
&= \mathbb{E}_{\tilde{g}} C_{11} \mathbb{E}_{g_1} \frac{C_{1\ell}}{C_{11}} \sum_{i=1}^{n} \frac{\partial}{\partial x_i} F\Big(\tilde{g}_1 + \frac{C_{1\ell}}{C_{11}} g_1, \dots, \tilde{g}_n + \frac{C_{1\ell}}{C_{11}} g_1 \Big) \\
&= \sum_{i=1}^{n} \mathbb{E}_{\tilde{g}} \mathbb{E}_{g_1} C_{1\ell} \frac{\partial}{\partial x_i} F\Big(\tilde{g}_1 + \frac{C_{1\ell}}{C_{11}} g_1, \dots, \tilde{g}_n + \frac{C_{1\ell}}{C_{11}} g_1 \Big) \\
&= \sum_{i=1}^{n} (\mathbb{E} g_1 g_i) \frac{\partial}{\partial x_i} F(g),
\end{aligned}$$

where $\mathbb{E}_{g_1}$ means the average with respect to g_1 and $\mathbb{E}_{\tilde{g}}$ means the average with respect to $\tilde{g}$. This formula holds if $\mathbb{E} |\frac{\partial}{\partial x_\ell} F(g)| < \infty$ or $\mathbb{E}(g_1 g_\ell) = 0$ for each $\ell \le n$. □

Integration by Parts Formulas for Gibbs Averages

In this section, we explain how Stein's lemma is applied to simplify averages with respect to the Gibbs measures. Let $x(\sigma)$ and $y(\sigma)$ be Gaussian processes indexed by a countably infinite set Σ. Furthermore, suppose that the covariances are uniformly bounded

$$\mathbb{E} x(\sigma)^2,\ \mathbb{E} y(\sigma)^2 \le a. \tag{2.4}$$

For a measure G on Σ, define the random change of density

$$G'(\sigma) = \frac{\exp y(\sigma)}{Z} G(\sigma) \qquad \text{and} \qquad Z = \sum_{\sigma \in \Sigma} \exp y(\sigma) G(\sigma).$$

Let $\langle \cdot \rangle$ denote the average with respect to the product measure $G'^{\otimes \infty}$; that is, for $f(\sigma^1, \dots, \sigma^\ell)$,

$$\langle f \rangle = \sum_{\sigma^1, \dots, \sigma^n \in \Sigma} f(\sigma^1, \dots, \sigma^n) G'(\sigma^1) \cdots G'(\sigma^n).$$

We have the following simplification of the Gibbs average.

Theorem 2.1.2 (*Gaussian Integration by Parts*)

If $C(\sigma^1, \sigma^2) = \mathbb{E}x(\sigma^1)y(\sigma^2)$ and $f = f(\sigma^1, \ldots, \sigma^n)$, then

$$\mathbb{E}\langle fx(\sigma^1)\rangle = \mathbb{E}\Big\langle f\Big(\sum_{\ell=1}^{n} C(\sigma^1, \sigma^\ell) - nC(\sigma^1, \sigma^{n+1})\Big)\Big\rangle. \tag{2.5}$$

Proof. We present a proof adapted from [76, Lemma 1.4 and Exercise 1.2.1].

Finite Set Case: When Σ is finite, the proof is a straightforward application of integration by parts. Starting on the left hand side, we have

$$\begin{aligned}\mathbb{E}\langle fx(\sigma^1)\rangle &= \mathbb{E}\sum_{\sigma^1,\ldots,\sigma^n} f(\sigma^1, \ldots, \sigma^n)x(\sigma^1)G'(\sigma^1)\cdots G'(\sigma^n) \\ &= \sum_{\sigma^1,\ldots,\sigma^n} f(\sigma^1, \ldots, \sigma^n)\mathbb{E}x(\sigma^1)\frac{\prod_{l\le n} e^{y(\sigma^l)}G(\sigma^l)}{(\sum_{\sigma^{n+1}} e^{y(\sigma^{n+1})}G(\sigma^{n+1}))^n}.\end{aligned}$$

We can now apply integration parts to the term on the inside of the expectation. Treating the variables $(y(\sigma^{n+1}))_{\sigma^{n+1}\in\Sigma}$ as a separate variable, when the derivative hits the term in the numerator, we have

$$\frac{d}{dy(\sigma^l)}\frac{\prod_{l\le n} e^{y(\sigma^l)}G(\sigma^l)}{(\sum_{\sigma^{n+1}} e^{y(\sigma^{n+1})}G(\sigma^{n+1}))^n} = \frac{\prod_{l\le n} e^{y(\sigma^l)}G(\sigma^l)}{(\sum_{\sigma^{n+1}} e^{y(\sigma^{n+1})}G(\sigma^{n+1}))^n} = G'(\sigma^1)\cdots G'(\sigma^n),$$

and when it hits the denominator, we have

$$\begin{aligned}\mathbb{E}\frac{d}{dy(\sigma^{n+1})}\frac{\prod_{l\le n} e^{y(\sigma^l)}G(\sigma^l)}{(\sum_{\sigma^{n+1}} e^{y(\sigma^{n+1})}G(\sigma^{n+1}))^n} &= -n\frac{e^{y(\sigma^{n+1})}G(\sigma^{n+1})\prod_{l\le n} e^{y(\sigma^l)}G(\sigma^l)}{(\sum_{\sigma^{n+1}} e^{y(\sigma^{n+1})}G(\sigma^{n+1}))^{n+1}} \\ &= -nG'(\sigma^1)\cdots G'(\sigma^{n+1}).\end{aligned}$$

Therefore, using Gaussian integration by parts to compute $\mathbb{E}x(\sigma^1)G'(\sigma^1)\cdots G'(\sigma^n)$

$$\begin{aligned}&\mathbb{E}x(\sigma^1)\frac{\prod_{l\le n} e^{y(\sigma^l)}G(\sigma^l)}{(\sum_{\sigma^{n+1}} e^{y(\sigma^{n+1})}G(\sigma^{n+1}))^n} \\ &\qquad = \sum_{l\le n} C(\sigma^1, \sigma^l)\mathbb{E}G'(\sigma^1)\cdots G'(\sigma^n) - n\sum_{\sigma^{n+1}} C(\sigma^1, \sigma^{n+1})\mathbb{E}G'(\sigma^1)\cdots G'(\sigma^{n+1}).\end{aligned}$$

Now summing over $\sigma^1, \ldots, \sigma^n$ and moving the expectation to the outside, we see

$$\begin{aligned}\mathbb{E}\langle fx(\sigma^1)\rangle &= \sum_{\sigma^1,\ldots,\sigma^n} f(\sigma^1, \ldots, \sigma^n)\sum_{l\le n} C(\sigma^1, \sigma^l)\mathbb{E}G'(\sigma^1)\cdots G'(\sigma^n) \\ &\quad - n\sum_{\sigma^1,\ldots,\sigma^{n+1}} f(\sigma^1, \ldots, \sigma^n)C(\sigma^1, \sigma^{n+1})\mathbb{E}G'(\sigma^1)\cdots G'(\sigma^{n+1}) \\ &= \mathbb{E}\Big\langle f\Big(\sum_{l=1}^{n} C(\sigma^1, \sigma^l) - nC(\sigma^1, \sigma^{n+1})\Big)\Big\rangle.\end{aligned}$$

Infinite Set Case: The infinite set case is the limiting case of the finite set case. Suppose that G is a

measure on a countably infinite set. If G_m is the restriction of G onto the set $A_m \subset \Sigma$ of the m largest weights, our result for the finite set case shows (2.5) holds for all G_m. If the variances are bounded (2.4), elementary estimates on the differences between between the sums over A_m and Σ can be used to show the result holds for $m \to \infty$. The details of such an argument appears in [76, Lemma 1.5]. □

Remark 2.1.1. The result in Lemma 2.1.2 can be easily extended to the case when the change of measure is not with respect to a centered Gaussian. For example,

$$G^{pert}(\sigma) = \frac{\exp(y(\sigma) + h(\sigma))}{Z} G(\sigma) \qquad \text{and} \qquad Z = \sum_{\sigma \in \Sigma} \exp(y(\sigma) + h(\sigma)) G(\sigma)$$

where $h(\sigma)$ is independent of $x(\sigma)$ and $y(\sigma)$. In this case, it also follows that

$$\mathbb{E}\langle x(\sigma)\rangle_{pert} = \mathbb{E}\langle C(\sigma^1, \sigma^1) - C(\sigma^1, \sigma^2)\rangle_{pert}.$$

The proof of this fact is straightforward. If we define $\tilde{G}(\sigma) := e^{h(\sigma)} G(\sigma)$, then we have

$$G^{pert}(\sigma) = \frac{\exp(y(\sigma))}{Z} \tilde{G}(\sigma) \qquad \text{and} \qquad Z = \sum_{\sigma \in \Sigma} \exp(y(\sigma)) \tilde{G}(\sigma).$$

The proof is now identical to the centered Gaussian case.

2.1.2 Gaussian Concentration

We begin by stating the classical Gaussian concentration inequality for Lipschitz functions and apply it to prove a concentration inequality for the free energy functionals appearing in spin glasses.

Consider a Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}$. That is, F satisfies

$$|F(x) - F(y)| \leq L|x - y| \qquad \text{for all} \qquad x, y \in \mathbb{R}^n.$$

The smallest constant L is called the Lipschitz seminorm $\|F\|_{\text{Lip}}$ of F. We state the classical Gaussian concentration inequality for Lipschitz functions of Gaussian vectors.

Theorem 2.1.3 (*Gaussian Concentration*)

If $g = (g_i)_{i \leq n}$ is a standard Gaussian vector on $\mathbb{R}^n$, then for any $t > 0$,

$$\mathbb{P}\left(|F(g) - \mathbb{E}F(g)| \geq t\right) \leq 2\exp\Big(-\frac{t^2}{4\|F\|^2_{\text{Lip}}}\Big). \tag{2.6}$$

Proof. A proof of this result using Gaussian interpolation is provided in [76, Theorem 1.1]. Another proof using the spherical isoperimetric inequality is proved in [13, Chapter 3]. □

Gaussian Concentration of Free Energies

We will use this result to control free energy type functionals. These will be used in the concentration inequalities when we define the regularizing perturbations in Chapter 2.3. Consider the random free

energy

$$\tilde{F}_N = \frac{1}{N} \log \sum_{\sigma \in \Sigma} \exp(g(\sigma)) G(\sigma), \tag{2.7}$$

where $g(\sigma)$ is a Gaussian process with bounded covariance

$$\mathbb{E} g(\sigma)^2 \leq Na \qquad \text{for all} \qquad \sigma \in \Sigma. \tag{2.8}$$

We have the following concentration inequality for the free energy functions.

Theorem 2.1.4 (*Concentration of Free Energy Functions*)

If (2.8) holds, then for all $t \geq 0$,

$$\mathbb{P}(|\tilde{F}_N - \mathbb{E}\tilde{F}_N| \geq t) \leq 2 \exp\Big(-\frac{Nt^2}{4a}\Big) \tag{2.9}$$

which implies that $\mathbb{E}(\tilde{F}_N - \mathbb{E}\tilde{F}_N)^2 \leq \frac{8a}{N}$.

Proof. Using an approximation argument, it suffices to prove this result for finite configuration spaces Σ. If g satisfies (2.8), then $F_N(g)$ is Lipschitz with constant $\sqrt{\frac{a}{N}}$, so the result follows immediately from the Gaussian concentration [Theorem 2.1.3]. The details can be found in [76, Theorem 1.2]. □

As another consequence of the Gaussian concentration inequality, we also show how it implies the concentration of the maximum value of a Gaussian process. When applied to the mixed p-Hamiltonians, the result will imply that the maximum value converges almost surely to its expected value.

Theorem 2.1.5 (*Borel–TIS Inequality*)

Let $X = (X_i)_{i \leq n}$ be a centered Gaussian random variable on $\mathbb{R}^n$ and let

$$M = \max_{i \leq n} \mathbb{E}(X_i^2). \tag{2.10}$$

Then for each $t > 0$,

$$\mathbb{P}\left(\left|\max_{i \leq n} X_i - \mathbb{E}\max_{i \leq n} X_i\right| > t\right) \leq 2e^{-\frac{t^2}{2M}}. \tag{2.11}$$

Proof. Let $X \sim N(0, \boldsymbol{C})$. Since $\boldsymbol{C}$ is positive semi-definite, there exists a matrix $\boldsymbol{A}$ such that $\boldsymbol{C} = \boldsymbol{A}^\mathsf{T}\boldsymbol{A}$. If $g \sim N(0, \boldsymbol{I})$, then it follows that $X \overset{d}{=} \boldsymbol{A}g$. Consider

$$f(x) = \max_{i \leq n} \boldsymbol{A}x.$$

We first check that f is Lipschitz continuous with Lipschitz constant

$$\sqrt{M} = \sqrt{\max_{i \leq n} (\boldsymbol{C})_{ii}} = \sqrt{\max_{i \leq n} \mathbb{E}(X_i^2)}.$$

Let $x, y \in \mathbb{R}^n$. Without loss of generality, suppose that the jth row is the maximum, i.e. $f(\boldsymbol{A}x) = (\boldsymbol{A}x)_j$ and that $f(\boldsymbol{A}x) \geq f(\boldsymbol{A}y)$. If z is the jth row of $\boldsymbol{A}$, we have

$$|f(\boldsymbol{A}x) - f(\boldsymbol{A}y)| = (\boldsymbol{A}x)_j - \max_{i \leq n} (\boldsymbol{A}y)_i \leq |(\boldsymbol{A}x)_j - (\boldsymbol{A}y)_j| = |z \cdot (x - y)|.$$

By the Cauchy–Schwarz inequality, we have

$$|f(\boldsymbol{A}x) - f(\boldsymbol{A}y)| \le \|z\| \cdot \|x - y\|.$$

Notice that

$$\|z\|^2 = \sum_{i \le n} \boldsymbol{A}_{ji}^2 = (\boldsymbol{A}^\mathsf{T}\boldsymbol{A})_{jj} \le \max_{i \le n} \boldsymbol{A}^\mathsf{T}\boldsymbol{A}$$

so

$$|f(\boldsymbol{A}x) - f(\boldsymbol{A}y)| \le \sqrt{\max_{i \le n} \boldsymbol{A}^\mathsf{T}\boldsymbol{A}}\|x - y\| = \sqrt{M}\|x - y\|.$$

Therefore, by Gaussian concentration [Theorem 2.1.3] and the union bound,

$$\begin{aligned}
\mathbb{P}\left(\left|\max_{i \le n} X_i - \mathbb{E}\max_{i \le n} X_i\right| > t\right) &\le \mathbb{P}\left(\max_{i \le n} X_i - \max_{i \le n} \mathbb{E}X_i > t\right) + \mathbb{P}\left(\max_{i \le n} X_i - \mathbb{E}\max_{i \le n} X_i < -t\right) \\
&= \mathbb{P}(f(Z) - \mathbb{E}f(Z) > t) + \mathbb{P}(f(Z) - \mathbb{E}f(Z) < -t) \\
&\le 2e^{-\frac{t^2}{2M}}.
\end{aligned}$$

□

2.2 Asymptotic Gibbs Measure

In this section, we will describe a class of probability measures that will encode the distribution of the overlap array in the limit. One of the key ideas in spin glass is the free energy (1.10) can be lower bounded by functionals of the distribution of the distribution of the overlap arrays $(R_{\ell,\ell'})_{\ell,\ell' \ge 1}$ through an approach called the *cavity computations*.

Consider a probability measure Γ on the unit sphere of a Hilbert space H. These measures represent the possible limiting distributions of the configurations in spin glasses. Let $\langle \cdot \rangle_\Gamma$ denote the average with respect to $\Gamma^{\otimes\infty}$. If $(\rho^\ell)_{\ell \ge 1}$ is an independent sample from Γ and let

$$(Q_{\ell,\ell'})_{\ell,\ell' \ge 1} = (\rho^\ell \cdot \rho^{\ell'})_{\ell,\ell' \ge 1}$$

denote the array of overlaps from this sample. Since Γ is on the unit sphere, this array is bounded and the diagonal elements are constant. Consider $(z(\rho))$ and $(y(\rho))$ denote centered Gaussian processes indexed by points in the support of Γ with covariance

$$\mathbb{E}z(\rho^1)z(\rho^2) = C_z(Q_{1,2}) \qquad \text{and} \qquad \mathbb{E}y(\rho^1)y(\rho^2) = C_y(Q_{1,2})$$

where $C_z, C_y : [-1, 1] \to \mathbb{R}$ are Lipschitz continuous bounded functions. Consider the functional

$$\Phi(\Gamma) = \mathbb{E}\log\left\langle \int_\Sigma e^{\sigma z(\rho)}\, d\mu(\sigma) \right\rangle_\Gamma - \mathbb{E}\log\left\langle e^{y(\rho)} \right\rangle_\Gamma. \tag{2.12}$$

Here $\Sigma \subseteq \mathbb{R}$ and μ is a finite measure on Σ. The next result says that the functional is a continuous function of the distribution of $(Q_{\ell,\ell'})_{\ell,\ell' \ge 1}$.

Theorem 2.2.1

For each $\varepsilon > 0$, there exists a continuous bounded function F_ε of finitely many coordinates of $(Q_{\ell,\ell'})_{\ell,\ell'\geq 1}$ that only depends on ε and the covariance functions C_z and C_y, such that

$$|\Phi(\Gamma) - \langle F_\varepsilon(Q)\rangle_\Gamma| \leq \varepsilon \tag{2.13}$$

uniformly over all possible choices of measures Γ. In particular, $\Phi(\Gamma)$ is a continuous function of the distribution of the distribution of $(Q_{\ell,\ell'})_{\ell,\ell'\geq 1}$ under $\Gamma^{\otimes\infty}$.

Proof. The proof of this fact uses truncation and Gaussian concentration to control the errors. After truncation, the functionals can be approximated by the polynomials and the finite moments of the Gaussian processes up to some error ε, which gives us a way to explicitly construct the functional F_ε. The details of this proof can be found in [76, Theorem 1.4].

We will explain how this result implies that Φ is a continuous function of the distribution of the overlaps. Let $\varepsilon > 0$. By the approximation result, we can find a functional $F_{\varepsilon/3}$ such that

$$|\Phi(\Gamma) - \langle F_{\varepsilon/3}(Q)\rangle_\Gamma| \leq \frac{\varepsilon}{3}.$$

Let d be a metric that metrizes convergence in distribution. Since $F_{\varepsilon/3}(Q)$ is a continuous and bounded function of finitely many coordinates of the overlap array, there exists a $\delta(\varepsilon)$ such that for any Γ_1, Γ_2 such that $d(\Gamma_1, \Gamma_2) \leq \delta$ implies

$$|\langle F_{\varepsilon/3}(Q)\rangle_{\Gamma_1} - \langle F_{\varepsilon/3}(Q)\rangle_{\Gamma_2}| \leq \frac{\varepsilon}{3}$$

by the definition of convergence in distribution (convergence in distribution of infinite arrays means convergence over all finite dimensional distributions). This implies for any $d(\Gamma_1, \Gamma_2) \leq \delta$,

$$|\Phi(\Gamma_1) - \Phi(\Gamma_2)| \leq |\Phi(\Gamma_1) - \langle F_{\varepsilon/3}(Q)\rangle_{\Gamma_1}| + |\Phi(\Gamma_2) - \langle F_{\varepsilon/3}(Q)\rangle_{\Gamma_2}| + |\langle F_{\varepsilon/3}(Q)\rangle_{\Gamma_1} - \langle F_{\varepsilon/3}(Q)\rangle_{\Gamma_2}| \leq \varepsilon.$$

□

To understand the limit of the free energy, it suffices to understand how the functional (2.12) acts on arrays generated from samples from the Gibbs measure in the limit as $N \to \infty$. In this section, we will state several results that will allow us to understand the limiting distribution of the overlap arrays. We call this the distribution of the overlap arrays under the *asymptotic Gibbs measure*.

2.2.1 Dovbysh–Sudakov Representation

We need a way to generate the limiting overlap array in the limit. Let $(\boldsymbol{\sigma}_N)_{\ell\geq 1}$ denote an independent sample from $G_N^{\otimes\infty}$. We want to describe the limiting distribution of $\boldsymbol{\sigma}_N$. This is a really difficult problem because the space $\boldsymbol{\sigma}_N$ takes values changes with N, and it is not even clear if such a limiting distribution will even exist. Instead, we will study the distribution of the *overlap array*

$$R^N = (R^N_{\ell,\ell'})_{\ell,\ell'\geq 1} = \Big(\frac{\boldsymbol{\sigma}^\ell_N \cdot \boldsymbol{\sigma}^{\ell'}_N}{N}\Big)_{\ell,\ell'\geq 1} \tag{2.14}$$

under $G_N^{\otimes\infty}$. The normalization means that the entries of the overlap array are bounded. By the selection theorem [82, Theorem 3.2], the product of overlaps is a compact set, so the space of probability measures on $(R^N_{\ell,\ell'})_{\ell,\ell'\geq 1}$ is tight. This implies that $R^N \to R^\infty$ along some subsequence.

We will use the symmetry obeyed by these overlap arrays to recover the analogue of the limit of $\boldsymbol{\sigma}_N$ in the limit. We will see that these arrays can be generated by samples from an Hilbert space, which will encode the asymptotic spin distributions.

Representation of Exchangeable Arrays

The overlap arrays have some nice symmetry properties that will allow us to characterize its construction. Consider an independent sample $(\boldsymbol{\sigma}^\ell)_{\ell\geq 1}$ from $G_N^{\otimes\infty}$. Since the replicas $(\boldsymbol{\sigma}^\ell)$ are independently sampled, $(\boldsymbol{\sigma}^{\pi(\ell)}) \stackrel{d}{=} (\boldsymbol{\sigma}^\ell)$ for any permutation π of finitely many coordinates. The overlap array

$$R = (R^N_{\ell,\ell'})_{\ell,\ell'\geq 1} = \Big(\frac{\boldsymbol{\sigma}^\ell \cdot \boldsymbol{\sigma}^{\ell'}}{N}\Big)_{\ell,\ell'\geq 1} \tag{2.15}$$

is symmetric, and therefore weakly exchangeable

$$\big(R_{\pi(\ell),\pi(\ell')}\big)_{\ell,\ell'\geq 1} \stackrel{d}{=} \big(R_{\ell,\ell'}\big)_{\ell,\ell'\geq 1}.$$

Since it is a Gram array, it will be positive semi-definite with probability one. Such arrays are called *Gram-de Finetti arrays*. It turns out that all arrays of this form have an explicit construction.

Theorem 2.2.2 (*Aldous–Hoover Representation*)

Let $w, (u_\ell), (x_{\{\ell,\ell'\}})$ be a sequence of independent uniform random variables on $[0,1]$. Any infinite weakly exchangeable array is equal in distribution to the array with entries

$$R_{\ell,\ell} = g(w, u_\ell) \qquad \text{and} \qquad R_{\ell,\ell'} = f(w, u_\ell, u_{\ell'}, x_{\{\ell,\ell'\}})$$

for some functions $g : [0,1]^2 \to \mathbb{R}$ and $f : [0,1]^4 \to \mathbb{R}$ where f is symmetric in the middle coordinates.

Proof. This is an extension of de Finetti's theorem for exchangeable sequences. The details can be found in [76, Chapter 1.4]. □

Theorem 2.2.2 implies that all Gram-de Finetti arrays can be generated similarly to the overlap arrays. The only difference is a correction for the diagonal terms. Let $H = L^2([0,1], dv)$ where dv denotes the Lebesgue measure on $[0,1]$. We have the following method to generate Gram-de Finetti arrays.

Theorem 2.2.3 (*Dovbysh–Sudakov Representation*)

There exists a random probability measure η on $H \times \mathbb{R}^+$ such that the array $R = (R_{\ell,\ell'})_{\ell,\ell'\geq 1}$ is equal in distribution to

$$(h_\ell \cdot h_{\ell'} + a_\ell \delta_{\ell,\ell'})_{\ell,\ell'}$$

where, conditionally on η, $(h_\ell, a_\ell)_{\ell\geq 1}$ is a sequence of i.i.d. random variables with distribution η and $h_1 \cdot h_2$ denotes a scalar product on H.

Proof. This is an extension of Theorem 2.2.2. The details can be found in [76, Chatper 1.5]. □

2.2.2 Ruelle Probability Cascades

In this section, we define a class of random measures on a Hilbert space that will generate special Gram-de Finetti arrays that will characterize the limiting distribution of the overlaps under the asymptotic Gibbs measure. The points in the supports of these random measures are indexed by infinite trees. The details of the construction can be found in [76, Chapter 2.3]. We will summarize the main points here for reference.

Let $r \geq 1$ denote the levels of an infinite tree. Every vertex of this infinite tree is denoted by a sequence $(n_1, \ldots, n_p) \in \mathbb{N}^p$ where $p \leq r-1$ encoding the parents of the tree. These vertices are connected by a path to the root of the tree

$$\emptyset \to n_1 \to (n_1, n_2) \to \cdots \to (n_1, \ldots, n_p).$$

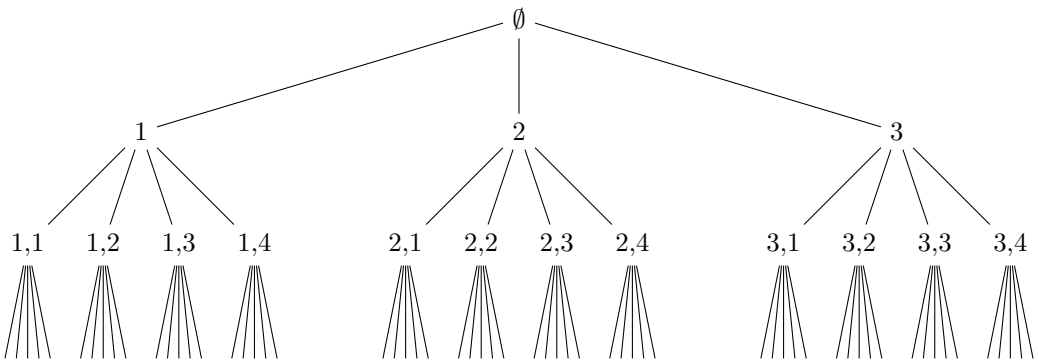

Figure 1. The labeling of the first two levels of the tree is displayed above. The vertices of the first level are indexed by natural numbers $\alpha_1 \in \mathbb{N}$. The vertices in the second level are indexed by pairs $(\alpha_1, \alpha_2) \in \mathbb{N}^2$. The first number denotes the parent vertex, and the second number denotes the child vertex.

The parameter of the Ruelle probability cascades are associated with a sequence of parameters

$$\begin{array}{l} 0 < x_0 < x_1 < \ldots < x_{r-2} < x_{r-1} < 1 \\ 0 = q_0 < q_1 < \ldots < q_{r-2} < q_{r-1} < \infty \end{array}.$$

These parameters can be encoded by a probability measure ζ,

$$\zeta(\{q_p\}) = x_p - x_{p-1} \qquad \text{for} \qquad 1 \leq p \leq r. \tag{2.16}$$

These parameters will define the weights of a random measure. We associate with every vertex of the tree with an independent Poisson process on $(0, \infty)$ with mean measure

$$\mu(dy) = x_k y^{-1-x_k}\, dy \qquad \text{where } k \text{ is the depth of the vertex.}$$

The Poisson process assigns more points near 0, so we arrange these points in decreasing order. In particular, every $\alpha = (\alpha_1, \ldots, \alpha_r) \in \mathbb{N}^r$ is associated with a sequence of weights from these Poisson

processes

$$(u_{\alpha_1}, u_{\alpha_1,\alpha_2}, \ldots, u_{\alpha_1,\alpha_2,\ldots,\alpha_r}).$$

Here u_{α_1} is the α_1th largest term from the points generated from the Poisson process with mean measure $x_1 y^{-1-x_1}$, u_{α_1,α_2} is the α_2th largest term generated from the Poisson process with mean measure $x_2 y^{-1-x_2}$ associated with the children of vertex α_1, etc. We define

$$v_\alpha = \frac{w_\alpha}{\sum_{\beta \in \mathbb{N}^r} w_\beta} \qquad w_\alpha = u_{\alpha_1} u_{\alpha_1,\alpha_2} \cdots u_{\alpha_1,\ldots,\alpha_r} \tag{2.17}$$

to be the random probability weight associated with $\alpha \in \mathbb{N}^r$. The sequence $(v_\alpha)_{\alpha \in \mathbb{N}^r}$ correspond to the weights of a random measure on a Hilbert space H supported on an ultrametric set indexed by $\alpha \in \mathbb{N}^r$.

We now explain how one would construct such a random measure with these weights. For each vertex $(\alpha_1, \ldots, \alpha_p)$ in the infinite tree, we assign it to a distinct orthonormal basis element $e_{\alpha_1,\ldots,\alpha_p}$ in H. For $\alpha \in \mathbb{N}^r$, we define

$$h_\alpha = e_{\alpha_1,\ldots,\alpha_r}\sqrt{q_r - q_{r-1}} + e_{\alpha_1,\ldots,\alpha_{r-1}}\sqrt{q_{r-1} - q_{r-2}} + \cdots + e_{\alpha_1}\sqrt{q_1 - q_0}.$$

We denote the number of common edges between leaves α and β by

$$\alpha \wedge \beta = \min\big\{0 \le p \le r \mid \alpha_1 = \beta_1, \ldots, \alpha_p = \beta_p, \alpha_{p+1} \neq \beta_{p+1}\big\}.$$

where $\alpha \wedge \beta = r$ if $\alpha = \beta$. From the definition of h_α, it follows that

$$h_\alpha \cdot h_\beta = q_{\alpha \wedge \beta}.$$

We say $G \sim \mathrm{RPC}(\zeta)$ if

$$G(h_\alpha) = v_\alpha \qquad \text{for} \qquad \alpha \in \mathbb{N}^r.$$

We call this measure G the *Reulle probability cascade* associated with the parameter ζ defined in (2.16). The Ruelle probability cascades satisfy several important properties that will allow us to understand the limiting distribution of the overlap array.

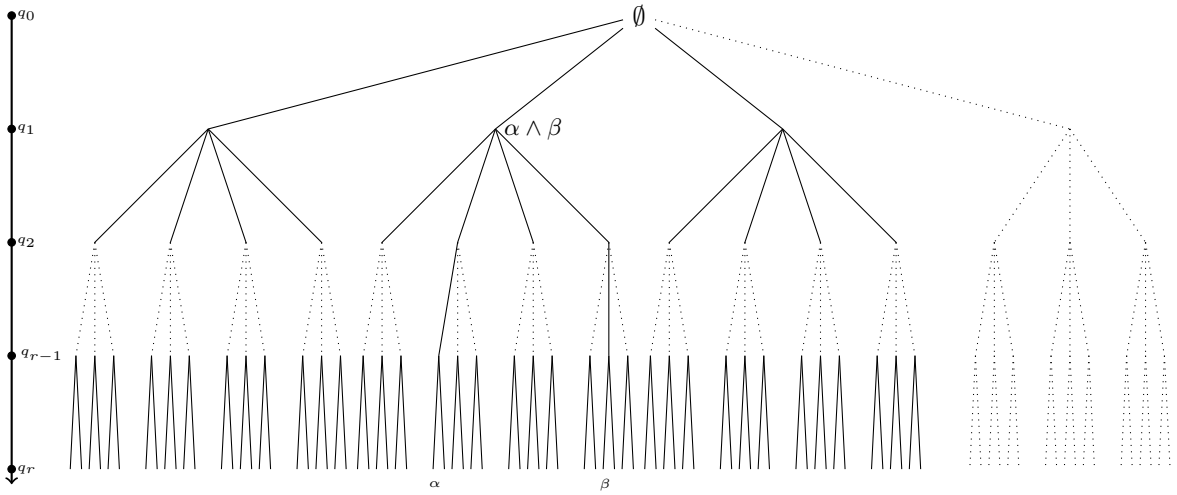

Figure 2. The leaves α and β in this infinite tree index the q_r clusters in the model. Such a tree is represents what is called *r-step replica symmetry breaking.*

Properties of the RPC

The first and most important property is the almost sure ultrametric support of the G. Consider a sample $(h_{\alpha^\ell})_{\ell\geq 1}$ from RPC(ζ). We denote the associated random Gram de-Finetti array by

$$(R_{\ell,\ell'})_{\ell,\ell'\geq 1} = (h_{\alpha^\ell}\cdot h_{\alpha^\ell})_{\ell,\ell'\geq 1} = (q_{\alpha^\ell\wedge\alpha^{\ell'}})_{\ell,\ell'\geq 1}. \tag{2.18}$$

By construction, this array is automatically ultrametric.

Theorem 2.2.4 (*Ultrametric Support*)

The array $(R_{\ell,\ell'})_{\ell,\ell'\geq 1}$ in (2.18) is almost surely ultrametric,

$$R_{1,2}\geq \min(R_{1,3},R_{2,3}). \tag{2.19}$$

Proof. This property follows from the fact that $\alpha^1\wedge\alpha^2\geq\min(\alpha^1\wedge\alpha^3,\alpha^2\wedge\alpha^3)$ and (q_p) is increasing. □

One of the convenient properties is the ability to compute averages with respect to the Ruelle probabiliy cascades recursively using the tree structure. Let $G\sim$ RPC(ζ), and a function $f(h_{\alpha^1},\dots,h_{\alpha^n})\to\mathbb{R}$, we let

$$\langle f\rangle = \sum_{\alpha^1,\dots,\alpha^n\in\mathbb{N}^r} f(h_{\alpha^1},\dots,h_{\alpha^n})G(h_\alpha^1)\cdots G(h_\alpha^n) = \sum_{\alpha^1,\dots,\alpha^n\in\mathbb{N}^r} f(h_{\alpha^1},\dots,h_{\alpha^n})v_{\alpha^1}\cdots v_{\alpha^n}$$

denote the average with respect to G. Let $(\omega_p)_{1\leq p\leq r}$ be a sequence of i.i.d. uniform random variables on $[0,1]$. We let

$$X_r(\omega_1,\dots,\omega_r) \tag{2.20}$$

denote a random variable X_r determined by r sources of randomness. Define recursively for $0\leq p\leq r$,

$$X_p = X_p(\omega_1,\dots,\omega_p) = \frac{1}{x_p}\log\mathbb{E}_{\omega_{p+1}}\exp x_pX_{p+1}. \tag{2.21}$$

In particular, X_0 is non-random. Let $\omega_i(\alpha) = (\omega^i_{\alpha_1},\dots,\omega^i_{\alpha_1,\dots,\alpha_r})$ for $i\leq N$ denote a sequence of independent uniform random variables on $[0,1]$ indexed by $\alpha\in\mathbb{N}^r$. We let

$$X_{r,i}(\alpha) = X(\omega_i(\alpha)) = X(\omega^i_{\alpha_1},\dots,\omega^i_{\alpha_1,\dots,\alpha_r}).$$

denote independent copies of X_r indexed by $\alpha\in\mathbb{N}^r$. We have the following recursive formula for the expected value of certain free energy type functionals defined via the Ruelle probability cascades.

Theorem 2.2.5 (*Recursive Expected Value*)

If X_0 is defined via the recursive construction (2.21), then

$$X_0 = \frac{1}{N}\mathbb{E}\log\left\langle \exp\left(\sum_{i=1}^{N} X_{r,i}(\alpha)\right)\right\rangle = \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r}\exp\left(\sum_{i=1}^{N} X_{r,i}(\alpha)\right)v_\alpha. \tag{2.22}$$

Proof. The proof follows from the homogeneity property satisfied by the mean measure of the Poisson Dirichlet processes [76, Theorem 2.6], and the fact the children of the vertices are independently generated. The details can be found in [76, Theorem 2.9]. □

Remark 2.2.1. This property applies to the sums of Gaussian random variables. For example, recall the random variable X_r introduced in (1.21). If we let F^{-1} be the quantile function of a standard Gaussian, then (1.21) is equal in distribution to

$$X_r(\omega_1,\ldots,\omega_r) = \log\sum_{\sigma\in\{-1,+1\}}\exp\left(\sigma\sum_{1\le j\le r}\sqrt{\xi'(q_j)-\xi'(q_{j-1})}F^{-1}(\omega_j)\right).$$

Averaging recursively with respect to ω_j is equivalent to averaging with respect to the Gaussian processes by the Coding Lemma [82, Lemma 5.1]. The encoding Gaussian random variables with uniform random variables will be used multiple times throughout this thesis whenever Theorem 2.2.5 is used.

The Bolthausen–Sznitman invariance property [76, Equation 2.26] states the expected value of functionals of the arrays (2.18) are invariant under certain changes of measures. The first consequence states that the paramter of the Ruelle probability cascades describes the distribution of the first off diagonal entry of the array.

Theorem 2.2.6 (*Functional Order Parameter*)

If (2.25) holds, then

$$\mathbb{E}\langle \mathbb{1}(R_{1,2}=q_p)\rangle = x_p - x_{p-1} \qquad \text{for} \qquad 0\le p\le r.$$

Proof. By the Bolthausen–Sznitman invariance principle for the Ruelle probability cascades [76, Lemma 2.5 and Equation (2.79)],

$$\mathbb{E}\langle\psi(R_{1,2})\rangle = \sum_{p=1}^{r}\psi(q_p)(x_p-x_{p-1}).$$

Taking $n=1$, $f=1$, and $\varphi(x)=\mathbb{1}(x=q_p)$ in (2.23) finishes the proof. □

The most important consequence of this invariance is the an identity called the Ghirlanda–Guerra identities. We will see later that this identity characterizes the distribution of arrays generated by Ruelle probability cascades.

Theorem 2.2.7 (*Ghirlanda–Guerra Identities*)

For any $n \geq 1$, any function f of the overlaps of n replica $R^n = (R_{\ell,\ell'})_{\ell,\ell' \leq n}$ and any function $\psi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\langle f(R^n)\psi(R_{1,n+1})\rangle = \mathbb{E}\langle f(R^n)\rangle\mathbb{E}\langle\psi(R_{1,n+1})\rangle + \frac{1}{n}\sum_{\ell=2}^{n}\mathbb{E}\langle f(R^n)\psi(R_{1,\ell})\rangle. \tag{2.23}$$

Proof. The proof follows from the Bolthausen-Sznitman invariance property and integration by parts. The details can be found in [76, Lemma 2.5 and Theorem 2.10]. □

Characterization of the RPCs

The Ghirlanda–Guerra identities characterizes arrays generated from samples from the Ruelle probability cascades. We will show that any measure that generates an array satisfying the Ghirlanda–Guerra identities must be a Ruelle probability cascade.

Let G be a random measure on a Hilbert space H with inner product $\langle\cdot,\cdot\rangle$, and let $\langle\cdot\rangle$ denote the average with respect to $G^{\otimes\infty}$. Let $(\sigma^\ell)_{\ell\geq 1}$ be an i.i.d. sample from G, and define the corresponding overlap array

$$(R_{\ell,\ell'})_{\ell,\ell'\geq 1} = (\langle\sigma^\ell, \sigma^{\ell'}\rangle)_{\ell,\ell'\geq 1}. \tag{2.24}$$

Suppose that the also satisfies an invariance property called the Ghirlanda–Guerra identities:

Definition 1. We say that an array satisfies the *Ghirlanda–Guerra identities* if for any $n \geq 1$, any function f of the overlaps of n replica $R^n = (R_{\ell,\ell'})_{\ell,\ell'\leq n}$, and any function $\psi : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}\langle f(R^n)\psi(R_{1,n+1})\rangle = \mathbb{E}\langle f(R^n)\rangle\mathbb{E}\langle\psi(R_{1,n+1})\rangle + \frac{1}{n}\sum_{\ell=2}^{n}\mathbb{E}\langle f(R^n)\psi(R_{1,\ell})\rangle. \tag{2.25}$$

We will show that if the array satisfies the Ghirlanda–Guerra identities (2.25), then the array must be generated from a Ruelle probability cascades. This characterization will follow from several consequences of arrays satisfying these identities.

The first few consequences states that the array resembles one generated by the Ruelle probability cascades. First, the array has fixed diagonal elements.

Theorem 2.2.8 (*Fixed Self Overlaps*)

Suppose that the Ghirlanda–Guerra identities (2.25) hold. If q^* is the largest point in the support of ζ then,

$$\mathbb{E}\langle\mathbb{1}(R_{1,1} = q^*)\rangle = 1.$$

Proof. See [76, Theorem 2.15] for the details. □

The arrays also must take non-negative values. This means that the infinite array is also positive in the sense that all its entries are non-negative.

Theorem 2.2.9 (*Talagrand's Positivity Principle*)

Suppose that the Ghirlanda–Guerra identities (2.25) hold. Then the overlap is non-negative

$$\zeta([0,\infty)) = 1.$$

Proof. See [76, Theorem 2.16] for the details. □

Panchenko proved that the Ghirlanda–Guerra identities imply that the overlap matrix array must be ultrametric [75]. This remarkable result implies that G must be supported on an ultrametric subset, and is the key property satisfied by the limiting distribution of the overlap arrays in the spin glass models.

This result follows from another consequence called the *duplication property*, which describes the infinite tree structure of the support. It essentially says that it is possible to sample a configuration from the same parent of the configuration with the largest overlap. Given $\varepsilon > 0$, let $a \approx b$ mean that $a - \varepsilon < b < a + \varepsilon$, and $\boldsymbol{A} \approx \boldsymbol{B}$ mean that all entires $A_{i,j} - \varepsilon < B_{i,j} < A_{i,j} + \varepsilon$. We have the following result.

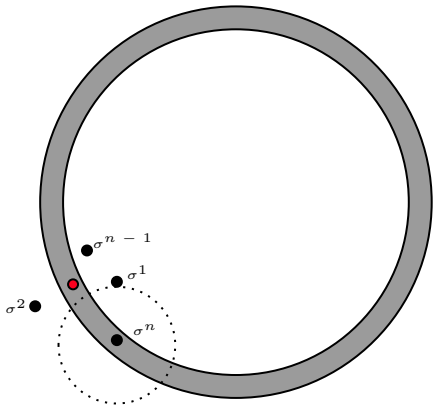

Figure 3. The duplication property states that we can find an $\boldsymbol{\sigma}^{n+1}$ that is the same distance from $\boldsymbol{\sigma}^1, \ldots, \boldsymbol{\sigma}^{n-1}$ as $\boldsymbol{\sigma}^n$, and is at least as far as the closest of the first $n-1$ replica from $\boldsymbol{\sigma}^n$. In this picture the gray area are the points on the sphere such that are the same distances away from the first $n-1$ replica as $\boldsymbol{\sigma}^n$. The red point is the duplicate, because it is the the gray area, but further away than the $\boldsymbol{\sigma}^1$, which is the closest point out of the first $n-1$ replica to $\boldsymbol{\sigma}^n$.

The duplication property interpreted as a fact about infinite trees is shown below.

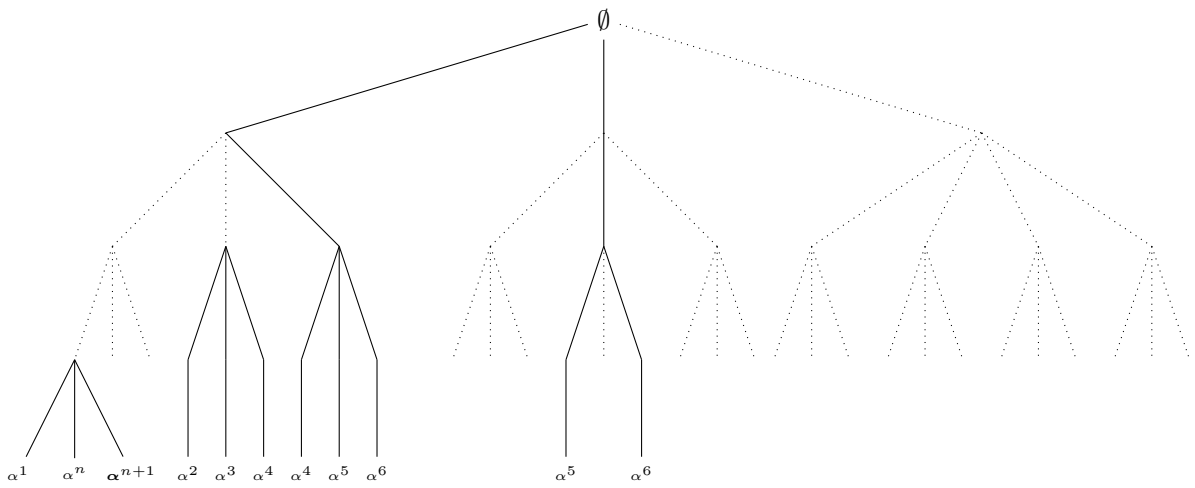

Figure 4. The infinite tree structure implies we can always find another index $\alpha^{n+1} \in \mathbb{N}^r$ such that $\alpha^n \wedge \alpha^{n+1} = \alpha^1 \wedge \alpha^n$ (the largest overlap value in the picture). Furthermore, this duplicate α^{n+1} is the same distance from as α^n from all the other leaves, i.e. $\alpha^\ell \wedge \alpha^{n+1} = \alpha^\ell \wedge \alpha^n$.

Theorem 2.2.10 (*Duplication Property*)

Suppose that the Ghirlanda–Guerra identities (2.25) hold. Given $\varepsilon > 0$, if the matrix $\boldsymbol{A}$ satisfies

$$\mathbb{E}\langle \mathbb{1}(\boldsymbol{R}^n \approx \boldsymbol{A})\rangle > 0,$$

and $A_n^* + \varepsilon := \max(A_{1,n}, \ldots, A_{n-1,n}) + \varepsilon < q^*$, then

$$\mathbb{E}\langle \mathbb{1}(\boldsymbol{R}^n \approx \boldsymbol{A}, R_{\ell,n+1} \approx A_{\ell,n} \text{ for } \ell \leq n-1, R_{n,n+1} < A_n^* + \varepsilon)\rangle > 0$$

Proof. See [76, Theorem 2.20] for the details. □

The duplication property is the key fact that allows us to prove the ultrametricity conjecture.

Theorem 2.2.11 (*The Parisi Ultrametricity Conjecture*)

Suppose that the Ghirlanda–Guerra identities (2.25) hold. Then, the overlap matrix $(R_{\ell,\ell'})_{\ell,\ell' \geq 1}$ is ultrametric,

$$R_{1,2} \geq \min(R_{1,3}, R_{2,3}). \tag{2.26}$$

Proof. The proof of ultrametricity shows that it is impossible for a non-ultrametric block matrix to appear in the overlap array. By the duplication property, we can generate duplicates so that the non-ultrametric block matrices can be arbitrary large. Examining the barycenters of configurations from the non-ultrametric block matrices will result in a contradiction because the duplication property will force one barycenter to be arbitrarily close to two distinct vectors simultaneously. We can conclude that the duplication property can only hold if the arrays were ultrametric. See [76, Chapter 2.5] for the details. □

The next consequence states that the parameter to generate the array is determined by the distribution of the first off diagonal element of the array. We care about the distributions of the off diagonal arrays, because it turns out that the diagonal entries of the array are almost surely constant. Let ζ denote the distribution of the first off diagonal entry of the array under $\mathbb{E}G^{\otimes 2}$

$$\zeta(A) = \mathbb{E}\langle \mathbb{1}(R_{1,2} \in A)\rangle. \tag{2.27}$$

The measure ζ is the corresponding parameter of the random measure.

Theorem 2.2.12 (*Uniqueness*)

Suppose that the Ghirlanda–Guerra identities (2.25) hold. Then the distribution of the entire overlap matrix $(R_{\ell,\ell'})_{\ell,\ell'\geq 1}$ under $\mathbb{E}G^{\otimes\infty}$ is uniquely determined by ζ in (2.27).

Proof. If the array is ultrametric, then the distribution of the entire array is encoded by the first off diagonal. See [76, Theorem 2.13] for the details. □

These properties seem to suggest that if such a random G were to exist, it must resemble the Ruelle probability cascades with parameter ζ encoded by the nonnegative sequence (2.2.2). The next result states that there exists such a measure satisfying the Ghirlanda–Guerra identities for every probability ζ on $\mathbb{R}^+$ with bounded support.

Theorem 2.2.13 (*Existence and Continuity*)

For every probability distribution ζ on $\mathbb{R}^+$ with bounded support there exists a random measure G on a separable Hilbert space H, which satisfies the Ghirlanda–Guerra identities (2.25) and such that (2.27) holds. Moreover, the distribution of the overlap array $(R_{\ell,\ell'})_{\ell,\ell'\geq 1}$ defined in (2.24) is continuous with respect to the weak convergence of ζ.

Proof. See [76, Theorem 2.17] for the details. □

This fact combined with the properties of the Ruelle probability cascades and the Dovbysh–Sudakov representation completes the characterization of the limiting distribution of arrays satisfying the Ghirlanda–Guerra identities (2.25). We summarize the results. Any array that satisfies (2.25) is has non-negative entires, ultrametric, and has a fixed diagonal [Theorem 2.2.8, Theorem 2.2.9, Theorem 2.2.11]. This array is also uniquely determined by the functional order parameter ζ corresponding to the distribution of the first off diagonal element [Theorem 2.2.12]. Given ζ, we can find a random measure G on a separable Hilbert space H [Theorem 2.2.13] that satisfies the Ghirlanda–Guerra identities. By the Dovbysh–Sudakov representation [Theorem 2.2.3], we can use this random measure to generate an array from the Ruelle probability cascades that is equal in distribution to the random array $(R_{\ell,\ell'})_{\ell,\ell'\geq 1}$.

2.3 Regularizing Perturbations of Gibbs Measures

In the previous section, we showed that the Ghirlanda–Guerra identities (2.25) characterizes the Ruelle probability cascades. In this section, we will show that the asympotic Gibbs measure satisfies the Ghirlanda–Guerra identities. This will allow us to approximate the limiting arrays with samples from the Ruelle probability cascades.

Unfortunately, the array generated by samples from the Gibbs measure might not satisfy the Ghirlanda–Guerra identities apriori. We will introduce a random perturbation of the Hamiltonian that will force the array to satisfy the Ghirlanda–Guerra identities in the limit (at least on average). This will allow us to apply the probabilistic method to find a deterministic perturbation of the Gibbs measure that will imply the array will satisfy the Ghirlanda–Guerra identities (2.25). The perturbation will be small enough so that the limit of the free energy will be unaffected.

2.3.1 Ghirlanda–Guerra Identities

In this section, we first explain the connection between the Ghirlanda–Guerra identities and the concentration (on average) of the free energy. The explanations here will be made precise in the next section. We first explain how the Ghirlanda–Guerra shows up naturally when one tests the concentration of the free energy functions. Roughly speaking, the Ghirlanda–Guerra identities are an approximate independence property for the pure p-spin Hamiltonian

$$\mathbb{E}\Big\langle f(R^n)\frac{H_N^p(\boldsymbol{\sigma}^1)}{N}\Big\rangle - \mathbb{E}\big\langle f(R^n)\big\rangle\mathbb{E}\Big\langle\frac{H_N^p(\boldsymbol{\sigma}^1)}{N}\Big\rangle \approx 0. \tag{2.28}$$

for bounded functions f of n replica. If we use Gaussian integration by parts [Theorem 2.1], then the left hand side can be simplified to

$$\mathbb{E}\Big\langle f(\sum_{\ell=1}^{n} R_{1,\ell}^p - nR_{1,n+1}^p)\Big\rangle - \mathbb{E}\langle f\rangle\mathbb{E}\langle 1 - R_{1,2}^p\rangle = \sum_{\ell=2}^{n}\mathbb{E}\langle fR_{1,\ell}^p\rangle - n\langle fR_{1,n+1}^p\rangle + \mathbb{E}\langle f\rangle\mathbb{E}\langle R_{1,2}^p\rangle \tag{2.29}$$

since $R_{1,1} = 1$ because the self overlaps are fixed. If (2.28), then then right hand side of (2.29) will equal to 0. For example, if we take $n = 3$ and $f = R_{1,2}^2$, the Ghirlanda–Guerra identities implies the 2-spin Hamiltonian satisfies

$$\mathbb{E}\langle R_{1,2}^2R_{1,3}^2\rangle - \frac{1}{2}\mathbb{E}\langle R_{1,2}^2\rangle\mathbb{E}\langle R_{1,3}^2\rangle - \frac{1}{2}\mathbb{E}\langle R_{1,2}^4\rangle \approx 0.$$

Taking different p and n implies these identities encodes information about the pth moment of the overlaps. The Ghirlanda–Guerra identities roughly describe the joint moments of the overlap array.

Concentration of the free energy functions actually gives us a way to see this approximate independence. One can show that on average over temperature,

$$\int_0^1 \mathbb{E}\Big\langle\Big|\frac{H_N^p(\boldsymbol{\sigma})}{N} - \mathbb{E}\Big\langle\frac{H_N^p(\boldsymbol{\sigma})}{N}\Big\rangle\Big|\Big\rangle\, d\beta = O(N^{1/4}) \tag{2.30}$$

where $\langle\cdot\rangle$ is the average with respect to the Gibbs measure $G_N(\beta) \propto e^{\beta H_N^p(\boldsymbol{\sigma})}$. If f is bounded, then on average (2.28) is bounded above by (2.30), so we have the approximate independence property, at least on average over for high temperature. It turns out that we can make hold without averaging over temperature, by adding a perturbation to the Gibbs measure. The temperature parameters in the perturbation will be replaced with independent uniform random variables, allowing us similar control as the average over temperature.

These identities describe a way to generate the off-diagonal entries of the infinite array. If ζ is the distribution of one overlap, then we can generated $R_{1,n+1}$ conditionally on $(R_{\ell,\ell'})_{\ell,\ell'\le n}$. The identities

imply that the distribution of $R_{1,n+1}$ is given by

$$\frac{1}{n}\zeta + \frac{1}{n}\sum_{\ell=2}^{n} \delta_{R_{1,\ell}}, \tag{2.31}$$

which means with probability $\frac{1}{n}$ the overlap of $\boldsymbol{\sigma}^{n+1}$ and $\boldsymbol{\sigma}^1$ will take an independent value according to ζ, or with probability $\frac{n-1}{n}$, it will take one of the existing values from $R_{1,2}, \ldots, R_{1,n}$ uniformly. This gives us a way to generate the marginal distributions of the overlap array, but it is insufficient to generate the joint distribution, so the characterization of the array is not a trivial result. As we have seen in the previous section showing ultrametricity of the array implies Theorem 2.2.13 and Theorem 2.2.12, these identities are enough to characterize the distribution of the arrays.

2.3.2 Perturbed Hamiltonian

We begin by introducing the classical Ghirlanda–Guerra identities. We will introduce some generalized forms of these identities in the next chapter, when we discuss the synchronization of vector spin models. We will introduce a Gaussian perturbation that will force the overlap array to satisfy the Ghirlanda–Guerra identities in the limit. This general procedure is explained in [76, Chapter 3.2], but we will recreate parts of the proofs here, because it will be used later.

Recall that the overlap is the normalized inner product defined by

$$R_{\ell,\ell'} = \frac{\boldsymbol{\sigma}^\ell \cdot \boldsymbol{\sigma}^{\ell'}}{N}.$$

In all spin glass models, the self overlaps are fixed and normalized, so the diagonal terms $R_{\ell,\ell} = 1$ for some fixed constant. We begin by introducing the Gaussian perturbation.

Lemma 2.3.1

There exists a centered Gaussian process $h_N(\boldsymbol{\sigma})$ with covariance

$$\mathbb{E}h_N(\boldsymbol{\sigma}^1)h_N(\boldsymbol{\sigma}^2) = \sum_{p\geq 1} 4^{-p}u_p^2 R_{1,2}^p \tag{2.32}$$

where $u_p \in [0,3]$.

Proof. We denote the right hand side of (2.32) by

$$(C_{\ell,\ell'})_{\ell,\ell'\geq 1} = \left(\sum_{p\geq 1} 4^{-p}u_p^2 R_{\ell,\ell'}^p\right)_{\ell,\ell'\geq 1}.$$

It suffices to show that the covariance matrix $(C_{\ell,\ell'})_{\ell,\ell'\geq 1} = (\mathbb{E}h_N(\boldsymbol{\sigma}^\ell)h_N(\boldsymbol{\sigma}^{\ell'}))_{\ell,\ell'\geq 1}$ is positive semidefinite; that is, $(C_{\ell,\ell'})_{\ell,\ell'\geq n}$ is positive semidefinite for all n. The array

$$R = (R_{\ell,\ell'})_{\ell,\ell'\geq 1} = \left(\frac{\boldsymbol{\sigma}^\ell \cdot \boldsymbol{\sigma}^{\ell'}}{N}\right)_{\ell,\ell'\geq 1}$$

is a Gram array generated by $(\boldsymbol{\sigma}^\ell)_{\ell\geq 1}$, so it is positive semidefinite. For any $p \geq 1$, this implies that

$$(C^p_{\ell,\ell'})_{\ell,\ell'\geq 1} := (R^p_{\ell,\ell'})_{\ell,\ell'\geq 1} = R^{\odot p} \tag{2.33}$$

is the Hadamard power of positive definite matrices, so it is also positive definite by the Schur product theorem [56, Theorem 7.5.3]. Lastly, since

$$(C_{\ell,\ell'})_{\ell,\ell'\geq 1} = \left(\sum_{p\geq 1} 4^{-p} u_p^2 C^p_{\ell,\ell'}\right)_{\ell,\ell'\geq 1}$$

is a linear combination of positive definite arrays, so $(C_{\ell,\ell'})_{\ell,\ell'\geq 1}$ is positive semidefinite. □

Remark 2.3.1. We can also explicitly construct a Gaussian process with the appropriate covariance. These Gaussian processes show up when we explicitly compute the covariance of p-spin Hamiltonians.

Let $H_N(\boldsymbol{\sigma})$ be an arbitrary Hamiltonian (either random or non-random) and let $h_N(\boldsymbol{\sigma})$ be an independent centered Gaussian with covariance given by (2.32). We define the *perturbed Hamilton* by

$$H_N^{\text{pert}}(\boldsymbol{\sigma}) = H_N(\boldsymbol{\sigma}) + s_N h_N(\boldsymbol{\sigma}). \tag{2.34}$$

For a specific choice of s_N, this perturbation will regularize the Gibbs measure, forcing it to satisfy some additional nice properties, without changing the limit of the free energy.

Proposition 2.3.1

If $\frac{s_N^2}{N} \to 0$, *then*

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\log \int_{\Sigma_N} e^{H_N(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma}) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\log \int_{\Sigma_N} e^{H_N^{\text{pert}}(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma})$$

where μ_N *is a probability measure on* Σ_N.

Proof. We first fix $N \geq 1$. Jensen's inequality implies that conditionally on $H_N(\boldsymbol{\sigma})$,

$$\begin{aligned}\frac{1}{N}\mathbb{E}\log \int_{\Sigma_N} e^{H_N^{\text{pert}}(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma}) &\leq \frac{1}{N}\mathbb{E}_{H_N}\log \int_{\Sigma_N} \mathbb{E}_{h_N} e^{H_N^{\text{pert}}(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma}) \\ &= \frac{1}{N}\mathbb{E}\log \int_{\Sigma_N} e^{H_N(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma}) + \frac{s_N^2}{2N}\operatorname{Var}(h_N),\end{aligned}$$

since the self overlaps are fixed, $\operatorname{Var}(h_N)$ is bounded and constant for all $\boldsymbol{\sigma} \in \Sigma_N$, so the variance term vanishes if $\frac{s_N^2}{N} \to 0$. We now apply Jensen's inequality to the integral conditionally on H_N. Let $Z_N = \int_{\Sigma_N} e^{H_N^{\text{pert}}(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma})$, we have

$$\begin{aligned}\frac{1}{N}\mathbb{E}\log \int_{\Sigma_N} e^{H_N^{\text{pert}}(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma}) &\geq \frac{1}{N}\mathbb{E}_{H_N}\mathbb{E}_{h_N}\log \int_{\Sigma_N} e^{s_N h_N(\boldsymbol{\sigma})} Z_N \frac{e^{H_N(\boldsymbol{\sigma})}}{Z_N}\, d\mu_N(\boldsymbol{\sigma}) \\ &= \frac{1}{N}\mathbb{E}_{H_N}\mathbb{E}_{h_N} \int_{\Sigma_N} (s_N h_N(\boldsymbol{\sigma}) + \log(Z_N)) \frac{e^{H_N(\boldsymbol{\sigma})}}{Z_N}\, d\mu_N(\boldsymbol{\sigma}) \\ &= \frac{1}{N}\mathbb{E}_{H_N}\log Z_N = \frac{1}{N}\mathbb{E}\log \int_{\Sigma_N} e^{H_N(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma})\end{aligned}$$

since h_N is a centered Gaussian process. Taking $N \to \infty$ and combining the upper and lower bounds finishes the proof. □

Proposition 2.3.1 gives an upper bound on the size of our perturbation. We now find a a lower bound on the size of our perturbation that will allow us to regularize the asymptotic Gibbs measure. If there is a gap between the upper and lower bounds, we can construct a perturbation that will regularize the Gibbs measure without affecting the limit of the free energy.

We define the perturbed Gibbs measure,

$$G_N^{\text{pert}}(\boldsymbol{\sigma}) = \frac{\exp(H_N^{\text{pert}})}{Z_N} \qquad \text{and} \qquad Z_N = \int_{\Sigma_N} e^{H_N^{\text{pert}}(\boldsymbol{\sigma})}\, d\mu_N(\boldsymbol{\sigma}). \tag{2.35}$$

Furthermore, suppose we also have the following control over the concentration of the corresponding perturbed free energy

$$\sup\left\{\mathbb{E}|\log Z_N - \mathbb{E}\log Z_N| \mid 0 \le u_p \le 3, p \ge 1\right\} \le v_N(s_N). \tag{2.36}$$

The dependence on s_N is through the perturbation term $s_N h_N$. For any $n \ge 2$, $p \ge 1$, and function f of the overlaps $R^n = (R_{\ell,\ell'})_{\ell,\ell' \le n}$, we define the quantity

$$\Delta(f,n,p) = \left|\mathbb{E}\langle f(R^n) R_{1,n+1}^p\rangle - \frac{1}{n}\mathbb{E}\langle f(R^n)\rangle \mathbb{E}\langle R_{1,2}^p\rangle - \frac{1}{n}\sum_{\ell=2}^{n} \mathbb{E}\langle f(R^n) R_{1,\ell}^p\rangle\right| \tag{2.37}$$

to denote the error of the Ghirlanda–Guerra identities defined in (2.25). Notice that $\Delta(f,n,p) = 0$ if and only if the array satisfies (2.25). In the definition of (2.32), we replace the deterministic constant u_p with random i.i.d. $u_p \in [1,2]$ uniformly distributed and independent from all other sources of randomness. If s_N is large enough, then we will show that the perturbed Hamiltonian satisfies the Ghirlanda–Guerra identities on average.

Theorem 2.3.1 (*The Ghirlanda–Guerra Identities*)

If s_N satisfies

$$\lim_{N\to\infty} s_N = \infty \qquad \textit{and} \qquad \lim_{N\to\infty} \frac{v_N(s_N)}{s_N^2} = 0, \tag{2.38}$$

then

$$\lim_{N\to\infty} \mathbb{E}_u \Delta(f,n,p) = 0$$

for any $p \ge 1$, $n \ge 2$ and any bounded measurable function $f(R^n)$. The average is with respect to the i.i.d. uniform random variables $u_p \in [1,2]$ appearing in the covariance (2.32).

Proof. The proof follows from a concentration inequality and integration by parts. The proof is found in [76, Chapter 3.2]. We will recreate the proof here, because we will be referencing a critical step later in the thesis.

To simplify notation, we let $h_p(\boldsymbol{\sigma})$ to be a Gaussian process with covariance

$$\mathbb{E}h_p(\boldsymbol{\sigma}^1)h_p(\boldsymbol{\sigma}^2) = R_{1,2}^p.$$

The perturbation term $h_N(\boldsymbol{\sigma})$ is a linear combination of such Gaussians. These Gaussian processes

provide information about the pth moment of the overlaps. By the convexity of the log partition functions and the bounded derivatives of the log partition functions with respect to s_N [76, Theorem 3.3], it can be shown that for any $p \geq 1$ and s_N such that $\frac{v_N(s_N)}{s_N^2} \leq 4^{-p}$,

$$\int_1^2 \mathbb{E}\langle |h_p(\boldsymbol{\sigma}) - \mathbb{E}\langle h_p(\boldsymbol{\sigma})| \, du\rangle \leq 2 + 48\sqrt{v_N(s_N)}. \tag{2.39}$$

The inequality provides control over the errors of the approximate Ghirlanda–Guerra identities if we have concentration of the perturbed log partition functions.

This control over the moments will allow us to prove the Ghirlanda–Guerra identities using a Gaussian concentration argument. We first do the argument conditionally on u_p. Let f be a bounded function of n replica. Without loss of generality, we can normalize f so that $|f| \leq 1$. Consider the concentration result for the pth power of the overlap

$$|\mathbb{E}\langle f h_p(\boldsymbol{\sigma})\rangle - \mathbb{E}\langle f\rangle \mathbb{E}\langle h_p(\boldsymbol{\sigma})\rangle| \leq \mathbb{E}\langle |h_p(\boldsymbol{\sigma}) - \mathbb{E}\langle h_p(\boldsymbol{\sigma})|\rangle \tag{2.40}$$

The bound (2.39) gives control over this error.

The left hand side of (2.40) is actually equal to the Ghirlanda–Guerra identities provided that the self overlaps are fixed. By Gaussian integration by parts (Lemma 2.1.2), and the fact that the covariance of the perturbation term $s_N h_N(\boldsymbol{\sigma})$ has a $s_N 2^{-p} u_p$ coefficient in front of $R_{1,2}^p$, we see that

$$\mathbb{E}\langle h_p(\boldsymbol{\sigma})\rangle = s_N 2^{-p} u_p (\mathbb{E}\langle R_{1,1}^p - R_{1,2}^p\rangle) = s_N 2^{-p} u_p (1 - \mathbb{E}\langle R_{1,2}^p\rangle). \tag{2.41}$$

Here we used the fact that the self overlap is constant, i.e. $R_{1,1} = 1$. Similarly, by Gaussian integration by parts

$$\mathbb{E}\langle f h_p(\boldsymbol{\sigma})\rangle = s_N 2^{-p} u_p \mathbb{E}\left\langle f \sum_{\ell=1}^n R_{1,\ell}^p - n f R_{1,n+1} \right\rangle. \tag{2.42}$$

Since $R_{1,1} = 1$, substituting (2.41) and (2.42) into the left hand side of (2.40) implies that

$$|\mathbb{E}\langle f h_p(\boldsymbol{\sigma})\rangle - \mathbb{E}\langle f\rangle \mathbb{E}\langle h_p(\boldsymbol{\sigma})\rangle| = s_N 2^{-p} u_p n \Delta(f, n, p) \tag{2.43}$$

where $\Delta(f, n, p)$ was defined in (2.37). A critical step of this proof was the fact that the self overlap terms canceled. Averaging both sides over $u_p \in [1, 2]$ implies that

$$s_N 2^{-p} n \mathbb{E}_{u_p} \Delta(f, n, p) \leq 2 + 48\sqrt{v_N(s_N)}$$

provided that $\frac{v_N(s_N)}{s_N^2} \leq 4^{-p}$. We can now average over all $u_p \in [1, 2]$ for $p \geq 1$ and rearrange terms to conclude that

$$\mathbb{E}_u \Delta(f, n, p) \leq \frac{2^p}{n}\left(\frac{2}{s_N} + 48\frac{\sqrt{v_N(s_N)}}{s_N}\right) \to 0$$

if (2.38) holds. □

The perturbation also implies a positivity property of the overlaps. This is important because the overlaps generated by the Ruelle probability cascades must be non-negative. The perturbation also implies that the overlaps are non-negative under the approximate Ghirlanda–Guerra identities. This

positivity is uniform over all choices of Hamiltonian $H_N(\boldsymbol{\sigma})$, so it is a very general result.

Theorem 2.3.2 (*Talagrand's Positivity Principle*)

If s_N satisfies

$$\lim_{N\to\infty} s_N = \infty, \tag{2.44}$$

then

$$\lim_{N\to\infty} \sup_{H_N(\boldsymbol{\sigma})} \mathbb{E}_u \mathbb{E} G_N^{\otimes 2}(R_{1,2} \leq -\varepsilon) = 0.$$

Proof. See [76, Chapter 3.3] for the details. □

2.3.3 Application to Mixed p-spin Models

We now show how this general perturbation principle can be applied to the mixed p-spin models appearing in spin glasses. The overlaps are bounded in mixed p-spin models, so we have the following bound uniform bound of $u_p \in [0,3]$ on the variance of the Hamiltonian

$$\mathbb{E}(H_N(\boldsymbol{\sigma}) + s_N h_N(\boldsymbol{\sigma}))^2 \leq N\xi'(1) + 9s_N^2. \tag{2.45}$$

By the Gaussian concentration of free energy type functions [Theorem 2.1.4 and Theorem 3.1.3], after multiplying both sides by N^2, we have

$$\mathbb{E}(\log Z_N - \mathbb{E}\log Z_N)^2 \leq 8(N\xi'(1) + 9s_N^2). \tag{2.46}$$

Therefore, by Jensen's inequality,

$$\mathbb{E}|\log Z_N - \mathbb{E}\log Z_N| \leq (\mathbb{E}(\log Z_N - \mathbb{E}\log Z_N)^2)^{1/2} \leq \sqrt{8(N\xi'(1) + 9s_N^2)}. \tag{2.47}$$

We conclude that the conditions of Theorem 2.3.1 and Theorem 2.3.2 hold for $s_N = N^\gamma$ for $1/4 < \gamma$. By Proposition 2.3.1 we also require that $\gamma < \frac{1}{2}$ for the free energy to remain unchanged. We can take $s_N = N^\gamma$ for $1/4 < \gamma < 1/2$, to ensure that the there exists a perturbation that is large enough to force the Ghirlanda–Guerra identities without changing the limit of the free energy.

If we take $s_N = N^\gamma$ where $1/4 < \gamma < 1/2$, then $\lim_{N\to\infty} \mathbb{E}_u \Delta(f,n,p) = 0$. We need this to hold for all $p \geq 1$ to get the full Ghirlanda–Guerra identities. If the Ghirlanda–Guerra identities hold for the following countable collection of functions,

$$\mathscr{F} = \Big\{(f,n,p) \mid p \geq 1, n \geq 2, f \text{is a monomial}\Big\} \tag{2.48}$$

then we can recover the full Ghirlanda–Guerra identities by approximation. Let (f_j, n_j, p_j) be an enumeration of the elements of $\mathscr{F}$. If we define

$$\Delta_N(u) = \sum_{j\geq 1} 2^{-j} \Delta(f_j, n_j, p_j), \tag{2.49}$$

then Theorem 2.3.1 implies that $\mathbb{E}_u \Delta_N(u) \to 0$ by the dominated convergence theorem, because $|\Delta(f_j, n_j, p_j)|$ is bounded by 2.

