Spherical Sherrington-Kirkpatrick models and the TAP approach

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Leon Fröber

2023

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Erstbetreuer: Prof. Dr. David Belius,
Zweitbetreuer: Prof. Dr. Jiří Černý und
externer Experte: Prof. Dr. Erwin Bolthausen

Basel, den 20. Juni 2023

Prof. Dr. Marcel Mayor, Dekan.

Contents

Acknowledgments

First, I want to thank my advisor David Belius for his guidance and helpful advice along the preparation of our research articles and this thesis. I also want to thank Jiří Černý for being my second advisor and Erwin Bolthausen for being the external expert. I thank Nicola Kistler for introducing me to the research of spin glasses, and more generally I would like to thank our working group including the former members for the pleasant atmosphere and all the interesting discussions we had, especially Marius Schmidt for always being enthusiastic about teaching me a lot about the field of spin glasses and its methods. I also want to thank Justin Ko, with whom it was a pleasure to work with. Furthermore, I want to thank Florin Boenkost, Jonas Peter and Elias Polak for making studying mathematics much more pleasant, and motivated me to proceed my studies even after my master's degree. I also want to express my gratitude for the colleagues and friends at the university of Basel, who were always helpful if I had mathematical questions outside my field. I am also grateful for all the support I received from my friends from Frankfurt am Main, my family in particular my parents, and especially Maren.

I am also grateful for the funding of the Schweizerische Nationalfonds (SNF), which without this project would not have been possible.

CHAPTER I

Introduction

Understanding the behavior of interacting particles is of great importance in physics and has motivated the investigation of *spin glass models*. They model the interactions of a large number N of particles $i = 1, \ldots, N$, where each particle has a *spin* σ. To each configuration $\sigma = (\sigma_1, ..., \sigma_N)$ of spins an energy level given by the so called *Hamiltonian* $\mathcal{H}_N(\sigma)$ is assigned. In the physically more realistic models such as the Ising model [Isi24,Bru67] the interacting particles are placed on the vertices of a lattice and the interaction strength decreases with their distance. Mean field models do not weigh the interaction strength based on the particles' locations, making them easier to study. These mean field models [KTJ76, CS92, CL04, Tal00, Tal06a] and related models [MPV87,KR98,MM09,DSS15,DS19] have furthermore attracted general interest within the fields of physics and mathematics as canonical examples of complex systems, as well as in the study of neural networks [She93,KR98].

One important mean field model is the *Sherrington-Kirkpatrick (SK) model* that was introduced in [SK75] as toy model of an exotic magnetic alloy. Another important model is its spherical variant, the *spherical Sherrington-Kirkpatrick (SSK) model*, that appeared in [KTJ76]. There exist several different approaches to analyze these models. We are particularly interested in the TAP method [TAP77], which has an elegant geometric interpretation (see Subsection 2.1), and is under active development of a stand-alone approach to general mean field spin glasses. In this thesis we analyze spherical 2-spin SSK models using a TAP method. Two specific cases will be studied, namely the spiked 2-spin spherical model (in Chapter II and III) and a multiple spin SSK model with constrained overlaps (in Chapter IV).

1. Preliminaries

1.1. Basic mean field models

Let us first introduce the basic terminology and fundamental questions of mean field spin glasses. Let Σ_N be a set of spin configurations (e.g. $\{-1,1\}^N$, $[-1,1]^N$ or $\{\sigma \in \mathbb{R}^N : \sum_{i=1}^N \sigma_i^2 = 1\}$), $\mathcal{H}_N : \Sigma_N \to \mathbb{R}$ be the Hamiltonian and $\beta > 0$ be the *inverse temperature*. Then define the *partition function* by

$$Z_N(\beta) = E\left[\exp(\beta \mathcal{H}_N(\sigma))\right] \tag{1.1}$$

where E denotes the expectation with respect to uniformly distributed $\sigma \in \Sigma_N$. Further define the *Gibbs measure* by

$$G_N(d\sigma) = \frac{\exp(\beta \mathcal{H}_N(\sigma))}{Z_N(\beta)} E[d\sigma],$$

which is a probability measure, its notion dating back to [LR69, Dob68]. It models the behavior of interacting spins, and represents the probability of each spin configuration $\sigma \in \Sigma_N$ in a given system at the temperature $\frac{1}{\beta}$ at equilibirum. Note that this measure has a bias towards configurations of higher energy $\mathcal{H}_N(\sigma)$. Configurations of highest energy are called ground states. Arguably the main goal of the study of spin glass models is to characterize the behavior of σ under this measure. (Note that in the physics literature a negative sign is placed in front of $\beta \mathcal{H}_N(\sigma)$ in (1.1). This is merely a notational difference, and reflects the physical convention that systems tend towards states of lower energy.)

An important feature of spin glass models is that the Hamiltonian is a random function. In the original (2-spin) SK model [SK75] the Hamiltonian was defined as

$$\mathcal{H}_N^{\mathrm{SK}}(\sigma) = \frac{1}{\sqrt{N}} \sum_{1 \le i < j \le N} g_{ij} \sigma_i \sigma_j, \tag{1.2}$$

where each spin $\sigma_i \in \{-1, +1\}$ and the g_{ij} are i.i.d. standard Gaussian random variables. One often also considers

$$\mathcal{H}_N^{\mathrm{SKe}}(\sigma) = \frac{1}{\sqrt{N}} \sum_{1 \le i < j \le N} g_{ij} \sigma_i \sigma_j + h \sum_{i=1}^N \sigma_i \tag{1.3}$$

in place of (1.2), for some $h \in \mathbb{R}$. The added term is an "external field" which favours $+1$ spins over -1 spins in the Gibbs measure (if $h > 0$).

An important step in studying spin glass models is the computation of the *free energy*, which is the exponential rate of growth of the partition function Z_N and given by

$$F_N(\beta) := \frac{1}{N} \log E\left[\exp\left(\beta \mathcal{H}_N(\sigma)\right)\right]. \tag{1.4}$$

In all of these models this converges to a deterministic limit as $N \to \infty$, despite the randomness of the Hamiltonian, see e.g. [T+03, Theorem 2.2.4. and Corollary 2.2.5.]. The free energy from (1.4) is also called the *quenched free energy* to distinguish it from another important quantity called the *annealed free energy*, given by

$$\frac{1}{N} \log \mathbb{E}\left[E\left[\exp\left(\beta \mathcal{H}_N(\sigma)\right)\right]\right], \tag{1.5}$$

where $\mathbb{E}$ denotes the expectation with respect to the randomness of the Hamiltonian. In all aforementioned models the quenched free energy concentrates and converges to a deterministic limit, which is asymptotically the same as

$$\frac{1}{N} \mathbb{E}\left[\log E\left[\exp\left(\beta \mathcal{H}_N(\sigma)\right)\right]\right]. \tag{1.6}$$

Note that the only difference between (1.5) and (1.6) is the position of the logarithm. By Jensen's inequality the annealed free energy is an upper bound for the quenched free energy. In some cases the upper bound is tight and they are asymptotically equal, which is convenient because the annealed free energy is much easier to compute.

The main results of this thesis are the computation of the fluctuations of the ground state in one model and the computation of the free energy in another.

1.2. Spherical models

An interesting modification of the SK model is to use a spherical configuration space, i.e. $\Sigma_N = \{\sigma \in \mathbb{R}^N : |\sigma| = 1\}$ where $|\cdot|$ denotes the Euclidian norm. This model is more accessible to explicit computations through random matrix arguments and was studied in [KTJ76] very shortly after the initial SK model was introduced, and continues to be studied [BL16, BCWLDW21]. A different modification of spin glass models in general is the addition of *spikes*. These are "non-linear external fields", and are one of the focal points of this thesis. Models with spike terms appear for example in statistical inference problems [RM14, LKZ17, LM19]. Models with a quadratic spike term were investigated for instance in [AMMN19]. Another important generalization of spin glass models are (mixed) p-spin models for $p \geq 2$ [Tal00, AA13, AAČ13, Sub17a, SZ17, CS17, JT17, BČNS22], which feature other types of spin interactions.

The spiked mixed p-spin SSK model is a generalization that combines all these modifications and is constructed as follows. The Hamiltonian is given by $\mathcal{H}_N(\sigma)$ which is a centered Gaussian process with covariance

$$E[\mathcal{H}_N(\sigma)\mathcal{H}_N(\sigma')] = N\xi\left(\tfrac{\sigma \cdot \sigma'}{N}\right) \tag{1.7}$$

for $\sigma, \sigma' \in \mathbb{R}^N$ with $|\sigma|, |\sigma'| < 1$ and a power series

$$\xi(x) = \sum_{p \geq 0} a_p x^p \text{ with } a_p \geq 0 \text{ and } \xi(1) < \infty. \tag{1.8}$$

The spiked Hamiltonian is defined as

$$\mathcal{H}_N^{\xi,f}(\sigma) = \mathcal{H}_N(\sigma) + f_N(\sigma \cdot u),$$

where $f_N : [-1,1] \to \mathbb{R}$ is the spike and $u \in \mathbb{R}^N$ with $|u| = 1$ is the direction of the spike. The model's free energy is given by

$$F_N^{\xi,f}(\beta) = \frac{1}{N} \log E\left[\exp\left(\beta \mathcal{H}_N^{\xi,f}(\sigma)\right)\right]. \tag{1.9}$$

While in this thesis we restrict ourselves to studying the 2-spin case $\xi(x) = x^2$, part of the motivation is the long-term goal of developing a TAP method that can handle the general model.

1.3. Multiple spin models

Another interesting adaptation is SSK models with constrained multiple spins [Pan18a, Ko19, AZ22]. They were introduced to study the so called *overlap distribution*, which is the distribution of the scalar product of two independently sampled configurations from the Gibbs measure, and are a focus of this thesis.

The Hamiltonian of a 2-spin SSK model with an external field given by the scalar product $h_N \cdot \sigma$, for some vector h_N, is defined by

$$\mathcal{H}_N^{\mathrm{SSKe}}(\sigma) = \frac{1}{\sqrt{N}} \sum_{1 \le i < j \le N} g_{ij}\sigma_i\sigma_j + h_N \cdot \sigma,$$

similarly to (1.3). (Note that the sum of the external field in (1.3) is written as a scalar product here. Due to the symmetry of the sphere any vector h_N of a fixed magnitude yields an equivalent model.) The Gibbs measure of this model is

$$G_N^{\beta,h}(d\sigma) = \frac{\exp(\beta\mathcal{H}_N^{\mathrm{SSKe}}(\sigma))d\sigma}{\int_{\mathcal{S}_{N-1}} \exp(\beta\mathcal{H}_N^{\mathrm{SSKe}}(\tau))d\tau},$$

where $d\sigma$ denotes the uniform measure on $\mathcal{S}_{N-1} = \{\sigma \in \mathbb{R}^N : |\sigma| = 1\}$. If n configurations with $(\sigma^1, ..., \sigma^n) \sim \bigotimes_{k=1}^n G_N^{\beta_k, h_N^{(k)}} =: G_N^n$ are independently sampled at possibly different inverse temperatures β_k and external field vectors $h_N^{(k)}$, then the probability that these overlaps $\sigma^k \cdot \sigma^\ell$ are within ε of $q_{k,\ell} \in [-1,1] \setminus \{0\}$ for all $k, \ell \in \{1, ..., n\}$ is

$$\begin{aligned} &G_N^n\left(\forall k, \ell \in \{1,...,n\} : \left|\sigma^k \cdot \sigma^\ell - q_{k,\ell}\right| < \varepsilon\right) \\ =& \frac{\int_{\mathcal{S}_{N-1}^n} \mathbb{1}_{\{|\sigma^k\cdot\sigma^\ell - q_{k,\ell}|_\infty < \varepsilon\}} \exp\left(\sum_{j=1}^n \beta\mathcal{H}_N(\sigma^{(j)})\right) d\sigma^{(1)} \dots d\sigma^{(n)}}{\int_{\mathcal{S}_{N-1}^n} \exp\left(\sum_{j=1}^n \beta\mathcal{H}_N(\sigma^{(j)})\right) d\sigma^{(1)} \dots d\sigma^{(n)}}. \end{aligned}$$

In matrix notation with $\boldsymbol{Q} = (q_{k,\ell})_{k,\ell}$ one then obtains with $|\cdot|_\infty$ denoting here the supremum norm

$$\frac{1}{N} \log G_N^n\left(\left|(\sigma^{(k)} \cdot \sigma^{(\ell)})_{k,\ell} - \boldsymbol{Q}\right|_\infty < \varepsilon\right) = F_N^\varepsilon(\beta, h, \boldsymbol{Q}) - \sum_{j=1}^n F_N(\beta^k, h^k),$$

where $F_N(\beta^k, h^k)$ denotes the free energy from (1.4) for different temperatures and external fields, and

$$F_N^\varepsilon(\beta, h, \boldsymbol{Q}) = \frac{1}{N} \log \int_{\mathcal{S}_{N-1}^n} \mathbb{1}_{\left\{\left|(\sigma^{(k)}\cdot\sigma^{(\ell)})_{k,\ell} - \boldsymbol{Q}\right|_\infty < \varepsilon\right\}} e^{\sum_{j=1}^n \beta\mathcal{H}_N(\sigma^{(j)})} d\sigma^{(1)} \dots d\sigma^{(n)} \tag{1.10}$$

defines a free energy of a multiple spin model with an overlap matrix $\boldsymbol{Q}$.

Computing the free energy (1.10) is of independent interest, and also a step in analyzing the Gibbs measure of the 2-spin SSK model. One of the results of this thesis is a computation of (1.10).

1.4. Methods to compute the free energy

A powerful but mathematically non-rigorous tool to calculate the limiting free energy is the *replica method* used by Parisi in [Par79, Par80]. Rigorous approaches to tackle this problem include the interpolation method of Guerra [Gue03] and the methods of Talagrand, Aizenman-Sims-Starr and Panchenko [ASS03, Tal06a, Tal06b, Con13, Pan14, Che13]. A very interesting and more geometric approach to solve the SK model was proposed by Thouless, Anderson and Palmer [TAP77], which we will call the *TAP approach*. This approach played a complementary role to other approaches in both physics [BM80, DDY83, GM84a, MPV87, KPV93, CS95, CGPM03] and mathematics [Cha10, Tal10, CP18, AJ19, CPS22], but has not been fully developed as a stand-alone solution of spin glass models yet.

Initial steps to develop such a stand-alone TAP theory were taken in [BK19, Bel22]. In [BK19] the limiting free energy of the 2-spin spherical SK model with (linear) external field was computed, and shown to be equal to the solution of a maximization problem involving the TAP free energy. In [Bel22] an upper bound involving the TAP free energy was shown for the free energy of spiked mixed p-spin models. This thesis is a contribution to this framework, which is described in more detail in Section 2.

Other frameworks for a mathematically rigorous TAP theory include that of Bolthausen [Bol14, Bol19, BY22], which relies on an iterative construction of TAP solutions, and that of Subag [Sub18, Sub21, CPS22], which uses properties of the limiting Gibbs measure in its analysis, and that

2. The TAP method

The TAP method aims to express the free energy as a maximization problem over a set of magnetization vectors m [TAP77], in the spherical case $\{m \in \mathbb{R}^n : |m| < 1\}$. Consider for instance the free energy of the 2-spin SSK model

$$F_N(\beta) = \frac{1}{N} \log E\left[\exp\left(\beta \mathcal{H}_N(\sigma) + hu \cdot \sigma\right)\right], \tag{2.1}$$

where $\mathcal{H}_N : \mathcal{S}_{N-1} \to \mathbb{R}$ is the standard 2-spin SSK Hamiltonian from (1.2), $h \in \mathbb{R}$ and u a unit vector. The TAP free energy of this model is

$$F_{\text{TAP}}(m) := \beta \mathcal{H}_N(m) + Nhu \cdot m + \frac{N}{2}\beta^2(1 - |m|^2)^2 + \frac{N}{2}\log(1 - |m|^2), \tag{2.2}$$

and [BK19] shows that

$$\left| F_N(\beta) - \sup_{\substack{|m|<1 \\ \beta(1-|m|^2) \leq \frac{1}{\sqrt{2}}}} \frac{1}{N} F_{\text{TAP}}(m) \right| \longrightarrow 0 \text{ in probability.} \tag{2.3}$$

The condition $\beta(1 - |m|^2) \leq \frac{1}{\sqrt{2}}$ in (2.3) that restricts the optimization space at low temperature is called the *Plefka condition* [Ple82a] and will be expained in more detail in Subsection 2.2. In Chapter II of this thesis we prove a generalization of (2.3) where the free energy has a spike term in place of the external field term.

More generally the TAP free energy of the spherical mixed p-spin model is

$$F_{\text{TAP}}^{\xi,h}(m) = \beta \mathcal{H}_N^{\xi}(m) + Nhu \cdot m + \frac{N}{2} \log\left(1 - |m|^2\right) + \frac{\beta^2}{2}\left(\xi(1) - \xi(|m|^2) - (1 - |m|^2)\xi'(|m|^2)\right)$$

and [Bel22, Theorem 1.2] proved that the free energy

$$F_N^{\xi,h}(\beta) = \frac{1}{N} \log E\left[\exp\left(\beta \mathcal{H}_N^{\xi}(\sigma) + Nhu \cdot m\right)\right]$$

is bounded from above by the supremum of the TAP free energy, i.e.

$$F_N^{\xi,h}(\beta) \leq \sup_{m \in B_N} \frac{1}{N} F_{\text{TAP}}^{\xi,h}(m) + o(1). \tag{2.4}$$

That article conjectures that the same optimization problem is also a lower bound at high temperatures, as is known to be true for the 2-spin case with a linear external field by (2.3). A future goal is to prove a matching lower bound for the free energy and also extend this result to low temperatures (with an appropriate Plefka condition). Here we will focus on spherical models with 2-spin interactions.

2.1. The slicing and annealing method

The main idea of the TAP approach of this thesis is to estimate the contribution to the partition function of certain subsets of the sphere $\mathcal{S}_{N-1}$, that can be thought of as "slices" around $m \in B_N$ (see the red circle in Figure I.1), by $\exp(F_{\text{TAP}}(m))$.

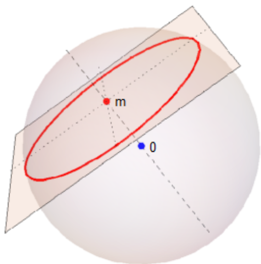

Figure I.1: Slice of the sphere around m.

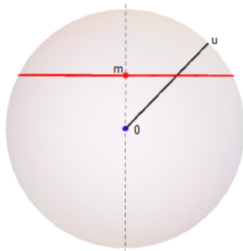

Figure I.2: Sideview with external field vector u.

Similarly to when one uses the Laplace method, it turns out that the leading order contribution comes from m that maximize $F_{\text{TAP}}(m)$. Furthermore one expects the slices of maximal TAP free energy to be the regions of the spin space charged by the Gibbs measure. The clear geometric intuition behind this analysis of the partition function is one of the advantages of the TAP approach.

Note that one can show for β small and no external field that the quenched and annealed free energies are asymptotically equal, i.e.

$$\frac{1}{N}\log E\left[\exp\left(\beta H_N(m)\right)\right] \approx \frac{1}{N}\log \mathbb{E}\left[E\left[\exp\left(\beta \mathcal{H}_N(\sigma)\right)\right]\right] \approx \frac{\beta^2}{2} \quad \text{if } \beta \le \frac{1}{\sqrt{2}} \tag{2.5}$$

[KTJ76], [T+03, Section 2.2]. The strategy of the TAP method is to use a "recentering" and "slicing" to find certain subsets around the magnetizations where the effective external field almost vanishes. If the effective temperature on that subset is high enough one can then apply a "quenched=annealed" approximation to compute the contribution of the partition function restricted to that subset.

We will briefly illustrate this for the 2-spin SSK model by sketching a proof for the lower bound on $F_N(\beta)$ of (2.3). In the 2-spin SSK model one can use the identity

$$H_N(\sigma) = H_N(m) + \nabla H_N(\sigma - m)\cdot\sigma + H_N(\sigma - m) \quad \text{for all } \sigma, m \in \mathbb{R}^N$$

which "recenters" the Hamiltonian, to write the partition function as

$$Z_N = \exp\left(\beta H_N(m) + Nhu\cdot m\right) E\left[\exp\left(\beta H_N(\sigma - m) + Nh^m\cdot(\sigma - m)\right)\right] \tag{2.6}$$

where

$$h^m = \frac{\beta}{N}\nabla H_N(m) + hu$$

is the "effective" external field. For a lower bound we can construct a slice by simply inserting an indicator function into (2.6). Defining $A_\varepsilon(m)$ as the subset of the sphere where the scalar product of $\sigma - m$ with any vectors in $\operatorname{span}\{m, h^m\}$ is bounded by $\varepsilon > 0$, one can verify that

$$h^m\cdot(\sigma - m) = \mathcal{O}(\varepsilon) \quad \text{and} \quad |\sigma - m|^2 \approx 1 - |m|^2$$

for all $\sigma \in A_\varepsilon(m)$. We can then write

$$Z_N \geq \exp\left(\beta H_N(m) + Nhu \cdot m + \mathcal{O}(\varepsilon N)\right) E\left[\mathbb{1}_{A_\varepsilon(m)} \exp\left(\beta H_N(\sigma - m)\right)\right]. \tag{2.7}$$

By rescaling the argument of the Hamiltonian using the change of variables $\hat{\sigma} = \frac{\sigma - m}{|\sigma - m|}$ one obtains that the expectation in (2.7) is approximately equal to

$$E\left[\mathbb{1}_{A_\varepsilon(m)}\right] E_{\mathrm{span}\{m,h^m\}^\perp}\left[\exp\left(\beta(1-|m|^2)H_N(\hat{\sigma})\right)\right] \tag{2.8}$$

where $E_{\mathrm{span}\{m,h^m\}^\perp}$ denotes the expectation with respect to $\hat{\sigma}$ uniformly distributed on the sphere restricted to $\mathrm{span}\{m, h^m\}^\perp$. The first expectation is a volume term and is approximately equal to $\exp(\frac{N}{2}\log(1-|m|^2))$. The second expectation is approximately the partition function on a $N-2$ dimensional unit sphere with effective inverse temperature $\beta_m := \beta(1-|m|^2)$ and no external field. The Plefka condition $\beta(1-|m|^2) \leq \frac{1}{\sqrt{2}}$ is precisely the high temperature condition $\beta_m \leq \frac{1}{\sqrt{2}}$ from (2.5) for this partition function, and if it is fulfilled we can use (2.5) to obtain

$$Z_N \geq \exp\Big(\underbrace{\beta H_N(m) + Nhu \cdot m + \tfrac{N}{2}\log(1-|m|^2) + \tfrac{N}{2}\beta^2(1-|m|^2)^2}_{F_{\mathrm{TAP}}(m)} + \mathcal{O}(\varepsilon N)\Big). \tag{2.9}$$

Since this applies for all m that satisfy Plefka's condition one obtains

$$Z_N \geq \sup_{|m|<1, \beta(1-|m|^2) \leq \frac{1}{\sqrt{2}}} \exp\left(F_{\mathrm{TAP}}(m)\right).$$

The argument for the upper bound is more intricate since one must show that regions that do not satisfy Plefka's condition can be ignored.
In Chapter II it will be shown that this method extends to general spike terms, and in Chapter IV to multiple spin models.

2.2. The Plefka condition

To further illustrate the importance of Plefka's condition let us consider m's for which it is not satisfied. In the spherical 2-spin case these m are

$$|m|^2 < 1 - \frac{1}{\sqrt{2}\beta}.$$

Note that the contribution to the partition function of the slice $A_\varepsilon(m)$ around $m \in B_N$ is given by the r.h.s. of (2.7). Using also the approximation (2.8) its log equals

$$\beta \mathcal{H}_N(m) + Nhu \cdot m + \log E\left[\mathbb{1}_{A_\varepsilon(m)}\right] + \log E_{\mathrm{span}\{m,h^m\}^\perp}\left[\exp\left(\beta H_N(\sigma - m)\right)\right] + \mathcal{O}(N\varepsilon). \tag{2.10}$$

We argued that this equals $F_{\mathrm{TAP}}(m)$ if m satisfies the Plefka condition, but this is not necessarily true in general. The Plefka condition appeared when applying the "annealed=quenched" approximation (2.5) to the term $E_{\mathrm{span}\{m,h^m\}^\perp}\left[\exp\left(\beta H_N(\sigma - m)\right)\right]$ to obtain the Onsager term $\frac{N}{2}\beta^2(1-|m|^2)^2$. As mentioned in Subsection 1.1 the annealed free energy is only an upper bound for the quenched free energy, so the TAP free energy for m not satisfying Plefka's condition might overestimate (2.10). A special property of the spherical 2-spin model is that it is possible to obtain an explicit formula for (2.10) even when the Plefka condition is not satisfied (see [KTJ76]), namely

$$\begin{aligned}\tilde{F}_{\mathrm{TAP}}(m) :=& \beta \mathcal{H}_N(m) + Nhu \cdot m + \frac{N}{2}\log(1-|m|^2) \\ &+ \frac{N}{2}\begin{cases} \sqrt{2}\beta\left(1-|m|^2\right) - \frac{1}{2}\log\left(\sqrt{2}\beta\left(1-|m|^2\right)\right) - \frac{3}{4}, & \text{for } \beta(1-|m|^2) > \frac{1}{\sqrt{2}}, \\ \frac{\beta^2}{2}\left(1-|m|^2\right)^2, & \text{for } \beta(1-|m|^2) \leq \frac{1}{\sqrt{2}}. \end{cases}\end{aligned}$$

Furthermore one can obtain formulas for $\frac{1}{N}\sup_{m:|m|^2=q} F_{\text{TAP}}(m)$ and $\frac{1}{N}\sup_{m:|m|^2=q} \tilde{F}_{\text{TAP}}(m)$ (for the former see (2.13) in the next subsection, the latter is derived similarly).

Figure I.3 shows $\frac{1}{N}\sup_{m:|m|^2=q} F_{\text{TAP}}(m)$ (orange) and $\frac{1}{N}\sup_{m:|m|^2=q} \tilde{F}_{\text{TAP}}(m)$ (blue) as functions of $q \in [0,1)$. The region where Plefka's condition is satisfied is to the right of the red vertical line.

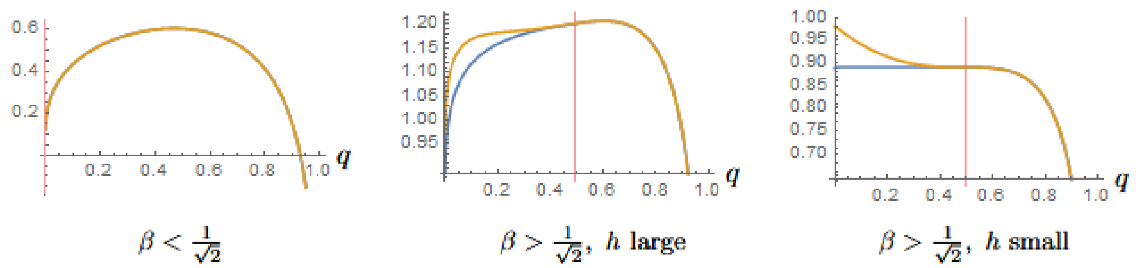

Figure I.3: Comparison F_{TAP} (orange) to $\tilde{F}_{\text{TAP}}(m)$ (blue)

We see that both functions coincide in the Plefka region, while $F_{\text{TAP}}(m)$ overestimates $\tilde{F}_{\text{TAP}}(m)$ in the non-Plefka region, and thus overestimates the contribution to the partition function of the slice $A_\varepsilon(m)$.

In Chapter IV we will derive the Plefka condition for the 2-spin spherical multiple spin model. At the moment it is however not clear what the correct Plefka conditions are in general. In (pure) p-spin models a possible condition was suggested in [BS22a, (1.6) and Lemma 2.1].