Since the convergence holds on average, by the probabilistic method [76, Lemma 3.3], there exists a deterministic sequence $u^N = (u_p^N)_{p\geq 1}$ such that

$$\lim_{N\to\infty} \Delta_N(u^N) = 0,$$

where G_N^{pert} defined in (2.35) is with respect to this deterministic $(u_p^N)_{p\geq 1}$. This sequence u^N can also be chosen so that the average perturbed Gibbs measures appearing in the cavity computations can be replaced by this sequence simultaneously.

The covariances $C_{\ell,\ell'}^p$ defined in (2.33) form a basis for bounded measurable functions, because the $C_{\ell,\ell'}^p$ are bounded monomials of $R_{\ell,\ell'}$. Furthermore, the entries of the array $(R_{\ell,\ell'})_{\ell,\ell'\geq 1}$ is bounded, so it is compact because it is a countable product of compact spaces. Therefore, the space of probability measures on these arrays is tight, so the selection theorem [82, Theorem 3.2] implies that there exists a subsequence such that the array $(R_{\ell,\ell'})_{\ell,\ell'\geq 1}$ under $\mathbb{E}G_N^{\otimes\infty}$ converges in distribution.

By the Dovbysh–Sudakov representation [Theorem 2.2.3], the limiting array is generated by the inner products of an i.i.d. samples from G a measure on the from the unit ball of a separable Hilbert space. If $\mathbb{E}$ is the average with respect to the measure G, then the limiting array must satisfy

$$\mathbb{E}f\psi_{1,n+1} = \frac{1}{n}\mathbb{E}f\mathbb{E}\psi_{1,2} + \frac{1}{n}\sum_{\ell=2}^{n}\mathbb{E}f\psi_{1,\ell} \tag{2.50}$$

where

$$\psi_{\ell,\ell'} = \psi(R_{\ell,\ell'})$$

and ψ is a bounded measurable function single variable function and f is a function of n overlaps. Arrays generated from the limiting Gibbs measure G will satisfy the Ghirlanda–Guerra identities exactly, so we can approximate it in distribution with arrays generated from the Ruelle probability cascades, giving us an explicit way to compute the free energy. This will be appear when we prove the lower bounds of the vector spin free energies in Chapter 4.

Chapter 3

Preliminaries for Spherical and Vector Spin Models

In this chapter, we explain how the machinery used to study classical spin glass models can be adapted to spherical models and vector spin models. Again, the proofs of these results can be found in various textbooks on these subjects, so we will not recreate the proofs of the classical results. We will show how these classical results can be applied to prove the key results in the study of spherical vector spin glasses.

In the previous section, our results were developed for finite or countable configuration spaces. In this section, we focus on more general configuration spaces such as S_N, the $N-1$ dimensional sphere in $\mathbb{R}^n$ with radius $\sqrt{N}$, and $(\mathbb{R}^n)^N$, the configuration space of n dimensional vector spins. We will prove several facts about these configuration spaces that will be used throughout the thesis.

Outline of the Chapter

In Section 3.1, we explore the geometry of high dimensional spheres and its connections with high dimensional Gaussians. Next in Section 3.1.2, we prove some technical regularity results that will be used to extend the results for countable configuration spaces covered in Chapter 2 to the setting of spherical spin glasses. We then move onto the study of vector spin models in Section 3.2 where we explain how a generalized class of Ghirlanda–Guerra identities imply a synchronization property in these models. Lastly, we conclude this chapter in Section 3.3 by stating several facts from matrix analysis that will be used to study the properties of the matrix functionals that appear in Chapter 4 and Chapter 5.

3.1 Geometry of High Dimensional Spheres

The primary objective of this thesis is to prove free energy formulas for a class of spherical spin glass models. This means that the methods used to study the Ising models in the classical spin glass models have to be adapted to the spherical setting. The Ising case is considerably simpler to study mathematically. The uniform measure on $\Sigma_N = \{-1,+1\}^N$ is supported on a finite set and the measure can be decomposed into a product measure on $\{-1,+1\}$. These nice properties are not automatically satisfied by the sphere, so some extra care must be taken in our analysis.

We begin by defining the spherical configuration space and the class of Gaussian processes on the

sphere. Let S_N^ρ denote the sphere in $\mathbb{R}^N$ with radius ρ,

$$S_N^\rho = \{\boldsymbol{\sigma} \in \mathbb{R}^N \mid \|\boldsymbol{\sigma}\|_2 = \rho\}. \tag{3.1}$$

To simplify notation, we denote the sphere of radius $\sqrt{N}$ by $S_N^{\sqrt{N}} := S_N$. Let λ_N^ρ and λ_N denote the respective normalized uniform measures on S_N^ρ and S_N, i.e. $\lambda_N^\rho(S_N^\rho) = 1$. Recall that the Hamiltonians (1.26) in spherical spin glasses are Gaussian processes indexed by points on S_N. In this section, we state several facts about high dimensional spheres that will be applied throughout the derivation of these free energy formulas.

3.1.1 Connection Between Spherical and Gaussian Distributions

A common theme throughout this thesis is the connection between high dimensional spheres and Gaussian random variables. Loosely speaking, a vector from a high dimensional Gaussian and a vector sampled uniformly on a high dimensional sphere are statistically indistinguishable.

Polar Coordinates

Let ν_N be the standard Gaussian measure on $\mathbb{R}^N$. We write each coordinate $\boldsymbol{\omega} \in \mathbb{R}^N$ in its polar coordinate form $\boldsymbol{\omega} = s\sigma$, where $s = \frac{\|\boldsymbol{\omega}\|}{\sqrt{N}} \in \mathbb{R}^+$ and $\sigma = \frac{\sqrt{N}\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|} \in S_N$. Let γ_N denote the law of s under ν_N. The standard Gaussian is rotationally invariant, so the distribution of $\boldsymbol{\sigma}$ is uniform on S_N and independent from the distribution of s.

Proposition 3.1.1

The law of $\boldsymbol{\sigma}$ under ν_N is λ_N and the law of s^2 under ν_N is $\Gamma(\frac{N}{2}, \frac{N}{2})$. Furthermore, $\boldsymbol{\sigma}$ and s are independent.

Proof. We find the distribution $\boldsymbol{\sigma}$ under ν_N. Consider a measurable $A \subset S_N$. Writing the Gaussian integral in polar coordinates,

$$\nu(A) = \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} \mathbb{1}(\boldsymbol{\sigma} \in A) e^{-\frac{\|\boldsymbol{\omega}\|^2}{2}}\, d\boldsymbol{\omega} = \frac{1}{(2\pi)^{N/2}} \int_0^\infty \int_{S_N^1} \mathbb{1}(\sqrt{N}\tilde{\boldsymbol{\sigma}} \in A) r^{N-1} e^{-\frac{r^2}{2}}\, dr d\lambda_N^1(\tilde{\boldsymbol{\sigma}}), \tag{3.2}$$

where λ_N^1 is the uniform measure on S_N^1, the $N-1$ dimensional unit sphere in $\mathbb{R}^N$. Since

$$\int_0^\infty r^{N-1} e^{-\frac{r^2}{2}}\, dr \overset{x=\frac{r^2}{2}}{=} 2^{\frac{N-1}{2}} \int_0^\infty x^{\frac{N-1}{2}} e^{-x} (2x)^{-\frac{1}{2}}\, dx = 2^{\frac{N}{2}-1}\Gamma\Big(\frac{N}{2}\Big),$$

if we do the change of variables $\boldsymbol{\sigma} = \sqrt{N}\tilde{\boldsymbol{\sigma}}$, then (3.2) simplifies to

$$\nu(A) = \frac{1}{2\pi^{N/2}}\Gamma\Big(\frac{N}{2}\Big) N^{-\frac{N-1}{2}} \int_{S_N} \mathbb{1}(\boldsymbol{\sigma} \in A) d\lambda_N^{\sqrt{N}}(\tilde{\boldsymbol{\sigma}}) = \lambda_N(A),$$

since the surface area of S_N is $\frac{2\pi^{N/2}}{\Gamma(\frac{N}{2})} N^{\frac{N-1}{2}}$, so the uniform distribution is appropriately normalized.

Similarly, to find the distribution s^2 under ν_N, we consider a measurable $B \subset \mathbb{R}^+$. Writing the

Gaussian integral in polar coordinates,

$$\nu(B) = \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} \mathbb{1}(s^2 \in B) e^{-\frac{\|\boldsymbol{\omega}\|^2}{2}} \, d\boldsymbol{\omega} = \frac{1}{(2\pi)^{N/2}} \int_0^\infty \int_{S_N^1} \mathbb{1}(N^{-1}r^2 \in B) s^{N-1} e^{-\frac{r^2}{2}} \, dr d\lambda_N^1(\tilde{\boldsymbol{\sigma}}).$$

Since the surface area of S_N^1 is $\frac{2\pi^{N/2}}{\Gamma(\frac{N}{2})}$, this integral simplifies to

$$\frac{1}{2^{\frac{N}{2}-1}\Gamma(\frac{N}{2})} \int_0^\infty \mathbb{1}(N^{-1}r^2 \in B) r^{N-1} e^{-\frac{r^2}{2}} \, dr \overset{s=N^{-\frac{1}{2}}r}{=} \frac{N^{\frac{N}{2}}}{2^{\frac{N}{2}-1}\Gamma(\frac{N}{2})} \int_0^\infty \mathbb{1}(s^2 \in B)(s^2)^{\frac{N-1}{2}} e^{-N\frac{s^2}{2}} \frac{1}{2s} \, ds^2$$

so s^2 has a $\Gamma(\frac{N}{2}, \frac{N}{2})$ distribution.

Independence follows from a similar computation, because the expression in terms of polar coordinates already decomposes $d\boldsymbol{\omega}$ into a product measure. We can also normalize both dr and $d\lambda_N^1$ to make them probability measures after a change of variables $\sqrt{N}^{-1} r \mapsto s$ and $\sqrt{N}\tilde{\boldsymbol{\sigma}} \mapsto \boldsymbol{\sigma}$. □

When N is big, s concentrates around 1 by the weak law of large numbers. This implies that a high dimensional Gaussian concentrates around a narrow band of the S_N. That is, a high dimensional Gaussian is approximately uniform on the sphere of radius N. Let

$$E_N(\delta) = \left\{\boldsymbol{\omega} \in \mathbb{R}^N \mid \|\boldsymbol{\omega}\|^2 \in [(1-\delta)N, (1+\delta)N]\right\} = \left\{\boldsymbol{\omega} \in \mathbb{R}^N \mid \|\boldsymbol{\omega}\| \in [\sqrt{(1-\delta)N}, \sqrt{(1+\delta)N}]\right\}.$$

We will show that the probability of this set is not exponentially small.

Proposition 3.1.2

For every $\delta > 0$,

$$\lim_{N\to\infty} \frac{\log \nu_N(E_N(\delta))}{N} = 0.$$

Proof. This follows immediately from the weak law of large numbers. Since $\|\boldsymbol{\omega}\|^2 = \sum_{i=1}^N g_i^2$ where (g_i) are i.i.d. standard Gaussians, for any $\delta > 0$

$$\lim_{N\to\infty} \nu_N(E_N(\delta)) = \lim_{N\to\infty} \mathbb{P}\left(\left|\sum_{i=1}^N g_i^2 - N\right| \le N\delta\right) = 1.$$

Taking logarithms and normalizing implies our result. □

In the next subsection, we will show that there is also an inverse relationship. That is, a high dimensional sphere is almost a high dimensional Gaussian in the sense that the projections of a sphere onto the a low dimensional space are approximately Gaussian.

The Poincaré Limit

To do cavity computations we will need to decompose integrals along spheres over its coordinates. In this subsection, we will present an approach to integrate spherical integrals conditionally along a subspace. This approach can be used to prove the Poincaré Lemma and the pure state expansion to derive the TAP equations in spherical models. Furthermore, we can also use this approach to prove the Gaussian concentration inequality from the classical isoperimetric inequality.

Consider a unit vector $\boldsymbol{m}$. Given $\boldsymbol{\sigma} \in S_N$, consider the following orthogonal decomposition onto $\boldsymbol{m}$

$$\boldsymbol{\sigma} = \alpha \boldsymbol{m} + \hat{\boldsymbol{\sigma}},$$

where $\alpha \in [-\sqrt{N}, \sqrt{N}]$ and $\hat{\boldsymbol{\sigma}} \perp \boldsymbol{m}$. The notation λ_N will denote the normalized surface measure on the sphere of radius $N-1$ in $\mathbb{R}^N$.

Lemma 3.1.1 (*Disintegration of the Surface Measure*)

For any function $f : S_N \to \mathbb{R}$,

$$\int_{S_N} f(\boldsymbol{\sigma})\, d\lambda_N(\boldsymbol{\sigma}) = \int_{-\sqrt{N}}^{\sqrt{N}} \int_{S_{N-1}} f\left(\alpha \boldsymbol{m} + \sqrt{\frac{N-\alpha^2}{N-1}}\hat{\boldsymbol{\sigma}}\right) \frac{\Gamma(\frac{N}{2})}{\sqrt{N}\sqrt{\pi}\Gamma(\frac{N-1}{2})} \left(1 - \frac{\alpha^2}{N}\right)^{\frac{N-3}{2}} d\lambda_{N-1}(\hat{\boldsymbol{\sigma}}) d\alpha, \tag{3.3}$$

where λ_{N-1} is the uniform surface measure on the $N-2$ dimensional sphere perpendicular to $\boldsymbol{m}$.

Proof. The proof is a straightforward calculus computation [74, Equation (2.22)]. Instead of repeating these computations, we will provide an intuitive interpretation of this formula in the context of the disintegration theorem and conditional probabilities [82, Theorem 1.10],

$$\int_{S_N} f(\boldsymbol{\sigma})\, d\lambda_N(\boldsymbol{\sigma}) = \int_{\pi(S_N)} \left(\int_{S_N} f(\boldsymbol{\sigma})\, d\lambda_{N,\alpha}(\boldsymbol{\sigma}) \right) d\lambda_N \circ \pi^{-1}(\alpha) \tag{3.4}$$

where π is the projection of $\boldsymbol{\sigma}$ onto $\boldsymbol{m}$ and $\lambda_{N,\alpha}(\boldsymbol{\sigma})$ is the conditional probability of $\boldsymbol{\sigma}$ given $\pi(\boldsymbol{\sigma}) = \alpha$.

Let λ_N^ρ denote the surface measure on the $N-1$ dimensional sphere in $\mathbb{R}^N$ with radius ρ and let $|S_N^\rho|$ denote the surface area of the $N-1$ dimensional sphere of radius ρ in $\mathbb{R}^N$. If we consider the projection map $\pi : S_N \to [-\sqrt{N}, \sqrt{N}]$ given by $\pi(\boldsymbol{\sigma}) = \langle \boldsymbol{\sigma}, \boldsymbol{m} \rangle$, then the distribution of $\alpha = \langle \boldsymbol{\sigma}, \boldsymbol{m} \rangle = \pi(\boldsymbol{\sigma})$ under λ_N is

$$\begin{aligned}
\lambda_N \circ \pi^{-1}(d\alpha) &= \mathbb{1}([-\sqrt{N}, \sqrt{N}])(\alpha) \frac{\lambda_N(\pi^{-1}(d\alpha))}{|S_N^{\sqrt{N}}|} \\
&= \mathbb{1}([-\sqrt{N}, \sqrt{N}])(\alpha) \frac{|S_{N-1}^{\sqrt{N-\alpha^2}}|}{|S_N^{\sqrt{N}}|} \frac{\sqrt{N}}{\sqrt{N-\alpha^2}} d\alpha \\
&= \mathbb{1}([-\sqrt{N}, \sqrt{N}])(\alpha) \frac{\Gamma(\frac{N}{2})}{\sqrt{N}\sqrt{\pi}\Gamma(\frac{N-1}{2})} \left(1 - \frac{\alpha^2}{N}\right)^{\frac{N-3}{2}} d\alpha
\end{aligned} \tag{3.5}$$

because $|S_N^\rho| = \frac{2\pi^{\frac{N}{2}}}{\Gamma(\frac{N}{2})} \rho^{N-1}$. Intuitively, the probability the density of the projection π of $\boldsymbol{\sigma}$ onto $\boldsymbol{m}$ is proportional to the volume of the perpendicular band around α. In polar coordinates, $\theta = \arccos(\frac{\langle \boldsymbol{\sigma}, \boldsymbol{m} \rangle}{\sqrt{N}}) = \arccos(\frac{\alpha}{\sqrt{N}})$ so

$$\cos(\theta) = \alpha \implies -\sin(\theta) d\theta = d\alpha \implies d\theta = \frac{-d\alpha}{\sqrt{N-\alpha^2}},$$

which implies the change in volume expressed in terms of $d\alpha$ is

$$\lambda_N(\pi^{-1}(d\alpha)) = -\sqrt{N} d\theta \cdot |S_{N-1}^{\sqrt{N-\alpha^2}}| = \frac{\sqrt{N}|S_{N-1}^{\sqrt{N-\alpha^2}}|}{\sqrt{N-\alpha^2}} d\alpha,$$

which gives us the second equation in (3.5).

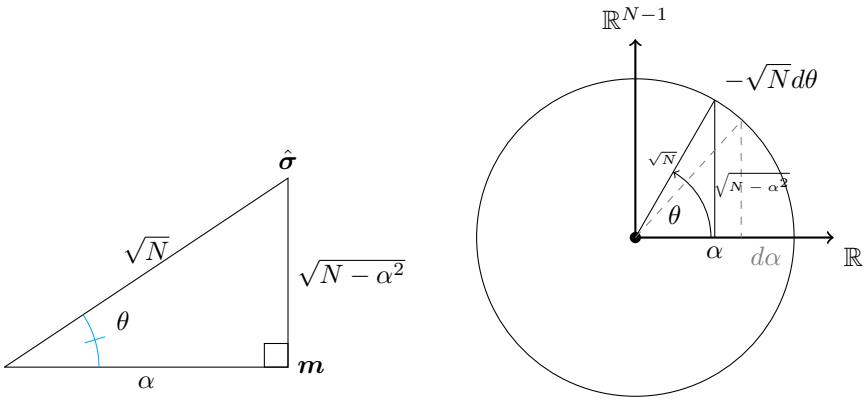

Figure 5. The value $\lambda_N(\pi^{-1}(d\alpha))$ is equal to the change of surface area of the band with width $d\alpha$. The change in surface area is given by $-\sqrt{N}d\theta \cdot |S_{N-1}^{\sqrt{N-\alpha^2}}|$.

By symmetry, the conditional probability of λ_N given α is uniform over perpendicular band $\pi^{-1}(\alpha)$, which is a $N-2$ dimensional sphere in $\mathbb{R}^N$ with radius $\sqrt{N-\alpha^2}$,

$$\lambda_{N,\alpha}(\boldsymbol{\sigma}) = \frac{d\lambda_{N-1}^{\sqrt{N-\alpha^2}}(\hat{\boldsymbol{\sigma}})}{|S_{N-1}^{\sqrt{N-\alpha^2}}|}$$

since $\hat{\boldsymbol{\sigma}}$ is the perpendicular component of $\boldsymbol{\sigma}$ in $\pi^{-1}(\alpha)$. The decomposition (3.4) can be expressed as

$$\int_{-\sqrt{N}}^{\sqrt{N}} \int_{S_{N-1}^{\sqrt{N-\alpha^2}}} f\left(\alpha \boldsymbol{m} + \sqrt{\frac{N-\alpha^2}{N-1}}\hat{\boldsymbol{\sigma}}\right) \frac{\Gamma(\frac{N}{2})}{\sqrt{N}\sqrt{\pi}\Gamma(\frac{N-1}{2})} \left(1-\frac{\alpha^2}{N}\right)^{\frac{N-3}{2}} \frac{d\lambda_{N-1}^{\sqrt{N-\alpha^2}}(\hat{\boldsymbol{\sigma}})}{|S_{N-1}^{\sqrt{N-\alpha^2}}|} d\alpha. \tag{3.6}$$

We can now do the change of variables [65, Proposition 10.16] of the form $\hat{\sigma} \mapsto \sqrt{\frac{N-1}{N-\alpha^2}}\hat{\sigma}$ to remove the dependence of the conditional probability on α to conclude (3.6) can be expressed as

$$\int_{S_N} f(\boldsymbol{\sigma})\, d\lambda_N(\boldsymbol{\sigma}) = \int_{-\sqrt{N}}^{\sqrt{N}} \int_{S_{N-1}} f\left(\alpha \boldsymbol{m} + \sqrt{\frac{N-\alpha^2}{N-1}}\hat{\boldsymbol{\sigma}}\right) \frac{\Gamma(\frac{N}{2})}{\sqrt{N}\sqrt{\pi}\Gamma(\frac{N-1}{2})} \left(1-\frac{\alpha^2}{N}\right)^{\frac{N-3}{2}} d\lambda_{N-1}(\hat{\boldsymbol{\sigma}}) d\alpha.$$

The Jacobian factor $(\frac{N-1}{N-\alpha^2})^{\frac{N-2}{2}}$ is canceled out by the normalization term in the surface integral. □

Remark 3.1.1. We will usually take $\boldsymbol{m} = e_N$. The result of Lemma 3.1.1 gives us a way to decouple the last coordinate of the surface measure. This step will be useful to compare the systems of different sizes in the cavity computations.

One of the main difficulties in the study of spherical spin glass models is the lack of symmetry across the spin coordinates. In models, the configuration spaces can be written as products. For example, Ising spin models $\Sigma_N = \{-1,+1\}^N$ which can be written as a product space of $\{-1,+1\}$ and its corresponding measure is a product of uniform distributions on $\{-1,+1\}$. The spherical models, however, do not have this nice symmetry. The Poincaré limit will give this symmetry in an approximate sense.

We first provide some intuition behind this result. Let $\boldsymbol{g} = (g_i)_{i \le N}$ be a sequence of independent standard Gaussians, and let $\|\boldsymbol{g}\|^2 = g_1^2 + \cdots + g_N^2$ be the sum of squares. From the rotational invariance

of the Gaussian vectors discussed in the previous section, for any measurable set $A \subset S_N$,

$$\mathbb{P}\left(\frac{\sqrt{N}}{\|\boldsymbol{g}\|_2}(g_1,\dots,g_N) \in A\right) \to \lambda_N(A).$$

Focusing on the first M coordinates the law of large numbers implies $\frac{\|\boldsymbol{g}\|_2}{\sqrt{N}} \to 1$, so for a measurable set $B \subset \mathbb{R}^M$,

$$\mathbb{P}\left(\frac{\sqrt{N}}{\|\boldsymbol{g}\|_2}(g_1,\dots,g_M) \in B\right) \to \nu_M(B),$$

where ν_M is the standard Gaussian measure on $\mathbb{R}^M$. We can make this statement precise.

Theorem 3.1.1 (*Poincaré Lemma*)

Let $\pi_{N,M} : \mathbb{R}^N \to \mathbb{R}^M$ be the projection onto $\mathbb{R}^M$. We have

$$\lambda_N(\pi_{N,M}^{-1}(A) \cap S_N) \to \nu_M(A) \text{ as } N \to \infty,$$

where ν_M is the standard Gaussian measure on $\mathbb{R}^M$.

Proof. We apply the decomposition in Lemma 3.1.1. Taking $\boldsymbol{m} = e_1$, and using Fubini's theorem, we see

$$\int_{S_N} f(\boldsymbol{\sigma})\, d\lambda_N(\boldsymbol{\sigma}) = \int_{S_{N-1}} \int_{-\sqrt{N}}^{\sqrt{N}} f\left(\alpha, \sqrt{\frac{N-\alpha^2}{N-1}}\hat{\boldsymbol{\sigma}}\right) \frac{\Gamma(\frac{N}{2})}{\sqrt{N}\sqrt{\pi}\Gamma(\frac{N-1}{2})}\left(1-\frac{\alpha^2}{N}\right)^{\frac{N-3}{2}} d\alpha d\lambda_{N-1}(\hat{\boldsymbol{\sigma}}). \quad (3.7)$$

Assuming f is integrable, we can use the dominated convergence theorem and the fact

$$\lim_{N\to\infty} \frac{\Gamma(\frac{N}{2})}{\sqrt{N}\sqrt{\pi}\Gamma(\frac{N-1}{2})}\left(1-\frac{\alpha^2}{N}\right)^{\frac{N-3}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{\alpha^2}{2}}$$

to conclude the statement for a projection onto a one dimensional subspace. Repeating the decomposition in Lemma 3.1.1 on λ_{N-1} by projecting it onto e_2, etc will prove the statement for general projections. □

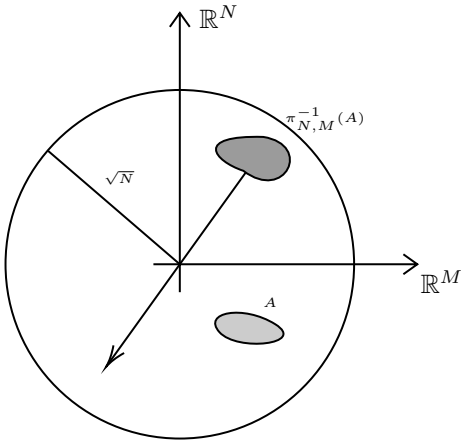

Figure 6. The volume of $A \subset \mathbb{R}^M$ under λ_N is approximately $\nu_M(A)$ when N is large.

The main consequence of this result is that the first M coordinates of a vector sampled uniformly from a high dimensional sphere is approximately Gaussian. This is essentially the converse of the polar decomposition result in the last section.

3.1.2 Regularity of Spherical Gaussian Processes

The results stated in Chapter 2 has been stated for configuration spaces with finite or countably many points. In this section, we state a regularity property of Gaussian processes on the sphere that will allow us to extend these results to spherical models. The main result will be a chaining result called Dudley's integral inequality, which will allow us to control the errors when discretizing the uniform measure on the sphere.

We first introduce the spherical models. Let S_N denote the sphere in $\mathbb{R}^N$ with radius $\sqrt{N}$,

$$S_N = \{\sigma \in \mathbb{R}^n \mid \|\sigma\|_2 = \sqrt{N}\}. \tag{3.8}$$

We define λ_N to be the uniform surface measure on S_N. Consider centered Gaussian process $(X(\boldsymbol{\sigma}))_{\boldsymbol{\sigma}\in S_N}$ taking values in $\mathbb{R}$ indexed by points on the sphere with covariance structure given by inner products. In particular, the covariance of $(X(\boldsymbol{\sigma}))_{\boldsymbol{\sigma}\in S_N}$ is given by

$$\mathbb{E}X(\boldsymbol{\sigma}^1)X(\boldsymbol{\sigma}^2) = NC_X\Big(\frac{\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2}{N}\Big) = NC_X(R_{1,2}) \qquad \text{where} \qquad R_{1,2} = \frac{\boldsymbol{\sigma}^1\cdot\boldsymbol{\sigma}^2}{N} \tag{3.9}$$

for some continuous bounded Lipschitz function $C_X : [-1,1] \to \mathbb{R}$. These Gaussian processes are the Hamiltonians appearing in spherical spin glass models.

To apply the results from Chapter 2, we will need to extend the results to Gaussian processes indexed by this configuration space with uncountably many points. We will use a chaining argument to partition the sphere into smaller parts and control the total deviation. We do this rigorously by using a classical result called Dudley's Integral inequality.

To this end, for $\varepsilon > 0$, we define the *covering number* $\mathcal{N}(K, d, \varepsilon)$ for a subset K of a metric space (T, d) to be the smallest number of open balls of radius ε whose union covers K. The covering number of the sphere is well known.

Lemma 3.1.2 (*Covering Number of the Euclidean Sphere*)

For any $\varepsilon > 0$, the covering numbers of the $N-1$ dimensional unit sphere S_N^1 in $\mathbb{R}^N$ satisfies

$$\mathcal{N}(S_N, \|\cdot\|_2, \varepsilon) \le \Big(\frac{2\sqrt{N}}{\varepsilon} + 1\Big)^N.$$

Proof. This proof of this result for the unit sphere can be found in [107, Corollay 4.2.13]. It states for any $\varepsilon > 0$, the covering numbers of the $N-1$ dimensional unit sphere S_N^1 in $\mathbb{R}^N$ satisfies

$$\mathcal{N}(S_N^1, \|\cdot\|_2, \varepsilon) \le \Big(\frac{2}{\varepsilon} + 1\Big)^N.$$

We use a scaling argument to state this in terms of S_N. Consider balls of radius $\frac{\varepsilon}{\sqrt{N}}$. The covering number $\mathcal{N}(S_N, \|\cdot\|_2, \varepsilon)$ is equivalent to the covering number $\mathcal{N}(S_N^1, \|\cdot\|_2, \frac{\varepsilon}{\sqrt{N}})$. Therefore, the estimate

above implies that

$$\mathcal{N}(S_N, \|\cdot\|_2, \varepsilon) = \mathcal{N}\Big(S_N^1, \frac{\varepsilon}{\sqrt{N}}\Big) \leq \Big(\frac{2\sqrt{N}}{\varepsilon} + 1\Big)^N.$$

□

Remark 3.1.2. If we assume that $\varepsilon \in (0, 2\sqrt{N}]$, then the upper bound can be simplified to

$$\mathcal{N}(S_N, \|\cdot\|_2, \varepsilon) \leq \Big(\frac{4N^{1/2}}{\varepsilon}\Big)^N. \tag{3.10}$$

This restriction on the radius ε is not really a restriction because if $\varepsilon > 2\sqrt{N}$, then the minimal covering is given by the trivial covering, so the covering number is 1.

Given a Gaussian process $(X_t)_{t\in T}$ indexed by points in T, we define the canonical norm on T induced by the Gaussian process by

$$\mathrm{d}(s,t) = \big(\mathbb{E}(X_s - X_t)^2\big)^{1/2}. \tag{3.11}$$

Dudley's integral inequality gives an upper bound on the deviations of a sub-Gaussian process defined on the metric space (T, d) in terms of its entropy and covariance.

Theorem 3.1.2 (*Dudley's Integral Inequality*)

Let $(X(\boldsymbol{\sigma}))_{\sigma\in T}$ be a mean zero random process indexed by the metric space (T, d). There exists an absolute constant L such that

$$\mathbb{E}\sup_{\mathrm{d}(\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2)\leq\delta} |X(\boldsymbol{\sigma}^1) - X(\boldsymbol{\sigma}^2)| \leq L\int_0^\delta \sqrt{\log \mathcal{N}(T, \mathrm{d}, \varepsilon)}\, d\varepsilon$$

where $\mathcal{N}(T, d, \varepsilon)$ is the covering number *of T.*

Proof. The proof of this result appears in [103, Appendix A.5] and a stronger result is proved in [107, Theorem 8.13]. The proof is a standard chaining argument. □

Dudley's Integral Inequality can be used to prove an important continuity property for Gaussian processes indexed by points on the sphere. Recall the covariance of $X(\sigma)$ defined in (3.9). Dudley's integral inequality implies that we can find a uniform control on the maximal distance between nearby points.

Corollary 3.1.1 (*Regularity of Spherical Gaussian Processes*)

There exists constant $L > 0$ independent of the covariance of $X(\sigma)$ such that for any $N \geq 2$ and $0 < \delta < 2\sqrt{N}$,

$$\mathbb{E}\Big(\sup_{\|\boldsymbol{\sigma}^1-\boldsymbol{\sigma}^2\|_2\leq\delta} |X(\boldsymbol{\sigma}^1) - X(\boldsymbol{\sigma}^2)|\Big) \leq L\|C_X\|_{\mathrm{Lip}}^{1/2} N^{\frac{1}{2}}\delta. \tag{3.12}$$

Proof. Since $X(\boldsymbol{\sigma}^1) - X(\boldsymbol{\sigma}^2) \sim N(0, 2N(C_X(1) - C_X(R_{1,2})))$ we have

$$\mathrm{d}(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) = \big(\mathbb{E}(X(\boldsymbol{\sigma}^1) - X(\boldsymbol{\sigma}^2))^2\big)^{1/2} = (2N(C_X(1) - C_X(R_{1,2})))^{1/2}.$$

Since C_X is Lipschitz, we have $C_X(1) - C_X(R_{1,2}) \leq \|C_X\|_{\mathrm{Lip}}(R_{1,1} - R_{1,2}) = \|C_X\|_{\mathrm{Lip}}(2N)^{-1}\|\boldsymbol{\sigma}^1 - \boldsymbol{\sigma}^2\|_2^2$. This implies that $\mathrm{d}(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) \leq \|C_X\|_{\mathrm{Lip}}^{1/2}\|\boldsymbol{\sigma}^1 - \boldsymbol{\sigma}^2\|_2$, so balls with respect to the $\|C_X\|_{\mathrm{Lip}}^{1/2}\mathrm{d}$ norm are

contained withing balls with respect to the $\|\cdot\|_2$ norm since

$$\|\boldsymbol{\sigma}^1 - \boldsymbol{\sigma}^2\|_2 \leq \delta \implies \|C_X\|_{\text{Lip}}^{1/2} \mathrm{d}(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) \leq \delta.$$

We can apply Dudley's integral inequality (Theorem 3.1.2) and the estimate of the covering number on the sphere (Lemma 3.1.2 and its simplification (3.10)) to conclude that

$$\begin{aligned}
\mathbb{E} \sup_{\|\boldsymbol{\sigma}^1 - \boldsymbol{\sigma}^2\|_2 \leq \delta} |X(\boldsymbol{\sigma}^1) - X(\boldsymbol{\sigma}^2)| &\leq \mathbb{E} \sup_{\mathrm{d}(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) \leq \|C_X\|_{\text{Lip}}^{1/2} \delta} |X(\boldsymbol{\sigma}^1) - X(\boldsymbol{\sigma}^2)| \\
&\leq L \int_0^{\|C_X\|_{\text{Lip}}^{1/2} \delta} \sqrt{\log \mathcal{N}(S_N, d, \varepsilon)}\, d\varepsilon \\
&\leq L \int_0^{\|C_X\|_{\text{Lip}}^{1/2} \delta} \sqrt{\log \mathcal{N}\big(S_N, \|\cdot\|, \|C_X\|_{\text{Lip}}^{-1/2} \varepsilon\big)}\, d\varepsilon \\
&\leq L \int_0^{\|C_X\|_{\text{Lip}}^{1/2} \delta} \sqrt{N \log\Big(\frac{4\|C_X\|_{\text{Lip}}^{1/2} N^{1/2}}{\varepsilon}\Big)}\, d\varepsilon \\
&\leq LN \|C_X\|_{\text{Lip}}^{1/2} \int_{\frac{4N^{1/2}}{\delta}}^{\infty} \frac{\sqrt{\log(s)}}{s^2}\, ds
\end{aligned} \tag{3.13}$$

by the change of variables $s = \frac{4\|C_X\|_{\text{Lip}}^{1/2} N^{1/2}}{\varepsilon}$. The integral satisfies the bound

$$\int_{\frac{4N^{1/2}}{\delta}}^{\infty} \frac{\sqrt{\log(s)}}{s^2}\, ds \leq \int_{\frac{4N^{1/2}}{\delta}}^{\infty} \frac{1}{s^{3/2}}\, ds \leq 2N^{-1/2}\delta$$

since $\frac{4N^{1/2}}{\delta} \geq 1$ for $\delta \in (0, 2\sqrt{N})$. Redefining L and using this bound in (3.13) proves the result. □

The regularity of spherical Gaussian processes will allow us to partition the configuration spaces into balls effectively reducing the spherical case to the finite dimensional case. We now extend the Gaussian concentration result and Gaussian integration results to spherical configuration spaces.

Gaussian Concentration

In this subsection, we extend the Gaussian concentration result for free energy functions (Theorem 2.1.4) to free energy functions on the sphere. In particular, we want to show that free energy functionals of the form

$$\tilde{F}_N = \frac{1}{N} \log \int_{S_N} e^{X(\boldsymbol{\sigma})}\, d\lambda_N(\boldsymbol{\sigma}), \tag{3.14}$$

satisfy a concentration inequality. We also assume that $X(\boldsymbol{\sigma})$ is centered with covaraince structure obeying the conditions in (3.9).

Theorem 3.1.3

The Gaussian concentration result Theorem 2.1.4 holds for spherical free energies of the form (3.14).

Proof. Theorem 2.1.4 proves this result in the case when the reference measure is countable. We want to extend the Gaussian concentration result for countable configuration spaces in Theorem 2.1.4 to the case when the reference measure is uniform on the sphere. It suffices to show that the spherical free

energies

$$\tilde{F}_N = \frac{1}{N} \log \int_{S_N} e^{X(\boldsymbol{\sigma})} \, d\lambda_N(\boldsymbol{\sigma}),$$

can be approximated by free energies by free energies over finite configuration spaces. We will use a covering argument and the bound on the expected deviations of $X(\boldsymbol{\sigma})$ on the small balls to control the errors when we partition S_N into small balls.

Given $\varepsilon > 0$, we define $M_\varepsilon = \mathcal{N}(S_N, \|\cdot\|_2, \varepsilon)$ to be the smallest number of balls of radius ε that cover S_N. We define B_i to be the intersection of the ith ball with S_N and without loss of generality, we can make the B_i disjoint by removing the intersections. For each B_i, we fix a $\sigma_i \in B_i$ and define

$$F_\varepsilon = \frac{1}{N} \log \sum_{i=1}^{M_\varepsilon} e^{X(\boldsymbol{\sigma}_i)} \lambda_N(B_i).$$

Consider a sequence $\varepsilon_n \to 0$ and the corresponding approximate free energies $F_n := F_{\varepsilon_n}$ constructed above. We show that $F_n \to \tilde{F}_N$ almost surely along a subsequence by showing that

$$\mathbb{E}|\tilde{F}_N - F_n| \to 0$$

as $n \to \infty$. By definition,

$$\mathbb{E}\left| \sum_{i=1}^{M_{\varepsilon_n}} \int_{B_i} e^{X(\boldsymbol{\sigma})} - e^{X(\boldsymbol{\sigma}_i)} \, d\lambda_N(\boldsymbol{\sigma}) \right| \le \sum_{i=1}^{M_{\varepsilon_n}} \int_{B_i} \mathbb{E}|e^{X(\boldsymbol{\sigma})} - e^{X(\boldsymbol{\sigma}_i)}| \, d\lambda_N(\boldsymbol{\sigma}). \tag{3.15}$$

To control the exponentials on the inside, we condition on the maximal value of $X(\boldsymbol{\sigma})$,

$$\int_{B_i} \mathbb{E}|e^{X(\boldsymbol{\sigma})} - e^{X(\boldsymbol{\sigma}_i)}| \, \mathbb{1}\left(\max_{\boldsymbol{\sigma} \in B_i} X(\boldsymbol{\sigma}) \le K \right) d\lambda_N(\boldsymbol{\sigma}) + \int_{B_i} \mathbb{E}|e^{X(\boldsymbol{\sigma})} - e^{X(\boldsymbol{\sigma}_i)}| \, \mathbb{1}\left(\max_{\boldsymbol{\sigma} \in B_i} X(\boldsymbol{\sigma}) > K \right) d\lambda_N(\boldsymbol{\sigma}). \tag{3.16}$$

On the first set, $\{\max_{\boldsymbol{\sigma} \in B_i} X(\boldsymbol{\sigma}) \le K\}$ the exponential function is locally Lipschitz, so Corollary 3.1.1 implies

$$\begin{aligned} \mathbb{E}|e^{X(\boldsymbol{\sigma})} - e^{X(\boldsymbol{\sigma}_i)}| \, \mathbb{1}\left(\max X(\boldsymbol{\sigma}) \le K \right) &\le \mathbb{E} e^a |X(\boldsymbol{\sigma}) - X(\boldsymbol{\sigma}_i)| \, \mathbb{1}\left(\max X(\boldsymbol{\sigma}) \le K \right) \\ &\le e^K \mathbb{E} \max_{\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_i\| < \varepsilon_n} |X(\boldsymbol{\sigma}) - X(\boldsymbol{\sigma}_i)| \\ &\le e^K L \|C_X\|_{\mathrm{Lip}} N^{\frac{1}{2}} \varepsilon_n. \end{aligned}$$

For any fixed K, we can control this term by taking ε_n sufficiently small.

On the second set, $\{\max_{\boldsymbol{\sigma} \in B_i} X(\boldsymbol{\sigma}) > K\}$ has exponentially small probability (see the tail bound on Dudley's Integral Inequality in [107, Theorem 8.1.6]). By the Cauchy–Schwarz inequality, we can compute an explicit upperbound using the formula for the moment generating function of a Gaussian,

$$\begin{aligned} \mathbb{E}|e^{X(\boldsymbol{\sigma})} - e^{X(\boldsymbol{\sigma}_i)}| \, \mathbb{1}\left(\max_{\boldsymbol{\sigma} \in B_i} X(\boldsymbol{\sigma}) > K \right) &\le \left(\mathbb{E}|e^{X(\boldsymbol{\sigma})} - e^{X(\boldsymbol{\sigma}_i)}|^2 \right)^{1/2} \left(\mathbb{P}\left(\max_{\boldsymbol{\sigma} \in S_N} X(\boldsymbol{\sigma}) > K \right) \right)^{1/2} \\ &\le 2 e^{\|C_X\|_\infty^2} \cdot e^{-C_1 (C_2 - K)^2}, \end{aligned}$$

for some positive universal constants C_1 and C_2 that do not depend on ε. This upperbound is uniform over all the partitions, so we can control it by taking K sufficiently large. Therefore, for any $\delta > 0$, we

can find K sufficiently large so that the second term in (3.16) is smaller than $\frac{\delta}{2}$ and ε_n sufficiently small so that the first term is smaller than (3.16) is smaller than $\frac{\delta}{2}$. Since $\sum_{i=1}^{M_{\varepsilon_n}} \lambda(B_i) = 1$, we have shown that (3.15) satisfies

$$\mathbb{E}\left| \sum_{i=1}^{M_{\varepsilon_n}} \int_{B_i} e^{X(\boldsymbol{\sigma})} - e^{X(\boldsymbol{\sigma}_i)} \, d\lambda_N(\boldsymbol{\sigma}) \right| \to 0$$

as $n \to \infty$. Therefore, we can take logarithms and normalize to conclude there exists a subsequence such that $F_n \to \tilde{F}_N$ almost surely.

The approximate free energy F_n can be estimated by

$$\frac{1}{N} \min_{\boldsymbol{\sigma} \in S_N} X(\boldsymbol{\sigma}) \leq F_n \leq \frac{1}{N} \int_{S_N} e^{X(\boldsymbol{\sigma})} \, d\lambda_N(\boldsymbol{\sigma}).$$

Since the covariance of $X(\boldsymbol{\sigma})$ is bounded the upper and lower bounds is integrable and since $X(\boldsymbol{\sigma})$ is symmetric, the expected value of the lower bound is equal to $-\mathbb{E} \max_{\boldsymbol{\sigma} \in S_N} X(\boldsymbol{\sigma})$ which is also finite by Dudley's integral inequality (see [107, Theorem 8.13]). The dominated convergence theorem implies that $\mathbb{E}F_n \to \mathbb{E}\tilde{F}_N$ along the same subsequence. Applying the concentration inequality Theorem 2.1.4 to F_n and taking limits finishes the proof. □

Gaussian Integration By Parts

In this subsection, we extend the Gaussian integration by parts formula for Gibbs averages to spheres. Let $X(\boldsymbol{\sigma})$ and $Y(\boldsymbol{\sigma})$ be centered jointly Gaussian random variables satisfying the covariance condition (3.9). We want to extend the integration by parts formula (Lemma 2.1.2) for Gibbs averages of the form

$$\mathbb{E}\big\langle f(\boldsymbol{\sigma}^1, \dots, \boldsymbol{\sigma}^n) X(\boldsymbol{\sigma}^1) \big\rangle \tag{3.17}$$

where $\langle \cdot \rangle$ is the average with respect to the measure with density

$$\frac{e^{Y(\boldsymbol{\sigma})}}{Z_N} d\lambda_N(\boldsymbol{\sigma}) \qquad \text{where} \qquad Z_N = \int_{S_N} e^{Y(\boldsymbol{\sigma})} \, d\lambda_N(\boldsymbol{\sigma}). \tag{3.18}$$

If $C(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) = \mathbb{E}X(\boldsymbol{\sigma}^1)Y(\boldsymbol{\sigma}^2)$, we want to show that

$$\mathbb{E}\langle f X(\boldsymbol{\sigma}^1) \rangle = E\left\langle f\left(\sum_{\ell=1}^{n} C(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^\ell) - nC(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^{n+1}) \right) \right\rangle. \tag{3.19}$$

Theorem 3.1.4 (*Gaussian Integration by Parts on the Sphere*)

The Gaussian integration by parts result Lemma 2.1.2 holds for Gibbs measures of the form (3.18).

Proof. Lemma 2.1.2 proves this result in the case when the reference measure is countable. We need to extend the integration by parts formula to the case when the reference measure is uniform on the sphere. It suffices to show that this Gibbs measure can be approximated by a reference measure with finite support. We use a covering procedure and Dudley's integral inequality (Theorem 3.1.2) to control the approximation errors.

We proceed with the same covering argument as in the proof of the Gaussian concentration inequality Theorem 3.1.3. Given a sequence $\varepsilon_n \to 0$, we define $M_n = \mathcal{N}(S_N, \| \cdot \|_2, \varepsilon_n)$ to be the smallest number

of balls of radius ε_n that cover S_N. We define B_i to be the intersection of the ith ball with S_N and without loss of generality, we can make the B_i disjoint by removing the intersections. For each B_i, we fix a $\sigma_i \in B_i$ and define the average over the discretized sphere as

$$\langle X(\boldsymbol{\sigma})\rangle_n = \frac{\sum_{i=1}^{M_n} X(\boldsymbol{\sigma}_i)e^{Y(\boldsymbol{\sigma}_i)}\lambda_N(B_i)}{Z_n} \qquad \text{where} \qquad Z_n = \sum_{i=1}^{M_n} e^{Y(\boldsymbol{\sigma}_i)}\lambda_N(B_i).$$

Since the integration by parts formula Lemma 2.1.2 applies to $\langle X(\boldsymbol{\sigma})\rangle_n$, it suffices to show that

$$|\mathbb{E}\langle X(\boldsymbol{\sigma})\rangle - \mathbb{E}\langle X(\boldsymbol{\sigma})\rangle_n| \to 0 \tag{3.20}$$

and apply the dominated convergence theorem. We accomplish this by showing that both the numerator and normalization converge almost surely along a subsequence.

(a) We want to find a bound on the normalizations appearing in (3.20). Using the exact same argument as the bound of (3.15) in the proof of Theorem 3.1.3, it follows that

$$\mathbb{E}|Z - Z_n| \le \sum_{i=1}^{M_n} \int_{B_i} |e^{Y(\boldsymbol{\sigma})} - e^{Y(\boldsymbol{\sigma}_i)}|\, d\lambda_N(\boldsymbol{\sigma}) \to 0.$$

Therefore, $Z_n \to Z$ almost surely along a subsequence.

(b) We want to find an upper bound of the terms appearing appearing in the numerator of (3.20),

$$\mathbb{E}\left|\sum_{i=1}^{M_n} \int_{B_i} X(\boldsymbol{\sigma})e^{Y(\boldsymbol{\sigma})} - X(\boldsymbol{\sigma}_i)e^{Y(\boldsymbol{\sigma}_i)}\, d\lambda_N(\boldsymbol{\sigma})\right| \le \sum_{i=1}^{M_n} \int_{B_i} \mathbb{E}|X(\boldsymbol{\sigma})e^{Y(\boldsymbol{\sigma})} - X(\boldsymbol{\sigma}_i)e^{Y(\boldsymbol{\sigma}_i)}|\, d\lambda_N(\boldsymbol{\sigma}). \tag{3.21}$$

We can proceed as in the derivation of the upper bound of (3.15). We can condition on the events that

$$\{\max_{\boldsymbol{\sigma}\in B_i} Y(\boldsymbol{\sigma}) \le K \text{ and } \max_{\boldsymbol{\sigma}\in B_i} X(\boldsymbol{\sigma}) \le K\} \quad \text{and} \quad \{\max_{\boldsymbol{\sigma}\in B_i} Y(\boldsymbol{\sigma}) > K \text{ or } \max_{\boldsymbol{\sigma}\in B_i} X(\boldsymbol{\sigma}) > K\}.$$

On the first set, $f(x, y) = xe^y$ has a bounded derivative, so it is Lipschitz and we can control the distances between $X(\boldsymbol{\sigma})$ and $Y(\boldsymbol{\sigma})$ on the balls using Corollary 3.1.1. On the second set, we can use the union bound and the fact that the probability of the tail $\{\max_{\boldsymbol{\sigma}\in B_i} Y(\boldsymbol{\sigma}) > K\}$ and $\{\max_{\boldsymbol{\sigma}\in B_i} X(\boldsymbol{\sigma}) > K\}$ are exponentially small so we can also control it. The same argument as in the derivation of the upper bound of (3.15) implies that

$$\mathbb{E}\left|\sum_{i=1}^{M_n} \int_{B_i} X(\boldsymbol{\sigma})e^{Y(\boldsymbol{\sigma})} - X(\boldsymbol{\sigma}_i)e^{Y(\boldsymbol{\sigma}_i)}\, d\lambda_N(\boldsymbol{\sigma})\right| \to 0,$$

so the numerators converge almost surely along a subsequence.

By taking a further subsequence, we can find a subsequence such that $\langle X(\boldsymbol{\sigma})\rangle_n \to \langle X(\boldsymbol{\sigma})\rangle$ almost surely. The covariances of $X(\boldsymbol{\sigma})$ and $Y(\boldsymbol{\sigma})$ are bounded, so we can apply the dominated convergence theorem to conclude that $\mathbb{E}\langle X(\boldsymbol{\sigma})\rangle_n \to \mathbb{E}\langle X(\boldsymbol{\sigma})\rangle$ along the same subsequence. The right hand side of the integration by parts formula follows from a similar approximation argument. □

3.2 Synchronization of Vector Spins

This section states the critical result that connects vector spin models with classical models. In vector spin models, the covariances of the Hamiltonians are given as convex functions of the an overlap array of block matrices. This is the critical result that allows us to reduce vector spin models with the classical one dimensional models, and apply the standard techniques.

3.2.1 Generalized Ghirlanda–Guerra Identities

To study the arrays in vector spin models, we need a form of the Ghirlanda–Guerra identities adapted to the matrix valued Gram arrays appearing in those models. We first show that there exists a Gaussian process with the appropriate covariance structure that will encode the moments of these matrix valued arrays.

The parameters of this covariance are $p, m, n_1, \ldots, n_m \geq 1$ and vectors $\vec{\lambda}^1, \ldots, \vec{\lambda}^m \in \mathbb{R}^n$. To simplify notation, let $\theta = (p, m, n_1, \ldots, n_m, \vec{\lambda}^1, \ldots, \vec{\lambda}^m)$ denote the list of parameters. We define

$$C^{\theta}_{\ell,\ell'} = \prod_{j \leq m} \left(\boldsymbol{R}^{\odot p}_{\ell,\ell'} \vec{\lambda}^j, \vec{\lambda}^j \right)$$

where $\boldsymbol{R}_{\ell,\ell'}$ is the overlap matrix in the vector spin models

$$\boldsymbol{R}_{\ell,\ell'} = \frac{1}{N} \sum_{i=1}^{N} \vec{\sigma}_i \otimes \vec{\sigma}_i. \tag{3.22}$$

These covariances form a basis for bounded measurable functions of the quadratic forms of overlaps. Without loss of generality, we will normalize the diagonal entries of this overlap matrix so that all entries are equal to 1. We also assume that the diagonals of this infinite array $\boldsymbol{R}_{\ell,\ell}$ are constant matrices for all $\ell \geq 1$. This will be the case (at least approximately) for the constrained vector spin models.

Lemma 3.2.1

For any choice of parameter θ, there exists a centered Gaussian process with covariance

$$\mathbb{E} h_\theta(\vec{\boldsymbol{\sigma}}^{\ell}) h_\theta(\vec{\boldsymbol{\sigma}}^{\ell'}) = C^{\theta}_{\ell,\ell'}. \tag{3.23}$$

Proof. We first show for any vector $\vec{\lambda} \in \mathbb{R}^n$ and $p \in \mathbb{N}$, $(\boldsymbol{R}^{\odot p}_{\ell,\ell'} \vec{\lambda}, \vec{\lambda})$ is positive semidefinite. Recall that

$$\boldsymbol{R}_{\ell,\ell'} = \frac{1}{N} \sum_{i=1}^{N} \vec{\sigma}^{\ell}_i \otimes \vec{\sigma}^{\ell'}_i.$$

By looking at each coordinate of the matrix, it is also easy to verify that

$$\boldsymbol{R}^{\odot p}_{\ell,\ell'} = \frac{1}{N^p} \sum_{i_1, \ldots, i_p} (\vec{\sigma}^{\ell}_{i_1} \odot \cdots \odot \vec{\sigma}^{\ell}_{i_p}) \otimes (\vec{\sigma}^{\ell'}_{i_1} \odot \cdots \odot \vec{\sigma}^{\ell'}_{i_p}).$$

For any $\vec{\lambda}$, we also have

$$\left((\vec{\sigma}^{\ell}_{i_1} \odot \cdots \odot \vec{\sigma}^{\ell}_{i_p}) \otimes (\vec{\sigma}^{\ell'}_{i_1} \odot \cdots \odot \vec{\sigma}^{\ell'}_{i_p}) \vec{\lambda}, \vec{\lambda} \right) = \left((\vec{\sigma}^{\ell}_{i_1} \odot \cdots \odot \vec{\sigma}^{\ell}_{i_p})^{\mathsf{T}} \vec{\lambda}, (\vec{\sigma}^{\ell'}_{i_1} \odot \cdots \odot \vec{\sigma}^{\ell'}_{i_p})^{\mathsf{T}} \vec{\lambda} \right)$$

so

$$(\boldsymbol{R}_{\ell,\ell'}^{\odot p}\vec{\lambda},\vec{\lambda}) = \frac{1}{N^p}\sum_{i_1,\dots,i_p}\Big((\vec{\sigma}_{i_1}^{\ell}\odot\cdots\odot\vec{\sigma}_{i_p}^{\ell})^{\mathsf{T}}\vec{\lambda},(\vec{\sigma}_{i_1}^{\ell'}\odot\cdots\odot\vec{\sigma}_{i_p}^{\ell'})^{\mathsf{T}}\vec{\lambda}\Big)$$

is positive semidefinite because it is the sum of Gram matrices. By the Schur product theorem [56, Theorem 7.5.3], it follows that

$$C_{\ell,\ell'}^{\theta} = \prod_{j\le m}(\boldsymbol{R}_{\ell,\ell'}^{\odot p}\vec{\lambda}_j,\vec{\lambda}_j)^{m_j}$$

is positive semidefinite because it is the Hadamard product of positive semidefinite matrices. □

If we restrict the vectors $\vec{\lambda}\in([-1,1]\cap\mathbb{Q})^n$, then the space of possible parameters for θ is countable. Let Θ denote the space of all possible parameters. We can find a bijective map $j(\theta):\Theta\to\mathbb{N}$ such that

$$\sum_{\theta\in\Theta}2^{-2j(\theta)}\cdot 9\cdot C_{1,2}^{\theta}\le 1,$$

since $C_{1,2}^{\theta}\le n^2$. Therefore, if we let h_θ be independent Gaussian processes with covariance given by (3.23) and u_θ be independent uniform random variables in $[1,2]$, the following perturbation Hamiltonian

$$h_N(\vec{\boldsymbol{\sigma}}) = \sum_{\theta\in\Theta}2^{-j(\theta)}u_\theta h_\theta(\vec{\boldsymbol{\sigma}}) \tag{3.24}$$

is almost surely a well defined Gaussian process with finite variance, because it is the linear combination of well defined Gaussian processes.

We define $H_N^{\mathrm{pert}} = H_N(\vec{\boldsymbol{\sigma}}) + s_N h_N(\vec{\boldsymbol{\sigma}})$, and let $\langle\cdot\rangle$ denote the average with respect to the perturbed measure. Let

$$\Delta(f,k,\theta) = \Big|\mathbb{E}\langle f(R^k)C_{1,k+1}^{\theta}\rangle - \frac{1}{k}\mathbb{E}\langle f(R^k)\rangle\mathbb{E}\langle C_{1,2}^{\theta}\rangle - \frac{1}{k}\sum_{\ell=2}^{k}\mathbb{E}\langle f(R^k)C_{1,\ell}^{\theta}\rangle\Big| \tag{3.25}$$

to denote the analogue of (2.37) to be the deviation from the Ghirlanda–Guerra identities defined in (2.25) for vector spin models. If we take s_N sufficiently large, then the generalized Ghirlanda–Guerra identities also hold.

Theorem 3.2.1 (*The Generalized Ghirlanda–Guerra Identities*)

If s_N satisfies

$$\lim_{N\to\infty}s_N=\infty \qquad \textit{and} \qquad \lim_{N\to\infty}\frac{v_N(s_N)}{s_N^2}=0, \tag{3.26}$$

then

$$\lim_{N\to\infty}\mathbb{E}_u\Delta(f,k,\theta)=0$$

for any $\theta\in\Theta$, $k\ge 2$ and any bounded measurable function $f(R^k)$. The average is with respect to the i.i.d. uniform random variables $u_\theta\in[1,2]$ appearing in the covariance (2.32).

Proof. The proof is identical to Theorem 2.3.1, with $R_{\ell,\ell'}^p$ replaced with $C_{\ell,\ell'}^{\theta}$. The key step in the proof is that in the constrained models, the self overlaps $C_{\ell,\ell}^{\theta}$ are constant, so the cancellation of the self overlaps in (2.43) when we substitute (2.41) and (2.42) also hold in this generalized setting. □

The covariances $C^{\theta}_{\ell,\ell'}$ form a basis for bounded measurable functions of quadratic forms, because the $C^{\theta}_{\ell,\ell'}$ are bounded and of the form of monomials of quadratic forms $(\boldsymbol{R}^{\odot p}_{\ell,\ell'}\vec{\lambda}^j, \vec{\lambda}^j)$. Bounded measurable functions can be approximated by linear combinations of monomials of its entries. Furthermore, the entries of the array $(\boldsymbol{R}_{\ell,\ell'})_{\ell,\ell'\geq 1}$ is bounded, so it is compact because it is a countable product of compact spaces. Therefore, the space of probability measures on these arrays is tight, so the selection theorem [82, Theorem 3.2] implies that there exists a subsequence such that the array $(\boldsymbol{R}_{\ell,\ell'})_{\ell,\ell'\geq 1}$ under $\mathbb{E}G_N^{\otimes\infty}$ converges in distribution.

If $\mathbb{E}$ is the average with respect to the limiting distribution of $(R_{\ell,\ell'})_{\ell,\ell'\geq 1}$, then the limiting array must satisfy

$$\mathbb{E}f\psi_{1,n+1} = \frac{1}{n}\mathbb{E}f\mathbb{E}\psi_{1,2} + \frac{1}{n}\sum_{\ell=2}^{n}\mathbb{E}f\psi_{1,\ell} \tag{3.27}$$

where

$$\psi_{\ell,\ell'} = \psi((\boldsymbol{R}^{\odot p}_{\ell,\ell'}\vec{\lambda}^1, \vec{\lambda}^1), \dots, (\boldsymbol{R}^{\odot p}_{\ell,\ell'}\vec{\lambda}^m, \vec{\lambda}^m))$$

and $\psi : \mathbb{R}^m \to \mathbb{R}$ is a bounded measurable function and f is a function of n overlaps. The same logic as in Section 2.3.3 can be adapted to show that for mixed p-spin models there exists a deterministic sequence of u_θ can be chosen so that the perturbed Gibbs measure satisfies the Ghirlanda–Guerra identities for all parameters θ simultaneously without spoiling the averages in the cavity computations.

In the next section, we will describe how these generalized Ghirlanda–Guerra identities (3.27) imply a powerful synchronization between the trace and the off-diagonal elements of the overlap matrices.

3.2.2 Vector Spin Synchronization

Matrix valued arrays satisfying the generalized Ghirlanda–Guerra identities (3.27), can be approximated by arrays generated from the Ruelle probability cascades. This synchronization property for vector spin models was discovered by Panchenko [81, 80] using a similar synchronization property for multi-species spin glass models [78].

The key consequence is the matrix entries are deterministic functions of the trace. The array of traces can be generated from the Ruelle probability cascades, so this connection allows us to recover the overlap matrix array using the tools in Chapter 2. Furthermore, these matrix valued arrays satisfy a much stronger positivity principle in the sense that the entries of the array must be positive semidefinite. This result is completely nontrivial, because generated by outer products (3.22), are not even symmetric. This consequence is essential because it will imply the Gaussian processes in the cavity computations will be well defined.

The synchronization is yet another consequence of the ultrametric structure of the limiting Gibbs distribution. We first state the main result.

Theorem 3.2.2 (*Synchronization of Overlap Matrices*)

Suppose a matrix valued array $(\boldsymbol{R}_{\ell,\ell'})_{\ell,\ell'\geq 1}$ satisfies (3.27) for all choices of parameters. There exists a function $\boldsymbol{\Phi}:\mathbb{R}^+\to\mathbb{S}_n^+$ such that

$$\boldsymbol{R}_{\ell,\ell'}=\boldsymbol{\Phi}(\mathrm{tr}(\boldsymbol{R}_{\ell,\ell'})) \tag{3.28}$$

almost surely. Furthermore, $\boldsymbol{\Phi}$ is nondecreasing (in the sense of the Loewner partial order)

$$x_1\leq x_2 \implies \boldsymbol{\Phi}(x_2)\leq\boldsymbol{\Phi}(x_1)$$

and Lipschitz continuous

$$\|\boldsymbol{\Phi}(x_2)-\boldsymbol{\Phi}(x_1)\|_1\leq L_n|x_2-x_1|$$

for some constant L_n that only depends on the dimension n.

Proof. The detailed proof of this result can be found in [81, Theorem 3]. Instead, we will sketch the main ideas in this proof and explain the intuition behind the argument.

Multispecies Synchronization: The first observation is Ghirlanda–Guerra identities imply a synchronization property satisfied by arrays of quadratic forms. This type of symmetry was also observed in multispecies spin glasses [78]. Given $p\geq 1$ and $\vec{\lambda}^1,\dots,\vec{\lambda}^m\in[-1,1]^n$, we define the following arrays

$$\Lambda^j_{\ell,\ell'}=(\boldsymbol{R}_{\ell,\ell'}\vec{\lambda}^j,\vec{\lambda}^j)\qquad\text{and}\qquad\Lambda_{\ell,\ell'}=\sum_{j=1}^m\Lambda^j_{\ell,\ell'}. \tag{3.29}$$

Ultrametricity implies that the arrays the sum of the arrays encodes all the information from the original arrays. This is a consequence of the fact that the Ghirlanda–Guerra identities imply that $\Lambda_{\ell,\ell'}$ is ultrametric in addition to $\Lambda^j_{\ell,\ell'}$. This joint ultrametric structure of the arrays implies a synchronization between the entries of the sums.

Lemma 3.2.2 (*Multispecies Synchronization*)

Suppose a matrix valued array $(\boldsymbol{R}_{\ell,\ell'})_{\ell,\ell'\geq 1}$ satisfies (3.27) for all choices of parameters.

1. *With probability one, if $\Lambda^j_{\ell,\ell'}>\Lambda^j_{\ell,\ell''}$ for some $j\leq m$, then $\Lambda^j_{\ell,\ell'}>\Lambda^j_{\ell,\ell''}$ for all $j\leq m$.*
2. *There exists a non-decreasing 1-Lipschitz function $L_j:\mathbb{R}^+\to\mathbb{R}^+$ for $j\leq m$ such that with probability one, $\Lambda^j_{\ell,\ell'}=L_j(\Lambda_{\ell,\ell'})$.*

The joint ultrametricity of the arrays states that the arrays $\Lambda^j_{\ell,\ell'}$ are synchronized in the sense that increases in one array, are simultaneously observed in all arrays. The second consequence states that the values of each individual array, are deterministic functions of its sum. The second result is very non-trivial because one would not expect to be able to recover the individual values of random variables after observing its sum. This is essentially the synchronization property for multispecies spin glasses, and it is discussed in [78, Theorem 4 and Lemma 2].

Trace Parametrization: If we take $p = 1$ and $\vec{\lambda}^j = e_j$, then (3.29)

$$\Lambda^j_{\ell,\ell'} = (R_{\ell,\ell'}\vec{\lambda}^j, \vec{\lambda}^j) = R^{j,j}_{\ell,\ell'} \qquad \text{and} \qquad \Lambda_{\ell,\ell'} = \sum_{j=1}^{m} \Lambda^j_{\ell,\ell'} = \operatorname{tr}(\boldsymbol{R}_{\ell,\ell'}). \tag{3.30}$$

The previous result implies the diagonal entries of the overlap arrays are encoded by its trace,

$$R^{j,j}_{\ell,\ell'} = L_k(\operatorname{tr}(\boldsymbol{R}_{\ell,\ell'})).$$

Next, to encode some information about the off diagonal elements, we take $\vec{\lambda}^1 = e_j + e_{j'}$ and $\vec{\lambda}^1 = e_j - e_{j'}$. This implies that

$$R^{k,k'}_{\ell,\ell'} + R^{k',k}_{\ell,\ell'} = L_{k,k'}(\operatorname{tr}(\boldsymbol{R}_{\ell,\ell'})).$$

We can also encode the sum of the squares by taking $p = 2$, to conclude that

$$(R^{k,k'}_{\ell,\ell'})^2 + (R^{k',k}_{\ell,\ell'})^2 = \tilde{L}_{k,k'}(\operatorname{tr}(\boldsymbol{R}_{\ell,\ell'})).$$

The Overlaps are Symmetric: The next step is to show that $R^{k,k'}_{\ell,\ell'} = R^{k',k}_{\ell,\ell'}$, so the arrays are almost surely symmetric. The results of the trace parametrization imply that

$$\begin{cases} R^{k,k'}_{\ell,\ell'} + R^{k',k}_{\ell,\ell'} = L_{k,k'}(\operatorname{tr}(\boldsymbol{R}_{\ell,\ell'})) \\ (R^{k,k'}_{\ell,\ell'})^2 + (R^{k',k}_{\ell,\ell'})^2 = \tilde{L}_{k,k'}(\operatorname{tr}(\boldsymbol{R}_{\ell,\ell'})) \end{cases} \tag{3.31}$$

Using the quadratic formula, this system can be solved which implies $R^{k,k'}_{\ell,\ell'}, R^{k',k}_{\ell,\ell'}$ can be encoded by at most a pair of functions $f_1(\operatorname{tr}(\boldsymbol{R}_{\ell,\ell'}))$ and $f_2((\operatorname{tr}(\boldsymbol{R}_{\ell,\ell'})))$. The key result is that both the functions must be the same, which implies $R^{k,k'}_{\ell,\ell'} = R^{k',k}_{\ell,\ell'}$. This result follows because the array of traces,

$$(T_{\ell,\ell'})_{\ell,\ell'\geq 1} = (\operatorname{tr}(\boldsymbol{R}_{\ell,\ell'}))_{\ell,\ell'\geq 1}$$

satisfy the classical Ghirlanda–Guerra identities (2.23), so they are generated by the Ruelle probability cascades by the arguments in Section 2.2.2.

Similarly to the proof of ultrametricity, the symmetry of the overlap matrices follows from the duplication property. By ultrametricity, it is possible to generate duplicates from the support of the distribution of $\operatorname{tr}(\boldsymbol{R}_{1,2})$. The duplication property implies that if a certain block matrix is in the support of the overlap array, then we can take these block matrices arbitrary large by duplicating configurations that generate these block matrices. Examining the barycenters of these configurations will imply that the block matrices must be symmetric.

The Overlaps are Positive Semidefinite: To show that $\boldsymbol{\Phi}$ is increasing, it suffices to show that

$$(\boldsymbol{\Phi}(x_1)\vec{\lambda}, \vec{\lambda}) \leq (\boldsymbol{\Phi}(x_2)\vec{\lambda}, \vec{\lambda})$$

for all $\vec{\lambda} \in [-1, 1]$ whenever $x_2 - x_1 > 0$. Furthermore, it is automatic that $\boldsymbol{\Phi}(0) = \mathbf{0}$, so this will also imply that the matrices are positive semidefinite for free. This monotonicity property for all quadratic forms is a consequence of the first consequence of the multispecies synchronization Lemma 3.2.2. □

3.3 Matrix Analysis

The analysis of vector spin models is a bit more complicated, because the single variable functionals in the Parisi formulas for classical models are replaced by functionals of matrix paths. To do calculus of variations on these functionals, one needs to be comfortable with several facts about matrices. We summarize the results here, because they will be useful moving forward. A good reference for these results are [56] and [87].

3.3.1 Matrix Directional Derivatives

We will be studying the critical points of functions defined with respect to matrix paths. To study these critical point, we will need some formulas for the directional derivatives of certain matrix functions appearing in the variational formulas. In this section, we state several facts about matrices that will be used in the derivation of the Crisanti–Sommers formula. Let $\mathbb{S}_+^n$ be the space of symmetric positive semidefinite real valued $n \times n$ matrices.

Let $\boldsymbol{C}$ be an arbitrary symmetric matrix we use the following notation to denote the matrix derivatives of various functions in the direction $\boldsymbol{C}$. Let $f : \mathbb{S}_+^n \to \mathbb{R}$, we define

$$\frac{d}{d\boldsymbol{A}} f(\boldsymbol{A}) := \frac{d}{dt} f(\boldsymbol{A} + t\boldsymbol{C})\Big|_{t=0} \in \mathbb{R}.$$

We summarize several matrix derivatives that are used to compute the partial derivative of the functional in this thesis. Let $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \operatorname{tr}(\boldsymbol{A}\boldsymbol{B})$ denote the Frobenius inner product on symmetric matrices. We have

1.
$$\frac{\partial}{\partial \boldsymbol{A}} \langle \boldsymbol{B}, \boldsymbol{A} \rangle = \frac{d}{dt} \operatorname{tr}(\boldsymbol{B}(\boldsymbol{A} + t\boldsymbol{C}))\Big|_{t=0} = \langle \boldsymbol{B}, \boldsymbol{C} \rangle. \tag{3.32}$$

2.
$$\frac{\partial}{\partial \boldsymbol{A}} \operatorname{Sum}(\boldsymbol{\xi}(\boldsymbol{A})) = \frac{d}{dt} \operatorname{Sum}(\boldsymbol{\xi}(\boldsymbol{A} + t\boldsymbol{C}))\Big|_{t=0} = \langle \boldsymbol{\xi}'(\boldsymbol{A}), \boldsymbol{C} \rangle. \tag{3.33}$$

3.
$$\frac{\partial}{\partial \boldsymbol{A}} \log |\boldsymbol{A}| = \frac{d}{dt} \log |\boldsymbol{A} + t\boldsymbol{C}|\Big|_{t=0} = \langle \boldsymbol{A}^{-1}, \boldsymbol{C} \rangle. \tag{3.34}$$

4.
$$\frac{\partial}{\partial \boldsymbol{A}} \langle \boldsymbol{A}^{-1}, \boldsymbol{B} \rangle = \frac{d}{dt} \operatorname{tr}(\boldsymbol{B}(\boldsymbol{A} + t\boldsymbol{C})^{-1})\Big|_{t=0} = -\langle \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}, \boldsymbol{C} \rangle. \tag{3.35}$$

Before proving these derivatives, we write several basic operations in terms of the Frobenius inner product for symmetric matrices,

1. Let $\boldsymbol{A}$ be a $n \times n$ matrix, and let $\vec{h} \in \mathbb{R}^n$, we have

$$(\vec{h}, \boldsymbol{A}\vec{h}) = \vec{h}^{\mathsf{T}} \boldsymbol{A} \vec{h} = \operatorname{tr}(\vec{h}\vec{h}^{\mathsf{T}} \boldsymbol{A}) = \langle \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{A} \rangle. \tag{3.36}$$

2. Let $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ be $n \times n$ matrices, we have

$$\langle \boldsymbol{A} \odot \boldsymbol{B}, \boldsymbol{C} \rangle = \operatorname{tr}((\boldsymbol{A} \odot \boldsymbol{B}) \times \boldsymbol{C}) = \operatorname{tr}(\boldsymbol{A} \times (\boldsymbol{B} \odot \boldsymbol{C})) = \langle \boldsymbol{A}, \boldsymbol{B} \odot \boldsymbol{C} \rangle. \tag{3.37}$$

3. Let $\mathbf{1}$ be the $n \times n$ matrix with all 1's, as a consequence of the above fact, we have

$$\mathrm{Sum}(\boldsymbol{A} \odot \boldsymbol{B}) = \mathrm{tr}(\mathbf{1} \times (\boldsymbol{A} \odot \boldsymbol{B})) = \mathrm{tr}((\mathbf{1} \odot \boldsymbol{A}) \times \boldsymbol{B})) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{B}) = \langle \boldsymbol{A}, \boldsymbol{B} \rangle. \tag{3.38}$$

We now compute the directional derivatives.

Proposition 3.3.1 (*Derivative of the Trace*)

For any matrix $\boldsymbol{B}$, the directional derivative of the trace in direction $\boldsymbol{C}$ is given by

$$\frac{\partial}{\partial \boldsymbol{A}} \mathrm{tr}(\boldsymbol{B}\boldsymbol{A}) = \frac{d}{dt} \mathrm{tr}(\boldsymbol{B} \times (\boldsymbol{A} + t\boldsymbol{C}))\Big|_{t=0} = \mathrm{tr}(\boldsymbol{B}\boldsymbol{C}).$$

Proof. By linearity, we have

$$\frac{d}{dt} \mathrm{tr}(\boldsymbol{B} \times (\boldsymbol{A} + t\boldsymbol{C}))\Big|_{t=0} = \frac{d}{dt} \mathrm{tr}(\boldsymbol{B}\boldsymbol{A}) + t\,\mathrm{tr}(\boldsymbol{B}\boldsymbol{C})\Big|_{t=0} = \mathrm{tr}(\boldsymbol{B}\boldsymbol{C}).$$

□

This immediately implies the directional derivative of a quadratic form.

Proposition 3.3.2 (*Derivative of quadratic form*)

For $\vec{h} \in \mathbb{R}^n$, the directional derivative of the quadratic form in direction $\boldsymbol{C}$ is given by

$$\frac{\partial}{\partial \boldsymbol{A}} (\vec{h}, \boldsymbol{A}\vec{h}) = \frac{d}{dt} (\vec{h}, (\boldsymbol{A} + t\boldsymbol{C})\vec{h})\Big|_{t=0} = \mathrm{tr}(\vec{h}\vec{h}^{\mathsf{T}}\boldsymbol{C}).$$

Proof. We can write the quadratic form as

$$(\vec{h}, \boldsymbol{A}\vec{h}) = \mathrm{tr}(\vec{h}\vec{h}^{\mathsf{T}}\boldsymbol{A}).$$

The property follows immediately from Proposition 3.3.1. □

Next, we compute the matrix derivative with respect to a smooth function $\boldsymbol{G}(\boldsymbol{A})$, where the each coordinate of $\boldsymbol{G}_{ij}(\boldsymbol{A})$ is a smooth function single variable function of $\boldsymbol{A}_{ij}$.

Proposition 3.3.3 (*Derivatives of Matrix Valued Functions*)

Suppose each coordinate of $G(\boldsymbol{A})$ only depends on the corresponding coordinate of $\boldsymbol{A}$. If $\boldsymbol{A}$ is positive semidefinite, then

$$\frac{\partial}{\partial \boldsymbol{A}} \mathrm{Sum}(\boldsymbol{G}(\boldsymbol{A})) = \frac{d}{dt} \mathrm{Sum}(\boldsymbol{G}(\boldsymbol{A} + t\boldsymbol{C}))\Big|_{t=0} = \mathrm{tr}(\boldsymbol{G}'(\boldsymbol{A})\boldsymbol{C}).$$

Proof. We first write

$$\mathrm{Sum}(\boldsymbol{G}(\boldsymbol{A})) = \mathrm{tr}(\mathbf{1} \times \boldsymbol{G}(\boldsymbol{A})).$$

Therefore, by Proposition 3.3.1,

$$\frac{d}{dt}\operatorname{tr}(\mathbf{1}\times \boldsymbol{G}(\boldsymbol{A}+t\boldsymbol{C}))\Big|_{t=0} = \operatorname{tr}(\mathbf{1}\times \frac{d}{dt}\boldsymbol{G}(\boldsymbol{A}+t\boldsymbol{C}))\Big|_{t=0} = \operatorname{tr}(\mathbf{1}\times \boldsymbol{G}'(\boldsymbol{A})\odot \boldsymbol{C}) = \operatorname{tr}(\boldsymbol{G}'(\boldsymbol{A})\boldsymbol{C}).$$

We used the fact that each coordinate of $\boldsymbol{G}(\boldsymbol{A})$ only depends on the corresponding coordinate of $\boldsymbol{A}$, so the matrices can be differentiated term by term. □

We now compute the derivative of the log determinant.

Proposition 3.3.4 (*Derivative of Log Determinant*)

For any matrix $\boldsymbol{A} > 0$, the directional derivative of the log determinant in direction $\boldsymbol{C}$ is given by

$$\frac{\partial}{\partial \boldsymbol{A}}\log|\boldsymbol{A}| = \frac{d}{dt}\log|\boldsymbol{A}+t\boldsymbol{C}|\Big|_{t=0} = \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{C}).$$

Proof. By the chain rule, we have

$$\frac{d}{dt}\log|\boldsymbol{A}+t\boldsymbol{C}|\Big|_{t=0} = \frac{1}{|\boldsymbol{A}+t\boldsymbol{C}|}\frac{d}{dt}|\boldsymbol{A}+t\boldsymbol{C}|\Big|_{t=0}.$$

By Jacobi's formula for invertible matrices, we have

$$\frac{d}{dt}|\boldsymbol{A}+t\boldsymbol{C}| = \operatorname{tr}(|\boldsymbol{A}+t\boldsymbol{C}|(\boldsymbol{A}+t\boldsymbol{C})^{-1}\boldsymbol{C}) = |\boldsymbol{A}+t\boldsymbol{C}|\operatorname{tr}((\boldsymbol{A}+t\boldsymbol{C})^{-1}\boldsymbol{C}),$$

and therefore

$$\frac{d}{dt}\log|\boldsymbol{A}+t\boldsymbol{C}|\Big|_{t=0} = \frac{|\boldsymbol{A}+t\boldsymbol{C}|}{|\boldsymbol{A}+t\boldsymbol{C}|}\operatorname{tr}((\boldsymbol{A}+t\boldsymbol{C})^{-1}\boldsymbol{C})\Big|_{t=0} = \operatorname{tr}(A^{-1}\boldsymbol{C}).$$

□

Lastly, we compute the matrix derivative of the inverse

Proposition 3.3.5 (*Derivative of Inverse*)

For any matrix $\boldsymbol{B}$ and $\boldsymbol{A} > 0$, the directional derivative of the inverse of $\boldsymbol{A}$ in direction $\boldsymbol{C}$ is given by

$$\frac{\partial}{\partial \boldsymbol{A}}\operatorname{tr}(\boldsymbol{B}\boldsymbol{A}^{-1}) = \frac{d}{dt}\operatorname{tr}(\boldsymbol{B}\times(\boldsymbol{A}+t\boldsymbol{C})^{-1})\Big|_{t=0} = -\operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}) = -\operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}).$$

Proof. By definition, we have

$$\frac{d}{dt}\operatorname{tr}(\boldsymbol{B}\times(\boldsymbol{A}+t\boldsymbol{C})^{-1})\Big|_{t=0} = \lim_{t\to 0}\frac{\operatorname{tr}(\boldsymbol{B}(\boldsymbol{A}+t\boldsymbol{C})^{-1}) - \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}^{-1})}{t} = \lim_{t\to 0}\operatorname{tr}\left(\frac{\boldsymbol{B}(\boldsymbol{A}+t\boldsymbol{C})^{-1} - \boldsymbol{B}\boldsymbol{A}^{-1}}{t}\right)$$

which simplifies to

$$\lim_{t\to 0}\operatorname{tr}\left(\boldsymbol{B}(\boldsymbol{A}+t\boldsymbol{C})^{-1}\frac{\boldsymbol{A}-(\boldsymbol{A}+t\boldsymbol{C})}{t}\boldsymbol{A}^{-1}\right) = -\operatorname{tr}(\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}) = -\operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}).$$

Since $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) = \operatorname{tr}(\boldsymbol{C}\boldsymbol{A}\boldsymbol{B})$, the derivative is also equal to $-\operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})$ □

Proposition 3.3.6 (*Quadratic Form of Inverse*)

For any vectors $\vec{h} \in \mathbb{R}^n$ and $\boldsymbol{A} > 0$, the directional derivative of the trace in direction $\boldsymbol{C}$ is given by

$$\frac{\partial}{\partial \boldsymbol{A}}(\vec{h}, \boldsymbol{A}^{-1}\vec{h}) = \frac{d}{dt}(\vec{h}, (\boldsymbol{A} + t\boldsymbol{C})^{-1}\vec{h})\Big|_{t=0} = \operatorname{tr}(\boldsymbol{A}^{-1}\vec{h}\vec{h}^{\mathsf{T}}\boldsymbol{A}^{-1}\boldsymbol{C}).$$

Proof. We can write the quadratic form as

$$(\vec{h}, \boldsymbol{A}\vec{h}) = \operatorname{tr}(\vec{h}\vec{h}^{\mathsf{T}}\boldsymbol{A}).$$

The property follows immediately from Proposition 3.3.5. □

Critical Point Conditions

The directional derivatives will be used to find critical point conditions satisfied by the the matrix paths. If $\boldsymbol{A}$ is an interior minimizer of f, then

$$\frac{d}{dt} f(\boldsymbol{A} + t\boldsymbol{C})\Big|_{t=0} = 0$$

for all directions $\boldsymbol{C}$. This is because both $\boldsymbol{C}$ and $-\boldsymbol{C}$ are admissible variations at an interior point. The following results will be used to derive the matrix equalities in the critical point conditions.

Proposition 3.3.7

If $\operatorname{tr}(\boldsymbol{AC}) \leq \operatorname{tr}(\boldsymbol{BC})$ for all symmetric matrices $\boldsymbol{C}$ then $\boldsymbol{A} \leq \boldsymbol{B}$.

Proof. Suppose that $\operatorname{tr}(\boldsymbol{AC}) \leq \operatorname{tr}(\boldsymbol{BC})$ for all symmetric matrices $\boldsymbol{C}$ but $\boldsymbol{B} - \boldsymbol{A}$ is not positive semidefinite. That is, there exists a vector v such that

$$v^{\mathsf{T}}(\boldsymbol{B} - \boldsymbol{A})v < 0.$$

Consider the symmetric matrix $\boldsymbol{C} = vv^{\mathsf{T}}$. Therefore, our assumption implies,

$$0 \leq \operatorname{tr}((\boldsymbol{B} - \boldsymbol{A})\boldsymbol{C}) = \operatorname{tr}((\boldsymbol{B} - \boldsymbol{A})vv^{\mathsf{T}}) = \operatorname{tr}(v^T(\boldsymbol{B} - \boldsymbol{A})v) < 0$$

which is a contradiction. Therefore, $\boldsymbol{B} - \boldsymbol{A}$ must be positive semidefinite. □

This implies the following two sided version of the result,

Proposition 3.3.8

If $\operatorname{tr}(\boldsymbol{AC}) = \operatorname{tr}(\boldsymbol{BC})$ for all symmetric matrices $\boldsymbol{C}$, then $\boldsymbol{A} = \boldsymbol{B}$.

Proof. If $\operatorname{tr}(\boldsymbol{AC}) = \operatorname{tr}(\boldsymbol{BC})$, then by Proposition 3.3.7,

$$\operatorname{tr}(\boldsymbol{AC}) \leq \operatorname{tr}(\boldsymbol{BC}) \implies \boldsymbol{A} \leq \boldsymbol{B} \qquad \text{and} \qquad \operatorname{tr}(\boldsymbol{BC}) \leq \operatorname{tr}(\boldsymbol{AC}) \implies \boldsymbol{B} \leq \boldsymbol{A}.$$

Therefore, $\boldsymbol{A} - \boldsymbol{B}$ is both positive semidefinite and negative semidefinite, so all of its eigenvalues are 0. This implies that $\boldsymbol{A} = \boldsymbol{B}$. □

3.3.2 Positive Semidefinite Matrices

We also summarize several elementary facts about positive semidefinite matrices that we will use throughout the thesis. We begin by stating facts about the convexity and concavity of functionals on the space of positive semidefinite matrices.

Convexity of Matrix Functionals

These convexity arguments will be used to prove inequalities involving the vector spin functionals.

Proposition 3.3.9 (*Log determinants are concave*)

Let $\boldsymbol{A} > 0$ and suppose $\boldsymbol{C}$ is a symmetric matrix. The function

$$f(x) = \log|\boldsymbol{A} + x\boldsymbol{C}|$$

is concave in its domain. In particular, we have

$$\log|\boldsymbol{A}| + \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{C}) \geq \log|\boldsymbol{A} + \boldsymbol{C}|$$

for all $\boldsymbol{C}$ such that $|\boldsymbol{A} + \boldsymbol{C}| > 0$.

Proof. It suffices to show $f''(x) \leq 0$ whenever $\boldsymbol{A} + x\boldsymbol{C} > 0$. These derivatives are the first and second directional derivatives of $\log|\boldsymbol{A} + x\boldsymbol{C}|$ in the direction $\boldsymbol{C}$,

$$f'(x) = \operatorname{tr}((\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}) \qquad f''(x) = -\operatorname{tr}((\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}(\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}).$$

Using the eigendecomposition of $(\boldsymbol{A}+x\boldsymbol{C})^{-1}$, we can express it as $(\boldsymbol{A}+x\boldsymbol{C})^{-1} = \boldsymbol{B}\boldsymbol{B}^{\mathsf{T}}$ for some matrix $\boldsymbol{B}$. Therefore,

$$\begin{aligned}\operatorname{tr}((\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}(\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}) &= \operatorname{tr}(\boldsymbol{C}(\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}(\boldsymbol{A} + x\boldsymbol{C})^{-1}) \\ &= \operatorname{tr}(\boldsymbol{C}(\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}\boldsymbol{B}\boldsymbol{B}^{\mathsf{T}}) \\ &= \operatorname{tr}((\boldsymbol{C}\boldsymbol{B})^{\mathsf{T}}(\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}\boldsymbol{B}).\end{aligned}$$

It is easy to see that $(\boldsymbol{C}\boldsymbol{B})^{\mathsf{T}}(\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}\boldsymbol{B}$ is a positive definite matrix because

$$v^{\mathsf{T}}(\boldsymbol{C}\boldsymbol{B})^{\mathsf{T}}(\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}\boldsymbol{B}v = (\boldsymbol{C}\boldsymbol{B}v)^{\mathsf{T}}(\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}\boldsymbol{B}v > 0$$

for all $v \in \mathbb{R}^n$ since $(\boldsymbol{A} + x\boldsymbol{C})^{-1}$ is positive definite. Therefore, the sum of its eigenvalues are positive, so

$$\operatorname{tr}((\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}(\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}) = \operatorname{tr}((\boldsymbol{C}\boldsymbol{B})^{\mathsf{T}}(\boldsymbol{A} + x\boldsymbol{C})^{-1}\boldsymbol{C}\boldsymbol{B}) > 0.$$

Since $f(x)$ is concave, it lies below its tangent lines so

$$f(t) \leq f(0) + f'(0)t \text{ for all } t \text{ in the domain of } f(t).$$

If $|\boldsymbol{A} + \boldsymbol{C}| > 0$ then we can take $t = 1$ to conclude

$$\log|\boldsymbol{A}| + \operatorname{tr}(\boldsymbol{A}^{-1}\boldsymbol{C}) \geq \log|\boldsymbol{A} + \boldsymbol{C}|.$$

□

The convexity of the covariances of vector spin Hamiltonians is a straightforward consequence of the convexity of the covariance functions in one dimension.

Proposition 3.3.10 (Sum($\boldsymbol{\xi}(\boldsymbol{A})$) *is convex*)

Suppose $\boldsymbol{C}$ is a symmetric matrix. The function

$$f(x) = \mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{A} + x\boldsymbol{C}))$$

is convex. In particular, we have

$$\mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{A})) + \mathrm{tr}(\boldsymbol{\xi}'(\boldsymbol{A})\boldsymbol{C}) \leq \mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{A} + \boldsymbol{C}))$$

for all $\boldsymbol{C}$.

Proof. Since β_p are positive and $\beta_p = 0$ for all odd p, $\boldsymbol{\xi}(\boldsymbol{A})$ is a convex function in each of its coordinates. Since the finite sum of convex functions are convex,

$$f(x) = \mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{A} + x\boldsymbol{C}))$$

is convex. Since each entry of $\boldsymbol{\xi}$ is a convex function, we have

$$\boldsymbol{\xi}(\boldsymbol{A})_{ij} + \boldsymbol{\xi}'(\boldsymbol{A})_{ij}\boldsymbol{C}_{ij} \leq \boldsymbol{\xi}(\boldsymbol{A} + \boldsymbol{C})_{ij}.$$

Summing over $i, j \leq n$ implies

$$\mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{A})) + \mathrm{tr}(\boldsymbol{\xi}'(\boldsymbol{A})\boldsymbol{C}) \leq \mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{A} + \boldsymbol{C})).$$

□

The last consequence is about the convexity of matrix inverses.

Proposition 3.3.11 (*Inverse Matrices are Convex*)

Let $\boldsymbol{A}$ be a positive definite matrix, and suppose that $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{S}_+^n$ also satisfy $(\boldsymbol{A} - \boldsymbol{B})^{-1} > 0$ and $(\boldsymbol{A} - \boldsymbol{C})^{-1} > 0$. Then inverting matrices are convex,

$$(\boldsymbol{A} - (\varepsilon\boldsymbol{B} + (1-\varepsilon)\boldsymbol{C}))^{-1} \leq \varepsilon(\boldsymbol{A} - \boldsymbol{B})^{-1} + (1-\varepsilon)(\boldsymbol{A} - \boldsymbol{C})^{-1}.$$

Proof. Let $v \in \mathbb{R}^n$ and $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}_+^n$. It suffices to show that

$$f(t) = \mathrm{tr}((\boldsymbol{A} - t\boldsymbol{B})^{-1}vv^{\mathsf{T}})$$

satisfies $f''(t) \geq 0$ for all t in the domain. By the chain rule and Proposition 3.3.5,

$$f'(t) = \mathrm{tr}((\boldsymbol{A} - t\boldsymbol{B})^{-1}\boldsymbol{B}(\boldsymbol{A} - t\boldsymbol{B})^{-1}vv^{\mathsf{T}})$$

and

$$f''(t) = 2\,\mathrm{tr}((\boldsymbol{A} - t\boldsymbol{B})^{-1}\boldsymbol{B}(\boldsymbol{A} - t\boldsymbol{B})^{-1}\boldsymbol{B}(\boldsymbol{A} - t\boldsymbol{B})^{-1}vv^{\mathsf{T}}).$$

Since $(\boldsymbol{A} - t\boldsymbol{B})^{-1}$ is positive definite, we have

$$f''(t) = 2\,\mathrm{tr}((\boldsymbol{B}(\boldsymbol{A} - t\boldsymbol{B})^{-1}v)^{\mathsf{T}}(\boldsymbol{A} - t\boldsymbol{B})^{-1}(\boldsymbol{B}(\boldsymbol{A} - t\boldsymbol{B})^{-1}v)) > 0.$$

The result in the proposition follows immediately. Let $v \in \mathbb{R}^n$, $\boldsymbol{A}$ be a positive definite matrix, and suppose that $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{S}^n_+$ also satisfy $(\boldsymbol{A} - \boldsymbol{B})^{-1} > 0$ and $(\boldsymbol{A} - \boldsymbol{C})^{-1} > 0$. We have shown that

$$g(t) = \mathrm{tr}((\boldsymbol{A} - \boldsymbol{C} - t(\boldsymbol{B} - \boldsymbol{C}))^{-1}vv^{\mathsf{T}}) = v^{\mathsf{T}}(\boldsymbol{A} - \boldsymbol{C} - t(\boldsymbol{B} - \boldsymbol{C}))^{-1}v$$

is a convex function in for $t \in [0, 1]$ so

$$v^{\mathsf{T}}(\boldsymbol{A} - (\varepsilon\boldsymbol{B} + (1-\varepsilon)\boldsymbol{C}))^{-1}v = g(\varepsilon) \le (1-\varepsilon)g(0) + \varepsilon g(1) = \varepsilon x^{\mathsf{T}}(\boldsymbol{A} - \boldsymbol{B})^{-1}v + (1-\varepsilon)v^{\mathsf{T}}(\boldsymbol{A} - \boldsymbol{C})^{-1}v.$$

This holds for all $v \in \mathbb{R}^n$, so

$$(\boldsymbol{A} - (\varepsilon\boldsymbol{B} + (1-\varepsilon)\boldsymbol{C}))^{-1} \le \varepsilon(\boldsymbol{A} - \boldsymbol{B})^{-1} + (1-\varepsilon)(\boldsymbol{A} - \boldsymbol{C})^{-1}.$$

□

Positive Definite Matrices

Positive definite matrices have some nice additional properties. Positive definite matrices represent the interior points of the positive semidefinite cone, so perturbations of these matrices always preserve positive definiteness.

We first state a straightforward relationship between the determinant and traces of positive definite matrices. It will be convenient to bound the determinants with the trace in our analysis. This is because the matrices are encoded by the deterministic functions of the trace after applying vector spin synchronizations.

Proposition 3.3.12 (*Upper Bound on the Determinant*)

If $\boldsymbol{A}$ is a positive definite, then

$$|\boldsymbol{A}| \le \Big(\frac{\mathrm{tr}(\boldsymbol{A})}{n}\Big)^n.$$

Proof. Since $\boldsymbol{A}$ is positive definite, its eigenvalues $\lambda_1, \ldots, \lambda_n$ are positive. Therefore, by the AM–GM inequality,

$$|\boldsymbol{A}|^{1/n} = \Big(\prod_{i=1}^n \lambda_j\Big)^{1/n} \le \frac{\sum_{i=1}^n \lambda_i}{n} = \frac{\mathrm{tr}(\boldsymbol{A})}{n}.$$

□

If the matrix path is strictly increasing, then all sufficiently small perturbations of the path will not violate monotonicity. This key fact is the main reason for introducing a logarithmic barrier in the derivation of the Cristanti–Sommers functionals as we will see later.

Proposition 3.3.13 (*Admissible Perturbations of Positive Definite Matrices*)

If $\boldsymbol{A}$ is a positive definite, then for all symmetric matrices $\boldsymbol{C}$, there exists a ε^* such that

$$\boldsymbol{A} + \varepsilon \boldsymbol{C}$$

is also positive definite for all $\varepsilon < \varepsilon^*$.

Proof. We will show that $v^\mathsf{T}(\boldsymbol{A}+\varepsilon\boldsymbol{C})v > 0$ for all $v \in \mathbb{R}^n$ and all ε sufficiently small. Since $\boldsymbol{A}$ is positive definite, we have

$$v^\mathsf{T}(\boldsymbol{A} + \varepsilon \boldsymbol{C})v \geq \lambda_{min}(\boldsymbol{A})\|v\|^2 - \varepsilon\|\boldsymbol{C}\|_\infty\|v\|^2 = (\lambda_{min}(\boldsymbol{A}) - \varepsilon\|\boldsymbol{C}\|_\infty)\|v\|^2.$$

Since $\lambda_{min}(\boldsymbol{A}) > 0$, setting $\varepsilon < \frac{\lambda_{min}(\boldsymbol{A})}{\|\boldsymbol{C}\|_\infty}$ guarantees $v^\mathsf{T}(\boldsymbol{A} + \varepsilon\boldsymbol{C})v > 0$. □

The following fact will imply that the covariances in the vector spin models are well defined.

Proposition 3.3.14 (*Hadamard Product of Positive Definite Matrices*)

If for each $j \leq n$, there exists a $p \geq 2$ such that $\beta_p(j) \neq 0$, and both $|\boldsymbol{Q}_\ell| > 0$ and $|\boldsymbol{Q}_\ell - \boldsymbol{Q}_{\ell-1}| > 0$, then $|\boldsymbol{\xi}'(\boldsymbol{Q}_\ell) - \boldsymbol{\xi}'(\boldsymbol{Q}_{\ell-1})| > 0$.

Proof. First recall the Hadamard product representation of $\boldsymbol{\xi}'(\boldsymbol{Q})$,

$$\boldsymbol{\xi}'(\boldsymbol{Q}) = \sum_{p\geq 2} p(\beta_p\beta_p^\mathsf{T}) \odot (\boldsymbol{Q})^{\odot(p-1)}.$$

Using the difference of powers formula to factor term by term, we have

$$\begin{aligned}\boldsymbol{\xi}'(\boldsymbol{Q}_\ell) - \boldsymbol{\xi}'(\boldsymbol{Q}_{\ell-1}) &= \sum_{p\geq 2} p(\beta_p\beta_p^\mathsf{T}) \odot \left(\boldsymbol{Q}_\ell^{\odot(p-1)} - \boldsymbol{Q}_{\ell-1}^{\odot(p-1)}\right) \\ &= (\boldsymbol{Q}_\ell - \boldsymbol{Q}_{\ell-1}) \odot \sum_{p\geq 2} p(\beta_p\beta_p^\mathsf{T}) \odot \sum_{0\leq k\leq p-2} \boldsymbol{Q}_\ell^{\odot(p-2)-k} \odot \boldsymbol{Q}_{\ell-1}^{\odot k}.\end{aligned}$$

By the Schur product theorem, the above is the Hadamard product of positive semidefinite matrices, so it must be positive semidefinite. By our assumption on β_p, there exists a M sufficiently large such that all entries of $\sum_{2\leq p\leq M} p(\beta_p\beta_p^\mathsf{T})$ are positive. Therefore,

$$\boldsymbol{\xi}'(\boldsymbol{Q}_\ell) - \boldsymbol{\xi}'(\boldsymbol{Q}_{\ell-1}) = (\boldsymbol{Q}_\ell - \boldsymbol{Q}_{\ell-1}) \odot \sum_{2\leq p\leq M} p(\beta_p\beta_p^\mathsf{T}) \odot \sum_{0\leq k\leq p-2} \boldsymbol{Q}_\ell^{\odot(p-2)-k} \odot \boldsymbol{Q}_{\ell-1}^{\odot k} \tag{3.39}$$

$$+ (\boldsymbol{Q}_\ell - \boldsymbol{Q}_{\ell-1}) \odot \sum_{p>M} p(\beta_p\beta_p^\mathsf{T}) \odot \sum_{0\leq k\leq p-2} \boldsymbol{Q}_\ell^{\odot(p-2)-k} \odot \boldsymbol{Q}_{\ell-1}^{\odot k}. \tag{3.40}$$

It suffices to show that the first matrix term (3.39) is positive definite, because the second matrix term (3.40) is positive semidefinite (the product of positive semidefinite matrices), and the sum of a positive definite matrix and a positive semidefinite matrix is positive definite. To prove this fact, we recall Oppenheim's inequality [56, Theorem 7.8.16],which states for all positive semidefinite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,

$$\det(\boldsymbol{A} \odot \boldsymbol{B}) \geq \det(\boldsymbol{A}) \prod_{j\leq n} \boldsymbol{B}_{jj}.$$

Therefore, we have the determinant of (3.39) is bounded below by

$$|\boldsymbol{Q}_\ell - \boldsymbol{Q}_{\ell-1}| \cdot \prod_{i\le n} \left(\sum_{2\le p\le M} p(\beta_p \beta_p^\mathsf{T}) \odot \sum_{0\le k\le p-2} \boldsymbol{Q}_\ell^{\odot(p-2)-k} \odot \boldsymbol{Q}_{\ell-1}^{\odot k} \right)_{ii}. \tag{3.41}$$

We claim that each diagonal element appearing above is strictly positive. Since $\boldsymbol{Q}_\ell > 0$, for each $p \ge 2$, we have

$$\boldsymbol{A}_p := \sum_{0\le k\le p-2} \boldsymbol{Q}_\ell^{\odot(p-2)-k} \odot \boldsymbol{Q}_{\ell-1}^{\odot k} = \boldsymbol{Q}_\ell^{\odot(p-2)} + \sum_{1\le k\le p-2} \boldsymbol{Q}_\ell^{\odot(p-2)-k} \odot \boldsymbol{Q}_{\ell-1}^{\odot k} > 0$$

since the first term is the Hadamard power of a positive definite matrix and hence positive definite by the Schur product theorem. Since the diagonal elements of a positive definite matrix are all strictly positive, we have for all $i \le n$,

$$\left(\sum_{2\le p\le M} p(\beta_p \beta_p^\mathsf{T}) \odot \boldsymbol{A}_p \right)_{ii} \ge \left(\sum_{2\le p\le M} p(\beta_p \beta_p^\mathsf{T}) \right)_{ii} \cdot \min_{2\le p\le M} (\boldsymbol{A}_p)_{ii} > 0.$$

Substituting this fact into (3.41) and using the fact $|\boldsymbol{Q}_\ell - \boldsymbol{Q}_{\ell-1}| > 0$ implies (3.39) is positive definite, so the superadditivity of the determinant for positive semidefinite matrices implies

$$|\boldsymbol{\xi}'(\boldsymbol{Q}_\ell) - \boldsymbol{\xi}'(\boldsymbol{Q}_{\ell-1})| > 0.$$

as required. □

The monotonicity property is useful for finding conditions satisfied by the optimal matrix path.

Proposition 3.3.15 (*Monotonicity of Products*)

If $\boldsymbol{A}, \boldsymbol{C} \in \mathbb{S}_+^n$, then

$$\operatorname{tr}(\boldsymbol{AC}) \ge 0.$$

In particular, if $\boldsymbol{B} \ge \boldsymbol{A}$, then

$$\operatorname{tr}(\boldsymbol{BC}) \ge \operatorname{tr}(\boldsymbol{AC}).$$

Proof. Consider the eigendecomposition $\boldsymbol{C} = \boldsymbol{R}\boldsymbol{\Lambda}\boldsymbol{R}^\mathsf{T}$, where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ is the diagonal matrix of eigenvalues and $\boldsymbol{R}$ is an orthogonal matrix. Since $\boldsymbol{\Lambda}$ is diagonal, we can write it as a sum of real valued vectors $v_1, \dots, v_n$,

$$\boldsymbol{\Lambda} = \sum_{i=1}^n v_i v_i^\mathsf{T},$$

where $v_i = \sqrt{\lambda_i} e_i$. Since $\operatorname{tr}(\boldsymbol{ABC}) = \operatorname{tr}(\boldsymbol{CAB})$, the rank 1 decomposition above implies that

$$\operatorname{tr}(\boldsymbol{AC}) = \operatorname{tr}(\boldsymbol{R}^\mathsf{T}\boldsymbol{A}\boldsymbol{R}\boldsymbol{\Lambda}) = \sum_{i=1}^n \operatorname{tr}(\boldsymbol{R}^\mathsf{T}\boldsymbol{A}\boldsymbol{R} v_i v_i^\mathsf{T}) = \sum_{i=1}^n \operatorname{tr}((\boldsymbol{R}v_i)^\mathsf{T}\boldsymbol{A}(\boldsymbol{R}v_i)) \ge 0,$$

since $\boldsymbol{A}$ is positive semidefinite. If $\boldsymbol{B} \ge \boldsymbol{A}$, then $\boldsymbol{B} - \boldsymbol{A}$ is positive semidefinite, so $\operatorname{tr}((\boldsymbol{B} - \boldsymbol{A})\boldsymbol{C}) \ge 0$, which implies that $\operatorname{tr}(\boldsymbol{BC}) \ge \operatorname{tr}(\boldsymbol{AC})$. □

Chapter 4

The Parisi Formula for Spherical Vector Spin Glasses

We now move onto the main contribution of this thesis. We will extend the ideas used to study the classical Ising spin, spherical, and vector spin models introduced in the previous chapters to prove a free energy formula for constrained copies of spherical spin glasses. This model is known as the *spherical vector spin glass*. Our goal for this chapter is to prove a free energy formula for this model.

The free energy for multiple systems of spherical spin glasses with constrained overlaps was first studied in [48]. In [83] the authors proved an upper bound of the constrained free energy using Guerra's interpolation. In this chapter, we prove this upper bound is sharp. Our approach combines the ideas of the Aizenman–Sims–Starr scheme in [30] and the synchronization mechanism used in the vector spin models in [80] and [81]. We derive a vector version of the Aizenman–Sims–Starr scheme for spherical spin glass and use the synchronization property of arrays obeying the overlap-matrix form of the Ghirlanda–Guerra identities to prove the matching lower bound.

The results of this chapter are adapted from [64] written by the author of this thesis.

Outline of the Chapter

We first review the motivation for these models in Section 4.1 in the context of the spherical spin glass models introduced in Chapter 1. We define the spherical vector model and remind the reader of the notation introduced in the earlier chapters. The main result of this chapter is the analogue of the Parisi formula for this high dimensional spin glass model.

The proof of the analogue of the Parisi formula begins with Section 4.2. We begin by using an analogue of Guerra's interpolation to prove the upper bound in Section 4.2. In Section 4.3, we prove the sharpness of functionals that appeared in the upper bound using classical large deviations. We begin the proof of the lower bound by using the Poincaré limit to derive an analogue of the Aizenman–Sims–Starr scheme for high dimensional spherical spin glass models in Section 4.4. In Section 4.5, we introduce a perturbation of the Hamiltonian that will force the overlaps under the asymptotic Gibbs measure to satisfy the synchronization properties used in the study of vector spin glass models. In Section 4.6 we combine all the results and finish the proof of the lower bound using standard cavity computations.

4.1 Introduction

In this section we will motivate the mathematical importance of the spherical vector spin glasses and present the first main contribution of this thesis.

4.1.1 Motivation

We will motivate the model as a natural generalization of the classical spherical spin glasses. Recall that the distribution of the overlap arrays encodes the limiting distribution of the free energies.

For $1 \leq j \leq n$, consider the usual spherical mixed p-spin (1.26) Hamiltonian defined on S_N,

$$H_N^j(\boldsymbol{\sigma}) = \sum_{p \geq 1} \beta_p^j H_{N,p}(\boldsymbol{\sigma}) \quad \text{where} \quad \frac{1}{N^{(p-1)/2}} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1 \ldots i_p} \sigma_{i_1} \cdots \sigma_{i_p}. \tag{4.1}$$

A key observation is that the random variables $g_{i_1 \ldots i_p}$ are the same for all Hamiltonians. The difference between the H_N^j is only in the temperature parameters β^j. For $j \leq n$ consider replicas n replicas $\boldsymbol{\sigma}(j)$ sampled at from the Gibbs measure

$$dG_N^j(\boldsymbol{\sigma}(j)) \propto e^{H_N^j(\boldsymbol{\sigma}) + \vec{h}(j) \sum_{i \leq N} \sigma_i(j)} \quad \text{where} \quad \vec{h}(j) \in \mathbb{R}$$

on S_N. We let $\langle \cdot \rangle$ denote the average with respect to the product measure $\prod_{i=1}^j G_N^j$.

The goal is to understand the distribution of the overlaps these n replica of spherical spin glasses sampled from these Gibbs measures

$$\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) = \left(\frac{\boldsymbol{\sigma}(i) \cdot \boldsymbol{\sigma}(j)}{N} \right)_{i,j \leq n}.$$

Let $\boldsymbol{Q} \in \mathbb{S}_n^+$ denote a possible value for $\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}})$. We define the constrained free energy by

$$F_N^{\varepsilon}(\boldsymbol{Q}) = \frac{1}{N} \mathbb{E} \log \int_{S_N^n} \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q}) \exp \Big(\sum_{j \leq n} H_N^j(\boldsymbol{\sigma}(j)) + \vec{h}(j) \sum_{i \leq N} \sigma_i(j) \Big) \, d\lambda_N^{\otimes n}(\vec{\boldsymbol{\sigma}}), \tag{4.2}$$

and the unconstrained 1 dimensional free energy with

$$F_N^j = \frac{1}{N} \mathbb{E} \log \int_{S_N} \exp \Big(\sum_{j \leq n} H_N^j(\boldsymbol{\sigma}(j)) + \vec{h}(j) \sum_{i \leq N} \sigma_i(j) \Big) \, d\lambda_N(\boldsymbol{\sigma}(j)) \tag{4.3}$$

Given $\varepsilon > 0$, we use the notation $\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q}$ to mean $\| \boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) - \boldsymbol{Q} \|_\infty \leq \varepsilon$ which says that all the entries of $\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}})$ is within an ε neighbourhood of $\boldsymbol{Q}$. By comparing the constrained free energy with the classical free energies, we can test the likelihood that $\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q}$. In particular, we can show that with high probability the overlap of n replica $\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}})$ cannot be in a configuration $\boldsymbol{Q}$ under the product Gibbs measure by comparing the free energies.

Proposition 4.1.1 (*Configurations of the Self Overlaps*)

Suppose for some $\delta > 0$,

$$\lim_{N\to\infty} F_N^{\varepsilon}(\boldsymbol{Q}) \leq \lim_{N\to\infty} \sum_{j\leq n} F_N - \delta. \tag{4.4}$$

There exists a constant $C(\delta)$ independent of N such that for N sufficiently large

$$\mathbb{E}\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\rangle \leq C(\delta) e^{-\frac{N}{C(\delta)}}.$$

Proof. This is a generalization of the argument in Proposition 13.4.3 in [102]. The argument follows by concentration of the free energy functions Theorem 3.1.3.

We define the random partition functions

$$Z_N(\boldsymbol{Q}) = \int_{S_N^n} \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q}) \exp\Big(\sum_{j\leq n} H_N^j(\boldsymbol{\sigma}(j)) + \vec{h}(j) \sum_{i\leq N} \sigma_i(j) \Big) d\lambda_N^{\otimes n}(\vec{\boldsymbol{\sigma}}), \tag{4.5}$$

and

$$Z_N^j = \int_{S_N} \exp\Big(\sum_{j\leq n} H_N^j(\boldsymbol{\sigma}(j)) + \vec{h}(j) \sum_{i\leq N} \sigma_i(j) \Big) d\lambda_N(\boldsymbol{\sigma}(j)). \tag{4.6}$$

After taking normalized logarithms and expectations, these quantities become (4.2) and (4.3). The ratio of these quantities give the probability that $\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q}$,

$$\mathbb{E}\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\rangle = \mathbb{E}\frac{Z_N(\boldsymbol{Q})}{\prod_{j\leq n} Z_N^j}.$$

For any $t > 0$, the concentration inequality for free energy functions Theorem 3.1.3 implies

$$\mathbb{P}(Z_N(\boldsymbol{Q}) \leq e^{NF_N(\boldsymbol{Q})+Nt}) = \mathbb{P}\left(\frac{1}{N}\log Z_N(\boldsymbol{Q}) - \frac{1}{N}\mathbb{E}\log Z_N(\boldsymbol{Q}) \leq t \right) \geq 1 - C(t)e^{-\frac{N}{C(t)}}, \tag{4.7}$$

and for all $j \leq n$,

$$\mathbb{P}(Z_N^j \geq e^{NF_N^j - Nt}) = \mathbb{P}\left(\frac{1}{N}\log Z_N - \frac{1}{N}\mathbb{E}\log Z_N \geq -t \right) \geq 1 - C(t)e^{-\frac{N}{C(t)}}, \tag{4.8}$$

where the constants $C(t)$ depends on the covariance structure of the Hamiltonian (which is bounded) and t. These random quantities are close to its averages with high probability, so condition (4.4) in the statement of the theorem holds with high probability. We define the events

$$A = \{Z_N(\boldsymbol{Q}) \leq e^{NF_N(\boldsymbol{Q})+N\delta}\} \quad \text{and} \quad B = \{Z_N^j \geq e^{NF_N^j - N\frac{\delta}{n}} \text{ for all } j \leq n\}$$

Then

$$\mathbb{E}\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\rangle = \mathbb{E}\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\, \mathbb{1}(A \cap B)\rangle + \mathbb{E}\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\, \mathbb{1}(A^c \cup B^c)\rangle$$

On the first set, condition (4.4) implies that

$$\begin{aligned}\mathbb{E}\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\, \mathbb{1}(A\cap B)\rangle &= \mathbb{E}\frac{Z_N(\boldsymbol{Q})}{\prod_{j\le n} Z_N^j}\, \mathbb{1}(A\cap B) \\ &\le \mathbb{E}\, \mathbb{1}(A\cap B) e^{N(F_N(\boldsymbol{Q})-\sum_{j\le n} F_N^j)} \\ &\le e^{-N\delta}. \end{aligned} \tag{4.9}$$

By the Gaussian concentration estimates (4.7) and (4.8), the other term is also exponentially small by the union bound,

$$\mathbb{E}\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\, \mathbb{1}(A^c\cup B^c)\rangle \le \mathbb{E}\langle \mathbb{1}(A^c)\rangle + \mathbb{E}\langle \mathbb{1}(B^c)\rangle = \mathbb{P}(A^c)+\mathbb{P}(B^c) \le \tilde{C}(\delta) e^{-\frac{N}{\tilde{C}(\delta)}} \tag{4.10}$$

for some constant $\tilde{C}(\delta)$ that only depends on δ and n. Redefining the constant and combining the estimates in (4.9) and (4.10) finishes the proof. □

A formula for the free energy will allow us to understand the support of n the overlaps generated by n replica by comparing it to the classical free energy formulas. The remainder of this chapter is devoted to defining the vector spin free energy and proving its corresponding variational formula.

4.1.2 Model Description

Fix $n \ge 1$. The main goal is to find a formula for the free energy of n constrained copies of mixed even p-spin spherical spin glasses. The copies are coupled by constraining their overlaps and can possibly exist at different temperatures. We start by reviewing the usual spherical spin glass model defined in Chapter 1.

Let S_N be the sphere in $\mathbb{R}^N$ of radius $\sqrt{N}$ and denote the configuration of the jth copy by

$$\boldsymbol{\sigma}(j) = \big(\vec{\sigma}_1(j),\dots,\vec{\sigma}_N(j)\big) \in S_N. \tag{4.11}$$

For $p \ge 2$, the p-spin Hamiltonian is denoted by

$$H_{N,p}(\boldsymbol{\sigma}(j)) = \frac{1}{N^{(p-1)/2}} \sum_{1\le i_1,\dots,i_p\le N} g_{i_1,\dots,i_p}\vec{\sigma}_{i_1}(j)\cdots\vec{\sigma}_{i_p}(j), \tag{4.12}$$

where $g_{i_1,\dots,i_p}$ are i.i.d. standard Gaussian for all $p \ge 2$ and indices $(i_1,\dots,i_p)$. The corresponding even mixed p-spin Hamiltonian for the jth copy at inverse temperature $(\vec{\beta}_p(j))_{p\ge 2}$ is denoted by

$$H_N^j(\vec{\boldsymbol{\sigma}}) = \sum_{p\ge 2} \vec{\beta}_p(j) H_{N,p}(\boldsymbol{\sigma}(j)). \tag{4.13}$$

We assume that the inverse temperatures satisfy $\sum_{p\ge 2} 2^p \vec{\beta}_p^{\,2}(j) < \infty$ for all $j \le n$, so that the covariance of (4.13) is well-defined, and that $\vec{\beta}_p(j) = 0$ for odd p.

We now introduce the model for a system of n copies of spherical spin glass. A configuration of n copies can be viewed as vector spins,

$$\vec{\boldsymbol{\sigma}} = (\vec{\sigma}_1,\dots,\vec{\sigma}_N) \in (\mathbb{R}^n)^N, \tag{4.14}$$

where the vector entries of $\vec{\boldsymbol{\sigma}}$ are denoted by

$$\vec{\sigma}_i = \big(\vec{\sigma}_i(1), \ldots, \vec{\sigma}_i(n)\big) \in \mathbb{R}^n. \tag{4.15}$$

The configurations $\vec{\boldsymbol{\sigma}}$ are restricted to the set

$$S_N^n = \big\{\vec{\boldsymbol{\sigma}} \in (\mathbb{R}^N)^n \mid \|\boldsymbol{\sigma}(j)\| = \sqrt{N} \text{ for all } j \leq n\big\}, \tag{4.16}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^N$. The Hamiltonian of n copies of even mixed p-spin models of spherical spin glasses is denoted by

$$H_N(\vec{\boldsymbol{\sigma}}) = \sum_{j \leq n} H_N^j(\vec{\boldsymbol{\sigma}}). \tag{4.17}$$

The upper indices $\ell \geq 1$ of the configurations $\vec{\boldsymbol{\sigma}}^\ell$ index sequences of spin configurations. The Hamiltonian is a Gaussian process indexed by $\vec{\boldsymbol{\sigma}}^\ell \in S_N^n$ with covariance given by functions of normalized inner products. The inner products, or overlaps, of the configurations of copy $\boldsymbol{\sigma}^\ell(j)$ and $\boldsymbol{\sigma}^{\ell'}(j')$ is denoted by

$$R_{\ell,\ell'}^{j,j'} = R_{\ell,\ell'}^{j,j'}\big(\vec{\boldsymbol{\sigma}}^\ell(j), \vec{\boldsymbol{\sigma}}^{\ell'}(j')\big) = \frac{1}{N}\sum_{i \leq N} \vec{\sigma}_i^\ell(j)\vec{\sigma}_i^{\ell'}(j'). \tag{4.18}$$

The overlaps of vector configurations $\vec{\boldsymbol{\sigma}}^\ell$ and $\vec{\boldsymbol{\sigma}}^{\ell'}$ are given by the overlap matrices

$$\boldsymbol{R}_{\ell,\ell'} = \boldsymbol{R}(\vec{\boldsymbol{\sigma}}^\ell, \vec{\boldsymbol{\sigma}}^{\ell'}) = \big(R_{\ell,\ell'}^{j,j'}\big)_{j,j' \leq n} = \frac{1}{N}\sum_{i \leq N} \vec{\sigma}_i^\ell \otimes \vec{\sigma}_i^{\ell'}. \tag{4.19}$$

The overlaps are always normalized by the dimension of the vectors in the inner product. Let $x \in \mathbb{R}^n$ and let $\boldsymbol{A} = (A_{j,j'})_{j,j' \leq n} \in \mathbb{R}^{n \times n}$. Consider the real valued convex function

$$\xi_{j,j'}(x) = \sum_{p \geq 2} \vec{\beta}_p(j)\vec{\beta}_p(j')x^p \tag{4.20}$$

and its matrix valued counterpart

$$\boldsymbol{\xi}(\boldsymbol{A}) = \big(\xi_{j,j'}(A_{j,j'})\big)_{j,j' \leq n} = \sum_{p \geq 2} (\vec{\beta}_p \otimes \vec{\beta}_p) \odot \boldsymbol{A}^{\odot p}, \tag{4.21}$$

where $\otimes$ is the outer product on vectors in $\mathbb{R}^n$ and $\odot$ is the Hadamard product on $n \times n$ matrices. It is easy to check that the mixed p-spin Hamiltonian of the copies (4.13) are centered Gaussian processes with covariance

$$\mathbb{E}H_N^j\big(\vec{\boldsymbol{\sigma}}^\ell\big)H_N^{j'}\big(\vec{\boldsymbol{\sigma}}^{\ell'}\big) = N\xi_{j,j'}\big(R_{\ell,\ell'}^{j,j'}\big), \tag{4.22}$$

and the Hamiltonian (4.17) is a centered Gaussian process with covariance

$$\mathbb{E}H_N(\vec{\boldsymbol{\sigma}}^\ell)H_N(\vec{\boldsymbol{\sigma}}^{\ell'}) = N\,\mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{R}_{\ell,\ell'})), \tag{4.23}$$

where the sum of all entries in a matrix is denoted by

$$\mathrm{Sum}(\boldsymbol{A}) = \sum_{j,j' \leq n} A_{j,j'}. \tag{4.24}$$

4.1.3 The Limit of the Free Energy

We now define the constrained free energy. Let $\boldsymbol{Q} = \left(Q^{j,j'}\right)_{j,j' \leq n}$ be a $n \times n$ symmetric positive semidefinite matrix with off-diagonals, $Q^{j,j'} \in [-1,1]$ for $j \neq j'$ and diagonals $Q^{j,j} = 1$. Given $\varepsilon > 0$, we denote the set of spins with constrained self overlaps by

$$Q_N^{\varepsilon} = \left\{ \vec{\boldsymbol{\sigma}} \in S_N^n \mid \| \boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) - \boldsymbol{Q} \|_{\infty} \leq \varepsilon \right\}, \tag{4.25}$$

where $\| \cdot \|_{\infty}$ is the infinity norm on $n \times n$ matrices. For an external field $\vec{h} = \left(\vec{h}(j)\right)_{j \leq n} \in \mathbb{R}^n$, we define the free energy as

$$F_N^{\varepsilon}(\vec{\boldsymbol{\beta}}, \boldsymbol{Q}) = \frac{1}{N} \mathbb{E} \log \int_{Q_N^{\varepsilon}} \exp\left(H_N(\vec{\boldsymbol{\sigma}}) + \sum_{j \leq n} \vec{h}(j) \sum_{i \leq N} \vec{\sigma}_i(j) \right) d\lambda_N^n(\vec{\boldsymbol{\sigma}}), \tag{4.26}$$

where the reference measure $\lambda_N^n = \lambda_N^{\otimes n}$ is the product of normalized uniform measures λ_N on S_N.

We will prove the limit of (5.8) can be expressed as a Parisi type functional. We begin by introducing some notation. Let

$$\Gamma_n = \left\{ \boldsymbol{A} \mid \boldsymbol{A} \text{ is a } n \times n \text{ positive-semidefinite matrix} \right\}, \tag{4.27}$$

denote the space of $n \times n$ matrices, and let

$$\Pi = \left\{ \boldsymbol{\pi} : [0,1] \to \Gamma_n \mid \boldsymbol{\pi} \text{ is left-continuous, } \boldsymbol{\pi}(x_1) \leq \boldsymbol{\pi}(x_2) \text{ for } x_1 \leq x_2 \right\} \tag{4.28}$$

denote the space of left-continuous monotone paths on Γ_n. The notation $\boldsymbol{\pi}(x_1) \leq \boldsymbol{\pi}(x_2)$ means $\boldsymbol{\pi}(x_2) - \boldsymbol{\pi}(x_1) \in \Gamma_n$. Distances between paths are given by the metric

$$\mathrm{d}(\boldsymbol{\pi}, \tilde{\boldsymbol{\pi}}) = \int_0^1 \| \boldsymbol{\pi}(x) - \tilde{\boldsymbol{\pi}}(x) \|_1 \, dx, \tag{4.29}$$

where $\|\boldsymbol{A}\|_1 = \sum_{j,j'} |A_{j,j'}|$. These paths are the functional order parameters of p-spin models with vector spins.

Consider a left continuous discrete path $\boldsymbol{\pi} \in \Pi$ connecting $\boldsymbol{0}$ and $\boldsymbol{Q}$,

$$\boldsymbol{\pi}(x) = \boldsymbol{Q}_k \text{ for } x_{k-1} < x \leq x_k \text{ for } 0 \leq k \leq r, \quad \boldsymbol{\pi}(0) = \boldsymbol{0}, \quad \boldsymbol{\pi}(1) = \boldsymbol{Q}. \tag{4.30}$$

This path can be encoded with a sequence of real numbers

$$0 = x_{-1} < x_0 < \cdots < x_r = 1, \tag{4.31}$$

and a monotone sequence of $n \times n$ symmetric positive semi-definite matrices

$$\boldsymbol{0} = \boldsymbol{Q}_0 \leq \boldsymbol{Q}_1 \leq \cdots \leq \boldsymbol{Q}_r = \boldsymbol{Q}. \tag{4.32}$$

Recall definition (4.21), and denote

$$\boldsymbol{\theta}(\boldsymbol{A}) = \left(\theta_{j,j'}(A_{j,j'})\right)_{j,j' \leq n} = \boldsymbol{A} \odot \boldsymbol{\xi}'(\boldsymbol{A}) - \boldsymbol{\xi}(\boldsymbol{A}), \tag{4.33}$$

where $\boldsymbol{\xi}'(\boldsymbol{A}) = (\xi'_{j,j'}(A_{j,j'}))_{j,j'\le n}$ is the matrix of entry wise derivatives of $\boldsymbol{\xi}$. The matrix given by

$$\boldsymbol{\Delta}_k = \boldsymbol{\xi}'(\boldsymbol{Q}_k) - \boldsymbol{\xi}'(\boldsymbol{Q}_{k-1}), \quad 1 \le k \le r, \tag{4.34}$$

is positive semidefinite. This can seen by applying the Schur product theorem to the Hadamard product representation (4.21).

Given a symmetric positive definite matrix $\boldsymbol{\Lambda}$, for $k \le r$ we define recursively

$$\boldsymbol{\Lambda}_r = \boldsymbol{\Lambda}, \quad \boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}_{k+1} - x_k \boldsymbol{\Delta}_{k+1} \text{ for } 0 \le k \le r-1. \tag{4.35}$$

Let $|\cdot|$ be the determinant of $n \times n$ matrices and consider the set

$$\mathcal{L} := \mathcal{L}(\boldsymbol{\pi}) = \{\boldsymbol{\Lambda} \in \Gamma_n \mid |\boldsymbol{\Lambda}_0| > 0\}. \tag{4.36}$$

For $\boldsymbol{\Lambda} \in \mathcal{L}$ and discrete $\boldsymbol{\pi} \in \Pi$, we define the following functional

$$\mathscr{P}_{\vec{\boldsymbol{\beta}},\boldsymbol{Q}}(\boldsymbol{\Lambda},\boldsymbol{\pi}) = \frac{1}{2}\Big[\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) - n - \log|\boldsymbol{\Lambda}| + (\boldsymbol{\Lambda}_0^{-1}\vec{h},\vec{h}) + \sum_{0\le k\le r-1} \frac{1}{x_k} \log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|} \tag{4.37}$$

$$- \sum_{0\le k\le r-1} x_k \cdot \operatorname{Sum}\big(\boldsymbol{\theta}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\theta}(\boldsymbol{Q}_k)\big)\Big]. \tag{4.38}$$

The dependence on $\vec{\boldsymbol{\beta}}$ is through the functions $\boldsymbol{\xi}$ and $\boldsymbol{\theta}$ defined in (4.21) and (4.33). The following is the main result:

Theorem 4.1.1

For $n \ge 1$ and $\vec{h} \in \mathbb{R}^n$, the limit of the free energy at inverse temperature $\vec{\boldsymbol{\beta}}$ and constraint $\boldsymbol{Q}$ is given by

$$\lim_{\varepsilon\to 0}\lim_{N\to\infty} F_N^{\varepsilon}(\vec{\boldsymbol{\beta}},\boldsymbol{Q}) = \inf_{\boldsymbol{\pi},\boldsymbol{\Lambda}} \mathscr{P}_{\vec{\boldsymbol{\beta}},\boldsymbol{Q}}(\boldsymbol{\Lambda},\boldsymbol{\pi}). \tag{4.39}$$

The infimum is over $\boldsymbol{\Lambda} \in \mathcal{L}$ and discrete paths given by (4.31) and (4.32) over all $r \ge 1$.

This formula implies a large deviations principle with explicit rate function for the overlap distribution under the Gibbs measure. If we define μ_N as the distribution of the overlaps between the n replica,

$$\mu_N(\boldsymbol{Q}) = \lim_{\varepsilon\to 0}\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\rangle$$

then μ_N satisfies a large deviations principle almost surely with rate function

$$I(\boldsymbol{Q}) = -\inf_{\boldsymbol{\pi},\boldsymbol{\Lambda}} \mathscr{P}_{\vec{\boldsymbol{\beta}},\boldsymbol{Q}}(\boldsymbol{\Lambda},\boldsymbol{\pi}) + \sum_{i\le n} \inf_x \mathscr{C}_{\vec{\boldsymbol{\beta}}(i)}(x) \tag{4.40}$$

where $\vec{\boldsymbol{\beta}}(i)$ is the sequence of the temperature parameters defining $\xi(q)$ in the mixed p-spin Hamiltonian appearing in the 1-dimensional Crisanti–Sommers formula (1.28). This follows from the fact that the normalization terms decouple, so

$$\lim_{\varepsilon\to 0}\frac{1}{N}\mathbb{E}\log\mu_N(\boldsymbol{Q}) = \lim_{\varepsilon\to 0}\frac{1}{N}\mathbb{E}\log\langle \mathbb{1}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}) \approx \boldsymbol{Q})\rangle = \lim_{\varepsilon\to 0} F_N(\vec{\boldsymbol{\beta}},\boldsymbol{Q}) - \sum_{j=1}^{n} F_N(\vec{\boldsymbol{\beta}}(i))$$

where $F_N(\vec{\boldsymbol{\beta}}(j))$ are the 1-dimensional spherical free energies defined in (1.33) with respect to the p-spin Hamiltonian with temperature parameters $\vec{\boldsymbol{\beta}}(j)$. The almost sure large deviations principle follows by Gaussian concentration [Proposition 4.1.1] and is stated precisely in [12, Theorem 1.13].

Remark 4.1.1. If $\det(\boldsymbol{Q}) = 0$, we show in Lemma 4.3.2 that for all fixed β and $\boldsymbol{\pi}$,

$$\inf_{\boldsymbol{\Lambda}} \mathscr{P}_{\vec{\boldsymbol{\beta}},\boldsymbol{Q}}(\boldsymbol{\Lambda},\boldsymbol{\pi}) = -\infty.$$

By concentration of measure, this implies that asymptotically degenerate sequences of vector configurations have exponentially low probability of appearing in the product Gibbs measure in the $N \to \infty$ limit. This is intuitive because the configurations with overlaps equal to degenerate matrices has almost no volume on the exponential scale.

Remark 4.1.2. Our form of the Parisi functional $\mathscr{P}_{\vec{\boldsymbol{\beta}},\boldsymbol{Q}}(\boldsymbol{\Lambda},\boldsymbol{\pi})$, is missing the $\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Delta}_1)$ that appears in [83]. This is because we assumed $x_0 > 0$ in (4.31) while $x_0 = 0$ in [83]. By applying L'Hôpital's rule and Jacobi's formula, this term can be recovered by observing

$$\begin{aligned}\lim_{x_0\to 0} \frac{1}{2x_0} \log \frac{|\boldsymbol{\Lambda}_1|}{|\boldsymbol{\Lambda}_1 - x_0\boldsymbol{\Delta}_1|} &= \lim_{x_0\to 0} \frac{1}{2}|\boldsymbol{\Lambda}_1 - x_0\boldsymbol{\Delta}_1|^{-1} \operatorname{tr}(|\boldsymbol{\Lambda}_1 - x_0\boldsymbol{\Delta}_1|(\boldsymbol{\Lambda}_1 - x_0\boldsymbol{\Delta}_1)^{-1}\boldsymbol{\Delta}_1) \\ &= \frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Delta}_1).\end{aligned}$$

We will reintroduce this term in Chapter 5 through continuity.

Remark 4.1.3. We will introduce a form of this formula without the Lagrange multiplier term $\boldsymbol{\Lambda}$ in Chapter 5. This is called the Crisanti–Sommers form of the free energy.

4.2 Upper Bound — Guerra's Interpolation

Remark 4.2.1. Throughout this chapter, we will denote by L any constant that depends only on the global parameters of the model such as the number of copies n, and the inverse temperature parameters $\vec{\boldsymbol{\beta}}$. The constant can change even within the same equation.

We begin by proving the upper bound of the free energy.

Lemma 4.2.1

For $n \geq 1$ and $\vec{h} \in \mathbb{R}^n$,

$$\lim_{\varepsilon\to 0} \limsup_{N\to\infty} F_N^{\varepsilon}(\vec{\boldsymbol{\beta}},\boldsymbol{Q}) \leq \inf_{\boldsymbol{\Lambda},\pi} \mathscr{P}_{\vec{\boldsymbol{\beta}},\boldsymbol{Q}}(\boldsymbol{\Lambda},\pi). \tag{4.41}$$

A version of this upper bound was proved in Section 2 of [83]. We will provide a different proof using the Ruelle probability cascades and Guerra's interpolation. The main difference is the following proof will hold without the condition that the diagonals of $\boldsymbol{\Lambda}$ are greater than 1.

Consider the sequence of real numbers

$$0 = x_{-1} < x_0 < \cdots < x_r = 1, \tag{4.42}$$

and the sequence of $n \times n$ positive semi definite matrices

$$\boldsymbol{0} = \boldsymbol{Q}_0 \leq \boldsymbol{Q}_1 \leq \cdots \leq \boldsymbol{Q}_r = \boldsymbol{Q}. \tag{4.43}$$

Let $(v_\alpha)_{\alpha\in\mathbb{N}^r}$ be the weights of the Ruelle probability cascades [90] corresponding to the sequence (4.42). For paths $\alpha^1, \alpha^2 \in \mathbb{N}^r$, we denote the common vertices by

$$\alpha^1 \wedge \alpha^2 = \min\big\{0 \le j \le r \mid \alpha_1^1 = \alpha_1^2, \dots, \alpha_j^1 = \alpha_j^2, \alpha_{j+1}^1 \ne \alpha_{j+1}^2\big\} \tag{4.44}$$

and $\alpha^1 \wedge \alpha^2 = r$ if $\alpha^1 = \alpha^2$. Consider independent centered Gaussian processes $Z(\alpha) = \big(Z^j(\alpha)\big)_{j\le n}$ and $Y(\alpha)$ indexed with $\alpha \in \mathbb{N}^r$ and covariances

$$\mathrm{Cov}(Z(\alpha^1), Z(\alpha^2)) = \boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1\wedge\alpha^2}), \tag{4.45}$$

$$\mathrm{Cov}(Y(\alpha^1), Y(\alpha^2)) = \mathrm{Sum}\,\big(\boldsymbol{\theta}(\boldsymbol{Q}_{\alpha^1\wedge\alpha^2})\big). \tag{4.46}$$

Let $Z_i(\alpha)$ be an independent copy of $Z(\alpha)$ also independent of $Y(\alpha)$. A Gaussian interpolation argument will bound the free energy with functions of these Gaussian processes.

Lemma 4.2.2

For all $N > 0$, there exists a constant L such that

$$\begin{aligned} F_N^\varepsilon(\vec{\boldsymbol{\beta}}, \boldsymbol{Q}) \le \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{Q_N^\varepsilon} \exp\Big(\sum_{i\le N}\sum_{j\le n} \vec{\sigma}_i(j)\big(Z_i^j(\alpha) + \vec{h}(j)\big)\Big)\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}) \\ - \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha \exp\sqrt{N}Y(\alpha) + L\varepsilon. \end{aligned} \tag{4.47}$$

Proof. The result follows from Gaussian interpolation. For $0 \le t \le 1$, we define the interpolating Hamiltonian

$$H_t(\vec{\boldsymbol{\sigma}}, \alpha) = \sqrt{t}H_N(\vec{\boldsymbol{\sigma}}) + \sum_{i\le N}\sum_{j\le n} \vec{\sigma}_i(j)\big(\sqrt{1-t}Z_i^j(\alpha) + \vec{h}(j)\big) + \sqrt{t}\sqrt{N}Y(\alpha),$$

on $S_N^n \times \mathbb{N}^r$. For a given a constraint $\boldsymbol{Q}$, we define the interpolating free energy function

$$\varphi(t) = \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{Q_N^\varepsilon} \exp H_t(\vec{\boldsymbol{\sigma}}, \alpha)\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}).$$

Let $\langle\cdot\rangle_t$ be the average on $Q_N^\varepsilon \times \mathbb{N}^r$ with respect to the Gibbs measure

$$G(d\vec{\boldsymbol{\sigma}}, \alpha) \propto v_\alpha \exp H_t(\vec{\boldsymbol{\sigma}}, \alpha)\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}).$$

A straightforward computation shows

$$\varphi'(t) = \frac{1}{N}\mathbb{E}\Big\langle \frac{\partial}{\partial t} H_t(\vec{\boldsymbol{\sigma}}, \alpha)\Big\rangle_t.$$

By Gaussian integration by parts [Theorem 3.1.4],

$$\frac{1}{N}\mathbb{E}\Big\langle \frac{\partial}{\partial t} H_t(\vec{\boldsymbol{\sigma}}, \alpha)\Big\rangle_t = \frac{1}{2}\mathbb{E}\Big\langle \mathrm{Sum}\,\Big(\boldsymbol{\xi}(\boldsymbol{R}_{1,1}) - \boldsymbol{R}_{1,1}\odot\boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1\wedge\alpha^1}) + \boldsymbol{\theta}(\boldsymbol{Q}_{\alpha^1\wedge\alpha^1})\Big)\Big\rangle_t \tag{4.48}$$

$$- \frac{1}{2}\mathbb{E}\Big\langle \mathrm{Sum}\,\Big(\boldsymbol{\xi}(\boldsymbol{R}_{1,2}) - \boldsymbol{R}_{1,2}\odot\boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1\wedge\alpha^2}) + \boldsymbol{\theta}(\boldsymbol{Q}_{\alpha^1\wedge\alpha^2})\Big)\Big\rangle_t. \tag{4.49}$$

We use convexity [Proposition 3.3.10] to bound (4.49). Since $\vec{\beta}_p = \vec{0}$ for odd p, $\xi_{j,j'}(x)$ is a convex function for all $j, j' \leq n$ and therefore lies above all its tangent lines. That is,

$$\xi_{j,j'}(a) - a\xi'_{j,j'}(b) + \theta_{j,j'}(b) \geq 0 \text{ for all } a, b \in \mathbb{R}.$$

which implies, $\text{Sum}\left(\boldsymbol{\xi}(\boldsymbol{R}_{1,2}) - \boldsymbol{R}_{1,2} \odot \boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1 \wedge \alpha^2}) + \boldsymbol{\theta}(\boldsymbol{Q}_{\alpha^1 \wedge \alpha^2})\right)$ is non-negative. To bound (4.48) we use definition (4.33) and notice (4.48) is equal to

$$\mathbb{E}\Big\langle \text{Sum}\left(\boldsymbol{\xi}(\boldsymbol{R}_{1,1}) - \boldsymbol{\xi}(\boldsymbol{Q}_{\alpha^1 \wedge \alpha^1}) - (\boldsymbol{R}_{1,1} - \boldsymbol{Q}_{\alpha^1 \wedge \alpha^1}) \odot \boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1 \wedge \alpha^1})\right)\Big\rangle_t. \tag{4.50}$$

The self overlaps are constrained, so $\|\boldsymbol{R}_{1,1} - \boldsymbol{Q}_{\alpha^1 \wedge \alpha^1}\|_\infty \leq \varepsilon$. The Lipschitz continuity of $\boldsymbol{\xi}$ implies (4.48) is bounded by $L\varepsilon$, for some constant L that does not depend on N.

These bounds on (4.48) and (4.49) imply

$$\varphi'(t) \leq L\varepsilon. \tag{4.51}$$

By the mean value theorem, (4.51) gives us the upper bound

$$\varphi(1) \leq \varphi(0) + L\varepsilon, \tag{4.52}$$

where

$$\varphi(1) = F_N^\varepsilon(\vec{\boldsymbol{\beta}}, \boldsymbol{Q}) + \frac{1}{N}\mathbb{E}\log \sum_{\alpha \in \mathbb{N}^r} v_\alpha \exp \sqrt{N} Y(\alpha), \tag{4.53}$$

$$\varphi(0) = \frac{1}{N}\mathbb{E}\log \sum_{\alpha \in \mathbb{N}^r} v_\alpha \int_{Q_N^\varepsilon} \exp\Big(\sum_{i \leq N}\sum_{j \leq n} \vec{\sigma}_i(j)\big(Z_i^j(\alpha) + \vec{h}(j)\big)\Big)\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}). \tag{4.54}$$

Rearranging terms finishes the proof of the upper bound. □

The terms in (4.47) containing $Y(\alpha)$ and $Z(\alpha)$ can be computed explicitly using the recursive construction of the Ruelle probability cascades [Theorem 2.2.5]. Recalling the covariance structure in (4.46), a recursive computation [76, Chapter 3] using Theorem 2.2.5 shows

$$\limsup_{N \to \infty} \frac{1}{N}\mathbb{E}\log \sum_{\alpha \in \mathbb{N}^r} v_\alpha \exp \sqrt{N} Y(\alpha) = \sum_{0 \leq k \leq r-1} x_k \cdot \text{Sum}\left(\boldsymbol{\theta}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\theta}(\boldsymbol{Q}_k)\right). \tag{4.55}$$

The term in (4.47) containing $Z(\alpha)$ can be computed similarly after decoupling the constraint on Q_N^ε using Lagrange multipliers and rotational invariance [83, Lemma 1]. Let ν_N be the standard Gaussian measure on $\mathbb{R}^N$. We write $\boldsymbol{\omega}(j) \in \mathbb{R}^N$ in its polar coordinate form $\boldsymbol{\omega}(j) = (s_j \boldsymbol{\sigma}(j))$, where $s_j = \frac{\|\boldsymbol{\omega}(j)\|}{\sqrt{N}} \in \mathbb{R}^+$ and $\boldsymbol{\sigma}(j) = \frac{\sqrt{N}\boldsymbol{\omega}(j)}{\|\boldsymbol{\omega}(j)\|} \in S_N$. Let γ_N denote the law of s_j under ν_N. By rotational invariance, the law of $\boldsymbol{\sigma}(j)$ under ν_j is λ_N, and $\boldsymbol{\sigma}(j)$ and s_j are independent [Proposition 3.1.1]. We express (4.54) in terms of a Gaussian integral.