2.3. Computing the maximal TAP free energy

Once an estimate of the type (2.3) has been obtained, one needs to compute the maximal TAP free energy. In the 2-spin case the maximal TAP free energy can be written as

$$\sup_{q\in[0,1],\beta(1-q)\leq\frac{1}{\sqrt{2}}} \left\{ \sup_{|\sigma|=1} \{\beta q H_N(\sigma) + N f(\sqrt{q}u\cdot\sigma)\} + \frac{N}{2}\beta^2(1-q)^2 + \frac{N}{2}\log(1-q) \right\} \tag{2.11}$$

using the change of variables $m = \sqrt{q}\sigma$ ($|\sigma| = 1$) and using that $H_N(x\sigma) = x^2 H_N(\sigma)$ for the 2-spin Hamiltonian (1.2). By defining a new inverse temperature $\tilde{\beta} = \beta q$ and a new spike term $\tilde{f}(\cdot) = f(\sqrt{q}\cdot)$ one obtains that the TAP free energy is a maximization problem over a ground state term plus a term that only depends on the magnitude of m. Thus, if we can compute the ground state for each $q \in [0,1]$, the N-dimensional TAP variational formula turns into a low dimensional variational formula.

In the case of a 2-spin SSK model with linear external field ($f(u\cdot\sigma) = hu\cdot\sigma$ for some $h \in \mathbb{R}$) it was shown in [BK19] that the ground state equals

$$N\sqrt{2\beta^2 + h^2} \tag{2.12}$$

to leading order, which was then used to conclude by the logic above that the maximal TAP free energy is equals

$$\sup_{q\in[0,1],\beta(1-q)\leq\frac{1}{\sqrt{2}}} N\left\{ \sqrt{2\beta^2 q^2 + h^2 q} + \frac{1}{2}\beta^2(1-q)^2 + \frac{1}{2}\log(1-q) \right\} \tag{2.13}$$

to leading order. One of the results of Chapter III of this thesis is an extension of this to non-linear spikes.

The degree of precision of the results presented so far is that which in principle allows for a rough description of the Gibbs measure G_N. To obtain more precise information about the Gibbs measure finer results are needed. In particular for the 2-spin model a more precise version of (2.3), and a more precise version of (2.13) would be needed, going beyond the leading order. In this thesis we take a first step in this direction by studying the fluctuations of (2.13), including for non-linear spikes.

3. Results

3.1. Spiked SSK

In Chapter II we will generalize (2.3) by showing that the TAP variational formula is indeed the limit of the free energy of the 2-spin SSK model with non-linear spikes, i.e.

$$\left| F_N^f(\beta) - \sup_{\substack{m\in B_N \\ \beta(1-|m|^2)\le \frac{1}{\sqrt{2}}}} \frac{1}{N} F_{\mathrm{TAP}}(m) \right| \longrightarrow 0 \text{ in probability.} \tag{3.1}$$

In Chapter III we will solve the variational problem for the spiked 2-spin SSK model and give a formula for the supremum in (3.1). Theorem III.1.1 computes the ground state, while Theorem III.1.2 (a) gives a formula like (2.13) for spiked 2-spin models. Let

$$\mathcal{L}(r,\alpha) := f(r\alpha) + \sqrt{2}\beta r^2\sqrt{1-\alpha^2} + \frac{1}{2}\beta^2(1-r^2)^2 + \frac{1}{2}\log(1-r^2)$$

and $p = \min\{0, 1-\frac{1}{\sqrt{2}\beta}\}$ then more precisely the first part of the second theorem implies that

$$\sup_{|m|<1, \beta(1-|m|^2)\le \frac{1}{\sqrt{2}}} \frac{1}{N} F_{\mathrm{TAP}}(m) \longrightarrow \sup_{r\in[\sqrt{p},1], \alpha\in(-1,1)} \mathcal{L}(r,\alpha). \tag{3.2}$$

Furthermore, Chapter III computes the fluctuations of the limiting TAP free energy (3.2). Theorem III.1.2 (b) implies that

$$\sup_{|m|<1, \beta(1-|m|^2)\le \frac{1}{\sqrt{2}}} \frac{1}{N} F_{\mathrm{TAP}}(m) = \sup_{r\in[\sqrt{p},1], \alpha\in(-1,1)} \mathcal{L}(r,\alpha) + \frac{1}{\sqrt{N}}\mathcal{X}_N + \frac{1}{N}\mathcal{Y}_N + o\left(\frac{1}{N}\right),$$

where $\mathcal{X}_N$ converges to a Gaussian and $\mathcal{Y}_N$ converges to a quadratic function of three Gaussians.

3.2. Multiple spin SSK

In Chapter IV we will show that a similar TAP variational formula is also the limit of the free energy of a multiple spin SSK model with linear external field. Recall the definitions from Subsection 1.3. We introduce the TAP free energy

$$F_{\mathrm{TAP}}(\boldsymbol{m}) = \frac{N}{2}\log|\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T}| + \sum_{k=1}^n \beta_k H_N(m^k) + N\sum_{k=1}^n h^k\cdot m^k + \frac{N}{2}\beta^\mathsf{T}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T})^{\odot 2}\beta$$

for this model, where $\boldsymbol{m} = (m^1,\dots,m^n)\in\mathbb{R}^{n\times N}$ are magnetization vectors, $|\cdot|$ denotes the determinant, and $\boldsymbol{A}^{\odot 2} = \boldsymbol{A}\odot\boldsymbol{A} = (A_{k,l}^2)_{k,l=1,\dots,n}$ denotes the Hadamard square of the entries of $\boldsymbol{A}$. We further introduce a Plefka condition for the vector spin model given by $\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta)$ for

$$\mathrm{Plef}_N(\boldsymbol{Q},\beta) \quad = \quad \left\{\boldsymbol{m}\in\mathbb{R}^{n\times N} : \boldsymbol{0}\le\boldsymbol{m}\boldsymbol{m}^\mathsf{T}<\boldsymbol{Q},\ \|\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T})\boldsymbol{\beta}^{\frac{1}{2}}\|_2\le\tfrac{1}{\sqrt{2}}\right\}$$

where $\boldsymbol{\beta} = \mathrm{diag}(\beta)\in\mathbb{R}^{n\times n}$, $\|\cdot\|_2$ denotes the spectral norm and $\le$ is the Loewner partial order on matrices (so that $\boldsymbol{A}\ge 0$ for $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ means that $\boldsymbol{A}$ is positive semi-definite).

Theorem IV.1.1 states that for any positive definite $\boldsymbol{Q} \in [-1,1]^{n\times n}$ with $Q_{k,k} = 1$ for $k = 1,\ldots,n$ it holds that

$$\lim_{\varepsilon\to 0} \limsup_{N\to\infty} |F_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) - \sup_{\boldsymbol{m}\in \mathrm{Plef}_N(\boldsymbol{Q},\beta)} \frac{1}{N} F_{\mathrm{TAP}}(\boldsymbol{m})| = 0,$$

where the limits are in probability.

Additionally, in Theorem IV.1.2 we present a formula for the ground state energy when $h^1,\ldots,h^n$ are multiples of a single vector. Defining for positive definite $\boldsymbol{A} \in \mathbb{R}^{n\times n}$

$$\mathrm{GSE}(\beta, h, \boldsymbol{A}) = \sqrt{2}\mathrm{Tr}\left(\sqrt{\left(\frac{1}{2}hh^{\mathsf{T}} + \boldsymbol{\beta}\boldsymbol{A}\boldsymbol{\beta}\right)^{\frac{1}{2}} \boldsymbol{A} \left(\frac{1}{2}hh^{\mathsf{T}} + \boldsymbol{\beta}\boldsymbol{A}\boldsymbol{\beta}\right)^{\frac{1}{2}}}\right),$$

the theorem implies that

$$\lim_{N\to\infty} \sup_{\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta)} \frac{1}{N} F_{\mathrm{TAP}}(\boldsymbol{m}) = \sup_{\boldsymbol{A}\in\mathrm{Plef}_n(\boldsymbol{Q},\beta)} \left(\mathrm{GSE}(\beta, h, \boldsymbol{A}) + \frac{1}{2}\log|\boldsymbol{Q} - \boldsymbol{A}| + \frac{1}{2}\beta^{\mathsf{T}}(\boldsymbol{Q} - \boldsymbol{A})^{\odot 2}\beta\right)$$

where the limits are in probability.

We will define all notation relevant to a chapter at its beginning. Some notation will differ between chapters.

CHAPTER II

Free energy of the Spherical SK model

1. Introduction

In this chapter we will extend the result from [BK19] to spiked 2-spin SSK models. Let $\mathcal{S}_{N-1}$ denote the surface of an $N-1$-dimensional unit sphere in $\mathbb{R}^N$ and $J \in \mathbb{R}^{N\times N}$ a symmetric GOE disorder matrix with

$$\operatorname{Var}(J_{ij}) = \begin{cases} 1, & \text{if } i \neq j, \\ 2, & \text{if } i = j. \end{cases}$$

Then for $\sigma \in \mathbb{R}^N$ we define the (2-spin) SK Hamiltonian

$$H_N(\sigma) := \sqrt{N}\sigma^T J\sigma = \sqrt{N}\sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j,$$

and the partition function

$$Z_N(\beta) := E\left[\exp\left(\beta H_N(\sigma) + Nf\left(\sigma\cdot u\right)\right)\right]$$

for $\beta \geq 0$, where $f \in C^2([-1,1])$ is the spike/external field function, $u \in \mathcal{S}_{N-1}$ the external field direction and E denotes the uniform measure on $\mathcal{S}_{N-1}$. Furthermore we define the free energy

$$F_N(\beta) = \frac{1}{N}\log Z_N(\beta)$$

and the TAP free energy

$$F_{\text{TAP}}(m) := \beta H_N(m) + Nf(u\cdot m) + \frac{N}{2}\beta^2(1-|m|^2)^2 + \frac{N}{2}\log(1-|m|^2)$$

for any $m \in B_N := \{x \in \mathbb{R} : |x| < 1\}$. Let the Plefka region be

$$\mathcal{M}_\beta := \left\{ m \in \mathbb{R}^N : \ |m| < 1, \ \beta(1-|m|^2) \leq \frac{1}{\sqrt{2}} \right\}.$$

The goal of this chapter is to prove the following.

Theorem 1.1. *It holds that*

$$\left| F_N(\beta) - \sup_{\substack{m\in B_N \\ \beta(1-|m|^2)\leq \frac{1}{\sqrt{2}}}} \frac{1}{N}F_{TAP}(m) \right| \longrightarrow 0 \quad \textit{in probability.}$$

We will follow the steps in the proof in [BK19] and make adaptations where necessary. In particular we will first prove a lower bound and then an upper bound.

2. Lower Bound

The goal of this section is to prove the lower bound of Theorem 1.1. We first adapt a useful lemma from [BK19].

Lemma 2.1. *Let $v_1, v_2, v_3 \in \mathbb{R}^N$ and $\langle v_1, v_2, v_3\rangle^\perp$ be the orthogonal complement of the span of v_1, v_2, v_3. It holds that*

$$\sup_{\beta\in[0,\frac{1}{\sqrt{2}}], v_1,v_2,v_3\in\mathbb{R}^N} \left| \frac{1}{N}\log E_{\langle v_1,v_2,v_3\rangle^\perp}\left[\exp\left(\beta H_N(\sigma)\right)\right] - \frac{\beta^2}{2} \right| \to 0,$$

where $E_{\langle v_1,v_2,v_3\rangle^\perp}$ denotes the expectation with respect to σ uniformly distributed on the unit sphere intersected with $\langle v_1, v_2, v_3\rangle^\perp$.

Proof. Note that [BK19, Lemma 5] shows a similar statement for two vectors $v_1, v_2 \in \mathbb{R}^N$ instead of three. However the argument of the proof easily extends to three linearly independent vectors by defining a orthonormal basis $w_1, ..., w_N$ such that $\langle w_{N-2}, w_{N-1}, w_N \rangle = \langle v_1, v_2, v_3 \rangle$ and then comparing $\sum_{i=1}^{N-3} b_i \sigma_i^2$ and $\sum_{i=1}^{N-3} a_i \sigma_i^2$, where the b_i are the eigenvalues of the disorder matrix J, and the a_i are the eigenvalues of the top left $(N-3) \times (N-3)$ minor of J when written in basis $w_1, ..., w_N$. □

Now we are ready to prove the lower bound of Theorem 1.1.

Proposition 2.2. *Let $f \in C^1([-1,1])$ and $\varepsilon > 0$. Then*

$$\mathbb{P}\left(F_N(\beta) \geq \sup_{\substack{m \in B_N \\ \beta(1-|m|^2) \leq \frac{1}{\sqrt{2}}}} \frac{1}{N} F_{TAP}(m) - \varepsilon \right) \longrightarrow 1$$

as $N \to \infty$.

Proof. Let m be an arbitrary vector in the N-dimensional unit sphere. By writing $\tilde{\sigma} = \sigma - m$ we recenter the spins around m, and recenter the Hamiltonian via the identity

$$\beta H_N(\sigma) + N f(u \cdot \sigma) = \beta H_N(m) + N f(u \cdot m) + \beta H_N(\tilde{\sigma}) + \xi(\sigma, m),$$

valid for all σ, m, where

$$\xi(\sigma, m) = \beta \nabla H_N(m) \cdot \tilde{\sigma} + N\left(f(u \cdot \sigma) - f(u \cdot m)\right)$$

denotes the effective external field. Let v_1, v_2, v_3 be orthonormal basis vectors of a 3-dimensional linear subspace of $\mathbb{R}^N$ that contains m, u and $\nabla H_N(m)$, and define for $\varepsilon > 0$

$$A_\varepsilon(m) = \{\sigma : |\tilde{\sigma} \cdot v_i| < \varepsilon, \text{ for } i = 1, 2, 3\}.$$

We have for $\sigma \in A_\varepsilon(m)$

$$|\tilde{\sigma} \cdot u| < \sqrt{3}\varepsilon \quad \text{ and if } m \neq 0: \ |\tilde{\sigma} \cdot \tfrac{m}{|m|}| < \sqrt{3}\varepsilon, \tag{2.1}$$

and since $|\frac{1}{N} \nabla H_N(m)|$ is bounded by some $c > 0$ with probability tending to 1 (see [BK19, (2.6)]) one obtains

$$|\tilde{\sigma} \cdot \tfrac{1}{N} \nabla H_N(m)| < c\varepsilon \quad \text{ with probability tending to 1.} \tag{2.2}$$

Furthermore it holds for $\sigma \in A_\varepsilon(m)$ that

$$|\tilde{\sigma}|^2 = |\sigma|^2 - |m|^2 - 2\tilde{\sigma} \cdot m = 1 - |m|^2 + \mathcal{O}(\varepsilon). \tag{2.3}$$

By restricting the partition function integral to $A_\varepsilon(m)$ we obtain

$$\begin{aligned} Z_N(\beta) &\geq E\left[\mathbb{1}_{A_\varepsilon(m)} \exp\left(\beta H_N(\sigma) + N f(u \cdot \sigma)\right)\right] \\ &= \exp\left(\beta H_N(m) + N f(u \cdot m)\right) E\left[\mathbb{1}_{A_\varepsilon(m)} \exp\left(\beta H_N(\tilde{\sigma}) + \xi(\sigma, m)\right)\right], \end{aligned} \tag{2.4}$$

where we will now show that $\xi(\sigma, m) = \mathcal{O}(N\varepsilon)$. First note that $f \in C^1([-1,1])$, so we have that

$$|f(u \cdot \sigma) - f(u \cdot m)| \leq |f'|_\infty |u \cdot \tilde{\sigma}| \overset{(2.1)}{<} |f'|_\infty \varepsilon \tag{2.5}$$

for any $\sigma \in A_\varepsilon(m)$. Using this and (2.2) gives us that $\xi(\sigma, m) = \mathcal{O}(\varepsilon N)$ and therefore it follows from (2.4) that

$$Z_N(\beta) \geq \exp\left(\beta H_N(m) + N f(u \cdot m)\right) E\left[\mathbb{1}_{A_\varepsilon(m)} \exp\left(\beta H_N(\tilde{\sigma})\right)\right] e^{\mathcal{O}(\varepsilon N)}. \tag{2.6}$$

Let $\gamma\sigma^\perp$ be the projection of $\tilde{\sigma}$ onto the hyperplane $\langle v_1, v_2, v_3\rangle^\perp$, where $\sigma^\perp$ is a unit vector and $\gamma \geq 0$ is the magnitude of the projection. Since we have for some $c > 0$ and all $\sigma \in A_\varepsilon(m)$ that

$$|\tilde{\sigma} - \gamma\sigma^\perp| \leq c\varepsilon, \tag{2.7}$$

we obtain by (2.3) that

$$\gamma^2 = 1 - |m|^2 + \mathcal{O}(\varepsilon). \tag{2.8}$$

Since the absolute values of the eigenvalues of $\frac{1}{\sqrt{N}}J$ are bounded by $\sqrt{2} + \varepsilon$ with probability tending to 1 (see [EYY12, Theorem 2.2]), we also have by (2.7) that

$$H_N(\tilde{\sigma}) = H_N(\gamma\sigma^\perp) + \mathcal{O}(\varepsilon N),$$

and furthermore by (2.8) that

$$H_N(\gamma\sigma^\perp) = \gamma^2 H_N(\sigma^\perp) = (1 - |m|^2) H_N(\sigma^\perp) + \mathcal{O}(\varepsilon N).$$

This gives us that the r.h.s. of (2.6) is equal to

$$\exp\left(\beta H_N(m) + N f(u \cdot m)\right) E\left[\mathbb{1}_{A_\varepsilon(m)} \exp\left(\beta(1 - |m|^2) H_N(\sigma^\perp)\right)\right] e^{\mathcal{O}(\varepsilon N)} \tag{2.9}$$

Because $\sigma^\perp$ is independent of $\sigma \cdot m$ and $\sigma \cdot u$ under E, and is uniform on the unit sphere intersected with $\langle v_1, v_2, v_3\rangle^\perp$, we obtain that (2.9) is equal to

$$\exp\left(\beta H_N(m) + N f(u \cdot m) + \mathcal{O}(\varepsilon N)\right) E[\mathbb{1}_{A_\varepsilon(m)}] E_{\langle v_1, v_2, v_3\rangle^\perp}\left[\exp\left(\beta(1 - |m|^2) H_N(\sigma)\right)\right],$$

where $E_{\langle v_1, v_2, v_3\rangle^\perp}$ denotes the expectation with respect to uniformly distributed σ on the unit sphere intersected with $\langle v_1, v_2, v_3\rangle^\perp$. By [BK19, (2.9)] the density of $\gamma\sigma^\perp$ with respect to the Lebesgue measure on $\langle v_1, v_2, v_3\rangle^\perp$ is

$$\frac{\Gamma\left(\frac{N}{2}\right)}{\pi\Gamma\left(\frac{N-3}{2}\right)} (1 - |\sigma|^2)^{\frac{N-5}{2}} d\sigma \;=\; \frac{N-3}{2\pi} (1 - |\sigma|^2)^{\frac{N-5}{2}} d\sigma$$

and together with (2.8), we obtain that for some $c > 0$

$$E[\mathbb{1}_{A_\varepsilon(m)}] \geq N c \varepsilon^2 (1 - |m|^2 - c\varepsilon)^{\frac{N-5}{2}}$$

for all m with $|m|^2 < 1 - \delta$ with $\delta \geq c\varepsilon$. By setting e.g. $\varepsilon = \frac{1}{\sqrt{N}}$ this equals

$$\exp\left(\frac{N}{2} \log\left(1 - |m|^2\right) + o(N)\right),$$

and thus $Z_N(\beta)$ is at least

$$\exp\left(\beta H_N(m) + N f(u \cdot m) + \frac{N}{2} \log\left(1 - |m|^2\right) + o(N)\right) E_{\langle v_1, v_2, v_3\rangle^\perp}\left[\exp\left(\beta(1 - |m|^2) H_N(\sigma)\right)\right],$$

for any m with $|m|^2 < 1 - \delta$, where the error term is $o(N)$ uniformly in m. Using Lemma 2.1 one obtains that $Z_N(\beta)$ must be at least

$$\exp\left(\beta H_N(m) + N f(u \cdot m) + \frac{N}{2} \log\left(1 - |m|^2\right) + N \frac{\beta^2}{2} (1 - |m|^2)^2 + o(N)\right), \tag{2.10}$$

provided

$$\beta(1 - |m|^2) \leq \frac{1}{\sqrt{2}}, \tag{2.11}$$

where the error term is $o(N)$ almost surely, uniformly in m that satisfy (2.11) and $|m|^2 < 1 - \delta$. Note that for ε and δ small enough the TAP free energy is smaller than cN for any given constant c for all m with $|m|^2 > 1 - \delta$ with probability tending to 1. Thus these m will not affect the supremum of (2.10) over all m satisfying (2.11). $\square$

3. Upper Bound

We will now prove the upper bound using a few tools from [BK19]. In [BK19, Section 4] an upper bound in the case of a linear external field is proved. First an upper bound for a coarse-grained Hamiltonian without external field is shown, and then later a recentering method is used to account for the external field. We adapt this approach to a non-linear spike.

First we will define an approximation of the Hamiltonian by binning similarly sized eigenvalues together. Let us define for the $K \geq 2$ equally spaced numbers $x_1, ..., x_K$ in $[-\sqrt{2}, \sqrt{2}]$ given by

$$-\sqrt{2} = x_1 < x_2 < ... < x_K < x_K = \sqrt{2} - \frac{2\sqrt{2}}{K} \quad \text{and} \quad x_{k+1} - x_k = \frac{2\sqrt{2}}{K},$$

and let $\theta_{1/N} < ... < \theta_{N/N}$ be the typical positions of the normalized GOE eigenvalues $\frac{1}{\sqrt{N}}\lambda_1 < ... < \frac{1}{\sqrt{N}}\lambda_N$ given by

$$\theta_s = \inf\left\{\theta : \int_{-\sqrt{2}}^{\theta} \frac{\sqrt{2-x^2}}{\pi} dx\right\} \quad \text{for } s \in [0,1].$$

We define a partition $I_1, ..., I_K$ of $\{1, ..., N\}$ by

$$I_k = \{i : x_k \leq \theta_{i/N} < x_{k+1}\},\ k = 1, ..., K-1 \quad \text{and} \quad I_K = \{i : x_K \leq \theta_{i/N}\},$$

collecting the indices of eigenvalues with roughly the same position, and the relative sizes of these bins

$$\mu_k = \frac{|I_k|}{N}.$$

Let us define

$$\mathcal{F}_K(\beta) = \beta\Lambda_K(\beta) - \frac{1}{2} - \frac{1}{2}\log(2\beta) - \frac{1}{2}h_K(\Lambda_K(\beta)),$$

where

$$h_K(\Lambda) = \sum_{k=1}^{K} \mu_k \log(\Lambda - x_k),$$

and for $\beta > 0$

$$\Lambda_K(\beta) \text{ is the unique solution of } h_K'(\Lambda) = 2\beta, \quad \Lambda \in (x_K, \infty). \tag{3.1}$$

With this setup we can use [BK19, Lemma 8 - Lemma 18] without changes.

Note that via diagonalization we have

$$F_N(\beta) = E\left[\exp\left(\beta\sum_{i=1}^{N}\lambda_i\sigma_i^2 + Nf(\sigma\cdot\tilde{u})\right)\right],$$

where $\tilde{u}$ is the external field vector u in diagonalized basis. Let $\tilde{H}_N(\sigma) = N\beta\sum_{i=1}^{N}\theta_{i/N}\sigma_i^2$ be a deterministic version of $H_N(\sigma)$ and define the deterministic free energy by

$$\tilde{F}_N^f(\beta, u) := \frac{1}{N}\log E\left[\exp\left(\tilde{H}_N(\sigma) + Nf(\sigma\cdot\tilde{u})\right)\right].$$

Note that the eigenvalues $\theta_N^1 \leq \ldots \leq \theta_N^N$ of $\frac{1}{\sqrt{N}}J$ satisfy

$$\theta_N^i = \theta_{i/N} + o(1),$$

where the $o(1)$ terms tend to zero in probability uniformly in i (see e.g. [BGK16, Theorem 2.9]). As a consequence we also have that

$$\lim_{N\to\infty}\frac{1}{N}\sup_{\sigma:|\sigma|=1}\left|\tilde{H}_N(\sigma) - \sum_{i=1}^{N}\lambda_i\sigma_i^2\right| = 0, \quad \mathbb{P}\text{-a.s.}, \tag{3.2}$$

so it suffices to show the upper bound for $\tilde{F}^f_N(\beta,u)$. Let us also define a modified TAP free energy by

$$\frac{1}{N}\tilde{F}^K_{\mathrm{TAP}}(m) = \beta\tilde{H}_N(\sigma) + f(m\cdot\tilde{u}) + \frac{1}{2}\log(1-|m|^2) + \mathcal{F}_K(\beta(1-|m|^2)).$$

The next lemma will adapt the statement from [BK19, Lemma 17] for non-linear spikes.

Lemma 3.1. *There exists a linear subspace M_N of $\mathbb{R}^N$ of dimension $M := \dim M_N = \lfloor N^{\frac{3}{4}}\rfloor$ such that*

$$\lim_{N\to\infty}\sup_{m\in M_N,|m|\le 1}\sup_{\hat{\sigma}\in M_N^{\perp},|\hat{\sigma}|\le 1}\sup_{\lambda\in\mathbb{R}:|\lambda|\le|f'|_\infty}\left(\beta\frac{1}{N}\nabla\tilde{H}_N(m) + N\lambda u\right)\cdot\hat{\sigma} = 0. \tag{3.3}$$

Proof. We create M_N from [BK19, Lemma 17] using $|f'|_\infty u$ in place of h_N for its construction. [BK19, Lemma 17] then implies that

$$\lim_{N\to\infty}\sup_{m\in\mathrm{M}_N,|m|\le 1}\sup_{\hat{\sigma}\in\mathrm{M}_N^{\perp},|\hat{\sigma}|\le 1}\left(\beta\frac{1}{N}\nabla\tilde{H}_N(m) + N|f'|_\infty u\right)\cdot\hat{\sigma} = 0. \tag{3.4}$$

Since M_N is a linear subspace the claim follows. □

We will now prove a statement similar to [BK19, Proposition 19].

Lemma 3.2. *We have for N large enough*

$$\tilde{F}^f_N(\beta,u) \le \frac{1}{N}\sup_{m:|m|<1}\tilde{F}^K_{TAP}(m) + \frac{c}{K}$$

for a universal constant $c > 0$.