Lemma 4.2.3

There exists a $\delta \in (0, \varepsilon)$, such that (4.54) is bounded above by

$$\frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{\Omega_N^{\varepsilon,\delta}} \exp\Big(\sum_{i\le N}\sum_{j\le n} \vec{\omega}_i(j)\big(Z_i^j(\alpha)+\vec{h}(j)\big)\Big)\, d\nu_N^n(\vec{\boldsymbol{\omega}}) - \frac{n\log \nu_N(E_N^\delta)}{N} + L\delta \tag{4.56}$$

where the δ shell around Q_N^ε is denoted by

$$\Omega_N^{\varepsilon,\delta} = \big\{\vec{\boldsymbol{\omega}} = (s_j\boldsymbol{\sigma}(j))_{j\le n} \in (\mathbb{R}^N)^n \mid \vec{\boldsymbol{\sigma}} \in Q_N^\varepsilon,\ s_j \in [\sqrt{1-\delta}, \sqrt{1+\delta}] \text{ for all } j \le n\big\} \tag{4.57}$$

and the δ neighbourhood of the radial component is denoted by

$$E_N^\delta = \{x \in \mathbb{R}^N \mid \|x\| \in [\sqrt{(1-\delta)N}, \sqrt{(1+\delta)N}]\}.$$

Proof. We will use a Gaussian interpolation argument. Let $\tilde{Z}_i^j(\alpha)$ be an independent copy of $Z_i^j(\alpha)$. For $0 \le t \le 1$, we define the interpolating Hamiltonian

$$H_t(\vec{\boldsymbol{\omega}}, \alpha) = \sqrt{t}\Big(\sum_{i\le N}\sum_{j\le n} \vec{\sigma}_i(j)\tilde{Z}_i^j(\alpha)\Big) + \sqrt{1-t}\Big(\sum_{i\le N}\sum_{j\le n} \vec{\omega}_i(j)Z_i^j(\alpha)\Big) + \sum_{i\le N}\sum_{j\le n} \vec{\sigma}_i(j)\vec{h}(j),$$

on $\Omega_N^{\varepsilon,\delta} \times \mathbb{N}^r$. The corresponding interpolating free energy function is denoted by

$$\varphi(t) = \frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{\Omega_N^{\varepsilon,\delta}} \exp H_t(\vec{\boldsymbol{\omega}}, \alpha)\, d\nu_N^n(\vec{\boldsymbol{\omega}}).$$

Let $\langle\cdot\rangle_t$ be the average on $\Omega_N^{\varepsilon,\delta} \times \mathbb{N}^r$ with respect to the Gibbs measure

$$G(d\vec{\boldsymbol{\omega}}, \alpha) \propto v_\alpha \exp H_t(\vec{\boldsymbol{\omega}}, \alpha)\, d\nu_N^n(\vec{\boldsymbol{\omega}}).$$

By Gaussian integration by parts,

$$\varphi'(t) = \frac{1}{N}\mathbb{E}\Big\langle \frac{\partial}{\partial t} H_t(\vec{\boldsymbol{\omega}}, \alpha)\Big\rangle_t = \mathbb{E}\Big\langle \mathbb{E}\frac{\partial H_t(\vec{\boldsymbol{\omega}}^1, \alpha^1)}{\partial t} \cdot H_t(\vec{\boldsymbol{\omega}}^1, \alpha^1) - \mathbb{E}\frac{\partial H_t(\vec{\boldsymbol{\omega}}^1, \alpha^1)}{\partial t} \cdot H_t(\vec{\boldsymbol{\omega}}^2, \alpha^2)\Big\rangle_t.$$

Computing the covariances, we get

$$\begin{aligned}\varphi'(t) = \frac{1}{2}\mathbb{E}\Big\langle \operatorname{Sum}\Big(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^1, \vec{\boldsymbol{\sigma}}^1) \odot \boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1\wedge\alpha^1}) - \boldsymbol{R}(\vec{\boldsymbol{\omega}}^1, \vec{\boldsymbol{\omega}}^1) \odot \boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1\wedge\alpha^1})\Big)\Big\rangle_t \\ - \frac{1}{2}\mathbb{E}\Big\langle \operatorname{Sum}\Big(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^1, \vec{\boldsymbol{\sigma}}^2) \odot \boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1\wedge\alpha^2}) - \boldsymbol{R}(\vec{\boldsymbol{\omega}}^1, \vec{\boldsymbol{\omega}}^2) \odot \boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1\wedge\alpha^2})\Big)\Big\rangle_t.\end{aligned}$$

Since $\vec{\omega}_i(j) = s_j\vec{\sigma}_i(j)$ and $s_j \in [\sqrt{1-\delta}, \sqrt{1+\delta}]$, we have the bound

$$\operatorname{Sum}\Big(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^1, \vec{\boldsymbol{\sigma}}^2) \odot \boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1\wedge\alpha^2}) - \boldsymbol{R}(\vec{\boldsymbol{\omega}}^1, \vec{\boldsymbol{\omega}}^2) \odot \boldsymbol{\xi}'(\boldsymbol{Q}_{\alpha^1\wedge\alpha^2})\Big) \le \delta n^2 \|\boldsymbol{\xi}'(1)\|_\infty.$$

By the triangle inequality,

$$|\varphi'(t)| \le n^2\delta\|\boldsymbol{\xi}'(1)\|_\infty = L\delta,$$

resulting in the bound

$$\varphi(1) \leq \varphi(0) + L\delta. \tag{4.58}$$

The ending term of the interpolation can be simplified using rotational invariance of ν_N,

$$\begin{aligned}
\varphi(1) &= \frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{\Omega_N^{\varepsilon,\delta}} \exp\Big(\sum_{i\leq N}\sum_{j\leq n} \vec{\sigma}_i(j)\big(\tilde{Z}_i^j(\alpha)+\vec{h}(j)\big)\Big)\, d\nu_N^n(\vec{\boldsymbol{\omega}}) \\
&= \frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{[\sqrt{1-\delta},\sqrt{1+\delta}]^n} \int_{Q_N^\varepsilon} \exp\Big(\sum_{i\leq N}\sum_{j\leq n} \vec{\sigma}_i(j)\big(\tilde{Z}_i^j(\alpha)+\vec{h}(j)\big)\Big)\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}) d\gamma_N^n(s) \\
&= \frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{Q_N^\varepsilon} \exp\Big(\sum_{i\leq N}\sum_{j\leq n} \vec{\sigma}_i(j)\big(\tilde{Z}_i^j(\alpha)+\vec{h}(j)\big)\Big)\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}) + \frac{n\log \nu_N(E_N^\delta)}{N}.
\end{aligned} \tag{4.59}$$

Substituting (4.59) into (4.58) gives the bound

$$\begin{aligned}
&\frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{\Omega_N^{\varepsilon,\delta}} \exp\Big(\sum_{i\leq N}\sum_{j\leq n} \vec{\sigma}_i(j)\big(\tilde{Z}_i^j(\alpha)+\vec{h}(j)\big)\Big)\, d\nu_N^n(\vec{\boldsymbol{\omega}}) \\
&\leq \frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{\Omega_N^{\varepsilon,\delta}} \exp\Big(\sum_{i\leq N}\sum_{j\leq n} \Big(s_j\vec{\sigma}_i(j)\tilde{Z}_i^j(\alpha) + \vec{\sigma}_i(j)\vec{h}(j)\Big)\Big)\, d\nu_N^n(\vec{\boldsymbol{\omega}}) \\
&\quad - \frac{n\log \nu_N(E_N^\delta)}{N} + L\delta.
\end{aligned} \tag{4.60}$$

On the set $\Omega_N^{\varepsilon,\delta}$, the Cauchy–Schwarz inequality implies

$$\Big|\sum_{j\leq n}\sum_{i\leq N}\Big(s_j\vec{\sigma}_i(j)\vec{h}(j) - \vec{\sigma}_i(j)\vec{h}(j)\Big)\Big| \leq \delta\sqrt{N}\sum_{j\leq n}\|\vec{\boldsymbol{\sigma}}(j)\|\cdot|\vec{h}(j)| \leq \delta LN.$$

Therefore, we can replace $\vec{\sigma}_i(j)\vec{h}(j)$ with $\vec{\omega}_i(j)\vec{h}(j)$ in the upper bound of (4.60) and absorb the error into $L\delta$ giving

$$\frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{\Omega_N^{\varepsilon,\delta}} \exp\Big(\sum_{i\leq N}\sum_{j\leq n} \vec{\omega}_i(j)\big(Z_i^j(\alpha)+\vec{h}(j)\big)\Big)\, d\nu_N^n(\vec{\boldsymbol{\omega}}) - \frac{n\log \nu_N(E_N^\delta)}{N} + L\delta,$$

the required upper bound in (4.56). □

We now explicitly compute the upper bound of (4.56). We denote the subset of $\mathbb{R}^{Nn}$ constrained by coupling the overlaps with,

$$\tilde{\Omega}_N^\varepsilon = \big\{\,\vec{\boldsymbol{\omega}} \in (\mathbb{R}^N)^n \mid R^{j,j'}(\vec{\boldsymbol{\omega}},\vec{\boldsymbol{\omega}}) \in [Q^{j,j'}-\varepsilon, Q^{j,j'}+\varepsilon] \text{ for all } j,j' \leq n\,\big\}. \tag{4.61}$$

For $\delta < \varepsilon$, $\Omega_N^{\varepsilon,\delta} \subset \tilde{\Omega}_N^{2\varepsilon}$ so (4.56) is bounded above by

$$\frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{\tilde{\Omega}_N^{2\varepsilon}} \exp\Big(\sum_{i\leq N}\sum_{j\leq n} \vec{\omega}_i(j)\big(Z_i^j(\alpha)+\vec{h}(j)\big)\Big)\, d\nu_N^n(\vec{\boldsymbol{\omega}}) - \frac{n\log \nu_N(E_N^\delta)}{N} + L\varepsilon. \tag{4.62}$$

For any $\vec{\boldsymbol{\omega}} \in \tilde{\Omega}_N^{2\varepsilon}$ and $\boldsymbol{\Lambda} \in \mathcal{L}$,

$$\Big\| \sum_{j,j' \le n} \Lambda^{j,j'} Q^{j,j'} - \frac{1}{N} \sum_{j,j' \le n} \sum_{i \le N} \Lambda^{j,j'} \vec{\omega}_i(j') \vec{\omega}_i(j) \Big\|_1 \le 2\varepsilon \|\boldsymbol{\Lambda}\|_1.$$

Therefore, adding and subtracting $\frac{1}{2}\sum_{i\le N}((\boldsymbol{\Lambda}-\boldsymbol{I})\vec{\omega}_i,\vec{\omega}_i)$ from the exponent implies (5.83) can be bounded above by

$$\begin{aligned} &\frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{\tilde{\Omega}_N^{2\varepsilon}} \exp\Big(\sum_{i\le N}\sum_{j\le n} \vec{\omega}_i(j)\big(Z_i^j(\alpha)+\vec{h}(j)\big) - \frac{1}{2}\sum_{i\le N}((\boldsymbol{\Lambda}-\boldsymbol{I})\vec{\omega}_i,\vec{\omega}_i)\Big)\, d\nu_N^n(\vec{\boldsymbol{\omega}}) \\ &+ \frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) - \frac{n}{2} - \frac{n\log \nu_N(E_N^\delta)}{N} + 2\varepsilon\|\boldsymbol{\Lambda}\|_1 - L\varepsilon. \end{aligned} \tag{4.63}$$

Since $\tilde{\Omega}_N^{2\varepsilon} \subset (\mathbb{R}^N)^n$, if we define the function,

$$Y_{r,i}(\alpha) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp\Big(\sum_{j\le n} \vec{\omega}_i(j)\big(Z_i^j(\alpha)+\vec{h}(j)\big) - \frac{1}{2}\sum_{j,j'\le n} \Lambda^{j,j'}\vec{\omega}_i(j)\vec{\omega}_i(j')\Big)\, d\vec{\omega}_i \tag{4.64}$$

then our upper bound (4.63) can be written as

$$\frac{1}{N}\mathbb{E}\log \sum_{\alpha\in\mathbb{N}^r} v_\alpha \prod_{i\le N} Y_{r,i}(\alpha) + \frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) - \frac{n}{2} - \frac{n\log \nu_N(E_N^\delta)}{N} + \varepsilon\|\boldsymbol{\Lambda}\|_1 - L\varepsilon. \tag{4.65}$$

The term containing $Y_{r,i}(\alpha)$ in (4.65) can be computed recursively. Let $\vec{\boldsymbol{z}}_k = (z_k^j)_{j\le n}$ be a Gaussian vector with covariance $\boldsymbol{\Delta}_k$ defined in (4.34) and let $\vec{\boldsymbol{z}}_k$ be independent for $1 \le k \le r$. For $i \le M$ let $\vec{\boldsymbol{z}}_{k,i}$ be an independent copy of $\vec{\boldsymbol{z}}_k$. We define the recursion starting with

$$Y_{r,i} = \log \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \exp\Big(\sum_{j\le n} \vec{\omega}_i(j)\Big(\sum_{1\le k\le r} z_{k,i}^j + \vec{h}(j)\Big) - \frac{1}{2}\sum_{j,j'\le n} \Lambda^{j,j'}\vec{\omega}_i(j)\vec{\omega}_i(j')\Big)\, d\vec{\omega}_i \tag{4.66}$$

with subsequent values for $0 \le k \le r-1$ given recursively by

$$Y_{k,i} = \frac{1}{x_k}\log \mathbb{E}_k \exp x_k Y_{k+1,i}, \tag{4.67}$$

where $\mathbb{E}_k$ refers to expectation with respect to the random vector $\vec{\boldsymbol{z}}_{k+1,i}$. The $\vec{\boldsymbol{z}}_i$ are i.i.d. so $Y_{0,i} = Y_{0,1}$ for all $i \le N$. The recursive representation of the average in [Theorem 2.2.5] implies (4.65) can be written as

$$Y_{0,1} + \frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) - \frac{n}{2} - \frac{n\log \nu_N(E_N^\delta)}{N} + \varepsilon\|\boldsymbol{\Lambda}\|_1 - L\varepsilon. \tag{4.68}$$

In this model, $Y_{0,1}$ has a closed form. Starting from the start of the recursion, a direct computation (see equation (2.17) in [83]) shows

$$Y_{r,1} = -\frac{1}{2}\log|\boldsymbol{\Lambda}| + \frac{1}{2}\Big(\boldsymbol{\Lambda}^{-1}\Big(\sum_{1\le k\le r} \vec{z}_{k,1} + \vec{h}\Big), \Big(\sum_{1\le k\le r} \vec{z}_{k,1} + \vec{h}\Big)\Big). \tag{4.69}$$

Here $(\cdot,\cdot)$ is the scalar product of vectors in $\mathbb{R}^n$. The first term is non-random and will propagate through the recursion. The second term can be computed recursively using the following result:

Lemma 4.2.4

Let $\vec{g}$ be a Gaussian vector with covariance $\boldsymbol{C}$. Then for any $y \in \mathbb{R}^n$ and $x \in (0,1]$,

$$\frac{1}{x}\log \mathbb{E}\exp\Big(\frac{x}{2}\big(\boldsymbol{A}^{-1}(y+\vec{g}),y+\vec{g}\big)\Big) = \frac{1}{2x}\log\frac{|\boldsymbol{A}|}{|\boldsymbol{A}-x\boldsymbol{C}|} + \frac{1}{2}\big((\boldsymbol{A}-x\boldsymbol{C})^{-1}y,y\big).$$

Proof. The one dimensional case was proven in [98, Lemma 3.5]. We will prove the analogous result for $\mathbb{R}^n$. The expectation can be computed explicitly as follows,

$$\begin{aligned}
&\mathbb{E}\exp\Big(\frac{x}{2}\big(\boldsymbol{A}^{-1}(y+\vec{g}),y+\vec{g}\big)\Big)\\
&= \left(\frac{|\boldsymbol{C}|^{-1}}{(2\pi)^n}\right)^{1/2}\int_{\mathbb{R}^n}\exp\Big(\frac{x}{2}\big(\boldsymbol{A}^{-1}(y+z),y+z\big)-\frac{1}{2}\big(\boldsymbol{C}^{-1}z,z\big)\Big)\,dz\\
&= \left(\frac{|\boldsymbol{C}|^{-1}}{(2\pi)^n}\right)^{1/2}\int_{\mathbb{R}^n}\exp\Big(\frac{x}{2}\big((\boldsymbol{A}-x\boldsymbol{C})^{-1}y,y\big)-\frac{1}{2}\big((\boldsymbol{C}^{-1}-x\boldsymbol{A}^{-1})(z-\boldsymbol{B}y),(z-\boldsymbol{B}y)\big)\Big)\,dz\\
&= \left(\frac{|\boldsymbol{C}|^{-1}}{|\boldsymbol{C}^{-1}-x\boldsymbol{A}^{-1}|}\right)^{1/2}\exp\Big(\frac{x}{2}\big((\boldsymbol{A}-x\boldsymbol{C})^{-1}y,y\big)\Big)
\end{aligned}$$

where the matrix $\boldsymbol{B}$ is given by

$$\boldsymbol{B} = x(\boldsymbol{C}^{-1}-x\boldsymbol{A}^{-1})^{-1}\boldsymbol{A}^{-1}.$$

The conclusion follows immediately if we rewrite the matrices in the normalizing constant as,

$$(\boldsymbol{C}^{-1}-x\boldsymbol{A}^{-1}) = \boldsymbol{C}^{-1}(\boldsymbol{A}-x\boldsymbol{C})\boldsymbol{A}^{-1},$$

which implies

$$|\boldsymbol{C}^{-1}-x\boldsymbol{A}^{-1}| = |\boldsymbol{C}|^{-1}|\boldsymbol{A}-x\boldsymbol{C}||\boldsymbol{A}|^{-1}.$$

□

Using Lemma 4.2.4 to compute the recursion gives the appropriate closed form.

Corollary 4.2.1

If $|\boldsymbol{\Lambda}_0| > 0$, then

$$Y_{0,1} = -\frac{1}{2}\log|\boldsymbol{\Lambda}| + \frac{1}{2}\Big(\boldsymbol{\Lambda}_0^{-1}\vec{h},\vec{h}\Big) + \frac{1}{2}\sum_{0\le k\le r-1}\frac{1}{x_k}\log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|}. \tag{4.70}$$

Proof. Using Lemma 4.2.4 to compute the expectation of the second term in (4.69) recursively implies

$$Y_{r-1,1} = -\frac{1}{2}\log|\boldsymbol{\Lambda}| + \frac{1}{2x_r}\log\frac{|\boldsymbol{\Lambda}_r|}{|\boldsymbol{\Lambda}_r - x_{r-1}\boldsymbol{\Delta}_r|} \tag{4.71}$$

$$+\frac{1}{2}\Big((\boldsymbol{\Lambda}_r - x_{r-1}\boldsymbol{\Delta}_r)^{-1}\Big(\sum_{1\le k\le r-1}\vec{z}_{k,1}+\vec{h}\Big),\Big(\sum_{1\le k\le r-1}\vec{z}_{k,1}+\vec{h}\Big)\Big). \tag{4.72}$$

Again, the terms in (4.71) are non-random, so they propagate through the recursion. Computing the terms in (4.72) inductively using repeated applications of Lemma 4.2.4 implies

$$Y_{0,1} = -\frac{1}{2}\log|\boldsymbol{\Lambda}| + \frac{1}{2}\Big(\boldsymbol{\Lambda}_0^{-1}\vec{h},\vec{h}\Big) + \frac{1}{2}\sum_{0\le k\le r-1}\frac{1}{x_k}\log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|}. \tag{4.73}$$

□

Notice ν_N concentrates around the sphere of radius $\sqrt{N}$ in high dimensions, so the Gaussian term will vanish in the limit by the weak law of large numbers. That is, for any $\delta > 0$,

$$\lim_{N\to\infty}\frac{n\log\nu_N(E_N^\delta)}{N} = \lim_{N\to\infty}\frac{n\log(\mathbb{P}(|\frac{1}{N}\sum_{i=1}^N g_i^2 - 1| \le \delta))}{N} = 0, \tag{4.74}$$

where $g_1, g_2, \dots$ are i.i.d. standard normals. The other terms vanish by taking $\varepsilon \to 0$, so combining (4.70) and (4.74) with (4.68), gives the bound

$$\begin{aligned}&\lim_{\varepsilon\to 0}\limsup_{N\to\infty}\frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha\int_{Q_N^\varepsilon}\exp\Big(\sum_{i\le N}\sum_{j\le n}\vec{\sigma}_i(j)\big(Z_i^j(\alpha)+\vec{h}(j)\big)\Big)\,d\lambda_N^n(\vec{\boldsymbol{\sigma}})\\ &\le\frac{1}{2}\Big(\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) - n - \log|\boldsymbol{\Lambda}| + (\boldsymbol{\Lambda}_0^{-1}\vec{h},\vec{h}) + \sum_{0\le k\le r-1}\frac{1}{x_k}\log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|}\Big).\end{aligned} \tag{4.75}$$

The upper bound in (4.75) holds for all $\boldsymbol{\Lambda} \in \mathcal{L}$. Applying the bounds (4.75) and (4.55) to (4.52) and taking the infimum over all discrete paths encoded by the monotone sequences (4.42) and (4.43) shows

$$\lim_{\varepsilon\to 0}\limsup_{N\to\infty} F_N^\varepsilon(\vec{\boldsymbol{\beta}},\boldsymbol{Q}) \le \inf_{\boldsymbol{\Lambda},\pi}\mathscr{P}_{\vec{\boldsymbol{\beta}},\boldsymbol{Q}}(\boldsymbol{\Lambda},\pi), \tag{4.76}$$

completing the proof of the upper bound.

4.3 Sharpness of the Upper Bound

We now prove for every fixed path π, the upper bound (4.75) is asymptotically sharp in the sense that it attains equality after minimizing over $\boldsymbol{\Lambda}$. This fact will be used again when a similar functional appears in the proof of the lower bound. The proof of this sharpness for the replica symmetric case can be found in [83, Lemma 4]. We will provide a proof of the general case below.

Let π be any fixed discrete monotone path characterized by the sequences (4.42) and (4.43) and denote the functional appearing in (4.56) by

$$f_N^1(\pi) = \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha\int_{\Omega_N^{\varepsilon,\delta}}\exp\Big(\sum_{i\le N}\sum_{j\le n}\vec{\omega}_i(j)\big(Z_i^j(\alpha)+\vec{h}(j)\big)\Big)\,d\nu_N^n(\vec{\boldsymbol{\sigma}}). \tag{4.77}$$

We will prove the matching lower bound of (4.75) by decoupling the functional $f_N^1(\pi)$ from the constraint $\boldsymbol{Q}$ and explicitly computing its value recursively.

Lemma 4.3.1

For all $0 < \delta < \varepsilon$,

$$\liminf_{N\to\infty} f_N^1(\pi) \geq \inf_{\boldsymbol{\Lambda}} \frac{1}{2}\Big(\operatorname{tr}(\boldsymbol{\Lambda Q}) - n - \log|\boldsymbol{\Lambda}| + (\boldsymbol{\Lambda}_0^{-1}\vec{h},\vec{h}) + \sum_{0\leq k\leq r-1} \frac{1}{x_k}\log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|}\Big).$$

Recall (4.61), the subset of $\mathbb{R}^{Nn}$ constrained by coupling the overlaps,

$$\tilde{\Omega}_N^\delta = \big\{\, \vec{\boldsymbol{\omega}} \in (\mathbb{R}^N)^n \mid \|\boldsymbol{R}(\vec{\boldsymbol{\omega}},\vec{\boldsymbol{\omega}}) - \boldsymbol{Q}\|_\infty \leq \delta \,\big\}. \tag{4.78}$$

Clearly, there exists a $\delta^* < \varepsilon$ such that $\Omega_N^{\varepsilon,\delta} \supseteq \tilde{\Omega}_N^{\delta^*}$, so

$$f_N^1(\pi) \geq \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{\tilde{\Omega}_N^{\delta^*}} \exp\Big(\sum_{i\leq N}\sum_{j\leq n} \vec{\omega}_i(j)\big(Z_i^j(\alpha) + \vec{h}(j)\big)\Big)\, d\nu_N^n(\vec{\boldsymbol{\omega}}). \tag{4.79}$$

We introduce the Lagrange multipliers $\boldsymbol{\Lambda} \in \mathcal{L}$ defined in (4.36). Like in the proof of the upper bound, since $\|\boldsymbol{R}(\vec{\boldsymbol{\omega}},\vec{\boldsymbol{\omega}}) - \boldsymbol{Q}\|_\infty \leq \delta^*$ for $\vec{\boldsymbol{\omega}} \in \tilde{\Omega}_N^{\delta^*}$, adding and subtracting the quadratic form $\frac{1}{2}\sum_{i\leq N}((\boldsymbol{\Lambda} - \boldsymbol{I})\vec{\omega}_i,\vec{\omega}_i)$ from the exponent implies (4.79) is bounded below by

$$\begin{aligned}&\frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_{\tilde{\Omega}_N^{\delta^*}} \exp\Big(\sum_{i\leq N}\sum_{j\leq n} \vec{\omega}_i(j)\big(Z_i^j(\alpha) + \vec{h}(j)\big) - \frac{1}{2}\sum_{i\leq N}((\boldsymbol{\Lambda} - \boldsymbol{I})\vec{\omega}_i,\vec{\omega}_i)\Big)\, d\nu_N^n(\vec{\boldsymbol{\omega}})\\ &+ \frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda Q}) - \frac{n}{2} - \delta^*\|\boldsymbol{\Lambda}\|_1.\end{aligned} \tag{4.80}$$

We view the quantity on the first line of (4.80) as a function of $\boldsymbol{\Lambda}$ and the region of integration. In general, we denote this integral over sets $V \subset (\mathbb{R}^N)^n$ by

$$\Phi_V(\boldsymbol{\Lambda}) = \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha \int_V \exp\Big(\sum_{i\leq N}\sum_{j\leq n} \vec{\omega}_i(j)\big(Z_i^j(\alpha) + \vec{h}(j)\big) - \frac{1}{2}\sum_{i\leq N}((\boldsymbol{\Lambda} - \boldsymbol{I})\vec{\omega}_i,\vec{\omega}_i)\Big)\, d\nu_N^n(\vec{\boldsymbol{\omega}}) \tag{4.81}$$

and the integral over the whole space by

$$F(\boldsymbol{\Lambda}) := \Phi_{\mathbb{R}^{Nn}}(\boldsymbol{\Lambda}). \tag{4.82}$$

The map $V \mapsto \Phi_V(\boldsymbol{\Lambda})$ is monotone and, in particular, $\Phi_V(\boldsymbol{\Lambda}) \leq F(\boldsymbol{\Lambda})$. Furthermore, the function $F(\boldsymbol{\Lambda})$ does not depend on N, and was computed using the recursion (4.67) giving the closed form in Corollary 4.2.1,

$$F(\boldsymbol{\Lambda}) = \frac{1}{2}\Big(- \log|\boldsymbol{\Lambda}| + \Big(\boldsymbol{\Lambda}_0^{-1}\vec{h},\vec{h}\Big) + \sum_{0\leq k\leq r-1} \frac{1}{x_k}\log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|}\Big). \tag{4.83}$$

We will prove that minimizing over $\boldsymbol{\Lambda}$ removes the dependence on the constraint $\boldsymbol{Q}$ asymptotically. We start by showing there exists a unique $\boldsymbol{\Lambda}_*$ that minimizes $\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda Q}) + F(\boldsymbol{\Lambda})$ if the lower bound (4.92) is finite.

Lemma 4.3.2

Given a positive semi-definite constraint $\boldsymbol{Q}$:

1. *If $\boldsymbol{Q}$ is degenerate, then*
$$\inf_{\boldsymbol{\Lambda}} \Big(\frac{1}{2} \operatorname{tr}(\boldsymbol{\Lambda Q}) + F(\boldsymbol{\Lambda}) \Big) = -\infty. \tag{4.84}$$

2. *If $\boldsymbol{Q}$ is non-degenerate, then there exists a $\boldsymbol{\Lambda}_* \in \mathcal{L}$ that minimizes $\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda Q}) + F(\boldsymbol{\Lambda})$ and satisfies*
$$\frac{\partial}{\partial t} \Big(\frac{1}{2} \operatorname{tr}\big((\boldsymbol{\Lambda}_* + t\boldsymbol{B})\boldsymbol{Q}\big) + F(\boldsymbol{\Lambda}_* + t\boldsymbol{B}) \Big) \Big|_{t=0} = 0 \tag{4.85}$$
for any symmetric matrix $\boldsymbol{B}$.

Proof. Consider the eigendecomposition of $\boldsymbol{\Lambda} \in \mathcal{L}$,

$$\boldsymbol{\Lambda} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}}.$$

Using this change of variables and (4.35), we see (4.83) can be rewritten in terms of $\boldsymbol{P}$ and $\boldsymbol{D}$ as,

$$\begin{aligned} \frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda Q}) + F(\boldsymbol{\Lambda}) = \frac{1}{2}\Bigg(& \operatorname{tr}\big(\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}}\boldsymbol{Q}\boldsymbol{P}\big) - \log|\boldsymbol{D}| \\ & + \Big(\Big(\boldsymbol{D} - \sum_{0 \le k < r} x_k \boldsymbol{P}^{\mathsf{T}} \boldsymbol{\Delta}_{k+1} \boldsymbol{P} \Big)^{-1} (\boldsymbol{P}^{\mathsf{T}}\vec{h}), (\boldsymbol{P}^{\mathsf{T}}\vec{h}) \Big) \qquad (4.86) \\ & + \sum_{0 \le k \le r-1} \frac{1}{x_k} \log \frac{|\boldsymbol{D} - \sum_{k+1 \le \ell < r} x_\ell \boldsymbol{P}^{\mathsf{T}} \boldsymbol{\Delta}_{\ell+1} \boldsymbol{P}|}{|\boldsymbol{D} - \sum_{k \le \ell < r} x_\ell \boldsymbol{P}^{\mathsf{T}} \boldsymbol{\Delta}_{\ell+1} \boldsymbol{P}|} \Bigg). \qquad (4.87) \end{aligned}$$

In this form, the infimum is over positive semidefinite diagonal matrices $\boldsymbol{D}$ and orthogonal matrices $\boldsymbol{P}$ such that $|\boldsymbol{D} - \sum_{0 \le \ell < r} x_\ell \boldsymbol{P}^{\mathsf{T}} \boldsymbol{\Delta}_{\ell+1} \boldsymbol{P}| > 0$,

$$\inf_{\boldsymbol{\Lambda}} \Big(\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda Q}) + F(\boldsymbol{\Lambda}) \Big) = \inf_{\boldsymbol{D},\boldsymbol{P}} \Big(\frac{1}{2}\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}}\boldsymbol{Q}) + F(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}}) \Big).$$

Case (1)*:* Suppose $\boldsymbol{Q}$ is degenerate, i.e. $|\boldsymbol{Q}| = 0$. There exists an orthogonal matrix $\boldsymbol{P}$, corresponding to the eigendecomposition of $\boldsymbol{Q}$, such that $\tilde{\boldsymbol{D}} = \boldsymbol{P}^{\mathsf{T}}\boldsymbol{Q}\boldsymbol{P}$ and $\tilde{\boldsymbol{D}}_{11} = 0$. Given this $\boldsymbol{P}$, we choose diagonal matrix $\boldsymbol{D}$ with diagonal entries large enough such that all the Gershgorin discs of $\boldsymbol{D} - \sum_{k \le \ell < r} x_\ell \boldsymbol{P}^{\mathsf{T}} \boldsymbol{\Delta}_{\ell+1} \boldsymbol{P}$ are contained in the positive real half plane for all $k \le r-1$. In particular for all $k \le r-1$, the smallest eigenvalue of $\boldsymbol{D} - \sum_{k \le \ell < r} x_\ell \boldsymbol{P}^{\mathsf{T}} \boldsymbol{\Delta}_{\ell+1} \boldsymbol{P}$ is strictly positive and will remain bounded away from zero if we increase the value of the first diagonal element. That is, there exists a $c > 0$ such that

$$\liminf_{D_{11} \to \infty} \lambda_{min} \Big(\boldsymbol{D} - \sum_{k \le \ell < r} x_\ell \boldsymbol{P}^{\mathsf{T}} \boldsymbol{\Delta}_{\ell+1} \boldsymbol{P} \Big) \ge c > 0 \text{ for all } k \le r-1. \tag{4.88}$$

We fix all entries $\boldsymbol{D}_{jj}$ for $2 \le j \le n$ and show (4.87) diverges to $-\infty$ as we take the first entry $\boldsymbol{D}_{11} \to \infty$.

For the above choice of $\boldsymbol{D}$ and $\boldsymbol{P}$, we have

$$\begin{aligned}\operatorname{tr}\big(\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}}\boldsymbol{Q}\boldsymbol{P}\big) - \log|\boldsymbol{D}| &= \sum_{i=1}^{n} \boldsymbol{D}_{ii}(\boldsymbol{P}^{\mathsf{T}}\boldsymbol{Q}\boldsymbol{P})_{ii} - \sum_{i=1}^{n} \log \boldsymbol{D}_{ii} \\ &= \sum_{i=2}^{n} \boldsymbol{D}_{ii}(\boldsymbol{P}^{\mathsf{T}}\boldsymbol{Q}\boldsymbol{P})_{ii} - \sum_{i=2}^{n} \log \boldsymbol{D}_{ii} - \log \boldsymbol{D}_{11},\end{aligned}$$

which implies

$$\lim_{D_{11}\to\infty} \Big(\operatorname{tr}\big(\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}}\boldsymbol{Q}\boldsymbol{P}\big) - \log|\boldsymbol{D}| \Big) = -\infty. \tag{4.89}$$

We will now show that the remaining terms of (4.87) are finite. Let ν_{min}^{k} denote the smallest eigenvalue of $\boldsymbol{D} - \sum_{k\le\ell<r} x_\ell \boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Delta}_{\ell+1}\boldsymbol{P}$. By (4.88), we have $\liminf_{D_{11}\to\infty} \nu_{min}^{k} \ge c$ for all $k \le r-1$. Bounding the quadratic form with the largest eigenvalue of the associated matrix implies

$$\lim_{D_{11}\to\infty} \Big(\Big(\boldsymbol{D} - \sum_{0\le k<r} x_k \boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Delta}_{k+1}\boldsymbol{P} \Big)^{-1} (\boldsymbol{P}^{\mathsf{T}}\vec{h}), (\boldsymbol{P}^{\mathsf{T}}\vec{h}) \Big) \le \big(\nu_{min}^{0}\big)^{-1} \|\vec{h}\|^2 \le c^{-1}\|\vec{h}\|^2 < \infty. \tag{4.90}$$

The logarithm terms in (4.87) can be bounded by the minimum eigenvalues in a similar manner. It suffices to show an arbitrary term in the sum is bounded, that is,

$$\lim_{D_{11}\to\infty} \frac{1}{x_k} \log \frac{|\boldsymbol{D} - \sum_{k+1\le\ell<r} x_\ell \boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Delta}_{\ell+1}\boldsymbol{P}|}{|\boldsymbol{D} - \sum_{k\le\ell<r} x_\ell \boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Delta}_{\ell+1}\boldsymbol{P}|} < \infty. \tag{4.91}$$

If we define the matrices $\boldsymbol{A}_k := \boldsymbol{D} - \sum_{k\le\ell<r} x_\ell \boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Delta}_{\ell+1}\boldsymbol{P}$ and $\boldsymbol{B}_k := x_k \boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Delta}_{k+1}\boldsymbol{P}$, then

$$\log \frac{|\boldsymbol{D} - \sum_{k+1\le\ell<r} x_\ell \boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Delta}_{\ell+1}\boldsymbol{P}|}{|\boldsymbol{D} - \sum_{k\le\ell<r} x_\ell \boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Delta}_{\ell+1}\boldsymbol{P}|} = \log \frac{|\boldsymbol{A}_k + \boldsymbol{B}_k|}{|\boldsymbol{A}_k|} = \log|\boldsymbol{A}_k^{-1}(\boldsymbol{A}_k + \boldsymbol{B}_k)| = \log|\boldsymbol{I} + \boldsymbol{A}_k^{-1}\boldsymbol{B}_k|.$$

Bounding this with the largest eigenvalue, we see

$$\log|\boldsymbol{I} + \boldsymbol{A}_k^{-1}\boldsymbol{B}_k| \le n \log \lambda_{max}(\boldsymbol{I} + \boldsymbol{A}_k^{-1}\boldsymbol{B}_k).$$

Using submultiplicativity of the spectral norm and the lower bound on the smallest eigenvalue of $\boldsymbol{A}_k$ in (4.88), we have

$$\lambda_{max}(\boldsymbol{I} + \boldsymbol{A}_k^{-1}\boldsymbol{B}_k) = 1 + \lambda_{max}(\boldsymbol{A}_k^{-1}\boldsymbol{B}_k) \le 1 + \lambda_{max}(\boldsymbol{A}_k^{-1})\lambda_{max}(\boldsymbol{B}_k) \le 1 + c^{-1}\lambda_{max}(\boldsymbol{B}_k) < \infty,$$

giving the required bound in (4.91).

Therefore, for a particular $\boldsymbol{P}$, we can construct a sequence of diagonal matrices $\boldsymbol{D}$ with arbitrary large first diagonal element such that $\frac{1}{2}\operatorname{tr}(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}}\boldsymbol{Q}) + F(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathsf{T}})$ is unbounded. In particular, we have

$$\inf_{\boldsymbol{\Lambda}} \Big(\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) + F(\boldsymbol{\Lambda}) \Big) = -\infty.$$

Case (2)*:* Consider the case when $\boldsymbol{Q}$ is positive definite. We will prove that (4.83) attains a minimum at some point $\boldsymbol{\Lambda}^* \in \mathcal{L}$. By Hölder's inequality, $F(\boldsymbol{\Lambda})$ is a convex function of $\boldsymbol{\Lambda}$, so any local minimizer is also a global minimizer. We will prove that the minimizer is attained in a compact subset of Γ_n under the spectral norm on symmetric matrices $\|\boldsymbol{\Lambda}\|_2 = \lambda_{max}(\boldsymbol{\Lambda})$.

Because $\boldsymbol{Q}$ is positive definite, the diagonal elements of $\boldsymbol{P}^\intercal \boldsymbol{Q}\boldsymbol{P}$ is positive and uniformly bounded away from 0 for all orthogonal matrices $\boldsymbol{P}$. That is, the first term in (4.87) can be bounded below by

$$\operatorname{tr}(\boldsymbol{D}\boldsymbol{P}^\intercal \boldsymbol{Q}\boldsymbol{P}) - \log|\boldsymbol{D}| = \sum_{j\le n}\Big(\boldsymbol{D}_{jj}(\boldsymbol{P}^\intercal \boldsymbol{Q}\boldsymbol{P})_{jj} - \log|\boldsymbol{D}_{jj}|\Big) \ge \sum_{j\le n}\Big(\boldsymbol{D}_{jj}\lambda_{min}(\boldsymbol{Q}) - \log|\boldsymbol{D}_{jj}|\Big)$$

which clearly diverges to ∞ if any diagonal element $\boldsymbol{D}_{jj} \to \infty$. The remaining terms in (4.87) are non-negative, so $\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) + F(\boldsymbol{\Lambda}) \to \infty$ if $\|\boldsymbol{\Lambda}\|_2 \to \infty$. Since $\boldsymbol{\Lambda} - \boldsymbol{\Lambda}_0 \ge 0$, we also have $\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) + F(\boldsymbol{\Lambda}) \to \infty$ if $\|\boldsymbol{\Lambda}_0\|_2 \to \infty$ by monotonicity.

By definition (4.35), we have $\boldsymbol{\Lambda}_{k+1} = \boldsymbol{\Lambda}_k + x_k\boldsymbol{\Delta}_{k+1}$. By submultiplicativity of $\|\cdot\|_2$,

$$\|\boldsymbol{\Delta}_{k+1}\|_2 = \|\boldsymbol{\Lambda}_k\boldsymbol{\Lambda}_k^{-1}\boldsymbol{\Delta}_{k+1}\|_2 \le \|\boldsymbol{\Lambda}_k\|_2\|\boldsymbol{\Lambda}_k^{-1}\boldsymbol{\Delta}_{k+1}\|_2,$$

so the last term in (4.83) can be bounded below by

$$\log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|} = \log|\boldsymbol{I} + x_k\boldsymbol{\Lambda}_k^{-1}\boldsymbol{\Delta}_{k+1}| \ge \log(1 + x_k\|\boldsymbol{\Lambda}_k^{-1}\boldsymbol{\Delta}_{k+1}\|_2) \ge \log(1 + x_k\|\boldsymbol{\Lambda}_k\|_2^{-1}\|\boldsymbol{\Delta}_{k+1}\|_2).$$

Let k^* be the smallest index such that $\boldsymbol{\Delta}_{k^*+1} \ne \boldsymbol{0}$, then it is clear the above term diverges as $\|\boldsymbol{\Lambda}_{k^*}\|_2 \to 0$. Since the sequence (4.43) is monotone, we have $\|\boldsymbol{\Lambda}_0\|_2 = \|\boldsymbol{\Lambda}_{k^*}\|_2$. The rest of the terms in (4.83) are bounded or positive for fixed π, so $\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) + F(\boldsymbol{\Lambda}) \to \infty$ if $\|\boldsymbol{\Lambda}_0\| \to 0$.

We have shown, $\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) + F(\boldsymbol{\Lambda})$ is unbounded if $\|\boldsymbol{\Lambda}_0\|_2 \to 0$ or $\|\boldsymbol{\Lambda}_0\|_2 \to \infty$. Therefore, there exists a $0 < c < C < \infty$ such that the minimizer is attained in the compact set

$$\mathcal{L}^* = \{\boldsymbol{\Lambda} \in \Gamma_n \mid c \le \|\boldsymbol{\Lambda}_0\|_2 \le C\} \subseteq \mathcal{L}.$$

The matrix $\boldsymbol{B} = \sum_{0\le k\le r} x_k\boldsymbol{\Delta}_k$ is a fixed positive semidefinite matrix, so the map $f : \Gamma_n \to \mathbb{R}$ defined by $f(\boldsymbol{\Lambda}) = \|\boldsymbol{\Lambda} - \boldsymbol{B}\|_2 = \|\boldsymbol{\Lambda}_0\|_2$ is continuous. Therefore, $\mathcal{L}^* = f^{-1}([c, C])$ is closed and clearly bounded, so it is compact. By the extreme value theorem, there exists a $\boldsymbol{\Lambda}_* \in \mathcal{L}$ such that $\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) + F(\boldsymbol{\Lambda})$ attains its minimum at $\boldsymbol{\Lambda}_*$. Furthermore, our function is convex and $\boldsymbol{\Lambda}_*$ is an interior point of $\mathcal{L}$, so the minimizer $\boldsymbol{\Lambda}_*$ is unique and satisfies the critical point condition,

$$\frac{\partial}{\partial t}\Big(\frac{1}{2}\operatorname{tr}\big((\boldsymbol{\Lambda}_* + t\boldsymbol{B})\boldsymbol{Q}\big) + F(\boldsymbol{\Lambda}_* + t\boldsymbol{B})\Big)\Big|_{t=0} = 0,$$

for all symmetric matrices $\boldsymbol{B}$. □

Lemma 4.3.1 is trivially satisfied when the infimum is $-\infty$, so we focus on the non-degenerate case moving forward. We now prove asymptotic sharpness of the upper bound using a standard large deviations calculation to decouple the constraints.

Lemma 4.3.3

For any $\delta^* > 0$ and positive definite $\boldsymbol{Q}$,

$$\liminf_{N\to\infty} \Phi_{\tilde{\Omega}_N^{\delta^*}}(\boldsymbol{\Lambda}) + \frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) \ge \inf_{\boldsymbol{\Lambda}}\Big(\frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) + F(\boldsymbol{\Lambda})\Big). \tag{4.92}$$

Proof. Consider the partition

$$(\mathbb{R}^N)^n = \tilde{\Omega}_N^{\delta^*} \cup \Big(\bigcup_{j,j' \le n} V_{j,j'}^+ \Big) \cup \Big(\bigcup_{j,j' \le n} V_{j,j'}^- \Big)$$

where

$$V_{j,j'}^+ = \Big\{ \vec{\boldsymbol{\omega}} \mid R^{j,j'}(\vec{\boldsymbol{\omega}}, \vec{\boldsymbol{\omega}}) \ge Q^{j,j'} + \delta^* \Big\}, \tag{4.93}$$

$$V_{j,j'}^- = \Big\{ \vec{\boldsymbol{\omega}} \mid R^{j,j'}(\vec{\boldsymbol{\omega}}, \vec{\boldsymbol{\omega}}) \le Q^{j,j'} - \delta^* \Big\}. \tag{4.94}$$

Recall the monotone function Φ_V defined in (4.81). For $\boldsymbol{\Lambda}_*$ satisfying (4.85), by considering values near this critical point, we will show there exists a constant $c > 0$ such that for all half-spaces V in (4.93) or (4.94)

$$\Phi_V(\boldsymbol{\Lambda}_*) \le F(\boldsymbol{\Lambda}_*) - c. \tag{4.95}$$

We only show this for $V = V_{j,j'}^-$ for $j \ne j'$. The proof for the other cases are similar. For all $t \ge 0$ and $\vec{\boldsymbol{\omega}} \in V_{j,j'}^-$,

$$tR^{j,j'}(\vec{\boldsymbol{\omega}}, \vec{\boldsymbol{\omega}}) \le t(Q^{j,j'} - \delta^*).$$

Let $\boldsymbol{B}$ be a matrix such that $B^{j,j'} = B^{j',j} = 1$ and is zero everywhere else. Adding and subtracting $\frac{1}{2}tNR^{j,j'}(\vec{\boldsymbol{\omega}}, \vec{\boldsymbol{\omega}})$ and $\frac{1}{2}tNR^{j',j}(\vec{\boldsymbol{\omega}}, \vec{\boldsymbol{\omega}})$ in the exponent, by symmetry of $\boldsymbol{Q}$, we have

$$\begin{aligned}
\Phi_V(\boldsymbol{\Lambda}_*) &\le t(Q^{j,j'} - \delta^*) + \Phi_V(\boldsymbol{\Lambda}_* + t\boldsymbol{B}) \\
&\le t(Q^{j,j'} - \delta^*) + F(\boldsymbol{\Lambda}_* + t\boldsymbol{B}) \\
&= -t\delta^* - \frac{1}{2}\operatorname{tr}(\boldsymbol{\Lambda}_*\boldsymbol{Q}) + \frac{1}{2}\operatorname{tr}((\boldsymbol{\Lambda}_* + t\boldsymbol{B})\boldsymbol{Q}) + F(\boldsymbol{\Lambda}_* + t\boldsymbol{B}) =: U(t).
\end{aligned} \tag{4.96}$$

Since $U(0) = F(\boldsymbol{\Lambda}_*)$, the critical point condition (4.85) implies $U'(0) = -\delta^*$. In particular, there is a t^* such that $U(t^*) < U(0)$. Since (4.96) holds for all $t > 0$, there is a c such that

$$\Phi_V(\boldsymbol{\Lambda}_*) \le U(t^*) \le U(0) - c = F(\boldsymbol{\Lambda}_*) - c.$$

Recall the sets (4.93), (4.94), and (4.78) form a partition of $(\mathbb{R}^N)^n$. A consequence of the recursion in the Ruelle probability cascades (see equation (118) in the proof of [80, Lemma 7] or the proof of [81, Lemma 6]) implies

$$F(\boldsymbol{\Lambda}_*) \le \frac{\log(2n^2+1)}{Nx_0} + \max\Big(\max_V \Phi_V(\boldsymbol{\Lambda}_*), \Phi_{\tilde{\Omega}_N^{\delta^*}}(\boldsymbol{\Lambda}_*) \Big)$$

where the maximum over V is over the halfspaces of the form (4.93), (4.94). Our bounds in (4.95) ensures we cannot have

$$F(\boldsymbol{\Lambda}_*) \le \frac{\log(2n^2+1)}{Nx_0} + \max_V \Phi_V(\boldsymbol{\Lambda}_*)$$

for N sufficiently large. Therefore, we must have

$$F(\boldsymbol{\Lambda}_*) \le \frac{\log(2n^2+1)}{Nx_0} + \Phi_{\tilde{\Omega}_N^{\delta^*}}(\boldsymbol{\Lambda}_*).$$

Taking $N \to \infty$ completes the proof. □

The proof of Lemma 4.3.1 follows by applying Lemma 4.3.3 to (4.80),

$$\lim_{\varepsilon\to 0}\liminf_{N\to\infty} f_N^1(\pi) \geq \inf_{\boldsymbol{\Lambda}} \frac{1}{2}\bigg(\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) - n - \log|\boldsymbol{\Lambda}| + (\boldsymbol{\Lambda}_0^{-1}\vec{h},\vec{h}) + \sum_{0\leq k\leq r-1} \frac{1}{x_k}\log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|}\bigg).$$

We have shown that Theorem 4.1.1 is trivially satisfied for degenerate constraint $\boldsymbol{Q}$ just from examining the upper bound. The case for positive definite constraint $\boldsymbol{Q}$ is much harder and will require some preliminary work before attempting the cavity computations. We begin by introducing a variant of the Aizenman–Sims–Starr scheme.

4.4 The Aizenman–Sims–Starr Scheme

Before we can complete the cavity computations to prove the lower bound, we first prove an analogue of the Aizenman–Sims–Starr scheme [2] for spherical spin glass models with vector spins. The extension to this model is non-trivial because the uniform measure on the sphere is not a product measure, so the usual proof of the scheme fails.

This section follows the proof of the Aizenman–Sims–Starr scheme adapted for spherical models in [30]. The main difference is the Aizenman–Sims–Starr scheme (see Lemma 4.5.2) will be with respect to a Gaussian reference measure as opposed to the surface measure in [30]. This form was chosen for convenience, because it matches the form of the functional (4.63).

To simplify notation, we first prove an analogue of the Aizenman–Sims–Starr scheme with no external field. We will explain how to reintroduce the external field at the end of Section 4.5. Consider the partition function with $\vec{h} = \vec{0}$ for a system of size N,

$$Z_N(\boldsymbol{Q},\varepsilon) = \int_{Q_N^\varepsilon} \exp\big(H_N(\vec{\boldsymbol{\sigma}})\big)\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}) \tag{4.97}$$

and the corresponding partition function for a system of size $M+N$,

$$Z_{M+N}(\boldsymbol{Q},\varepsilon) = \int_{Q_{M+N}^\varepsilon} \exp\big(H_{N+M}(\vec{\boldsymbol{\rho}})\big)\, d\lambda_{M+N}^n(\vec{\boldsymbol{\rho}}). \tag{4.98}$$

We denote spin configurations from the system of size $M+N$ with $\vec{\boldsymbol{\rho}} = (\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\omega}}) \in S_{M+N}^n$ where $\vec{\boldsymbol{\sigma}} \in \mathbb{R}^N$ denotes the bulk coordinates and $\vec{\boldsymbol{\omega}} \in \mathbb{R}^M$ denotes the cavity coordinates.

We proceed like the traditional Aizenman–Sims–Starr scheme and split the Hamiltonian into the cavity fields [76, Section 3.5]

$$H_{M+N}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\omega}}) \stackrel{d}{=} \sum_{j\leq n} H_{M,N}^j(\vec{\boldsymbol{\sigma}}) + \sum_{i\leq M}\sum_{j\leq n} \vec{\omega}_i(j) Z_i^j(\vec{\boldsymbol{\sigma}}) + r(\vec{\boldsymbol{\rho}}), \tag{4.99}$$

$$H_N(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\omega}}) \stackrel{d}{=} \sum_{j\leq n} H_{M,N}^j(\vec{\boldsymbol{\sigma}}) + \sqrt{M}\sum_{j\leq n} Y^j(\vec{\boldsymbol{\sigma}}). \tag{4.100}$$

Here, $H_{M,N}(\vec{\boldsymbol{\sigma}}) := \sum_{j\leq n} H_{M,N}^j(\vec{\boldsymbol{\sigma}})$ is defined like $H_N(\vec{\boldsymbol{\sigma}})$ but with normalization $(M+N)^{-(p-1)/2}$. The covariance of this Hamiltonian is given by

$$\mathbb{E} H_{M,N}^j(\vec{\boldsymbol{\sigma}}^\ell) H_{M,N}^{j'}(\vec{\boldsymbol{\sigma}}^{\ell'}) = (M+N)\xi_{j,j'}\Big(\frac{N}{M+N} R_{\ell,\ell'}^{j,j'}\Big). \tag{4.101}$$

The cavity fields $Z(\vec{\boldsymbol{\sigma}})$ and $Y(\vec{\boldsymbol{\sigma}})$ in (4.99) and (4.100) are centered Gaussian processes with covariances:

$$\mathbb{E}Z_i^j(\vec{\boldsymbol{\sigma}}^\ell)Z_{i'}^{j'}(\vec{\boldsymbol{\sigma}}^{\ell'}) = \delta_{i,i'}\xi'_{j,j'}\big(R_{\ell,\ell'}^{j,j'}\big) + \mathscr{O}\Big(\frac{M}{N}\Big), \tag{4.102}$$

$$\mathbb{E}Y^j(\vec{\boldsymbol{\sigma}}^\ell)Y^{j'}(\vec{\boldsymbol{\sigma}}^{\ell'}) = \theta_{j,j'}\big(R_{\ell,\ell'}^{j,j'}\big) + \mathscr{O}\Big(\frac{M}{N}\Big), \tag{4.103}$$

and the remainder term $r(\vec{\boldsymbol{\rho}})$ has covariance,

$$\mathbb{E}r(\vec{\boldsymbol{\rho}}^{\,\ell})r(\vec{\boldsymbol{\rho}}^{\,\ell'}) = \mathscr{O}\Big(\frac{M^2}{M+N}\Big). \tag{4.104}$$

We will prove that we can replace the cavity fields $Z(\vec{\boldsymbol{\sigma}})$ and $Y(\vec{\boldsymbol{\sigma}})$ with centered Gaussian fields $z_i(\vec{\boldsymbol{\sigma}})$ and $y(\vec{\boldsymbol{\sigma}})$ taking values in $\mathbb{R}^n$ indexed by $\vec{\boldsymbol{\sigma}} \in S_N^n$, with covariances

$$\mathbb{E}z_i^j(\vec{\boldsymbol{\sigma}}^\ell)z_{i'}^{j'}(\vec{\boldsymbol{\sigma}}^{\ell'}) = \delta_{i,i'}\xi'_{j,j'}\big(R_{\ell,\ell'}^{j,j'}\big), \tag{4.105}$$

$$\mathbb{E}y^j(\vec{\boldsymbol{\sigma}}^\ell)y^{j'}(\vec{\boldsymbol{\sigma}}^{\ell'}) = \theta_{j,j'}\big(R_{\ell,\ell'}^{j,j'}\big). \tag{4.106}$$

Let $\langle\cdot\rangle_{M,N}$ be the average with respect to the Gibbs measure,

$$G_{M,N}(d\vec{\boldsymbol{\sigma}}) = \frac{\exp H_{M,N}(\vec{\boldsymbol{\sigma}})\, d\lambda_N^n(\vec{\boldsymbol{\sigma}})}{Z_{M,N}(\boldsymbol{Q},\varepsilon)}, \tag{4.107}$$

on Q_N^ε, with normalization

$$Z_{M,N}(\boldsymbol{Q},\varepsilon) = \int_{Q_N^\varepsilon} \exp\big(H_{M,N}(\vec{\boldsymbol{\sigma}})\big)\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}). \tag{4.108}$$

We start as usual with the inequality

$$\begin{aligned}\liminf_{N\to\infty}\frac{1}{N}\mathbb{E}\log Z_N(\boldsymbol{Q},\varepsilon) &\geq \frac{1}{M}\liminf_{N\to\infty}\big(\mathbb{E}\log Z_{M+N}(\boldsymbol{Q},\varepsilon) - \mathbb{E}\log Z_N(\boldsymbol{Q},\varepsilon)\big)\\ &= \frac{1}{M}\liminf_{N\to\infty}\left(\mathbb{E}\log\frac{Z_{M+N}(\boldsymbol{Q},\varepsilon)}{Z_{M,N}(\boldsymbol{Q},\varepsilon)} - \mathbb{E}\log\frac{Z_N(\boldsymbol{Q},\varepsilon)}{Z_{M,N}(\boldsymbol{Q},\varepsilon)}\right).\end{aligned} \tag{4.109}$$

The surface measure λ_{M+N} appearing in Z_{M+N} is not a product measure, so the standard proof of the Aizenman–Sims–Starr does not apply after this point. Instead, recall that the δ shell around Q_M^ε is denoted by

$$\Omega_M^{\varepsilon,\delta} = \big\{\vec{\boldsymbol{\omega}} = (s_j\boldsymbol{\tau}(j))_{j\leq n} \in (\mathbb{R}^M)^n \mid \vec{\boldsymbol{\tau}} \in Q_M^\varepsilon,\ s_j \in [\sqrt{1-\delta}, \sqrt{1+\delta}] \text{ for all } j \leq n\big\}, \tag{4.110}$$

where $s_j \in \mathbb{R}^+$ and $\tau(j) \in S_M$ are the radial and angular components of the polar form of $\boldsymbol{\omega}(j)$ (see after (4.55) for the formulas). We will prove the following lower bound of (4.109).

Lemma 4.4.1

Let ν_M be the standard normal distribution on $\mathbb{R}^M$. There exists a constant L such that for any $\varepsilon > 0$ and $M \geq 1$, there exists a $\delta \in (0, \varepsilon)$ such that (4.109) is bounded below by

$$\frac{1}{M} \liminf_{N\to\infty} \Bigg(\mathbb{E} \log \Big\langle \int_{\Omega_M^{\varepsilon/2,\delta}} \exp\Big(\sum_{i\leq M} \sum_{j\leq n} \vec{\omega}_i(j) z_i^j(\vec{\boldsymbol{\sigma}}) \Big) \, d\nu_M^n(\vec{\boldsymbol{\omega}}) \Big\rangle_{M,N} \\ - \mathbb{E} \log \Big\langle \exp \sqrt{M} y(\vec{\boldsymbol{\sigma}}) \Big\rangle_{M,N} \Bigg) - L\delta. \tag{4.111}$$

The main difference between the bound (4.111) and the traditional Aizenman–Sims–Starr representation is the Gaussian reference measure appearing in the first cavity field. This measure appears as a consequence of the Poincaré limit, which states that the standard Gaussian measure in $\mathbb{R}^M$ is the limiting distribution of projected uniform distributions on S_{N+M} as N tends to infinity.

4.4.1 Poincaré limit

We first explain a method to asymptotically decouple λ_{M+N} into an approximate product measure over the spheres $S_N \times S_M$. The distribution of the projection of S_{N+M} onto $\mathbb{R}^M$ under λ_{N+M} converges weakly to the Gaussian distribution ν_M on $\mathbb{R}^M$ in the Poincaré limit [88]. In particular, the distribution of the cavity coordinates under the normalized surface measure will be approximately Gaussian for large N. For large M, ν_M will concentrate around S_M. We first introduce some notation and state this result in one dimension.

For $K \geq 1$, we denote the unit sphere in $\mathbb{R}^K$ with S_K^1 and $|S_K^1|$ its surface area. Let

$$A_{M,N} = \prod_{j=1}^{M} \Big[-\sqrt{M+N+1-j}, \sqrt{M+N+1-j} \Big], \tag{4.112}$$

be a subset of $\mathbb{R}^M$ representing the domain of the cavity coordinates. We define the density on $\mathbb{R}^M$,

$$d\nu_{M,N}(\boldsymbol{x}) = f_{M,N}(\boldsymbol{x})\, d\boldsymbol{x},$$

where

$$f_{M,N}(\boldsymbol{x}) = b_{M,N} \prod_{j=1}^{M} \left(1 - \frac{x_j^2}{M+N+1-j} \right)^{\frac{M+N-j-2}{2}}, \tag{4.113}$$

with normalizing coefficient

$$b_{M,N} = \prod_{j=1}^{M} \frac{|S_{M+N-j}^1|}{|S_{M+N+1-j}^1| \sqrt{M+N+1-j}}.$$

The pointwise limit of (4.113) converges to the standard normal distribution on $\mathbb{R}^M$

$$d\nu_M(\boldsymbol{x}) = f_M(\boldsymbol{x})\, d\boldsymbol{x},$$

where

$$f_M(\boldsymbol{x}) := \lim_{N\to\infty} f_{M,N}(\boldsymbol{x}) = \left(\frac{1}{2\pi}\right)^{M/2} \exp\left(-\frac{\|\boldsymbol{x}\|^2}{2}\right).$$

Lastly, we define the coefficients

$$a_1 = 1, \quad a_\ell(\boldsymbol{x}) = \prod_{j=1}^{\ell-1} \sqrt{1 + \frac{1-x_j^2}{M+N-j}} \text{ for } 1 < \ell \leq M+1 \tag{4.114}$$

and the corresponding map for $\psi : S_N \times A_{M,N} \to S_{M+N}$ given by

$$\psi(\boldsymbol{\sigma}, \boldsymbol{\omega}) = \big(\sigma_1 a_{M+1}(\boldsymbol{\omega}), \ldots, \sigma_N a_{M+1}(\boldsymbol{\omega}), \omega_1 a_1(\boldsymbol{\omega}), \ldots, \omega_M a_M(\boldsymbol{\omega})\big)$$

for $\boldsymbol{\sigma} \in S_N$ and $\boldsymbol{\omega} \in A_{M,N}$. The surface measure on S_{M+N} can be decoupled as follows:

Lemma 4.4.2

Suppose g is a nonnegative function defined on S_{M+N}. Then for $\boldsymbol{\rho} = (\boldsymbol{\sigma}, \boldsymbol{\omega}) \in S_{M+N}$ we have

$$\int_{S_{M+N}} g(\boldsymbol{\rho})\, d\lambda_{M+N}(\boldsymbol{\rho}) = \int_{A_{M,N}} \int_{S_N} g(\psi(\boldsymbol{\sigma}, \boldsymbol{\omega}))\, d\lambda_N(\boldsymbol{\sigma}) d\nu_{M,N}(\boldsymbol{\omega}).$$

Proof. This result can be found in [30, Lemma 3]. It follows from Lemma 3.1.1 applied fist to $\boldsymbol{m} = e_{M+N}$, then inductively to $e_{M+N-1}, \ldots, e_{M+1}$ □

We will need a multidimensional version of this argument. To simplify notation, for n copies of S_{N+M}, we define $a_\ell^j := a_\ell(\boldsymbol{\omega}(j))$ keeping the dependence on the cavity coordinate $\boldsymbol{\omega}(j)$ implicit. Similarly, we define

$$\Psi(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\omega}}) = \Big(\psi\big(\boldsymbol{\sigma}(j), \boldsymbol{\omega}(j)\big)\Big)_{j\leq n} = \Big(a_{M+1}^j \vec{\sigma}_1(j), \ldots, a_{M+1}^j \vec{\sigma}_N(j), a_1^j \vec{\omega}_1(j), \ldots, a_M^j \vec{\omega}_M(j)\Big)_{j\leq n},$$

to represent the transformation applied coordinate-wise. The following result explains how the surface measure on S_{M+N} decouples asymptotically.

Corollary 4.4.1

Suppose g is a nonnegative function defined on S_{M+N}^n. Then for $\vec{\boldsymbol{\rho}} = (\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\omega}}) \in S_{M+N}^n$ we have

$$\int_{S_{M+N}^n} g(\vec{\boldsymbol{\rho}})\, d\lambda_{M+N}^n(\vec{\boldsymbol{\rho}}) = \int_{S_N^n} \int_{A_{M,N}^n} g\big(\Psi(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\omega}})\big)\, d\nu_{M,N}^n(\vec{\boldsymbol{\omega}}) d\lambda_N^n(\vec{\boldsymbol{\sigma}}). \tag{4.115}$$

Proof. We apply Lemma 4.4.2 to each coordinate $j \leq n$. The region of integration is a product set and we are integrating a non-negative function, so we are able to rearrange the order of integration by Fubini's Theorem. □

We apply Corollary 4.4.1 to lower bound $Z_{M+N}(\boldsymbol{Q}, \varepsilon)$ with an integral over the product set of bulk and cavity coordinates. To simplify notation, we denote the transformed coordinates with

$$\tilde{\boldsymbol{\rho}} = (\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\omega}}) := \Psi(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\omega}}) \tag{4.116}$$

where

$$\tilde{\boldsymbol{\sigma}} = \big(a^j_{M+1}\vec{\sigma}_1(j), \dots, a^j_{M+1}\vec{\sigma}_N(j)\big)_{j\le n} \text{ and } \tilde{\boldsymbol{\omega}} = \big(a^j_1\vec{\omega}_1(j), \dots, a^j_M\vec{\omega}_M(j)\big)_{j\le n} \tag{4.117}$$

are the respective transformed bulk and cavity coordinates. For an arbitrary non-negative function g on S^n_{M+N}, (4.115) implies

$$\begin{aligned}\int_{Q^\varepsilon_{M+N}} g(\vec{\boldsymbol{\rho}})\, d\lambda^n_{M+N}(\vec{\boldsymbol{\rho}}) &= \int_{S^n_{M+N}} \mathbb{1}_{Q^\varepsilon_{M+N}}(\vec{\boldsymbol{\rho}}) g(\vec{\boldsymbol{\rho}})\, d\lambda^n_{M+N}(\vec{\boldsymbol{\rho}}) \\ &= \int_{S^n_N}\int_{A^n_{M,N}} \mathbb{1}_{Q^\varepsilon_{M+N}}(\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\omega}}) g(\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\omega}})\, d\nu^n_{M,N}(\vec{\boldsymbol{\omega}}) d\lambda^n_N(\vec{\boldsymbol{\sigma}}).\end{aligned} \tag{4.118}$$

We first split the integral over Q^ε_{M+N} into a product set over $Q^\varepsilon_N \times \Omega^{\varepsilon/2,\delta}_M$ for suitably chosen δ.

Lemma 4.4.3

For any $\varepsilon > 0$ and N sufficiently large, there exists a $\delta \in (0,\varepsilon)$ such that

$$\mathbb{1}_{Q^\varepsilon_{M+N}}(\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\omega}}) \ge \mathbb{1}_{Q^\varepsilon_N}(\vec{\boldsymbol{\sigma}})\, \mathbb{1}_{\Omega^{\varepsilon/2,\delta}_M}(\vec{\boldsymbol{\omega}}). \tag{4.119}$$

Proof. We will find conditions on δ such that (4.119) holds. Let $\delta > 0$ and take $\vec{\boldsymbol{\sigma}} \in Q^\varepsilon_N$ and $\vec{\boldsymbol{\omega}} \in \Omega^{\varepsilon/2,\delta}_M$. The overlaps of the transformed coordinates $\tilde{\boldsymbol{\rho}}$ defined in (4.116) satisfy

$$\boldsymbol{R}(\tilde{\boldsymbol{\rho}},\tilde{\boldsymbol{\rho}}) = \frac{N}{M+N}\boldsymbol{R}(\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\sigma}}) + \frac{M}{M+N}\boldsymbol{R}(\tilde{\boldsymbol{\omega}},\tilde{\boldsymbol{\omega}}). \tag{4.120}$$

The set $\Omega^{\varepsilon/2,\delta}_M$ is bounded, so the corresponding transformed overlaps $\boldsymbol{R}(\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\sigma}})$ and $\boldsymbol{R}(\tilde{\boldsymbol{\omega}},\tilde{\boldsymbol{\omega}})$, can be approximated by the standard overlaps $\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}})$ and $\boldsymbol{R}(\vec{\boldsymbol{\tau}},\vec{\boldsymbol{\tau}})$ of configurations $\vec{\boldsymbol{\sigma}}$ and $\vec{\boldsymbol{\tau}}$ on the spheres S_N and S_M respectively.

Firstly, since $\|\vec{\boldsymbol{\omega}}(j)\|^2 < M(1+\delta)$ for all $\vec{\boldsymbol{\omega}} \in \Omega^{\varepsilon/2,\delta}_M$ and $j \le n$, we have the relation

$$\lim_{N\to\infty}\big(N - a^j_{M+1}a^{j'}_{M+1}N\big) = \frac{\|\boldsymbol{\omega}(j)\|^2 + \|\boldsymbol{\omega}(j')\|^2}{2} - M \le M\delta. \tag{4.121}$$

Since $R^{j,j'}(\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\sigma}}) = a^j_{M+1}a^{j'}_{M+1}R^{j,j'}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}})$, for all N sufficiently large

$$\Big\|\frac{N}{N+M}\boldsymbol{R}(\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\sigma}}) - \frac{N}{N+M}\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}})\Big\|_\infty \le \frac{LM\delta}{N+M}. \tag{4.122}$$

Secondly, $\Omega^{\varepsilon/2,\delta}_M$ is a compact set so $\lim_{N\to\infty} a_\ell(\vec{\boldsymbol{\omega}}) = 1$ uniformly for all $1 < \ell \le M$. Therefore,

$$\lim_{N\to\infty}\|\boldsymbol{R}(\tilde{\boldsymbol{\omega}},\tilde{\boldsymbol{\omega}}) - \boldsymbol{R}(\vec{\boldsymbol{\omega}},\vec{\boldsymbol{\omega}})\|_\infty = 0,$$

uniformly on $\Omega^{\varepsilon/2,\delta}_M$. Likewise, on $\Omega^{\varepsilon/2,\delta}_M$, $\|\vec{\boldsymbol{\omega}}(j) - \vec{\boldsymbol{\tau}}(j)\|^2 = \|s_j\vec{\boldsymbol{\tau}}(j) - \vec{\boldsymbol{\tau}}(j)\|^2 \le \delta M$ for all $j \le n$, so

$$\|\boldsymbol{R}(\vec{\boldsymbol{\omega}},\vec{\boldsymbol{\omega}}) - \boldsymbol{R}(\vec{\boldsymbol{\tau}},\vec{\boldsymbol{\tau}})\|_\infty \le \delta.$$

Therefore, the triangle inequality implies for all $\vec{\boldsymbol{\omega}} \in \Omega_M^{\varepsilon/2,\delta}$,

$$\Big\| \frac{M}{N+M} \boldsymbol{R}(\tilde{\boldsymbol{\omega}}, \tilde{\boldsymbol{\omega}}) - \frac{M}{N+M} \boldsymbol{R}(\vec{\boldsymbol{\tau}}, \vec{\boldsymbol{\tau}}) \Big\|_\infty \le \frac{LM\delta}{N+M}. \tag{4.123}$$

For any $\vec{\boldsymbol{\sigma}} \in Q_N^\varepsilon$ and $\vec{\boldsymbol{\omega}} \in \Omega_M^{\varepsilon/2,\delta}$, (4.122) and (4.123) imply for N sufficiently large

$$\begin{aligned}
\big\| \boldsymbol{R}(\tilde{\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}) - \boldsymbol{Q} \big\|_\infty &= \Big\| \frac{N}{M+N} \big(\boldsymbol{R}(\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\sigma}}) - \boldsymbol{Q} \big) + \frac{M}{M+N} \big(\boldsymbol{R}(\tilde{\boldsymbol{\omega}}, \tilde{\boldsymbol{\omega}}) - \boldsymbol{Q} \big) \Big\|_\infty \\
&\le \Big\| \frac{N}{M+N} \big(\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) - \boldsymbol{Q} \big) + \frac{M}{M+N} \big(\boldsymbol{R}(\vec{\boldsymbol{\tau}}, \vec{\boldsymbol{\tau}}) - \boldsymbol{Q} \big) \Big\|_\infty + \frac{LM\delta}{N+M} \\
&\le \frac{N\varepsilon}{N+M} + \frac{M\varepsilon}{2(N+M)} + \frac{LM\delta}{M+N} \\
&= \varepsilon + \frac{M}{N+M} \Big(L\delta - \frac{\varepsilon}{2} \Big).
\end{aligned}$$

Choosing $\delta \le \frac{\varepsilon}{2L}$, we have

$$\big\| \boldsymbol{R}(\tilde{\boldsymbol{\rho}}, \tilde{\boldsymbol{\rho}}) - \boldsymbol{Q} \big\|_\infty \le \varepsilon.$$

In particular, this means

$$\Big\{ (\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\omega}}) \in Q_{N+M}^\varepsilon \Big\} \supseteq \Big\{ \vec{\boldsymbol{\sigma}} \in Q_N^\varepsilon \Big\} \times \Big\{ \vec{\boldsymbol{\omega}} \in \Omega_M^{\varepsilon/2,\delta} \Big\},$$

completing the proof. □

Applying Lemma 4.4.3 to (4.118) and taking $g(\vec{\boldsymbol{\rho}}) = \exp H_{M+N}(\vec{\boldsymbol{\rho}})$, we have for N sufficiently large

$$\mathbb{E} \log Z_{M+N}(\boldsymbol{Q}, \varepsilon) \ge \mathbb{E} \log \int_{Q_N^\varepsilon} \int_{\Omega_M^{\varepsilon/2,\delta}} \exp \Big(\sum_{j \le n} H_{M+N}^j (\tilde{\boldsymbol{\sigma}}, \tilde{\boldsymbol{\omega}}) \Big) \, d\nu_{M,N}^n(\vec{\boldsymbol{\omega}}) d\lambda_N^n(\vec{\boldsymbol{\sigma}}). \tag{4.124}$$

Consequently, we are able to decouple the surface measure, which resolves the first major obstacle in the proof of the Aizenman–Sims–Starr representation.

4.4.2 Proof of the Vector Spin Aizenman–Sims–Starr Scheme

We now prove proof of Lemma 4.4.1. Using (4.124), we can derive a lower bound for the first term in (4.109).

Lemma 4.4.4

For every $\varepsilon > 0$, there exists a $\delta \in (0, \varepsilon)$ such that

$$\begin{aligned}
\liminf_{N\to\infty} \mathbb{E} \log \frac{Z_{M+N}(\boldsymbol{Q}, \varepsilon)}{Z_{M,N}(\boldsymbol{Q}, \varepsilon)} \ge \liminf_{N\to\infty} \mathbb{E} \log \Big\langle \int_{\Omega_M^{\varepsilon/2,\delta}} \exp \Big(\sum_{i \le M} \sum_{j \le n} \vec{\omega}_i(j) z_i^j(\vec{\boldsymbol{\sigma}}) \Big) \, d\nu_M^n(\vec{\boldsymbol{\omega}}) \Big\rangle_{M,N} \\
- LM\delta.
\end{aligned} \tag{4.125}$$

Proof. We start by splitting the left hand side of (4.125) in three parts

$$\mathbb{E}\log\frac{Z_{M+N}(\boldsymbol{Q},\varepsilon)}{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\,d\nu_{M,N}^n(\vec{\boldsymbol{\omega}})}+\mathbb{E}\log\frac{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\,d\nu_{M,N}^n(\vec{\boldsymbol{\omega}})}{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\,d\nu_{M}^n(\vec{\boldsymbol{\omega}})}+\mathbb{E}\log\frac{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\,d\nu_{M}^n(\vec{\boldsymbol{\omega}})}{Z_{M,N}(\boldsymbol{Q},\varepsilon)}, \tag{4.126}$$

where

$$J_{M,N}:=J_{M,N}(\vec{\boldsymbol{\omega}})=\int_{Q_N^\varepsilon}\exp\Big(\sum_{j\le n}H_{M,N}^j(\vec{\boldsymbol{\sigma}})+\sum_{j\le n}\sum_{i\le M}\vec{\omega}_i(j)z_i^j(\vec{\boldsymbol{\sigma}})\Big)\,d\lambda_N^n(\vec{\boldsymbol{\sigma}}).$$

We bound each of the terms in (4.126) separately.

Step 1: We show the first term in (4.126) satisfies

$$\liminf_{N\to\infty}\mathbb{E}\log\frac{Z_{M+N}(\boldsymbol{Q},\varepsilon)}{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\,d\nu_{M,N}^n(\vec{\boldsymbol{\omega}})}\ge -LM\delta. \tag{4.127}$$

Recall the lower bound (4.124)

$$\mathbb{E}\log Z_{M+N}(\boldsymbol{Q},\varepsilon)\ge\mathbb{E}\log\int_{Q_N^\varepsilon}\int_{\Omega_M^{\varepsilon/2,\delta}}\exp\Big(\sum_{j\le n}H_{M+N}^j\big(\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\omega}}\big)\Big)\,d\nu_{M,N}^n(\vec{\boldsymbol{\omega}})d\lambda_N^n(\vec{\boldsymbol{\sigma}}). \tag{4.128}$$

We now use Gaussian interpolation to control the term on the right hand side.

Recall the Gaussian fields in (4.99) and (4.105). We define the interpolating Hamiltonian

$$H_t(\vec{\boldsymbol{\rho}})=\sum_{j\le n}H_{1,t}^j(\vec{\boldsymbol{\rho}})+H_{2,t}^j(\vec{\boldsymbol{\rho}})+H_{3,t}^j(\vec{\boldsymbol{\rho}}). \tag{4.129}$$

For $\vec{\boldsymbol{\sigma}}\in Q_N^\varepsilon$, $\vec{\boldsymbol{\omega}}\in\Omega_M^{\varepsilon/2,\delta}$, using the corresponding transformed coordinates (4.117), the Gaussian processes in (4.129) are given by

$$H_{1,t}^j(\vec{\boldsymbol{\rho}})=\sqrt{t}H_{M,N}^j\big(\tilde{\boldsymbol{\sigma}}\big)+\sqrt{1-t}\tilde{H}_{M,N}^j(\vec{\boldsymbol{\sigma}}), \tag{4.130}$$

$$H_{2,t}^j(\vec{\boldsymbol{\rho}})=\sum_{i\le M}\vec{\omega}_i(j)\Big(\sqrt{t}a_i^jZ_i^j(\tilde{\boldsymbol{\sigma}})+\sqrt{1-t}z_i^j(\vec{\boldsymbol{\sigma}})\Big), \tag{4.131}$$

$$H_{3,t}^j(\vec{\boldsymbol{\rho}})=\sqrt{t}\,r\big(\tilde{\boldsymbol{\sigma}},\tilde{\boldsymbol{\omega}}\big). \tag{4.132}$$

The Gaussian process $\tilde{H}_{M,N}^j$ is an independent copy of $H_{M,N}^j$. Let

$$\varphi(t)=\mathbb{E}\log\int_{Q_N^\varepsilon}\int_{\Omega_M^{\varepsilon/2,\delta}}\exp\Big(H_t(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\omega}})\Big)\,d\nu_{M,N}^n(\vec{\boldsymbol{\omega}})d\lambda_N^n(\vec{\boldsymbol{\sigma}}) \tag{4.133}$$

be the corresponding interpolating Hamiltonian. By Gaussian integration by parts,

$$\varphi'(t)=\frac{1}{2}\sum_{j,j'\le n}\sum_{k\le 3}\mathbb{E}\Big\langle\mathbb{E}\frac{\partial H_{k,t}^j(\vec{\boldsymbol{\rho}}^1)}{\partial t}\cdot H_{k,t}^{j'}(\vec{\boldsymbol{\rho}}^1)-\mathbb{E}\frac{\partial H_{k,t}^j(\vec{\boldsymbol{\rho}}^1)}{\partial t}\cdot H_{k,t}^{j'}(\vec{\boldsymbol{\rho}}^2)\Big\rangle_t \tag{4.134}$$

where $\langle\cdot\rangle_t$ is the average with respect to the Gibbs measure on $Q_N^\varepsilon\times\Omega_M^{\varepsilon/2,\delta}$ proportional to $\exp(H_t)$ with respect to the reference measure $\lambda_N^n\times\nu_{M,N}^n$. We now compute the covariances of the cavity fields

in (4.129).

For all $\ell \le M+1$ and $j \le n$, $a_\ell^j(\vec{\boldsymbol{\omega}}) \to 1$ uniformly on $\Omega_M^{\varepsilon/2,\delta}$ by compactness. The leading terms of (4.102) and (4.104) do not grow in N, so by continuity the terms (4.131) and (4.132) in (4.134) vanish in the limit.

We now compute the covariances containing $H_{1,t}^j$. The covariance of (4.101) is order N, so it is not obvious that the differences of the covariances are small. We resolve this by using identity (4.121) and applying the mean value theorem. If we let $a_i^j(\boldsymbol{\omega}^1) := a_i(\boldsymbol{\omega}^1(j))$, then

$$\begin{aligned}
&\lim_{N\to\infty} \Big|\mathbb{E}\frac{\partial H_{1,t}^j(\vec{\boldsymbol{\rho}}^1)}{\partial t} \cdot H_{1,t}^{j'}(\vec{\boldsymbol{\rho}}^2)\Big| \\
&= \lim_{N\to\infty} (M+N)\Big|\xi'_{j,j'}\Big(\frac{N}{M+N} a_{M+1}^j(\boldsymbol{\omega}^1) a_{M+1}^{j'}(\boldsymbol{\omega}^2) R_{1,2}^{j,j'}\Big) - \xi'_{j,j'}\Big(\frac{N}{M+N} R_{1,2}^{j,j'}\Big)\Big| \\
&\le \lim_{N\to\infty} \|\boldsymbol{\xi}'(1)\|_1 \Big|N\Big(a_{M+1}^j(\boldsymbol{\omega}^1) a_{M+1}^{j'}(\vec{\boldsymbol{\omega}}^2) - 1\Big)\Big| \\
&\le \|\boldsymbol{\xi}'(1)\|_1 \delta M,
\end{aligned}$$

uniformly on $Q_N^\varepsilon \times \Omega_M^{\varepsilon/2,\delta}$ for all $j, j' \le n$. In particular, $\lim_{N\to\infty} \sup_{0\le t\le 1} |\varphi'(t)| \le LM\delta$. The mean value theorem implies

$$\liminf_{N\to\infty} \mathbb{E}\log \frac{Z_{M+N}(\boldsymbol{Q},\varepsilon)}{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\, d\nu_{M,N}^n(\vec{\boldsymbol{\omega}})} \ge \liminf_{N\to\infty} \big(\varphi(1) - \varphi(0)\big) \ge -\limsup_{N\to\infty} \sup_{0\le t\le 1} |\varphi'(t)| \ge -LM\delta,$$

finishing the bound of (4.127).

Step 2: We show the second term in (4.126) satisfies

$$\liminf_{N\to\infty} \mathbb{E}\log \frac{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\, d\nu_{M,N}^n(\vec{\boldsymbol{\omega}})}{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\, d\nu_M^n(\vec{\boldsymbol{\omega}})} \ge 0.$$

This proof is identical to the proof of Lemma 5 in [30]. The key observation is $f_{M,N}$, the density of $\nu_{M,N}$, converges to $f_M(\vec{\boldsymbol{\omega}})$, the density of ν_M. Since $\log(1-x) \ge -x - x^2$ for $x < 0.5$, for N sufficiently large, we have

$$\begin{aligned}
\log \frac{f_{M,N}(\boldsymbol{x})}{f_M(\boldsymbol{x})} &= \log b_{M,N} + \sum_{j=1}^M \frac{M+N-2-j}{2} \cdot \log\Big(1 - \frac{x_j^2}{M+N+1-j}\Big) - \frac{M}{2}\log\frac{1}{2\pi} + \frac{\|\boldsymbol{x}\|^2}{2} \\
&\ge \log b_{M,N} + \sum_{j=1}^M \left(\frac{M+N+1-j}{2}\right) \cdot \left(-\frac{x_j^2}{M+N+1-j} - \frac{x_j^4}{(M+N+1-j)^2}\right) \\
&\quad - \frac{M}{2}\log\frac{1}{2\pi} + \frac{\|\boldsymbol{x}\|^2}{2} \\
&\ge \log \frac{b_{M,N}}{(2\pi)^{-M/2}} - \frac{\|\boldsymbol{x}\|^4}{2N}.
\end{aligned}$$

To simplify notation, let $f_M^n(\vec{\boldsymbol{\omega}}) := \prod_{j\le n} f_M(\vec{\boldsymbol{\omega}}(j))$ and $f_{M,N}^n(\vec{\boldsymbol{\omega}}) := \prod_{j\le n} f_{M,N}(\vec{\boldsymbol{\omega}}(j))$. Jensen's

inequality implies

$$\begin{aligned}\mathbb{E}\log\frac{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\, d\nu_{M,N}^n(\vec{\boldsymbol{\omega}})}{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\, d\nu_M^n(\vec{\boldsymbol{\omega}})} &= \mathbb{E}\log\frac{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}(\vec{\boldsymbol{\omega}}) f_M^n(\vec{\boldsymbol{\omega}})\frac{f_{M,N}^n(\vec{\boldsymbol{\omega}})}{f_M^n(\vec{\boldsymbol{\omega}})} d\vec{\boldsymbol{\omega}}}{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}(\vec{\boldsymbol{\omega}}) f_M^n(\vec{\boldsymbol{\omega}}) d\vec{\boldsymbol{\omega}}}\\ &\geq \mathbb{E}\frac{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}(\vec{\boldsymbol{\omega}}) f_M^n(\vec{\boldsymbol{\omega}})\log\frac{f_{M,N}^n(\vec{\boldsymbol{\omega}})}{f_M^n(\vec{\boldsymbol{\omega}})} d\vec{\boldsymbol{\omega}}}{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}(\vec{\boldsymbol{\omega}}) f_M^n(\vec{\boldsymbol{\omega}})\, d\vec{\boldsymbol{\omega}}}\\ &\geq n\log\frac{b_{M,N}}{(2\pi)^{-M/2}} - \frac{n(1+\delta)^2M^2}{2N}.\end{aligned}$$

Since $b_{M,N}\to(2\pi)^{-M/2}$, we have

$$\liminf_{N\to\infty}\mathbb{E}\log\frac{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\, d\nu_{M,N}^n(\vec{\boldsymbol{\omega}})}{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\, d\nu_M^n(\vec{\boldsymbol{\omega}})} \geq \liminf_{N\to\infty} n\log\frac{b_{M,N}}{(2\pi)^{-M/2}} - \frac{n(1+\delta)^2M^2}{2N} = 0.$$

Step 3: By definition of $\langle\cdot\rangle_{M,N}$, the last term in (4.126) satisfies

$$\begin{aligned}&\liminf_{N\to\infty}\mathbb{E}\log\frac{\int_{\Omega_M^{\varepsilon/2,\delta}} J_{M,N}\, d\nu_M^n(\vec{\boldsymbol{\omega}})}{Z_{M,N}(\boldsymbol{Q},\varepsilon)}\\ &= \liminf_{N\to\infty}\mathbb{E}\log\Big\langle\int_{\Omega_M^{\varepsilon/2,\delta}}\exp\Big(\sum_{i\leq M}\sum_{j\leq n}\vec{\omega}_i(j)z_i^j(\vec{\boldsymbol{\sigma}})\Big)\, d\nu_M^n(\vec{\boldsymbol{\tau}})\Big\rangle_{M,N}.\end{aligned}$$

Step 4: Combining the inequalities in Step 1, Step 2, and Step 3 with the factorization (4.126) finishes the proof. □

We now derive a lower bound for the second term appearing in (4.109).

Lemma 4.4.5

We have

$$\lim_{N\to\infty} -\mathbb{E}\log\frac{Z_N(\boldsymbol{Q},\varepsilon)}{Z_{M,N}(\boldsymbol{Q},\varepsilon)} \geq \lim_{N\to\infty} -\mathbb{E}\log\Big\langle\exp\sqrt{M}y(\vec{\boldsymbol{\sigma}})\Big\rangle_{M,N}.$$

Proof. The proof by Gaussian interpolation is standard, see for example [30, Lemma 2]. Consider the interpolating Hamiltonian,

$$H_t(\vec{\boldsymbol{\sigma}}) = \sum_{j\leq n}\Big(H_{M,N}^j(\vec{\boldsymbol{\sigma}}) + \sqrt{t}\sqrt{M}Y^j(\vec{\boldsymbol{\sigma}}) + \sqrt{1-t}\sqrt{M}y^j(\vec{\boldsymbol{\sigma}})\Big),$$

Recalling (4.100), consider the corresponding interpolating function,

$$\varphi(t) = \mathbb{E}\log\int_{Q_N^\varepsilon}\exp H_t(\vec{\boldsymbol{\sigma}})d\lambda_N(\vec{\boldsymbol{\sigma}}).$$

Differentiating φ, we have

$$\varphi'(t) = \frac{1}{2}\mathbb{E}\Big\langle\mathbb{E}\frac{\partial H_t(\vec{\boldsymbol{\sigma}}^1)}{\partial t}\cdot H_t(\vec{\boldsymbol{\sigma}}^1) - \mathbb{E}\frac{\partial H_t(\vec{\boldsymbol{\rho}}^1)}{\partial t}\cdot H_t(\vec{\boldsymbol{\sigma}}^2)\Big\rangle_t$$

where $\langle\cdot\rangle_t$ is the Gibbs average on Q_N^ε with respect to the Hamiltonian $H_t(\vec{\boldsymbol{\sigma}})$. The covariances are given

by

$$\mathbb{E}\frac{\partial H_t(\vec{\boldsymbol{\sigma}}^1)}{\partial t}\cdot H_t(\vec{\boldsymbol{\sigma}}^2)=M\sum_{j,j'\le n}\Big(\mathbb{E}Y^j(\vec{\boldsymbol{\sigma}}^1)Y^{j'}(\vec{\boldsymbol{\sigma}}^2)-\mathbb{E}y^j(\vec{\boldsymbol{\sigma}}^1)y^{j'}(\vec{\boldsymbol{\sigma}}^2)\Big)=\mathscr{O}\Big(\frac{M}{N}\Big),$$

for any $\vec{\boldsymbol{\sigma}}^1,\vec{\boldsymbol{\sigma}}^2\in Q_N^{\varepsilon}$. Integrating $\varphi'(t)$, we have

$$\varphi(1)=\varphi(0)+\mathscr{O}\Big(\frac{M}{N}\Big).$$

Notice (4.100) implies $\varphi(1)=Z_N(\boldsymbol{Q},\varepsilon)$. Taking $N\to\infty$, and normalizing both sides by $Z_{M,N}(\boldsymbol{Q},\varepsilon)$ finishes the proof. □

The proof of Lemma 4.4.1 is now immediate.

Proof of Lemma 4.4.1. Applying Lemma 4.4.4 and Lemma 4.4.5 to (4.109), we have the lower limit of (4.109) is bounded below by

$$\frac{1}{M}\liminf_{N\to\infty}\Big(\mathbb{E}\log\Big\langle\int_{\Omega_M^{\varepsilon/2,\delta}}\exp\Big(\sum_{i\le M}\sum_{j\le n}\vec{\omega}_i(j)z_i^j(\vec{\boldsymbol{\sigma}})\Big)\,d\nu_M^n(\vec{\boldsymbol{\omega}})\Big\rangle_{M,N}\\ -\mathbb{E}\log\Big\langle\exp\sqrt{M}y(\vec{\boldsymbol{\sigma}})\Big\rangle_{M,N}\Big)\Big)-L\delta,$$

finishing the proof of Lemma 4.4.1. □

4.5 Perturbation and the Ghirlanda–Guerra identities

Using the Aizenman–Sims–Starr scheme, we can approximate the lower bound of the free energy with continuous functionals of the distribution of the overlap array. In particular, we have the terms

$$\mathbb{E}\log\Big\langle\int_{\Omega_M^{\varepsilon/2,\delta}}\exp\Big(\sum_{i\le M}\sum_{j\le n}\vec{\omega}_i(j)z_i^j(\vec{\boldsymbol{\sigma}})\Big)\,d\nu_M^n(\vec{\boldsymbol{\omega}})\Big\rangle_{M,N}\text{ and }\mathbb{E}\log\Big\langle\exp\sqrt{M}y(\vec{\boldsymbol{\sigma}})\Big\rangle_{M,N},$$

appearing in Lemma 4.4.1 are continuous functionals of the distributions of the overlap array $(R_{\ell,\ell'})_{\ell,\ell'\ge1}$ under the Gibbs measure $\mathbb{E}(G_{M,N})^{\otimes\infty}$ by [81, Theorem 1.3] using an extension of Theorem 2.2.1. Before computing the value of the lower bound in the limit, we must first understand the limiting distribution of this overlap. Our main tool is a perturbation of the Gibbs measure that, in the limit, will force the overlaps to satisfy the matrix version of the Ghirlanda–Guerra identities defined in [80, Theorem 3] and Section 3.2.1 that in turn imply a powerful synchronization property [80, Theorem 4] explained in Theorem 3.2.2 in addition to the main consequences of the usual identities explained [76, Section 3] and Section 2.2. These consequences will be summarized at the end of this section.

In this section, we introduce this perturbation of the Hamiltonian. We face two main obstacles. Firstly, the usual proof of the Ghirlanda–Guerra identities requires the self-overlaps $\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}})$ to be constant, which is not immediate in our setting because self-overlaps are only constrained to lie within an ε window $\boldsymbol{Q}$. Secondly, we need to find a suitable perturbation to give us the matrix version of the Ghirlanda–Guerra identities. Both of these issues are resolved in detail in Section 4 and Section 5 of [80]. They can be adapted to our setting with a few minor modifications.

4.5.1 Modified Coordinates

We begin by introducing a transformation of the coordinates that was used to control the self overlaps in the vector spin models [80, Section 3]. This transformation will fix the self overlaps allowing us to apply the usual proof of the Ghirlanda–Guerra identities.

We use essentially the same change of variables as defined in [80, Section 3] with two main differences. Firstly, since we only need to find a bound for positive definite constraints $\boldsymbol{Q}$, we do not need to truncate the constraints like in [80]. Secondly, the spins $\vec{\sigma}_i$ are bounded by a universal constant in the vector spin models, while the individual spins in the spherical models have entries bounded by N. In our setting, we will need to use a slightly different approach to obtain the relevant bounds on the distortion.

Let $\lambda_{min}(\boldsymbol{Q}) > 0$ denote the smallest eigenvalue of $\boldsymbol{Q}$. We first state this transformation as it appears in [80, Lemma 4] in Section 3 of [80].

Lemma 4.5.1

Let $\varepsilon < \lambda_{min}(\boldsymbol{Q})$. For each positive definite matrix $\boldsymbol{R}$ such that $\|\boldsymbol{R} - \boldsymbol{Q}\|_\infty \leq \varepsilon$, there exists a positive semidefinite matrix $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{R})$ such that $\boldsymbol{A}\boldsymbol{R}\boldsymbol{A}^\mathsf{T} = \boldsymbol{Q}$.

Furthermore, we have the bounds

$$\operatorname{tr}\left((\boldsymbol{A} - \boldsymbol{I})\boldsymbol{R}(\boldsymbol{A} - \boldsymbol{I})^\mathsf{T}\right) \leq L\sqrt{\varepsilon} \tag{4.135}$$

and, for any $\boldsymbol{R}_1$, $\boldsymbol{R}_2$ such that both $\|\boldsymbol{R}_1 - \boldsymbol{Q}\|_\infty \leq \varepsilon$ and $\|\boldsymbol{R}_2 - \boldsymbol{Q}\|_\infty \leq \varepsilon$,

$$\|\boldsymbol{A}(\boldsymbol{R}_1) - \boldsymbol{A}(\boldsymbol{R}_2)\|_\infty \leq \frac{L}{\varepsilon}\|\boldsymbol{R}_1 - \boldsymbol{R}_2\|_\infty. \tag{4.136}$$

In the spherical model, we will also need uniform control on $\|\boldsymbol{A}(\boldsymbol{R})\|_\infty$. Since our constant $\boldsymbol{Q}$ is positive definite, this fact follows as an immediate consequence of (4.135) and (4.136).

Corollary 4.5.1

If $\varepsilon \leq 1$, each matrix $A(\boldsymbol{R})$ constructed in Lemma 4.5.1, also satisfies the bound

$$\|A(\boldsymbol{R})\|_\infty \leq L. \tag{4.137}$$

Proof. We first find a bound on $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{Q})$. Since $\boldsymbol{Q}$ is positive definite, by the Cholesky decomposition, there exists an invertible matrix $\boldsymbol{B}$ such that $\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{B}^\mathsf{T}$. By (4.135), we have

$$\operatorname{tr}\Big((\boldsymbol{A} - \boldsymbol{I})\boldsymbol{Q}(\boldsymbol{A} - \boldsymbol{I})^\mathsf{T}\Big) = \operatorname{tr}\Big((\boldsymbol{A} - \boldsymbol{I})\boldsymbol{B}\boldsymbol{B}^\mathsf{T}(\boldsymbol{A} - \boldsymbol{I})^\mathsf{T}\Big) = \|(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{B}\|_F^2 \leq L\sqrt{\varepsilon}.$$

By norm equivalence, we see

$$\|(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{B}\|_2 \leq \|(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{B}\|_F^2 \leq L\sqrt{\varepsilon}.$$

Since $\boldsymbol{B}$ is invertible and the $\|\cdot\|_2$ norm is sub-multiplicative we have,

$$\|\boldsymbol{A} - \boldsymbol{I}\|_2 = \|(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{B}\boldsymbol{B}^{-1}\|_2 \leq \|(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{B}\|_2\|\boldsymbol{B}^{-1}\|_2.$$

Therefore, by norm equivalence, if we assume $\varepsilon \le 1$,

$$\|\boldsymbol{A}-\boldsymbol{I}\|_\infty \le \sqrt{n}\|\boldsymbol{A}-\boldsymbol{I}\|_2 \le L\sqrt{n}\sqrt{\varepsilon}\|\boldsymbol{B}^{-1}\|_2 \le L\sqrt{n}\|\boldsymbol{B}^{-1}\|_2,$$

which implies $\boldsymbol{A}(\boldsymbol{Q})$ is uniformly bounded for all $\varepsilon \le 1$.

Furthermore, by (4.136), for any $\boldsymbol{A}(\boldsymbol{R})$ such that $\|\boldsymbol{R}-\boldsymbol{Q}\|_\infty \le \varepsilon$, we have

$$\|\boldsymbol{A}(\boldsymbol{R})-\boldsymbol{A}(\boldsymbol{Q})\|_\infty \le \frac{L}{\varepsilon}\|\boldsymbol{R}-\boldsymbol{Q}\|_\infty \le L. \tag{4.138}$$

Therefore, all matrices $\boldsymbol{A}(\boldsymbol{R})$ lie within a closed ball around $\boldsymbol{A}(\boldsymbol{Q})$, which implies that $\|\boldsymbol{A}(\boldsymbol{R})\|_\infty$ is uniformly bounded for all $\varepsilon \le 1$. □

Remark 4.5.1. Corollary 4.5.1 also holds if we assume is $\boldsymbol{Q}$ is only positive semidefinite. The matrix $\boldsymbol{A}(\boldsymbol{R})$ has an explicit construction in the proof of Lemma 4 in [80], that only depended on a subset of the eigenvalues of $\boldsymbol{Q}$. Therefore, there are finitely many possible constructions of $\boldsymbol{A}(\boldsymbol{Q})$, so we can apply the bound (4.138) to each possible values $\boldsymbol{A}(\boldsymbol{Q})$ to conclude the uniform bound $\|A(\boldsymbol{R})\|_\infty \le L$.

Lemma 4.5.1 implies there exists a coordinate transform that fixes the self overlaps. For each $\vec{\boldsymbol{\sigma}} \in Q_N^\varepsilon$, suppose $\boldsymbol{A}_\sigma = \boldsymbol{A}(\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}}))$ is chosen as in Lemma 4.5.1. Denote the modified coordinates by $\hat{\boldsymbol{\sigma}} = (A_\sigma\sigma_i)_{i\le n} := \boldsymbol{A}_\sigma\vec{\boldsymbol{\sigma}}$ and observe the corresponding modified overlap satisfies

$$\boldsymbol{R}(\hat{\boldsymbol{\sigma}},\hat{\boldsymbol{\sigma}}) = \boldsymbol{R}(\boldsymbol{A}_\sigma\vec{\boldsymbol{\sigma}},\boldsymbol{A}_\sigma\vec{\boldsymbol{\sigma}}) = \frac{1}{N}\sum_{i\le N}(\boldsymbol{A}_\sigma\sigma_i)(\boldsymbol{A}_\sigma\sigma_i)^{\mathsf{T}} = \boldsymbol{A}\boldsymbol{R}(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\sigma}})\boldsymbol{A}^{\mathsf{T}} = \boldsymbol{Q}. \tag{4.139}$$

The bounds (4.135), (4.136), and (4.137) are used to show the modified overlap matrix is close to the usual overlap. Notice that,

$$\|\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^\ell,\hat{\boldsymbol{\sigma}}^{\ell'})-\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^\ell,\vec{\boldsymbol{\sigma}}^{\ell'})\|_\infty \le \|\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^\ell,\hat{\boldsymbol{\sigma}}^{\ell'})-\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^\ell,\hat{\boldsymbol{\sigma}}^{\ell'})\|_\infty + \|\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^\ell,\hat{\boldsymbol{\sigma}}^{\ell'})-\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^\ell,\vec{\boldsymbol{\sigma}}^{\ell'})\|_\infty.$$

To control the first term, by the Cauchy–Schwarz inequality we have,

$$\begin{aligned}
\|\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^\ell,\hat{\boldsymbol{\sigma}}^{\ell'})-\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^\ell,\hat{\boldsymbol{\sigma}}^{\ell'})\|_\infty &\le \sup_{j,j'\le n}\frac{1}{N}\Big|\sum_{i=1}^N \boldsymbol{A}_\sigma\vec{\sigma}_i^\ell(j)\boldsymbol{A}_\sigma\vec{\boldsymbol{\sigma}}_i^{\ell'}(j') - \sigma_i^\ell(j)\boldsymbol{A}_\sigma\vec{\boldsymbol{\sigma}}_i^{\ell'}(j')\Big| \\
&\le \frac{1}{N}\sup_{j,j'\le n}\|(\boldsymbol{A}_\sigma-\boldsymbol{I})\vec{\boldsymbol{\sigma}}^\ell(j)\|\|\boldsymbol{A}_\sigma\vec{\boldsymbol{\sigma}}^{\ell'}(j')\| \\
&\le \sup_{j\le n}\frac{\|(\boldsymbol{A}_\sigma-\boldsymbol{I})\vec{\boldsymbol{\sigma}}^\ell(j)\|}{\sqrt{N}}\|\boldsymbol{A}_\sigma\|_\infty \\
&\le \|\boldsymbol{A}_\sigma\|_\infty \operatorname{tr}(\boldsymbol{R}((\boldsymbol{A}_\sigma-\boldsymbol{I})\vec{\boldsymbol{\sigma}}^\ell,(\boldsymbol{A}_\sigma-\boldsymbol{I})\vec{\boldsymbol{\sigma}}^\ell))^{1/2}.
\end{aligned}$$

Using observation (4.139), the bounds (4.135) and (4.137) imply

$$\|\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^\ell,\hat{\boldsymbol{\sigma}}^{\ell'})-\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^\ell,\hat{\boldsymbol{\sigma}}^{\ell'})\|_\infty \le L\varepsilon^{1/4}.$$

A similar computation applied to the second term gives a similar bound,

$$\|\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^\ell,\hat{\boldsymbol{\sigma}}^{\ell'})-\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^\ell,\vec{\boldsymbol{\sigma}}^{\ell'})\|_\infty \le L\varepsilon^{1/4}.$$

Therefore, the modified overlap only differs from the overlap by a factor of $\varepsilon^{1/4}$,

$$\|\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^{\ell}, \hat{\boldsymbol{\sigma}}^{\ell'}) - \boldsymbol{R}(\vec{\boldsymbol{\sigma}}^{\ell}, \vec{\boldsymbol{\sigma}}^{\ell'})\|_{\infty} \leq L\varepsilon^{1/4}. \tag{4.140}$$

The bounds (4.140) and (4.136) will ensure this change of variables will not affect the limiting values in the perturbed Aizenman–Sims–Starr scheme that we introduce next.

4.5.2 Perturbed Hamiltonian

We now define the perturbation that will force the overlaps to satisfy the matrix version of the Ghirlanda–Guerra identities in [80] using the Gaussian process defined in Theorem 3.2.1. This perturbation is identical to the one introduced in Section 5 of [80] and Section 3.2.1. We summarize the key steps below.

We denote the family of parameters

$$\theta = (p, m, n_1, \ldots, n_m, \vec{\nu}^{\,1}, \ldots, \vec{\nu}^{\,m}). \tag{4.141}$$

For each θ, there exists Gaussian processes $h_\theta(\vec{\boldsymbol{\sigma}})$ indexed by $\vec{\boldsymbol{\sigma}} \in S_N^n$ with mean 0 and covariance

$$C^{\theta}_{\ell,\ell'} = \mathrm{Cov}\big(h_\theta(\vec{\boldsymbol{\sigma}}^{\ell}), h_\theta(\vec{\boldsymbol{\sigma}}^{\ell'})\big) = \prod_{j \leq m} \big(R^{\odot p}_{\ell,\ell'} \vec{\nu}^{\,j}, \vec{\nu}^{\,j}\big)^{n_j}. \tag{4.142}$$

Furthermore, for $\nu \in [-1,1]^n$ and $\vec{\boldsymbol{\sigma}} \in S_N^n$, the covariance is bounded by $n^{2p(n_1+\cdots+n_m)}$. We denote the countable set of parameters with

$$\Theta = \{\theta \mid p \geq 1, m \geq 1, n_1, \ldots, n_m \geq 1, \vec{\nu}^{\,1}, \ldots, \vec{\nu}^{\,m} \in ([-1,1] \cap \mathbb{Q})^n\}. \tag{4.143}$$

Let $j_0 : ([-1,1] \cap \mathbb{Q})^n \to \mathbb{N}$ be a one-to-one function. We denote an enumeration of $\theta \in \Theta$ with

$$j(\theta) = p + n_1 + \cdots + n_m + j_0(\vec{\nu}^{\,1}) + \cdots + j_0(\vec{\nu}^{\,m}) + 22m. \tag{4.144}$$

Let $(u_\theta)_{\theta \in \Theta}$ be a random sequence of i.i.d. uniform random variables in $[1,2]$. We define the interpolating Hamiltonian,

$$h_N(\vec{\boldsymbol{\sigma}}) = \sum_{\theta \in \Theta} 2^{-j(\theta)} n^{2(n_1+\cdots+n_m)} u_\theta h_\theta(\vec{\boldsymbol{\sigma}}). \tag{4.145}$$

The covariance of this process is bounded by 1, and given explicitly by

$$\mathrm{Cov}\big(h_N(\vec{\boldsymbol{\sigma}}^{\ell}), h_N(\vec{\boldsymbol{\sigma}}^{\ell'})\big) = \sum_{\theta \in \Theta} 2^{-2j(\theta)} n^{4(n_1+\cdots+n_m)} u_\theta^2 \prod_{j \leq m} \big(R^{\odot p}_{\ell,\ell'} \vec{\nu}^{\,j}, \vec{\nu}^{\,j}\big)^{n_j}. \tag{4.146}$$

For $\frac{1}{4} < \gamma < \frac{1}{2}$, we denote the sequence $s_N = N^{\gamma}$. Recall the modified coordinates defined in the previous section denoted with $\hat{\boldsymbol{\sigma}} = (\boldsymbol{A}_\sigma \vec{\sigma}_i)_{i \leq N}$. We define the perturbed Hamiltonian

$$H_N^{pert}(\vec{\boldsymbol{\sigma}}) = H_N(\vec{\boldsymbol{\sigma}}) + s_N h_N(\hat{\boldsymbol{\sigma}}), \tag{4.147}$$

and the corresponding perturbed partition function

$$Z_N^{pert}(\boldsymbol{Q},\varepsilon)=\int_{Q_N^\varepsilon}\exp\Big(H_N^{pert}(\vec{\boldsymbol{\sigma}})+\sum_{i\le N}\sum_{j\le n}\vec{h}(j)\vec{\sigma}_i(j)\Big)\,d\lambda_N^n(\vec{\boldsymbol{\sigma}}). \tag{4.148}$$

Since $\frac{s_N^2}{N}\to 0$, a straightforward Gaussian interpolation argument shows

$$\liminf_{N\to\infty}\frac{1}{N}\mathbb{E}\log Z_N(\boldsymbol{Q},\varepsilon)=\liminf_{N\to\infty}\frac{1}{N}\mathbb{E}\log Z_N^{pert}(\boldsymbol{Q},\varepsilon). \tag{4.149}$$

4.5.3 Perturbed Aizenman–Sims–Starr Scheme

The Aizenman–Sims–Starr scheme proved in Section 4.4 has to be modified slightly to account for the extra perturbation term in the Hamiltonian. Let $\langle\cdot\rangle_{pert}$ be the average on Q_N^ε with respect to the Gibbs measure

$$G_N^{pert}(\vec{\boldsymbol{\sigma}})=\frac{\exp\Big(H_{M,N}^{pert}(\vec{\boldsymbol{\sigma}})+\sum_{i\le N}\vec{h}(j)\vec{\sigma}_i(j)\Big)}{Z_{M,N}^{pert}(\boldsymbol{Q},\varepsilon)}, \tag{4.150}$$

where $H_{M,N}^{pert}(\vec{\boldsymbol{\sigma}})=H_{M,N}(\vec{\boldsymbol{\sigma}})+s_N h_N(\hat{\boldsymbol{\sigma}})$ and

$$Z_N^{pert}(\boldsymbol{Q},\varepsilon)=\int_{Q_N^\varepsilon}\exp\Big(H_{M,N}^{pert}(\vec{\boldsymbol{\sigma}})+\sum_{i\le N}\sum_{j\le n}\vec{h}(j)\vec{\sigma}_i(j)\Big)\,d\lambda_N^n(\vec{\boldsymbol{\sigma}}). \tag{4.151}$$

The following modification of Lemma 4.4.1 will be used in the proof of lower bound.

Lemma 4.5.2

For $s_N=N^\gamma$, $\vec{h}=\vec{0}$ and $\hat{\boldsymbol{\sigma}}=(\boldsymbol{A}_\sigma\vec{\sigma}_i)_{i\le N}$ we have

$$\liminf_{N\to\infty}\frac{1}{N}\mathbb{E}\log Z_N^{pert}\ge\frac{1}{M}\liminf_{N\to\infty}\Bigg(\mathbb{E}\log\Big\langle\int_{\Omega_M^{\varepsilon/2,\delta}}\exp\Big(\sum_{i\le M}\sum_{j\le n}\vec{\omega}_i(j)z_i^j(\hat{\boldsymbol{\sigma}})\Big)\,d\nu_M^n(\vec{\boldsymbol{\omega}})\Big\rangle_{pert}$$
$$-\mathbb{E}\log\Big\langle\exp\sqrt{M}y(\hat{\boldsymbol{\sigma}})\Big\rangle_{pert}\Bigg)-L\delta-L\varepsilon^{1/4}. \tag{4.152}$$

Proof. Only a small modification needs to be made to adapt the proof of Lemma 4.4.1 to this setting. We start from the bound in (4.109),

$$\liminf_{N\to\infty}\frac{1}{N}\mathbb{E}\log Z_N^{pert}(\boldsymbol{Q},\varepsilon)\ge\frac{1}{M}\liminf_{N\to\infty}\big(\mathbb{E}\log Z_{M+N}^{pert}(\boldsymbol{Q},\varepsilon)-\mathbb{E}\log Z_N^{pert}(\boldsymbol{Q},\varepsilon)\big). \tag{4.153}$$

Since the $s_N h_N(\hat{\boldsymbol{\sigma}})$ terms in the perturbed Hamiltonian are independent with all other Gaussian processes, we can leave the $s_N h_N$ terms untouched by the interpolations in the proof of Lemma 4.4.1. The exact same computations imply that Lemma 4.4.1 can be applied in this setting to conclude that (4.153)

is bounded below by

$$\frac{1}{M}\liminf_{N\to\infty}\Bigg(\mathbb{E}\log\int_{Q_N^\varepsilon}\int_{\Omega_M^{\varepsilon/2,\delta}}\exp\Big(\sum_{i\le M}\sum_{j\le n}\vec{\omega}_i(j)z_i^j(\vec{\boldsymbol{\sigma}})+H_{M,N}(\vec{\boldsymbol{\sigma}}) \\ +s_{N+M}h_{N+M}(\boldsymbol{A}_\rho\tilde{\boldsymbol{\rho}})\Big)\,d\nu_M^n(\vec{\boldsymbol{\omega}})\,d\lambda_N^n(\vec{\boldsymbol{\sigma}}) \\ -\mathbb{E}\log\int_{Q_N^\varepsilon}\exp\Big(\sqrt{M}y(\vec{\boldsymbol{\sigma}})+H_{M,N}(\vec{\boldsymbol{\sigma}})+s_Nh_N(\boldsymbol{A}_\sigma\vec{\boldsymbol{\sigma}})\Big)\,d\lambda_N^n(\vec{\boldsymbol{\sigma}})\Bigg)-L\delta, \tag{4.154}$$

after canceling the normalization terms. The perturbation term $s_{N+M}h_{N+M}(\boldsymbol{A}_\rho\tilde{\boldsymbol{\rho}})$ needs to appear as $s_Nh_N(\hat{\boldsymbol{\sigma}})$ in the normalization (4.151). We can use an interpolation to show that we can replace perturbation term at the cost of a small error term that vanishes as $N\to\infty$.

Consider the interpolating Hamiltonian

$$H_t(\vec{\boldsymbol{\rho}})=\sum_{i\le M}\sum_{j\le n}\vec{\omega}_i(j)z_i^j(\vec{\boldsymbol{\sigma}})+H_{M,N}(\vec{\boldsymbol{\sigma}})+\sqrt{t}s_{N+M}h_{N+M}\big(\boldsymbol{A}_\rho\tilde{\boldsymbol{\rho}}\big)+\sqrt{1-t}s_Nh_N(\hat{\boldsymbol{\sigma}}),$$

and the interpolating free energy

$$\varphi(t)=\mathbb{E}\log\int_{Q_N^\varepsilon}\int_{\Omega_M^{\varepsilon/2,\delta}}\exp\Big(H_t(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\omega}})\Big)\,d\nu_{M,N}^n(\vec{\boldsymbol{\omega}})d\lambda_N^n(\vec{\boldsymbol{\sigma}}). \tag{4.155}$$

Conditionally on u_θ, to show that $|\varphi'(t)|=o(1)$ after integrating by parts, we will need to control

$$\begin{aligned}\Big|\mathbb{E}\frac{dH_t(\vec{\boldsymbol{\rho}}^1)}{dt}H_t(\vec{\boldsymbol{\rho}}^2)\Big|&=\Big|s_{N+M}^2\mathbb{E}h_{N+M}\big(\boldsymbol{A}_{\rho^1}\tilde{\boldsymbol{\rho}}^1\big)h_{N+M}\big(\boldsymbol{A}_{\rho^2}\tilde{\boldsymbol{\rho}}^2\big)-s_N^2\mathbb{E}h_N(\hat{\boldsymbol{\sigma}}^1)h_N(\hat{\boldsymbol{\sigma}}^2)\Big|\\&=\Big|(N+M)^{2\gamma}g(\boldsymbol{R}(\boldsymbol{A}_{\rho^1}\tilde{\boldsymbol{\rho}}^1,\boldsymbol{A}_{\rho^2}\tilde{\boldsymbol{\rho}}^2))-N^{2\gamma}g(\boldsymbol{R}(\boldsymbol{A}_{\sigma^1}\vec{\boldsymbol{\sigma}}^1,\boldsymbol{A}_{\sigma^2}\vec{\boldsymbol{\sigma}}^2))\Big|,\end{aligned} \tag{4.156}$$

where g is the covariance function of h_N given by (4.146). The function g and its derivatives is bounded on compacts uniformly for all parameters u_θ. Using (4.120) and (4.121)

$$\begin{aligned}&\|\boldsymbol{R}(\tilde{\boldsymbol{\rho}}^1,\tilde{\boldsymbol{\rho}}^2)-\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^1,\vec{\boldsymbol{\sigma}}^2)\|_\infty\\&=\sup_{j,j'\le n}\Big|\frac{(a_{M+1}^j(\boldsymbol{\omega}^1)a_{M+1}^{j'}(\boldsymbol{\omega}^2)N-N)R^{j,j'}(\vec{\boldsymbol{\sigma}}^1,\vec{\boldsymbol{\sigma}}^2)}{M+N}-\frac{MR^{j,j'}(\vec{\boldsymbol{\sigma}}^1,\vec{\boldsymbol{\sigma}}^2)}{M+N}+\frac{MR^{j,j'}(\tilde{\boldsymbol{\omega}}^1,\tilde{\boldsymbol{\omega}}^2)}{M+N}\Big|\\&=\mathscr{O}(N^{-1}),\end{aligned}$$

and therefore, by Lemma 4.5.1,

$$\begin{aligned}\|\boldsymbol{R}(\boldsymbol{A}_{\rho^1}\tilde{\boldsymbol{\rho}}^1,\boldsymbol{A}_{\rho^2}\tilde{\boldsymbol{\rho}}^2)-\boldsymbol{R}(\boldsymbol{A}_{\sigma^1}(\vec{\boldsymbol{\sigma}}^1),\boldsymbol{A}_{\sigma^2}(\vec{\boldsymbol{\sigma}}^2))\|_\infty&=\|\boldsymbol{A}_{\rho^1}\boldsymbol{R}(\tilde{\boldsymbol{\rho}}^1,\tilde{\boldsymbol{\rho}}^2)\boldsymbol{A}_{\rho^2}^\mathsf{T}-\boldsymbol{A}_{\sigma^1}\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^1,\vec{\boldsymbol{\sigma}}^2)\boldsymbol{A}_{\sigma^2}^\mathsf{T}\|_\infty\\&\le\frac{L\|\boldsymbol{R}(\tilde{\boldsymbol{\rho}}^1,\tilde{\boldsymbol{\rho}}^2)-\boldsymbol{R}(\vec{\boldsymbol{\sigma}}^1,\vec{\boldsymbol{\sigma}}^2)\|_\infty}{\varepsilon}\\&=\mathscr{O}((N\varepsilon)^{-1}).\end{aligned}$$

Using the Taylor series of $g(\boldsymbol{R}(\boldsymbol{A}_{\rho^1}\tilde{\boldsymbol{\rho}}^1,\boldsymbol{A}_{\rho^2}\tilde{\boldsymbol{\rho}}^2))$ around $\boldsymbol{R}(\boldsymbol{A}_{\sigma^1}\vec{\boldsymbol{\sigma}}^1,\boldsymbol{A}_{\sigma^2}\vec{\boldsymbol{\sigma}}^2)$, we see

$$(N+M)^{2\gamma}g(\boldsymbol{R}(\boldsymbol{A}_{\rho^1}\tilde{\boldsymbol{\rho}}^1,\boldsymbol{A}_{\rho^2}\tilde{\boldsymbol{\rho}}^2))=(N+M)^{2\gamma}g(\boldsymbol{R}(\boldsymbol{A}_{\sigma^1}\vec{\boldsymbol{\sigma}}^1,\boldsymbol{A}_{\sigma^2}\vec{\boldsymbol{\sigma}}^2))+\mathscr{O}(N^{-1-2\gamma}/\varepsilon).$$

Since $(N+M)^{2\gamma} - N^{\gamma} = \mathscr{O}(N^{-1-2\gamma})$ we see (4.156) satifies

$$\Big|(N+M)^{2\gamma} g(\boldsymbol{R}(\boldsymbol{A}_{\rho^1}\tilde{\boldsymbol{\rho}}^1, \boldsymbol{A}_{\rho^2}\tilde{\boldsymbol{\rho}}^2)) - N^{2\gamma} g(\boldsymbol{R}(\boldsymbol{A}_{\sigma^1}\vec{\boldsymbol{\sigma}}^1, \boldsymbol{A}_{\sigma^2}\vec{\boldsymbol{\sigma}}^2))\Big| = \mathscr{O}(N^{-(1-2\gamma)}/\varepsilon).$$

The above bound holds uniformly for u_θ, so combined with the fact $\gamma < \frac{1}{2}$, means that we can replace $s_{N+M}h_{N+M}(\boldsymbol{A}_\rho\tilde{\boldsymbol{\rho}})$ with $s_N h_N(\hat{\boldsymbol{\sigma}})$ and the error introduced vanishes as $N \to \infty$. Normalizing both terms in (4.154) by $\mathbb{E}\log Z^{pert}_{M,N}(\boldsymbol{Q}, \varepsilon)$ implies

$$\begin{aligned}\liminf_{N\to\infty} \frac{1}{N}\mathbb{E}\log Z_N^{pert} \geq \frac{1}{M}\liminf_{N\to\infty} \Bigg(&\mathbb{E}\log\Big\langle \int_{\Omega_M^{\varepsilon/2,\delta}} \exp\Big(\sum_{i\leq M}\sum_{j\leq n} \vec{\omega}_i(j) z_i^j(\vec{\boldsymbol{\sigma}})\Big)\, d\nu_M^n(\vec{\boldsymbol{\omega}})\Big\rangle_{pert} \\ &- \mathbb{E}\log\Big\langle \exp\sqrt{M}y(\vec{\boldsymbol{\sigma}})\Big\rangle_{pert}\Bigg) - L\delta. \end{aligned} \tag{4.157}$$

When we characterize the limiting distribution of the overlap array, we will require the self overlaps to be constant. Replacing $\vec{\boldsymbol{\sigma}}$ with the modified coordinates $\hat{\boldsymbol{\sigma}}$ in the cavity fields achieves this. Starting from (4.157), an interpolation argument will prove that the cavity fields can be replaced with

$$\mathbb{E}\log\Big\langle \int_{\Omega_M^{\varepsilon/2,\delta}} \exp\Big(\sum_{i\leq M}\sum_{j\leq n} \vec{\omega}_i(j) z_i^j(\hat{\boldsymbol{\sigma}})\Big)\, d\nu_M^n(\vec{\boldsymbol{\omega}})\Big\rangle_{pert}, \tag{4.158}$$

and

$$\mathbb{E}\log\Big\langle \exp\sqrt{M}y(\hat{\boldsymbol{\sigma}})\Big\rangle_{pert} \tag{4.159}$$

at the cost of $L\varepsilon^{1/4}$ error. We only prove (4.158) because the proof of (4.159) is almost identical.

Consider the Hamiltonian,

$$Z_i^j(\vec{\boldsymbol{\sigma}}; t) = \sqrt{t} z_i^j(\vec{\boldsymbol{\sigma}}) + \sqrt{1-t} z_i^j(\hat{\boldsymbol{\sigma}}),$$

and the corresponding interpolating function

$$\varphi(t) = \mathbb{E}\log\Big\langle \int_{Q_M^{\varepsilon/2}} \exp\Big(\sum_{i\leq M}\sum_{j\leq n} \vec{\tau}_i(j) Z_i^j(\vec{\boldsymbol{\sigma}}; t)\Big)\, d\lambda_M^n(\vec{\boldsymbol{\tau}})\Big\rangle_{pert}.$$

Let $\hat{\boldsymbol{R}}_{\ell,\ell'} := R(\hat{\boldsymbol{\sigma}}^\ell, \hat{\boldsymbol{\sigma}}^{\ell'})$, a standard integration by parts computation will show

$$\begin{aligned}|\varphi'(t)| &\leq \Big\| \boldsymbol{R}_{1,1} \odot \Big(\boldsymbol{\xi}'(\boldsymbol{R}_{1,1}) - \boldsymbol{\xi}'(\hat{\boldsymbol{R}}_{1,1})\Big) - \boldsymbol{R}_{1,2} \odot \Big(\boldsymbol{\xi}'(\boldsymbol{R}_{1,2}) - \boldsymbol{\xi}'(\hat{\boldsymbol{R}}_{1,2})\Big)\Big\|_\infty \\ &\leq n^2\xi'(1)\Big(\|\boldsymbol{R}_{1,1} - \hat{\boldsymbol{R}}_{1,1}\|_\infty + \|\boldsymbol{R}_{1,2} - \hat{\boldsymbol{R}}_{1,2}\|_\infty\Big) \\ &\leq L\varepsilon^{1/4}\end{aligned}$$

since $\|\boldsymbol{R}_{1,1} - \hat{\boldsymbol{R}}_{1,1}\|_\infty \leq L\varepsilon^{1/4}$ by Lemma 4.5.1 and (4.140). Integrating the quantity above implies

$$\varphi(0) \geq \varphi(1) - \sup_{t\in[0,1]} |\varphi'(t)| \geq \varphi(1) - L\varepsilon^{1/4}.$$

The bound for (4.159) is similar to above, and is proved using the interpolation

$$Y(\vec{\boldsymbol{\sigma}};t) = \sqrt{t}y(\vec{\boldsymbol{\sigma}}) + \sqrt{1-t}y(\hat{\boldsymbol{\sigma}}).$$

Applying the bounds (4.158) and (4.159) to (4.157) finishes the proof. □

Remark 4.5.2. We assumed $\vec{h} = \vec{0}$ in the computations above to simplify notation. If $\vec{h}$ was non-zero, then the lower bound (4.152) in Lemma 4.5.2 is of the form

$$\frac{1}{M}\liminf_{N\to\infty}\Big(\mathbb{E}\log\Big\langle\int_{\Omega_M^{\varepsilon/2,\delta}}\exp\Big(\sum_{i\le M}\sum_{j\le n}\vec{\omega}_i(j)\big(z_i^j(\hat{\boldsymbol{\sigma}})+\vec{h}(j)\big)\Big)\,d\nu_M^n(\vec{\boldsymbol{\omega}})\Big\rangle_{pert} \tag{4.160}$$

$$-\mathbb{E}\log\Big\langle\exp\sqrt{M}y(\hat{\boldsymbol{\sigma}})\Big\rangle_{pert}\Big) - L\delta. \tag{4.161}$$

where $\langle\cdot\rangle_{pert}$ is the average on Q_N^ε with respect to the Gibbs measure with external field,

$$G_N(d\vec{\boldsymbol{\sigma}}) \propto \exp\Big(H_{M,N}^{pert}(\hat{\boldsymbol{\sigma}}) + \sum_{i\le N}\sum_{j\le n}\vec{h}(j)\vec{\sigma}_i(j)\Big)\,d\lambda_N^n(\vec{\boldsymbol{\sigma}}). \tag{4.162}$$

The bound (4.161) follows by a simple modification of the above proof. The external field can be decoupled into its cavity and non-cavity coordinates immediately,

$$\sum_{i\le M+N}\sum_{j\le n}\vec{h}(j)\vec{\rho}_i(j) = \sum_{i\le N}\sum_{j\le n}\vec{h}(j)\vec{\sigma}_i(j) + \sum_{i\le M}\sum_{j\le n}\vec{h}(j)\vec{\omega}_i(j). \tag{4.163}$$

The first summation appears in the Gibbs average (4.162) and the second summation appears in the cavity field term (4.160). However, the external field in the exponent of (4.124) will appear as

$$\sum_{i\le N}\sum_{j\le n}\vec{h}(j)a_{M+1}^j\vec{\sigma}_i(j) + \sum_{i\le M}\sum_{j\le n}\vec{h}(j)a_i^j\vec{\omega}_i(j) \tag{4.164}$$

in Step 1 of the proof of Lemma 4.4.4. To resolve this issue, notice for $\vec{\boldsymbol{\omega}} \in \Omega_M^{\varepsilon/2,\delta}$ each term $a_\ell^j(\vec{\boldsymbol{\omega}}) \to 1$ uniformly on $\Omega_M^{\varepsilon/2,\delta}$ for all $\ell \le M$. For the $M+1$ coefficient, we also have

$$\lim_{N\to\infty} N|a_{M+1}^j - 1| = \left(\frac{M}{2} - \frac{\|\vec{\boldsymbol{\omega}}(j)\|^2}{2}\right) \le \frac{M\delta}{2}.$$

Therefore, by the Cauchy–Schwarz inequality, for all $(\vec{\boldsymbol{\sigma}},\vec{\boldsymbol{\omega}}) \in Q_N^\varepsilon \times \Omega_M^{\varepsilon/2,\delta}$ and $j \le n$ we have

$$\begin{aligned}
&\Big\|\sum_{i\le N}\vec{h}(j)a_{M+1}^j\vec{\sigma}_i(j) + \sum_{i\le M}\vec{h}(j)a_i^j\vec{\omega}_i(j) - \sum_{i\le N}\vec{h}(j)\vec{\sigma}_i(j) - \sum_{i\le M}\vec{h}(j)\vec{\omega}_i(j)\Big\|_\infty \\
&\le \|\vec{h}\|_\infty N|a_{M+1}^j - 1| + M\|\vec{h}\|_\infty \sup_{i\le M}|a_i^j - 1| \\
&\le LM\delta
\end{aligned}$$

for N sufficiently large. Therefore, we can replace the external field in (4.164) with (4.163) and absorb the $LM\delta$ error into the right hand side of (4.157).

4.5.4 Consequences of the Perturbation

The lower bound (4.161) is a continuous functional of the distribution of the modified arrays $(\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^{\ell}, \hat{\boldsymbol{\sigma}}^{\ell'}))_{\ell,\ell' \geq 1}$ under the Gibbs average $\mathbb{E}(G_N^{pert})^{\otimes \infty}$ [81, Lemma 8], so it suffices to study the distribution of the modified array. To this end, we state matrix version of the Ghirlanda–Guerra identities and several of its consequences. These are identical to [80, Section 5] and can now be applied in this setting with no modification.

The entries of the overlaps are in $[-1, 1]$, so the probability distributions on finite dimensional subsets of the infinite array are tight. Therefore, by the selection theorem, there exists a subsequence such that all finite dimensional distributions of $(\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^{\ell}, \hat{\boldsymbol{\sigma}}^{\ell'}))_{\ell,\ell' \geq 1}$ converge weakly. Furthermore, there exists a non-random sequence of parameters (u_θ^N) (see [80, Lemma 5] and [76, Lemma 3.3]), possibly changing in N, such that the limiting array, denoted by $(\hat{\boldsymbol{R}}_{\ell,\ell'})_{\ell,\ell' \geq 1}$ also satisfies a matrix version of the Ghirlanda–Guerra identities.

Consider k replica of this limiting array, $\hat{R}^k = (\hat{\boldsymbol{R}}_{\ell,\ell'})_{\ell,\ell' \leq k}$, we have the following analogue of the Ghirlanda–Guerra identities [80, Theorem 3] .

Lemma 4.5.3

Given any measurable function $\varphi : \mathbb{R}^m \to \mathbb{R}$ and $f = f(R^k)$, the array satisfies the Ghirlanda–Guerra identities

$$\mathbb{E} f(\hat{R}^k)\boldsymbol{C}_{1,k+1} = \frac{1}{k}\mathbb{E} f(\hat{R}^k)\mathbb{E}\boldsymbol{C}_{1,2} + \frac{1}{k}\sum_{\ell=2}^{k} \mathbb{E} f(\hat{R}^k)\boldsymbol{C}_{1,\ell}, \tag{4.165}$$

where

$$\boldsymbol{C}_{\ell,\ell'} = \varphi\Big((\hat{\boldsymbol{R}}_{\ell,\ell'}^{\odot p}\nu^1, \nu^1), \ldots, (\hat{\boldsymbol{R}}_{\ell,\ell'}^{\odot p}\nu^m, \nu^m)\Big). \tag{4.166}$$

We have two main consequences of Lemma 4.5.3. If we take $\vec{\nu}_i = e_i$ the standard basis vectors in $\mathbb{R}^n$, (4.165) implies the traces of the overlap array, denoted by $(T_{\ell,\ell'})_{\ell,\ell' \geq 1} = (\mathrm{tr}(\hat{\boldsymbol{R}}_{\ell,\ell'}))_{\ell,\ell' \geq 1}$, satisfy the usual Ghirlanda–Guerra identities,

$$\mathbb{E} f(T^k) g(T_{1,k+1}) = \frac{1}{k}\mathbb{E} f(T^k)\mathbb{E} g(T_{1,2}) + \frac{1}{k}\sum_{\ell=2}^{k} \mathbb{E} f(T^k) g(T_{1,\ell}), \quad g : \mathbb{R} \to \mathbb{R}. \tag{4.167}$$

where $T^k = (T_{\ell,\ell'})_{\ell,\ell' \leq k}$ is a sample of k replicas from the array of traces and g is a measurable function. In particular, we are able to apply all the consequences of the standard Ghirlanda–Guerra identities to $(T_{\ell,\ell'})_{\ell,\ell' \geq 1}$.

Furthermore, (4.165) implies a synchronization property [Theorem 3.2.2] for overlap matrices [80, 81].

Lemma 4.5.4

There is a function $\boldsymbol{\Phi} : \mathbb{R}^+ \to \Gamma_n$ such that

$$\tilde{\boldsymbol{R}}_{\ell,\ell'} = \boldsymbol{\Phi}\big(\mathrm{tr}(\tilde{\boldsymbol{R}}_{\ell,\ell'})\big) \text{ a.s.} \tag{4.168}$$

Furthermore, this function is non-decreasing, $\boldsymbol{\Phi}(x_1) \leq \boldsymbol{\Phi}(x_2)$ for all $x_1 \leq x_2$, and Lipshitz continuous, $\|\boldsymbol{\Phi}(x_2) - \boldsymbol{\Phi}(x_1)\|_1 \leq L|x_2 - x_1|$.

Lemma 4.5.3 and Lemma 4.5.4 will allow us to characterize the distribution of the limiting array in

the final step of the proof of the lower bound.

4.6 Lower Bound — Cavity Computations

We now have the tools to prove the lower limit of the free energy. The remainder of the proof is standard and almost identical to other spin glass models (see Chapter 3 of [76] or the proof of the lower bound in [81] and [80] or the results in Chapter 2 and Chapter 3). We will summarize the steps and reiterate the importance of the synchronization mechanism.

Let $\boldsymbol{Q}$ be a positive definite constraint. Starting from the Aizenman–Sims–Starr scheme (4.161), we have $\liminf_{N\to\infty} F_N^{\varepsilon}(\vec{\boldsymbol{\beta}}, \boldsymbol{Q})$ is bounded below by

$$\frac{1}{M}\liminf_{N\to\infty}\Bigg(\mathbb{E}\log\Big\langle\int_{\Omega_M^{\varepsilon/2,\delta}}\exp\Big(\sum_{i\le M}\sum_{j\le n}\vec{\omega}_i(j)\big(z_i^j(\hat{\boldsymbol{\sigma}})+\vec{h}(j)\big)\Big)\,d\nu_M^n(\vec{\boldsymbol{\omega}})\Big\rangle_{pert} \tag{4.169}$$
$$-\mathbb{E}\log\Big\langle\exp\sqrt{M}y(\hat{\boldsymbol{\sigma}})\Big\rangle_{pert}\Bigg)-L\delta-L\varepsilon^{1/4}.$$

From an extension of Theorem 2.2.1 defined in [81, Lemma 8], the averages on (4.169) are continuous functionals of the distribution of the modified infinite array $(\hat{\boldsymbol{R}}^{N,M}_{\ell,\ell'})_{\ell,\ell'\ge 1} := (\boldsymbol{R}(\hat{\boldsymbol{\sigma}}^{\ell},\hat{\boldsymbol{\sigma}}^{\ell'}))_{\ell,\ell'\ge 1}$. To compute this lower bound explicitly, it suffices to understand the limiting distribution of the array $(\hat{\boldsymbol{R}}^{N,M}_{\ell,\ell'})_{\ell,\ell'\ge 1}$ under the perturbed Gibbs measure $\mathbb{E}(G_N^{pert})^{\otimes\infty}$ defined in (4.150) for a deterministic choice of parameters (u_θ^N) such that Lemma 4.5.3 holds. By the selection theorem, there exists a subsequence such that

$$\big(\tilde{\boldsymbol{R}}^{N,M}_{\ell,\ell'}\big)_{\ell,\ell'\ge 1}\xrightarrow{d}(\tilde{\boldsymbol{R}}^{M}_{\ell,\ell'})_{\ell,\ell'\ge 1}.$$

The diagonal elements of this array are constant, so by Lemma 4.5.3, the limiting array $(\tilde{\boldsymbol{R}}^{M}_{\ell,\ell'})_{\ell,\ell'}$ satisfies the generalized Ghirlanda–Guerra identities (4.165) and the synchronization property (4.168). In particular, there exists a function $\boldsymbol{\Phi}:[0,1]\to\Gamma_n$ such that

$$\tilde{\boldsymbol{R}}^{M}_{\ell,\ell'}=\boldsymbol{\Phi}(\mathrm{tr}(\tilde{\boldsymbol{R}}^{M}_{\ell,\ell'}))$$

almost surely. Recall that $\boldsymbol{\Phi}$ is non-decreasing and Lipschitz. This allows us to approximate its distribution with a random measure generated by the Ruelle probability cascades.

We begin by characterizing the array $(\mathrm{tr}(\tilde{\boldsymbol{R}}^{M}_{\ell,\ell'}))_{\ell,\ell'\ge 1}$ consisting of the traces of the limiting array. As a consequence of the generalized Ghirlanda–Guerra identities (4.167), the array of traces also satisfies the usual Ghirlanda–Guerra identities. We denote the distribution of $\mathrm{tr}(\tilde{\boldsymbol{R}}^{M}_{1,2})$ with

$$\mu(q)=\mathbb{P}\big(\mathrm{tr}(\tilde{\boldsymbol{R}}^{M}_{1,2})\le q\big). \tag{4.170}$$

Following the usual proof of the lower bound (see Chapter 3 of [76]) there exists a sequence of cumulative distribution functions $(\mu_k)_{k\ge 1}$ such that $\mu_k\to\mu$ in L^1,

$$\lim_{k\to\infty}\int_0^n|\mu_k(q)-\mu(q)|\,dq=0.$$

For each k, we can encode the discrete probability measures with a sequences of parameters

$$\begin{aligned} x_{-1} = 0 < x_0 < x_1 < \ldots < x_r = 1 \\ 0 = q_0 < q_1 < \ldots < q_r = n = \mathrm{tr}(\boldsymbol{Q}) \end{aligned} \tag{4.171}$$

such that

$$\mu_k(q) = x_p \text{ for } q_p \leq q < q_{p+1}. \tag{4.172}$$

Let $(v_\alpha)_{\alpha \in \mathbb{N}^r}$ be the Ruelle probability cascades corresponding to (4.171). Let $(\alpha^\ell)_{\ell \geq 1}$ be an i.i.d. sample from $\mathbb{N}^r$ according to the weights $(v_\alpha)_{\alpha \in \mathbb{N}^r}$, it follows that the array

$$(T^k_{\ell,\ell'})_{\ell,\ell' \geq 1} = (q_{\alpha^\ell \wedge \alpha^{\ell'}})_{\ell,\ell' \geq 1}$$

also converges to $(\mathrm{tr}(\tilde{\boldsymbol{R}}^M_{\ell,\ell'}))_{\ell,\ell' \geq 1}$ by Theorem 2.2.12 and Theorem 2.2.13.

From here, we use the synchronization mechanism to recover a sequence of monotone paths in Π that describes the distribution of the limiting overlap matrix array $(\tilde{\boldsymbol{R}}^{N,M}_{\ell,\ell'})_{\ell,\ell' \geq 1}$. We define

$$\boldsymbol{Q}^k_{\ell,\ell'} = \boldsymbol{\Phi}(T^k_{\ell,\ell'}),$$

and observe $(\boldsymbol{Q}^k_{\ell,\ell'})_{\ell,\ell' \geq 1}$ converges to the distribution of $(\tilde{\boldsymbol{R}}^M_{\ell,\ell'})_{\ell,\ell' \geq 1}$ because $\boldsymbol{\Phi}$ is Lipschitz. It also follows that the discrete path

$$\pi_k(x) = \boldsymbol{Q}_k \text{ for } x_{k-1} < x \leq x_k \text{ for } 0 \leq k \leq r, \quad \pi(0) = \boldsymbol{0}, \quad \pi(1) = \boldsymbol{Q}. \tag{4.173}$$

induced by

$$\begin{aligned} x_{-1} = 0 < \ x_0 < \ x_1 < \ldots < \ x_r = \ 1 \\ 0 = \boldsymbol{Q}_0 \leq \boldsymbol{Q}_1 \leq \ldots \leq \boldsymbol{Q}_r = \boldsymbol{Q}. \end{aligned} \tag{4.174}$$

where $\boldsymbol{Q}_\ell = \boldsymbol{\Phi}(q_\ell)$ for $0 \leq \ell \leq r$ is a discretization of the path associated with the limiting array. To see this, recall (4.170) and define

$$\pi(x) := \boldsymbol{\Phi}(\mu^{-1}(x)) \in \Pi,$$

where $\mu^{-1} : [0,1] \to \mathbb{R}^+$ is the quantile distribution of μ. Similarly, for discrete μ_k given by (4.172), the paths

$$\pi_k(x) := \boldsymbol{\Phi}(\mu_k^{-1}(x)) \in \Pi,$$

are a discrete approximation of π [81, Equation (71)],

$$\mathrm{d}(\pi, \pi_k) = \int_0^1 \|\pi(x) - \pi_k(x)\|_1 \, dx \leq n \int_0^1 |\,\mathrm{tr}(\pi(x)) - \mathrm{tr}(\pi_k(x))| \, dx = n \int_0^1 |\mu(x) - \mu_k(x)| \, dx.$$

In particular, we have $\mathrm{d}(\pi, \pi_k) \to 0$ as $\mu_k \to \mu$ in L^1.