Proof. We will write c for fixed universal constants, where the exact value of c can change over the course of this proof. Let M_N be the set from Lemma 3.1. For any $\sigma\in\mathbb{R}^N$ let m be the projection of σ onto M_N and $\hat{\sigma} = \sigma - m \in \mathrm{M}_N^{\perp}$. Recentering the Hamiltonian around m and using that $|f(\sigma\cdot u) - f(m\cdot u)| \le |f'|_\infty|\hat{\sigma}\cdot u|$ one obtains as in [BK19, (4.44)]

$$\begin{aligned}
&E\left[\exp\left(\beta\tilde{H}_N(\sigma) + Nf(\sigma\cdot u)\right)\right]\\
=&E\left[\exp\left(\beta\tilde{H}_N(m) + Nf(m\cdot u) + \beta\nabla\tilde{H}_N(m)\cdot\hat{\sigma} + (f(\sigma\cdot u) - f(m\cdot u)) + \beta\tilde{H}_N(\hat{\sigma})\right)\right]\\
\le&E\left[\exp\left(\beta\tilde{H}_N(m) + Nf(m\cdot u) + N\left(\frac{\beta}{N}\nabla\tilde{H}_N(m) + \mathrm{sign}(\hat{\sigma}\cdot u)|f'|_\infty u\right)\cdot\hat{\sigma} + \beta\tilde{H}_N(\hat{\sigma})\right)\right],
\end{aligned}$$

which by (3.3) is at most

$$e^{o(N)}E\left[\exp\left(\beta\tilde{H}_N(m) + Nf(m\cdot u) + \beta\tilde{H}_N(\hat{\sigma})\right)\right]. \tag{3.5}$$

Let us now condition on the projection m. Since the $E[\cdot|m]$-law of $\hat{\sigma}$ is the uniform distribution on the sphere in the subspace $\mathrm{M}_N^{\perp}$ of radius $\sqrt{1-|m|^2}$, we have that (3.5) is equal to

$$e^{o(N)}E\left[\exp\left(\beta\tilde{H}_N(m) + Nf(m\cdot u)\right)E_{\mathrm{M}_N^{\perp}}\left[\exp\left(\beta(1-|m|^2)\tilde{H}_N(\sigma)\right)\right]\right]. \tag{3.6}$$

[BK19, Lemma 18] implies that that for any $C > 0$ and $K > 0$

$$\limsup_{N\to\infty}\sup_{\beta\in[0,C]}\left|\frac{1}{N}\log E_{\mathrm{M}_N^{\perp}}\left[\exp\left(\tilde{H}_N(\sigma)\right)\right] - \mathcal{F}_K(\beta)\right| \le \frac{c}{K},$$

so one obtains that (3.6) is at most

$$E\left[\exp\left(\beta\tilde{H}_N(m) + Nf(m\cdot u) + N\mathcal{F}_K(\beta(1-|m|^2))\right)\right]e^{o(N)+\frac{c}{K}N}. \tag{3.7}$$

By [BK19, (2.9)] the projection of $\sigma \in \mathcal{S}_{N-1}$ onto $\mathrm{M}_N^\perp$ has density

$$\underbrace{\frac{1}{\pi^{\frac{N-M}{2}}}\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{M}{2}\right)}}_{=e^{o(N)}}(1-|m|^2)^{\frac{N-M-2}{2}}dm,$$

so we get that (3.7) is equal to

$$e^{o(N)+\frac{c}{K}N}\int_{m:|m|<1}\exp\left(\tilde{F}^K_{\text{TAP}}(m)-(M+2)\log(1-|m|^2)\right)dm, \tag{3.8}$$

which in turn is bounded from above by

$$e^{o(N)+\frac{c}{K}N}\exp\left(\sup_{m:|m|<1}\left\{\tilde{F}^K_{\text{TAP}}(m)-(M+2)\log(1-|m|^2)\right\}\right)\underbrace{\int_{m:|m|<1}dm}_{=\mathcal{O}(1)}. \tag{3.9}$$

Note that we can find a δ depending only on β, u and f such that the supremum is always achieved for $|m|<1-\delta$, because

$$N\beta\tilde{H}_N(m)+Nf(m\cdot u)+N\mathcal{F}_K(\beta(1-|m|^2))\le cN$$

for some constant $c>0$, and therefore we obtain that the supremum in (3.9) is at most

$$\sup_{m:|m|<1}\tilde{F}^K_{\text{TAP}}(m)+cM, \tag{3.10}$$

which implies that $\tilde{F}^f_N(\beta,u)$ is bounded by $\sup_{m:|m|<1}\tilde{F}^K_{\text{TAP}}(m)+o(N)+\frac{cN}{K}$. □

The following lemma will prove the upper bound for low temperatures. We will follow the steps of [BK19, Proposition 7]. Let us define a deterministic variant of the TAP free energy

$$\tilde{F}_{\text{TAP}}(m)=\beta\tilde{H}_N(m)+Nf(m\cdot\tilde{u})+\frac{N}{2}\beta^2(1-|m|^2)^2+\frac{N}{2}\log(1-|m|^2),$$

where by (3.2) it suffices to show an upper bound with this deterministic variant.

Proposition 3.3. *If $f\in C^2([-1,1])$ then*

$$\tilde{F}^f_N(\beta,u)\le\frac{1}{N}\sup_{m\in\mathcal{M}_\beta}\tilde{F}_{TAP}(m)+o(1).$$

Proof. Fix $K\ge 2$. For any $N\ge 1$, we have that any local maximum m of $\tilde{F}^K_{\text{TAP}}(m)$ must satisfy

$$\nabla\tilde{F}^K_{\text{TAP}}(m)=0,$$

and

$$\nabla^2\tilde{F}^K_{\text{TAP}}(m)\;\text{ is negative semi-definite.} \tag{3.11}$$

Recall (3.1) and note that by [BK19, Lemma 12]

$$\mathcal{F}'_K(\beta)=\Lambda_K(\beta)-\frac{1}{2\beta},$$

so it holds that

$$\begin{aligned}\nabla\tilde{F}^K_{\text{TAP}}(m)=&\beta\nabla\tilde{H}_N(m)+Nf'(m\cdot u)u-Nm\left(\frac{1}{1-|m|^2}+2\beta\mathcal{F}'_K(\beta(1-|m|^2))\right)\\=&\beta\nabla\tilde{H}_N(m)+Nf'(m\cdot u)u-N2\beta m\Lambda_K(\beta(1-|m|^2)).\end{aligned}$$

Thus the Hessian $\nabla^2 \tilde{F}^K_{\mathrm{TAP}}(m)$ is equal to

$$\beta\nabla^2 \tilde{H}_N(m) + Nf''(m\cdot u)uu^T - N2\beta\Lambda_K(\beta(1-|m|^2))I + N4\beta^2\Lambda_K'(\beta(1-|m|^2))mm^T,$$

where I denotes the identity matrix. For any local maximum m let

$$A = \frac{1}{N}\beta\nabla^2 \tilde{H}_N(m) - 2\beta\Lambda_K(\beta(1-|m|^2))I$$

and

$$B = f''(m\cdot u)uu^T + 4\beta^2\Lambda_K'(\beta(1-|m|^2))mm^T.$$

Since the two matrices are symmetric and B is at most of rank 2, it follows by [Ful00, (11)] for $N \geq 5$ that the third largest eigenvalue a_{N-2} of A is bounded above by the largest eigenvalue of $A+B$. By $\nabla^2 \tilde{F}^K_{\mathrm{TAP}}(m) = N(A+B)$ and (3.11) one obtains that all eigenvalues of $A+B$ must be non-positive, and therefore $a_{N-2} \leq 0$. Furthermore since $\frac{1}{N}\beta\nabla^2 \tilde{H}_N(m)$ is a diagonal matrix with $2\beta\theta_{i/N}$, $i = 1, ..., N$ on its diagonal, the eigenvalues of A are $2\beta\theta_{i/N} - 2\beta\Lambda_K(\beta(1-|m|^2))$. This shows that

$$\Lambda_K(\beta(1-|m|^2)) \geq \theta_{1-\frac{2}{N}},$$

at all m which are local maxima. [BK19, Lemma 13] states that if $\Lambda_K(\beta) \geq \sqrt{2} - \varepsilon$ for some $\varepsilon \in (0, \frac{2\sqrt{2}}{K})$, then $\beta \leq \frac{1}{\sqrt{2}}$, so because $\theta_{1-\frac{2}{N}} = \sqrt{2} + o(1)$ it follows that

$$\beta(1-|m|^2) \leq \frac{1}{\sqrt{2}}$$

provided that N is large enough depending on K. [BK19, Lemma 14 + (4.28)] state that

$$\lim_{K\to\infty} \sup_{\beta\in[0,\frac{1}{\sqrt{2}}]} \left|\mathcal{F}_K(\beta) - \frac{\beta^2}{2}\right|,$$

so that

$$\mathcal{F}_K(\beta(1-|m|^2)) \leq \frac{\beta^2}{2}(1-|m|^2)^2 + \varepsilon_K,$$

where $\lim_{K\to\infty}\varepsilon_K = 0$. By Lemma 3.2 we then obtain

$$\tilde{F}^f_N(\beta, u) \leq \frac{1}{N}\sup_{m\in\mathcal{M}_\beta} \tilde{F}_{\mathrm{TAP}}(m) + \varepsilon_K + \frac{c}{K}.$$

Since both $\tilde{F}_{\mathrm{TAP}}(m)$ and $\tilde{F}^f_N(\beta, u)$ are independent of K we can take the limit in K and obtain the claim. □

Proposition 2.2 and Proposition 3.3 together prove (1.1).

CHAPTER III

Fluctuations of the groundstate in a spiked SSK model

Fluctuations of the ground state of the spiked spherical Sherrington-Kirkpatrick model

David Belius, Leon Fröber

The Sherrington-Kirkpatrick Hamiltonian is a random quadratic function on the high-dimensional sphere. This article studies the ground state (i.e. maximum) of this Hamiltonian with external field, or more generally with a non-linear "spike" term. We compute the level of the maximum to leading order, and under appropriate condition its first- and second-order fluctuations. The equivalent results are also derived for the maximum of the model's TAP free energy on the ball.

1. Introduction

This article studies the maximum of a natural random quadratic optimization problem in N variables over the sphere or ball in $\mathbb{R}^N$, in the presence of a possibly non-linear "spike" term. We prove a leading order law of large numbers as $N \to \infty$, and study the fluctuations around the limit. In the context of spin glasses [SK75, MPV87, Tal10, Pan13b] the maximum on the sphere that we study is precisely the ground state of the spherical Sherrington-Kirkpatrick Hamiltonian [KTJ76] with external field, or more generally with a non-linear "spike". Our result on maximum on the ball applies to the TAP free energy [TAP77, CS95, BK19] of this Hamiltonian.

The random quadratic optimization problem $\sup_{\sigma\in\mathbb{R}^N:|\sigma|=1}\{\sigma^T J\sigma + \sigma \cdot v\}$ for an $N \times N$ random matrix J and vector $v \in \mathbb{R}^N$ constitutes arguably the most basic yet interesting high-dimensional random optimization problem and merits special attention. The case where J is a GOE random matrix is representative. The large deviations of this maximum has been studied in [FLD14, DZ15]. A natural generalization is to replace the linear "external field" term $\sigma \cdot v$ with $f(\sigma \cdot v)$ for some non-linear "spike" function f [RM14, LKZ17, LM19, AMMN19]. The present paper determines the leading order of the maximum for general f, and gives a precise description of its fluctuations (i.e. its "typical deviations"). In particular Theorem 1.1 provides both a law of large numbers that computes the order N asymptotic of the maximum, and under appropriate assumptions on f also determines first- and second-order subleading fluctuation terms of order $N^{1/2}$ and 1 respectively.

Our main motivation comes from mean-field spin glasses, and concerns the maximum of the TAP free energy, which is a function of the form $m \mapsto m^T Jm + f(m \cdot v) + g(|m|)$ defined on the unit ball, for a certain function g that we recall below. Theorem 1.2 computes the leading order and fluctuations of the maximum of such a function on the ball, for a general g. Below we discuss the spin-glass motivation in more detail.

To formally state our results, define the Sherrington-Kirkpatrick *Hamiltonian*

$$H_N(\sigma) = \sqrt{N}\sigma^T J\sigma \text{ for } \sigma \in \mathbb{R}^N \tag{1.1}$$

where J is an $N \times N$ GOE random matrix, i.e. a symmetric matrix with centered Gaussian entries $J_{i,j}$ mutually independent for $i \le j$, and $\mathrm{Var}(J_{i,j}) = \frac{1}{2}(1 + \delta_{i=j})$. Let $f : [-1, 1] \to \mathbb{R}$ be a real function, $\beta > 0$ a constant which we call the *inverse temperature* and $v \in \mathbb{R}^N, |v| = 1$, a unit vector giving the direction of the spike. The *ground state* is the maximum

$$L_N = \sup_{|\sigma|=1}\{\beta H_N(\sigma) + Nf(v\cdot\sigma)\} \tag{1.2}$$

over the unit sphere. Let $\xrightarrow{\mathbb{P}}$ denote convergence in probability, $\xrightarrow{d}$ convergence in distribution, and $\mathcal{N}(\mu, \sigma^2)$ the Gaussian distribution with mean μ and variance σ^2. Our result about the maximum on the sphere is the following.

Theorem 1.1 (Maximum on sphere). *Let $f \in C^0([-1,1])$ and*

$$\mathcal{B}(\alpha) = f(\alpha) + \beta\sqrt{2(1-\alpha^2)}. \tag{1.3}$$

(a) *(Leading order) It holds that*

$$\frac{1}{N} L_N \xrightarrow{\mathbb{P}} \sup_{\alpha\in[-1,1]} \mathcal{B}(\alpha). \tag{1.4}$$

(b) *(Fluctuations) If additionally $f \in C^3([-1,1])$ and $\mathcal{B}(\alpha)$ has a unique global maximizer $\hat{\alpha} \neq 0$ with $\mathcal{B}''(\hat{\alpha}) < 0$, then there exist a constant κ and a matrix G such that*

$$L_N - N\mathcal{B}(\hat{\alpha}) - \sqrt{N}\kappa U_N - \left(\kappa\Lambda_N - \frac{1}{2}\begin{pmatrix} U_N \\ U_N' \end{pmatrix}^T G \begin{pmatrix} U_N \\ U_N' \end{pmatrix}\right) \xrightarrow{\mathbb{P}} 0, \tag{1.5}$$

where U_N, U_N', Λ_N are stochastically bounded random variables defined by

$$U_N = \sqrt{N}\left(v^T G_N v - \frac{\mathrm{Tr} G_N}{N}\right),\ U_N^{'} = -\sqrt{N}\left(v^T G_N^2 v - \frac{\mathrm{Tr} G_N^2}{N}\right),\ \Lambda_N = \frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{l} - \lambda_i} - \hat{z},$$

for $\hat{z} = \sqrt{2(1-\hat{\alpha})^2}$, $\hat{l} = \frac{2-\hat{\alpha}^2}{\hat{z}}$, $G_N = \left(\hat{l}\cdot I - \frac{J}{\sqrt{N}}\right)^{-1}$.
The random variables satisfy

$$(U_N, U_N', \Lambda_N) \xrightarrow{d} (U, U', \Lambda),$$

where

$$U \sim \mathcal{N}\left(0, \frac{\hat{z}^4}{\hat{\alpha}^2}\right), \quad U' \sim \mathcal{N}\left(0, \frac{\hat{z}^6(2+\hat{\alpha}^2+\hat{\alpha}^4)}{\hat{\alpha}^{10}}\right), \quad \Lambda \sim \mathcal{N}\left(\frac{\hat{z}^3}{2\hat{\alpha}^4}, \frac{\hat{z}^4}{\hat{\alpha}^8}\right) \tag{1.6}$$

with (U, U') and Λ independent and

$$Cov(U, U') = -\frac{\hat{z}^5(1+\hat{\alpha}^2)}{\hat{\alpha}^6}. \tag{1.7}$$

The constant and matrix are given by

$$\kappa = \frac{\beta\hat{\alpha}^2}{\hat{z}^2}, \qquad G = \beta\left(\frac{8\beta\hat{\alpha}^2}{\hat{z}^8\mathcal{B}''(\hat{\alpha})}\begin{pmatrix} 2 & \frac{\hat{\alpha}^4}{\hat{z}} \\ \frac{\hat{\alpha}^4}{\hat{z}} & \frac{\hat{\alpha}^8}{2\hat{z}^2} \end{pmatrix} + \begin{pmatrix} \frac{2\hat{\alpha}^2}{\hat{z}^3} & 0 \\ 0 & 0 \end{pmatrix}\right).$$

The same holds if $\mathcal{B}(\alpha)$ has a pair of global unique maximizers $\pm\hat{\alpha} \neq 0$ with $\mathcal{B}(\hat{\alpha}) = \mathcal{B}(-\hat{\alpha}), \mathcal{B}''(\hat{\alpha}) = \mathcal{B}''(-\hat{\alpha}) < 0$.

Part (a) for a linear spike functions $f(x) = hx, h \in \mathbb{R}$, appears in [BK19, Lemma 20] and is implicit in [DZ15, Theorem 1.3]. In that case the maximizer $\hat{\alpha}$ is unique and $\mathcal{B}(\hat{\alpha}) = \sqrt{2\beta^2 + h^2}$. The first-order fluctuation result of part (b) in the same linear-spike case, namely the convergence in law of $N^{-1/2}(L_N - \sqrt{2\beta^2 + h^2})$ to a centered Gaussian, is implied also by [CS17, Theorem 5] as explained in Remark 7.3. This corresponds to the first-order fluctuation term $\sqrt{N}\kappa U_N$ in (1.5).

Part (b) of the theorem covers the regime where the fluctuations are determined by the central limit-type behavior of sums over eigenvalues and entries of the spike vector v, and for this reason requires $\hat{\alpha} \neq 0$. When $\hat{\alpha} = 0$ the fluctuations should instead be determined by the fluctuations of the extreme eigenvalues of J (indeed for $f = 0$ the maximum is exactly the largest eigenvalue, which has non-Gaussian fluctuations [TW96]).

In Section 5 and 7 we give more explicit formulas for leading order and fluctuations for monomial spike functions f, and for these determine critical inverse temperatures β where the behavior of the ground state changes.

Our second main results concerns the fluctuations of the maximum on the ball of combinations of H_N with a spike and a deterministic radial function. For functions $f : [-1,1] \to \mathbb{R}$ and $g : [0,1] \to \mathbb{R}$ define

$$\tilde{L}_N = \sup_{m\in B_N(\mathcal{R})} \{\beta H_N(m) + Nf(v\cdot m) + Ng(|m|)\}, \tag{1.8}$$

where $B_N(\mathcal{R}) = \{m \in \mathbb{R}^N : |m| \in \mathcal{R}\}$ and $\mathcal{R} \subset [0,1]$. The prototypical example is the maximum of the TAP free energy, where the function g takes a particular form and the maximum is taken only over m with $|m|^2$ in a certain range, which is why we include the set $\mathcal{R}$ in the formulation (see the discussion after the theorem).

Theorem 1.2 (Maximum on ball)**.** *For $f \in C^0([-1,1])$, $\mathcal{R} \subset [0,1]$ closed and $g \in C^0(\mathcal{R})$ let*

$$\tilde{\mathcal{B}}(\alpha, r) = f(r\alpha) + g(r) + \beta r^2 \sqrt{2(1-\alpha^2)}. \tag{1.9}$$

(a) *(Leading order) It holds that*

$$\frac{1}{N}\tilde{L}_N \xrightarrow{\mathbb{P}} \sup_{r\in\mathcal{R}, \alpha\in[-1,1]} \tilde{\mathcal{B}}(\alpha, r). \tag{1.10}$$

(b) *(Fluctuations) If additionally $f \in C^3([-1,1]), g \in C^3(\mathcal{R})$ and $\tilde{\mathcal{B}}(\alpha, r)$ has a unique global maximizer $(\hat{\alpha}, \hat{r})$ in the interior of $\mathcal{R} \times [-1,1]$ with $\hat{\alpha} \neq 0, \hat{r} \neq 0$, and the Hessian matrix $\nabla^2 \tilde{\mathcal{B}}(\hat{\alpha}, \hat{r})$ is negative definite, then there is a matrix $\tilde{G}$ such that*

$$\tilde{L}_N - \tilde{\mathcal{B}}(\hat{\alpha}, \hat{r}) - \sqrt{N}\kappa U_N - \left(\kappa \Lambda_N - \frac{1}{2}\begin{pmatrix} U_N \\ U_N' \end{pmatrix}^T \tilde{G} \begin{pmatrix} U_N \\ U_N' \end{pmatrix}\right) \xrightarrow{\mathbb{P}} 0, \tag{1.11}$$

where $U_N, U_N', \Lambda_N, \kappa$ are as in Theorem 1.1. The matrix is given in terms of $\hat{z} = \sqrt{2(1-\hat{\alpha}^2)}$ by

$$\tilde{G} = K^T \left(\nabla^2 \mathcal{B}(\hat{\alpha}, \hat{r})\right)^{-1} K + \begin{pmatrix} 2\beta \frac{\hat{r}^2\hat{\alpha}^2}{\hat{z}^3} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{where} \quad K = \frac{2\beta\hat{r}\hat{\alpha}}{\hat{z}^2}\begin{pmatrix} \frac{2\hat{r}}{\hat{z}^2} & \frac{\hat{r}\hat{\alpha}^4}{\hat{z}^3} \\ \hat{\alpha} & 0 \end{pmatrix}.$$

The same holds if $\tilde{\mathcal{B}}(\alpha, r)$ has a pair of global unique maximizers $(\pm\hat{\alpha}, \hat{r})$ in the interior of $\mathcal{R} \times [-1,1]$ with $\hat{\alpha} \neq 0$, $\tilde{\mathcal{B}}(\hat{\alpha}, \hat{r}) = \tilde{\mathcal{B}}(-\hat{\alpha}, \hat{r})$ and $\nabla^2 \tilde{\mathcal{B}}(\hat{\alpha}, \hat{r}) = \nabla^2 \tilde{\mathcal{B}}(-\hat{\alpha}, \hat{r})$ negative-definite.

In the Thouless-Andersson-Palmer (TAP) [TAP77] approach to spin glasses one aims to extract important information about spin glass models from their *TAP free energy*, which is a random function arising from the Hamiltonian H_N of the model. For the spiked spherical Sherrington-Kirkpatrick model of this article it is given by [CS95, BK19]

$$F_{\mathrm{TAP}}(m) = \beta H_N(m) + Nf(v\cdot m) + \frac{N}{2}\log(1-|m|^2) + \frac{N}{2}\beta^2(1-|m|^2)^2, \ |m| < 1. \tag{1.12}$$

Only m satisfying certain conditions are believed to be "relevant" [TAP77, Ple82a, Ple82b, Sub18, BK19]. For the spherical Sherrington-Kirkpatrick model the only needed condition is *Plefka's condition*, requiring that $\sqrt{2}\beta(1-|m|^2) \leq 1$. [BK19]. The maximal TAP free energy over m that satisfy Plefka's condition is of the form (1.8) with $g(r) = \frac{1}{2}\left(\log(1-r^2) + \beta^2(1-r^2)^2\right)$ and $\mathcal{R} = \{r : r^2 \geq 1 - \frac{1}{\sqrt{2}\beta}\}$. In Sections 5 and 7 we determine more concretely for this g and monomial f when the conditions of Theorem 1.1 and 1.2 are satisfied and what the resulting formulas for leading order and fluctuations are.

1.1. Fluctuations and the TAP approach

The SK model and its variants consist of a high-dimensional spin space such the sphere $\{\sigma \in \mathbb{R}^N : |\sigma| = 1\}$ and a random energy such as $\beta H_N(\sigma) + Nf(\sigma \cdot v)$ associated to each spin configuration vector σ, where $H_N(\sigma)$ is a high-dimensional Gaussian field of which $H_N(\sigma)$ from (1.1) is a special case. From this energy one constructs the Gibbs measure, which in the case of a spherical spin space is the probability measure with density proportional to the Gibbs factor $\exp(\beta H_N(\sigma) + Nf(\sigma \cdot v))$ with respect to the uniform measure on the sphere. The normalizing factor of the measure is known as the partition function and usually denoted by Z_N. The vector σ sampled according to the Gibbs measure models the spins of exotic magnet materials, or other complex phenomena in related models [MPV87, MM09]. The ultimate goal of the area is to describe the behavior of σ sampled according to the Gibbs measure.

For the general class of mixed p-spin Hamiltonians H_N [Der80, GM84b, Tal00, AA13] this is a formidable task that is far from being accomplished. In the general case the "geometry" of the random landscape H_N is extraordinarily complex [Fyo15, AA13, AAČ13, Sub17a], and this is expected to be reflected in the behavior of the Gibbs measure. The Sherrington-Kirkpatrick Hamiltonian (1.1) is the special case of a 2-*spin* Hamiltonian, which when combined with a spherical spin space has significantly simpler behavior, and is much easier to study due to the spherical symmetry and quadratic nature of the Hamiltonian allowing many explicit calculations that are impossible in general. As such the 2-spin setting provides a valuable testing ground for new ideas and techniques. The motivation for this paper is to use the 2-spin spherical Hamiltonian as a starting point to explore fluctuations in spin glasses via a TAP approach.

A first step in understanding the Gibbs measure is computing the *free energy* which is the limit of $\frac{1}{N}\log Z_N$ as $N \to \infty$, i.e. the rate of exponential growth of the partition function. Knowledge of the free energy morally speaking corresponds to knowledge of which regions of the spin space have probability at least $e^{-o(N)}$ under the Gibbs measure, rather than exponentially small probability. Finer estimates for the free energy, such as lower order corrections and fluctuations, morally correspond to finer knowledge of the Gibbs measure. There are several approaches to computing the free energy [Par80, Gue03, ASS03, Tal06a, Tal06b, Con13, Pan14, Che13]. In the TAP approach one expects that the free energy is roughly speaking given by the maximum of the TAP free energy $F_{\mathrm{TAP}}(m)$ of the model. The final term of $F_{\mathrm{TAP}}(m)$ is called the "Onsager term", and the $F_{\mathrm{TAP}}(m)$ of general mixed p-spin spherical models coincides with (1.12) but with a more general Onsager term. The TAP approach for general models is under active investigation [Bol14, Bol19, BY22, Sub17b, CPS22, Sub21, Bel22] and the correspondence between free energy and maximal TAP free energy is proven mathematically rigorously without appealing to powerful machinery like the Parisi formula only in a few cases [Sub21, BK19, BFK23]. One of these is the spherical 2-spin case of this paper, where the free energy was computed completely within a TAP approach in [BK19].

From the point of view of the TAP approach the fluctuations of the free energy should arise on the one hand from the fluctuations of the maximum of F_{TAP}, and on the other hand from the fluctuations of certain "local" integrals (over "slices" in the terminology of [BK19, BFK23, Bel22] and over "bands" in the terminology of [Sub17b, Sub18, CPS22]; the Onsager term of F_{TAP} describes the leading order behavior of these integrals). In this article we completely determine the former kind of fluctuations for the spherical 2-spin model, to the highest degree of precision that is plausibly relevant for the study of the fluctuations of the free energy and Gibbs measure. The analysis of the latter type of fluctuations, and consequences for the fluctuations of the free energy, are left to future work.

See for instance [ALR87, BKL02, Cha09, BL16, CS17, SZ17, BCWDW20, Lan20, LS20, BB21, BS22b] for work on fluctuations in spin glasses from a non-TAP point of view.

1.2. Sketch of proof

In this subsection we give a brief sketch of our arguments. To prove Theorem 1.1 we diagonalize the matrix J and obtain that

$$\frac{1}{N}L_N = \sup_{|\sigma|=1}\left\{\beta\frac{1}{N}H_N(\sigma) + f(v\cdot\sigma)\right\} \stackrel{d}{=} \sup_{|\sigma|=1}\left\{\beta\sum_{i=1}^N \lambda_i\sigma_i^2 + f\left(\sum_{i=1}^N u_i\sigma_i\right)\right\}, \tag{1.13}$$

where $\lambda_1 < \ldots < \lambda_N$ are the eigenvalues of $\frac{1}{\sqrt{N}}J$ and u is the spike vector v written in the diagonal basis. By the orthogonal invariance of J the vector u is uniform on the sphere and independent of the λ_i. Next we decompose the maximization in (1.13) according to the value of $\sum_{i=1}^N u_i\sigma_i$ to obtain

$$\frac{1}{N}L_N \stackrel{d}{=} \sup_{\alpha\in[-1,1]}\left\{f(\alpha) + \beta\sup_{\substack{|\sigma|=1\\ \sigma\cdot u=\alpha}}\sum_{i=1}^N \lambda_i\sigma_i^2\right\}. \tag{1.14}$$

In the proof of Theorem 1.2 we use the similar identity (3.4) for $\frac{1}{N}\tilde{L}_N$ where the outer supremum is also over r.