Recall the Gaussian processes $Z^j_i(\alpha)$ and $Y(\alpha)$ defined in the (4.45) and (4.46) and consider the

following functionals of the discrete paths associated with the approximating arrays $(\boldsymbol{Q}_{\alpha^\ell \wedge \alpha^{\ell'}})_{\ell,\ell' \geq 1}$:

$$f_M^1(\pi) = \frac{1}{M}\mathbb{E}\log \sum_{\alpha \in \mathbb{N}^r} v_\alpha \int_{\Omega_M^{\varepsilon/2,\delta}} \exp\Big(\sum_{i \leq M} \sum_{j \leq n} \vec{\omega}_i(j)\big(Z_i^j(\alpha) + \vec{h}(j)\big)\Big)\, d\nu_M^n(\vec{\boldsymbol{\omega}}), \tag{4.175}$$

$$f_M^2(\pi) = \frac{1}{M}\mathbb{E}\log \sum_{\alpha \in \mathbb{N}^r} v_\alpha \exp \sqrt{M} Y(\alpha). \tag{4.176}$$

The covariances of $Z_i^j(\alpha)$ and $z_i^j(\hat{\boldsymbol{\sigma}})$, and $Y(\alpha)$ and $y(\hat{\boldsymbol{\sigma}})$ are given by the same functions of arrays so the difference of the functionals (4.175), (4.176) and the functional appearing in (4.169) can be approximated by the same continuous bounded function of the array [81, Lemma 8]. In summary, by choosing a discretization μ_k close enough to μ in L^1, we can find a corresponding discrete path $\hat{\pi}_M := \pi_k$ encoded by the sequences (4.174) such that

$$\liminf_{N \to \infty} F_N^\varepsilon(\vec{\boldsymbol{\beta}}, \boldsymbol{Q}) \geq f_M^1(\hat{\pi}_M) - f_M^2(\hat{\pi}_M) - L\delta - L\varepsilon^{1/4}. \tag{4.177}$$

The lower bound holds for all M, so we can take a sub-sequential limit as $M \to \infty$. However, we cannot apply Lemma 4.3.1 to compute the lower bound, because the paths $\hat{\pi}_M$ may change in M. To resolve this, notice that by monotonicity of the paths, $\hat{\pi}_M \to \hat{\pi}$ along some subsequence [81, Section 7]. Furthermore, there exists a discretization $\hat{\pi}^\varepsilon$ of $\hat{\pi}$ such that $\mathrm{d}(\hat{\pi}, \hat{\pi}^\varepsilon) \leq \varepsilon^{1/4}$. This approximation will introduce at most $L\varepsilon^{1/4}$ error by the Lipschitz continuity of $f_M^1(\pi)$ and $f_M^2(\pi)$, so

$$\liminf_{N \to \infty} F_N^\varepsilon(\vec{\boldsymbol{\beta}}, \boldsymbol{Q}) \geq \liminf_{M \to \infty} \Big(f_M^1(\hat{\pi}^\varepsilon) - f_M^2(\hat{\pi}^\varepsilon)\Big) - L\delta - L\varepsilon^{1/4}.$$

These paths are now fixed, so we can now compute its limit as $M \to \infty$. Applying Lemma 4.3.1 to decouple the constraint on $\boldsymbol{Q}$ asymptotically shows

$$\liminf_{M \to \infty} f_M^1(\hat{\pi}^\varepsilon) \geq \inf_{\boldsymbol{\Lambda}} \frac{1}{2}\Bigg(\mathrm{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) - n - \log|\boldsymbol{\Lambda}| + (\boldsymbol{\Lambda}_0^{-1}\vec{h}, \vec{h}) + \sum_{0 \leq k \leq r-1} \frac{1}{x_k} \log \frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|}\Bigg),$$

where $(\boldsymbol{\Lambda}_k)_{0 \leq k \leq r}$ are defined with respect to the sequences $(x_k)_{-1 \leq k \leq r}$ and $(\boldsymbol{Q}_k)_{0 \leq k \leq r}$ encoded by $\hat{\pi}^\varepsilon$. By the recursive computations (4.55),

$$\lim_{M \to \infty} f_M^2(\hat{\pi}^\varepsilon) = \sum_{0 \leq k \leq r-1} x_k \cdot \mathrm{Sum}\big(\boldsymbol{\theta}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\theta}(\boldsymbol{Q}_k)\big).$$

Taking $\varepsilon \to 0$ and consequently $\delta \to 0$ removes all the error terms, so we conclude

$$\lim_{\varepsilon \to 0} \liminf_{N \to \infty} F_N^\varepsilon(\vec{\boldsymbol{\beta}}, \boldsymbol{Q}) \geq \inf_{\boldsymbol{\Lambda}, \pi} \mathscr{P}_{\vec{\boldsymbol{\beta}}, \boldsymbol{Q}}(\pi, \boldsymbol{\Lambda}). \tag{4.178}$$

Chapter 5

The Crisanti–Sommers Formula for Spherical Vector Spin Glasses

We now move onto the second main contribution of this thesis. We will partially solve the optimization of the vector form of the Parisi functional from Theorem 4.1.1 proved in Chapter 4. Our goal for this chapter is to prove the analogue of the Crisanti–Sommers functional from Theorem 1.33 for the vector version of the Parisi functional. This formula is derived from the discrete Parisi variational formula for the limit of the free energy of constrained copies of spherical spin glasses. In vector spin models, the variations of the functional order parameters must preserve the monotonicity of matrix paths which introduces a new challenge in contrast to the derivation of the classical Crisanti–Sommers formula [98, Section 4].

Following the methodology of the proof in the one dimensional case, we will prove that both the discrete Crisanti–Sommers and Parisi functionals take the same values at the critical points. The minimizers will satisfy the same critical point equations that will allow us to reduce one formula to the other and vice versa.

The main difficulty is the minimizer of the Parisi functional in vector spin models may not be an interior point of the domain. In the one dimensional case, we could assume that our the discretization of the paths are strictly monotone, i.e. $q_1 < q_2 < \cdots < q_r$. This allowed us to differentiate with respect to q_k to recover the critical point conditions immediately. In the vector spin case, the increments $\boldsymbol{Q}_p - \boldsymbol{Q}_{p-1}$ may occur on the boundary of the positive definite cone so the directional derivatives are not necessarily equal to 0 at the critical points. This is our main obstacle, because the system of equations in the critical point conditions becomes a system of inequalities unless we can show that the increments $\boldsymbol{Q}_p - \boldsymbol{Q}_{p-1}$ are positive definite.

To fix this, we will introduce a positive definite barrier to the discrete functionals that impose a large penalty if the increments of the matrix path $\underline{\boldsymbol{Q}}$ are degenerate. This will force the functionals to take a minimum at an interior point, allowing us to use variational calculus to find approximate critical point conditions. This approach will allow us to reduce the Parisi functional into an approximate Crisanti–Sommers form and vice versa. We will use convexity to show that the approximations become exact as the size of the positive definite barrier tends to 0.

The results of this chapter are adapted from [63] written by the author of this thesis.

Outline of the Chapter

In Section 5.1 We start by reviewing the Parisi functional $\mathscr{P}_r$ that was introduced in Chapter 4. We will introduce the discrete Crisanti–Sommers analogue of this formula $\mathscr{C}_r$ and its integral form.

The proof of the main result of this chapter begins in Section 5.2. We first show that

$$\inf_{r,\Lambda,x,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) \geq \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}),$$

by using a barrier function to derive approximate critical point conditions for the Parisi functional. We will use the critical point conditions to reduce the Parisi functional into its approximate Crisanti–Sommers form, and we will use convexity to show that the approximate discrete Crisanti–Sommers functional is lower bounded by the usual Crisanti–Sommers functional evaluated at a different point.

We can use a similar argument with a barrier term to prove the upper bound in Section 5.3,

$$\inf_{r,\Lambda,x,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, Q) \leq \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}).$$

This direction of the argument uses the critical point conditions for the Crisanti–Sommers functional. Miraculously, the minimizers of $\mathscr{C}_r$ satisfies almost exactly the same critical point conditions as the minimizers of $\mathscr{P}_r$ which allows us to reduce $\mathscr{C}_r$ to $\mathscr{P}_r$ in the opposite direction.

In Section 5.4 we will prove that (5.19) is locally Lipschitz to conclude that it is the correct extension of the discrete Crisanti–Sommers formula.

5.1 Introduction

We define the spherical vector model and remind the reader of the notation introduced in the earlier chapters. The main result of this chapter is the analogue of the Crisanti–Sommers formula for this high dimensional spin glass model.

5.1.1 The Limit of the Free Energy and the Parisi Formula

We start by describing the spherical spin glass model with vector spins and the Parisi formula for the limit of its free energy. Fix integer $n \geq 1$. Let S_N be the sphere in $\mathbb{R}^N$ of radius $\sqrt{N}$. A configuration of n copies of spherical spin glasses can be viewed as vector spins with coordinates restricted to lie on S_N,

$$\vec{\boldsymbol{\sigma}} = (\vec{\sigma}_1, \dots, \vec{\sigma}_N) \in S_N^n \quad \text{where} \quad S_N^n = \big\{\vec{\boldsymbol{\sigma}} \in (\mathbb{R}^N)^n \mid \boldsymbol{\sigma}(j) \in S_N \text{ for all } j \leq n\big\}. \tag{5.1}$$

The jth coordinate of $\vec{\boldsymbol{\sigma}}$ is denoted by $\boldsymbol{\sigma}(j)$ and the vector entries of $\vec{\boldsymbol{\sigma}}$ are denoted by

$$\vec{\sigma}_i = \big(\vec{\sigma}_i(1), \dots, \vec{\sigma}_i(n)\big) \in \mathbb{R}^n. \tag{5.2}$$

For $p \geq 2$, the p-spin Hamiltonian is denoted by

$$H_{N,p}(\boldsymbol{\sigma}(j)) = \frac{1}{N^{(p-1)/2}} \sum_{1 \leq i_1, \dots, i_p \leq N} g_{i_1,\dots,i_p} \vec{\sigma}_{i_1}(j) \cdots \vec{\sigma}_{i_p}(j), \tag{5.3}$$

where $g_{i_1,\ldots,i_p}$ are i.i.d. standard Gaussians for all $p \geq 2$ and indices $(i_1, \ldots, i_p)$. The corresponding mixed p-spin Hamiltonian for the jth copy at inverse temperatures $(\vec{\beta}_p)_{p\geq 2}$ is denoted by

$$H_N^j(\vec{\boldsymbol{\sigma}}) = \sum_{p\geq 2} \vec{\beta}_p(j) H_{N,p}(\boldsymbol{\sigma}(j)). \tag{5.4}$$

We assume that the inverse temperatures satisfy $\sum_{p\geq 2} 2^p \vec{\beta}_p^{\,2}(j) < \infty$ for all $j \leq n$, so that (5.4) is well-defined, and that $\vec{\beta}_p = \vec{0}$ for odd p. The Hamiltonian of n copies of these even mixed p-spin models of spherical spin glasses is denoted by

$$H_N(\vec{\boldsymbol{\sigma}}) = \sum_{j\leq n} H_N^j(\vec{\boldsymbol{\sigma}}). \tag{5.5}$$

The overlaps between the vector configurations $\vec{\boldsymbol{\sigma}}^\ell$ and $\vec{\boldsymbol{\sigma}}^{\ell'}$ are given by the overlap matrices

$$\boldsymbol{R}_{\ell,\ell'} = \boldsymbol{R}(\vec{\boldsymbol{\sigma}}^\ell, \vec{\boldsymbol{\sigma}}^{\ell'}) = \frac{1}{N} \sum_{i\leq N} \vec{\sigma}_i^\ell \otimes \vec{\sigma}_i^{\ell'} \tag{5.6}$$

where $\otimes$ is the outer product on vectors in $\mathbb{R}^n$.

The constraint $\boldsymbol{Q}$ is a $n \times n$ symmetric positive definite matrix with off-diagonals $Q^{j,j'} \in [-1, 1]$ and diagonals $Q^{j,j} = 1$. Given $\varepsilon > 0$, we denote the set of spins with constrained self overlaps by

$$Q_N^\varepsilon = \big\{\vec{\boldsymbol{\sigma}} \in S_N^n \mid \|\boldsymbol{R}(\vec{\boldsymbol{\sigma}}, \vec{\boldsymbol{\sigma}}) - \boldsymbol{Q}\|_\infty \leq \varepsilon\big\}, \tag{5.7}$$

where $\|\cdot\|_\infty$ is the infinity norm on $n \times n$ matrices. For an external field $\vec{h} \in \mathbb{R}^n$, we define the free energy as

$$F_N^\varepsilon(\boldsymbol{Q}) = \frac{1}{N}\mathbb{E}\log \int_{Q_N^\varepsilon} \exp\Big(H_N(\vec{\boldsymbol{\sigma}}) + \sum_{j\leq n} \vec{h}(j) \sum_{i\leq N} \vec{\sigma}_i(j)\Big)\, d\lambda_N^n(\vec{\boldsymbol{\sigma}}), \tag{5.8}$$

where the reference measure $\lambda_N^n = \lambda_N^{\otimes n}$ is the product of normalized uniform measures λ_N on S_N.

The limit of (5.8) can be expressed as a Parisi type functional. The Parisi functional is a Lipschitz function of discrete monotone matrix paths encoded by an increasing sequence of real numbers and monotone sequence of $n \times n$ symmetric positive semidefinite matrices,

$$\begin{aligned} 0 = x_0 \leq x_1 \leq \ldots \leq x_{r-2} \leq x_{r-1} \leq 1 \\ \boldsymbol{0} = \boldsymbol{Q}_0 \leq \boldsymbol{Q}_1 \leq \ldots \leq \boldsymbol{Q}_{r-2} \leq \boldsymbol{Q}_{r-1} \leq \boldsymbol{Q}_r = \boldsymbol{Q} \end{aligned}. \tag{5.9}$$

To lighten notation, we will denote these sequences with $\underline{x} = (x_k)_{k=0}^{r-1}$ and $\underline{\boldsymbol{Q}} = (\boldsymbol{Q}_k)_{k=1}^r$.

For $\boldsymbol{A} \in \mathbb{S}_+^n$, where $\mathbb{S}_+^n$ is the space of $n \times n$ positive semidefinite matrices, we define the functions

$$\boldsymbol{\xi}(\boldsymbol{A}) = \sum_{p\geq 2} (\vec{\beta}_p \otimes \vec{\beta}_p) \odot \boldsymbol{A}^{\odot p}, \tag{5.10}$$

and

$$\boldsymbol{\xi}'(\boldsymbol{A}) = \sum_{p\geq 2} p(\vec{\beta}_p \otimes \vec{\beta}_p) \odot \boldsymbol{A}^{\odot(p-1)} \quad \text{and} \quad \boldsymbol{\theta}(\boldsymbol{A}) = \sum_{p\geq 2} (p-1)(\vec{\beta}_p \otimes \vec{\beta}_p) \odot \boldsymbol{A}^{\odot p}, \tag{5.11}$$

where $\odot$ is the Hadamard product on $n \times n$ matrices and $\boldsymbol{A}^{\odot p}$ is the pth Hadamard power of $\boldsymbol{A}$. Since $\vec{\beta}_p = \vec{0}$ for odd p, $\boldsymbol{\xi}(\cdot)$ is an even convex function in each of its coordinates. The r step discretization of

the Parisi functional is defined by

$$\mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) = \frac{1}{2}\Big[\langle \vec{h}\vec{h}^{\intercal}, \boldsymbol{\Lambda}_1^{-1}\rangle + \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - n - \log|\boldsymbol{\Lambda}| + \sum_{1\le k\le r-1} \frac{1}{x_k}\log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|} + \langle \boldsymbol{\xi}'(\boldsymbol{Q}_1), \boldsymbol{\Lambda}_1^{-1}\rangle$$
$$- \sum_{1\le k\le r-1} x_k \cdot \mathrm{Sum}\big(\boldsymbol{\theta}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\theta}(\boldsymbol{Q}_k)\big)\Big] \tag{5.12}$$

where $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \mathrm{tr}(\boldsymbol{A}\boldsymbol{B})$ is the Frobenius inner product on symmetric matrices, $|\cdot|$ is the determinant and

$$\boldsymbol{\Lambda}_r = \boldsymbol{\Lambda}, \quad \boldsymbol{\Lambda}_p = \boldsymbol{\Lambda} - \sum_{p\le k\le r-1} x_k\big(\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_k)\big) \text{ for } 1 \le p \le r-1. \tag{5.13}$$

The domain of the Parisi functional are all sequences (5.9) and Lagrange multipliers $\boldsymbol{\Lambda}$ such that $|\boldsymbol{\Lambda}_1| > 0$. This condition also implies that $|\boldsymbol{\Lambda}_p| > 0$ for all $1 \le p \le r-1$ so (5.12) is well defined. It was proven in Theorem 4.1.1 that the limit of the free energy (5.8) is given by minimizing (5.12).

Theorem 5.1.1

The limit of the free energy with self overlaps constrained to $\boldsymbol{Q}$ equals

$$\lim_{\varepsilon\to 0}\lim_{N\to\infty} F_N^{\varepsilon}(\boldsymbol{Q}) = \inf_{r,\boldsymbol{\Lambda},x,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}). \tag{5.14}$$

The infimum is over sequences of the form (5.9), $\boldsymbol{\Lambda}$ such that $|\boldsymbol{\Lambda}_1| > 0$, and all $r \ge 1$.

Remark 5.1.1. In this chapter, we assume $x_0 = 0$ and $\boldsymbol{Q}_0 = 0$ to include the boundary cases in the computations. This implies the form of the Parisi functional (5.12) differs slightly from the one proved in Chapter 4. See Remark 4.1.2 for the equivalence of (5.12) and (4.37).

5.1.2 Discrete Form of the Crisanti–Sommers Formula

We will show that discrete the Parisi formula has a discrete Crisanti–Sommers representation. The discrete form of the Crisanti–Sommers functional is derived by examining the critical points of the discrete Parisi functional (5.12). For $r \ge 1$ and the sequence of parameters $\underline{x}$ and $\underline{\boldsymbol{Q}}$ defined in (5.9), the discrete Crisanti–Sommers representation is given by

$$\mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}) = \frac{1}{2}\Big[\langle \vec{h}\vec{h}^{\intercal}, \boldsymbol{D}_1\rangle + \frac{1}{x_{r-1}}\log|\boldsymbol{Q} - \boldsymbol{Q}_{r-1}| - \sum_{1\le k\le r-2}\frac{1}{x_k}\log\frac{|\boldsymbol{D}_{k+1}|}{|\boldsymbol{D}_k|} + \langle \boldsymbol{Q}_1, \boldsymbol{D}_1^{-1}\rangle$$
$$+ \sum_{1\le k\le r-1} x_k \cdot \mathrm{Sum}\big(\boldsymbol{\xi}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}(\boldsymbol{Q}_k)\big)\Big], \tag{5.15}$$

where,

$$\boldsymbol{D}_p = \sum_{p\le k\le r-1} x_k\big(\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k\big) \text{ for } 1 \le p \le r-1. \tag{5.16}$$

We require the additional constraint that $|\boldsymbol{Q} - \boldsymbol{Q}_{r-1}| > 0$ otherwise $\mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}})$ will be positive infinity. This condition implies that $|\boldsymbol{D}_p| > 0$ for all $1 \le p \le r-1$ so $\mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}})$ is well defined.

We will prove that the representations (5.12) and (5.15) are equal at its minimizers.

Theorem 5.1.2

For all positive definite constraints $\boldsymbol{Q}$, we have

$$\inf_{r,\Lambda,x,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) = \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}),$$

where the first infimum is over sequences (5.9) and $\boldsymbol{\Lambda} \in \mathbb{S}_+^n$ such such that $|\boldsymbol{\Lambda}_1| > 0$ and the second infimum is over sequences (5.9) such that $|\boldsymbol{D}_{r-1}| > 0$.

Remark 5.1.2. Reassuringly, in the one dimensional case, these formulas agree with the usual discretizations of the Parisi functional and Crisanti–Sommers functional (See [98, Section 4]).

5.1.3 The Integral Form of the Crisanti–Sommers Representation

The main goal of this thesis is to prove that the free energy can be obtained by minimizing a functional closely resembling the Cristanti–Sommers functional for one dimensional spherical spin glasses. The parameters of the functional is the c.d.f. of the trace of the overlap matrix, and the synchronized matrix path identifying the overlap matrix with its trace [80, Theorem 4]. Let

$$x(t) : [0, n] \to [0, 1] \quad \text{such that} \quad x(0) = 0 \quad \text{and} \quad x(n) = 1 \tag{5.17}$$

denote a right continuous non-decreasing function and

$$\boldsymbol{\Phi}(t) : [0, n] \to \mathbb{S}_+^n \quad \text{such that} \quad \operatorname{tr}(\boldsymbol{\Phi}(t)) = t \quad \text{and} \quad \boldsymbol{\Phi}(0) = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{\Phi}(n) = \boldsymbol{Q} \tag{5.18}$$

denote a 1-Lipschitz monotone matrix path in the space of $n \times n$ positive semidefinite matrices parametrized by its trace. A monotone matrix path is one with positive semidefinite increments, $\boldsymbol{\Phi}(t_2) - \boldsymbol{\Phi}(t_1) \in \mathbb{S}_+^n$ for $t_2 \geq t_1$. Since $\boldsymbol{\Phi}$ is 1-Lipschitz in each of its coordinates its coordinate wise derivative $\boldsymbol{\Phi}'$ exists almost everywhere and is bounded by 1 almost everywhere.

The largest point in the support of the measure associated with the c.d.f. $x(t)$ is denoted by

$$t_x := x^{-1}(1) = \inf\{t \in [0, n] \mid 1 \leq x(t)\}.$$

Assuming that $|\boldsymbol{Q} - \boldsymbol{\Phi}(t_x)| > 0$, we define the quantity

$$\mathscr{C}(x, \boldsymbol{\Phi}) = \frac{1}{2}\left(\int_0^n x(t) \langle \boldsymbol{\xi}'(\boldsymbol{\Phi}(t)) + \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{\Phi}'(t) \rangle \, dt + \log |\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(t_x)| + \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t) \rangle \, dt \right), \tag{5.19}$$

where $\hat{\boldsymbol{\Phi}}(t) : [0, n] \to \mathbb{R}^{n \times n}$ is a decreasing matrix path given by

$$\hat{\boldsymbol{\Phi}}(t) = \int_t^n x(s) \boldsymbol{\Phi}'(s) \, ds. \tag{5.20}$$

Because $\hat{\boldsymbol{\Phi}}(t) = \boldsymbol{Q} - \boldsymbol{\Phi}(t)$ for $t \geq t_x$, the functional does not depend on t_x. More precisely, if $\hat{t} \geq t_x$ and

$|\boldsymbol{Q} - \boldsymbol{\Phi}(\hat{t})| > 0$, then

$$\begin{aligned}\int_0^{\hat{t}} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle \, dt &= \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle \, dt + \int_{t_x}^{\hat{t}} \langle (\boldsymbol{Q} - \boldsymbol{\Phi}(t))^{-1}, \boldsymbol{\Phi}'(t)\rangle \, dt \\ &= \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle \, dt - \log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(\hat{t})| + \log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(t_x)|\end{aligned}$$

which implies

$$\log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(\hat{t})| + \int_0^{\hat{t}} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle \, dt = \log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(t_x)| + \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle \, dt. \tag{5.21}$$

Our main result will be that the limit of the free energy (5.8) is given by minimizing (5.19).

Theorem 5.1.3

The limit of the free energy with self overlaps constrained to $\boldsymbol{Q}$ is

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} F_N^{\varepsilon}(\boldsymbol{Q}) = \inf_{x, \boldsymbol{\Phi}} \mathscr{C}(x, \boldsymbol{\Phi}). \tag{5.22}$$

The infimum is over $x(t)$ and $\boldsymbol{\Phi}(t)$ defined in (5.17) and (5.18) such that $|\boldsymbol{Q} - \boldsymbol{\Phi}(t_x)| > 0$.

Remark 5.1.3. When $n = 1$, (5.19) is identical to the usual 1-dimensional Crisanti–Sommers formula. This is because the only trace parametrization of a one dimensional monotone path is $\boldsymbol{\Phi}(t) = t$.

The Crisanti–Sommers form of the functional has some properties that makes it easier to analyze over the Parisi form. First of all, the Lagrange multiplier that appears in (5.12) is absent in (5.19) and the fixed parameters of the model, $(\vec{\beta}_p)_{p \geq 1}$ and $\vec{h}$ only appear in the first two terms of (5.19). We will also prove that $\mathscr{C}(x, \boldsymbol{\Phi})$ is a locally Lipschitz (See Lemma 5.4.2) with respect to the norm

$$\|x_1 + x_2\|_1 + \|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\|_\infty := \int_0^n |x_1(t) - x_2(t)| \, dt + \max_{i,j \leq n} \left(\sup_{t \in [0,n]} |\boldsymbol{\Phi}_1^{i,j}(t) - \boldsymbol{\Phi}_2^{i,j}(t)| \right).$$

The space of parameters is compact under these norms as a consequence of Prokhorov's theorem and the Arzelà–Ascoli theorem. We will use this observation to show that the functional attains its minimum in the interior of its domain for each fixed path $\boldsymbol{\Phi}$, which will allow us to use variational methods to study the minimizers of $\mathscr{C}(x, \boldsymbol{\Phi})$. This also implies that for fixed $\boldsymbol{\Phi}$, $\mathscr{C}(x, \boldsymbol{\Phi})$ is a convex function in x so there exists a unique minimizer $x_{\boldsymbol{\Phi}}$. Joint convexity in both coordinates remains an open problem.

5.2 The Lower Bound of the Parisi Functional

In this section, we will prove that the infimum of the Crisanti–Sommers functional is a lower bound of the Parisi functional:

Lemma 5.2.1

For any positive definite constraint $\boldsymbol{Q}$, we have

$$\inf_{r,\Lambda,x,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) \geq \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}),$$

where the first infimum is over sequences (5.9) and $\boldsymbol{\Lambda} \in \mathbb{S}_+^n$ such such that $|\boldsymbol{\Lambda}_1| > 0$ and the second infimum is over sequences (5.9) such that $|\boldsymbol{D}_{r-1}| > 0$.

Without loss of generality, we will assume that $\vec{\beta}_2 > 0$. This assumption implies that all entries of $\boldsymbol{\xi}''(\boldsymbol{A})$ are positive for all symmetric matrices $\boldsymbol{A}$, so we don't have to worry about dividing by 0 in the derivation of the critical point conditions. We can make this assumption because both the infimums of $\mathscr{C}_r$ and $\mathscr{P}_r$ are uniformly continuous with respect to $(\vec{\beta}_p)_{p\geq 2}$ [Proposition 5.3.1 and Proposition 5.2.1] so we can send $\vec{\beta}_2 \to 0$ to recover the result in the general case.

To simplify notation, we may also fix $x_{r-1} = 1$. This won't affect the global infimum because the closure of paths satisfying $x_{r-1} = 1$ is equal to the closure of paths satisfying $x_{r-1} \leq 1$.

It remains to prove that for fixed sequences $\underline{x}$ and $r \geq 2$,

$$\inf_{\Lambda,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) \geq \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}).$$

We prove this by examining the behavior of $\mathscr{P}_r$ at its critical point. We will perturb $\mathscr{P}_r$ by adding a logarithmic penalty at the boundary to force the minimizer of $\mathscr{P}_r$ to have positive definite increments. The minimizer will satisfy an interior critical point condition that will allow us to reduce the perturbed functional $\mathscr{P}_r^\varepsilon$ into a perturbed $\mathscr{C}_r^\varepsilon$ functional. These perturbed functionals will converge to $\mathscr{P}_r$ and $\mathscr{C}_r$ in the limit as the size of the barrier tends to 0 using a convexity argument.

5.2.1 Adding a Positive Definite Barrier

We fix $r \geq 2$ and let $\underline{\boldsymbol{Q}} = (\boldsymbol{Q}_k)_{k=0}^r$ denote the monotone sequence of matrices such that $\boldsymbol{Q}_0 = \boldsymbol{0}$ and $\boldsymbol{Q}_r = \boldsymbol{Q}$. We begin our proof by modifying $\mathscr{P}_r$ with a logarithmic barrier term that assigns infinitely large penalties if $\underline{\boldsymbol{Q}}$ is not strictly increasing. Let $\varepsilon > 0$, and consider the barrier function

$$\mathscr{B}_r(\underline{\boldsymbol{Q}}) = -\sum_{0\leq k\leq r-1} \log|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k|.$$

Since $|\boldsymbol{A}| \leq (\frac{\operatorname{tr}(A)}{n})^n$ [Proposition 3.3.12] for all $k \leq r-1$, we have $|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| \leq 1$ so $\mathscr{B}_r \geq 0$. Furthermore, $\mathscr{B}_r \to +\infty$ if $|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| \to 0$ for some $0 \leq k \leq r-1$.

For a fixed strictly increasing sequence such that

$$0 = x_0 < x_1 < \cdots < x_{r-1} = 1, \tag{5.23}$$

we define the functional,

$$\begin{aligned}\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \frac{1}{2}\Big[\langle \vec{h}\vec{h}^\intercal, \boldsymbol{\Lambda}_1^{-1}\rangle + \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - n - \log|\boldsymbol{\Lambda}| + \sum_{1\leq k\leq r-1} \frac{1}{x_k}\log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|} + \langle \boldsymbol{\xi}'(\boldsymbol{Q}_1), \boldsymbol{\Lambda}_1^{-1}\rangle \\ - \sum_{1\leq k\leq r-1} x_k \cdot \operatorname{Sum}\big(\boldsymbol{\theta}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\theta}(\boldsymbol{Q}_k)\big) - \varepsilon \sum_{0\leq k\leq r-1} \log|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k|\Big].\end{aligned} \tag{5.24}$$

Notice that $\mathscr{P}_r^\varepsilon = \mathscr{P}_r + \varepsilon\mathscr{B}_r$ and it decreases pointwise to $\mathscr{P}(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}})$ on its domain as $\varepsilon \to 0$, where $\underline{x}$ is the fixed monotone sequence (5.23). The barrier term forces the minimizers to lie in the interior of the positive definite cone, since $\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) \to +\infty$ if one of the increments $|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| \to 0$. We now examine the behavior of $\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$ at its minimizers and recover a system of critical point equations.

5.2.2 Critical Point Conditions

We will study the first variation of $\mathscr{P}_r^\varepsilon$ to recover critical point conditions for its minimizer. These critical point conditions will relate the increments of $(\boldsymbol{\Lambda}_k)_{k=1}^r$ and $(\boldsymbol{D}_k)_{k=1}^{r-1}$. We want to minimize the function

$$\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) := \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \boldsymbol{Q}_1, \dots, \boldsymbol{Q}_{r-1})$$

over the parameters

$$\boldsymbol{\Lambda} \in \mathscr{L} := \big\{\boldsymbol{\Lambda} \in \mathbb{S}_+^n \mid |\boldsymbol{\Lambda}_1| > 0\big\}$$

and

$$(\boldsymbol{Q}_k)_{k=1}^{r-1} \in \mathcal{Q}_r := \big\{\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_{r-1} \in \mathbb{S}_+^n \mid |\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| > 0,\ \forall\ 0 \le k \le r-1\big\}$$

where $\mathbb{S}_+^n$ is the space of positive semidefinite $n \times n$ matrices. By compactness, $\mathscr{P}_r^\varepsilon$ attains its minimum at some $\boldsymbol{\Lambda} \in \mathscr{L}$ and $(\boldsymbol{Q}_k)_{k=1}^{r-1} \in \mathcal{Q}_r$. Since $\mathscr{B}(\underline{\boldsymbol{Q}}) = \infty$ if the increments are not positive definite, $(\boldsymbol{Q}_k)_{k=1}^{r-1}$ must have also have positive definite increments,

$$|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| > 0 \quad \forall\ 0 \le k \le r-1.$$

This implies that symmetric matrices are admissible variations of $\boldsymbol{\Lambda}$ and $\underline{\boldsymbol{Q}}$ [Proposition 3.3.13]. In particular, if $\boldsymbol{C}$ is a symmetric matrix, then for all t sufficiently small,

$$\boldsymbol{\Lambda} + t\boldsymbol{C} \in \mathscr{L},$$

and for $1 \le k \le r-1$,

$$(\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_k + t\boldsymbol{C}, \dots \boldsymbol{Q}_{r-1}) \in \mathcal{Q}_r.$$

If $\boldsymbol{\Lambda} \in \mathscr{L}$ and $\underline{\boldsymbol{Q}} \in \mathcal{Q}_r$ is a minimizer of $\mathscr{P}_r^\varepsilon$, then for all $1 \le k \le r-1$,

$$\frac{d}{dt}\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda} + t\boldsymbol{C}, \underline{\boldsymbol{Q}})\Big|_{t=0} = 0 \quad \text{and} \quad \frac{d}{dt}\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \boldsymbol{Q}_1, \dots, \boldsymbol{Q}_k + t\boldsymbol{C}, \dots \boldsymbol{Q}_{r-1})\Big|_{t=0} = 0.$$

We can conclude the directional derivatives must be equal to 0 because both $\boldsymbol{C}$ and $-\boldsymbol{C}$ are admissible variations. We can compute the first variation of the functionals explicitly by computing the matrix derivatives of $\mathscr{P}_r^\varepsilon$ and derive some critical point conditions on the minimizers.

To simplify notation, we use summation by parts to write $\mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}})$ as

$$\begin{aligned}\mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) = \frac{1}{2}\Big[\operatorname{tr}(\boldsymbol{\Lambda}\boldsymbol{Q}) - n + \operatorname{tr}((\vec{h}\vec{h}^\intercal + \boldsymbol{\xi}'(\boldsymbol{Q}_1))\boldsymbol{\Lambda}_1^{-1}) - \sum_{2 \le k \le r}\Big(\frac{1}{x_k} - \frac{1}{x_{k-1}}\Big)\log|\boldsymbol{\Lambda}_k| - \frac{1}{x_1}\log|\boldsymbol{\Lambda}_1| \\ - \sum_{1 \le k \le r-1} x_k \cdot \operatorname{tr}\left(\mathbf{1} \times (\boldsymbol{\theta}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\theta}(\boldsymbol{Q}_k))\right)\Big].\end{aligned} \tag{5.25}$$

Since $(\boldsymbol{\Lambda}_k)_{k=1}^r$ is a function of $(\boldsymbol{Q}_k)_{k=1}^r$, for $1 \le p \le r-1$ we have

(1) If $p < \ell$

$$\frac{d\mathbf{\Lambda}_\ell}{d\boldsymbol{Q}_p} = \mathbf{0}$$

(2) If $p > \ell$

$$\frac{d\mathbf{\Lambda}_\ell}{d\boldsymbol{Q}_p} = (x_p - x_{p-1})(\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C})$$

(3) If $p = \ell$

$$\frac{d\mathbf{\Lambda}_\ell}{d\boldsymbol{Q}_p} = x_p(\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C}).$$

Using the formulas in Section 3.3 and the chain rule on (5.25), the derivatives in direction $2\boldsymbol{C}$ (the constant 2 is to cancel the constant factor of $\frac{1}{2}$ in front of $\mathscr{P}_r$) for $2 \le p \le r-1$ are given by

$$\begin{aligned}
\partial_{\boldsymbol{Q}_p}\mathscr{P}_r &= -(x_p - x_{p-1})\operatorname{tr}\Big(\mathbf{\Lambda}_1^{-1}\big(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1)\big)\mathbf{\Lambda}_1^{-1} \times (\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C})\Big) \\
&\quad - (x_p - x_{p-1})\sum_{1\le \ell < p}\Big(\frac{1}{x_\ell} - \frac{1}{x_{\ell-1}}\Big)\operatorname{tr}\big(\mathbf{\Lambda}_\ell^{-1}(\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C})\big)\big) \\
&\quad - x_p\Big(\frac{1}{x_p} - \frac{1}{x_{p-1}}\Big)\operatorname{tr}\big(\mathbf{\Lambda}_p^{-1}(\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C})\big) - \frac{(x_p - x_{p-1})}{x_1}\operatorname{tr}(\mathbf{\Lambda}_1^{-1}(\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C})) \\
&\quad + (x_p - x_{p-1})\operatorname{tr}(\boldsymbol{Q}_p(\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C})) \\
&= -(x_p - x_{p-1})\operatorname{tr}\Big(\mathbf{\Lambda}_1^{-1}\big(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1)\big)\mathbf{\Lambda}_1^{-1} \times (\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C})\Big) \\
&\quad - (x_p - x_{p-1})\sum_{1\le k\le p-1}\frac{1}{x_k}\Big(\operatorname{tr}((\mathbf{\Lambda}_k^{-1} - \mathbf{\Lambda}_{k+1}^{-1}) \times (\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C}))\Big) \\
&\quad + (x_p - x_{p-1})\operatorname{tr}(\boldsymbol{Q}_p \times (\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C})),
\end{aligned} \tag{5.26}$$

and the derivative for $p = 1$ is given by

$$\partial_{\boldsymbol{Q}_1}\mathscr{P}_r = -x_1\operatorname{tr}\Big(\mathbf{\Lambda}_1^{-1}\big(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1)\big)\mathbf{\Lambda}_1^{-1} \times (\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C})\Big) + x_1\operatorname{tr}(\boldsymbol{Q}_1 \times (\boldsymbol{\xi}''(\boldsymbol{Q}_p) \odot \boldsymbol{C})). \tag{5.27}$$

It is also straightforward to see that

$$\partial_{\mathbf{\Lambda}}\mathscr{P}_r^\varepsilon = \langle \boldsymbol{Q}, \boldsymbol{C}\rangle - \langle \mathbf{\Lambda}^{-1}, \boldsymbol{C}\rangle - \langle \mathbf{\Lambda}_1^{-1}\big(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1)\big)\mathbf{\Lambda}_1^{-1}, \boldsymbol{C}\rangle + \sum_{1\le k\le r-1}\frac{1}{x_k}\langle \mathbf{\Lambda}_{k+1}^{-1} - \mathbf{\Lambda}_k^{-1}, \boldsymbol{C}\rangle.$$

Derivation of Critical Point Equations

From the formulas for the partial derivatives, we can derive a system of implicit equations satisfied by the minimizers of $\mathscr{P}_r^\varepsilon$:

(a) The directional derivatives of $\mathscr{P}_r^\varepsilon$ with respect to $\mathbf{\Lambda}$ in the symmetric direction $2\boldsymbol{C}$ is

$$\partial_{\mathbf{\Lambda}}\mathscr{P}_r^\varepsilon = \langle \boldsymbol{Q}, \boldsymbol{C}\rangle - \langle \mathbf{\Lambda}^{-1}, \boldsymbol{C}\rangle - \langle \mathbf{\Lambda}_1^{-1}\big(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1)\big)\mathbf{\Lambda}_1^{-1}, \boldsymbol{C}\rangle + \sum_{1\le k\le r-1}\frac{1}{x_k}\langle \mathbf{\Lambda}_{k+1}^{-1} - \mathbf{\Lambda}_k^{-1}, \boldsymbol{C}\rangle. \tag{5.28}$$

At the minimizer, we require

$$\partial_{\mathbf{\Lambda}}\mathscr{P}_r^\varepsilon := \frac{d}{dt}\mathscr{P}_r^\varepsilon(\mathbf{\Lambda} + 2t\boldsymbol{C}, \underline{\boldsymbol{Q}})\Big|_{t=0} = 0.$$

This equality holds for all symmetric directions $\boldsymbol{C}$, so the minimizer must satisfy the equation [Proposition 3.3.8]

$$\boldsymbol{Q} = \boldsymbol{\Lambda}^{-1} + \boldsymbol{\Lambda}_1^{-1}\big(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1)\big)\boldsymbol{\Lambda}_1^{-1} + \sum_{1\le k\le r-1} \frac{1}{x_k}(\boldsymbol{\Lambda}_k^{-1} - \boldsymbol{\Lambda}_{k+1}^{-1}). \tag{5.29}$$

(b) For $2 \le p \le r-1$, the directional derivatives of $\mathscr{P}_r^\varepsilon$ with respect to $\boldsymbol{Q}_p$ in the symmetric direction $2\boldsymbol{C}$ is [Equation 5.26]

$$\begin{aligned}\partial_{\boldsymbol{Q}_p}\mathscr{P}_r^\varepsilon = &-(x_p - x_{p-1})\langle \boldsymbol{\Lambda}_1^{-1}\big(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1)\big)\boldsymbol{\Lambda}_1^{-1}, \boldsymbol{\xi}''(\boldsymbol{Q}_p)\odot \boldsymbol{C}\rangle \\ &- (x_p - x_{p-1})\sum_{1\le k\le p-1}\frac{1}{x_k}\langle \boldsymbol{\Lambda}_k^{-1} - \boldsymbol{\Lambda}_{k+1}^{-1}, \boldsymbol{\xi}''(\boldsymbol{Q}_p)\odot \boldsymbol{C}\rangle \\ &+ (x_p - x_{p-1})\langle \boldsymbol{Q}_p, \boldsymbol{\xi}''(\boldsymbol{Q}_p)\odot \boldsymbol{C}\rangle \\ &+ \varepsilon\langle (\boldsymbol{Q}_{p+1} - \boldsymbol{Q}_p)^{-1}, \boldsymbol{C}\rangle - \varepsilon\langle (\boldsymbol{Q}_p - \boldsymbol{Q}_{p-1})^{-1}, \boldsymbol{C}\rangle.\end{aligned} \tag{5.30}$$

At the minimizer, we require

$$\partial_{\boldsymbol{Q}_p}\mathscr{P}_r^\varepsilon := \frac{d}{dt}\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \boldsymbol{Q}_1, \dots, \boldsymbol{Q}_p + 2t\boldsymbol{C}, \dots, \boldsymbol{Q}_{r-1})\Big|_{t=0} = 0.$$

This equality holds for all symmetric directions $\boldsymbol{C}$, so the minimizer satisfies the critical point equation

$$\boldsymbol{Q}_p = \boldsymbol{\Lambda}_1^{-1}\big(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1)\big)\boldsymbol{\Lambda}_1^{-1} + \sum_{1\le k\le p-1}\frac{1}{x_k}(\boldsymbol{\Lambda}_k^{-1} - \boldsymbol{\Lambda}_{k+1}^{-1}) - \varepsilon\boldsymbol{E}_p \tag{5.31}$$

where

$$\boldsymbol{E}_p := \frac{1}{x_p - x_{p-1}}\Big((\boldsymbol{Q}_{p+1} - \boldsymbol{Q}_p)^{-1} - (\boldsymbol{Q}_p - \boldsymbol{Q}_{p-1})^{-1}\Big) \oslash \boldsymbol{\xi}''(\boldsymbol{Q}_p). \tag{5.32}$$

The notation $\oslash$ refers to the Hadamard division operation (entry-wise division). $\boldsymbol{E}_p$ is well defined since the fixed $(x_p)_{p=1}^r$ in (5.23) is strictly monotone and $\vec{\beta}_2 > 0$ so all entries of $\boldsymbol{\xi}''(\boldsymbol{A})$ is positive.

(c) For $p = 1$, the directional derivatives of $\mathscr{P}_r^\varepsilon$ with respect to $\boldsymbol{Q}_1$ in the symmetric direction $2\boldsymbol{C}$ is [Equation 5.27]

$$\begin{aligned}\partial_{\boldsymbol{Q}_1}\mathscr{P}_r^\varepsilon = &-x_1\langle \boldsymbol{\Lambda}_1^{-1}\big(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1)\big)\boldsymbol{\Lambda}_1^{-1}, \boldsymbol{\xi}''(\boldsymbol{Q}_1)\odot \boldsymbol{C}\rangle + x_1\langle \boldsymbol{Q}_1, \boldsymbol{\xi}''(\boldsymbol{Q}_1)\odot \boldsymbol{C}\rangle \\ &+ \varepsilon\langle (\boldsymbol{Q}_2 - \boldsymbol{Q}_1)^{-1}, \boldsymbol{C}\rangle - \varepsilon\langle \boldsymbol{Q}_1^{-1}, \boldsymbol{C}\rangle.\end{aligned} \tag{5.33}$$

At the minimizer, we require

$$\partial_{\boldsymbol{Q}_1}\mathscr{P}_r^\varepsilon := \frac{d}{dt}\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \boldsymbol{Q}_1 + 2t\boldsymbol{C}, \dots, \boldsymbol{Q}_{r-1})\Big|_{t=0} = 0.$$

This equality holds for all symmetric directions $\boldsymbol{C}$, so the minimizer satisfies the critical point equation

$$\boldsymbol{Q}_1 = \boldsymbol{\Lambda}_1^{-1}\big(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1)\big)\boldsymbol{\Lambda}_1^{-1} - \varepsilon\boldsymbol{E}_1, \tag{5.34}$$

where $\boldsymbol{E}_1$ is given by the formula in (5.32) with $p = 1$.

For $1 \le p \le r-1$, the critical point equations (5.31) and (5.34) can be expressed as

$$\boldsymbol{Q}_p = \boldsymbol{\Lambda}_1^{-1}(\vec{h}\vec{h}^T + \boldsymbol{\xi}'(\boldsymbol{Q}_1))\boldsymbol{\Lambda}_1^{-1} + \sum_{1\le k\le p-1} \frac{1}{x_k}(\boldsymbol{\Lambda}_k^{-1} - \boldsymbol{\Lambda}_{k+1}^{-1}) - \varepsilon \boldsymbol{E}_p \tag{5.35}$$

where $\boldsymbol{E}_r := \boldsymbol{0}$ and $\boldsymbol{E}_p$ was defined in (5.32). These critical point conditions can be used to relate $\boldsymbol{\Lambda}_k$ in $\mathscr{P}_r^\varepsilon$ with the $\boldsymbol{D}_k$ terms in $\mathscr{C}_r$. Taking differences of the critical point conditions (5.29) and (5.35), we can conclude that,

$$\begin{aligned} x_{r-1}(\boldsymbol{Q} - \boldsymbol{Q}_{r-1}) &= x_{r-1}\boldsymbol{\Lambda}^{-1} + (\boldsymbol{\Lambda}_{r-1}^{-1} - \boldsymbol{\Lambda}^{-1}) - \varepsilon x_{r-1}(\boldsymbol{E}_r - \boldsymbol{E}_{r-1}) \\ &= \boldsymbol{\Lambda}_{r-1}^{-1} - \varepsilon x_{r-1}(\boldsymbol{E}_r - \boldsymbol{E}_{r-1}) \end{aligned}$$

since $x_{r-1} = 1$, and for $1 \le p \le r-2$, we can conclude that

$$x_p(\boldsymbol{Q}_{p+1} - \boldsymbol{Q}_p) = \boldsymbol{\Lambda}_p^{-1} - \boldsymbol{\Lambda}_{p+1}^{-1} - \varepsilon x_p(\boldsymbol{E}_{p+1} - \boldsymbol{E}_p).$$

Taking sums of the above, we have for $1 \le p \le r-1$,

$$\sum_{p\le k\le r-1} x_k(\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k) + \varepsilon \sum_{p\le k\le r-1} x_k(\boldsymbol{E}_{k+1} - \boldsymbol{E}_k) = \boldsymbol{D}_p + \varepsilon \sum_{p\le k\le r-1} x_k(\boldsymbol{E}_{k+1} - \boldsymbol{E}_k) = \boldsymbol{\Lambda}_p^{-1}. \tag{5.36}$$

We will summarize this critical point condition in the following lemma.

Lemma 5.2.2

For fixed $r \ge 2$, if $\boldsymbol{\Lambda}$ and $\underline{\boldsymbol{Q}}$ is a minimizer of $\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$ and $\vec{\beta}_2 > 0$, then $\boldsymbol{\Lambda}$ and $\underline{\boldsymbol{Q}}$ satisfy the following critical point equations

$$\boldsymbol{\Lambda}_p^{-1} = \boldsymbol{D}_p(\varepsilon) \qquad \text{for } 1 \le p \le r-1 \tag{5.37}$$

where

$$\boldsymbol{D}_p(\varepsilon) = \boldsymbol{D}_p + \varepsilon\overline{\boldsymbol{E}}_p \qquad \text{and} \qquad \overline{\boldsymbol{E}}_p = \sum_{p\le k\le r-1} x_k(\boldsymbol{E}_{k+1} - \boldsymbol{E}_k). \tag{5.38}$$

5.2.3 Reduction to an approximate Crisanti–Sommers functional

In this subsection, we will reduce $\mathscr{P}_r^\varepsilon$ defined in (5.24) to an approximate Cristanti–Sommers functional. If $\boldsymbol{\Lambda}$ and $\underline{\boldsymbol{Q}}$ satisfy the critical point conditions (5.37), we will show that $\mathscr{P}_r^\varepsilon$ can be reduced to

$$\begin{aligned} \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}}) = \frac{1}{2}\Big[&\langle \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{D}_1(\varepsilon)\rangle + \frac{1}{x_{r-1}}\log|\boldsymbol{D}_{r-1}(\varepsilon)| - \sum_{1\le k\le r-2}\frac{1}{x_k}\log\frac{|\boldsymbol{D}_{k+1}(\varepsilon)|}{|\boldsymbol{D}_k(\varepsilon)|} + \langle \boldsymbol{Q}_1, \boldsymbol{D}_1^{-1}(\varepsilon)\rangle \\ &+ \sum_{1\le k\le r-1} x_k \cdot \mathrm{Sum}\left(\boldsymbol{\xi}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}(\boldsymbol{Q}_k)\right) \\ &- \varepsilon \sum_{1\le k\le r-2}\langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle - \sum_{1\le k\le r-2}\frac{\varepsilon}{x_k}\langle \boldsymbol{D}_{k+1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle \\ &- \varepsilon\langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}), \overline{\boldsymbol{E}}_{r-1}\rangle - \varepsilon \sum_{0\le k\le r-1}\log|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k|\Big]. \end{aligned} \tag{5.39}$$

Notice that $\mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}})$ is of the same form as $\mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}})$, but with $\boldsymbol{D}_k$ replaced by $\boldsymbol{D}_k(\varepsilon)$ and some additional error terms. If we set $\varepsilon = 0$, then the error terms in the second line all vanish and we are left with the usual $\mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}})$ functional. In the next subsection, we will show that we can bound the minimum of $\mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}})$ with $\mathscr{C}_r$ evaluated at a different path to remove the error terms.

Lemma 5.2.3

If $\boldsymbol{\Lambda}$ and $\underline{\boldsymbol{Q}}$ satisfy the critical point conditions (5.37), then

$$\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}}).$$

Proof. The reduction of $\mathscr{P}_r^\varepsilon$ to $\mathscr{C}_r^\varepsilon$ is a straightforward, but tedious computation. We will show

$$2(\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) - \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}})) = 0.$$

If (5.37) holds, then

$$\boldsymbol{\Lambda}_{k+1} - \boldsymbol{\Lambda}_k = x_k(\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_k)) \qquad 1 \le k \le r-1 \tag{5.40}$$

$$\boldsymbol{D}_k - \boldsymbol{D}_{k+1} = x_k(\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k) \qquad 1 \le k \le r-1 \tag{5.41}$$

$$\boldsymbol{\Lambda}_k^{-1} = \boldsymbol{D}_k(\varepsilon) \qquad 1 \le k \le r-1 \tag{5.42}$$

where $\boldsymbol{D}_r := \boldsymbol{0}$. These identities will be used multiple times throughout this proof.

We begin by observing that the external fields cancel if (5.37) holds,

$$\langle \vec{h}\vec{h}^\mathsf{T}, \boldsymbol{\Lambda}_1^{-1} \rangle \overset{(5.42)}{=} \langle \vec{h}\vec{h}^\mathsf{T}, \boldsymbol{D}_1(\varepsilon) \rangle. \tag{5.43}$$

Next, we simplify the summation of the logarithm terms in $\mathscr{P}_r^\varepsilon$ using the fact $x_{r-1} = 1$,

$$\begin{aligned} -\log|\boldsymbol{\Lambda}| + \sum_{1\le k\le r-1} \frac{1}{x_k} \log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|} &= -\log|\boldsymbol{\Lambda}| + \frac{1}{x_{r-1}}(\log|\boldsymbol{\Lambda}| - \log|\boldsymbol{\Lambda}_{r-1}|) + \sum_{1\le k\le r-2} \frac{1}{x_k} \log\frac{|\boldsymbol{\Lambda}_{k+1}|}{|\boldsymbol{\Lambda}_k|} \\ &\overset{(5.42)}{=} \frac{1}{x_{r-1}} \log|\boldsymbol{D}_{r-1}(\varepsilon)| - \sum_{1\le k\le r-2} \frac{1}{x_k} \log\frac{|\boldsymbol{D}_{k+1}(\varepsilon)|}{|\boldsymbol{D}_k(\varepsilon)|}. \end{aligned} \tag{5.44}$$

Therefore, the log determinant terms in $\mathscr{P}_r^\varepsilon$ and $\mathscr{C}_r^\varepsilon$ also cancel.

Since $\boldsymbol{\theta}(\boldsymbol{A}) = \boldsymbol{A} \odot \boldsymbol{\xi}'(\boldsymbol{A}) - \boldsymbol{\xi}(\boldsymbol{A})$ and $\mathrm{Sum}(\boldsymbol{A} \odot \boldsymbol{B}) = \langle \boldsymbol{A}, \boldsymbol{B} \rangle$, the remaining terms in $2(\mathscr{P}_r^\varepsilon - \mathscr{C}_r^\varepsilon)$ are

$$-\sum_{1\le k\le r-2} x_k\Big(\langle \boldsymbol{Q}_{k+1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle - \langle \boldsymbol{Q}_k, \boldsymbol{\xi}'(\boldsymbol{Q}_k)\rangle\Big) - \Big(\langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q})\rangle - \langle \boldsymbol{Q}_{r-1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle\Big) \tag{5.45}$$

$$+\varepsilon \sum_{1\le k\le r-2} \langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle + \sum_{1\le k\le r-2} \frac{\varepsilon}{x_k} \langle \boldsymbol{D}_{k+1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle \tag{5.46}$$

$$+\langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - n + \langle \boldsymbol{\xi}'(\boldsymbol{Q}_1), \boldsymbol{\Lambda}_1^{-1}\rangle - \langle \boldsymbol{Q}_1, \boldsymbol{D}_1^{-1}(\varepsilon)\rangle + \varepsilon\langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle - \varepsilon\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}), \overline{\boldsymbol{E}}_{r-1}\rangle. \tag{5.47}$$

We will show that (5.45) will cancel (5.46) and (5.47) at the critical point. We start by simplifying the

summation term in (5.45),

$$-\sum_{1\le k\le r-2} x_k\Big(\langle \boldsymbol{Q}_{k+1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle - \langle \boldsymbol{Q}_k, \boldsymbol{\xi}'(\boldsymbol{Q}_k)\rangle\Big) \tag{5.48}$$

$$\begin{aligned}
&= -\sum_{1\le k\le r-2}\Big(\langle x_k(\boldsymbol{Q}_{k+1}-\boldsymbol{Q}_k), \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle + \langle \boldsymbol{Q}_k, x_k(\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})-\boldsymbol{\xi}'(\boldsymbol{Q}_k))\rangle\Big)\\
&\overset{(5.40)(5.41)}{=} -\sum_{1\le k\le r-2}\Big(\langle \boldsymbol{D}_k-\boldsymbol{D}_{k+1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle + \langle \boldsymbol{Q}_k, \boldsymbol{\Lambda}_{k+1}-\boldsymbol{\Lambda}_k\rangle\Big)\\
&\overset{(5.42)}{=} -\sum_{1\le k\le r-2}\Big(\langle \boldsymbol{D}_k(\varepsilon)-\boldsymbol{D}_{k+1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle + \langle \boldsymbol{Q}_k, \boldsymbol{\Lambda}_{k+1}-\boldsymbol{\Lambda}_k\rangle\Big) - \varepsilon\sum_{1\le k\le r-2}\langle \overline{\boldsymbol{E}}_{k+1}-\overline{\boldsymbol{E}}_k, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle.
\end{aligned} \tag{5.49}$$

Using summation by parts and (5.37), the first summation in (5.49) is equal to

$$-\sum_{1\le k\le r-2}\Big(\langle \boldsymbol{D}_k(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})-\boldsymbol{\xi}'(\boldsymbol{Q}_k)\rangle + \langle \boldsymbol{D}_{k+1}^{-1}(\varepsilon), \boldsymbol{Q}_k-\boldsymbol{Q}_{k+1}\rangle\Big) \tag{5.50}$$

$$-\langle \boldsymbol{D}_1(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_1)\rangle + \langle \boldsymbol{D}_{r-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle - \langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \boldsymbol{Q}_{r-1}\rangle + \langle \boldsymbol{D}_1^{-1}(\varepsilon), \boldsymbol{Q}_1\rangle. \tag{5.51}$$

The critical point conditions (5.37) implies

$$\begin{aligned}
\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})-\boldsymbol{\xi}'(\boldsymbol{Q}_k) &\overset{(5.40)}{=} \frac{1}{x_k}(\boldsymbol{\Lambda}_{k+1}-\boldsymbol{\Lambda}_k)\\
&\overset{(5.42)}{=} \frac{1}{x_k}\boldsymbol{D}_k^{-1}(\varepsilon)(\boldsymbol{D}_k(\varepsilon)-\boldsymbol{D}_{k+1}(\varepsilon))\boldsymbol{D}_{k+1}^{-1}(\varepsilon)\\
&\overset{(5.41)}{=} \boldsymbol{D}_k^{-1}(\varepsilon)(\boldsymbol{Q}_{k+1}-\boldsymbol{Q}_k)\boldsymbol{D}_{k+1}^{-1}(\varepsilon) + \frac{\varepsilon}{x_k}\boldsymbol{D}_k^{-1}(\varepsilon)(\overline{\boldsymbol{E}}_k-\overline{\boldsymbol{E}}_{k+1})\boldsymbol{D}_{k+1}^{-1}(\varepsilon),
\end{aligned}$$

which combined with the fact $\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{CAB})$ implies the summation term (5.50) simplifies to

$$-\sum_{1\le k\le r-2}\frac{\varepsilon}{x_k}\langle \boldsymbol{D}_{k+1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k-\overline{\boldsymbol{E}}_{k+1}\rangle. \tag{5.52}$$

Substituting (5.52) into (5.49) and adding the boundary terms (5.51) implies that

$$\begin{aligned}
(5.48) = &-\varepsilon\sum_{1\le k\le r-2}\langle \overline{\boldsymbol{E}}_{k+1}-\overline{\boldsymbol{E}}_k, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle - \varepsilon\sum_{1\le k\le r-2}\frac{1}{x_k}\langle \boldsymbol{D}_{k+1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k-\overline{\boldsymbol{E}}_{k+1}\rangle\\
&-\langle \boldsymbol{D}_1(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_1)\rangle + \langle \boldsymbol{D}_{r-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle - \langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \boldsymbol{Q}_{r-1}\rangle + \langle \boldsymbol{D}_1^{-1}(\varepsilon), \boldsymbol{Q}_1\rangle.
\end{aligned} \tag{5.53}$$

Substituting (5.53) into (5.45) implies

$$\begin{aligned}2(\mathscr{P}_r^\varepsilon - \mathscr{C}_r^\varepsilon) &= \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - n + \langle \boldsymbol{\xi}'(\boldsymbol{Q}_1), \boldsymbol{\Lambda}_1^{-1}\rangle - \langle \boldsymbol{Q}_1, \boldsymbol{D}_1^{-1}(\varepsilon)\rangle + \varepsilon\langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle - \varepsilon\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}), \overline{\boldsymbol{E}}_{r-1}\rangle \\ &\quad - \langle \boldsymbol{D}_1(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_1)\rangle + \langle \boldsymbol{D}_{r-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle - \langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \boldsymbol{Q}_{r-1}\rangle + \langle \boldsymbol{D}_1^{-1}(\varepsilon), \boldsymbol{Q}_1\rangle \\ &\quad - \langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q})\rangle + \langle \boldsymbol{Q}_{r-1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle \\ &\overset{(5.42)}{=} \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - \operatorname{tr}(\boldsymbol{I}) + \varepsilon\langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle - \varepsilon\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}), \overline{\boldsymbol{E}}_{r-1}\rangle \\ &\quad + \langle \boldsymbol{D}_{r-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle - \langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \boldsymbol{Q}_{r-1}\rangle - \langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q})\rangle + \langle \boldsymbol{Q}_{r-1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle \\ &= \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - \operatorname{tr}(\boldsymbol{I}) + \varepsilon\langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle + \langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle - \langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \boldsymbol{Q}_{r-1}\rangle - \langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q})\rangle.\end{aligned} \tag{5.54}$$

since $\boldsymbol{D}_{r-1}(\varepsilon) = \boldsymbol{Q} - \boldsymbol{Q}_{r-1} + \varepsilon\overline{\boldsymbol{E}}_{r-1}$. We will show that the $\langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \boldsymbol{Q}_{r-1}\rangle$ term cancels all the remaining terms. Using the critical point condition and the definitions of $\boldsymbol{\Lambda}_{r-1}$ defined in (5.13) and $\boldsymbol{D}_{r-1}(\varepsilon)$ defined in (5.16) and (5.38), we get

$$\begin{aligned}\boldsymbol{D}_{r-1}^{-1}(\varepsilon)\boldsymbol{Q}_{r-1} &= \boldsymbol{D}_{r-1}^{-1}(\varepsilon)(-\boldsymbol{D}_{r-1}(\varepsilon) + \boldsymbol{Q} - \varepsilon\overline{\boldsymbol{E}}_{r-1}) \\ &= -\boldsymbol{I} + \boldsymbol{D}_{r-1}^{-1}(\varepsilon)\boldsymbol{Q} + \varepsilon\boldsymbol{D}_{r-1}^{-1}(\varepsilon)\overline{\boldsymbol{E}}_{r-1} \\ &\overset{(5.42)}{=} -\boldsymbol{I} + (\boldsymbol{\Lambda} - \boldsymbol{\xi}'(\boldsymbol{Q}) + \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}))\boldsymbol{Q} + \varepsilon\boldsymbol{D}_{r-1}^{-1}(\varepsilon)\overline{\boldsymbol{E}}_{r-1}.\end{aligned}$$

Taking the trace and using the fact $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA})$ implies

$$\langle \boldsymbol{D}_{r-1}^{-1}, \boldsymbol{Q}_{r-1}\rangle = -\operatorname{tr}(\boldsymbol{I}) + \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - \langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q})\rangle + \langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle + \varepsilon\langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle. \tag{5.55}$$

Substituting (5.55) into (5.54) cancels out all remaining terms, so

$$2\big(\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) - \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}})\big) = 0.$$

□

5.2.4 Removing the Error Terms

We now bound the minimum of the perturbed functional $\mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}})$ defined in (5.39) with $\mathscr{C}_r$ evaluated at a different path of matrices. We can't simply send $\varepsilon \to 0$ to remove the error terms, because we do not know that $\varepsilon\boldsymbol{E}_k \to \boldsymbol{0}$ since $\boldsymbol{E}_k$ depends on ε. Consider the monotone path encoded by the sequences

$$\begin{aligned} x_{-1} = 0 = \ x_0 < \ x_1 < \ldots < \ x_{r-2} < \ x_{r-1} = \ \ 1 \\ \boldsymbol{0} = \tilde{\boldsymbol{Q}}_0 < \tilde{\boldsymbol{Q}}_1 < \ldots < \tilde{\boldsymbol{Q}}_{r-2} < \tilde{\boldsymbol{Q}}_{r-1} < \tilde{\boldsymbol{Q}}_r = \boldsymbol{Q}\end{aligned} \tag{5.56}$$

where $\tilde{\boldsymbol{Q}}_p = \boldsymbol{Q}_p + \varepsilon\boldsymbol{E}_p$ for $1 \le p \le r$. We first note that $(\tilde{\boldsymbol{Q}}_k)_{k=1}^r \in \mathcal{Q}_r$. By definition,

$$\tilde{\boldsymbol{D}}_p := \sum_{p\le k\le r-1} x_k(\tilde{\boldsymbol{Q}}_{k+1} - \tilde{\boldsymbol{Q}}_k) = \sum_{p\le k\le r-1} x_k(\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k) + \varepsilon \sum_{p\le k\le r-1} x_k(\boldsymbol{E}_{k+1} - \boldsymbol{E}_k) = \boldsymbol{D}_p(\varepsilon). \tag{5.57}$$

Since $|\boldsymbol{Q}_k - \boldsymbol{Q}_{k-1}| > 0$ implies $|\boldsymbol{\xi}'(\boldsymbol{Q}_k) - \boldsymbol{\xi}'(\boldsymbol{Q}_{k-1})| > 0$ [Proposition 3.3.14], the critical point condition (5.37) implies the path $(\tilde{\boldsymbol{Q}}_k)_{k=1}^r$ has positive definite increments for $1 \le k \le r-1$,

$$x_k(\tilde{\boldsymbol{Q}}_{k+1} - \tilde{\boldsymbol{Q}}_k) = \tilde{\boldsymbol{D}}_k - \tilde{\boldsymbol{D}}_{k+1} = \boldsymbol{\Lambda}_k^{-1} - \boldsymbol{\Lambda}_{k+1}^{-1} > 0.$$

The boundary conditions are also satisfied since $\boldsymbol{E}_r = 0$ implies that $\tilde{\boldsymbol{Q}}_r = \boldsymbol{Q}_r + \boldsymbol{E}_r = \boldsymbol{Q}$ and the critical point condition for $\boldsymbol{Q}_1$ (5.34) implies that

$$\tilde{\boldsymbol{Q}}_1 = \boldsymbol{Q}_1 + \varepsilon \boldsymbol{E}_1 = \boldsymbol{\Lambda}_1^{-1}(\vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}_1))\boldsymbol{\Lambda}_1^{-1} > 0.$$

Using convexity, we will prove that the perturbed functional $\mathscr{C}_r^\varepsilon$ can be lower bounded by $\mathscr{C}_r$ evaluated at the path encoded by (5.56),

$$\mathscr{C}_r^\varepsilon((\boldsymbol{Q}_k)_{k=1}^r) \ge \mathscr{C}_r(\underline{x}, (\tilde{\boldsymbol{Q}}_k)_{k=1}^r) \tag{5.58}$$

provided that $(\boldsymbol{Q}_k)_{k=1}^r$ satisfies the critical point conditions (5.37). Since the sequences of matrices $(\tilde{\boldsymbol{Q}}_k)_{k=1}^r$ is in $\mathcal{Q}_r$, we get the obvious lower bound,

$$\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, (\boldsymbol{Q}_k)_{k=1}^r) = \mathscr{C}_r^\varepsilon(\underline{x}, (\boldsymbol{Q}_k)_{k=1}^r) \ge \mathscr{C}_r(\underline{x}, (\tilde{\boldsymbol{Q}}_k)_{k=1}^r) \ge \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}).$$

The lower bound does not depend on the discretization r, ε, nor the fixed sequence (5.23). Therefore, we can minimize the upper bound over sequences (5.23), r and ε to prove the required lower bound,

$$\inf_{r,\Lambda,x,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) \ge \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}).$$

We now prove the lower bound (5.58).

Lemma 5.2.4

For all $\varepsilon > 0$, if $\underline{\boldsymbol{Q}}$ satisfies the critical point conditions (5.37), then

$$\mathscr{C}_r^\varepsilon((\boldsymbol{Q}_k)_{k=1}^r) \ge \mathscr{C}_r(\underline{x}, (\tilde{\boldsymbol{Q}}_k)_{k=1}^r).$$

Proof. Since $\tilde{\boldsymbol{D}}_p = \boldsymbol{D}_p(\varepsilon)$ and the barrier $\mathscr{B}_r \ge 0$, it remains to show that

$$\begin{aligned}\langle \boldsymbol{Q}_1, \boldsymbol{D}_1^{-1}(\varepsilon)\rangle + \sum_{1\le k\le r-1} x_k \cdot \mathrm{Sum}\left(\boldsymbol{\xi}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}(\boldsymbol{Q}_k)\right) - \varepsilon\langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}), \overline{\boldsymbol{E}}_{r-1}\rangle \\ - \varepsilon \sum_{1\le k\le r-2} \langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle - \sum_{1\le k\le r-2} \frac{\varepsilon}{x_k}\langle \boldsymbol{D}_{k+1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle\end{aligned} \tag{5.59}$$

is bounded below by

$$\begin{aligned}&\langle \tilde{\boldsymbol{Q}}_1, \tilde{\boldsymbol{D}}_1^{-1}\rangle + \sum_{1\le k\le r-1} x_k \cdot \mathrm{Sum}\left(\boldsymbol{\xi}(\tilde{\boldsymbol{Q}}_{k+1}) - \boldsymbol{\xi}(\tilde{\boldsymbol{Q}}_k)\right) \\ &= \langle \tilde{\boldsymbol{Q}}_1, \tilde{\boldsymbol{D}}_1^{-1}\rangle + \sum_{1\le k\le r-1} (x_{k-1} - x_k)\,\mathrm{Sum}(\boldsymbol{\xi}(\tilde{\boldsymbol{Q}}_k)) + x_{r-1}\,\mathrm{Sum}(\boldsymbol{\xi}(\tilde{\boldsymbol{Q}}_r)).\end{aligned}$$

We will use convexity of the $\boldsymbol{\xi}$ terms to absorb the ε error terms in (5.59). The definition of $\overline{\boldsymbol{E}}_k$ in

(5.38) implies that

$$\overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1} = x_k(\boldsymbol{E}_{k+1} - \boldsymbol{E}_k) \qquad 1 \le k \le r-1. \tag{5.60}$$

Using summation by parts and (5.60) the last four ε terms in (5.59) can be simplified to

$$\begin{aligned}
&- \varepsilon \sum_{1\le k\le r-2} \langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle - \varepsilon \sum_{1\le k\le r-2} \frac{1}{x_k}\langle \boldsymbol{D}_{k+1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle \\
&- \varepsilon\langle \boldsymbol{D}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}), \overline{\boldsymbol{E}}_{r-1}\rangle \\
&\overset{(5.42)(5.60)}{=} \varepsilon \sum_{1\le k\le r-2} \langle x_k(\boldsymbol{E}_{k+1} - \boldsymbol{E}_k), \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle - \varepsilon \sum_{1\le k\le r-2} \langle \boldsymbol{\Lambda}_{k+1}, \boldsymbol{E}_{k+1} - \boldsymbol{E}_k\rangle \\
&- \varepsilon\langle \boldsymbol{\Lambda}_{r-1}, \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}), \overline{\boldsymbol{E}}_{r-1}\rangle \\
&= \varepsilon \sum_{1\le k\le r-2} \langle x_k(\boldsymbol{E}_{k+1} - \boldsymbol{E}_k), \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle + \varepsilon \sum_{1\le k\le r-2} \langle \boldsymbol{\Lambda}_{k+1} - \boldsymbol{\Lambda}_k, \boldsymbol{E}_k\rangle \\
&- \varepsilon\langle \boldsymbol{\Lambda}_{r-1}, \boldsymbol{E}_{r-1}\rangle + \varepsilon\langle \boldsymbol{\Lambda}_1, \boldsymbol{E}_1\rangle - \varepsilon\langle \boldsymbol{\Lambda}_{r-1}, \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}), \overline{\boldsymbol{E}}_{r-1}\rangle \\
&\overset{(5.40)}{=} \varepsilon \sum_{1\le k\le r-2} \langle x_k(\boldsymbol{E}_{k+1} - \boldsymbol{E}_k), \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle + \varepsilon \sum_{1\le k\le r-2} \langle x_k(\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_k)), \boldsymbol{E}_k\rangle \\
&- \varepsilon\langle \boldsymbol{\Lambda}_{r-1}, \boldsymbol{E}_{r-1}\rangle + \varepsilon\langle \boldsymbol{\Lambda}_1, \boldsymbol{E}_1\rangle - \varepsilon\langle \boldsymbol{\Lambda}_{r-1}, \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}), \overline{\boldsymbol{E}}_{r-1}\rangle \\
&= \varepsilon \sum_{1\le k\le r-1} x_k\big(\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}), \boldsymbol{E}_{k+1}\rangle - \langle \boldsymbol{\xi}'(\boldsymbol{Q}_k), \boldsymbol{E}_k\rangle\big) + \varepsilon\langle \boldsymbol{\Lambda}_1, \boldsymbol{E}_1\rangle
\end{aligned}$$

since $\boldsymbol{E}_r = 0$ and $\overline{\boldsymbol{E}}_{r-1} = \boldsymbol{E}_r - \boldsymbol{E}_{r-1} = -\boldsymbol{E}_{r-1}$. Therefore, excluding the leftover $\langle \boldsymbol{Q}_1, \boldsymbol{D}_1^{-1}(\varepsilon)\rangle + \varepsilon\langle \boldsymbol{\Lambda}_1, \boldsymbol{E}_1\rangle$ term, (5.59) is equal to

$$\begin{aligned}
&\sum_{1\le k\le r-1} x_k \cdot \mathrm{Sum}\left(\boldsymbol{\xi}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}(\boldsymbol{Q}_k)\right) + \varepsilon \sum_{1\le k\le r-1} x_k \cdot \left(\langle \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}), \boldsymbol{E}_{k+1}\rangle - \langle \boldsymbol{\xi}'(\boldsymbol{Q}_k), \boldsymbol{E}_k\rangle\right) \\
&= \sum_{1\le k\le r-1} (x_{k-1} - x_k)\,\mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{Q}_k) + \varepsilon\boldsymbol{\xi}'(\boldsymbol{Q}_k) \odot \boldsymbol{E}_k) + x_{r-1}\,\mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{Q}_r + \varepsilon\boldsymbol{E}_r)).
\end{aligned} \tag{5.61}$$

Since $\boldsymbol{\xi}(\boldsymbol{A})$ is convex [Proposition 3.3.10] and $(x_{k-1} - x_k) \le 0$, we also have

$$(x_{k-1} - x_k)\,\mathrm{Sum}(\boldsymbol{\xi}(\tilde{\boldsymbol{Q}}_k)) = (x_{k-1} - x_k)\,\mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{Q}_k + \varepsilon\boldsymbol{E}_k)) \le (x_{k-1} - x_k)\,\mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{Q}_k) + \varepsilon\boldsymbol{\xi}'(\boldsymbol{Q}_k) \odot \boldsymbol{E}_k). \tag{5.62}$$

Furthermore, the leftover terms satisfy

$$\langle \boldsymbol{D}_1^{-1}(\varepsilon), \boldsymbol{Q}_1\rangle + \varepsilon\langle \boldsymbol{\Lambda}_1, \boldsymbol{E}_1\rangle = \langle \tilde{\boldsymbol{D}}_1^{-1}, \boldsymbol{Q}_1\rangle + \langle \tilde{\boldsymbol{D}}_1^{-1}, \varepsilon\boldsymbol{E}_1\rangle = \langle \tilde{\boldsymbol{D}}_1^{-1}, \tilde{\boldsymbol{Q}}_1\rangle. \tag{5.63}$$

Applying (5.62) and (5.63) to (5.61) and the left over terms implies that (5.59) is bounded below by

$$\langle \tilde{\boldsymbol{Q}}_1, \tilde{\boldsymbol{D}}_1^{-1}\rangle + \sum_{1\le k\le r-1} (x_{k-1} - x_k)\,\mathrm{Sum}(\boldsymbol{\xi}(\tilde{\boldsymbol{Q}}_k)) + x_{r-1}\,\mathrm{Sum}(\boldsymbol{\xi}(\tilde{\boldsymbol{Q}}_r)),$$

which is what we needed to show. □

5.2.5 Summary of the Proof

We first recall a basic continuity result satisfied by the Parisi functional.