We then solve the constrained optimization problem in (1.14) using Lagrange multipliers, and obtain the identity

$$\sup_{\substack{|\sigma|=1 \\ \sigma\cdot u=\alpha}} \sum_{i=1}^{N} \lambda_i \sigma_i^2 = \inf_{l>\lambda_N} \left\{ l - \frac{\alpha^2}{s_{\lambda,u}(l)} \right\} \quad \text{for} \quad s_{\lambda,u}(l) = \sum_{i=1}^{N} \frac{u_i^2}{\lambda_i - l}, \tag{1.15}$$

provided $|\alpha| \geq |u_N|$ over $l > \lambda_N$. This reduces the high-dimensional optimization over $\sigma \in \mathbb{R}^N$ to a low-dimensional one. We recognize the random function $s_{\lambda,u}(l)$ as the Stieltjes transform of the empirical spectral distribution of J weighted by u_i^2. It is easy to see that it converges to the Stieltjes transform of the semi-circle law $s(l)$. In our normalization it is given by $s(l) = l - \sqrt{l^2 - 2}$, and also $\lambda_N \to \sqrt{2}$ in probability. We thus obtain from (1.15) a limiting optimization problem which is explicitly solvable:

$$\inf_{l>\sqrt{2}} \left\{ l - \frac{\alpha^2}{s(l)} \right\} = \sqrt{2(1-\alpha^2)}, \tag{1.16}$$

cf. (1.3). To prove the leading order results Theorem 1.1 (a) and Theorem 1.2 (a) it suffices to approximate the infimum in (1.15) by that in (1.16). For this purpose we obtain in Section 4 sufficiently uniform estimates for the convergence of $s_{\lambda,u}(l)$ to $s(l)$, and combine these with a simple ad-hoc argument for $|\alpha| \leq |u_N|$ to prove Theorem 1.1 (a) and Theorem 1.2 (a).

For the fluctuation result Theorem 1.1 (b) the assumption that $\hat{\alpha} \neq 0$ makes the identity (1.15) hold in a neighborhood $[\hat{\alpha} - \varepsilon, \hat{\alpha} + \varepsilon]$ of the unique maximizer $\hat{\alpha}$ with high probability, and using this the maximum can be written exactly as the minimax

$$\frac{1}{N} L_N = \sup_{\alpha\in[\hat{\alpha}-\varepsilon,\hat{\alpha}+\varepsilon]} \inf_{l>\lambda_N} h(\alpha, l, s_{\lambda,u}(l)) \quad \text{for} \quad h(\alpha, l, g) = f(\alpha) + \beta\left(l - \frac{\alpha^2}{g}\right).$$

A similar function $h((\alpha, r), l, g)$ gives a similar "high probability" identity for $\frac{1}{N}\tilde{L}_N$ (see (6.2)). Therefore both Theorem 1.1 (b) and Theorem 1.2 (b) can be proved by studying fluctuations of

$$\sup_{y\in\mathcal{Y}} \inf_{l\in\mathcal{L}} h(y, l, s_{\lambda,u}(l)), \tag{1.17}$$

for a general function $h(y, l, g)$ where $y \in \mathcal{Y} \subset \mathbb{R}^n, n \geq 1$ and $\mathcal{L} \subset (\sqrt{2}, \infty)$, under the assumption that the limiting minimax $\sup_{y\in\mathcal{Y}} \inf_{l\in\mathcal{L}} h(y, l, s(l))$ has a unique optimizer $(\hat{y}, \hat{l})$. In Section 4 we study the fluctuations of $s_{\lambda,u}(l)$ around $s(l)$ using a combination of central limit theorems for sums over eigenvalues and the entries of the spike vector v. We then expand $h(y, l, s_{\lambda,u}(l))$ quadratically in these fluctuations and in y, l, around the point $\hat{y}, \hat{l}, s(\hat{l})$. The first- and second-order fluctuations of (1.17) are obtained by solving the minimax optimization for this approximating quadratic, leading to the proof of (1.5) and (1.11).

1.3. Organization

In the preliminary Section 2 we recall some useful results about the GOE random matrix and its eigenvalues. In Section 3 we use Lagrange multipliers to reduce the optimizations over σ in L_N and $\tilde{L}_N$ to low-dimensional optimization as described in the sketch above. In Section 4 we prove uniform leading order estimates for the convergence of $s_{\lambda,u}(l)$ to $s(l)$, and deduce from these the leading order estimates Theorem 1.1 (a) and Theorem 1.2 (a). Then in Section 5 we provide some concrete examples of f and g to which the leading order results apply. In Section 6 we study the fluctuations of $s_{\lambda,u}$, and use this and the quadratic expansion described in the sketch to prove the fluctuation results Theorem 1.1 (b) and Theorem 1.2 (b). Finally in Section 7 we apply these to study the fluctuation for the examples of Section 5.

1.4. Notation

We use the following notations, in addition to those already introduced before Theorem 1.1. The unit sphere is denoted $\mathcal{S}_{N-1} = \{\sigma \in \mathbb{R}^N : |\sigma| = 1\}$. Furthermore we write $O_{\mathbb{P}}$ and $o_{\mathbb{P}}$ for probabilistic versions of the standard notation for the order of quantities as $N \to \infty$. More precisely we write $X_N = O_{\mathbb{P}}(T(N))$ if $X_N/T(N)$ is stochastically bounded, i.e. if

$$\lim_{x\to\infty} \limsup_{N\to\infty} \mathbb{P}\left(\frac{|X_N|}{T(N)} \geq x\right) = 0, \tag{1.18}$$

and $X_N = o_{\mathbb{P}}(T(N))$ if

$$\frac{|X_N|}{T(N)} \xrightarrow{\mathbb{P}} 0. \tag{1.19}$$

2. Random matrix preliminaries

In this section we recall some standard results about the eigenvalues of the GOE. We denote the semi-circle law on $[-\sqrt{2}, \sqrt{2}]$ by

$$\mu_{\mathrm{sc}}(dx) = \frac{\sqrt{2-x^2}}{\pi} dx. \tag{2.1}$$

Let $\theta_{1/N}, ..., \theta_{N/N} \in [-\sqrt{2}, \sqrt{2}]$ be given by

$$\int_{-\sqrt{2}}^{\theta_{k/N}} \mu_{\mathrm{sc}}(dx) = \frac{k}{N}, \tag{2.2}$$

which are sometimes called the *classical locations* of the eigenvalues of J. From e.g. [EYY12, Theorem 2.2] we know that the eigenvalues concentrate around these, i.e.:

Lemma 2.1. *For any $\varepsilon > 0$ and all $k \in \{1, ..., N\}$*

$$|\lambda_k - \theta_{k/N}| \leq N^{-\frac{2}{3}+\varepsilon} \min\left\{k^{-\frac{1}{3}}, (N-k)^{-\frac{1}{3}}\right\}$$

with probability tending to one as $N \to \infty$.

In particular

$$\lambda_N \xrightarrow{\mathbb{P}} \sqrt{2} \text{ and } \lambda_1 \xrightarrow{\mathbb{P}} -\sqrt{2}. \tag{2.3}$$

It is elementary to estimate sums of the classical locations with integrals over the semi-circle law. The next lemma records this.

Lemma 2.2. *For all $w \in C^1\left([-\sqrt{2}-\varepsilon, \sqrt{2}+\varepsilon]\right)$ it holds that*

$$\left|\frac{1}{N}\sum_{i=1}^{N} w\left(\theta_{i/N}\right) - \int_{-\sqrt{2}}^{\sqrt{2}} w\left(x\right)\mu_{sc}\left(dx\right)\right| \leq \frac{2\sqrt{2}|w'|_\infty}{N}, \tag{2.4}$$

Proof. It follows from (2.2) that

$$\begin{aligned}\int_{-\sqrt{2}}^{\sqrt{2}} w(x)\mu_{sc}(dx) &= \sum_{i=1}^{N} \int_{\theta_{(i-1)/N}}^{\theta_{i/N}} w(x)\mu_{sc}(dx) \\ &= \sum_{i=1}^{N} w(\theta_{i/N})\frac{1}{N} + \xi(w)\frac{1}{N}\sum_{i=1}^{N}(\theta_{i/N} - \theta_{(i-1)/N}),\end{aligned}$$

where $\xi(w) \in [-|w'|_\infty, |w'|_\infty]$. □

The next lemma is concerned with fluctuations of sums over the eigenvalues.

Lemma 2.3. *If $\varepsilon > 0$ and $w \in C^1([-\sqrt{2}-\varepsilon, \sqrt{2}+\varepsilon])$*

$$\sum_{i=1}^{N} w(\lambda_i) - N \int_{-\sqrt{2}}^{\sqrt{2}} w(x)\mu_{sc}(dx) \xrightarrow{d} \mathcal{N}\left(m(w), v(w)\right),$$

where

$$\begin{aligned} m(w) =& \frac{w(\sqrt{2}) + w(-\sqrt{2})}{4} - \frac{1}{2\pi} \int_{-\sqrt{2}}^{\sqrt{2}} w(x) \frac{1}{\sqrt{2-x^2}} dx \\ v(w) =& \frac{1}{2\pi^2} \int_{-\sqrt{2}}^{\sqrt{2}} \int_{-\sqrt{2}}^{\sqrt{2}} \left(\frac{w(x)-w(y)}{x-y} \right)^2 \frac{2-xy}{\sqrt{2-x^2}\sqrt{2-y^2}} dxdy. \end{aligned}$$

Proof. Let $\tilde{\lambda}_1, ..., \tilde{\lambda}_N \stackrel{d}{=} \sqrt{2}\lambda_1, ..., \sqrt{2}\lambda_N$ and $\tilde{\mu}_{\text{sc}}$ the measure of the semi-circle law on the interval $[-2, 2]$. By [BY05, Theorem 1.1] with $\kappa = \sigma^2 = 2$ and $\beta = 0$ it holds that for differentiable w

$$\sum_{i=1}^{N} w(\tilde{\lambda}_i) - N \int_{-2}^{2} w(x)\tilde{\mu}_{\text{sc}}(dx) \xrightarrow{d} \mathcal{N}\left(\tilde{m}(w), \tilde{v}(w)\right)$$

with expectation

$$\tilde{m}(w) = \frac{w(2) + w(-2)}{4} - \frac{1}{2\pi} \int_{-1}^{1} w(2t) \frac{1}{\sqrt{1-t^2}} dt$$

and variance

$$\tilde{v}(w) = \frac{1}{2\pi^2} \int_{-2}^{2} \int_{-2}^{2} w'(s)w'(t) \log\left(\frac{4 - ts + \sqrt{4-s^2}\sqrt{4-t^2}}{4 - ts - \sqrt{4-s^2}\sqrt{4-t^2}} \right) dsdt$$

By a change of variables we immediately get $m(w)$ from $\tilde{m}(w(\frac{1}{\sqrt{2}}\cdot))$. For the variance we can show that the expressions match by using integration by parts twice. Note that

$$\frac{\partial}{\partial y} \log\left(\frac{4 - xy + \sqrt{4-x^2}\sqrt{4-y^2}}{4 - xy - \sqrt{4-x^2}\sqrt{4-y^2}} \right) = 2 \frac{\sqrt{4-x^2}}{\sqrt{4-y^2}(x-y)}$$

and

$$\frac{\partial}{\partial x} \frac{\sqrt{4-x^2}}{\sqrt{4-y^2}(x-y)} = -\frac{4-xy}{\sqrt{4-x^2}\sqrt{4-y^2}(x-y)^2},$$

which gives

$$\begin{aligned} &\int_{-2}^{2} \int_{-2}^{2} \left(w(x) - w(y)\right)^2 \tfrac{4-xy}{(x-y)^2\sqrt{4-x^2}\sqrt{4-y^2}} dxdy \\ =& \int_{-2}^{2} \left(\left[\left(w(x) - w(y)\right)^2 \tfrac{-\sqrt{4-x^2}}{\sqrt{4-y^2}(y-x)} \right]_{x=-2}^{2} - \int_{-2}^{2} 2\left(w(x) - w(y)\right) w'(x) \tfrac{-\sqrt{4-x^2}}{\sqrt{4-y^2}(y-x)} dx \right) dy \\ =& \int_{-2}^{2} \left(\left[\left(w(x) - w(y)\right) w'(x) \log\left(\tfrac{4-xy+\sqrt{4-x^2}\sqrt{4-y^2}}{4-xy-\sqrt{4-x^2}\sqrt{4-y^2}} \right) \right]_{y=-2}^{2} \right. \\ &\qquad \left. + \int_{-2}^{2} w'(y)w'(x) \log\left(\tfrac{4-xy+\sqrt{4-x^2}\sqrt{4-y^2}}{4-xy-\sqrt{4-x^2}\sqrt{4-y^2}} \right) dy \right) dx \\ =& \int_{-2}^{2} \int_{-2}^{2} w'(y)w'(x) \log\left(\tfrac{4-xy+\sqrt{4-x^2}\sqrt{4-y^2}}{4-xy-\sqrt{4-x^2}\sqrt{4-y^2}} \right) dydx. \end{aligned}$$

By a change of variables we thus get the expression $v(w)$ from $v(w(\frac{1}{\sqrt{2}}\cdot))$. ∎

3. Reduction to a low-dimensional optimization

In this section we start the proof of Theorem 1.1 and Theorem 1.2 by applying the method of Lagrange multipliers to the original high-dimensional optimization problem and as a result reduce it to a low-dimensional optimization problem.

Recall from (1.2) that

$$L_N = \sup_{|\sigma|=1} \{\beta H_N(\sigma) + N f(v \cdot \sigma)\} \tag{3.1}$$

where v is a fixed unit vector. We have

$$\frac{1}{N} L_N = \sup_{|\sigma|=1} \left\{ \beta \sum_{i=1}^{N} \lambda_i \sigma_i^2 + f\left(\sum_{i=1}^{N} \sigma_i u_i \right) \right\}, \tag{3.2}$$

where $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_N$ are the eigenvectors of $\frac{1}{\sqrt{N}} J$ and $u = (u_1, ..., u_N)$ is v in the diagonalizing basis of J. Note that u is a random unit vector uniform on the sphere, independent of $\lambda_1, ..., \lambda_N$. We can rewrite (3.2) as

$$\frac{1}{N} L_N = \sup_{\alpha \in [-1,1]} \left\{ f(\alpha) + \beta \sup_{\substack{|\sigma|=1 \\ \sigma \cdot u = \alpha}} \sum_{i=1}^{N} \lambda_i \sigma_i^2 \right\}. \tag{3.3}$$

Similarly, using the substitution $m = r\sigma$ with $r = |m|$ for $|\sigma| = 1$ in (1.8),

$$\frac{1}{N} \tilde{L}_N = \sup_{r \in \mathcal{R}, \alpha \in [-1,1]} \left\{ f(\alpha r) + g(r) + \beta r^2 \sup_{\substack{|\sigma|=1 \\ \sigma \cdot u = \alpha}} \sum_{i=1}^{N} \lambda_i \sigma_i^2 \right\}. \tag{3.4}$$

The next lemma will in turn rewrite the inner supremum of (3.3), (3.4) in terms of the Stieltjes transform of the weighted empirical spectral measure

$$\mu_{\lambda,u} = \sum_{i=1}^{N} u_i \delta_{\lambda_i}. \tag{3.5}$$

Recall that the Stieltjes transform of a measure μ on $\mathbb{R}$ is given by

$$s_\mu(l) = \int_{\mathbb{R}} \frac{1}{l - \lambda} \mu(d\lambda), \tag{3.6}$$

for l outside the support of μ, so that

$$s_{\mu_{\lambda,u}}(l) = \sum_{i=1}^{N} \frac{u_i^2}{l - \lambda_i}. \tag{3.7}$$

In the interest of compact notation we drop the μ and write

$$s_{\lambda,u} = s_{\mu_{\lambda,u}}. \tag{3.8}$$

We can now formulate our result on the inner optimization in (3.3), which is an exact identity if $|\alpha| \geq |u_N|$ and a bound that is sufficient for our purposes if $|\alpha| < |u_N|$. This and all further results in this section hold deterministically for any $u \in \mathcal{S}_{N-1}$ with $u_1^2, ..., u_N^2 \in (0, 1)$ and $-\infty < \lambda_1 < ... < \lambda_N < \infty$.

Lemma 3.1. *For any $\lambda_1 < ... < \lambda_N$ and $u_1, ..., u_N \in (-1, 1) \setminus \{0\}$ with $\sum_{i=1}^{N} u_i^2 = 1$ it holds that if $1 > |\alpha| \geq |u_N|$ then*

$$\sup_{\substack{|\sigma|=1 \\ \sigma \cdot u = \alpha}} \sum_{i=1}^{N} \lambda_i \sigma_i^2 = \inf_{l > \lambda_N} \left\{ l - \frac{\alpha^2}{s_{\lambda,u}(l)} \right\}. \tag{3.9}$$

If $\alpha \in [-|u_N|, |u_N|]$ then

$$\lambda_N - \frac{2u_N^2}{\sqrt{1 - u_N^2}} (\lambda_N - \lambda_1) \leq \sup_{\substack{|\sigma|=1 \\ \sigma \cdot u = \alpha}} \sum_{i=1}^{N} \lambda_i \sigma_i^2 \leq \lambda_N. \tag{3.10}$$

In the proof and later we will consider the function $\varphi_N : [\lambda_N, \infty) \mapsto \mathbb{R}^+$ given by

$$\varphi_N(l) = -\frac{1}{s_{\lambda,u}(l)} \overset{(3.7)}{=} -\left(\sum_{i=1}^{N} \frac{u_i^2}{l-\lambda_i}\right)^{-1} \text{ for } l > \lambda_N, \tag{3.11}$$

which satisfies

$$\varphi_N(l) \to 0 \text{ as } l \downarrow \lambda_N \tag{3.12}$$

when $u_N \neq 0$, so that defining $\varphi_N(\lambda_N) = 0$ makes φ_N a continuous function. Note that for $l > \lambda_N$

$$\varphi_N'(l) = \frac{s_{\lambda,u}^{(1)}(l)}{s_{\lambda,u}(l)^2} \overset{(3.7)}{=} -\frac{\sum_{i=1}^N \frac{u_i^2}{(l-\lambda_i)^2}}{\left(\sum_{i=1}^N \frac{u_i^2}{l-\lambda_i}\right)^2}, \tag{3.13}$$

so if $u_N \neq 0$

$$\varphi_N'(l) \to -\frac{1}{u_N^2} \text{ for } l \downarrow \lambda_N, \tag{3.14}$$

so that φ_N is also differentiable on $[\lambda_N, \infty)$.

Proof of Lemma 3.1. Starting with the main case (3.9), note that introducing Lagrange multipliers we have

$$\sup_{\substack{|\sigma|=1\\ \sigma\cdot u=\alpha}} \sum_{i=1}^N \lambda_i \sigma_i^2 \leq \inf_{l,r\in\mathbb{R}} \sup_{\sigma\in\mathbb{R}^N} \mathscr{L}(\sigma, l, r), \tag{3.15}$$

where

$$\begin{aligned} \mathscr{L}(\sigma, l, r) &= \textstyle\sum_{i=1}^N \lambda_i\sigma_i^2 - l\left(\sum_{i=1}^N \sigma_i^2 - 1\right) - r\left(\sum_{i=1}^N \sigma_i u_i - \alpha\right) \\ &= \textstyle\sum_{i=1}^N \left((\lambda_i - l)\sigma_i^2 - r u_i \sigma_i\right) + l + \alpha r. \end{aligned}$$

Furthermore if there are some $l, r \in \mathbb{R}, \sigma \in \mathbb{R}^N$ achieving the minimax on the r.h.s. of (3.15), then these (σ, l, r) are a critical point of $\mathscr{L}$ and in fact (3.15) with equality.

When $l < \lambda_N$ then $\sup_{\sigma\in\mathbb{R}^N} \mathscr{L}(\sigma, l, r) = \infty$. The same is true for $l = \lambda_N$ and $r \neq 0$. For $l \geq \lambda_N, r = 0$ we have $\sup_{\sigma\in\mathbb{R}^N} \mathscr{L}(\sigma, l, 0) = \lambda_N$. Thus

$$\inf_{l,r: l\leq\lambda_N \text{ or } r=0} \sup_{\sigma\in\mathbb{R}^N} \mathscr{L}(\sigma, l, r) = \lambda_N. \tag{3.16}$$

Now consider the remaining case $l > \lambda_N, r \neq 0$. In this case $\sup_{\sigma\in\mathbb{R}^N} \mathscr{L}(\sigma, r, l)$ is maximized by

$$\sigma_i^*(l, r) = \frac{1}{2}\frac{r u_i}{\lambda_i - l},$$

for which

$$\mathscr{L}(\sigma^*(r, l), r, l) = l + \alpha r + \frac{1}{4}\sum_{i=1}^N \frac{r^2 u_i^2}{l - \lambda_i}.$$

Since $\alpha \neq 0$ by assumption we have

$$\inf_{r\neq 0} \sup_{\sigma\in\mathbb{R}^N} \mathscr{L}(\sigma, l, r) = \inf_{r\neq 0}\left\{l + \alpha r + \frac{r^2}{4}\sum_{i=1}^N \frac{u_i^2}{l-\lambda_i}\right\} = l - \frac{\alpha^2}{\sum_{i=1}^N \frac{u_i^2}{l-\lambda_i}},$$

where the infimum is attained at $r^*(l) = -2\frac{\alpha}{\sum_{i=1}^N \frac{u_i^2}{l-\lambda_i}} \neq 0$, for which

$$\mathscr{L}(\sigma^*(l, r^*), l, r^*(l)) = l + \alpha^2\varphi_N(l). \tag{3.17}$$

We have

$$l + \alpha^2\varphi_N(l) \geq l - \alpha^2(l - \lambda_1) \overset{|\alpha|<1}{\to} \infty \text{ as } l \uparrow \infty, \tag{3.18}$$

and recalling (3.12) we have $l + \alpha^2\varphi_N(l) \to \lambda_N$ as $l \downarrow \lambda_N$. Furthermore by (3.14)

$$\frac{d}{dl}\left\{l + \alpha^2\varphi_N(l)\right\} \to 1 - \frac{\alpha^2}{u_N^2} \overset{|\alpha|>|u_N|}{<} 0 \text{ as } l \downarrow \lambda_N, \tag{3.19}$$

so the infimum of (3.17) over $l > \lambda_N$ is attained at some $l^* > \lambda_N$, and we obtain

$$\inf_{l,r:l>\lambda_N,r\neq 0}\sup_{\sigma\in\mathbb{R}^N}\mathscr{L}(\sigma,l,r) = \mathscr{L}(\sigma^*(l^*,r^*),l^*,r^*(l^*)) = \inf_{l>\lambda_N}\left\{l - \frac{\alpha^2}{\sum_{i=1}^N \frac{u_i^2}{\lambda_i - l}}\right\} < \lambda_N.$$

Together with (3.16) this proves that the minimax in (3.15) is indeed attained at some $l^*, r^* \in \mathbb{R}, \sigma^* \in \mathbb{R}^N$, so (3.15) holds in equality and (3.9) follows.

Next considering (3.10) note that the upper bound is trivial, and the lower bound follows by plugging in

$$\sigma = \sqrt{\tfrac{1-\alpha^2}{1-u_N^2}}e_N + \left(\alpha - \sqrt{\tfrac{1-\alpha^2}{1-u_N^2}}u_N\right)u \tag{3.20}$$

where $e_N = (0, \ldots, 0, 1)$, which satisfies $|\sigma| = 1$, $\sigma \cdot u = \alpha$ and $\sigma_N^2 \geq 1 - u_N^2\left(1 + \frac{1}{\sqrt{1-u_N^2}}\right)$ so that

$$\sum_{i=1}^N \lambda_i\sigma_i^2 \geq \lambda_N\sigma_N^2 + \lambda_1(1-\sigma_N^2) = \lambda_N - u_N^2\left(1 + \frac{1}{\sqrt{1-u_N^2}}\right)(\lambda_N - \lambda_1). \tag{3.21}$$

□

We now prove a few results about this minimization problem. The next lemma shows that the map $l \to -s_\mu(l)^{-1}$ is convex for any measure μ, so in particular $\varphi_N(l)$ is convex. Note that for any μ and $k \in \mathbb{N}$

$$s_\mu^{(k)}(l) \overset{(3.6)}{=} (-1)^k k! \int_{\mathbb{R}} \frac{1}{(l-x)^{k+1}}\mu(dx). \tag{3.22}$$

Lemma 3.2 (Convexity)**.** *For any $\lambda \in \mathbb{R}$ and measure μ on $\mathbb{R}$ with support contained in $(-\infty, \lambda]$ the map $l \to -\frac{1}{s_\mu(l)}$ is convex in (λ, ∞). If the support of μ is not a singleton it is strictly convex.*

Proof. For $l > \lambda$ the second derivative equals

$$\left(-\frac{1}{s_\mu(l)}\right)'' = \frac{s_\mu''(l) - 2s_\mu'(l)^2 s_\mu(l)}{s_\mu(l)^3}.$$

Letting $w(x) = \frac{1}{l-x}$ and using (3.22) the numerator equals

$$2\int w(x)^3\mu(dx) - 2\left(\int w(x)^2\mu(dx)\right)^2 \int w(x)\mu(dx).$$

Since $w(x) > 0$ on the support of μ it holds that $\int w(x)\mu(dx) > 0$, and dividing through by this quantity we obtain

$$2\left(\frac{\int w(x)^3\mu(dx)}{\int w(x)\mu(dx)} - \left(\frac{\int w(x)^2\mu(dx)}{\int w(x)\mu(dx)}\right)^2\right).$$

This is non-positive by the Cauchy-Schwartz inequality, and equals zero only if $w(x)^2$ is constant on the support of μ, which is only the case if the support of μ is a singleton. □

The previous lemma implies the following about a general version of the minimization in (3.9).

Lemma 3.3 (Uniqueness)**.** *Let μ be a real measure with support which is not a singleton and is contained in $[\lambda_-, \lambda_+]$ for $-\infty < \lambda_- < \lambda_+ < \infty$. For any $\alpha^2 < 1$ there is a unique $l^* \geq \lambda_+$ that achieves the infimum of*

$$\inf_{l>\lambda_+}\left\{l - \frac{\alpha^2}{s_\mu(l)}\right\},$$

and $l^* > \lambda_+$ iff $\alpha^2 > \lim_{l\downarrow\lambda_+}\frac{s_\mu(l)^2}{-s_\mu'(l)}$. If $\alpha = \pm 1$ then the infimum equals $\int \lambda\mu(d\lambda)$ and is achieved for $l \to \infty$.

Proof. If $\alpha = 0$ then $l^* = \lambda_+$ is the unique minimizer. If $\alpha^2 \in (0,1)$ then $l + \alpha^2/s_\mu(l)$ is strictly convex for $l \in (\lambda_+, \infty)$ by Lemma 3.2, and similarly to (3.18) it holds that $l - \alpha^2/s_{\mu(l)} \geq l - \alpha^2(l - \lambda_-) \to \infty$ for $l \to \infty$. This implies that there is a unique minimizer in $[\lambda_+, \infty)$. The minimizer is λ_+ iff $\lim_{l \downarrow \lambda_N} \frac{d}{dl}\{l - \alpha^2/s_\mu(l)\} \geq 0$ and

$$\frac{d}{dl}\left\{l - \frac{\alpha^2}{s_\mu(l)}\right\} = 1 + \alpha^2 \frac{s_\mu'(l)}{s_\mu(l)^2}, \tag{3.23}$$

giving the condition in the statement.

For $\alpha = \pm 1$ it follows from (3.6) and (3.22) with $k = 1$ that the r.h.s of (3.23) converges to 0 for $l \to \infty$, which together with the convexity shows that the infimum is achieved for $l \to \infty$. Taylor expanding $\frac{1}{l-\lambda}$ yields

$$s_\mu(l) = \frac{1}{l} + \frac{1}{l^2}\int \lambda\mu(d\lambda) + O\left(\frac{\lambda_+^2}{l^2}\right) \text{ for } l \geq \lambda_+ + 1, \tag{3.24}$$

from which one can verify that the limit for $l \to \infty$ is $\int \lambda\mu(d\lambda)$. □

In particular for the minimization in (3.9) we obtain the following from the previous lemma and (3.13)-(3.14).

Corollary 3.4 (Uniqueness for φ_N)**.** *For any $\alpha^2 < 1, \lambda_1 < ... < \lambda_N$, $u_1^2, ..., u_N^2 \in (0,1)$ with $\sum_{i=1}^N u_i^2 = 1$ there is a unique $l^* \geq \lambda_N$ that achieves the infimum of*

$$\inf_{l \geq \lambda_N} \left\{l + \alpha^2 \varphi_N(l)\right\},$$

and $l^* > \lambda_N$ iff $\alpha^2 > u_N^2$. If $\alpha = \pm 1$ then the infimum equals $\sum_{i=1}^N u_i^2 \lambda_i$ and is achieved for $l \to \infty$.

From Lemma 3.1 and Corollary 3.4 we obtain the following.

Corollary 3.5. *For any $\lambda_1 < ... < \lambda_N$, $u_1^2, ..., u_N^2 \in (0,1)$ with $\sum_{i=1}^N u_i^2 = 1$ it holds that*

$$\sup_{\alpha \in [-1,1]} \left| \sup_{\substack{|\sigma|=1 \\ \sigma \cdot u = \alpha}} \sum_{i=1}^N \lambda_i \sigma_i^2 - \inf_{l > \lambda_N} \left\{l - \frac{\alpha^2}{s_{\lambda,u}(l)}\right\} \right| \leq 2(\lambda_N - \lambda_1) \frac{u_N^2}{\sqrt{1 - u_N^2}}.$$

Proof. Note that the difference is exactly zero for $|\alpha| \geq |u_N|$ by (3.9). For $|\alpha| < |u_N|$ it follows from (3.10) that

$$\left| \sup_{\sigma \cdot u = \alpha} \sum_{i=1}^N \lambda_i \sigma_i^2 - \lambda_N \right| \leq 2(\lambda_N - \lambda_1) \frac{u_N^2}{\sqrt{1 - u_N^2}}.$$

Also Corollary 3.4 implies that if $|\alpha| < |u_N|$ then $l^* = \lambda_N$ and therefore

$$\inf_{l > \lambda_N} \left\{l - \frac{\alpha^2}{s_{\lambda,u}(l)}\right\} = \lambda_N.$$

□

4. Leading order behavior

In this section we will study the behavior of L_N and $\tilde{L}_N$ to leading order, proving Theorem 1.1 (a) and Theorem 1.2 (a). These are in fact immediate consequences of (3.3), (3.4) and the following proposition.

Proposition 4.1. *It holds that*

$$\sup_{\alpha \in [-1,1]} \left| \sup_{\sigma \cdot u = \alpha} \sum_{i=1}^N \lambda_i \sigma_i^2 - \sqrt{2(1 - \alpha^2)} \right| \xrightarrow{\mathbb{P}} 0. \tag{4.1}$$

Thanks to Corollary 3.5 and the facts that $u_N \xrightarrow{\mathbb{P}} 0$ for u uniform on the unit sphere, and that λ_1, λ_N are stochastically bounded, this in turn is a direct consequence of

$$\sup_{\alpha\in[-1,1]} \left| \inf_{l>\lambda_N} \left\{ l - \frac{\alpha^2}{s_{\lambda,u}(l)} \right\} - \sqrt{2(1-\alpha^2)} \right| \xrightarrow{\mathbb{P}} 0. \tag{4.2}$$

The goal of the section is thus to prove (4.2) and therefore Proposition 4.1.

To do so we will show laws of large numbers for $s_{\lambda,u}(l)$ and its derivatives in the first subsection, and in the second subsection use them to compute the infimum in (4.2).

4.1. Law of large numbers for weighted Stieltjes transform

In this subsection we give a leading order estimate for $s_{\lambda,u}(l)$, showing roughly speaking that $s_{\lambda,u}(l) \to s_{\mu_{\mathrm{sc}}}(l)$. The following notations and results will also be useful later to handle the fluctuations of $s_{\lambda,u}(l)$ and $L_N, \tilde{L}_N$ in Section 6. To approximate $s_{\lambda,u}(l)$ by $s_{\mu_{\mathrm{sc}}}(l)$ we use the Stieltjes transforms of the measures

$$\mu_\lambda = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}, \qquad \mu_\theta = \frac{1}{N}\sum_{i=1}^N \delta_{\theta_{i/N}}, \qquad \mu_{\theta,u} = \frac{1}{N}\sum_{i=1}^N u_i^2 \delta_{\theta_{i/N}}, \tag{4.3}$$

where the first two are empirical measures of random eigenvalues and deterministic classical locations (recall (2.2)) respectively, and $\mu_{\theta,u}$ is a randomly weighted version of μ_θ, cf. (3.7). As we already have for the Stieltjes transform of $\mu_{\lambda,u}$ we use the abbreviations (see (2.1), (3.6))

$$\begin{aligned} s(l) = s_{\mu_{\mathrm{sc}}}(l) &= \int_{-\sqrt{2}}^{\sqrt{2}} \frac{\frac{1}{\pi}\sqrt{2-x^2}}{l-x} dx, \qquad & s_\lambda(l) = s_{\mu_\lambda}(l) &= \frac{1}{N}\sum_{i=1}^N \frac{1}{l-\lambda_i}, \\ s_\theta(l) = s_{\mu_\theta}(l) &= \frac{1}{N}\sum_{i=1}^N \frac{1}{l-\theta_{i/N}}, \qquad & s_{\theta,u}(l) = s_{\mu_{\theta,u}}(l) &= \frac{1}{N}\sum_{i=1}^N \frac{u_i^2}{l-\theta_{i/N}}. \end{aligned} \tag{4.4}$$

The integral for $s(l)$ in (4.4) can be computed explicitly yielding the following useful identities

$$\begin{aligned} s(l) &= l - \sqrt{l^2-2}, \qquad & s^{(1)}(l) &= -\tfrac{l-\sqrt{l^2-2}}{\sqrt{l^2-2}}, \\ s^{(2)}(l) &= \tfrac{2}{(l^2-2)^{\frac{3}{2}}}, \qquad & s^{(3)}(l) &= -\tfrac{6l}{(l^2-2)^{\frac{5}{2}}}, \end{aligned} \tag{4.5}$$

for all $k \in \mathbb{N}$ and $l > \sqrt{2}$. The two identities on the top row play a role in the study of the leading order here, and the higher derivatives on the bottom row will play a role in the study of the fluctuations in Section 6. Note that

$$s(\sqrt{2}) = \sqrt{2}, \qquad s(l) \text{ is decreasing on } [\sqrt{2},\infty), \text{ and } \qquad \lim_{l\to\infty} s(l) = 0. \tag{4.6}$$

We record the following direct consequence of Lemma 2.1, comparing weighted sums over eigenvalues with the corresponding sum over classical locations.

Lemma 4.2. *For any $\delta > 0$ we have*

$$\mathbb{P}\left(\forall w \in C^1([-\sqrt{2}-\varepsilon, \sqrt{2}+\varepsilon]), u \in \mathcal{S}_{N-1} : \left| \sum_{i=1}^N u_i^2 w(\lambda_i) - \sum_{i=1}^N u_i^2 w(\theta_{i/N}) \right| \le |w'|_\infty N^{-\frac{2}{3}+\delta} \right) \to 1. \tag{4.7}$$

The following approximations are a consequence of the previous lemma and Lemma 2.2.

Lemma 4.3. *Let $\varepsilon, \delta > 0$ and $k \in \mathbb{N}$. It holds uniformly for all $l > \sqrt{2} + \varepsilon$ that*

$$\left| s_\theta^{(k)}(l) - s^{(k)}(l) \right| = O\left(\frac{1}{N}\right), \tag{4.8}$$

and

$$s_\theta^{(k)}(l) = s_\lambda^{(k)}(l) + O_{\mathbb{P}}\left(N^{-\frac{2}{3}+\delta}\right), \tag{4.9}$$

$$s_{\theta,u}^{(k)}(l) = s_{\lambda,u}^{(k)}(l) + O_{\mathbb{P}}\left(N^{-\frac{2}{3}+\delta}\right). \tag{4.10}$$

Proof. Let $w(l,\theta) = \frac{1}{l-\theta}$ and fix some $k \in \mathbb{N}$. By Lemma 2.2

$$\left|s_\theta^{(k)}(l) - s^{(k)}(l)\right| \leq \frac{\sup_{x\in[-\sqrt{2},\sqrt{2}]} |w^{(k+1)}(l,x)|}{N} \leq \frac{1}{N}\frac{(k+1)!}{(l-\sqrt{2})^{k+2}} \leq \frac{1}{N}\frac{(k+1)!}{\varepsilon^{k+2}} \tag{4.11}$$

for all $l \geq \sqrt{2}+\varepsilon$, which implies (4.8). On the event that $\lambda_N \leq \sqrt{2}+\frac{\varepsilon}{2}$ we have by (4.7) that for any $\delta > 0$

$$\mathbb{P}\left(\forall l \geq \sqrt{2}+\varepsilon: \ \left|s_{\theta,u}^{(k)}(l) - s_{\lambda,u}^{(k)}(l)\right| \leq \frac{2(k+1)!}{\varepsilon^{k+2}} N^{-\frac{2}{3}+\delta}\right) \to 1,$$

implying (4.10). The same argument for $|s_\theta^{(k)}(l) - s_\lambda^{(k)}(l)|$ proves (4.9). □

The following lemma gives a law of large numbers for sums over the classical locations or eigenvalues, weighted by the random $u_1^2, ..., u_N^2$. It implies in particular that $s_{\theta,u}^{(k)}(l) \to s^{(k)}(l)$ and $s_{\lambda,u}^{(k)}(l) \to s^{(k)}(l)$ in probability.

Lemma 4.4. *Let $\varepsilon > 0$ and $w \in C^1([-\sqrt{2}-\varepsilon, \sqrt{2}+\varepsilon])$. Let u be a random vector uniformly distributed on the sphere. Then as $N \to \infty$*

$$\sum_{i=1}^N w(\theta_{i/N}) u_i^2 \xrightarrow{\mathbb{P}} \int_{-\sqrt{2}}^{\sqrt{2}} w(x)\mu_{sc}(dx), \tag{4.12}$$

and

$$\sum_{i=1}^N w(\lambda_i) u_i^2 \xrightarrow{\mathbb{P}} \int_{-\sqrt{2}}^{\sqrt{2}} w(x)\mu_{sc}(dx). \tag{4.13}$$

Proof. Construct u by setting $u_i = \frac{\tilde{u}_i}{|\tilde{u}|}$ with $\tilde{u}_1, ..., \tilde{u}_N \sim \mathcal{N}(0, \frac{1}{N})$ i.i.d.. We then have

$$\mathbb{E}\left[\sum_{i=1}^N w(\theta_{i/N})\tilde{u}_i^2\right] = \frac{1}{N}\sum_{i=1}^N w(\theta_{i/N}) = \int_{-\sqrt{2}}^{\sqrt{2}} w(x)\mu_{\mathrm{sc}}(dx) + o(1),$$

by Lemma 2.2 and since $\mathrm{Var}\left(N\tilde{u}_i^2\right) = 2$

$$\mathrm{Var}\left(\sum_{i=1}^N w(\theta_{i/N})\tilde{u}_i^2\right) = \frac{2}{N^2}\sum_{i=1}^N w(\theta_{i/N}) = O\left(\frac{1}{N}\right). \tag{4.14}$$

Therefore

$$\sum_{i=1}^N w(\theta_{i/N})\tilde{u}_i^2 \xrightarrow{\mathbb{P}} \int_{-\sqrt{2}}^{\sqrt{2}} w(x)\mu_{\mathrm{sc}}(dx) \text{ as } N \to \infty,$$

and since also $|\tilde{u}| \to 1$ in probability the claim (4.12) follows. The second claim then follows from Lemma 4.2. □

The previous lemma implies the following uniform convergence of $s_{\theta,u}^{(k)}(l)$ to $s^{(k)}(l)$.

Lemma 4.5. *Let $\varepsilon, \delta > 0$ and $k \in \mathbb{N}$. For any $\varepsilon > 0, L > 2$*

$$\sup_{l\in[\sqrt{2}+\varepsilon, L]} \left|s_{\theta,u}^{(k)}(l) - s^{(k)}(l)\right| = o_{\mathbb{P}}(1).$$

Proof. Firstly, by Lemma 4.4 and a union bound it holds for all $\delta > 0$ that

$$\lim_{N\to\infty} \sup_{l\in[\sqrt{2}+\varepsilon,L]\cap\delta\mathbb{Z}} \mathbb{P}\left(\left|s_{\theta,u}^{(k)}(l) - s^{(k)}(l)\right| \geq \delta\right) = 0.$$

Secondly, since $l \to \frac{1}{(l-\theta_{i/N})^k}$ is Lipschitz for $l \geq \sqrt{2}+\varepsilon$ so are $s_{\theta,u}^{(k)}(l)$ and $s^{(k)}(l)$. These two facts imply the claim. ☐

Remark 4.6. *(a) Though we do not need it here, it is easy to argue that the convergence is uniform on* $[\sqrt{2}+\varepsilon,\infty)$*, since* $\lim_{l\to\infty} s_\mu^{(k)}(l) = 0$ *for all* k *and* μ *with compact support. (b) In Section 6.2 we strengthen the bound to* $O_{\mathbb{P}}(N^{-1/2})$*, as this is needed to study the fluctuations of* L_N *and* $\tilde{L}_N$ *(see (6.26)).*

The estimate (4.10) and Lemma 4.5 together imply that for all $\varepsilon > 0, L > 2$,

$$s_{\lambda,u}(l) \to s(l) \text{ uniformly in probability on } [\sqrt{2}+\varepsilon, L]. \tag{4.15}$$

The next lemma deduces from this that also $s_{\lambda,u}(l)^{-1} \to s(l)^{-1}$ uniformly, and here we do take care to prove it for an unbounded interval.

Lemma 4.7. *For all* $\varepsilon > 0$ *it holds that*

$$\sup_{l\geq\sqrt{2}+\varepsilon} \left|\frac{1}{s_{\lambda,u}(l)} - \frac{1}{s(l)}\right| \xrightarrow{\mathbb{P}} 0.$$

Proof. From (4.5) it follows that $s(l) = l^{-1} + O\left(l^{-3}\right)$ for l large. Similarly from (3.24)

$$s_{\lambda,u}(l) = \frac{1}{l} + \frac{O\left(\sum_{i=1}^N u_i^2\lambda_i\right)}{l^2} + O\left(\frac{\max(|\lambda_1|,|\lambda_N|)^3}{l^3}\right),$$

for all $u \in \mathcal{S}_{N-1}$ and $l \geq \lambda_N + 1$. By Lemma 4.4 with $w(x) = x$ it holds that $\sum_{i=1}^N u_i^2\lambda_i \to \int_{-\sqrt{2}}^{\sqrt{2}} x\mu_{\mathrm{sc}}(dx) = 0$ in probability, and since also λ_1, λ_N are stochastically bounded it follows that for each $\eta > 0$ there is a large enough L such that

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{l\geq L}\left|\frac{1}{s_{\lambda,u}(l)} - \frac{1}{s(l)}\right| \geq \frac{\eta}{2}\right) = 0.$$

Furthermore (4.15) implies that

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{l\in[\sqrt{2}+\varepsilon,L]}\left|\frac{1}{s_{\lambda,u}(l)} - \frac{1}{s(l)}\right| \geq \frac{\eta}{2}\right) = 0,$$

giving the claim. ☐

4.2. Leading order estimate for Lagrange optimization

We now use the laws of large numbers to study the optimization problem

$$\inf_{l>\lambda_N}\left\{l - \frac{\alpha^2}{s_{\lambda,u}(l)}\right\}, \tag{4.16}$$

from (4.2). The law of large numbers $s_{\lambda,u}(l) \to s(l)$ leads us to consider the limiting optimization problem

$$\inf_{l>\sqrt{2}}\left\{l - \frac{\alpha^2}{s(l)}\right\}. \tag{4.17}$$

The next lemma solves this limiting optimization.

Lemma 4.8. *For all $\alpha \in [-1,1]$*

$$\inf_{l>\sqrt{2}} \left\{ l - \frac{\alpha^2}{s(l)} \right\} = \sqrt{2(1-\alpha^2)}, \tag{4.18}$$

and if $\alpha \in (-1,1)$ the unique minimizer is

$$\hat{l}(\alpha) = \frac{2-\alpha^2}{\sqrt{2(1-\alpha^2)}}, \tag{4.19}$$

while if $\alpha = \pm 1$ the infimum is achieved for $l \to \infty$.

Proof. We have

$$\frac{d}{dl}\left\{ l - \frac{\alpha^2}{s(l)} \right\} = 1 + \alpha^2 \frac{s'(l)}{s(l)^2} \overset{(4.5)}{=} 1 - \alpha^2 \frac{1}{\left(l - \sqrt{l^2-2}\right)\sqrt{l^2-2}} = 1 - \frac{\alpha^2}{1-x^2}, \tag{4.20}$$

where the last equality comes from the change of variables $l = \frac{1}{\sqrt{2}}\left(x + x^{-1}\right)$ for which $\sqrt{l^2-2} = \frac{1}{\sqrt{2}}\left(x - x^{-1}\right)$. If $\alpha \in (-1,1)$ the critical point equation thus has unique solution $x = \sqrt{1-\alpha^2}$ which yields (4.19). The claim for $\alpha^2 = 1$ follows from the general Lemma 3.3, or since the derivative (4.20) is negative for all $l > \sqrt{2}$. □

We recognize on the r.h.s. of (4.18) the term that (4.2) claims is the limit of (4.16). To prove (4.2) we thus need to approximate the random optimization (4.16) by the limiting (4.17).

From the explicit formula (4.19) it follows that minimizer in the limiting problem (4.17) is bounded away from $\sqrt{2}$ if α is bounded away from zero, and bounded if α is bounded away from ± 1. Formally, for all δ there exists a $\varepsilon > 0$ such that $\hat{l}(\alpha) \geq \sqrt{2} + \varepsilon$ if $|\alpha| \geq \delta$ and $\hat{l}(\alpha) \leq \varepsilon^{-1}$ if $|\alpha| \leq 1 - \delta$, and thus

$$\inf_{\lambda>\sqrt{2}} \left\{ l - \frac{\alpha^2}{s(l)} \right\} = \begin{cases} \inf\limits_{\lambda \geq \sqrt{2}+\varepsilon} \left\{ l - \frac{\alpha^2}{s(l)} \right\} & \text{if } |\alpha| \geq \delta, \\ \inf\limits_{\lambda \in [\sqrt{2}, \varepsilon^{-1}]} \left\{ l - \frac{\alpha^2}{s(l)} \right\} & \text{if } |\alpha| \leq 1-\delta. \end{cases} \tag{4.21}$$

The next lemma shows that this also holds for the random optimization problem (4.16).

Lemma 4.9. *Let $y(\alpha, l) = l - \frac{\alpha^2}{s_{\lambda,u}(l)}$. For each $\delta > 0$ there is an $\varepsilon > 0$ such that*

$$\lim_{N\to\infty} \mathbb{P}\left(\inf_{l \geq \lambda_N} y(\alpha, l) = \inf_{l \geq \sqrt{2}+\varepsilon} y(\alpha, l),\ \forall \alpha : |\alpha| \geq \delta \right) = 1 \tag{4.22}$$

and

$$\lim_{N\to\infty} \mathbb{P}\left(\inf_{l \geq \lambda_N} y(\alpha, l) = \inf_{l \in [\lambda_N, \varepsilon^{-1}]} y(\alpha, l),\ \forall \alpha : |\alpha| \leq 1-\delta \right) = 1. \tag{4.23}$$

Proof. For any $l > \sqrt{2}$ and α

$$\partial_l y\left(\alpha, l\right) = 1 + \alpha^2 \frac{s^{(1)}_{\lambda,u}(l)}{s_{\lambda,u}(l)^2} \xrightarrow[\text{Lem } 4.4]{\mathbb{P}} 1 + \alpha^2 \frac{s^{(1)}(l)}{s(l)^2} = 1 - \frac{\alpha^2}{1 - \frac{1}{2}s(l)^2} =: t(\alpha, l)$$

where the final expression follows by (4.20), since inverting the change of variables $l = \frac{1}{\sqrt{2}}(x + x^{-1})$ used there yields $x = \frac{1}{\sqrt{2}}(l - \sqrt{l^2-2}) = \frac{1}{\sqrt{2}}s(l)$. By (4.6) the r.h.s. tends to $-\infty$ if $l \downarrow \sqrt{2}$ and $\alpha \neq 0$, and to $1 - \alpha^2 > 0$ if $l \uparrow \infty$ and $|\alpha| < 1$. Thus there is an $\varepsilon > 0$ small enough so that

$$t(\delta, \sqrt{2}+\varepsilon) < 0 \qquad \text{and} \qquad t(1-\delta, \varepsilon^{-1}) > 0.$$

Since $\frac{s^{(1)}_{\lambda,u}(l)}{s_{\lambda,u}(l)^2}$ is negative for all $l > \lambda_N$ (see e.g. (3.13)) we have $\partial_l y(\alpha, \sqrt{2}+\varepsilon) \leq \partial_l y(\delta, \sqrt{2}+\varepsilon)$ for $|\alpha| \geq \delta$ on the event $\sqrt{2}+\varepsilon > \lambda_N$ (which has probability tending to one). It follows that

$$\lim_{N\to\infty} \mathbb{P}\left(\partial_l y(\alpha, \sqrt{2}+\varepsilon) < 0, \quad \forall \alpha : |\alpha| \geq \delta \right) = 1.$$

Since $y(\alpha, l)$ is almost surely convex in $l > \lambda_N$ by Lemma 3.2 the claim (4.22) follows. The claim (4.23) follows similarly since $\partial_l y(\alpha, \sqrt{2}+\varepsilon) \geq \partial_l y(1-\delta, \sqrt{2}+\varepsilon)$ for $|\alpha| \leq 1-\delta$ (if $\sqrt{2}+\varepsilon > \lambda_N$), so that

$$\lim_{N\to\infty} \mathbb{P}\left(\partial_l y(\alpha, \varepsilon^{-1}) > 0, \quad \forall \alpha : |\alpha| \leq 1-\delta\right) = 1.$$

□

We can now compute (4.16) for α bounded away from zero.

Lemma 4.10. *For all $\delta > 0$*

$$\sup_{\alpha\in[-1,1]:\delta\leq|\alpha|} \left| \inf_{l>\lambda_N} \left\{ l - \frac{\alpha^2}{s_{\lambda,u}(l)} \right\} - \sqrt{2(1-\alpha^2)} \right| \xrightarrow{\mathbb{P}} 0. \tag{4.24}$$

Proof. If we pick ε small enough depending on δ then by Lemma 4.7 and (4.22)

$$\sup_{\alpha\in[-1,1]:|\alpha|\geq\delta} \left| \inf_{l>\lambda_N} \left\{ l - \frac{\alpha^2}{s_{\lambda,u}(l)} \right\} - \inf_{l\geq\sqrt{2}+\varepsilon} \left\{ l - \frac{\alpha^2}{s(l)} \right\} \right| \xrightarrow{\mathbb{P}} 0,$$

while by (4.21) and (4.18) also

$$\inf_{l\geq\sqrt{2}+\varepsilon} \left\{ l - \frac{\alpha^2}{s(l)} \right\} = \inf_{l>\sqrt{2}} \left\{ l - \frac{\alpha^2}{s(l)} \right\} = \sqrt{2(1-\alpha^2)} \text{ for all } |\alpha| \geq \delta.$$

□

Next we estimate (4.16) for α close to zero.

Lemma 4.11. *There is a universal constant c such that for all $\delta > 0$*

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{\alpha\in[-1,1]:|\alpha|\leq\delta} \left| \inf_{l>\lambda_N} \left\{ l - \frac{\alpha^2}{s_{\lambda,u}(l)} \right\} - \sqrt{2(1-\alpha^2)} \right| \geq c\delta \right) = 0. \tag{4.25}$$

Proof. If $|\alpha| \leq \delta$ then

$$\sqrt{2} \geq \inf_{l>\lambda_N} \left\{ l - \frac{\alpha^2}{s_{\lambda,u}(l)} \right\} \geq \inf_{l>\lambda_N} \left\{ l - \frac{\delta^2}{s_{\lambda,u}(l)} \right\} \xrightarrow{\mathbb{P}} \sqrt{2(1-\delta^2)} = \sqrt{2} + O(\delta),$$

where we used Lemma 4.10. Since also $\sqrt{2(1-\alpha^2)} = \sqrt{2} + O(\delta)$ this implies (4.25). □

Proof of Proposition 4.1. The convergence (4.2) is a consequence of Lemma 4.10 and Lemma 4.11. By Corollary 3.5 this proves (4.1), since $u_N \to 0$ in probability and λ_1, λ_N are stochastically bounded (see Lemma 2.1). □

This also completes the proofs of Theorem 1.1 (a) and Theorem 1.2 (a), since as already mentioned the former follows from (3.3) and Proposition 4.1, while the latter follows from (3.4) and Proposition 4.1

5. Examples: Leading order

In this section we consider some important special cases where specific choices are made for f and g and characterize the maximizing α and r as explicitly as possible. Later after proving Theorems 1.1 (b) resp. 1.2 (b) about fluctuations we will see that they also apply to these examples.

Recall

$$\mathcal{B}(\alpha) = f(\alpha) + \sqrt{2}\beta\sqrt{1-\alpha^2}.$$

We will first consider the ground state L_N on the sphere for monomials $f(x) = hx^k$. Define for $h > 0$

$$\beta_c(k,h) := \begin{cases} \infty & \text{for } k = 1, \\ \sqrt{2}h & \text{for } k = 2, \\ \frac{h}{\sqrt{2}}\frac{k-1}{k-2}\left(1 - \frac{1}{(k-1)^2}\right)^{\frac{k}{2}} & \text{for } k \geq 3. \end{cases} \tag{5.1}$$

Let also for $k \geq 3$

$$\tilde{\beta}_c(k,h) = \frac{hk}{\sqrt{2}} \frac{(k-2)^{\frac{k-2}{2}}}{(k-1)^{\frac{k-1}{2}}} > \beta_c(k,h). \tag{5.2}$$

The next lemma shows for monomial f that $\mathcal{B}(\alpha)$ has a unique maximizer $\hat{\alpha}$ for $\beta \neq \beta_c(k,h)$, where $\hat{\alpha} = 0$ if $\beta > \beta_c(k,h)$ and $\hat{\alpha} > 0$ if $\beta < \beta_c(k,h)$.

Lemma 5.1 (Ground state on sphere for monomials)**.** *When $k = 1$ then for all $\beta > 0$*

$$\sup_{\alpha \in [-1,1]} \mathcal{B}(\alpha) = \sqrt{h^2 + 2\beta^2},$$

and the unique local and global maximizer of $\mathcal{B}(\alpha)$ is $\alpha = \frac{h}{\sqrt{h^2+2\beta^2}}$.

When $k = 2$ and $\beta \geq \beta_c(2,h)$ the unique local and global maximizer of $\mathcal{B}(\alpha)$ is $\alpha = 0$ and when $\beta < \beta_c(2,h)$

$$\sup_{\alpha \in [-1,1]} \mathcal{B}(\alpha) = h + \frac{\beta^2}{2h},$$

and the unique local and global maximizers of $\mathcal{B}(\alpha)$ are $\alpha = \pm\sqrt{1 - \frac{\beta^2}{2h^2}}$.