Proposition 5.2.1 (*Uniform Continuity of $\mathscr{P}$ with respect to Temperature*)

Let $\mathscr{P}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}})$ and $\mathscr{P}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}})$ denote the Parisi functional (5.12) with respect to $\vec{\boldsymbol{\beta}}_1$ and $\vec{\boldsymbol{\beta}}_2$. If

$$\sum_{p\geq 2} \|\vec{\beta}_p^1 \otimes \vec{\beta}_p^1 - \vec{\beta}_p^2 \otimes \vec{\beta}_p^2\|_1 \leq \delta, \tag{5.64}$$

then

$$\Big| \inf_{r,\Lambda,x,Q} \mathscr{P}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) - \inf_{r,\Lambda,x,Q} \mathscr{P}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) \Big| \leq \delta.$$

Proof. Recall that the Parisi functional is a the limit of the free energy,

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E} \log \int_{\boldsymbol{Q}} \exp H_N^{\vec{\boldsymbol{\beta}}}(\sigma)\, d\lambda_N = \inf_{r,\Lambda,x,Q} \mathscr{P}_{\vec{\boldsymbol{\beta}}}(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}).$$

We can use Gaussian interpolation to prove uniform continuity. Consider the Hamiltonian,

$$H_t(\sigma) = \sqrt{t} H_N^{\vec{\boldsymbol{\beta}}_1}(\sigma) + \sqrt{1-t} H_N^{\vec{\boldsymbol{\beta}}_2}(\sigma),$$

and the interpolating free energy,

$$\varphi(t) = \frac{1}{N} \mathbb{E} \log \int_{\boldsymbol{Q}} \exp H_t(\sigma)\, d\lambda_N.$$

Differentiating with respect to t and integrating by parts, we see that

$$\varphi'(t) = \frac{1}{N} \Big\langle \frac{d}{dt} H_t(\sigma) \Big\rangle_t = \frac{1}{2} \Big\langle \operatorname{Sum}(\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{R}_{1,1}) - \boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{R}_{1,1})) - \operatorname{Sum}(\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{R}_{1,2}) - \boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{R}_{1,2}))) \Big\rangle_t$$

where $\langle \cdot \rangle_t$ is the Gibbs average proportional to $e^{H_t(\sigma)}$. Since $\|\boldsymbol{R}\|_\infty \leq 1$, the assumption (5.64) implies

$$\Big| \operatorname{Sum}\left(\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{R}) - \boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{R})\right) \Big| \leq \sum_{p\geq 2} \sum_{i,j\leq n} \Big| (\vec{\beta}_p^1(i)\vec{\beta}_p^1(j) - \vec{\beta}_p^2(i)\vec{\beta}_p^2(j)) \Big| \leq \delta.$$

Therefore, $|\varphi'(t)| \leq \delta$, so

$$\Big| \frac{1}{N} \mathbb{E} \log \int_{\boldsymbol{Q}} \exp H_N^{\vec{\boldsymbol{\beta}}_1}(\sigma)\, d\lambda_N - \frac{1}{N} \mathbb{E} \log \int_{\boldsymbol{Q}} \exp H_N^{\vec{\boldsymbol{\beta}}_2}(\sigma)\, d\lambda_N \Big| \leq \delta,$$

and taking limits implies

$$\Big| \inf_{r,\Lambda,x,Q} \mathscr{P}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) - \inf_{r,\Lambda,x,Q} \mathscr{P}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) \Big| \leq \delta.$$

□

The following result will allow us to take $\varepsilon \to 0$ to remove the effect of the perturbation to recover a statement about $\mathscr{P}$.

Proposition 5.2.2

Recall the functions $\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$ defined in (5.24) and $\mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}})$ defined in (5.66). We have

$$\lim_{\varepsilon\to 0}\inf_{\Lambda,Q} \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \inf_{\Lambda,Q}\lim_{\varepsilon\to 0} \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \inf_{\Lambda,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$$

and

$$\lim_{\varepsilon\to 0}\inf_{Q} \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}}) = \inf_{Q}\lim_{\varepsilon\to 0} \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}}) = \inf_{Q} \mathscr{C}_r(\underline{\boldsymbol{Q}}).$$

Proof. Since $\mathscr{P}_r^\varepsilon = \mathscr{P} + \varepsilon\mathscr{B}_r$ and $\mathscr{B}_r \geq 0$, $\mathscr{P}_r^\varepsilon$ is decreasing in ε. The functional $\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$ is only defined for strictly monotone sequences $\underline{\boldsymbol{Q}}$. If we restrict $\mathscr{P}_r$ to strictly monotone sequences, and take the infimums only over $\underline{\boldsymbol{Q}}$ with strictly increasing increments then

$$\lim_{\varepsilon\to 0}\inf_{\Lambda,Q} \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \inf_{\varepsilon,\Lambda,Q} \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \inf_{\Lambda,Q}\lim_{\varepsilon\to 0} \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \inf_{\Lambda,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$$

since $\inf_{\Lambda,Q} \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$ is also decreasing in ε. Furthermore, since $\mathscr{P}_r$ is continuous and the space of $\underline{\boldsymbol{Q}}$ with increasing increments is the closure of the paths with strictly increasing elements we can take the infimum over all paths of the form (5.9) without changing the value of $\inf_{\Lambda,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$.

The proof for $\mathscr{C}_r^\varepsilon$ is identical. □

We now summarize the proof of the lower bound.

Proof of Lemma 5.2.1. Assuming that $\vec{\beta}_2 > 0$, for $\varepsilon > 0$ and fixed sequence (5.23), the minimizer $\boldsymbol{\Lambda}^\varepsilon$, $\underline{\boldsymbol{Q}}^\varepsilon$ of $\mathscr{P}_r^\varepsilon$ satisfies the critical point conditions (5.37) by Lemma 5.2.2. From Lemma 5.2.3 and Lemma 5.2.4, these critical point conditions results in the following chain of inequalities,

$$\inf_{\Lambda,Q} \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}^\varepsilon, \underline{\boldsymbol{Q}}^\varepsilon) = \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}}^\varepsilon) \geq \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}).$$

Since $\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$ is decreasing in ε for fixed $\boldsymbol{\Lambda}$ and $\underline{\boldsymbol{Q}}$ and $\mathscr{P}_r(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$ is continuous, we can interchange the limit with the infimum [Proposition 5.2.2], so

$$\lim_{\varepsilon\to 0}\inf_{\Lambda,Q} \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \inf_{\Lambda,Q}\lim_{\varepsilon\to 0} \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \inf_{\Lambda,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) \geq \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}).$$

The lower bound does not depends on r nor the sequence (5.23), so we can take the infimum of $\mathscr{P}_r$ over all sequences of the form (5.23) and all discretizations to finish the proof of the lower bound.

This proves the case of the lower bound under the additional assumption that $\beta_2 > 0$. To conclude the general case, suppose that $(\vec{\beta}_p)_{p\geq 2}$ is a sequence of positive inverse temperature parameters such that $\beta_p = 0$ if p is odd. We can modify the temperature by adding a small positive perturbation to the second term, $(\vec{\beta}_p^\delta)_{p\geq 2} = (\vec{\beta}_2 + \delta\vec{1}, \vec{\beta}_4, \dots)$. Consider $\mathscr{P}_r^\delta$ and $\mathscr{C}_r^\delta$ defined with respect to $(\vec{\beta}_p^\delta)_{p\geq 2}$. We have

$$\inf_{r,\Lambda,x,Q} \mathscr{P}_r^\delta(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) \geq \inf_{r,x,Q} \mathscr{C}_r^\delta(\underline{x}, \underline{\boldsymbol{Q}}).$$

This holds for all $\delta > 0$, so we can use the fact that both $\inf \mathscr{P}_r^\delta$ and $\inf \mathscr{C}_r^\delta$ are uniformly continuous

functions of the temperature [Proposition 5.3.1] and send $\delta \to 0$ to conclude

$$\inf_{r,\Lambda,x,Q} \mathscr{P}_r(\mathbf{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) \geq \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}).$$

□

Remark 5.2.1. The exact formula for the error terms $\boldsymbol{E}_k$ was not needed in our computations. We are free to choose any barrier $\mathscr{B}_r$ that assigns infinitely large penalties to degenerate increments to prove the lower bound. The logarithmic barrier was chosen because its derivatives are easy to compute explicitly.

5.3 The Upper Bound of the Parisi Functional

We now use a similar procedure to prove the matching upper bound. To simplify notation, several terms such as $\mathscr{P}_r^\varepsilon$, $\mathscr{C}_r^\varepsilon$, and $\overline{\boldsymbol{E}}$ that appeared Section 5.2 will be redefined in this section. In this section, we will prove that the infimum of the Parisi Functional is a lower bound of the Crisanti–Sommers functional:

Lemma 5.3.1

For any positive definite constraint $\boldsymbol{Q}$, we have

$$\inf_{r,\Lambda,x,Q} \mathscr{P}_r(\mathbf{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}) \leq \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}),$$

where the first infimum is over sequences (5.9) and $\mathbf{\Lambda} \in \mathbb{S}_+^n$ such such that $|\mathbf{\Lambda}_1| > 0$ and the second infimum is over sequences (5.9) such that $|\boldsymbol{D}_{r-1}| > 0$.

Like the lower bound, we prove this by examining the behavior of $\mathscr{C}_r$ at its critical points. We will perturb $\mathscr{C}_r$ by adding a logarithmic penalty at the boundary to force the minimizer of $\mathscr{C}_r$ to have positive increments. The minimizers will satisfy an interior critical point condition that will allow us to reduce the perturbed functional $\mathscr{C}_r^\varepsilon$ into a perturbed $\mathscr{P}_r^\varepsilon$ functional. These perturbed functionals will converge to $\mathscr{P}_r$ and $\mathscr{C}_r$ in the limit as the size of the barrier tends to 0. The main difference is the convexity argument used in the proof of Lemma 5.2.4 does not work in this direction. Instead, we use a concavity argument to absorb the error terms into the Lagrange multiplier term.

5.3.1 Adding a Positive Definite Barrier

We fix $r \geq 2$ and let $\underline{\boldsymbol{Q}} = (\boldsymbol{Q}_k)_{k=0}^r$ denote the monotone sequence of matrices. We will add a logarithmic barrier to $\mathscr{C}_r$ that introduces a large penalty when $\underline{\boldsymbol{Q}}$ is not strictly increasing. Let $\varepsilon > 0$ and consider the barrier term

$$\mathscr{B}(\underline{\boldsymbol{Q}}) := -\varepsilon \sum_{0 \leq k \leq r-1} \log |\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k|.$$

Since $|\boldsymbol{A}| \leq (\frac{\operatorname{tr}(A)}{n})^n$ [Proposition 3.3.12] we have $|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| \leq 1$ so $\mathscr{B}_r \geq 0$. Furthermore, $\mathscr{B} \to +\infty$ if $|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| \to 0$ for some $0 \leq k \leq r-1$.

For a fixed strictly increasing path such that

$$0 = x_0 < x_1 < \cdots < x_{r-1} = 1, \tag{5.65}$$

we define the functional,

$$\mathscr{C}_r^{\varepsilon}(\underline{\boldsymbol{Q}}) = \frac{1}{2}\Big[\log|\boldsymbol{Q} - \boldsymbol{Q}_{r-1}| + \langle \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{D}_1\rangle - \sum_{1\le k\le r-2} \frac{1}{x_k}\log\frac{|\boldsymbol{D}_{k+1}|}{|\boldsymbol{D}_k|} + \langle \boldsymbol{D}_1^{-1}, \boldsymbol{Q}_1\rangle \\ + \sum_{1\le k\le r-1} x_k \cdot \mathrm{Sum}\big(\boldsymbol{\xi}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}(\boldsymbol{Q}_k)\big) - \varepsilon \sum_{0\le k\le r-1} \log|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k|\Big]. \tag{5.66}$$

Notice that $\mathscr{C}_r^{\varepsilon} = \mathscr{C}_r + \varepsilon\mathscr{B}_r$ decreases pointwise to $\mathscr{C}(\underline{x}, \underline{\boldsymbol{Q}})$ as $\varepsilon \to 0$, where $\underline{x}$ is the monotone sequence (5.65). The barrier term forces the minimizer to lie in the interior of the positive definite cone, since $\mathscr{C}_r^{\varepsilon}(\underline{\boldsymbol{Q}}) \to +\infty$ if one of the increments $|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| \to 0$. We now examine the behavior of $\mathscr{C}_r^{\varepsilon}(\underline{\boldsymbol{Q}})$ at its minimizers and recover a system of critical point equations.

5.3.2 Critical Point Conditions

We will study the first variation of $\mathscr{C}_r^{\varepsilon}$ to recover critical point conditions for its minimizer. We want to minimize the function

$$\mathscr{C}_r^{\varepsilon}(\underline{\boldsymbol{Q}}) := \mathscr{C}_r^{\varepsilon}(\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_{r-1})$$

over the parameters

$$(\boldsymbol{Q}_k)_{k=1}^{r-1} \in \mathcal{Q}_r := \big\{\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_{r-1} \in \mathbb{S}_+^n \mid |\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| > 0,\ \forall\, 0 \le k \le r-1\big\}.$$

By compactness, $\mathscr{C}_r^{\varepsilon}$ attains its minimum at some $(\boldsymbol{Q}_k)_{k=1}^{r-1} \in \mathcal{Q}_r$. Since $\mathscr{B}_r(\underline{\boldsymbol{Q}}) = \infty$ if the increments are not positive definite, $(\boldsymbol{Q}_k)_{k=1}^{r-1}$ must have positive definite increments,

$$|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| > 0 \quad \forall\, 0 \le k \le r-1.$$

This implies that symmetric matrices are admissible variations of $(\boldsymbol{Q}_k)_{k=1}^{r-1}$ [Proposition 3.3.13]. In particular, if $\boldsymbol{C}$ is a symmetric matrix, then for all t sufficiently small,

$$(\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_p + t\boldsymbol{C}, \dots \boldsymbol{Q}_{r-1}) \in \mathcal{Q}_r \text{ for } 1 \le p \le r-1.$$

If $(\boldsymbol{Q}_k)_{k=1}^{r-1} \in \mathcal{Q}_r$ is a minimizer of $\mathscr{C}_r^{\varepsilon}$, then for all $1 \le p \le r-1$,

$$\frac{d}{dt}\mathscr{C}_r^{\varepsilon}(\boldsymbol{\Lambda}, \boldsymbol{Q}_1, \dots, \boldsymbol{Q}_p + t\boldsymbol{C}, \dots \boldsymbol{Q}_{r-1})\Big|_{t=0} = 0.$$

We can compute the first variation of the functionals explicitly by computing the matrix derivatives of $\mathscr{C}_r^{\varepsilon}$ and derive some critical point conditions on the minimizers.

To simplify the computations, we use summation by parts to write $\mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}})$ as

$$\mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}) = \frac{1}{2}\Big[\mathrm{tr}(\vec{h}\vec{h}^{\mathsf{T}}\boldsymbol{D}_1) + \mathrm{tr}(\boldsymbol{Q}_1\boldsymbol{D}_1^{-1}) + \frac{1}{x_1}\log|\boldsymbol{D}_1| + \sum_{2\le k\le r-1} \log|\boldsymbol{D}_k|\Big(\frac{1}{x_k} - \frac{1}{x_{k-1}}\Big) \\ + \sum_{1\le k\le r-1} x_k \cdot \mathrm{tr}\Big(\mathbf{1} \times \big(\boldsymbol{\xi}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}(\boldsymbol{Q}_k)\big)\Big)\Big]. \tag{5.67}$$

Since $(\boldsymbol{D}_k)_{k=1}^{r}$ is a function of $(\boldsymbol{Q}_k)_{k=1}^{r}$, for $1 \le p \le r-1$ we have

(1) If $p < \ell$

$$\frac{d\boldsymbol{D}_\ell}{d\boldsymbol{Q}_p} = 0$$

(2) If $p > \ell$

$$\frac{d\boldsymbol{D}_\ell}{d\boldsymbol{Q}_p} = (x_{p-1} - x_p)\boldsymbol{C}$$

(3) If $p = \ell$

$$\frac{d\boldsymbol{D}_\ell}{d\boldsymbol{Q}_p} = -x_p\boldsymbol{C}.$$

Using the formulas in Section 3.3 and the chain rule on (5.67), the derivatives in direction $2\boldsymbol{C}$ (the constant 2 is to cancel the constant factor of $\frac{1}{2}$ in front of $\mathscr{C}_r$) for $2 \le p \le r-1$ are given by

$$\begin{aligned}
\partial_{\boldsymbol{Q}_p}\mathscr{C}_r &= (x_{p-1} - x_p)\operatorname{tr}(\vec{h}\vec{h}^T\boldsymbol{C}) - (x_{p-1} - x_p)\operatorname{tr}(\boldsymbol{D}_1^{-1}\boldsymbol{Q}_1\boldsymbol{D}_1^{-1}\boldsymbol{C}) + \frac{1}{x_1}(x_{p-1} - x_p)\operatorname{tr}(\boldsymbol{D}_1^{-1}\boldsymbol{C}) \\
&\quad + \sum_{2\le k<p}(x_{p-1} - x_p)\Big(\frac{1}{x_k} - \frac{1}{x_{k-1}}\Big)\operatorname{tr}(\boldsymbol{D}_k^{-1}\boldsymbol{C}) - x_p\Big(\frac{1}{x_p} - \frac{1}{x_{p-1}}\Big)\operatorname{tr}(\boldsymbol{D}_p^{-1}\boldsymbol{C}) \\
&\quad + (x_{p-1} - x_p)\operatorname{tr}(\boldsymbol{\xi}'(\boldsymbol{Q}_p)\boldsymbol{C}) \\
&= (x_{p-1} - x_p)\operatorname{tr}(\vec{h}\vec{h}^T\boldsymbol{C}) - (x_{p-1} - x_p)\operatorname{tr}(\boldsymbol{D}_1^{-1}\boldsymbol{Q}_1\boldsymbol{D}_1^{-1}\boldsymbol{C}) \\
&\quad - (x_{p-1} - x_p)\sum_{1\le k\le p-1}\frac{1}{x_k}\operatorname{tr}((\boldsymbol{D}_{k+1}^{-1} - \boldsymbol{D}_k^{-1})\boldsymbol{C}) + (x_{p-1} - x_p)\operatorname{tr}(\boldsymbol{\xi}'(\boldsymbol{Q}_p)\boldsymbol{C}),
\end{aligned} \tag{5.68}$$

and the derivative for $p = 1$ is given by

$$\partial_{\boldsymbol{Q}_1}\mathscr{C}_r = -x_1\operatorname{tr}(\vec{h}\vec{h}^T\boldsymbol{C}) + x_1\operatorname{tr}(\boldsymbol{D}_1^{-1}\boldsymbol{Q}_1\boldsymbol{D}_1^{-1}\boldsymbol{C}) - x_1\operatorname{tr}(\boldsymbol{\xi}'(\boldsymbol{Q}_1)\boldsymbol{C}). \tag{5.69}$$

Derivation of Critical Point Equations

From the formulas for the partial derivatives, we can derive a system of implicit equations satisfied by the minimizers of $\mathscr{C}_r^\varepsilon$:

(a) For $2 \le p \le r-1$, the directional derivatives of $\mathscr{C}_r^\varepsilon$ with respect to $\boldsymbol{Q}_p$ in the symmetric direction $2\boldsymbol{C}$ is [Equation 5.68]

$$\begin{aligned}
\partial_{\boldsymbol{Q}_p}\mathscr{C}_r^\varepsilon &= (x_{p-1} - x_p)\langle\vec{h}\vec{h}^T, \boldsymbol{C}\rangle - (x_{p-1} - x_p)\langle\boldsymbol{D}_1^{-1}\boldsymbol{Q}_1\boldsymbol{D}_1^{-1}, \boldsymbol{C}\rangle \\
&\quad - (x_{p-1} - x_p)\sum_{1\le k\le p-1}\frac{1}{x_k}\langle\boldsymbol{D}_{k+1}^{-1} - \boldsymbol{D}_k^{-1}, \boldsymbol{C}\rangle + (x_{p-1} - x_p)\langle\boldsymbol{\xi}'(\boldsymbol{Q}_p), \boldsymbol{C}\rangle \\
&\quad + \varepsilon\langle(\boldsymbol{Q}_{p+1} - \boldsymbol{Q}_p)^{-1}, \boldsymbol{C}\rangle - \varepsilon\langle(\boldsymbol{Q}_p - \boldsymbol{Q}_{p-1})^{-1}, \boldsymbol{C}\rangle.
\end{aligned} \tag{5.70}$$

At the minimizer, we require

$$\partial_{\boldsymbol{Q}_p}\mathscr{C}_r^\varepsilon := \frac{d}{dt}\mathscr{C}_r^\varepsilon(\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_p + 2t\boldsymbol{C}, \dots \boldsymbol{Q}_{r-1})\Big|_{t=0} = 0.$$

This equality holds for all symmetric directions $\boldsymbol{C}$, so the minimizer satisfies the critical point equation

$$\boldsymbol{\xi}'(\boldsymbol{Q}_p) = -\vec{h}\vec{h}^T + \boldsymbol{D}_1^{-1}\boldsymbol{Q}_1\boldsymbol{D}_1^{-1} + \sum_{1\le k\le p-1}\frac{1}{x_k}(\boldsymbol{D}_{k+1}^{-1} - \boldsymbol{D}_k^{-1}) + \varepsilon\boldsymbol{E}_p, \tag{5.71}$$

where

$$\boldsymbol{E}_p := \frac{1}{x_p - x_{p-1}}\Big((\boldsymbol{Q}_{p+1} - \boldsymbol{Q}_p)^{-1} - (\boldsymbol{Q}_p - \boldsymbol{Q}_{p-1})^{-1}\Big). \tag{5.72}$$

$\boldsymbol{E}_p$ is well defined because we fixed a strictly increasing sequence $(x_p)_{p=1}^r$ in (5.65).

(b) For $p = 1$, the directional derivatives of $\mathscr{C}_r^\varepsilon$ with respect to $\boldsymbol{Q}_1$ in the symmetric direction $2\boldsymbol{C}$ is [Equation 5.69]

$$\begin{aligned}\partial_{\boldsymbol{Q}_1}\mathscr{C}_r^\varepsilon = &-x_1\langle \vec{h}\vec{h}^T, \boldsymbol{C}\rangle + x_1\langle \boldsymbol{D}_1^{-1}\boldsymbol{Q}_1\boldsymbol{D}_1^{-1}, \boldsymbol{C}\rangle - x_1\langle \boldsymbol{\xi}'(\boldsymbol{Q}_1), \boldsymbol{C}\rangle \\ &+ \varepsilon\langle (\boldsymbol{Q}_2 - \boldsymbol{Q}_1)^{-1}, \boldsymbol{C}\rangle - \varepsilon\langle \boldsymbol{Q}_1^{-1}, \boldsymbol{C}\rangle.\end{aligned} \tag{5.73}$$

At the minimizer, we require

$$\partial_{\boldsymbol{Q}_1}\mathscr{C}_r^\varepsilon := \frac{d}{dt}\mathscr{C}_r^\varepsilon(\boldsymbol{Q}_1 + 2t\boldsymbol{C}, \dots, \boldsymbol{Q}_{r-1})\Big|_{t=0} = 0.$$

This equality holds for all symmetric directions $\boldsymbol{C}$, so the minimizer satisfies the critical point equation

$$\boldsymbol{\xi}'(\boldsymbol{Q}_1) = -\vec{h}\vec{h}^T + \boldsymbol{D}_1^{-1}\boldsymbol{Q}_1\boldsymbol{D}_1^{-1} + \varepsilon\boldsymbol{E}_1, \tag{5.74}$$

where $\boldsymbol{E}_1$ is given by the formula in (5.72) with $p = 1$.

For $1 \le p \le r-1$, the critical point equations (5.71) and (5.74) can be expressed as

$$\boldsymbol{\xi}'(\boldsymbol{Q}_p) = -\vec{h}\vec{h}^T + \boldsymbol{D}_1^{-1}\boldsymbol{Q}_1\boldsymbol{D}_1^{-1} + \sum_{1\le k\le p-1}\frac{1}{x_k}(\boldsymbol{D}_{k+1}^{-1} - \boldsymbol{D}_k^{-1}) + \varepsilon\boldsymbol{E}_p$$

where $\boldsymbol{E}_r := 0$ and $\boldsymbol{E}_p$ was defined in (5.72). By subtracting these equations, we can conclude for $1 \le p \le r-2$,

$$x_p(\boldsymbol{\xi}'(\boldsymbol{Q}_{p+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_p)) = \boldsymbol{D}_{p+1}^{-1} - \boldsymbol{D}_p^{-1} + \varepsilon x_p(\boldsymbol{E}_{p+1} - \boldsymbol{E}_p). \tag{5.75}$$

Consider $\boldsymbol{\Lambda}$ given by $\boldsymbol{\Lambda} := (\boldsymbol{Q} - \boldsymbol{Q}_{r-1})^{-1} + \boldsymbol{\xi}'(\boldsymbol{Q}) - \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}) - \varepsilon(\boldsymbol{E}_r - \boldsymbol{E}_{r-1})$. For this choice of $\boldsymbol{\Lambda}$, we have

$$\boldsymbol{\Lambda}_{r-1}(\varepsilon) = \boldsymbol{\Lambda} - (\boldsymbol{\xi}'(\boldsymbol{Q}) - \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})) + \varepsilon(\boldsymbol{E}_r - \boldsymbol{E}_{r-1}) = \boldsymbol{D}_{r-1}^{-1}.$$

Subtracting (5.75) from $\boldsymbol{\Lambda}_{r-1}(\varepsilon)$, we conclude that

$$\boldsymbol{\Lambda}_p + \sum_{p\le k\le r-1} x_k(\boldsymbol{E}_{k+1} - \boldsymbol{E}_k) = \boldsymbol{D}_p^{-1} \text{ for } 1 \le p \le r-1. \tag{5.76}$$

The critical point conditions implicitly implies that $\boldsymbol{\Lambda}_1(\varepsilon) > 0$ and $\boldsymbol{D}_k^{-1} = \boldsymbol{\Lambda}_k(\varepsilon) < \boldsymbol{\Lambda}_{k+1}(\varepsilon) = \boldsymbol{D}_{k+1}^{-1}$. We summarize the critical point condition in the following lemma.

Lemma 5.3.2

For fixed $r \geq 2$, if $\underline{\boldsymbol{Q}}$ is a minimizer of $\mathscr{C}_r^\varepsilon$ and

$$\boldsymbol{\Lambda} := (\boldsymbol{Q} - \boldsymbol{Q}_{r-1})^{-1} + \boldsymbol{\xi}'(\boldsymbol{Q}) - \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}) - \varepsilon(\boldsymbol{E}_r - \boldsymbol{E}_{r-1}) \tag{5.77}$$

then $\underline{\boldsymbol{Q}}$ satisfies the following critical point equations

$$\boldsymbol{D}_p^{-1} = \boldsymbol{\Lambda}_p(\varepsilon) \qquad \text{for } 1 \leq p \leq r-1 \tag{5.78}$$

where

$$\boldsymbol{\Lambda}_p(\varepsilon) = \boldsymbol{\Lambda}_p + \varepsilon \overline{\boldsymbol{E}}_p \qquad \text{and} \qquad \overline{\boldsymbol{E}}_p = \sum_{p \leq k \leq r-1} x_k(\boldsymbol{E}_{k+1} - \boldsymbol{E}_k). \tag{5.79}$$

5.3.3 Reduction to an approximate Parisi functional

In this subsection, we will reduce $\mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}})$ defined in (5.66) to an approximate Parisi functional. If $\boldsymbol{\Lambda}$ equals (5.77) and $\underline{\boldsymbol{Q}}$ satisfies the critical point conditions (5.78), then $\mathscr{C}_r^\varepsilon$ can be reduced to

$$\begin{aligned}
\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) = \frac{1}{2}\Big[&\langle \boldsymbol{\Lambda}, \boldsymbol{Q} \rangle - n - \log|\boldsymbol{\Lambda}| + \sum_{1 \leq k \leq r-1} \frac{1}{x_k} \log \frac{|\boldsymbol{\Lambda}_{k+1}(\varepsilon)|}{|\boldsymbol{\Lambda}_k(\varepsilon)|} \\
&+ \langle \boldsymbol{\Lambda}_1^{-1}(\varepsilon), \vec{h}\vec{h}^\intercal + \boldsymbol{\xi}'(\boldsymbol{Q}_1) \rangle - \sum_{1 \leq k \leq r-1} x_k \cdot \mathrm{Sum}\big(\boldsymbol{\theta}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\theta}(\boldsymbol{Q}_k)\big) \\
&- \varepsilon \sum_{1 \leq k \leq r-2} \langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{Q}_k \rangle + \sum_{1 \leq k \leq r-2} \frac{\varepsilon}{x_k} \langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1} \rangle \\
&+ \varepsilon \langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1} \rangle + \varepsilon \langle \boldsymbol{Q}_{r-1}, \overline{\boldsymbol{E}}_{r-1} \rangle - \varepsilon \sum_{0 \leq k \leq r-1} \log|\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k| \Big].
\end{aligned} \tag{5.80}$$

Notice that $\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$ is of the same form as $\mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}})$, but with $\boldsymbol{\Lambda}_k$ replaced by $\boldsymbol{\Lambda}_k(\varepsilon)$ and some additional error terms. If we set $\varepsilon = 0$, then the error terms in the second line all vanish and we are left with the usual $\mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}})$ functional. In the next subsection, we will show that we can bound the minimum of $\mathscr{P}_r^\varepsilon$ with $\mathscr{P}_r$ evaluated at a different parameter to remove the error terms.

Lemma 5.3.3

For fixed r, if $\boldsymbol{\Lambda}$ and $(\boldsymbol{Q}_k)_{k=1}^r$ satisfy the critical point conditions (5.77) and (5.78), then

$$\mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}}) = \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}).$$

Proof. The proof is a straightforward but tedious computation. The computation is almost identical to the proof of Lemma 5.2.3. Assuming that (5.77) and (5.78) hold, we will show that

$$2(\mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) - \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}})) = 0.$$

We will use the following identities multiple times throughout the proof,

$$\boldsymbol{\Lambda}_{k+1} - \boldsymbol{\Lambda}_k = x_k(\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_k)) \qquad 1 \le k \le r-1 \tag{5.81}$$

$$\boldsymbol{D}_k - \boldsymbol{D}_{k+1} = x_k(\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k) \qquad 1 \le k \le r-1 \tag{5.82}$$

$$\boldsymbol{D}_k^{-1} = \boldsymbol{\Lambda}_k(\varepsilon) \qquad 1 \le k \le r-1 \tag{5.83}$$

where $\boldsymbol{D}_r := \boldsymbol{0}$. These identities will allow us to simplify $\mathscr{C}_r^\varepsilon$ into $\mathscr{P}_r^\varepsilon$.

We start by observing that the external fields cancel if (5.78) holds,

$$\langle \vec{h}\vec{h}^\intercal, \boldsymbol{\Lambda}_1^{-1}(\varepsilon)\rangle \overset{(5.83)}{=} \langle \vec{h}\vec{h}^\intercal, \boldsymbol{D}_1(\varepsilon)\rangle. \tag{5.84}$$

Next, we simplify the summation fo the logarithm terms in $\mathscr{P}_r^\varepsilon$. Equation (5.78) applied the $r-1$ term implies that the boundary term in the first summation of (5.80) simplifies to

$$\frac{1}{x_{r-1}}(\log|\boldsymbol{\Lambda}_r(\varepsilon)| - \log|\boldsymbol{\Lambda}_{r-1}(\varepsilon)|) \overset{(5.83)}{=} \log|\boldsymbol{\Lambda}| - \log|\boldsymbol{D}_{r-1}^{-1}| = \log|\boldsymbol{\Lambda}| + \frac{1}{x_{r-1}}\log|\boldsymbol{Q} - \boldsymbol{Q}_{r-1}|,$$

since $\boldsymbol{D}_{r-1} = \boldsymbol{Q} - \boldsymbol{Q}_{r-1}$ and $x_{r-1} = 1$. Applying (5.78) again to $\boldsymbol{\Lambda}_{k+1}(\varepsilon)$ for $1 \le k \le r-1$ implies

$$-\log|\boldsymbol{\Lambda}| + \sum_{1\le k\le r-1}\frac{1}{x_k}\log\frac{|\boldsymbol{\Lambda}_{k+1}(\varepsilon)|}{|\boldsymbol{\Lambda}_k(\varepsilon)|} \overset{(5.83)}{=} \frac{1}{x_{r-1}}\log|\boldsymbol{Q} - \boldsymbol{Q}_{r-1}| - \sum_{1\le k\le r-2}\frac{1}{x_k}\log\frac{|\boldsymbol{D}_{k+1}|}{|\boldsymbol{D}_k|},$$

Therefore, the log determinant terms in $\mathscr{P}_r^\varepsilon$ and $\mathscr{C}_r^\varepsilon$ also cancel.

Since $\boldsymbol{\theta}(\boldsymbol{A}) = \boldsymbol{A}\odot\boldsymbol{\xi}'(\boldsymbol{A}) - \boldsymbol{\xi}(\boldsymbol{A})$ and $\mathrm{Sum}(\boldsymbol{A}\odot\boldsymbol{B}) = \langle \boldsymbol{A}, \boldsymbol{B}\rangle$, the remaining terms in $2(\mathscr{P}_r^\varepsilon - \mathscr{C}_r^\varepsilon)$ are

$$-\sum_{1\le k\le r-2} x_k\Big(\langle \boldsymbol{Q}_{k+1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle - \langle \boldsymbol{Q}_k, \boldsymbol{\xi}'(\boldsymbol{Q}_k)\rangle\Big) - \Big(\langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q})\rangle - \langle \boldsymbol{Q}_{r-1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle\Big) \tag{5.85}$$

$$-\varepsilon\sum_{1\le k\le r-2}\langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{Q}_k\rangle + \sum_{1\le k\le r-2}\frac{\varepsilon}{x_k}\operatorname{tr}\langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle \tag{5.86}$$

$$+\langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - n + \langle \boldsymbol{\xi}'(\boldsymbol{Q}_1), \boldsymbol{\Lambda}_1^{-1}(\varepsilon)\rangle + \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{Q}_{r-1}, \overline{\boldsymbol{E}}_{r-1}\rangle - \langle \boldsymbol{Q}_1, \boldsymbol{D}_1^{-1}\rangle. \tag{5.87}$$

We will show that (5.85) will cancel (5.86) and (5.87) at the critical point.

We start by simplifying the first summation term in (5.85) using (5.78),

$$-\sum_{1\le k\le r-2} x_k\Big(\langle \boldsymbol{Q}_{k+1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle - \langle \boldsymbol{Q}_k, \boldsymbol{\xi}'(\boldsymbol{Q}_k)\rangle\Big) \tag{5.88}$$

$$= -\sum_{1\le k\le r-2}\Big(\langle x_k(\boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k), \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle + \langle \boldsymbol{Q}_k, x_k(\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_k))\rangle\Big)$$

$$\overset{(5.81)(5.82)}{=} -\sum_{1\le k\le r-2}\Big(\langle \boldsymbol{D}_k - \boldsymbol{D}_{k+1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1})\rangle + \langle \boldsymbol{Q}_k, \boldsymbol{\Lambda}_{k+1}(\varepsilon) - \boldsymbol{\Lambda}_k(\varepsilon)\rangle\Big) + \varepsilon\sum_{1\le k\le r-2}\langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{Q}_k\rangle. \tag{5.89}$$

Using summation by parts and (5.78), the first summation (5.89) is equal to

$$
\begin{aligned}
&- \sum_{1\le k\le r-2} \Big(\langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_k)\rangle + \langle \boldsymbol{\Lambda}_{k+1}(\varepsilon), \boldsymbol{Q}_k - \boldsymbol{Q}_{k+1}\rangle \Big) \\
&- \langle \boldsymbol{\Lambda}_1^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_1)\rangle + \langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle - \langle \boldsymbol{\Lambda}_{r-1}(\varepsilon), \boldsymbol{Q}_{r-1}\rangle + \langle \boldsymbol{\Lambda}_1(\varepsilon), \boldsymbol{Q}_1\rangle.
\end{aligned} \tag{5.90}
$$

From the critical point condition (5.78), we have

$$
\begin{aligned}
\boldsymbol{Q}_k - \boldsymbol{Q}_{k+1} &\overset{(5.82)}{=} \frac{1}{x_k}(\boldsymbol{D}_{k+1} - \boldsymbol{D}_k) \\
&\overset{(5.83)}{=} \frac{1}{x_k}\boldsymbol{\Lambda}_k^{-1}(\varepsilon)(\boldsymbol{\Lambda}_k(\varepsilon) - \boldsymbol{\Lambda}_{k+1}(\varepsilon))\boldsymbol{\Lambda}_{k+1}^{-1}(\varepsilon) \\
&\overset{(5.81)}{=} \boldsymbol{\Lambda}_k^{-1}(\varepsilon)(\boldsymbol{\xi}'(\boldsymbol{Q}_k) - \boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}))\boldsymbol{\Lambda}_{k+1}^{-1}(\varepsilon) + \frac{\varepsilon}{x_k}\boldsymbol{\Lambda}_k^{-1}(\varepsilon)(\overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1})\boldsymbol{\Lambda}_{k+1}^{-1}(\varepsilon),
\end{aligned}
$$

which combined with the fact $\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{CAB})$ and equation (5.90) implies that

$$
\begin{aligned}
(5.88) =& \varepsilon \sum_{1\le k\le r-2} \langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{Q}_k\rangle - \sum_{1\le k\le r-2} \frac{\varepsilon}{x_k}\langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle \\
&- \langle \boldsymbol{\Lambda}_1^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_1)\rangle + \langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle - \langle \boldsymbol{\Lambda}_{r-1}(\varepsilon), \boldsymbol{Q}_{r-1}\rangle + \langle \boldsymbol{\Lambda}_1(\varepsilon), \boldsymbol{Q}_1\rangle.
\end{aligned} \tag{5.91}
$$

Substituting (5.91) into (5.85) implies that

$$
\begin{aligned}
2(\mathscr{P}_r^\varepsilon - \mathscr{C}_r^\varepsilon) &= \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - n + \langle \boldsymbol{\xi}'(\boldsymbol{Q}_1), \boldsymbol{\Lambda}_1^{-1}(\varepsilon)\rangle + \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{Q}_{r-1}, \overline{\boldsymbol{E}}_{r-1}\rangle - \langle \boldsymbol{Q}_1, \boldsymbol{D}_1^{-1}\rangle \\
&\quad - \langle \boldsymbol{\Lambda}_1^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_1)\rangle + \langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle - \langle \boldsymbol{\Lambda}_{r-1}(\varepsilon), \boldsymbol{Q}_{r-1}\rangle + \langle \boldsymbol{\Lambda}_1(\varepsilon), \boldsymbol{Q}_1\rangle \\
&\quad - \langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q})\rangle + \langle \boldsymbol{Q}_{r-1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle \\
&\overset{(5.83)}{=} \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - n + \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{Q}_{r-1}, \overline{\boldsymbol{E}}_{r-1}\rangle \\
&\quad + \langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle - \langle \boldsymbol{\Lambda}_{r-1}(\varepsilon), \boldsymbol{Q}_{r-1}\rangle - \langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q})\rangle + \langle \boldsymbol{Q}_{r-1}, \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle \\
&= \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - \mathrm{tr}(\boldsymbol{I}) + \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle + \langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle \\
&\quad - \langle \boldsymbol{\Lambda}, \boldsymbol{Q}_{r-1}\rangle + \langle \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{Q}_{r-1}\rangle - \langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q})\rangle.
\end{aligned} \tag{5.92}
$$

since $\boldsymbol{\Lambda}_{r-1}(\varepsilon) = \boldsymbol{\Lambda} - (\boldsymbol{\xi}'(\boldsymbol{Q}) - \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})) + \varepsilon\overline{\boldsymbol{E}}_{r-1}$. We will show that the $\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle$ term will cancel all remaining terms. Using the definition of $\boldsymbol{\Lambda}_{r-1}(\varepsilon)$ defined in (5.13) and (5.79),

$$
\begin{aligned}
\boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon)\boldsymbol{\xi}'(\boldsymbol{Q}_{r-1}) &= \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon)(\boldsymbol{\Lambda}_{r-1}(\varepsilon) - \boldsymbol{\Lambda} + \boldsymbol{\xi}'(\boldsymbol{Q}) - \varepsilon\overline{\boldsymbol{E}}_{r-1}) \\
&= \boldsymbol{I} - \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon)\boldsymbol{\Lambda} + \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon)\boldsymbol{\xi}'(\boldsymbol{Q}) - \varepsilon\boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon)\overline{\boldsymbol{E}}_{r-1}.
\end{aligned}
$$

From (5.78) and the fact $\boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon) = \boldsymbol{D}_{r-1} = \boldsymbol{Q} - \boldsymbol{Q}_{r-1}$, we have

$$
-\boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon)\boldsymbol{\Lambda} + \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon)\boldsymbol{\xi}'(\boldsymbol{Q}) = -(\boldsymbol{Q} - \boldsymbol{Q}_{r-1})\boldsymbol{\Lambda} + (\boldsymbol{Q} - \boldsymbol{Q}_{r-1})\boldsymbol{\xi}'(\boldsymbol{Q}).
$$

Since $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$, taking the trace implies

$$
\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \boldsymbol{\xi}'(\boldsymbol{Q}_{r-1})\rangle = \mathrm{tr}(\boldsymbol{I}) - \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle + \langle \boldsymbol{\Lambda}, \boldsymbol{Q}_{r-1}\rangle + \langle \boldsymbol{Q}, \boldsymbol{\xi}'(\boldsymbol{Q})\rangle - \langle \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{Q}_{r-1}\rangle - \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle. \tag{5.93}
$$

Substituting (5.93) into (5.92) cancels out all remaining terms, so

$$2\big(\mathscr{C}_r^{\varepsilon}(\underline{\boldsymbol{Q}}) - \mathscr{P}_r^{\varepsilon}(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})\big) = 0.$$

□

5.3.4 Removing the Error Terms

Like the case of the upper bound, we will use concavity of the terms of $\mathscr{P}_r^{\varepsilon}(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$ defined in (5.80) to bound the minimizer with $\mathscr{P}_r$ evaluated at a different Lagrange multiplier parameter. To this end, we define

$$\tilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda} + \varepsilon \overline{\boldsymbol{E}}_1$$

and for $1 \le p \le r-1$,

$$\tilde{\boldsymbol{\Lambda}}_p = \tilde{\boldsymbol{\Lambda}} - \sum_{p \le k \le r-1} x_k(\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_k)).$$

We first note that $\tilde{\boldsymbol{\Lambda}} \in \mathscr{L}$. By definition, we have

$$\tilde{\boldsymbol{\Lambda}}_1 := \tilde{\boldsymbol{\Lambda}} - \sum_{1 \le k \le r-1} x_k(\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_k)) = \boldsymbol{\Lambda} - \sum_{1 \le k \le r-1} x_k(\boldsymbol{\xi}'(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}'(\boldsymbol{Q}_k)) + \varepsilon \overline{\boldsymbol{E}}_1 = \boldsymbol{\Lambda}_1(\varepsilon), \quad (5.94)$$

is positive definite at the critical point because (5.77) implies $\boldsymbol{\Lambda}_1(\varepsilon) = \boldsymbol{D}_1^{-1} > 0$. By monotonicity, this implies that $\tilde{\boldsymbol{\Lambda}}_p > 0$ for all $1 \le p \le r-1$.

We will use concavity of the log determinants to prove that the original Parisi functional evaluated at $\tilde{\boldsymbol{\Lambda}}$ is a lower bound of $\mathscr{P}_r^{\varepsilon}(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}})$,

$$\mathscr{P}_r^{\varepsilon}(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) \ge \mathscr{P}_r(\tilde{\boldsymbol{\Lambda}}, \underline{\boldsymbol{Q}}) \quad (5.95)$$

provided that $(\boldsymbol{Q}_k)_{k=1}^r$ satisfies the critical point conditions (5.78). Since both $\tilde{\boldsymbol{\Lambda}}$ and the path $\underline{\boldsymbol{Q}}$ are elements in the sets we minimize over, we get the obvious lower bound,

$$\mathscr{C}_r^{\varepsilon}(\underline{\boldsymbol{Q}}) = \mathscr{P}_r^{\varepsilon}(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) \ge \mathscr{P}_r(\tilde{\boldsymbol{\Lambda}}, \underline{\boldsymbol{Q}}) \ge \inf_{r, \Lambda, x, Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}).$$

The lower bound does not depend on the discretization r, ε, nor the fixed sequence (5.65). In particular, we can minimize $\mathscr{C}_r^{\varepsilon}$ over sequences (5.65), r and ε to prove the required upper bound,

$$\inf_{r, x, Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}) \ge \inf_{r, \Lambda, x, Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}).$$

We now prove the lower bound (5.95).

Lemma 5.3.4

If $\underline{x}$ is equal to (5.65), $\underline{\boldsymbol{Q}}$ satisfies the critical point conditions (5.78) and $\boldsymbol{\Lambda}$ equals (5.77), then

$$\mathscr{P}_r^{\varepsilon}(\boldsymbol{\Lambda}, \underline{\boldsymbol{Q}}) \ge \mathscr{P}_r(\tilde{\boldsymbol{\Lambda}}, \underline{x}, \underline{\boldsymbol{Q}}).$$

Proof. Since $(\boldsymbol{Q}_k)_{k=1}^r$ is unchanged and $\mathscr{B}_r \geq 0$, it remains to show that

$$\begin{aligned}
&\langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - n - \log|\boldsymbol{\Lambda}| + \sum_{1\leq k\leq r-1} \frac{1}{x_k} \log \frac{|\boldsymbol{\Lambda}_{k+1}(\varepsilon)|}{|\boldsymbol{\Lambda}_k(\varepsilon)|} + \langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}_1), \boldsymbol{\Lambda}_1^{-1}(\varepsilon)\rangle \\
&- \varepsilon \sum_{1\leq k\leq r-2} \langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{Q}_k\rangle + \sum_{1\leq k\leq r-2} \frac{\varepsilon}{x_k} \langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle \\
&+ \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{Q}_{r-1}, \overline{\boldsymbol{E}}_{r-1}\rangle \\
&= \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - n + \sum_{2\leq k\leq r-1} \Big(\frac{1}{x_{k-1}} - \frac{1}{x_k}\Big) \log|\boldsymbol{\Lambda}_k(\varepsilon)| - \frac{1}{x_1}\log|\boldsymbol{\Lambda}_1(\varepsilon)| + \langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}_1), \boldsymbol{\Lambda}_1^{-1}(\varepsilon)\rangle \\
&- \varepsilon \sum_{1\leq k\leq r-2} \langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{Q}_k\rangle + \sum_{1\leq k\leq r-2} \frac{\varepsilon}{x_k} \langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle \\
&+ \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle + \varepsilon\langle \boldsymbol{Q}_{r-1}, \overline{\boldsymbol{E}}_{r-1}\rangle
\end{aligned} \tag{5.96}$$

is bounded below by

$$\begin{aligned}
&\langle \tilde{\boldsymbol{\Lambda}}, \boldsymbol{Q}\rangle - n - \log|\tilde{\boldsymbol{\Lambda}}| + \sum_{1\leq k\leq r-1} \frac{1}{x_k} \log \frac{|\tilde{\boldsymbol{\Lambda}}_{k+1}|}{|\tilde{\boldsymbol{\Lambda}}_k|} + \langle \vec{h}\vec{h}^{\mathsf{T}}, \tilde{\boldsymbol{\Lambda}}_1^{-1}\rangle + \langle \boldsymbol{\xi}'(\boldsymbol{Q}_1), \tilde{\boldsymbol{\Lambda}}_1^{-1}\rangle \\
&= \langle \tilde{\boldsymbol{\Lambda}}, \boldsymbol{Q}\rangle - n + \sum_{2\leq k\leq r-1} \Big(\frac{1}{x_{k-1}} - \frac{1}{x_k}\Big) \log|\tilde{\boldsymbol{\Lambda}}_k| - \frac{1}{x_1}\log|\tilde{\boldsymbol{\Lambda}}_1| + \langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}_1), \tilde{\boldsymbol{\Lambda}}_1^{-1}\rangle.
\end{aligned}$$

We will use concavity of the log determinant terms to absorb the error terms in (5.96). We use summation by parts to write the error terms in (5.96) as

$$\begin{aligned}
&\varepsilon\langle \boldsymbol{Q}_{r-1}, \overline{\boldsymbol{E}}_{r-1}\rangle - \varepsilon \sum_{1\leq k\leq r-2} \langle \overline{\boldsymbol{E}}_{k+1} - \overline{\boldsymbol{E}}_k, \boldsymbol{Q}_k\rangle + \sum_{1\leq k\leq r-2} \frac{\varepsilon}{x_k} \langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle + \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle \\
&= \varepsilon \sum_{1\leq k\leq r-2} \langle \boldsymbol{Q}_{k+1} - \boldsymbol{Q}_k, \overline{\boldsymbol{E}}_{k+1}\rangle + \varepsilon\langle \boldsymbol{Q}_1, \overline{\boldsymbol{E}}_1\rangle + \sum_{1\leq k\leq r-2} \frac{\varepsilon}{x_k} \operatorname{tr}\langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle + \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle \\
&\overset{(5.82)}{=} \sum_{1\leq k\leq r-2} \frac{\varepsilon}{x_k} \langle \boldsymbol{D}_k - \boldsymbol{D}_{k+1}, \overline{\boldsymbol{E}}_{k+1}\rangle + \varepsilon\langle \boldsymbol{Q}_1, \overline{\boldsymbol{E}}_1\rangle + \sum_{1\leq k\leq r-2} \frac{\varepsilon}{x_k} \langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle + \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle \\
&\overset{(5.83)}{=} \sum_{1\leq k\leq r-2} \frac{\varepsilon}{x_k} \langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon) - \boldsymbol{\Lambda}_{k+1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{k+1}\rangle + \varepsilon\langle \boldsymbol{Q}_1, \overline{\boldsymbol{E}}_1\rangle + \sum_{1\leq k\leq r-2} \frac{\varepsilon}{x_k} \langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k - \overline{\boldsymbol{E}}_{k+1}\rangle + \varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon), \overline{\boldsymbol{E}}_{r-1}\rangle \\
&= -\varepsilon \sum_{2\leq k\leq r-1} \Big(\frac{1}{x_{k-1}} - \frac{1}{x_k}\Big) \langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon), \overline{\boldsymbol{E}}_k\rangle + \frac{\varepsilon}{x_1}\langle \boldsymbol{\Lambda}_1^{-1}(\varepsilon), \overline{\boldsymbol{E}}_1\rangle + \varepsilon\langle \boldsymbol{Q}_1, \overline{\boldsymbol{E}}_1\rangle.
\end{aligned}$$

Adding and subtracting $\sum_{2\le k\le r-1}\left(\frac{1}{x_{k-1}}-\frac{1}{x_k}\right)\varepsilon\langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon),\overline{\boldsymbol{E}}_1\rangle$ and using the fact that

$$\begin{aligned}
&-\sum_{2\le k\le r-1}\Big(\frac{1}{x_{k-1}}-\frac{1}{x_k}\Big)\varepsilon\langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon),\overline{\boldsymbol{E}}_1\rangle+\frac{\varepsilon}{x_1}\langle \boldsymbol{\Lambda}_1^{-1}(\varepsilon),\overline{\boldsymbol{E}}_1\rangle+\varepsilon\langle \boldsymbol{Q}_1,\overline{\boldsymbol{E}}_1\rangle\\
&=\varepsilon\langle \boldsymbol{\Lambda}_{r-1}^{-1}(\varepsilon),\overline{\boldsymbol{E}}_1\rangle+\sum_{1\le k\le r-2}\frac{\varepsilon}{x_k}\langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon)-\boldsymbol{\Lambda}_{k+1}^{-1}(\varepsilon),\overline{\boldsymbol{E}}_1\rangle+\varepsilon\langle \boldsymbol{Q}_1,\overline{\boldsymbol{E}}_1\rangle\\
&\overset{(5.83)}{=}\varepsilon\langle \boldsymbol{D}_{r-1},\overline{\boldsymbol{E}}_1\rangle+\sum_{1\le k\le r-2}\frac{\varepsilon}{x_k}\langle \boldsymbol{D}_k-\boldsymbol{D}_{k+1},\overline{\boldsymbol{E}}_1\rangle+\varepsilon\langle \boldsymbol{Q}_1,\overline{\boldsymbol{E}}_1\rangle\\
&\overset{(5.82)}{=}\varepsilon\langle \boldsymbol{Q}-\boldsymbol{Q}_{r-1},\overline{\boldsymbol{E}}_1\rangle+\sum_{1\le k\le r-2}\varepsilon\langle \boldsymbol{Q}_{k+1}-\boldsymbol{Q}_k,\overline{\boldsymbol{E}}_1\rangle+\varepsilon\langle \boldsymbol{Q}_1,\overline{\boldsymbol{E}}_1\rangle\\
&=\varepsilon\langle \boldsymbol{Q},\overline{\boldsymbol{E}}_1\rangle
\end{aligned}$$

if the critical point condition (5.78) holds, we see that (5.96) is equal to

$$\begin{aligned}
&\langle \boldsymbol{\Lambda},\boldsymbol{Q}\rangle-n+\sum_{2\le k\le r-1}\Big(\frac{1}{x_{k-1}}-\frac{1}{x_k}\Big)\log|\boldsymbol{\Lambda}_k(\varepsilon)|-\frac{1}{x_1}\log|\boldsymbol{\Lambda}_1(\varepsilon)|+\langle \vec{h}\vec{h}^{\mathsf{T}}+\boldsymbol{\xi}'(\boldsymbol{Q}_1),\boldsymbol{\Lambda}_1^{-1}(\varepsilon)\rangle\\
&\quad+\sum_{2\le k\le r-1}\Big(\frac{1}{x_{k-1}}-\frac{1}{x_k}\Big)\varepsilon\langle \boldsymbol{\Lambda}_k^{-1}(\varepsilon),-\overline{\boldsymbol{E}}_k+\overline{\boldsymbol{E}}_1\rangle+\varepsilon\langle \boldsymbol{Q},\overline{\boldsymbol{E}}_1\rangle.
\end{aligned}\tag{5.97}$$

We can now use concavity of the log determinant terms and the first trace term to absorb the error terms. Since $x_{k-1}<x_k$, the concavity of the log determinant [Proposition 3.3.9] implies

$$\Big(\frac{1}{x_{k-1}}-\frac{1}{x_k}\Big)\log|\boldsymbol{\Lambda}_k(\varepsilon)|+\Big(\frac{1}{x_{k-1}}-\frac{1}{x_k}\Big)\varepsilon\langle \boldsymbol{\Lambda}_k^{-1},-\overline{\boldsymbol{E}}_k+\overline{\boldsymbol{E}}_1\rangle\ge\Big(\frac{1}{x_{k-1}}-\frac{1}{x_k}\Big)\log|\tilde{\boldsymbol{\Lambda}}_k|\tag{5.98}$$

because $\boldsymbol{\Lambda}_k(\varepsilon)-\varepsilon\overline{\boldsymbol{E}}_k+\varepsilon\overline{\boldsymbol{E}}_1=\tilde{\boldsymbol{\Lambda}}_k$. The linearity of the trace implies

$$\langle \boldsymbol{\Lambda},\boldsymbol{Q}\rangle+\varepsilon\langle \boldsymbol{Q},\overline{\boldsymbol{E}}_1\rangle=\langle \boldsymbol{\Lambda}+\varepsilon\overline{\boldsymbol{E}}_1,\boldsymbol{Q}\rangle=\langle \tilde{\boldsymbol{\Lambda}},\boldsymbol{Q}\rangle.\tag{5.99}$$

The inequalities (5.98) and (5.99) and the fact $\boldsymbol{\Lambda}_1(\varepsilon)=\tilde{\boldsymbol{\Lambda}}_1$ shown in (5.94) implies that (5.96) is bounded below by

$$\langle \tilde{\boldsymbol{\Lambda}},\boldsymbol{Q}\rangle-n+\sum_{2\le k\le r-1}\Big(\frac{1}{x_{k-1}}-\frac{1}{x_k}\Big)\log|\tilde{\boldsymbol{\Lambda}}_k|-\frac{1}{x_1}\log|\tilde{\boldsymbol{\Lambda}}_1|+\langle \vec{h}\vec{h}^{\mathsf{T}}+\boldsymbol{\xi}'(\boldsymbol{Q}_1),\tilde{\boldsymbol{\Lambda}}_1^{-1}\rangle,$$

which is what we needed to show. □

5.3.5 Summary of the Proof

We first recall a basic continuity fact about the Crisanti–Sommers functional,

Proposition 5.3.1 (*Uniform Continuity of $\mathscr{C}$ with respect to Temperature*)

Let $\mathscr{C}_{\vec{\boldsymbol{\beta}}_1}(\underline{x}, \underline{\boldsymbol{Q}})$ and $\mathscr{C}_{\vec{\boldsymbol{\beta}}_2}(\underline{x}, \underline{\boldsymbol{Q}})$ denote the Crisanti–Sommers functional (5.15) with respect to $\vec{\boldsymbol{\beta}}_1$ and $\vec{\boldsymbol{\beta}}_2$. If

$$\sum_{p\geq 2} \|\vec{\beta}_p^1 \otimes \vec{\beta}_p^1 - \vec{\beta}_p^2 \otimes \vec{\beta}_p^2\|_1 \leq \delta, \tag{5.100}$$

then for any $\vec{x}, \underline{\boldsymbol{Q}}$,

$$|\mathscr{C}_{\vec{\boldsymbol{\beta}}_1}(\underline{x}, \underline{\boldsymbol{Q}}) - \mathscr{C}_{\vec{\boldsymbol{\beta}}_2}(\underline{x}, \underline{\boldsymbol{Q}})| \leq 2\delta.$$

Proof. Since $\mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}})$ only depends on temperature through $\boldsymbol{\xi}$, we only have to find a bound for

$$\frac{1}{2}\Big| \sum_{1\leq k\leq r-1} x_k \cdot \mathrm{Sum}\left(\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{Q}_k)\right) - \sum_{1\leq k\leq r-1} x_k \cdot \mathrm{Sum}\left(\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{Q}_k)\right)\Big|, \tag{5.101}$$

where

$$\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{Q}) = \sum_{p\geq 2} (\vec{\beta}_p^1 \otimes \vec{\beta}_p^1) \odot \boldsymbol{Q}^{\odot p} \qquad \text{and} \qquad \boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{Q}) = \sum_{p\geq 2} (\vec{\beta}_p^2 \otimes \vec{\beta}_p^2) \odot \boldsymbol{Q}^{\odot p}.$$

Since $(\boldsymbol{Q}_p)_{ij} \leq 1$ for any matrix $\mathbf{0} \leq \boldsymbol{Q}_p \leq \boldsymbol{Q}$, the assumption (5.100) implies

$$\left|\mathrm{Sum}\left(\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{Q}_p) - \boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{Q}_p)\right)\right| \leq \sum_{p\geq 2} \sum_{i,j\leq n} \left|(\vec{\beta}_p^1(i)\vec{\beta}_p^1(j) - \vec{\beta}_p^2(i)\vec{\beta}_p^2(j))\right| \leq \delta.$$

Using summation by parts, we see that

$$\sum_{1\leq k\leq r-1} x_k \cdot \mathrm{Sum}\left(\boldsymbol{\xi}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}(\boldsymbol{Q}_k)\right) = -\sum_{1\leq k\leq r-1} (x_k - x_{k-1}) \,\mathrm{Sum}\left(\boldsymbol{\xi}(\boldsymbol{Q}_k)\right) + x_{r-1}\,\mathrm{Sum}(\boldsymbol{\xi}(\boldsymbol{Q}_r)),$$

so (5.101) is bounded by

$$\sum_{1\leq k\leq r-1} (x_k - x_{k-1})\left|\mathrm{Sum}\left(\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{Q}_k)\right) - \mathrm{Sum}\left(\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{Q}_k)\right)\right| + x_{r-1}|\,\mathrm{Sum}(\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_1}(\boldsymbol{Q}_r)) - \mathrm{Sum}(\boldsymbol{\xi}_{\vec{\boldsymbol{\beta}}_2}(\boldsymbol{Q}_r))| \leq 2\delta.$$

□

We now summarize the proof of the upper bound.

Proof of Lemma 5.3.1. For $\varepsilon > 0$ and fixed sequence (5.65), if we define $\boldsymbol{\Lambda}^\varepsilon$ to be equal to (5.77), then the minimizer $\underline{\boldsymbol{Q}}^\varepsilon$ of $\mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}})$ satisfies the critical point conditions (5.78) by Lemma 5.3.2. From Lemma 5.3.3 and Lemma 5.3.4, these critical point conditions implies the following chain of inequalities,

$$\inf_Q \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}}) = \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}}^\varepsilon) = \mathscr{P}_r^\varepsilon(\boldsymbol{\Lambda}^\varepsilon, \underline{\boldsymbol{Q}}^\varepsilon) \geq \inf_{r,\Lambda,x,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}).$$

Since $\mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}})$ is decreasing in ε for fixed $\underline{\boldsymbol{Q}}$ and $\mathscr{C}_r(\underline{\boldsymbol{Q}})$ is continuous, we can interchange the limit with the infimum [Proposition 5.2.2], so

$$\lim_{\varepsilon\to 0} \inf_Q \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}}) = \inf_Q \lim_{\varepsilon\to 0} \mathscr{C}_r^\varepsilon(\underline{\boldsymbol{Q}}) = \inf_Q \mathscr{C}_r(\underline{\boldsymbol{Q}}) \geq \inf_{r,\Lambda,x,Q} \mathscr{P}_r(\boldsymbol{\Lambda}, \underline{x}, \underline{\boldsymbol{Q}}).$$

The lower bound does not depends on r nor the sequence (5.65), so we can take the infimum of $\mathscr{C}_r$ over all sequences of the form (5.23) and all discretizations to finish the proof of the upper bound. □

Remark 5.3.1. If we can show that the minimizers of $\mathscr{P}_r$ and $\mathscr{C}_r$ have positive definite increments, then the equality of $\mathscr{P}_r$ and $\mathscr{C}_r$ at its critical points can be proved using the same proof as the one-dimensional case without adding the positive definite barrier.

5.4 Integral Form of the Crisanti–Sommers functional

We will derive the integral form for the analogue of the Crisanti–Sommers formula for spherical spin glasses with vector spins. Recall the monotone functions (5.17) and (5.18),

$$x(t) : [0, n] \to [0, 1] \quad \text{such that} \quad x(0) = 0 \quad \text{and} \quad x(n) = 1$$

and

$$\boldsymbol{\Phi}(t) : [0, n] \to \mathbb{S}^n_+ \quad \text{such that} \quad \operatorname{tr}(\boldsymbol{\Phi}(t)) = t \quad \text{and} \quad \boldsymbol{\Phi}(0) = \mathbf{0} \quad \text{and} \quad \boldsymbol{\Phi}(n) = \boldsymbol{Q}.$$

For $t_x := x^{-1}(1) = \inf\{t \in [0, n] \mid 1 \le x(t)\}$ and paths such that $|\boldsymbol{Q} - \boldsymbol{\Phi}(t_x)| > 0$ the analogue of the Crisanti–Sommers functional (5.19) was defined by

$$\begin{aligned}\mathscr{C}(x, \boldsymbol{\Phi}) &= \frac{1}{2}\bigg(\int_0^n x(t)\langle \boldsymbol{\xi}'(\boldsymbol{\Phi}(t)) + \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{\Phi}'(t)\rangle \, dt + \log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(t_x)| + \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle \, dt \bigg) \\ &= \frac{1}{2}\bigg(\langle \vec{h}\vec{h}^{\mathsf{T}}, \hat{\boldsymbol{\Phi}}(0)\rangle + \int_0^n x(t)\langle \boldsymbol{\xi}'(\boldsymbol{\Phi}(t)), \boldsymbol{\Phi}'(t)\rangle \, dt + \log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(t_x)| + \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle \, dt \bigg),\end{aligned} \tag{5.102}$$

where $\hat{\boldsymbol{\Phi}}(t) : [0, n] \to \mathbb{R}^{n \times n}$ is a decreasing function given by

$$\hat{\boldsymbol{\Phi}}(t) = \int_t^n x(s)\boldsymbol{\Phi}'(s) \, ds. \tag{5.103}$$

This functional is the continuous Lipschitz extension of the discrete functional (5.15) we proved in the last section. We first observe that $\mathscr{C}(x, \boldsymbol{\Phi})$ agrees with the discrete formula when $x(t)$ corresponds to a discrete probability measure on the trace.

Lemma 5.4.1

Let $\boldsymbol{\Phi}(t)$ be a fixed monotone matrix path. Let $x(t)$ be a step function with $r - 1$ steps,

$$x(t) = x_k \quad \text{for} \quad t_k \le t < t_{k+1}$$

for $1 \le k \le r - 1$ with boundary terms

$$x(t) = 0 \quad \text{for} \quad 0 \le t < t_1 \quad \text{and} \quad x(t) = 1 \quad \text{for} \quad t_x := t_{r-1} \le t \le 1.$$

If we define $\boldsymbol{Q}_k := \boldsymbol{\Phi}(t_k)$, then

$$\mathscr{C}(x, \boldsymbol{\Phi}) = \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}).$$

Proof. We first observe for $t_p \le t < t_{p+1}$ that,

$$\begin{aligned}\hat{\boldsymbol{\Phi}}(t) = \int_t^n x(t)\boldsymbol{\Phi}'(t)\,dt &= \sum_{k=p+1}^{r-2} x_k \int_{t_k}^{t_{k+1}} \boldsymbol{\Phi}'(t)\,dt + x_p \int_t^{t_{p+1}} \boldsymbol{\Phi}'(t)\,dt \\ &= \sum_{k=p+1}^{r-2} x_k(\boldsymbol{\Phi}(t_{k+1}) - \boldsymbol{\Phi}(t_k)) + x_p(\boldsymbol{\Phi}(t_{p+1}) - \boldsymbol{\Phi}(t)) \\ &= \boldsymbol{D}_{p+1} + x_p(\boldsymbol{\Phi}(t_{p+1}) - \boldsymbol{\Phi}(t)). \end{aligned} \tag{5.104}$$

We now compute each of the terms in $\mathscr{C}(x, \boldsymbol{\Phi})$ when $x(t)$ is piecewise constant.