When $k \geq 3$ and $\beta \geq \tilde{\beta}_c(k,h)$ the unique local and global maximizer of $\mathcal{B}(\alpha)$ is $\alpha = 0$. When $\beta < \tilde{\beta}_c(k,h)$ let $\hat{\alpha}$ be the largest solution to

$$\alpha^{2(k-2)}\left(1 - \alpha^2\right) = 2\left(\frac{\beta}{hk}\right)^2, \tag{5.3}$$

which is the unique solution to the equation in $\left(\sqrt{\frac{k-2}{k-1}}, 1\right)$. Then $\alpha = 0, \alpha = \hat{\alpha}$ are the only local maximizers of $\mathcal{B}(\alpha)$ in $[0,1]$. When $\beta > \beta_c(k,h)$ the global maximizer is $\alpha = 0$ and when $\beta = \beta_c(k,h)$ both $\alpha = 0$ and $\alpha = \hat{\alpha}$ are global maximizers, and when $\beta < \beta_c(k,h)$ the global maximizer in $[0,1]$ is $\hat{\alpha}$.

When $k \geq 4$ and k even then $\alpha = -\hat{\alpha}$ is also local resp. global maximizer and the unique one in $[-1,0)$, and if $k \geq 3$ and k odd then there are no local maximizers in $[-1,0)$.

Remark 5.2. *Also when $k = 1,2$ and $\beta > \beta_c(k,h)$ the unique global maximizer is a solution of* (5.3) *(in fact the unique solution).*

Proof. Since $\mathcal{B}'(\alpha) \to -\infty$ for $\alpha \to \pm 1$ a non-negative maximizer must exist and it must be a local maximizer of $\mathcal{B}(\alpha)$ in $(-1,1)$. We have

$$\mathcal{B}'(\alpha) = hk\alpha^{k-1} - \sqrt{2}\beta\frac{\alpha}{\sqrt{1-\alpha^2}}.$$

For k odd we have $\mathcal{B}'(\alpha) < 0$ for $\alpha \in (-1,0)$, so there are no local maximizers in that interval. If k is even and thus $\mathcal{B}$ is symmetric, every local or global maximizer $-\alpha < 0$ must correspond to $+\alpha > 0$ that is also a local resp. global maximizum of $\mathcal{B}$. Thus we may now restrict attention to $\alpha \in [0,1]$.

For $k = 1$ and all $\beta > 0$ we have that $\mathcal{B}'(\alpha) = 0 \iff h - \sqrt{2}\beta\frac{\alpha}{\sqrt{1-\alpha^2}} = 0$ has the unique solution $\frac{h}{\sqrt{h^2+2\beta^2}}$ which must then be the unique local and global maximizer of $\mathcal{B}(\alpha)$, and indeed $\mathcal{B}(\frac{h}{\sqrt{h^2+2\beta^2}}) = h^2 + 2\beta^2$. This completes the proof in the case $k = 1$.

For $k \geq 2$ we will use that

$$\mathcal{B}''(\alpha) = hk(k-1)\alpha^{k-2} - \sqrt{2}\beta\frac{1}{(1-\alpha^2)^{3/2}}.$$

When $k=2$ then $\mathcal{B}'(0)=0$ for all β. If $\beta\geq\beta_c(k,h)$ then $\mathcal{B}'(\alpha)=0\iff 2h\alpha-\sqrt{2}\beta\frac{\alpha}{\sqrt{1-\alpha^2}}=0$ has no non-zero solutions, so $\alpha=0$ is the unique local and global maximizer. If $k=2$ and $\beta<\beta_c(k,h)$ then the unique positive solution of $\mathcal{B}'(\alpha)=0$ is $\sqrt{1-\frac{\beta^2}{2h^2}}$, and

$$\mathcal{B}\left(\sqrt{1-\frac{\beta^2}{2h^2}}\right)=h+\frac{\beta^2}{2h^2}=\frac{\beta}{\sqrt{2}}\left(\frac{\sqrt{2}h}{\beta}+\frac{\beta}{\sqrt{2}h}\right)>\sqrt{2}\beta=\mathcal{B}(0),$$

so this is the global maximum. Also $\mathcal{B}''(0)=2h-\sqrt{2}\beta>0$ so $\alpha=0$ is a local minimizer. This completes the proof in the case $k=2$.

If $k\geq 3$ then $\alpha=0$ is always a local maximizer of $\mathcal{B}(\alpha)$. Also the l.h.s. of (5.3) is maximized at $\alpha=\sqrt{\frac{k-2}{k-1}}$, so when $\beta>\tilde{\beta}_c(k,h)$ then using (5.2) the l.h.s. of (5.3) is smaller than the r.h.s. for all α, so the equation has no solutions and $\alpha=0$ is the unique maximizer. When $\beta=\tilde{\beta}_c(k,h)$ it has a single solution at $\alpha=\sqrt{\frac{k-2}{k-1}}$ and otherwise one in $\left(0,\sqrt{\frac{k-2}{k-1}}\right)$ and one in $\left(\sqrt{\frac{k-2}{k-1}},1\right)$. At any solution α of $\mathcal{B}'(\alpha)=0$ we have that

$$\begin{aligned}\mathcal{B}''(\alpha)&=(k-1)\sqrt{2}\beta\tfrac{1}{\sqrt{1-\alpha^2}}-\sqrt{2}\beta\tfrac{1}{(1-\alpha^2)^{3/2}}\\&=\tfrac{\sqrt{2}\beta}{\sqrt{1-\alpha^2}}\left(k-1-\tfrac{1}{1-\alpha^2}\right)\begin{cases}<0 & \text{if } \alpha>\sqrt{\frac{k-2}{k-1}},\\=0 & \text{if } \alpha=\sqrt{\frac{k-2}{k-1}},\\>0 & \text{if } \alpha<\sqrt{\frac{k-2}{k-1}}.\end{cases}\end{aligned}$$

This shows that when $\beta=\tilde{\beta}_c(k,h)$ we have that $\alpha=\sqrt{\frac{k-2}{k-1}}$ is a saddle point (using that $\alpha=0$ is a local maximizer and $\mathcal{B}'(\alpha)\to-\infty$ for $\alpha\to 1$), and when $\beta<\tilde{\beta}_c(k,h)$ the smaller solution is a local minimizer and the larger one is local maximizer. It only remains to check which of the two local maximizers is the global maximizer when $\beta<\tilde{\beta}_c(k,h)$.

To this end note that

$$\mathcal{B}(\alpha)>\mathcal{B}(0)\iff\frac{\alpha^k}{1-\sqrt{1-\alpha^2}}>\frac{\sqrt{2}\beta}{h}.$$

The left-hand side is uniquely maximized at $\tilde{\alpha}=\frac{\sqrt{k(k-2)}}{k-1}$. Thus if $\beta>\beta_c(k,h)$ so that $\frac{\tilde{\alpha}^k}{1-\sqrt{1-\tilde{\alpha}^2}}<\frac{\sqrt{2}\beta}{h}$ the global maximizer is $\alpha=0$, and if $\beta=\beta_c(k,h)$ we have $\mathcal{B}(\tilde{\alpha})=\mathcal{B}(0)$ and $\mathcal{B}(\alpha)<\mathcal{B}(0)$ for all $\alpha\in(0,1)\backslash\{\tilde{\alpha}\}$ so both $\alpha=0$ and $\alpha=\tilde{\alpha}$ are global maximizers, and the latter is the aforementioned non-zero local maximizer. Lastly if $\beta<\beta_c(k,h)$ then the global maximizer is non-zero and is the aforementioned non-zero local maximizer. This completes the proof for $k\geq 3$. □

We will now study an important special case of $\tilde{L}_N$. Recall the TAP free energy

$$F_{\mathrm{TAP}}(m)=\beta H_N(m)+Nf(u\cdot m)+Ng(|m|),$$

where $\beta\geq 0$ and

$$g(x)=\frac{1}{2}\log\left(1-x^2\right)+\frac{\beta^2}{2}(1-x^2)^2 \text{ for } x\geq 0.$$

Let $q_P=\max(1-\frac{1}{\sqrt{2}\beta},0)$ and define the *Plefka region*

$$\mathrm{Plef}(\beta)=[\sqrt{q_P},1]\subset[0,1],\tag{5.4}$$

and denote its interior by $\mathrm{Plef}(\beta)^{\circ}$. In TAP analysis one is interested in the maximum of F_{TAP} for m such that $|m|\in\mathrm{Plef}(\beta)$, that is in $\tilde{L}_N$ for this g and $\mathcal{R}=\mathrm{Plef}(\beta)$. Let $h>0, f(x)=hx^k$ for $k\geq 1$ and define

$$\tilde{\mathcal{B}}(\alpha,r)=f(r\alpha)+\sqrt{2}\beta r^2\sqrt{1-\alpha^2}+g(r),\tag{5.5}$$

so that by (1.10)

$$\frac{1}{N}\tilde{L}_N \xrightarrow{\mathbb{P}} \sup_{r\in \mathrm{Plef}(\beta),\alpha\in[-1,1]} \tilde{\mathcal{B}}(\alpha,r).$$

In the rest of the section we will compute the r.h.s. explicitly as possible, and show that except for critical values of β, h it has a unique maximizer.

Lemma 5.3 (TAP maximizer with linear external field)**.** *Let $h>0$, $\beta>0$ and $f(x)=hx$. It holds hat*

$$\sup_{r\in \mathrm{Plef}(\beta),\alpha\in[-1,1]} \tilde{\mathcal{B}}(\alpha,r) = \sup_{q\in[q_P,1)} \mathscr{B}(q)\,, \tag{5.6}$$

where

$$\mathscr{B}(q) = \sqrt{h^2 q + 2\beta^2 q^2} + \frac{1}{2}\log(1-q) + \frac{\beta^2}{2}(1-q)^2,$$

is a concave function in $[q_P,1)$ whose unique maximizer $\hat{q}$ is the unique solution to

$$\frac{q}{h^2+2q\beta^2} = (1-q)^2 \tag{5.7}$$

in $(q_P,1)$. Furthermore the unique maximizer of the l.h.s. of (5.6) is $\hat{r}=\sqrt{\hat{q}}$ and $\hat{\alpha}=\frac{h}{\sqrt{h^2+2\beta^2\hat{q}}}$.

Proof. We will first maximize $\tilde{\mathcal{B}}$ in α for fixed $r\neq 0$. Since

$$\partial_\alpha \tilde{\mathcal{B}}(\alpha,r) = hr - \sqrt{2}\beta r^2 \frac{\alpha}{\sqrt{1-\alpha^2}} \to -\infty \text{ for } \alpha\to\pm 1$$

a maximizer must exist and be a critical point. The critical point equation $\partial_\alpha \tilde{\mathcal{B}}(\alpha,r)=0$ has the unique solution

$$\alpha_r := \frac{h}{\sqrt{h^2+2\beta^2 r^2}} \tag{5.8}$$

which maximizes $\tilde{\mathcal{B}}(\cdot,r)$. This implies

$$\sup_{\alpha\in[-1,1]} \tilde{\mathcal{B}}(\alpha,r) = \tilde{\mathcal{B}}(\alpha_r,r) = r\sqrt{h^2+2\beta^2 r^2} + g(r), \tag{5.9}$$

(also when $r=0$ since then all three expressions are identically $\beta^2/2$). With the change of variables $q=r^2$ we get

$$\tilde{\mathcal{B}}(\alpha_r,r) = \mathscr{B}(q) := t(q) + g(\sqrt{q}) \tag{5.10}$$

where

$$t(q) = \sqrt{h^2 q + 2\beta^2 q^2},$$

and

$$g(\sqrt{q}) = \frac{1}{2}\log(1-q) + \frac{\beta^2}{2}(1-q)^2. \tag{5.11}$$

We have thus proved (5.6).

Furthermore we have

$$t'(q) = \frac{h^2+4\beta^2 q}{2\sqrt{h^2 q+2\beta^2 q^2}} \text{ and } t''(q) = \frac{2\beta^2}{\sqrt{h^2 q+2\beta^2 q^2}} - \frac{(h^2+4\beta^2 q)^2}{4(h^2 q+2\beta^2 q^2)^{\frac{3}{2}}}.$$

Since $2\beta^2(h^2 q+2\beta^2 q^2) < (h^2+4\beta^2 q)^2$ for all $q\in[0,1]$ one sees that $t''(q)<0$, so t is strictly concave. Also

$$\frac{\partial}{\partial q} g(\sqrt{q}) = -\beta^2(1-q) - \frac{1}{2(1-q)} \quad\text{and}\quad \frac{\partial^2}{\partial q^2} g(\sqrt{q}) = \beta^2 - \tfrac{1}{2(1-q)^2},$$

and the latter is negative for $q\in(q_P,1)$, so

$$q\to g(\sqrt{q}) \text{ is strictly concave in } [q_p,1]. \tag{5.12}$$

Thus also $\mathscr{B}(q)$ is strictly concave in $[q_P,1)$. This implies that $\tilde{\mathcal{B}}(q)$ has a unique maximizer $\hat{q}$ in $[q_P,1)$, and $\hat{r}=\sqrt{\hat{q}}$ is the unique maximizer of $r\to\tilde{\mathcal{B}}(\alpha_r,r)$ in Plef (β), and $(\sqrt{\hat{q}},\frac{h}{\sqrt{h^2+2\beta^2\hat{q}}})$ is the unique maximizer of the l.h.s. of (5.6).

Thus it only remains to derive the equation (5.7) for $\hat{q}$. For this it suffices to note that with $v(x)=x+x^{-1}$ we have the identities

$$t'(q)=\frac{\beta}{\sqrt{2}}v\left(\frac{\sqrt{2q}\beta}{\sqrt{h^2+2q\beta^2}}\right)\text{ and }\frac{\partial}{\partial q}g(\sqrt{q})=-\frac{\beta}{\sqrt{2}}v\left(\sqrt{2}\beta\,(1-q)\right).\tag{5.13}$$

Therefore the critical point equation $\mathscr{B}'(q)=0$ is equivalent to $v\left(\frac{\sqrt{2q}\beta}{\sqrt{h^2+2q\beta^2}}\right)=v(\sqrt{2}\beta\,(1-q))$ and since $v(x)$ is a bijection for $x\in[0,1]$ this is in turn equivalent to $\frac{\sqrt{2q}\beta}{\sqrt{h^2+2q\beta^2}}=\sqrt{2}\beta\,(1-q)$ and (5.7). Since a solution to (5.7) always exists a unique critical point always exists in $(q_P,1)$, and by concavity it is the unique local and global maximum. □

For the cases $k\geq 2$ the following fact will be useful.

Lemma 5.4. *For all f,β,h it holds that $\tilde{\mathcal{B}}(0,r)$ is strictly decreasing in r.*

Proof. We have

$$\tilde{\mathcal{B}}(0,r)=\sqrt{2}\beta r^2+\frac{\beta^2}{2}(1-r^2)^2+\frac{1}{2}\log(1-r^2),\tag{5.14}$$

and

$$\partial_r\tilde{\mathcal{B}}(0,r)=-2r\left(\beta^2(1-r^2)-\sqrt{2}\beta+\frac{1}{2(1-r^2)}\right)=-2r\left(\beta\sqrt{1-r^2}-\frac{1}{\sqrt{2(1-r^2)}}\right)^2\leq 0,\tag{5.15}$$

with equality only at a single point, implying the claim. □

We are now ready to study the case $k=2$. Define for $\beta>0$

$$\mathcal{F}(\beta)=\sup_{r\in[\sqrt{q_P},1)}\tilde{\mathcal{B}}(0,r)=\tilde{\mathcal{B}}(0,\sqrt{q_P})=\begin{cases}\frac{\beta^2}{2} & \text{for } \beta\leq\frac{1}{\sqrt{2}},\\ \sqrt{2}\beta-\frac{3}{4}-\frac{1}{2}\log(\sqrt{2}\beta) & \text{for } \beta\geq\frac{1}{\sqrt{2}}.\end{cases}\tag{5.16}$$

Lemma 5.5 (TAP maximizer with quadratic spike)**.** *Let $f(x)=hx^2$. If $h>\frac{1}{2}$ and $\beta<\sqrt{2}h$ then*

$$\sup_{r\in\mathrm{Plef}(\beta),\alpha\in[-1,1]}\tilde{\mathcal{B}}(\alpha,r)=\frac{\beta^2}{8h^2}(4h-1)+h-\frac{1}{2}\left(1+\log(2h)\right),\tag{5.17}$$

and the unique maximizers of the l.h.s. are

$$\left(\sqrt{1-\tfrac{1}{2h}},\pm\sqrt{1-\tfrac{\beta^2}{2h^2}}\right).\tag{5.18}$$

If either $h\leq\frac{1}{2}$ or $\beta\geq\sqrt{2}h$ then

$$\sup_{r\in\mathrm{Plef}(\beta),\alpha\in[-1,1]}\tilde{\mathcal{B}}(\alpha,r)=\mathcal{F}(\beta),\tag{5.19}$$

where the maximum is attained at $(q_P,0)$ (uniquely if $\beta>\frac{1}{\sqrt{2}}$ and otherwise also on $\{0\}\times[0,1]$).

Proof. We first maximize in α for fixed r. The critical point equation in α for r fixed is

$$r^2\left(2h\alpha-\sqrt{2}\beta\frac{\alpha}{\sqrt{1-\alpha^2}}\right)=0.\tag{5.20}$$

Thus when $r\neq 0$ the only critical points are $\alpha=0$ and if $\beta<\sqrt{2}h$ also

$$\alpha_r=\pm\sqrt{1-\frac{\beta^2}{2h^2}}.\tag{5.21}$$

Note that if $\beta < \sqrt{2}h$ and $r \neq 0$ we also have

$$\tilde{\mathcal{B}}(\alpha_r, r) - \tilde{\mathcal{B}}(0, r) = \left(\frac{2h^2+\beta^2}{2h} - \sqrt{2}\beta\right) r^2 = \frac{r^2}{2h}\left(\sqrt{2}h - \beta\right)^2 > 0,$$

so that the maximizing α for fixed $r \neq 0$ is

$$\alpha = \begin{cases} 0 & \text{if } \beta \geq \sqrt{2}h, \\ \pm\sqrt{1-\frac{\beta^2}{2h^2}} & \text{if } \beta < \sqrt{2}h. \end{cases}$$

Thus with $q = r^2$ and recalling (5.11) we have

$$\sup_{\alpha\in[-1,1]} \tilde{\mathcal{B}}(\alpha, r) = \begin{cases} \mathscr{B}(q) := \frac{2h^2+\beta^2}{2h} q + g(\sqrt{q}) & \text{if } \beta < \sqrt{2}h, \\ \tilde{\mathcal{B}}(0, \sqrt{q}) & \text{if } \beta \geq \sqrt{2}h. \end{cases} \tag{5.22}$$

If $\beta \geq \sqrt{2}h$ all claims thus follow by Lemma 5.4 and (5.16).

If $\beta < \sqrt{2}h$, since the first term $\mathscr{B}(q)$ is linear (5.12) implies that $\mathscr{B}(q)$ is strictly concave in $q \in [q_P, 1)$, and so it has a unique maximizer. Note that

$$\mathscr{B}'(q) = \frac{2h^2+\beta^2}{2h} + \frac{\partial}{\partial q} g(\sqrt{q}) = \frac{\beta}{\sqrt{2}}\left(v\left(\frac{\beta}{\sqrt{2}h}\right) - s\left(\sqrt{2}\beta(1-q)\right)\right),$$

recalling the second part of (5.13) and the function $v(x) = x + x^{-1}$ from $(0, 1]$ to $[2, \infty)$ which is an increasing bijection. Therefore $\mathscr{B}'(q) = 0$ is equivalent to

$$\frac{\beta}{\sqrt{2}h} = \sqrt{2}\beta(1-q) \iff q = 1 - \frac{1}{2h}.$$

Now if $h > \frac{1}{2}$, we have that $1 - \frac{1}{2h} \in (q_P, 1)$ so that $1 - \frac{1}{2h}$ is a critical point in $(q_P, 1)$ and by concavity it is the unique local and global maximum. It is easy to check that $\mathscr{B}\left(1 - \frac{1}{2h}\right)$ equals the r.h.s. of (5.17), completing the proof when $h > \frac{1}{2}$ and $\beta < \sqrt{2}h$. If $h \leq \frac{1}{2}$ the maxmizer is $q = \sqrt{q}_P$, since $\mathscr{B}(q) \to -\infty$ for $q \to 1$, and $\mathscr{B}(q) = \tilde{\mathcal{B}}(0, \sqrt{q_P}) = \mathcal{F}(\beta)$, giving the claims. □

The result on maximizers of $\tilde{\mathcal{B}}$ for monomial f with $k \geq 3$ is less explicit, and the analysis more complicated. We first show that the global maximum of $\tilde{\mathcal{B}}$ on $[0, 1] \times [\sqrt{q_p}, 1]$ is either achieved at a critical point of in the interior $(\sqrt{q_p}, 1) \times (0, 1)$ or at $(\sqrt{q_p}, 0)$.

Lemma 5.6. *For any $f \in C^1([-1, 1])$ we have that $\tilde{\mathcal{B}}(\alpha, r)$ for $(\alpha, r) \in [0, 1] \times [\sqrt{q_P}, 1]$ is maximized in the interior $(\sqrt{q_P}, 1) \times (0, 1)$ or at the point $(\alpha, r) = (0, \sqrt{q_P})$.*

Proof. Note that we have $\tilde{\mathcal{B}}(\alpha, 1) = -\infty$ and

$$\frac{\partial}{\partial\alpha}\tilde{\mathcal{B}}(\alpha, r) = rf'(r\alpha) - \frac{\sqrt{2}\beta r^2}{\sqrt{1-\alpha^2}} \longrightarrow -\infty \quad \text{as } \alpha \to 1, \tag{5.23}$$

so (α, r) with $r = 1$ or $\alpha = 1$ can not be maximizers. Lemma 5.4 shows the only possible maximizer with $r \in \left[\sqrt{q_P}, 1\right], \alpha = 0$ is $\left(\sqrt{q_P}, 0\right)$. If $\beta \leq \frac{1}{\sqrt{2}}$ then $q_P = 0$, and $\tilde{\mathcal{B}}(\alpha, 0) = f(0) + g(0)$ for all α, so if a point on the remaining boundary $r = \sqrt{q_P}, \alpha \in [0, 1]$ is a maximizer then so is $\left(\sqrt{q_P}, 0\right)$.

Lastly if $\beta > \frac{1}{\sqrt{2}}$ then any critical point of

$$\tilde{\mathcal{B}}(\alpha, \sqrt{q_P}) = f(\sqrt{q_P}\alpha) + \sqrt{2}\beta q_P\sqrt{1-\alpha^2} + g(\sqrt{q_P})$$

is a solution of

$$\sqrt{q_P} f'(\sqrt{q_P}\alpha) - \sqrt{2}\beta q_P \frac{\alpha}{\sqrt{1-\alpha^2}} = 0 \quad \Leftrightarrow \quad f'(\sqrt{q_P}\alpha) = \frac{\sqrt{2}\beta\sqrt{q_P}\alpha}{\sqrt{1-\alpha^2}}. \tag{5.24}$$

However, in any such point the derivative of $\tilde{\mathcal{B}}$ in r is

$$\begin{aligned} \alpha f'(\sqrt{q_P}\alpha)+2\sqrt{2}\beta\sqrt{q_P}\sqrt{1-\alpha^2}+g'(\sqrt{q_P}) &\overset{(5.24)}{=} \sqrt{2}\beta\sqrt{q_P}\left(\tfrac{\alpha^2}{\sqrt{1-\alpha^2}}+2\sqrt{1-\alpha^2}\right)-2\sqrt{2}\beta\sqrt{q_P} \\ &= \sqrt{2}\beta\sqrt{q_P}\left(\tfrac{2-\alpha^2}{\sqrt{1-\alpha^2}}-2\right), \end{aligned} \tag{5.25}$$

which is equal to zero for $\alpha=0$ and positive for all $\alpha\in(0,1)$. Therefore, if some $\alpha>0$ maximizes $\tilde{\mathcal{B}}\left(\alpha,\sqrt{q_P}\right)$ then there are larger values in the neighborhood of that point, and thus $(\sqrt{q_P},\alpha)$ cannot be a global maximizer. □

Define

$$h_c(k,\beta)=\begin{cases}0 & \text{for } k=1,\\ \min\{\frac{1}{2},\frac{\beta}{\sqrt{2}}\} & \text{for } k=2,\\ \mathcal{W}(k,\beta) & \text{for } k\geq 3,\end{cases} \tag{5.26}$$

where

$$\mathcal{W}(k,\beta)=\inf_{r\in\mathrm{Plef}(\beta)}\left\{\frac{\mathcal{F}(\beta)-g(r)-2\beta^2r^2(1-r^2)}{\left(r\sqrt{1-2\beta^2(1-r^2)^2}\right)^k}\right\}. \tag{5.27}$$

We now show that if $h>h_c(k,\beta)$ for $k\geq 3$ then there is a unique maximizer in the interior $(q_P,1)\times(0,1)$, while for $h<h_c(k,\beta)$ the point $(\sqrt{q_P},0)$ is the unique maximizer.

Lemma 5.7 (TAP maximizer with degree $k\geq 3$ spike)**.** *Let $k\geq 3,\beta>0,h>0$ and $f(x)=hx^k$. It holds that*

$$\sup_{r\in\mathrm{Plef}(\beta),\alpha\in[-1,1]}\tilde{\mathcal{B}}(\alpha,r)=\sup_{r\in\mathrm{Plef}(\beta)}\left\{hr^k\left(1-2\beta^2\left(1-r^2\right)^2\right)^{\frac{k}{2}}+2\beta^2r^2\left(1-r^2\right)+g\left(r\right)\right\}. \tag{5.28}$$

If $h<h_c\left(k,\beta\right)$ then the unique maximizer of the l.h.s. is $(\sqrt{q_P},0)$ and the l.h.s. equals $\mathcal{F}(\beta)$, and if $h>h_c\left(k,\beta\right)$ it the unqiue maximizer is $(\hat{r},\sqrt{1-2\beta^2\left(1-\hat{r}^2\right)^2})$ where $\hat{r}$ is the largest of the two solutions of

$$\left(1-r^2\right)\left(r^2\left(1-2\beta^2(1-r^2)^2\right)\right)^{\frac{k-2}{2}}=\frac{1}{hk} \tag{5.29}$$

in $(\sqrt{q_P},1)$.

Proof. By Lemma 5.6 the maximizer of the l.h.s. of (5.28) is either $(\sqrt{q}_P,0)$ or a critical point of $\tilde{\mathcal{B}}$ in $(1,\sqrt{q}_P)\times(0,1)$. The critical point equations are

$$0 = hk\alpha^kr^{k-1}+2\sqrt{2}\beta r\sqrt{1-\alpha^2}+g'(r) \tag{5.30}$$

$$0 = hk\alpha^{k-1}r^k-\sqrt{2}\beta r^2\tfrac{\alpha}{\sqrt{1-\alpha^2}}. \tag{5.31}$$

Any solution to (5.31) must satisfy $hk\alpha^kr^{k-1}=r\sqrt{2}\beta r^2\frac{\alpha^2}{1-\alpha^2}$, and plugging this into (5.30) we get that any critical point must satisfy

$$\frac{\alpha^2}{\sqrt{1-\alpha^2}}+2\sqrt{1-\alpha^2}=c(r), \tag{5.32}$$

where

$$c(r)=\frac{g'(r)}{\sqrt{2}\beta r}=\frac{1}{\sqrt{2}\beta(1-r^2)}+\sqrt{2}\beta(1-r^2).$$

The quadratic (5.32) in α^2 has the solutions $\frac{-(c(r)^2-4)\pm c(r)\sqrt{c(r)^2-4}}{2}$ which are well-defined since $c(r)>2$ for $r>\sqrt{q}_P$. Since only one is non-negative and using $\sqrt{(x+x^{-1})^2-4}=x^{-1}-x$ for $x\in(0,1)$ we obtain that any critical point must satisfy

$$\alpha^2=\frac{c(r)\sqrt{c(r)^2-4}-(c(r)^2-4)}{2}=1-2\beta^2(1-r^2)^2. \tag{5.33}$$

The r.h.s. lies in $[0,1]$ for all $\beta > 0$ and $r \in \mathrm{Plef}(\beta)$. Thus

$$\sup_{r\in\mathrm{Plef}(\beta),\alpha\in[-1,1]} \tilde{\mathcal{B}}(\alpha,r) = \sup_{r\in\mathrm{Plef}(\beta)} \tilde{\mathcal{B}}(\sqrt{1-2\beta^2(1-r^2)^2}, r), \tag{5.34}$$

noting that when r is the left-end point $\sqrt{q_P}$ of $\mathrm{Plef}(\beta)$ the r.h.s. is $\tilde{\mathcal{B}}(0,\sqrt{q}_P)$. The r.h.s. of (5.34) equals the r.h.s. of (5.28), so (5.28) is proved.

Next note that

$$\begin{array}{ll} & \exists r \in (0,1) : \tilde{\mathcal{B}}(\sqrt{1-2\beta^2(1-r^2)^2}, r) > \tilde{\mathcal{B}}(0,\sqrt{q_P}) = \mathcal{F}(\beta) \\ \Leftrightarrow & \exists r \in (0,1) : h > \frac{\mathcal{F}(\beta)-2\beta^2 r^2(1-r^2)-g(r)}{(r^2(1-2\beta^2(1-r^2)^2))^{\frac{k}{2}}} \\ \Leftrightarrow & h > \mathcal{W}(k,\beta). \end{array} \tag{5.35}$$

Thus indeed for $h < h_c(k,\beta)$ the unique maximizer is $(\sqrt{q}_P, 0)$. When $h > h_c(k,\beta)$ the maximizer is a critical point $(\hat{r}, \sqrt{1-2\beta^2(1-\hat{r}^2)^2})$ in the interior $(\sqrt{q}_P, 1)\times(0,1)$. It remains to characterize this point and prove its uniqueness.

Firstly, plugging (5.33) into (5.31) one sees that any critical point (α, r) of $\tilde{\mathcal{B}}$ and critical point of the expression on the r.h.s. of (5.28) with $r \in (\sqrt{q}_P, 1)$ must satisfy (5.29). When $h > h_c(k,\beta)$ there is a local and global maximum, so the equation must have at least one solution. Let

$$T(q) = (1-q)\left(q(1-2\beta^2(1-q)^2)\right)^{\frac{k-2}{2}}, \tag{5.36}$$

so that the l.h.s. of (5.29) is $T(r^2)$. Note that $T(q)$ is non-negative for all $q \in (q_P, 1)$ and zero for $q \in \{q_P, 1\}$. Furthermore

$$\begin{aligned} \tfrac{\partial}{\partial q}\log T(q) &= -\frac{1}{1-q} + \frac{k-2}{2}\frac{1}{q} - (k-2)\frac{2\beta^2(1-q)}{1-2\beta^2(1-q)^2} \\ &= \frac{k-2-kq-(k-2)\,q\frac{2\beta^2(1-q)^2}{1-2\beta^2(1-q)^2}}{2q(1-q)} \\ &= \frac{k-2-q\left\{k-(k-2)\left(\frac{1}{1-2\beta^2(1-q)^2}-1\right)\right\}}{2q(1-q)}. \end{aligned}$$

Since $\frac{1}{1-2\beta^2(1-q)^2} - 1$ is negative and decreasing in $(q_P, 1)$, we have that the numerator is decreasing. Therefore $\frac{\partial}{\partial q}\log T(q)$ can switch sign only once in $(q_P, 1)$, showing that $T(q)$ has exactly one critical point in $(q_P, 1)$, so the equation (5.29) has zero, one or two solutions. We have already excluded the possibility of it having zero solutions. Thus the expression on the r.h.s. of (5.28) has one or two critical points, of which at least one is a local maximum.

To determine the number and type of the critical point(s) it is useful to note that the expression on the r.h.s. of (5.28) is always decreasing in r in a neighborhood of q_P. Indeed when $\beta < \frac{1}{\sqrt{2}}$ so that $q_P = 0$ this follows by expanding the expression around $r = 0$ as $\frac{\beta^2}{2} + (\beta^2 - \frac{1}{2})r^2 + O(r^3)$. When $\beta = \frac{1}{\sqrt{2}}$ similarly the expression expands as $\frac{\beta^2}{2} - r^4 + O(r^5)$. When $\beta > \frac{1}{\sqrt{2}}$ we can make the change of variables $1-2\beta^2\left(1-r^2\right) = z$ and expand the expression around $z = 0$ as $\mathcal{F}(\beta) + \frac{1}{2}(1-\sqrt{2}\beta)z + O(z^2)$, which is decreasing in z in neighborhood of 0 and therefore decreasing in r in a neighborhood of $\sqrt{q_P}$.

Thus since the expression is decreasing in a neighbourhood of $r = \sqrt{q_P}$ the left-most critical point cannot be a local maximum. Thus there are two critical points and (5.29) has two solutions, the smaller which corresponds to a local minimum, and the larger of which corresponds to a local maximum which is also the global maximum. □

6. Fluctuations

In this section we prove Theorem 1.1 (b) and Theorem 1.2 (b) about the fluctuations of L_N resp. $\tilde{L}_N$. We do so by studying the fluctuations of minimax expressions of the type

$$\sup_y \inf_l h(y, l, s_{\lambda,u}(l)).$$

The next lemma shows that under the assumptions of Theorem 1.1 (b) and Theorem 1.2 (b) the quantities L_N and $\tilde{L}_N$ equal such minimax expressions with probability tending to one. Recall $\mathcal{B}(\alpha)$ and $\mathcal{B}(\alpha, r)$ from (1.3) and (1.9).

Lemma 6.1. *(a) If $\mathcal{B}(\alpha)$ has finitely many global maximizers $\hat{\alpha}_i$, $i = 1, ..., m$ which are all non-zero then for all $\varepsilon > 0$ small enough*

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{1}{N}L_N = \max_{i=1,\dots,m} \sup_{\alpha\in[\hat{\alpha}_i-\varepsilon,\hat{\alpha}_i+\varepsilon]} \inf_{l\in[\sqrt{2}+\varepsilon,\varepsilon^{-1}]} h(\alpha, l, s_{\lambda,u}(l))\right) = 1, \tag{6.1}$$

where

$$h(\alpha, l, g) = f(\alpha) + \beta\left(l - \frac{\alpha^2}{g}\right).$$

(b) If $\tilde{\mathcal{B}}(\alpha, r)$ has finitely many global maximizers $(\hat{\alpha}_i, \hat{r}_i)$, $i = 1, ..., m$, all lying in the interior $[-1, 1] \times \mathcal{R}$ with $\hat{\alpha}_i, \hat{r}_i \neq 0$, then for all $\varepsilon > 0$ small enough

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{1}{N}\tilde{L}_N = \max_{i=1,\dots,n} \sup_{\alpha\in[\hat{\alpha}_i-\varepsilon,\hat{\alpha}_i+\varepsilon],r\in[\hat{r}_i-\varepsilon,\hat{r}_i+\varepsilon]} \inf_{l\in[\sqrt{2}+\varepsilon,\varepsilon^{-1}]} h((\alpha, r), l, s_{\lambda,u}(l))\right) = 1, \tag{6.2}$$

where

$$h((\alpha, r), l, g) = f(\alpha r) + g(r) + \beta r^2\left(l - \frac{\alpha^2}{g}\right).$$

Proof. By (3.3) we have

$$\frac{1}{N}L_N = \sup_{\alpha\in[-1,1]} \left\{f(\alpha) + \beta \sup_{\substack{|\sigma|=1\\ \sigma\cdot u=\alpha}} \sum_{i=1}^N \lambda_i\sigma_i^2\right\}$$

and by Proposition 4.1

$$f(\alpha) + \beta \sup_{\substack{|\sigma|=1\\ \sigma\cdot u=\alpha}} \sum_{i=1}^N \lambda_i\sigma_i^2 = \mathcal{B}(\alpha) + o_{\mathbb{P}}(1)$$

for all $\alpha \in [-1, 1]$ uniformly, so for any $\varepsilon > 0$ a global maximizer α^* of the l.h.s. must lie in a ε-neighborhood of one of the $\hat{\alpha}_i \neq 0$ with probability tending to 1. Thus by Lemma 3.1

$$\frac{1}{N}L_N = \max_{i=1,\dots,m} \sup_{\alpha\in[\hat{\alpha}_i-\varepsilon,\hat{\alpha}_i+\varepsilon]} \inf_{l>\lambda_N} \left\{l - \frac{\alpha^2}{s_{\lambda,u}(l)}\right\},$$

with probability tending to one. Since $f \in C^1([-1, 1])$ and the derivative of $\sqrt{2(1-\alpha^2)}$ diverges for $\alpha^2 \to 1$, neither 1 nor -1 can be a maximizer, so the α_i^* are bounded away from ± 1 with probability tending to one. By Lemma 4.9 the minimizer in l of $h(\hat{\alpha}_i, l, s_{\lambda,u}(l))$ must lie in $[\sqrt{2}+\varepsilon, \varepsilon^{-1}]$ with probability tending to one for each i, after possibly decreasing ε, proving (a).

The claim (b) follows similarly using (3.4) and Proposition 4.1. □

6.1. General minimax optimization involving $s_{\lambda,u}$

In the rest of the section we will study the fluctuations of $\inf_{y\in\mathcal{Y}}\sup_{l\in\mathcal{L}} h(y,l,s_{\lambda,u}(l))$ under the assumptions that

$$\mathcal{Y}\subset\mathbb{R}^n, \mathcal{L}\subset(\sqrt{2},\infty), \mathcal{G} \text{ are compact with } s(\mathcal{L})\subset\mathcal{G}^o \tag{6.3}$$

(where A^o denotes the interior of a set A)

$$h:\mathcal{Y}\times\mathcal{L}\times\mathcal{G}\to\mathbb{R} \text{ is three times continuously differentiable,} \tag{6.4}$$

$$y\to\mathcal{B}(y) \text{ is uniquely maximized at a } \hat{y}\in\mathcal{Y}^o, \text{ where } \mathcal{B}(y)=\inf_{l\in\mathcal{L}} h(y,l,s(l))\}, \tag{6.5}$$

$$l\to h(\hat{y},l,s(l)) \text{ is uniquely minimized at a } \hat{l}\in\mathcal{L}^o, \tag{6.6}$$

$$\partial_{ll}h(\hat{y},l,s(l))|_{l=\hat{l}}>0, \tag{6.7}$$

$$\nabla^2\mathcal{B}(\hat{y}) \text{ is negative definite.} \tag{6.8}$$

The existence of the derivatives in (6.8) is guaranteed by the formula (4.18) for the specific h from Lemma 6.1 (a) (b). It also follows from the other assumptions by the implicit function theorem. The latter argument is included in the following two lemmas, which will be needed also later.

Lemma 6.2. *Let $n\geq 1, A\subset\mathbb{R}^n$, $\eta>0$ and $t:A\times[-\eta,\eta]\to\mathbb{R}$ be twice continuously differentiable. If $\partial_{bb}t(a,b)>0$ for all $a\in A, b\in[-\eta,\eta]$, and $\partial_b t(a,-\eta)<0, \partial_b t(a,\eta)>0$, for all $a\in A$ then,* $\mathrm{argmin}_{b\in[-\eta,\eta]}t(a,b)$ *is unique for all $a\in A$ and $b^*(a)=\mathrm{argmin}_{b\in[-\eta,\eta]}t(a,b)$ is continuously differentiable in A with*

$$\nabla b^*(a)=-\frac{\partial_b\nabla_a t(a,b)}{\partial_{bb}t(a,b)}|_{b=b^*(a)} \tag{6.9}$$

for all $a\in A$. Furthermore for all $a\in A$

$$\nabla_a\{t(a,b^*(a))\}=\{\nabla_a t\}(a,b^*(a)), \tag{6.10}$$

and

$$\nabla_a^2\{t(a,b^*(a))\}=\nabla_a^2 t(a,b^*(a))-\frac{\partial_b\{\nabla_a t\}(a,b^*(a))\,(\partial_b\{\nabla_a t\}(a,b^*(a)))^T}{\partial_{bb}t(a,b)}. \tag{6.11}$$

Proof. The assumption $\partial_{bb}t(a,b)>0$ implies that $\mathrm{argmin}_{b\in[-\eta,\eta]}t(a,b)$ is unique. Then the assumption $\partial_b t(a,-\eta)<0, \partial_b t(a,\eta)>0$, implies that $b^*(a)$ lies in $(-\eta,\eta)$ and is the unique solution of $\partial_b t(a,b)=0$ in this interval. Finally by the implicit function theorem applied to $\partial_b t(a,b)=0$ the solution $b^*(a)$ to this equation for b is continuously differentiable and satisfies $\nabla b^*(a)=-\frac{\partial_b\nabla_a t(a,b)}{\partial_{bb}t(a,b)}$, using again that $\partial_{bb}t(a,b)>0$. Furthermore

$$\nabla_a\{t(a,b^*(a))\}=\{\nabla_a t\}(a,b^*(a))+\underbrace{\partial_b t(a,b^*(a))}_{=0} \tag{6.12}$$

for all $a\in\mathcal{A}$, which shows (6.13). By taking the derivative of (6.12) one obtains

$$\nabla_a^2\{t(a,b^*(a))\}=\nabla_a^2 t(a,b^*(a))-\partial_b\{\nabla_a t\}(a,b^*(a))\ \nabla b^*(a)^T$$

and by using (6.9) this shows (6.11). □

Applied to $(y,l)\to h(y,l,s(l))$ the lemma yields that $\mathcal{B}(y)$ is differentiable in a neighborhood and the following relation between the derivatives of $\mathcal{B}(y)$ and the derivatives of $h(y,l,s(l))$.

Lemma 6.3. *Assume (6.3)-(6.8). Then there is a neighborhood $\mathcal{U}$ of $\hat{y}$ such that $\inf_{l\in\mathcal{L}} h(y,l,s(l))$ is uniquely maximized at a $\hat{l}(y)$ for $y \in \mathcal{U}$, $\mathcal{B}$ from (6.5) is three times continuously differentiable in $\mathcal{U}$, and for all $y \in \mathcal{U}$*

$$\nabla\mathcal{B}(y) = \nabla_y h(y, \hat{l}(y), s(\hat{l}(y))) \tag{6.13}$$

and

$$\nabla^2\mathcal{B}(y) = \nabla_y^2 h(y, \hat{l}(y), s(\hat{l}(y))) - \frac{\partial_l \{\nabla_y h(y,l,s(l))\} (\partial_l \{\nabla_y h(y,l,s(l))\})^T}{\partial_{ll} h(y,l,s(l))}\Big|_{l=\hat{l}(y)}. \tag{6.14}$$

Proof. Using (6.4) and (6.7) it follows that there is a neighborhood $\mathcal{U}$ of $\hat{y}$ and $[\hat{l}-\eta, \hat{l}+\eta]$ of $\hat{l}$ where $\partial_{ll} h(y,l,s(l)) > 0$ for all $y \in \mathcal{U}, l \in [\hat{l}-\eta, \hat{l}+\eta]$, and by (6.6) one can in addition ensure that $\partial_l h(y,l,s(l))|_{l=\hat{l}-\eta} < 0$ and $\partial_l h(y,l,s(l))|_{l=\hat{l}+\eta} > 0$. By Lemma 6.2 applied to $t(a,b) = h(\hat{y}+a, \hat{l}+b, s(\hat{l}+b))$ one obtains (6.13) and (6.14). Since all terms on the r.h.s. of (6.14) are continuously differentiable it follows that $\mathcal{B} \in C^3(\mathcal{U})$. □

6.2. Fluctuations of $s_{\lambda,u}$ around s

We will calculate the fluctuations of $\sup_{y\in\mathcal{Y}} \inf_{l\in\mathcal{L}} h(\alpha, l, s_{\lambda,u}(l))$ by quadratically expanding h around $(\hat{y}, \hat{l}, s(\hat{l}))$. To this end we start by studying the fluctuations of $s^{(k)}_{\lambda,u}(l)$ around $s^{(k)}(l)$. Note that for all $l \in \mathcal{L}$

$$s^{(k)}_{\lambda,u}(l) = s^{(k)}(l) + \frac{1}{\sqrt{N}} W_N^{(k)}(l) + \frac{1}{N}\Lambda_N^{(k)}(l) + \frac{1}{N} R_N^{(k)}(l), \tag{6.15}$$

where

$$W_N(l) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{N u_i^2 - 1}{l - \theta_{i/N}}, \tag{6.16}$$

and

$$\Lambda_N(l) = \sum_{i=1}^N \frac{1}{l-\lambda_i} - N s(l),$$

as well as

$$R_N(l) = \sum_{i=1}^N (N u_i^2 - 1)\left(\frac{1}{l-\lambda_i} - \frac{1}{l-\theta_{i/N}}\right).$$

We also define

$$U_N(l) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{N u_i^2 - 1}{l - \lambda_i},$$

which equals U_N from Theorem 1.1 for $l = \frac{2-\hat{\alpha}^2}{\hat{z}}$ (with $\hat{\alpha}$ and $\hat{z}$ as in the theorem), recalling from below (3.2) that u is the vector v in the diagonalizing basis of J and λ_i are the eigenvalues of $\frac{1}{\sqrt{N}} J_N$. The derivative $U_N'(l)$ for $l = \frac{2-\hat{\alpha}^2}{\hat{z}}$ also equals U_N' from Theorem 1.1. Later we will use that

$$U_N^{(k)}(l) - W_N^{(k)}(l) = \frac{1}{\sqrt{N}} R_N^{(k)}(l). \tag{6.17}$$

The next lemma shows that the error term $R_N^{(k)}$ in (6.15) and (6.17) is small.

Lemma 6.4. *For all k and $\varepsilon > 0$ it holds that $\sup_{l \geq \sqrt{2}+\varepsilon} |R_N^{(k)}(l)| = o_{\mathbb{P}}(1)$.*

Proof. Let $w(l,x) = \frac{1}{l-x}$ and denote by $w^{(k)}(l,x)$ the k-th derivative in l. Let $\delta > 0$ and define the event

$$\mathcal{E}_\delta = \left\{ \sup_{i=1,\dots,N} |\lambda_i - \theta_{i/N}| \leq N^{-\frac{2}{3}+\delta} \right\}, \tag{6.18}$$

whose probability converges to one for any choice of δ by Lemma 2.1, and define the σ-algebra

$$\sigma_\Lambda = \sigma(\lambda_1, ..., \lambda_N).$$

First consider

$$X := \frac{1}{N}\sum_{i=1}^{N}\left(N\tilde{u}_i^2 - 1\right)\left(w^{(k)}(l,\lambda_i) - w^{(k)}(l,\theta_{i/N})\right),$$

where $\tilde{u}_i$ are i.i.d with law $\mathcal{N}(0,1)$ and independent of J, as in the proof of Lemma 4.4. Then $\mathbb{E}[X|\sigma_\Lambda] = 0$ and

$$\begin{aligned}
\mathbb{E}[X^2|\sigma_\Lambda]1_{\mathcal{E}_\delta} &= \tfrac{1}{N^2}\mathbb{E}\left[\left(\sum_{i=1}^N \left(N\tilde{u}_i^2 - 1\right)\left(w^{(k)}(l,\lambda_i) - w^{(k)}(l,\theta_{i/N})\right)\right)^2 |\sigma_\Lambda\right]1_{\mathcal{E}_\delta} \\
&= \tfrac{2}{N^2}\sum_{i=1}^N \left(w^{(k)}(l,\lambda_i) - w^{(k)}(l,\theta_{i/N})\right)^2 1_{\mathcal{E}_\delta} \\
&\overset{(6.18)}{\leq} \tfrac{2|w^{(k+1)}|_\infty}{N^2} N N^{-\frac{4}{3}+2\delta}1_{\mathcal{E}_\delta} = N^{-7/3+2\delta}1_{\mathcal{E}_\delta},
\end{aligned}$$

which implies via Chebyshev's inequality that

$$\begin{aligned}
\mathbb{P}\left(|X| \geq \tfrac{1}{N\log N}\right) &= \mathbb{E}\left[\mathbb{P}\left(|X| \geq \tfrac{1}{N\log N}|\sigma_\Lambda\right)\right] \\
&\leq \mathbb{E}\left[\mathbb{P}\left(|X| \geq \tfrac{1}{N\log N}|\sigma_\Lambda\right)1_{\mathcal{E}_\delta}\right] + \mathbb{P}(\mathcal{E}_\delta^c) \\
&\leq \mathbb{E}\left[\mathbb{E}[X^2|\sigma_\Lambda](N\log N)^2 1_{\mathcal{E}_\delta}\right] + \mathbb{P}(\mathcal{E}_\delta^c) \\
&\leq (\log N)^2 N^{-\frac{1}{3}+2\delta} + \mathbb{P}(\mathcal{E}_\delta^c).
\end{aligned}$$

By choosing $\delta < \frac{1}{6}$ this probability converges to zero, and so $X = o_{\mathbb{P}}(\frac{1}{N})$.
Constructing the vector u via $u = \tilde{u}/|\tilde{u}|$ we then have

$$\begin{aligned}
&\sum_{i=1}^N u_i^2\left(w^{(k)}(l,\lambda_i) - w^{(k)}(l,\theta_{i/N})\right) - \sum_{i=1}^N \tilde{u}_i^2\left(w^{(k)}(l,\lambda_i) - w^{(k)}(l,\theta_{i/N})\right) \\
= \;& (1 - |\tilde{u}|^2)\sum_{i=1}^N u_i^2\left(w^{(k)}(l,\lambda_i) - w^{(k)}(l,\theta_{i/N})\right),
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}\left(\left|(1-|\tilde{u}|^2)\sum_{i=1}^N u_i^2\left(w^{(k)}(l,\lambda_i) - w^{(k)}(l,\theta_{i/N})\right)\right| \leq \tfrac{1}{N\log N}\right) \\
&\geq \mathbb{P}\left(\left|1-|\tilde{u}|^2\right| \leq \tfrac{1}{N^{\frac{1}{3}+\delta}\log N},\ \left|\sum_{i=1}^N u_i^2\left(w^{(k)}(l,\lambda_i) - w^{(k)}(l,\theta_{i/N})\right)\right| \leq N^{-\frac{2}{3}+\delta}\right) \\
&\geq \mathbb{P}\left(\left|1-|\tilde{u}|^2\right| \leq \tfrac{1}{N^{\frac{1}{3}+\delta}\log N}\right) - \mathbb{P}\left(\sup_{i=1,\ldots,N}|\lambda_i - \theta_{i/N}| \geq \tfrac{1}{|w^{(k+1)}|_\infty}N^{-\frac{2}{3}+\delta}\right),
\end{aligned}$$

(if $|w'|_\infty = 0$ the claim of the lemma is of course trivial) which for $\delta < \frac{1}{6}$ converges to 1 by the CLT on the first probability and Lemma 2.1 on the second. ☐

It thus holds that

$$s_{\lambda,u}^{(k)}(l) = s^{(k)}(l) + \frac{1}{\sqrt{N}}W_N^{(k)}(l) + \frac{1}{N}\Lambda_N^{(k)}(l) + o_{\mathbb{P}}(N^{-1}) \text{ uniformly in } l \geq \sqrt{2}+\varepsilon, \tag{6.19}$$

for any $\varepsilon > 0$ and $k \in \mathbb{N}$. The next lemma shows that $W^{(k)}(l)$ of (6.19) is of order $O_{\mathbb{P}}(1)$ uniformly.

Lemma 6.5. *For all k and $\varepsilon > 0$ it holds that $\sup_{l\geq\sqrt{2}+\varepsilon}|W_N^{(k)}(l)| = O_{\mathbb{P}}(1)$.*

Proof. We construct u by setting $u_i = \frac{\tilde{u}_i}{|\tilde{u}|}$ with $\tilde{u}_1, ..., \tilde{u}_N \sim \mathcal{N}(0, \frac{1}{N})$ i.i.d.. We then have

$$\begin{aligned}
W_N(l) &= \frac{1}{\sqrt{N}}\frac{1}{|\tilde{u}|^2}\sum_{i=1}^N \frac{N\tilde{u}_i^2 - |\tilde{u}|^2}{l-\theta_{i/N}} = \frac{1}{\sqrt{N}}\frac{1}{|\tilde{u}|^2}\sum_{i=1}^N\left(N\tilde{u}_i^2 - |\tilde{u}|^2\right)\left(\frac{1}{l-\theta_{i/N}} - s_\theta(l)\right) \\
&= \frac{1}{\sqrt{N}}\frac{1}{|\tilde{u}|^2}\sum_{i=1}^N\left(N\tilde{u}_i^2 - 1\right)\left(\frac{1}{l-\theta_{i/N}} - s_\theta(l)\right)
\end{aligned} \tag{6.20}$$

and similarly

$$W_N^{(k)}(l) = \frac{1}{|\tilde{u}|^2}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(N\tilde{u}_i^2-1)\left(\frac{k!(-1)^k}{(l-\theta_{i/N})^{k+1}} - s_\theta^{(k)}(l)\right). \tag{6.21}$$

Note that $\frac{1}{|\tilde{u}|^2} = 1 + o_{\mathbb{P}}(1)$ and that by using (4.8) and a CLT we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(N\tilde{u}_i^2-1)s_\theta^{(k)}(l) = O_{\mathbb{P}}(1).$$

Thus it only remains to show that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(N\tilde{u}_i^2-1\right)w^{(k)}(\theta_{i/N},l) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{N}}\right)$, i.e.

$$\lim_{z\to\infty}\lim_{N\to\infty}\mathbb{P}\left(\sup_{l\geq\sqrt{2}+\varepsilon}\frac{1}{N}\sum_{i=1}^{N}\frac{N\tilde{u}_i^2-1}{(l-\theta_{i/N})^k}\geq z\right) = 0. \tag{6.22}$$

Note that for $x\in(0,1)$

$$\frac{1}{(1-x)^k} = \sum_{j=0}^{\infty}x^jC_j(k) \tag{6.23}$$

where $C_j(k) = \frac{k(k+1)\dots(k+1-j)}{j!}$, so that that we have for $l\geq\sqrt{2}+\varepsilon$ and all $x\in[-\sqrt{2}-\frac{\varepsilon}{2},\sqrt{2}+\frac{\varepsilon}{2}]$

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\frac{N\tilde{u}_i^2-1}{(l-\theta_{i/N})^k} &= \frac{1}{N}\sum_{j=0}^{\infty}\sum_{i=1}^{N}\frac{C_j(k)\theta_{i/N}^j}{l^{j+k}}\left(N\tilde{u}_i^2-1\right)\\ &= \frac{1}{N}\sum_{j=0}^{\infty}\frac{C_j(k)}{l^k}\left(\frac{\sqrt{2}+\frac{\varepsilon}{2}}{l}\right)^j\sum_{i=1}^{N}\left(N\tilde{u}_i^2-1\right)\left(\frac{\theta_{i/N}}{\sqrt{2}+\frac{\varepsilon}{2}}\right)^j.\end{aligned}$$

Let

$$\psi_N(j) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(N\tilde{u}_i^2-1\right)\left(\frac{\theta_{i/N}}{\sqrt{2}+\frac{\varepsilon}{2}}\right)^j.$$

Since

$$C_j(k) = \binom{k+j-1}{k-1}\leq(j+k)^{k-1}\quad\text{and}\quad\left|\frac{\sqrt{2}+\frac{\varepsilon}{2}}{l}\right|<q$$

for some $q\in(0,1)$, there exists some $c_1 = c_1(k,q)>0$ such that for fixed $k\in\mathbb{N}_0$

$$\frac{C_j(k)}{l^k}\left(\frac{\sqrt{2}+\frac{\varepsilon}{2}}{l}\right)^j\leq c_1q^j$$

uniformly for all j and $l>\sqrt{2}+\varepsilon$, and so

$$\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{N\tilde{u}_i^2-1}{(l-\theta_{i/N})^k}\right|\leq\sum_{j=0}^{\infty}c_1q^j|\psi_N(j)|.$$

We have

$$\mathrm{Var}(\psi_N(j)) = \sum_{i=1}^{N}\left(\frac{\theta_{i/N}}{\sqrt{2}+\frac{\varepsilon}{2}}\right)^{2j}\mathbb{E}\left[(1-N\tilde{u}_i^2)^2\right] = \frac{2}{N}\sum_{i=1}^{N}\left(\frac{\theta_{i/N}}{\sqrt{2}+\frac{\varepsilon}{2}}\right)^{2j}\leq 2\left(\frac{\sqrt{2}}{\sqrt{2}+\frac{\varepsilon}{2}}\right)^{2j}.$$

For any $x\in\mathbb{R}^+$ via Chebyshev's inequality

$$\mathbb{P}\left(\exists j\geq 1:|\psi_N(j)|\geq x\right)\leq\sum_{j=1}^{\infty}\frac{\mathrm{Var}(\psi_N(j))}{x^2}\leq\frac{1}{x^2}\sum_{j=1}^{\infty}2\left(\frac{\sqrt{2}}{\sqrt{2}+\frac{\varepsilon}{2}}\right)^{2j}\leq\frac{c_2}{x^2} \tag{6.24}$$

for some $c_2 = c_2(\varepsilon) > 0$. Thus the probability in (6.22) is bounded from above by

$$\mathbb{P}\left(\sum_{j=0}^{\infty} c_1 q^j |\psi_N(j)| \geq z\right) \leq \mathbb{P}\left(\sup_j |\psi_N(j)| \geq \frac{z}{2c_1 \sum_{j=0}^{\infty} q^j}\right) \overset{(6.24)}{\leq} \frac{4c_1^2 c_2}{(1-q)^2} \frac{1}{z^2}$$

for all N. Taking the limts $N \to \infty$ then $z \to \infty$ completes the proof. □

The following lemma shows that $\Lambda_N^{(k)}(l)$ from (6.19) is of order $O_{\mathbb{P}}(1)$ for fixed l, and the suboptimal but sufficient bound $O_{\mathbb{P}}(N^{\frac{2}{5}})$ uniformly in l.