(a) The identity (5.104) implies $\hat{\boldsymbol{\Phi}}(0) = \boldsymbol{D}_1$ since $x_0 = 0$, so

$$\langle \vec{h}\vec{h}^{\mathsf{T}}, \hat{\boldsymbol{\Phi}}(0)\rangle = \langle \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{D}_1\rangle.$$

(b) Since $x(t) = x_k$ for $t_k \le t < t_{k+1}$, [Proposition 3.3.3] implies the second term in (5.102) simplifies to

$$\int_0^n x(t)\langle \boldsymbol{\xi}'(\boldsymbol{\Phi}(t)), \boldsymbol{\Phi}'(t)\rangle\,dt = \sum_{k=0}^{r-1} x_k \int_{t_k}^{t_{k+1}} \langle \boldsymbol{\xi}'(\boldsymbol{\Phi}(t)), \boldsymbol{\Phi}'(t)\rangle\,dt = \sum_{1\le k\le r-1} x_k \cdot \mathrm{Sum}\left(\boldsymbol{\xi}(\boldsymbol{Q}_{k+1}) - \boldsymbol{\xi}(\boldsymbol{Q}_k)\right).$$

(c) Since $\boldsymbol{\Phi}(t_x) = \boldsymbol{Q}_{r-1}$ by definition,

$$\log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(t_x)| = \log|\boldsymbol{Q} - \boldsymbol{Q}_{r-1}|.$$

(d) For almost every $t_p < t < t_{p+1}$ the identity (5.104) and Proposition 3.3.4 implies

$$\frac{d}{dt}\left(-\frac{1}{x_p}\log|\hat{\boldsymbol{\Phi}}(t)|\right) = \frac{d}{dt}\left(-\frac{1}{x_p}\log|\boldsymbol{D}_{p+1} + x_p(\boldsymbol{\Phi}(t_{p+1}) - \boldsymbol{\Phi}(t))|\right) = \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle$$

so the fundamental theorem of calculus implies that

$$\int_{t_1}^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle\,dt = \sum_{k=1}^{r-2}\int_{t_k}^{t_{k+1}} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle\,dt = -\sum_{1\le k\le r-2}\frac{1}{x_k}\log\frac{|\boldsymbol{D}_{k+1}|}{|\boldsymbol{D}_k|}$$

and since $x(t) = 0$ for $0 \le t < t_1$, $\hat{\boldsymbol{\Phi}}(t) = \boldsymbol{D}_1$ the boundary term is

$$\int_0^{t_1} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle\,dt = \langle \boldsymbol{D}_1^{-1}, \boldsymbol{\Phi}(t_1)\rangle - \langle \boldsymbol{D}_1^{-1}, \boldsymbol{\Phi}(0)\rangle = \langle \boldsymbol{D}_1^{-1}, \boldsymbol{Q}_1\rangle.$$

Substituting the formulas derived in (a) to (d) into $\mathscr{C}(x, \boldsymbol{\Phi})$ finishes the proof. □

Lemma 5.4.1 implies that $\mathscr{C}(x, \boldsymbol{\Phi})$ evaluated at a piecewise constant c.d.f. corresponds to $\mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}})$ evaluated at some sequence of the form (5.105). To see that every discrete path encoded by $(\underline{x}, \underline{\boldsymbol{Q}})$ corresponds to some $(x, \boldsymbol{\Phi})$, consider the sequences

$$\begin{aligned} x_{-1} = 0 \le\ & x_0 \le\ x_1 \le \ldots \le\ x_{r-2} \le\ x_{r-1} =\ 1 \\ \mathbf{0} = \boldsymbol{Q}_0 \le\ & \boldsymbol{Q}_1 \le \ldots \le \boldsymbol{Q}_{r-2} \le \boldsymbol{Q}_{r-1} < \boldsymbol{Q}_r = \boldsymbol{Q} \end{aligned}. \tag{5.105}$$

Taking $t_k := \mathrm{tr}(\boldsymbol{Q}_k)$ we define a Lipschitz path $\boldsymbol{\Phi}$ by taking $\boldsymbol{\Phi}(t_k) = \boldsymbol{Q}_k$ at each point t_k and interpolate linearly,

$$\boldsymbol{\Phi}(t_k) = \boldsymbol{Q}_k, \qquad \boldsymbol{\Phi}(t) = \frac{t_{k+1} - t}{t_{k+1} - t_k}\boldsymbol{\Phi}(t_k) + \frac{t - t_k}{t_{k+1} - t_k}\boldsymbol{\Phi}(t_{k+1}) \qquad \text{for } t_k \le t < t_{k+1},$$

and a piecewise constant c.d.f. $x(t) = x_k$ for $t_k \le x_k < t_{k+1}$. Applying Lemma 5.4.1 implies that $\mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}})$ evaluated at any sequence of the form (5.105) corresponds to $\mathscr{C}(x, \boldsymbol{\Phi})$ for some $(x, \boldsymbol{\Phi})$. This implies that

$$\inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}) \ge \inf_{x,\boldsymbol{\Phi}} \mathscr{C}(x, \boldsymbol{\Phi}), \tag{5.106}$$

since infimum on the right is over all c.d.f.s and not necessarily piecewise constant ones.

The opposite inequality is a bit trickier to show. We first show that $\mathscr{C}(x, \boldsymbol{\Phi})$ is locally Lipschitz, which will imply that the integral form of the functional is the Lipschitz extension of the functional evaluated on discrete paths. The functional is not well defined when $|\boldsymbol{Q} - \boldsymbol{\Phi}(t_x)| = 0$ because of the log determinant term, so we will show that the functional is Lipschitz if we restrict the domain to matrix paths such that $|\boldsymbol{Q} - \boldsymbol{\Phi}(t_x)|$ is uniformly bounded away from 0.

Let $T \in [0, 1)$ and $L > 0$. Consider the compact set

$$A_{T,L} = \{(x, \boldsymbol{\Phi}) \mid x(t) = 1 \text{ for } t \ge T, \|(\boldsymbol{Q} - \boldsymbol{\Phi}(t_x))^{-1}\|_\infty \le L\}.$$

This set is closed because any convergent sequence $(x_n, \boldsymbol{\Phi}_n)$ must satisfy the uniform bounds,

$$x_n(t) = 1 \text{ for } t \ge T \quad \text{and} \quad \|(\boldsymbol{Q} - \boldsymbol{\Phi}_n(x_n^{-1}(1)))^{-1}\|_\infty \le L,$$

for all n, so its limit point must as well. Furthermore, the product of the space of c.d.f.s on $[0, n]$ equipped with the $\|\cdot\|_1$ norm and the space of Lipschitz paths with fixed endpoints equipped with the $\|\cdot\|_\infty$ norm is compact by Prokhorov's theorem and the Arzelà–Ascoli theorem. Since $A_{T,L}$ is a closed subset of a compact set it is compact. We will show that $\mathscr{C}(x, \boldsymbol{\Phi})$ is Lipschitz on the compact set $A_{T,L}$.

Lemma 5.4.2

Let $(x_1, \boldsymbol{\Phi}_1), (x_2, \boldsymbol{\Phi}_2) \in A_{T,L}$. There exists a constant C_L that only depends on the fixed parameters of the model and the uniform bound L on $\|(\boldsymbol{Q} - \boldsymbol{\Phi}_n(x_n^{-1}(1)))^{-1}\|_\infty$ such that

$$|\mathscr{C}(x_1, \boldsymbol{\Phi}_1) - \mathscr{C}(x_2, \boldsymbol{\Phi}_2)| \le C_L(\|x_1 - x_2\|_1 + \|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\|_\infty),$$

where

$$\|x_1 - x_2\|_1 = \int_0^n |x_1(t) - x_2(t)|\, dt \quad \text{and} \quad \|\boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2\|_\infty = \max_{i,j \le n}\Big(\sup_{t \in [0,n]} |\boldsymbol{\Phi}_1^{i,j}(t) - \boldsymbol{\Phi}_2^{i,j}(t)| \Big).$$

Proof. Without loss of generality, suppose that $x_1^{-1}(1) \le x_2^{-1}(1)$. If this is the case, then $\mathscr{C}(x_1, \boldsymbol{\Phi}_2)$ is also well defined since $\boldsymbol{Q} - \boldsymbol{\Phi}_2(x_1^{-1}(1)) \ge \boldsymbol{Q} - \boldsymbol{\Phi}_2(x_2^{-1}(1))$ by monotonicity, so $(x_1, \boldsymbol{\Phi}_2) \in A_{T,L}$. Therefore,

$$|\mathscr{C}(x_1, \boldsymbol{\Phi}_1) - \mathscr{C}(x_2, \boldsymbol{\Phi}_2)| \le |\mathscr{C}(x_1, \boldsymbol{\Phi}_1) - \mathscr{C}(x_1, \boldsymbol{\Phi}_2)| + |\mathscr{C}(x_1, \boldsymbol{\Phi}_2) - \mathscr{C}(x_2, \boldsymbol{\Phi}_2)|.$$

Therefore, it suffices to show that the functional is Lipschitz in each of its coordinates,

$$|\mathscr{C}(x,\mathbf{\Phi}_1)-\mathscr{C}(x,\mathbf{\Phi}_2)|\le C_L\|\mathbf{\Phi}_1-\mathbf{\Phi}_2\|_\infty \quad\text{and}\quad |\mathscr{C}(x_1,\mathbf{\Phi})-\mathscr{C}(x_2,\mathbf{\Phi})|\le C_L\|x_1-x_2\|_1.$$

We start by showing the first inequality. The computation to show the functional is Lipschitz in x for fixed $\mathbf{\Phi}$ follows the similar computations.

Lipschitz in $\mathbf{\Phi}$: Fix $x(t)$ and consider $(x,\mathbf{\Phi}_1),(x,\mathbf{\Phi}_2)\in A_{T,L}$. We first show that the functional is Lipschitz with respect to the infinity norm on matrix paths,

$$|\mathscr{C}(x,\mathbf{\Phi}_1)-\mathscr{C}(x,\mathbf{\Phi}_2)|\le C_L\|\mathbf{\Phi}_1-\mathbf{\Phi}_2\|_\infty. \tag{5.107}$$

We will show that each term in $\mathscr{C}(x,\mathbf{\Phi})$ is Lipschitz in $\mathbf{\Phi}$ for fixed x.

(a) The integrand consists of functions of bounded variation, so we can integrate by parts to conclude

$$\begin{aligned}\langle\vec{h}\vec{h}^{\mathsf{T}},\hat{\mathbf{\Phi}}(0)\rangle &= \int_0^n x(t)\cdot\frac{d}{dt}\langle\vec{h}\vec{h}^{\mathsf{T}},\mathbf{\Phi}(t)\rangle\,dt\\ &= \langle\vec{h}\vec{h}^{\mathsf{T}},\boldsymbol{Q}\rangle-\int_0^n\langle\vec{h}\vec{h}^{\mathsf{T}},\mathbf{\Phi}(t)\rangle dx(t).\end{aligned}$$

Since $x(0)=0$ and $x(n)=1$, we have

$$\begin{aligned}|\langle\vec{h}\vec{h}^{\mathsf{T}},\hat{\mathbf{\Phi}}_1(0)\rangle-\langle\vec{h}\vec{h}^{\mathsf{T}},\hat{\mathbf{\Phi}}_2(0)\rangle| &\le \int_0^n\Big|\langle\vec{h}\vec{h}^{\mathsf{T}},\mathbf{\Phi}_1(t)-\mathbf{\Phi}_2(t)\rangle\Big|\,dx(t)\\ &\le n^2\|\vec{h}\vec{h}^{\mathsf{T}}\|_\infty\|\mathbf{\Phi}_1-\mathbf{\Phi}_2\|_\infty.\end{aligned}$$

(b) The second term can be bounded in a similar manner using integration by parts,

$$\begin{aligned}\int_0^n x(t)\langle\boldsymbol{\xi}'(\mathbf{\Phi}(t)),\mathbf{\Phi}'(t)\rangle\,dt &= \int_0^n x(t)\cdot\frac{d}{dt}\operatorname{Sum}(\boldsymbol{\xi}(\mathbf{\Phi}(t)))\,dt\\ &= \operatorname{Sum}(\boldsymbol{\xi}(\boldsymbol{Q}))-\int_0^n\operatorname{Sum}(\boldsymbol{\xi}(\mathbf{\Phi}(t)))\,dx(t).\end{aligned}$$

Since $\boldsymbol{\xi}(t)$ is a power series and $\mathbf{\Phi}$ is bounded, we can conclude

$$\left|\int_0^n x(t)\langle\boldsymbol{\xi}'(\mathbf{\Phi}_1(t)),\mathbf{\Phi}_1'(t)\rangle\,dt-\int_0^n x(t)\langle\boldsymbol{\xi}'(\mathbf{\Phi}_2(t)),\mathbf{\Phi}_2'(t)\rangle\,dt\right|\le n^2\|\boldsymbol{\xi}'(1)\|_\infty\|\mathbf{\Phi}_1-\mathbf{\Phi}_2\|_\infty.$$

(c) The condition $\|(\boldsymbol{Q}-\mathbf{\Phi}(t_x))^{-1}\|_\infty\le L$ and equivalence of the infinity norm and operator norm on $\mathbb{R}^{n\times n}$ implies that

$$\lambda_{min}(\boldsymbol{Q}-\mathbf{\Phi}(t_x))=\frac{1}{\lambda_{max}((\boldsymbol{Q}-\mathbf{\Phi}(t_x))^{-1})}\ge\frac{1}{\sqrt{n}\|(\boldsymbol{Q}-\mathbf{\Phi}(t_x))^{-1}\|_\infty}\ge\frac{1}{\sqrt{n}L}.$$

The determinant is the product of eigenvalues so $|\boldsymbol{Q}-\mathbf{\Phi}(t_x)|\ge(\sqrt{n}L)^{-n}>0$. Furthermore, $\log(t)$ is Lipschitz on $[(\sqrt{n}L)^{-n},\infty)$ and $|\boldsymbol{A}|$ is a polynomial of the entires of $\boldsymbol{A}$, so there exists universal constants C_1,C_2 that depends only on L and the dimension n such that

$$|\log|\boldsymbol{Q}-\mathbf{\Phi}_1(t_x)|-\log|\boldsymbol{Q}-\mathbf{\Phi}_2(t_x)||\le C_1\big||\boldsymbol{Q}-\mathbf{\Phi}_1(t_x)|-|\boldsymbol{Q}-\mathbf{\Phi}_2(t_x)|\big|\le C_2\|\mathbf{\Phi}_1-\mathbf{\Phi}_2\|_\infty.$$

(d) To show the last term is Lipschitz, we will show that all of its unit directional derivatives are uniformly bounded and apply the mean value theorem to conclude Lipschitz continuity. Let $\boldsymbol{\Phi}, \boldsymbol{\Psi}$ be arbitrary matrices such that $(x, \boldsymbol{\Phi}), (x, \boldsymbol{\Psi}) \in A_{T,L}$. By monotonicity $(1-\varepsilon)\boldsymbol{\Phi} + \varepsilon\boldsymbol{\Psi}$ is also a Lipschitz monotone path and since the matrix inverse is convex [Proposition 3.3.11],

$$\big(\boldsymbol{Q} - (\varepsilon\boldsymbol{\Phi}(t_x) + (1-\varepsilon)\boldsymbol{\Psi}(t_x))\big)^{-1} \leq \varepsilon(\boldsymbol{Q} - \boldsymbol{\Phi}(t_x)))^{-1} + (1-\varepsilon)(\boldsymbol{Q} - \boldsymbol{\Phi}(t_x)))^{-1}$$

so $(x, (1-\varepsilon)\boldsymbol{\Phi} + \varepsilon\boldsymbol{\Psi}) \in A_{T,L}$ for all $\varepsilon \in [0,1]$.

If we set $\boldsymbol{\Theta}(t) = \frac{\boldsymbol{\Psi}(t) - \boldsymbol{\Phi}(t)}{\|\boldsymbol{\Psi} - \boldsymbol{\Phi}\|_\infty}$, then for all $\varepsilon \in [0, \|\boldsymbol{\Psi} - \boldsymbol{\Phi}\|_\infty]$,

$$(x, \boldsymbol{\Phi} + \varepsilon\boldsymbol{\Theta}) = (x, (1-\varepsilon)\boldsymbol{\Phi} + \varepsilon\boldsymbol{\Psi}) \in A_{t,L},$$

Consider the function

$$f(\boldsymbol{\Phi}) = \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t) \rangle \, dt.$$

We will show that the directional derivatives of f in the admissible unit direction $\boldsymbol{\Theta}$ is uniformly bounded by some constant C that only depends on the fixed parameters of the model and the bound L,

$$\left| \frac{d}{d\varepsilon} f(\boldsymbol{\Phi} + \varepsilon\boldsymbol{\Theta}) \Big|_{\varepsilon=0} \right| \leq C.$$

Using [Proposition 3.3.5] to compute the derivative of the inverse,

$$\begin{aligned}
\frac{d}{d\varepsilon} f(\boldsymbol{\Phi} + \varepsilon\boldsymbol{\Theta}) \Big|_{\varepsilon=0} &= \frac{d}{d\varepsilon} \int_0^{t_x} \langle (\hat{\boldsymbol{\Phi}}(t) + \varepsilon\hat{\boldsymbol{\Theta}})^{-1}, \boldsymbol{\Phi}'(t) + \varepsilon\boldsymbol{\Theta}'(t) \rangle \, dt \Big|_{\varepsilon=0} \\
&= -\int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1} \hat{\boldsymbol{\Theta}}(t) \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t) \rangle \, dt + \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Theta}'(t) \rangle \, dt \\
&= -\int_0^{t_x} \int_t^n x(s) \langle \hat{\boldsymbol{\Phi}}(t)^{-1} \boldsymbol{\Theta}'(s) \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t) \rangle \, ds dt + \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Theta}'(t) \rangle \, dt.
\end{aligned}$$

Since $\operatorname{tr}(\boldsymbol{ABC}) = \operatorname{tr}(\boldsymbol{CAB})$, we integrate by parts to conclude

$$\begin{aligned}
&-\int_0^{t_x} \int_t^n x(s) \langle \hat{\boldsymbol{\Phi}}(t)^{-1} \boldsymbol{\Phi}'(t) \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Theta}'(s) \rangle \, ds dt \\
&= \int_0^{t_x} x(t) \langle \hat{\boldsymbol{\Phi}}(t)^{-1} \boldsymbol{\Phi}'(t) \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Theta}(t) \rangle \, dt + \int_0^{t_x} \int_t^n \langle \hat{\boldsymbol{\Phi}}(t)^{-1} \boldsymbol{\Phi}'(t) \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Theta}(s) \rangle \, dx(s) dt
\end{aligned}$$

and

$$\int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Theta}'(t) \rangle \, dt = -\langle \hat{\boldsymbol{\Phi}}(t_x)^{-1}, \boldsymbol{\Theta}(t_x) \rangle - \int_0^{t_x} x(t) \langle \hat{\boldsymbol{\Phi}}(t)^{-1} \boldsymbol{\Phi}'(t) \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Theta}(t) \rangle \, dt.$$

Since $\|\boldsymbol{\Theta}\|_\infty = 1$, $\|\boldsymbol{\Phi}'\|_\infty \leq 1$, and $\|\hat{\boldsymbol{\Phi}}^{-1}(t)\|_\infty \leq \|\boldsymbol{Q} - \boldsymbol{\Phi}(t_x)\|_\infty \leq L$, we can replace each entry of the matrices in the integrand with its highest possible value to get the crude upper bound

$$\left| \frac{d}{d\varepsilon} f(\boldsymbol{\Phi} + \varepsilon\boldsymbol{\Theta}) \Big|_{\varepsilon=0} \right| \leq 4n^4 L^2.$$

This upper bound holds for any starting point $\boldsymbol{\Phi}$ and all admissible directions $\boldsymbol{\Theta}$. We will now show

that this implies that our functional is Lipschitz. Given $\mathbf{\Phi}_1$ monotone paths $\mathbf{\Phi}_2$, we have

$$\mathbf{\Phi}_2 = \mathbf{\Phi}_1 + \|\mathbf{\Phi}_1 - \mathbf{\Phi}_2\|_\infty \mathbf{\Theta},$$

where $\mathbf{\Theta} = \frac{\mathbf{\Phi}_2 - \mathbf{\Phi}_1}{\|\mathbf{\Phi}_1 - \mathbf{\Phi}_2\|_\infty}$. Consider the function $g : [0, \|\mathbf{\Phi}_1 - \mathbf{\Phi}_2\|_\infty] \to \mathbb{R}$,

$$g(t) = f(\mathbf{\Phi}_1 + t\mathbf{\Theta}).$$

First notice that $\mathbf{\Phi}_1 + t\mathbf{\Theta}$ is a monotone path and $(x, \mathbf{\Phi}_1 + t\mathbf{\Theta}) \in A_{T,L}$ for $t \in [0, \|\mathbf{\Phi}_1 - \mathbf{\Phi}_2\|_\infty]$. For $t \in (0, \|\mathbf{\Phi}_1 - \mathbf{\Phi}_2\|_\infty)$ the uniform bound on the directional derivative implies

$$|g'(t)| = \left| \frac{d}{d\varepsilon} f(\mathbf{\Phi}_1 + t\mathbf{\Theta} + \varepsilon\mathbf{\Theta}) \Big|_{\varepsilon=0} \right| \le 4n^4 L^2.$$

Since $g(t)$ is continuous, the mean value theorem implies that

$$\left| \int_0^{t_x} \langle \hat{\mathbf{\Phi}}_1(t)^{-1}, \mathbf{\Phi}_1'(t) \rangle \, dt - \int_0^{t_x} \langle \hat{\mathbf{\Phi}}_2(t)^{-1}, \mathbf{\Phi}_2'(t) \rangle \, dt \right| = |g(0) - g(\|\mathbf{\Phi}_2 - \mathbf{\Phi}_1\|_\infty)| \le 4n^4 L^2 \|\mathbf{\Phi}_1 - \mathbf{\Phi}_2\|_\infty,$$

which proves that $f(\mathbf{\Phi})$ is Lipschitz.

Combining the bounds proved in (a) to (d) completes the proof for (5.107).

Lipschitz in x: Showing the functional is Lipschitz in x follows from a similar computation. Fix $\mathbf{\Phi}(t)$ and consider $(x_1, \mathbf{\Phi}), (x_2, \mathbf{\Phi}) \in A_{T,L}$. We now show that the functional is Lipschitz with respect to the L^1 norm on monotone functions,

$$|\mathscr{C}(x_1, \mathbf{\Phi}) - \mathscr{C}(x_2, \mathbf{\Phi})| \le C_L \|x_1 - x_2\|_1. \tag{5.108}$$

We will show that each term in $\mathscr{C}(x, \mathbf{\Phi})$ is Lipschitz in x for fixed $\mathbf{\Phi}$. To make the dependence of $\hat{\mathbf{\Phi}}$ on x_1 and x_2 explicit, we define

$$\hat{\mathbf{\Phi}}_{x_1}(t) := \int_t^n x_1(s) \mathbf{\Phi}'(s) \, ds \quad \text{and} \quad \hat{\mathbf{\Phi}}_{x_2}(t) := \int_t^n x_2(s) \mathbf{\Phi}'(s) \, ds.$$

(a) The matrix path satisfies $\|\mathbf{\Phi}\|_\infty \le 1$, so

$$\begin{aligned} |\langle \vec{h}\vec{h}^{\mathsf{T}}, \hat{\mathbf{\Phi}}_{x_1}(0) \rangle - \langle \vec{h}\vec{h}^{\mathsf{T}}, \hat{\mathbf{\Phi}}_{x_2}(0) \rangle| &\le \int_0^n |x_1 - x_2| \cdot |\langle \vec{h}\vec{h}^{\mathsf{T}}, \mathbf{\Phi}'(t) \rangle| \, dt \\ &\le n^2 \|\vec{h}\vec{h}^{\mathsf{T}}\|_\infty \|x_1 - x_2\|_1. \end{aligned}$$

(b) The matrix path satisfies $\|\boldsymbol{\xi}'(\mathbf{\Phi})\|_\infty \le \|\boldsymbol{\xi}'(1)\|_\infty$ and $\|\mathbf{\Phi}'\|_\infty \le 1$, so

$$\left| \int_0^n x_1(t) \langle \boldsymbol{\xi}'(\mathbf{\Phi}(t)), \mathbf{\Phi}'(t) \rangle \, dt - \int_0^n x(t) \langle \boldsymbol{\xi}'(\mathbf{\Phi}(t)), \mathbf{\Phi}'(t) \rangle \, dt \right| \le n^2 \|\boldsymbol{\xi}'(1)\|_\infty \|x_1 - x_2\|_1.$$

(c) By observation (5.21), we can replace the bound with $\hat{t} = \sup(\{t \le T \mid \|(\boldsymbol{Q} - \mathbf{\Phi}(t))^{-1}\|_\infty \le L\})$.

We need to show

$$f(x) := \log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(\hat{t})| + \int_0^{\hat{t}} \langle \hat{\boldsymbol{\Phi}}_x(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle \, dt,$$

is Lipschitz in x. The log determinant term is independent of x, so we only need to show that the integral is Lipschitz in x. We will show that all directional derivatives of x with respect to an admissible unit direction is bounded. Let $y(t)$ be another monotone function such that $y(t) = 1$ for all $t \geq \hat{t}$. It is easy to see that $((1-\varepsilon)x(t) + \varepsilon y(t), \boldsymbol{\Phi}) \in A_{T,L}$, so $z(t) = \frac{y(t)-x(t)}{\|x-y\|_\infty}$ is an admissible unit direction. Since $z(t) = 0$ for $t \geq \hat{t}$, Fubini's theorem implies that

$$\begin{aligned}\frac{d}{d\varepsilon} f(x(t) + \varepsilon z(t))\Big|_{\varepsilon=0} &= -\int_0^{\hat{t}} \int_t^{\hat{t}} z(s) \langle \hat{\boldsymbol{\Phi}}_x(t)^{-1} \boldsymbol{\Phi}'(s) \hat{\boldsymbol{\Phi}}_x(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle \, ds dt \\ &= -\int_0^{\hat{t}} \int_0^{s} z(s) \langle \hat{\boldsymbol{\Phi}}_x(t)^{-1} \boldsymbol{\Phi}'(t) \hat{\boldsymbol{\Phi}}_x(t)^{-1}, \boldsymbol{\Phi}'(s)\rangle \, dt ds.\end{aligned}$$

Since $\|z\|_1 = 1$, $\|\boldsymbol{\Phi}'\|_\infty \leq 1$, and $\|\hat{\boldsymbol{\Phi}}_x^{-1}(t)\|_\infty \leq \|\boldsymbol{Q} - \boldsymbol{\Phi}(\hat{t})\|_\infty \leq L$ we can replace each entry of the matrices in the integrand with its highest possible value to get the crude upper bound

$$\left| \frac{d}{d\varepsilon} f(x + \varepsilon z(t))\Big|_{\varepsilon=0} \right| \leq n^4 L^2.$$

The mean value theorem implies that

$$|f(x_1(t)) - f(x_2(t))| \leq n^4 L^2 \|x_1 - x_2\|_1.$$

Combining the bounds proved in (a) to (c) completes the proof for (5.108). □

Since $\mathscr{C}(x, \boldsymbol{\Phi})$ restricted to $A_{T,L}$ is Lipschitz continuous by Lemma 5.4.2, the extreme value theorem implies that $\mathscr{C}$ attains its minimum at some $(x_{T,L}, \boldsymbol{\Phi}_{T,L}) \in A_{T,L}$. We will show the global minimizer of $\mathscr{C}(x, \boldsymbol{\Phi})$ over its domain lies in $A_{\hat{T},\hat{L}}$ for some $\hat{T}$ and $\hat{L}$ that only depends on the fixed parameters of the model,

$$\inf\{\mathscr{C}(x, \boldsymbol{\Phi}) \mid (x, \boldsymbol{\Phi}) \in A_{T,L} \text{ for some } T \in [0, n) \text{ and } L > 0\} = \inf\{\mathscr{C}(x, \boldsymbol{\Phi}) \mid (x, \boldsymbol{\Phi}) \in A_{\hat{T},\hat{L}}\}.$$

This fact is enough to conclude that

$$\inf_{x,\boldsymbol{\Phi}} \mathscr{C}(x, \boldsymbol{\Phi}) = \inf_{A_{\hat{T},\hat{L}}} \mathscr{C}(x, \boldsymbol{\Phi}) = \inf_{r, A_{\hat{T},\hat{L}}} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}) \geq \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}). \tag{5.109}$$

In the second inequality, we used the fact that $\mathscr{C}(x, \boldsymbol{\Phi})$ is Lipchitz on $A_{\hat{T},\hat{L}}$, so its value agrees with the limit points of discrete c.d.f. The bounds (5.106) and (5.109) implies that

$$\inf_{x,\boldsymbol{\Phi}} \mathscr{C}(x, \boldsymbol{\Phi}) = \inf_{r,x,Q} \mathscr{C}_r(\underline{x}, \underline{\boldsymbol{Q}}).$$

It remains to find the parameters $\hat{T}$ and $\hat{L}$. There quantities are derived from the necessary conditions satisfied by the minimizer of $\mathscr{C}(x, \boldsymbol{\Phi})$ obtained by perturbing the critical points of $\mathscr{C}(x, \boldsymbol{\Phi})$. For fixed $\boldsymbol{\Phi}$, we define

$$\hat{T} = \sup\{t \leq n \mid \|(\boldsymbol{Q} - \boldsymbol{\Phi}(t))^{-1}\|_\infty \leq \hat{L}\} \tag{5.110}$$

to denote the largest feasible point in the support of the probability measure corresponding to x for a given Φ and $\hat{L}$.

Proposition 5.4.1

For fixed $\mathbf{\Phi}$, we define

$$x(t) = \operatorname{argmin}_x \mathscr{C}(x, \mathbf{\Phi}),$$

If μ is the probability measure on $[0, n]$ associated with x,

$$x(t) = \mu([0, t)).$$

then μ must be supported on points less than or equal to $\hat{T}$ that maximize the function

$$f(s) := \int_0^s \langle \mathbf{\Psi}(t), \mathbf{\Phi}'(t) \rangle \, dt$$

where

$$\mathbf{\Psi}(t) = \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\mathbf{\Phi}(t)) - \int_0^t \hat{\mathbf{\Phi}}(s)^{-1} \mathbf{\Phi}'(s) \hat{\mathbf{\Phi}}(s)^{-1} \, ds.$$

Proof. For fixed $\mathbf{\Phi}$, let x be the minimizer of $\mathscr{C}(x, \mathbf{\Phi})$ in $A_{\hat{T},\hat{L}}$. The proof involves examining the critical point condition of $\mathscr{C}(x, \mathbf{\Phi})$ by perturbing the c.d.f. Recall that (5.110) denotes the largest feasible point in the of support the measure μ corresponding to x. If $y(t)$ is another c.d.f. such that $y(t) = 1$ for $t \geq \hat{T}$, then $(1-\varepsilon)x(t) + \varepsilon y(t) = x(t) + \varepsilon(y(t) - x(t))$ satisfies the condition $(1-\varepsilon)x(t) + \varepsilon y(t) = 1$ for $t \geq \hat{T}$, so

$$((1-\varepsilon)x(t) + \varepsilon y(t), \mathbf{\Phi}) \in A_{\hat{T},\hat{L}} \quad \text{for all} \quad \varepsilon \in [0, 1].$$

In particular, if we define $z(t) = y(t) - x(t)$, then the right derivative

$$\frac{d}{d\varepsilon} \mathscr{C}(x(t) + \varepsilon z(t))\Big|_{\varepsilon=0} \geq 0$$

since a perturbation of the minimizer in an admissible direction must be non-negative. Taking the directional derivative and using the independence of $\mathscr{C}$ on t_x explained in (5.21), we see that

$$\begin{aligned}
&\frac{d}{d\varepsilon} \mathscr{C}(x(t) + \varepsilon z(t))\Big|_{\varepsilon=0} \\
&= \frac{1}{2}\left(\int_0^n z(t) \langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\mathbf{\Phi}(t)), \mathbf{\Phi}'(t) \rangle \, dt - \int_0^{\hat{T}} \left\langle \hat{\mathbf{\Phi}}(t)^{-1} \left(\int_t^n z(s) \mathbf{\Phi}'(s) \, ds \right) \hat{\mathbf{\Phi}}(t)^{-1}, \mathbf{\Phi}'(t) \right\rangle dt \right).
\end{aligned}$$

Since $z(t) = 0$ for $t \geq \hat{T}$, the second integral can be simplified using Fubini's theorem,

$$\begin{aligned}
&\int_0^{\hat{T}} \left\langle \hat{\boldsymbol{\Phi}}(t)^{-1} \left(\int_t^n z(s) \boldsymbol{\Phi}'(s)\, ds \right) \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t) \right\rangle dt \\
&= \int_0^{\hat{T}} \int_t^{\hat{T}} z(s) \langle \hat{\boldsymbol{\Phi}}(t)^{-1} \boldsymbol{\Phi}'(t) \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(s) \rangle\, ds dt && \operatorname{tr}(\boldsymbol{ABC}) = \operatorname{tr}(\boldsymbol{CAB}) \\
&= \int_0^{\hat{T}} \int_0^{s} z(s) \langle \hat{\boldsymbol{\Phi}}(t)^{-1} \boldsymbol{\Phi}'(t) \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(s) \rangle\, dt ds && 0 \leq t \leq s \leq \hat{T} \\
&= \int_0^{n} \int_0^{t} z(t) \langle \hat{\boldsymbol{\Phi}}(s)^{-1} \boldsymbol{\Phi}'(s) \hat{\boldsymbol{\Phi}}(s)^{-1}, \boldsymbol{\Phi}'(t) \rangle\, ds dt. && \text{relabel } s \text{ and } t
\end{aligned}$$

If we define the matrix,

$$\boldsymbol{\Psi}(t) = \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{\Phi}(t)) - \int_0^t \hat{\boldsymbol{\Phi}}(s)^{-1} \boldsymbol{\Phi}'(s) \hat{\boldsymbol{\Phi}}(s)^{-1}\, ds \tag{5.111}$$

then our computations above implies that

$$\frac{d}{d\varepsilon} \mathscr{C}(x(t) + \varepsilon z(t))\Big|_{\varepsilon=0} = \frac{1}{2} \int_0^n z(t) \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt.$$

Since $z(t) = y(t) - x(t)$, the critical point condition $\frac{d}{d\varepsilon} \mathscr{C}(x(t) + \varepsilon z(t))\big|_{\varepsilon=0} \geq 0$ implies that

$$\int_0^n y(t) \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt \geq \int_0^n x(t) \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt$$

for all functions $y(t)$. From this critical point condition, we are able to recover the support of $\mu(t)$, the measure corresponding to $x(t)$. In particular, if we define

$$y(t) = \nu([0,t]) = \int_0^t d\nu(s),$$

then Fubini's theorem implies that

$$\int_0^n y(t) \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt = \int_0^n \int_0^t \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, d\nu(s) dt = \int_0^n \int_s^n \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt d\nu(s).$$

The critical point condition implies that

$$\int_0^n \int_s^n \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt d\nu(s) \geq \int_0^n \int_s^n \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt d\mu(s). \tag{5.112}$$

Since $\int_0^n d\nu(s) = 1$ and $\int_0^n d\mu(s) = 1$,

$$\int_0^n \int_0^n \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt d\nu(s) = \int_0^n \int_0^n \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt d\mu(s),$$

we can subtract (5.112) to conclude

$$\int_0^n \int_0^s \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt d\nu(s) \leq \int_0^n \int_0^s \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle\, dt d\mu(s)$$

for all measures ν such that $\nu([0,t]) = 1$ for all $t \geq \hat{T}$. In particular, μ must be supported on points less than or equal to $\hat{T}$ that maximize the function

$$f(s) := \int_0^s \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle \, dt.$$

□

The result of Proposition 5.4.1, can derive some conditions on the largest value in the support of the minimizer $x_{\boldsymbol{\Phi}}$ for a given path $\boldsymbol{\Phi}$. We define

$$T_{\boldsymbol{\Phi}} = \inf\{t \leq n \mid |\boldsymbol{Q} - \boldsymbol{\Phi}(t)| = 0\}. \tag{5.113}$$

We will show that there exists a $S_{\boldsymbol{\Phi}} < T_{\boldsymbol{\Phi}}$ such that the minimizer $x_{\boldsymbol{\Phi}}(t) = 1$ for all $t > S_{\boldsymbol{\Phi}}$, i.e. the maximum point in the support of the minimizing measure corresponding to $\boldsymbol{\Phi}$ cannot be arbitrarily close to the largest feasible value in the support $T_{\boldsymbol{\Phi}}$. This also implies that the minimizer is attained where the $\mathscr{C}$ is well defined, proving the validity of the integral form of the functional.

Corollary 5.4.1

For fixed $\boldsymbol{\Phi}$, if we define

$$x_{\boldsymbol{\Phi}}(t) = \operatorname{argmin}_x \mathscr{C}(x, \boldsymbol{\Phi}),$$

then there exists a $S_{\boldsymbol{\Phi}}$ such that $x_{\boldsymbol{\Phi}}(t) = 1$ for all $t > S_{\boldsymbol{\Phi}}$. The $S_{\boldsymbol{\Phi}}$ is given explicitly by

$$S_{\boldsymbol{\Phi}} = \inf\{s \leq T \mid 3\langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{\Phi}(s) \rangle + \log|\boldsymbol{Q}| - 2\log|\boldsymbol{Q} - \boldsymbol{\Phi}(T^*)| + \log|\boldsymbol{Q} - \boldsymbol{\Phi}(s)| < 0\}$$

where $T^* < T_{\boldsymbol{\Phi}}$ is given by

$$T^* = \inf\{t < T_{\boldsymbol{\Phi}} \mid 2\boldsymbol{\Phi}(t) - \boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}) \geq 0\}.$$

Remark 5.4.1. Since $\boldsymbol{\Phi}(s) \to \boldsymbol{Q}$ and $\langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{\Phi}(s) \rangle \leq \langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{Q} \rangle$, the function

$$3\langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{\Phi}(s) \rangle + \log|\boldsymbol{Q}| - 2\log|\boldsymbol{Q} - \boldsymbol{\Phi}(T^*)| + \log|\boldsymbol{Q} - \boldsymbol{\Phi}(s)| \to -\infty$$

as $s \to T_{\boldsymbol{\Phi}}$. This implies that $S_{\boldsymbol{\Phi}} < T_{\boldsymbol{\Phi}}$, so $(x_{\boldsymbol{\Phi}}, \boldsymbol{\Phi}) \in A_{S_{\boldsymbol{\Phi}}, \|(\boldsymbol{Q} - \boldsymbol{\Phi}(S_{\boldsymbol{\Phi}}))^{-1}\|}$ and $\mathscr{C}(x_{\boldsymbol{\Phi}}, \boldsymbol{\Phi})$ is well defined and locally Lipchitz around the minimizer.

Proof. For fixed $\boldsymbol{\Phi}$, we will show that the function

$$f(s) := \int_0^s \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle \, dt \quad \text{where} \quad \boldsymbol{\Psi}(t) = \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{\Phi}(t)) - \int_0^t \hat{\boldsymbol{\Phi}}(s)^{-1} \boldsymbol{\Phi}'(s) \hat{\boldsymbol{\Phi}}(s)^{-1} \, ds.$$

appearing in Proposition 5.4.1 is strictly negative for s sufficiently close to $T_{\boldsymbol{\Phi}}$. Since $f(0) = 0$, Proposition 5.4.1 implies that the support of the minimizer μ cannot contain points such that $f(s) < 0$. This will prove that for fixed $\boldsymbol{\Phi}$, that the minimizer $x_{\boldsymbol{\Phi}}(t)$ cannot contain a point too close to $T_{\boldsymbol{\Phi}}$.

Recall that $T_{\boldsymbol{\Phi}}$ defined in (5.113) denotes the largest feasible point in the support. Since $x_{\boldsymbol{\Phi}}(t) \leq 1$,

we can multiply by this term to explicitly compute the integral,

$$\begin{aligned}
\int_0^s \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t)\rangle\, dt &= \int_0^s \langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{\Phi}'(t)\rangle\, dt - \int_0^s \int_0^t \langle \hat{\boldsymbol{\Phi}}(r)^{-1}\boldsymbol{\Phi}'(r)\hat{\boldsymbol{\Phi}}(r)^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dr\, dt \\
(x(r) \le 1) \quad &\le \int_0^s \langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{\Phi}'(t)\rangle\, dt - \int_0^s \int_0^t x(r)\langle \hat{\boldsymbol{\Phi}}(r)^{-1}\boldsymbol{\Phi}'(r)\hat{\boldsymbol{\Phi}}(r)^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dr\, dt \\
&= \int_0^s \langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{\Phi}'(t)\rangle\, dt - \int_0^s \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dt + \int_0^s \langle \hat{\boldsymbol{\Phi}}(0)^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dt \\
&\le \int_0^s \langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{\Phi}'(t)\rangle\, dt - \int_0^s \langle (\boldsymbol{Q} - \boldsymbol{\Phi}(t))^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dt + \int_0^s \langle \hat{\boldsymbol{\Phi}}(0)^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dt \\
&= \langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}) + \hat{\boldsymbol{\Phi}}(0)^{-1}, \boldsymbol{\Phi}(s)\rangle - \log|\boldsymbol{Q}| + \log|\boldsymbol{Q} - \boldsymbol{\Phi}(s)|. \qquad (5.114)
\end{aligned}$$

We want to remove the dependence of this bound from the $x(t)$. We do this by proving an upper bound on $\hat{\boldsymbol{\Phi}}(0)^{-1}$ independent of $x(t)$. Let $y(t) \equiv 1$, i.e. the measure is concentrated at 0. In this case, the free energy can be simplified explicitly, and this value is independent of $\boldsymbol{\Phi}$. In this case $\hat{\boldsymbol{\Phi}}(t) = \boldsymbol{Q} - \boldsymbol{\Phi}(t)$ and $t_y = 0$, so

$$\begin{aligned}
\mathscr{C}(y, \boldsymbol{\Phi}) &= \frac{1}{2}\bigg(\int_0^n y(t)\langle \boldsymbol{\xi}'(\boldsymbol{\Phi}(t)) + \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{\Phi}'(t)\rangle\, dt + \log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(t_y)| + \int_0^{t_y} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dt \bigg) \\
&= \frac{1}{2}\bigg(\langle \boldsymbol{\xi}(\boldsymbol{Q}) + \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{Q}\rangle + \log|\boldsymbol{Q}| \bigg).
\end{aligned}$$

We want to now compute a lower bound on $\mathscr{C}(x, \boldsymbol{\Phi})$ for general $x(t)$. Because both $\boldsymbol{\xi}'(\boldsymbol{\Phi}(t))$ and $\vec{h}\vec{h}^{\mathsf{T}}$ are positive semidefinite,

$$\int_0^n x(t)\langle \boldsymbol{\xi}'(\boldsymbol{\Phi}(t)) + \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{\Phi}'(t)\rangle\, dt \ge 0.$$

This implies

$$\mathscr{C}(x, \boldsymbol{\Phi}) \ge \log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(t_x)| + \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dt \qquad (5.115)$$

Since $\hat{\boldsymbol{\Phi}}(t) = \boldsymbol{Q} - \boldsymbol{\Phi}(t)$ for $t_x \le t \le T_{\boldsymbol{\Phi}}$,

$$\log|\boldsymbol{\Phi}(n) - \boldsymbol{\Phi}(t_x)| + \int_0^{t_x} \langle \hat{\boldsymbol{\Phi}}(t)^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dt = \int_0^{T_{\boldsymbol{\Phi}}} \langle \hat{\boldsymbol{\Phi}}(t)^{-1} - (\boldsymbol{Q} - \boldsymbol{\Phi}(t))^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dt + \log|\boldsymbol{Q}|.$$

Because $\hat{\boldsymbol{\Phi}}(t) \le \boldsymbol{Q} - \boldsymbol{\Phi}(t)$ we also have

$$\langle \hat{\boldsymbol{\Phi}}(t)^{-1} - (\boldsymbol{Q} - \boldsymbol{\Phi}(t))^{-1}, \boldsymbol{\Phi}'(t)\rangle \ge 0 \text{ for all } t < T_{\boldsymbol{\Phi}},$$

so the integral is well defined. For any $T^* < T_{\boldsymbol{\Phi}}$ we have the lower bound,

$$\begin{aligned}
\int_0^{T_{\boldsymbol{\Phi}}} \langle \hat{\boldsymbol{\Phi}}(t)^{-1} - (\boldsymbol{Q} - \boldsymbol{\Phi}(t))^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dt &\ge \int_0^{T^*} \langle \hat{\boldsymbol{\Phi}}(0)^{-1} - (\boldsymbol{Q} - \boldsymbol{\Phi}(t))^{-1}, \boldsymbol{\Phi}'(t)\rangle\, dt \\
&= \langle \hat{\boldsymbol{\Phi}}(0)^{-1}, \boldsymbol{\Phi}(T^*)\rangle + \log|\boldsymbol{Q} - \boldsymbol{\Phi}(T^*)| - \log|\boldsymbol{Q}|
\end{aligned}$$

since $\hat{\boldsymbol{\Phi}}(0) \geq \hat{\boldsymbol{\Phi}}(t)$ for all $t < T_{\boldsymbol{\Phi}}$. Applying this to (5.115) implies that

$$\mathscr{C}(x, \boldsymbol{\Phi}) \geq \frac{1}{2}\bigg(\langle \hat{\boldsymbol{\Phi}}(0)^{-1}, \boldsymbol{\Phi}(T^*) \rangle + \log |\boldsymbol{Q} - \boldsymbol{\Phi}(T^*)| \bigg)$$

Since we require $\inf_{x,\boldsymbol{\Phi}} \mathscr{C}(x, \boldsymbol{\Phi}) \leq \mathscr{C}(y, \boldsymbol{\Phi})$, we must have

$$\langle \hat{\boldsymbol{\Phi}}(0)^{-1}, \boldsymbol{\Phi}(T^*) \rangle \leq -\log |\boldsymbol{Q} - \boldsymbol{\Phi}(T^*)| + \langle \boldsymbol{\xi}(\boldsymbol{Q}) + \vec{h}\vec{h}^{\mathsf{T}}, \boldsymbol{Q} \rangle + \log |\boldsymbol{Q}|, \tag{5.116}$$

if $x_{\boldsymbol{\Phi}}(t)$ is a minimizer. This bound holds for all T^*, so we can define

$$T^* = \inf\{t < T_{\boldsymbol{\Phi}} \mid 2\boldsymbol{\Phi}(t) - \boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}) \geq 0\}.$$

We will verify that $T^* < T_{\boldsymbol{\Phi}}$ at the end of this proof. For $s \geq T^*$, this particular choice implies

$$\boldsymbol{\Phi}(s) - \boldsymbol{\Phi}(T^*) \leq \boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}) - \boldsymbol{\Phi}(T^*) \leq \boldsymbol{\Phi}(T^*),$$

so

$$\langle \hat{\boldsymbol{\Phi}}(0)^{-1}, \boldsymbol{\Phi}(s) \rangle = \langle \hat{\boldsymbol{\Phi}}(0)^{-1}, \boldsymbol{\Phi}(T^*) \rangle + \langle \hat{\boldsymbol{\Phi}}(0)^{-1}, \boldsymbol{\Phi}(s) - \boldsymbol{\Phi}(T^*) \rangle \leq 2\langle \hat{\boldsymbol{\Phi}}(0)^{-1}, \boldsymbol{\Phi}(T^*) \rangle. \tag{5.117}$$

Therefore, if we can show that $T^* < T_{\boldsymbol{\Phi}}$, then (5.114) combined with the bounds (5.116) and (5.117) implies that the largest point s in the support of the minimizer must satisfy

$$\int_0^s \langle \Psi(t), \boldsymbol{\Phi}'(t) \rangle \, dt \leq 3\langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{\Phi}(s) \rangle + \log |\boldsymbol{Q}| - 2\log |\boldsymbol{Q} - \boldsymbol{\Phi}(T^*)| + \log |\boldsymbol{Q} - \boldsymbol{\Phi}(s)|$$

Since $\boldsymbol{\Phi}(s) \to \boldsymbol{Q}$, we have $\log |\boldsymbol{Q} - \boldsymbol{\Phi}(s)| \to -\infty$, so

$$\int_0^s \langle \boldsymbol{\Psi}(t), \boldsymbol{\Phi}'(t) \rangle \, dt < 0$$

for s sufficiently large. This will contradict Proposition 5.4.1 for any fixed $\boldsymbol{\Phi}$ and L sufficiently large, so the support of the minimizer cannot contain points too close to $T_{\boldsymbol{\Phi}}$. In particular, we have proved that there exists a $S_{\boldsymbol{\Phi}}$ that depends on the fixed parameters of the model and $\boldsymbol{\Phi}$ such that $x_{\boldsymbol{\Phi}}(t) = 1$ for $t > S_{\boldsymbol{\Phi}}$. The $S_{\boldsymbol{\Phi}}$ is given explicitly by

$$S_{\boldsymbol{\Phi}} = \inf\{s \leq T_{\boldsymbol{\Phi}} \mid 3\langle \vec{h}\vec{h}^{\mathsf{T}} + \boldsymbol{\xi}'(\boldsymbol{Q}), \boldsymbol{\Phi}(s) \rangle + \log |\boldsymbol{Q}| - 2\log |\boldsymbol{Q} - \boldsymbol{\Phi}(T^*)| + \log |\boldsymbol{Q} - \boldsymbol{\Phi}(s)| < 0\},$$

proving the result.

It remains to show that $T^* < T_{\boldsymbol{\Phi}}$, so that $\log |\boldsymbol{Q} - \boldsymbol{\Phi}(T^*)|$ is well defined. To see that $T^* < T_{\boldsymbol{\Phi}}$, we need to show that there exists at $T^* < T_{\boldsymbol{\Phi}}$ such that

$$\vec{x}^{\mathsf{T}}(2\boldsymbol{\Phi}(t) - \boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}))\vec{x} = \vec{x}^{\mathsf{T}}\boldsymbol{\Phi}(T^*)\vec{x} - \vec{x}^{\mathsf{T}}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}) - \boldsymbol{\Phi}(T^*))\vec{x} \geq 0$$

for all $\|\vec{x}\| = 1$. We first notice that by monotonicity,

$$\vec{x}^{\mathsf{T}}\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}})\vec{x} = 0 \implies \vec{x}^{\mathsf{T}}\boldsymbol{\Phi}(t)\vec{x} = 0$$

for all $\vec{x}$. Since $\vec{x}^{\mathsf{T}}\boldsymbol{A}\vec{x} = 0 \iff \vec{x} \in \mathrm{Null}(\boldsymbol{A})$, this means $\mathrm{Null}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}})) \subseteq \mathrm{Null}(\boldsymbol{\Phi}(t))$. In fact, because

$\boldsymbol{\Phi}(t) \to \boldsymbol{\Phi}(T_{\boldsymbol{\Phi}})$ we have $\operatorname{Null}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}})) = \operatorname{Null}(\boldsymbol{\Phi}(t))$ for large enough t. This is because by continuity, both $\boldsymbol{\Phi}(t)$ and $\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}})$ must eventually have the same rank. Therefore, $\operatorname{Null}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}})) = \operatorname{Null}(\boldsymbol{\Phi}(t))$ for large enough $T_{\boldsymbol{\Phi}}$. This implies that

$$\vec{x}^{\mathsf{T}}\boldsymbol{\Phi}(T^*)\vec{x} - \vec{x}^{\mathsf{T}}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}) - \boldsymbol{\Phi}(T^*))\vec{x} = 0 \tag{5.118}$$

whenever $x \in \operatorname{Null}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}))$. Therefore, we only have to consider $\vec{x}$ that is perpendicular to the null space of $\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}})$. This is because

$$(\vec{x} + \vec{y})^{\mathsf{T}}\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}})(\vec{x} + \vec{y}) = \vec{y}^{\mathsf{T}}\boldsymbol{\Phi}(t)\vec{y}$$

if $\vec{x} \in \operatorname{Null}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}))$ and $\vec{y} \in \operatorname{Null}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}))^{\perp}$. This means that we can optimize the vectors in $\operatorname{Null}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}))^{\perp}$ instead of all vectors satisfying $\|\vec{x}\| = 1$.

Furthermore, using the Rayleigh quotient to find a lower bound on the quadratic forms in terms of its non-zero eigenvalues, we can use continuity of $\boldsymbol{\Phi}$ to find T^* sufficiently large such that for unit vectors $\vec{x} \in \operatorname{Null}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}))^{\perp}$,

$$\vec{x}^{\mathsf{T}}\boldsymbol{\Phi}(T^*)\vec{x} - \vec{x}^{\mathsf{T}}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}) - \boldsymbol{\Phi}(T^*))\vec{x} > \lambda_{\min}\boldsymbol{\Phi}(T^*) - \lambda_{\max}(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}) - \boldsymbol{\Phi}(T^*)) \geq 0$$

where $\lambda_{\min}$ is the smallest non-zero eigenvalue. We can conclude that $T^* < T$ because $\lambda_{\min}\boldsymbol{\Phi}(t) \to \lambda_{\min}\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}) > 0$ and $(\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}) - \boldsymbol{\Phi}(t)) \to 0$ as $t \to T_{\boldsymbol{\Phi}}$. From (5.118), we can also take $T^* < T$ slightly larger to ensure that $\boldsymbol{\Phi}(T^*)$ and $\boldsymbol{\Phi}(T_{\boldsymbol{\Phi}})$ have the same rank. We have shown that there exists a $T^* < T_{\boldsymbol{\Phi}}$ such that

$$2\boldsymbol{\Phi}(T^*) - \boldsymbol{\Phi}(T_{\boldsymbol{\Phi}}) \geq 0,$$

as required. □

Remark 5.4.2. In comparison to the one dimensional argument of this fact [98, Lemma 2.1 and Proposition 2.1], the proof in the vector spin case is much longer and trickier. The main difficulty is the fact that

$$\int_0^s \int_0^t \langle \hat{\boldsymbol{\Phi}}(r)^{-1}\boldsymbol{\Phi}'(r)\hat{\boldsymbol{\Phi}}(r)^{-1}, \boldsymbol{\Phi}'(t)\rangle \, dr \, dt$$

is not a monotone function of $\boldsymbol{\Phi}$, so extra care has to be made to prove a suitable upper bound of $f(s)$.

By Corollary 5.4.1 have shown that for every fixed $\boldsymbol{\Phi}$, there exists a $S_{\boldsymbol{\Phi}}$ such that the minimizer $x_{\boldsymbol{\Phi}}(t) = 1$ for $t > S_{\boldsymbol{\Phi}}$. This implies that

$$\inf_x \mathscr{C}(x, \boldsymbol{\Phi}) = \inf_{x \in A_{S_{\boldsymbol{\Phi}}, \infty}} \mathscr{C}(x, \boldsymbol{\Phi})$$

so we can restrict the domain of $\mathscr{C}(x, \boldsymbol{\Phi})$ to the region where it is well defined. Furthermore, it is easy to see that the map $x \mapsto \mathscr{C}(x, \boldsymbol{\Phi})$ is strictly convex for the same reasons the one dimensional Crisanti–Sommers functional is strictly convex. Therefore, for each fixed $\boldsymbol{\Phi}$, the optimization problem $\inf_x \mathscr{C}(x, \boldsymbol{\Phi})$ has a unique minimizer $x_{\boldsymbol{\Phi}}$. Therefore, the function $\mathscr{C}(\boldsymbol{\Phi}) := \mathscr{C}(x_{\boldsymbol{\Phi}}, \boldsymbol{\Phi})$ is a well defined continuous function with respect to $\boldsymbol{\Phi}$ and the space of monotone functions $\boldsymbol{\Phi}$ is a convex compact set so $\inf_{x,\boldsymbol{\Phi}} \mathscr{C}(x, \boldsymbol{\Phi})$ is attained at some $(x^*, \boldsymbol{\Phi}^*)$ by the extreme value theorem. This minimizer may not be unique, but we can pick one such minimizer and define $\hat{L} = \|(\boldsymbol{Q} - \boldsymbol{\Phi}^*(T_{\boldsymbol{\Phi}^*}))^{-1}\|_{\infty}$ and $\hat{T} = S_{\boldsymbol{\Phi}^*}$ in (5.109) to conclude that the integral form $\mathscr{C}(x, \boldsymbol{\Phi})$ is the correct continuous extension of the discrete

Crisanti–Sommers functional.

Remark 5.4.3. We have shown that the minimizer of $\mathscr{C}(x, \mathbf{\Phi})$ is attained where the functional $\mathscr{C}(x, \mathbf{\Phi})$ is well defined. The functional $\mathscr{C}(x, \mathbf{\Phi})$ is locally Lipschitz continuous in a neighbourhood of the region where the minimizer is attained, so it can be approximated by the discrete Crisanti–Sommers functional. This proves equivalence of the integral and discrete forms of the Crisanti–Sommers functionals.

Bibliography

[1] M. Aizenman and P. Contucci. On the stability of the quenched state in mean-field spin-glass models. *Journal of Statistical Physics*, 92(5-6):765–783, Sept. 1998.

[2] M. Aizenman, R. Sims, and S. L. Starr. Extended variational principle for the sherrington-kirkpatrick spin-glass model. *Phys. Rev. B*, 68:214403, Dec 2003.

[3] A. E. Alaoui, A. Montanari, and M. Sellke. Optimization of mean-field spin glasses, 2020.

[4] P. W. Anderson. Spin Glass I: A Scaling Law Rescued. *Physics Today*, 41(1):9, Jan. 1988.

[5] P. W. Anderson. Spin Glass II: Is There a Phase Transition? *Physics Today*, 41(3):9, Jan. 1988.

[6] P. W. Anderson. Spin Glass III: Theory Raises its Head. *Physics Today*, 41(6):9, Jan. 1988.

[7] P. W. Anderson. Spin Glass IV: Glimmerings of trouble. *Physics Today*, 41(9):9–11, Sept. 1988.

[8] P. W. Anderson. Spin Glass V: Real Power Brought to Bear. *Physics Today*, 42(7):9, Jan. 1989.

[9] P. W. Anderson. Spin Glass VI: Spin Glass As Cornucopia. *Physics Today*, 42(9):9, Jan. 1989.

[10] P. W. Anderson. Spin Glass VII: Spin Glass as Paradigm. *Physics Today*, 43(3):9, Jan. 1990.

[11] L.-P. Arguin and M. Aizenman. On the structure of quasi-stationary competing particle systems. *Ann. Probab.*, 37(3):1080–1113, 2009.

[12] G. B. Arous and A. Jagannath. Spectral gap estimates in mean field spin glasses. *arXiv preprint arXiv:1705.04243*, 2017. Communications in Mathematical Physics, to appear.

[13] S. Artstein-Avidan, A. Giannopoulos, and V. D. Milman. Asymptotic geometric analysis, part i. *Mathematical Surveys and Monographs*, 202, 2015.

[14] A. Auffinger, G. Ben Arous, and J. Černý. Random matrices and complexity of spin glasses. *Comm. Pure Appl. Math.*, 66(2):165–201, 2013.

[15] A. Auffinger and W.-K. Chen. On properties of Parisi measures. *Probab. Theory Related Fields*, 161(3-4):817–850, 2015.

[16] A. Auffinger and W.-K. Chen. The Parisi formula has a unique minimizer. *Comm. Math. Phys.*, 335(3):1429–1444, 2015.

[17] A. Auffinger and W.-K. Chen. Parisi formula for the ground state energy in the mixed p-spin model. *Ann. Probab.*, 45(6B):4617–4631, 2017.

[18] A. Auffinger and W.-K. Chen. Parisi formula for the ground state energy in the mixed p-spin model. *Ann. Probab.*, 45(6B):4617–4631, 2017.

[19] A. Auffinger and W.-K. Chen. On the energy landscape of spherical spin glasses. *Adv. Math.*, 330:553–588, 2018.

[20] A. Auffinger and Y. Zhou. On properties of the spherical mixed vector p-spin model. 2020.

[21] D. Belius and N. Kistler. The TAP-Plefka variational principle for the spherical SK model. *Comm. Math. Phys.*, 367(3):991–1017, 2019.

[22] G. Ben Arous, A. Dembo, and A. Guionnet. Cugliandolo-Kurchan equations for dynamics of spin-glasses. *Probab. Theory Related Fields*, 136(4):619–660, 2006.

[23] E. Bolthausen and A.-S. Sznitman. On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.*, 197(2):247–276, 1998.

[24] A. J. Bray and M. A. Moore. Replica theory of quantum spin glasses. *Journal of Physics C: Solid State Physics*, 13(24):L655–L660, aug 1980.

[25] S. Chatterjee. Chaos, concentration, and multiple valleys, 2008.

[26] S. Chatterjee. Disorder chaos and multiple valleys in spin glasses, 2009.

[27] S. Chatterjee. *Superconcentration and related topics*. Springer Monographs in Mathematics. Springer, Cham, 2014.

[28] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg, 2011.

[29] W.-K. Chen. Chaos in the mixed even-spin models. *Communications in Mathematical Physics*, 328, 11 2012.

[30] W.-K. Chen. The Aizenman-Sims-Starr scheme and Parisi formula for mixed p-spin spherical models. *Electron. J. Probab.*, 18:no. 94, 14, 2013.

[31] W.-K. Chen. Variational representations for the Parisi functional and the two-dimensional Guerra-Talagrand bound. *Ann. Probab.*, 45(6A):3929–3966, 2017.

[32] W.-K. Chen, H.-W. Hsieh, C.-R. Hwang, and Y.-C. Sheu. Disorder chaos in the spherical mean-field model. *J. Stat. Phys.*, 160(2):417–429, 2015.

[33] W.-K. Chen and D. Panchenko. Temperature chaos in some spherical mixed p-spin models. *J. Stat. Phys.*, 166(5):1151–1162, 2017.

[34] W.-K. Chen and D. Panchenko. On the TAP free energy in the mixed p-spin models. *Comm. Math. Phys.*, 362(1):219–252, 2018.

[35] W.-K. Chen, D. Panchenko, and E. Subag. The generalized tap free energy, 2018.

[36] W.-K. Chen, D. Panchenko, and E. Subag. The generalized tap free energy ii, 2019.

[37] W.-K. Chen and A. Sen. Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed p-spin models. *Comm. Math. Phys.*, 350(1):129–173, 2017.

[38] W.-K. Chen and A. Sen. Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed p-spin models. *Comm. Math. Phys.*, 350(1):129–173, 2017.

[39] P. Contucci and E. Mingione. A multi-scale Sherrington-Kirkpatrick model. *arXiv preprint arXiv:1804.00629*, 2018.

[40] A. 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.

[41] J. R. L. de Almeida and D. J. Thouless. Stability of the sherrington-kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, 11(5):983–990, may 1978.

[42] A. Dembo, A. Guionnet, and C. Mazza. Limiting dynamics for spherical models of spin glasses at high temperature. *J. Stat. Phys.*, 128(4):847–881, 2007.

[43] B. Derrida. Random-energy model: limit of a family of disordered models. *Phys. Rev. Lett.*, 45(2):79–82, 1980.

[44] B. Derrida. Random-energy model: an exactly solvable model of disordered systems. *Phys. Rev. B (3)*, 24(5):2613–2626, 1981.

[45] B. Derrida. A generalization of the Random Energy Model which includes correlations between energies. *Journal de Physique Lettres*, 46(9):401–407, 1985.

[46] B. Derrida and E. Gardner. Solution of the generalised random energy model. *Journal of Physics C: Solid State Physics*, 19(13):2253–2274, may 1986.

[47] D. Elderfield and D. Sherrington. The curious case of the potts spin glass. *Journal of Physics C: Solid State Physics*, 16(15):L497–L503, may 1983.

[48] S. Franz, G. Parisi, and M. Virasoro. The replica method on and off equilibrium. *http://dx.doi.org/10.1051/jp1:1992115*, 2, 10 1992.

[49] S. Franz, G. Parisi, and M. A. Virasoro. Free-energy cost for ultrametricity violations in spin glasses. *Europhysics Letters (EPL)*, 22(6):405–411, may 1993.

[50] S. K. Ghatak and D. Sherrington. Crystal field effects in a general s ising spin glass. *Journal of Physics C: Solid State Physics*, 10(16):3149–3156, aug 1977.

[51] R. Gheissari and A. Jagannath. On the spectral gap of spherical spin glass dynamics. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(2):756–776, 2019.

[52] S. Ghirlanda and F. Guerra. General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity. *J. Phys. A*, 31(46):9149–9155, 1998.

[53] F. Guerra. About the overlap distribution in mean field spin glass models. *International Journal of Modern Physics B*, 10(13n14):1675–1684, 1996.

[54] F. Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in mathematical physics*, 233(1):1–12, 2003.

[55] F. Guerra and F. L. Toninelli. The thermodynamic limit in mean field spin glass models. *Comm. Math. Phys.*, 230(1):71–79, 2002.

[56] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[57] A. Jagannath, J. Ko, and S. Sen. Max κ-cut and the inhomogeneous Potts spin glass. *Ann. Appl. Probab.*, 28(3):1536–1572, 2018.

[58] A. Jagannath, P. Lopatto, and L. Miolane. Statistical thresholds for tensor pca, 2018.

[59] A. Jagannath and I. Tobasco. A dynamic programming approach to the Parisi functional. *Proc. Amer. Math. Soc.*, 144(7):3135–3150, 2016.

[60] A. Jagannath and I. Tobasco. Low temperature asymptotics of spherical mean field spin glasses. *Comm. Math. Phys.*, 352(3):979–1017, 2017.

[61] A. Jagannath and I. Tobasco. Some properties of the phase diagram for mixed p-spin glasses. *Probab. Theory Related Fields*, 167(3-4):615–672, 2017.

[62] A. Jagannath and I. Tobasco. Bounds on the complexity of replica symmetry breaking for spherical spin glasses. *Proc. Amer. Math. Soc.*, 146(7):3127–3142, 2018.

[63] J. Ko. The crisanti–sommers formula for spherical spin glasses with vector spins. *arXiv preprint arXiv:1911.04355*, 2019.

[64] J. Ko. Free energy of multiple systems of spherical spin glasses with constrained overlaps. *Electron. J. Probab.*, 25:34 pp., 2020.

[65] J. M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2013.

[66] M. Lelarge and L. Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields*, 173(3-4):859–929, 2019.

[67] T. Lesieur, F. Krzakala, and L. Zdeborová. Constrained low-rank matrix estimation: phase transitions, approximate message passing and applications. *J. Stat. Mech. Theory Exp.*, (7):073403, 86, 2017.

[68] M. Mézard and A. Montanari. *Information, physics, and computation*. Oxford Graduate Texts. Oxford University Press, Oxford, 2009.

[69] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro. Nature of the spin-glass phase. *Phys. Rev. Lett.*, 52:1156–1159, Mar 1984.

[70] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro. Replica symmetry breaking and the nature of the spin glass phase. *Journal de Physique*, 45(5):843–854, 1984.

[71] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin glass theory and beyond*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific Publishing Co., Inc., Teaneck, NJ, 1987.

[72] A. Montanari. Optimization of the sherrington-kirkpatrick hamiltonian, 2018.

[73] J.-C. Mourrat and D. Panchenko. Extending the parisi formula along a hamilton-jacobi equation. *Electron. J. Probab.*, 25:17 pp., 2020.

[74] D. Panchenko. Cavity method in the spherical SK model. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(4):1020–1047, 2009.

[75] D. Panchenko. The Parisi ultrametricity conjecture. *Ann. of Math. (2)*, 177(1):383–393, 2013.

[76] D. Panchenko. *The Sherrington-Kirkpatrick model*. Springer Monographs in Mathematics. Springer, New York, 2013.

[77] D. Panchenko. The Parisi formula for mixed p-spin models. *Ann. Probab.*, 42(3):946–958, 2014.

[78] D. Panchenko. The free energy in a multi-species Sherrington-Kirkpatrick model. *Ann. Probab.*, 43(6):3494–3513, 2015.

[79] D. Panchenko. Chaos in temperature in generic $2p$-spin models. *Comm. Math. Phys.*, 346(2):703–739, 2016.

[80] D. Panchenko. Free energy in the mixed p-spin models with vector spins. *Ann. Probab.*, 46(2):865–896, 2018.

[81] D. Panchenko. Free energy in the Potts spin glass. *Ann. Probab.*, 46(2):829–864, 2018.

[82] D. Panchenko. *Lecture Notes on Probability Theory*. 2019.

[83] D. Panchenko and M. Talagrand. On the overlap in the multiple spherical SK models. *Ann. Probab.*, 35(6):2321–2355, 2007.

[84] G. Parisi. An Infinite Number of Order Parameters for Spin Glasses. *Phys. Rev. Lett.*, 43:1754–1756, 1979.

[85] G. Parisi. A sequence of approximated solutions to the s-k model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115–L121, apr 1980.

[86] G. Parisi. Order parameter for spin-glasses. *Phys. Rev. Lett.*, 50:1946–1948, Jun 1983.

[87] K. B. Petersen and M. S. Pedersen. The matrix cookbook, nov 2012. Version 20121115.

[88] H. Poincaré. *Calcul des probabilités*. Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics]. Éditions Jacques Gabay, Sceaux, 1987. Reprint of the second (1912) edition.

[89] T. Rizzo. Ultrametricity between states at different temperatures in spin-glasses. *The European Physical Journal B-Condensed Matter and Complex Systems*, 29(3):425–435, 2002.

[90] D. Ruelle. A mathematical reformulation of Derrida's REM and GREM. *Comm. Math. Phys.*, 108(2):225–239, 1987.

[91] I. J. Schoenberg. Positive definite functions on spheres. *Duke Math. J.*, 9(1):96–108, 03 1942.

[92] S.Franz and G.Parisi. Recipes for metastable states in spin glasses. 1995.

[93] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Phys. Rev. Lett.*, 35:1792–1796, Dec 1975.

[94] E. Subag. The geometry of the Gibbs measure of pure spherical spin glasses. *Invent. Math.*, 210(1):135–209, 2017.

[95] E. Subag. Following the ground-states of full-rsb spherical spin glasses. *arXiv preprint arXiv:1812.04588*, 2018.

[96] E. Subag. Following the ground-states of full-rsb spherical spin glasses, 2018.

[97] E. Subag. Free energy landscapes in spherical spin glasses. *arXiv preprint arXiv:1804.10576*, 2018.

[98] M. Talagrand. Free energy of the spherical mean field model. *Probab. Theory Related Fields*, 134(3):339–382, 2006.

[99] M. Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.

[100] M. Talagrand. Parisi measures. *J. Funct. Anal.*, 231(2):269–286, 2006.

[101] M. Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010.

[102] M. Talagrand. *Mean field models for spin glasses. Volume II*, volume 55 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Heidelberg, 2011. Advanced replica-symmetry and low temperature.

[103] M. Talagrand. *Upper and lower bounds for stochastic processes*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Heidelberg, 2014. Modern methods and classical problems.

[104] D. J. Thouless, P. W. Anderson, and R. G. Palmer. Solution of 'solvable model of a spin glass'. *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics*, 35(3):593–601, 1977.

[105] F. L. Toninelli. About the almeida-thouless transition line in the sherrington-kirkpatrick mean-field spin glass model. *Europhysics Letters (EPL)*, 60(5):764–767, dec 2002.

[106] J. L. van Hemmen and R. G. Palmer. The replica method and solvable spin glass model. *Journal of Physics A: Mathematical and General*, 12(4):563–580, apr 1979.

[107] R. Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018. An introduction with applications in data science, With a foreword by Sara van de Geer.

[108] L. Zdeborov and F. Krzakala. Statistical physics of inference: thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.