Lemma 6.6. *For all k and l it holds that $|\Lambda_N^{(k)}(l)| = O_{\mathbb{P}}(1)$, and $\sup_{l \geq \sqrt{2}+\varepsilon} |\Lambda_N^{(k)}(l)| = O_{\mathbb{P}}(N^{\frac{2}{5}})$ for all $\varepsilon > 0$.*

Proof. Lemma 2.3 implies that $|\Lambda_N^{(k)}(l)| = O_{\mathbb{P}}(1)$.
Let $w(l,x) = \frac{1}{l-x}$ and let $w^{(k)}$ denote the k-th derivative in l. It holds that

$$\Lambda_N^{(k)}(l) = \left(\sum_{i=1}^{N} w^{(k)}(l, \lambda_i) - \sum_{i=1}^{N} w^{(k)}(l, \theta_{i/N})\right) + \left(\sum_{i=1}^{N} w^{(k)}(l, \theta_{i/N}) - N s(l)\right), \tag{6.25}$$

where the left most term on the r.h.s. is bounded by

$$N |w^{(k+1)}(l) 1_{\{l \geq \sqrt{2}+\varepsilon\}}|_\infty \sup_{i=1,\ldots,N} |\lambda_i - \theta_{i/N}|,$$

which is of order $O_{\mathbb{P}}(N^{\frac{2}{5}})$ by Lemma 2.1. The right-most term of (6.25) is of order $O_{\mathbb{P}}(1)$ by Lemma 2.2. □

In particular we have from (6.19) that

$$s_{\lambda,u}^{(k)}(l) = s^{(k)}(l) + O_{\mathbb{P}}(N^{-1/2}) \text{ uniformly in } l \geq \sqrt{2} + \varepsilon, \tag{6.26}$$

for any $\varepsilon > 0$ and $k \in \mathbb{N}$.

6.3. Quadratic expansion and fluctuations of minimax

We are now ready to expand $h(y, l, s_{\lambda,u}(l))$ quadratically around $(\hat{y}, \hat{l}, s(\hat{l}))$. To formulate the result one needs to take various partial derivatives of h, such as $\partial_l \{\{\partial_g h\}(y, l, s(l))\}|_{(\hat{y},\hat{l})}$. To keep the typographical size of expressions manageable we define the shorthand notation

$$h_{\underbrace{g\ldots g}_{j \text{ times}}}(y, l, s(l)) = \{\partial_g^j h\}(y, l, s(l)) \tag{6.27}$$

for first taking the g derivative j times and then substituting $s(l)$ for g, and

$$h_{g\ldots g} = h_{g\ldots g}(\hat{y}, \hat{l}, s(\hat{l})) \text{ for } j \in \{0, 1, \ldots\}, \tag{6.28}$$

for in addition substituting $(\hat{y}, \hat{l})$ for (y, l) at the end. Furthermore for $V = \{l\}, V = \{y\}$ or $V = \{l, y\}$ the notation

$$h_{V,g\ldots g} = \nabla_V \{h_{g\ldots g}(y, l, s(l))\}|_{(\hat{y},\hat{l})} \in \mathbb{R}^{|V|}, \tag{6.29}$$

is the gradient (viewed as a column vector) in some combination of l and y *after* taking g derivatives and substituting $s(l)$, evaluated at $(\hat{y}, \hat{l})$. Lastly for $V, V' = \{l\}, \{y\}$ or $\{y, l\}$

$$h_{V',V,g\ldots g} = \nabla_{V'} \nabla_V \{h_{g\ldots g}(y, l, s(l))\}|_{(\hat{y},\hat{l})} \in \mathbb{R}^{|V| \times |V'|}, \tag{6.30}$$

is a matrix of mixed derivatives in y, l obtained in the same way. Then e.g. $h_{l,g} = h_{\{l\},g} = \partial_l \{\{\partial_g h\}(y, l, s(l))\}|_{(\hat{y},\hat{l})}$, or $h_y = h_{\{y\}} = \{\nabla_y h\}(\hat{y}, \hat{l}, s(\hat{l})) \in \mathbb{R}^n$ or $h_{y,g} = h_{\{y\},g} = \{\nabla_y \partial_g\} h(\hat{y}, \hat{l}, s(\hat{l})) \in \mathbb{R}^n$. In the statement and proof below $h, h_g, h_{gg}, h_{l,g} \in \mathbb{R}$, $h_{\{y,l\},g} \in \mathbb{R}^{n+1}$ (column vector), $h_{y,g} \in \mathbb{R}^n$ (column vector) and $h_{\{y,l\},\{y,l\}} \in \mathbb{R}^{(n+1)\times(n+1)}$ (matrix) appear.

Similarly, we write for short

$$W_N^{(k)} = W_N^{(k)}(\hat{l}) \quad \text{and} \quad \Lambda_N^{(k)} = \Lambda_N^{(k)}(\hat{l}) \quad \text{for} \quad k \in \mathbb{N}. \tag{6.31}$$

We now state the quadratic expansion.

Lemma 6.7. *Let $h, \mathcal{Y}, \mathcal{L}$ be as in (6.3)-(6.8). Writing $\Delta = (y - \hat{y}, l - \hat{l})^T \in \mathbb{R}^{n+1}$ (a column vector) it holds that*

$$h(y, l, s_{\lambda,u}(l)) = p_N(\Delta) + O_{\mathbb{P}}\left(|\Delta|^3\right) + o_{\mathbb{P}}(N^{-1}), \tag{6.32}$$

uniformly in all $(y, l) \in \mathcal{Y} \times \mathcal{L}$, for the random quadratic

$$p_N(\Delta) = h + \frac{h_g}{\sqrt{N}} A_N + \frac{1}{N} C_N + \frac{1}{\sqrt{N}} \Delta \cdot V_N + \frac{1}{2}\Delta^T D \Delta, \tag{6.33}$$

where the sequences C_N, V_N of random variables are stochastically bounded and given by

$$C_N = \Lambda_N h_g + \frac{W_N^2}{2} h_{gg}, \quad V_N = \begin{pmatrix} h_{y,g} & 0 \\ h_{l,g} & h_g \end{pmatrix} \begin{pmatrix} W_N \\ W_N' \end{pmatrix} \in \mathbb{R}^{n+1}, \quad D = h_{\{y,l\},\{y,l\}} \in \mathbb{R}^{(n+1)\times(n+1)}. \tag{6.34}$$

Proof. We start by Taylor expanding in $s_{\lambda,u}(l) - s(l)$ and obtain

$$\begin{aligned} h(y, l, s_{\lambda,u}(l)) &= h(y, l, s(l)) + \partial_g h(y, l, s(l))(s_{\lambda,u}(l) - s(l)) \\ &\quad + \tfrac{1}{2}\partial_{gg} h(y, l, s(l))\left(s_{\lambda,u}(l) - s(l)\right)^2 + O\left(|s_{\lambda,u}(l) - s(l)|^3\right), \end{aligned} \tag{6.35}$$

where we used (6.4) and therefore the constant in the O term depends on $h, \mathcal{Y}, \mathcal{L}$ (as in several estimates below). Using (6.19) and Lemmas 6.5 and 6.6 it follows that

$$\begin{aligned} h(y, l, s_{\lambda,u}(l)) &= h(y, l, s(l)) + \partial_g h(y, l, s(l))\left(\tfrac{1}{\sqrt{N}} W_N(l) + \tfrac{1}{N}\Lambda_N(l)\right) \\ &\quad + \tfrac{1}{2}\partial_{gg} h(y, l, s(l))\tfrac{1}{N} W_N(l)^2 + o_{\mathbb{P}}\left(N^{-1}\right), \end{aligned} \tag{6.36}$$

uniformly in $l \in \mathcal{L}$.

Next we Taylor expand $h(y, l, s(l))$ around $(\hat{y}, \hat{l})$, giving with the shorthand notation (6.28)

$$h(y, l, s(l)) = h + h_{\{y,l\}} \cdot \Delta + \frac{1}{2}\Delta^T h_{\{y,l\},\{y,l\}}\Delta + O(|\Delta|^3). \tag{6.37}$$

Note that $h_l = \partial_l\{h(y, l, s(l))\}|_{y=\hat{y}, l=\hat{l}} = 0$ by (6.6), and $h_y = \nabla\mathcal{B}(\hat{y}) = 0$ by (6.5) and (6.13), so

$$h_{\{y,l\}} = (h_y, h_l) = 0. \tag{6.38}$$

Similarly Taylor expanding $\partial_g h(y, s(l))$ and $\partial_{gg} h(y, s(l))$ around $(\hat{y}, \hat{l})$ gives

$$\partial_g h(y, l, s(l)) = h_g + h_{\{y,l\},g} \cdot \Delta + O(|\Delta|^2) \text{ and } \partial_{gg} h(y, l, s(l)) = h_{gg} + O(|\Delta|). \tag{6.39}$$

Finally Taylor expanding $W_N(l)$ around $\hat{l}$ and using Lemma 6.5 gives that

$$W_N(l) = W_N + W_N' \Delta_l + O_{\mathbb{P}}(|\Delta^2|), \tag{6.40}$$

(recall (6.31)) uniformly in $l \in \mathcal{L}$, and using Lemma 6.6 that $\Lambda_N(l) = \Lambda_N + O_{\mathbb{P}}(|\Delta| + |\Delta|^2 N^{\frac{2}{5}})$, so that

$$\Lambda_N(l) = \Lambda_N + O_{\mathbb{P}}(|\Delta| + N|\Delta|^2 N^{-\frac{1}{2}}) \tag{6.41}$$

Combining (6.36)-(6.41) and noting that $|\Delta|^a (N^{-1/2})^b = O(|\Delta|^3 + N^{-3/2})$ for $a + b \leq 3$ we obtain (6.32). □

The next lemma computes the minimax of $p_N(\Delta)$ from (6.33).

Lemma 6.8. *For any h satisfying (6.3)-(6.8) there exist constants E_1, E_2 and a stochastically bounded sequence of random variables F_N such that p_N from (6.33) a.s. satisfies*

$$\sup_{y\in\mathbb{R}^n}\inf_{l\in\mathbb{R}} p_N(\Delta_y,\Delta_l) = E_1 + \frac{1}{\sqrt{N}}E_2 W_N + \frac{1}{N}F_N. \tag{6.42}$$

Furthermore, E_1, E_2, F_N are explicit in terms of the derivatives of h at $\hat{y}, \hat{l}$ and equal

$$E_1 = h = h(\hat{y}, s(\hat{l})) = \mathcal{B}(\hat{y}), \qquad E_2 = h_g, \qquad E_3 = h_{gg}, \quad (E_1, E_2, E_3 \in \mathbb{R}), \tag{6.43}$$

and

$$F_N = E_2\Lambda_N - \frac{1}{2}\begin{pmatrix} W_N \\ W_N' \end{pmatrix}^T G \begin{pmatrix} W_N \\ W_N' \end{pmatrix} \in \mathbb{R}, \tag{6.44}$$

where

$$G = H - \begin{pmatrix} E_3 & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}, \quad H = K^T J K + \frac{ww^T}{h_{l,l}} \in \mathbb{R}^{2\times 2}, \quad J = \nabla^2\mathcal{B}(\hat{y}) \in \mathbb{R}^{n\times n}, \tag{6.45}$$

$$K = L - \frac{h_{l,y}w^T}{h_{l,l}} \in \mathbb{R}^{n\times 2}, \quad w = \begin{pmatrix} h_{l,g} \\ E_2 \end{pmatrix} = \begin{pmatrix} h_{l,g} \\ h_g \end{pmatrix} \in \mathbb{R}^{2\times 1}, \quad L = \begin{pmatrix} h_{y,g} & 0 \end{pmatrix} \in \mathbb{R}^{n\times 2}, \tag{6.46}$$

where we view w as a column vector, and recall from (6.27)-(6.29) that $h, h_g, h_{gg}, h_{l,l}, h_{l,g} \in \mathbb{R}$ are scalars, that $h_{y,g}, h_{l,y} \in \mathbb{R}^{n\times 1}$ are column vectors and $h_{y,l} \in \mathbb{R}^{1\times n}$ is a row vector.

Proof. The expressions D and V_N from (6.34) can be written as

$$D = h_{\{y,l\},\{y,l\}} = \begin{pmatrix} h_{y,y} & h_{l,y} \\ h_{y,l} & h_{l,l} \end{pmatrix} \text{ and } V_N = \begin{pmatrix} V_{y,N} \\ V_{l,N} \end{pmatrix} \overset{(6.46)}{=} \underbrace{\begin{pmatrix} L \\ w^T \end{pmatrix}}_{\in\mathbb{R}^{(n+1)\times 2}} \begin{pmatrix} W_N \\ W_N' \end{pmatrix}, \tag{6.47}$$

(for $V_{y,N} \in \mathbb{R}^n$ and $V_{l,N} \in \mathbb{R}$; note that $h_{y,l} = h_{l,y}^T$) and Δ from (6.34) as $\Delta = \begin{pmatrix} \Delta_y \\ \Delta_l \end{pmatrix}$ for $\Delta_l \in \mathbb{R}$ and row vector $\Delta_y \in \mathbb{R}^n$. With this notation p_N can be written as

$$p_N(\Delta_y,\Delta_l) = X_N + \frac{1}{\sqrt{N}}\Delta_y\cdot v_y + \frac{1}{\sqrt{N}}\Delta_l v_l + \frac{1}{2}\Delta_y^T h_{y,y}\Delta_y + \Delta_l h_{y,l}\Delta_y + \frac{1}{2}\Delta_l^2 h_{l,l},$$

for (recalling (6.43))

$$X_N = E_1 + \frac{E_2}{\sqrt{N}}W_N + \frac{1}{N}C_N. \tag{6.48}$$

Collecting the terms involving Δ_l we can furthermore write

$$p_N(\Delta_y,\Delta_l) = X_N + \frac{1}{\sqrt{N}}\Delta_y\cdot V_{y,l} + \frac{1}{2}\Delta_y^T h_{y,y}\Delta_y + \Delta_l\left(\frac{1}{\sqrt{N}}V_{l,N} + \Delta_y^T h_{l,y}\right) + \frac{1}{2}\Delta_l^2 h_{l,l}. \tag{6.49}$$

Recalling that $h_{l,l}$ is positive by (6.7) the quadratic $\Delta_l \to p_N(\Delta_y,\Delta_l)$ with Δ_y fixed is minimized by

$$\hat{\Delta}_l = -\frac{1}{h_{l,l}}\left(\frac{1}{\sqrt{N}}V_{l,N} + \Delta_y^T h_{l,y}\right), \tag{6.50}$$

and plugging this into (6.49) gives

$$\begin{aligned} p_N(\Delta_y) := p_N(\Delta_y,\hat{\Delta}_l) &= X_N + \tfrac{1}{\sqrt{N}}\Delta_y^T V_{y,N} + \tfrac{1}{2}\Delta_y^T h_{y,y}\Delta_y - \tfrac{1}{2}\tfrac{1}{h_{l,l}}\left(\tfrac{1}{\sqrt{N}}V_{l,N} + \Delta_y^T h_{l,y}\right)^2 \\ &= X_N + \tfrac{1}{\sqrt{N}}\Delta_y^T\left(V_{y,N} - \tfrac{1}{h_{l,l}}V_{l,N}h_{l,y}\right) + \tfrac{1}{2}\Delta_y^T\left(h_{y,y} - \tfrac{h_{l,y}h_{l,y}^T}{h_{l,l}}\right)\Delta_y - \tfrac{1}{2}\tfrac{1}{N}\tfrac{1}{h_{l,l}}V_{l,N}^2 \\ &= X_N + \tfrac{1}{\sqrt{N}}\Delta_y^T K\begin{pmatrix} W_N \\ W_N' \end{pmatrix} + \tfrac{1}{2}\Delta_y^T J\Delta_y - \tfrac{1}{2}\tfrac{1}{N}\tfrac{1}{h_{l,l}}\begin{pmatrix} W_N \\ W_N' \end{pmatrix}^T ww^T\begin{pmatrix} W_N \\ W_N' \end{pmatrix}, \end{aligned} \tag{6.51}$$

where the last representation follows since by Lemma 6.3

$$J = h_{y,y} - \frac{h_{l,y} h_{l,y}^T}{h_{ll}},$$

and

$$V_{y,N} - \frac{1}{h_{ll}} V_{l,N} h_{l,y} \overset{(6.46),(6.47)}{=} \left(L - \frac{h_{l,y} w^T}{h_{ll}} \right) \begin{pmatrix} W_N \\ W_N' \end{pmatrix} = K \begin{pmatrix} W_N \\ W_N' \end{pmatrix},$$

and

$$V_{l,N}^2 \overset{(6.46),(6.47)}{=} \left(w^T \begin{pmatrix} W_N \\ W_N' \end{pmatrix} \right)^2 = \begin{pmatrix} W_N \\ W_N' \end{pmatrix}^T w w^T \begin{pmatrix} W_N \\ W_N' \end{pmatrix}.$$

We now maximize $p_N(\Delta_y)$ in Δ_y. Recall that J is negative definite by assumption. It is easily seen that $p_N(\Delta_y)$ is maximized by

$$\hat{\Delta}_y = -\frac{1}{\sqrt{N}} J^{-1} \begin{pmatrix} W_N \\ W_N' \end{pmatrix}, \tag{6.52}$$

and plugging this in yields

$$\begin{aligned} p_N(\hat{\Delta}_y) &= X_N - \tfrac{1}{2}\tfrac{1}{N} \begin{pmatrix} W_N \\ W_N' \end{pmatrix}^T K^T J^{-1} K \begin{pmatrix} W_N \\ W_N' \end{pmatrix} - \tfrac{1}{2}\tfrac{1}{N}\tfrac{1}{h_{ll}} \begin{pmatrix} W_N \\ W_N' \end{pmatrix}^T w w^T \begin{pmatrix} W_N \\ W_N' \end{pmatrix} \\ &\overset{(6.45)}{=} X_N - \tfrac{1}{N}\tfrac{1}{2} \begin{pmatrix} W_N \\ W_N' \end{pmatrix}^T H \begin{pmatrix} W_N \\ W_N' \end{pmatrix} \\ &\overset{(*)}{=} E_1 + \tfrac{1}{\sqrt{N}} E_2 W_N + \tfrac{1}{N} F_N \qquad (*:\ \text{by } (6.34),\ (6.45),\ (6.48)) \end{aligned} \tag{6.53}$$

Thus have we have proved (6.42). □

The following lemma shows that we can reduce the optimization region $\mathcal{Y} \times \mathcal{L}$ to a small neighborhood of $(\hat{y}, \hat{l})$. Let

$$\mathcal{Y}(\varepsilon) = \{ y \in \mathcal{Y} : |y - \hat{y}| < \varepsilon \} \qquad \text{and} \qquad \mathcal{L}(\varepsilon) = \left\{ l \in \mathcal{L} : |l - \hat{l}| < \varepsilon \right\}.$$

Lemma 6.9. *For all h that satisfy (6.3)-(6.8), and all $\varepsilon_1 > 0$ there is a $\delta = \delta(\varepsilon_1)$ such that if $0 < \varepsilon_2 \le \delta$ then*

$$\lim_{N\to\infty} \mathbb{P} \left(\sup_{y \in \mathcal{Y}} \inf_{l \in \mathcal{L}} h(y, l, s_{\lambda,u}(l)) = \sup_{y \in \mathcal{Y}(\varepsilon_2)} \inf_{l \in \mathcal{L}(\varepsilon_1)} h(y, l, s_{\lambda,u}(l)) \right) = 1.$$

Proof. By the continuity of h, the compactness of $\mathcal{L} \backslash \mathcal{L}(\varepsilon_1)$ and (6.6) it holds for any $\varepsilon_1 > 0$ that

$$\inf_{l \in \mathcal{L} \backslash \mathcal{L}(\varepsilon_1)} h(\hat{y}, l, s(l)) > h(\hat{y}, \hat{l}, s(\hat{l})). \tag{6.54}$$

Using uniform continuity of h on the compact $\mathcal{Y} \times (\mathcal{L} \backslash \mathcal{L}(\varepsilon_1))$ there is some $\delta > 0$ such that if $0 < \varepsilon_2 \le \delta$ then in addition

$$\inf_{l \in \mathcal{L} \backslash \mathcal{L}(\varepsilon_1)} h(y, l, s(l)) > h(y, \hat{l}, s(\hat{l})) \text{ for all } y \in \mathcal{Y}(\varepsilon_2). \tag{6.55}$$

By Lemma 4.5, (6.4) and compactness it follows that $h(y, l, s_{\lambda,u}(l)) \to h(y, l, s(l))$ in probability uniformly in $\mathcal{Y} \times \mathcal{L}$, so that (6.55) holds with $s_{\lambda,u}$ in place of s, with probability tending to one. This implies that

$$\lim_{N\to\infty} \mathbb{P} \left(\inf_{l \in \mathcal{L}} h(y, l, s_{\lambda,u}(l)) = \inf_{l \in \mathcal{L}(\varepsilon_1)} h(y, l, s_{\lambda,u}(l)) \text{ for all } y \in \mathcal{Y}(\varepsilon_2) \right) = 1. \tag{6.56}$$

Similarly to (6.54) it also follows from (6.5) that

$$\sup_{y \in \mathcal{Y} \backslash \mathcal{Y}(\varepsilon_2)} \inf_{l \in \mathcal{L}} h(y, l, s(l)) < h(\hat{y}, \hat{l}, s(\hat{l})),$$

and similary by the uniform convergence of $h(y,l,s_{\lambda,u}(l))$ it follows that

$$\lim_{N\to\infty}\mathbb{P}\left(\sup_{y\in\mathcal{Y}}\inf_{l\in\mathcal{L}}h(y,l,s_{\lambda,u}(l))=\sup_{y\in\mathcal{Y}(\varepsilon_2)}\inf_{l\in\mathcal{L}}h(y,l,s_{\lambda,u}(l))\right)=1. \tag{6.57}$$

The claim then follows by (6.56) and (6.57). ☐

Let ∂_a denote the directional derivative in the direction of a vector a. The next lemma gives conditions under which the optimizer of a minimax is given by a unique critical point.

Lemma 6.10. *Let $n\geq 1, d>0, \eta>0$, $A(d)=\{a\in\mathbb{R}^n : |a|\leq d\}$ and $t : A(d)\times[-\eta,\eta]\to\mathbb{R}$ be twice continuously differentiable. Assume $\partial_{bb}t(a,b)>0$ for all $a\in A(d), b\in[-\eta,\eta]$, and $\partial_b t(a,\eta)>0, \partial_b t(a,-\eta)<0$ for all $a\in A(d)$, and*

$$\lambda_{\max}\left(\nabla_a^2 t(a,b)-\frac{1}{\partial_{bb}t(a,b)}\partial_b\nabla_a t(a,b)\left(\partial_b\nabla_a t(a,b)\right)^T\right)<0$$

for $a\in A(d), b\in[-\eta,\eta]$ (where $\lambda_{\max}$ denotes the largest eigenvalue), and that $\partial_a\left\{\inf_{b\in[-\eta,\eta]}t(a,b)\right\}$ exists and is negative for all a with $|a|=d$. Then t has a unique critical point in $A(d)\times[-\eta,\eta]$ and $\sup_{a\in A(d)}\inf_{b\in[-\eta,\eta]}t(a,b)$ is uniquely achieved at this critical point.

Proof. This t satisfies the assumptions of Lemma 6.2, so the map $a\to b^*(a):=\operatorname{argmin}_{b\in[-\eta,\eta]}t(a,b)$ is well defined and continuously differentiable, and $\nabla_a\{t(a,b^*(a))\}=\{\nabla_a t\}(a,b^*(a))$ and

$$\nabla_a^2\{t(a,b^*(a))\}=\nabla_a^2 t(a,b^*(a))-\frac{1}{\partial_{bb}t(a,b)}\partial_b\nabla_a t(a,b^*(a))\left(\partial_b\nabla_a t(a,b^*(a))\right)^T \text{ for all } a\in A(d).$$

By assumption this is negative-definite for all $a\in A(d)$, implying that if $a\to t(a,b^*(a))$ is concave and therefore if not maximized on the boundary of $A(d)$, it has a unique critical point in the interior which is the maximizer. Since the assumption $\partial_a\left\{\inf_{b\in[-\eta,\eta]}t(a,b)\right\}<0$ rules out the maximizer lying on the boundary, and (a,b) is a critical point of t iff $b=b^*(a)$ and a is a critical point of $a\to t(a,b^*(a))$, this proves the claim. ☐

We can now strengthen Lemma 6.9.

Lemma 6.11. *It holds that*

$$\lim_{N\to\infty}\mathbb{P}\left(\sup_{y\in\mathcal{Y}}\inf_{l\in\mathcal{L}}h(y,l,s_{\lambda,u}(l))=\sup_{y\in\mathcal{Y}\left(\frac{\log N}{\sqrt{N}}\right)}\inf_{l\in\mathcal{L}\left(\frac{(\log N)^2}{\sqrt{N}}\right)}h(y,l,s_{\lambda,u}(l))\right)=1.$$

Proof. By Lemma 6.9 there is for for each $\varepsilon_2>0$ small enough an $\varepsilon_1>0$ small enough so that

$$\lim_{N\to\infty}\mathbb{P}\left(\sup_{y\in\mathcal{Y}}\inf_{l\in\mathcal{L}}h(y,l,s_{\lambda,u}(l))=\sup_{y\in\mathcal{Y}(\varepsilon_2)}\inf_{l\in\mathcal{L}(\varepsilon_1)}h(y,l,s_{\lambda,u}(l))\right)=1. \tag{6.58}$$

Furthermore, for each $\varepsilon_2>0$ small enough, there is an $\varepsilon_1>0$ small enough such that

$$\partial_{ll}\{h(y,l,s(l))\}>0,\ \partial_l\{h(y,l,s(l))\}\big|_{\hat{l}-\varepsilon_1}>0,\ \partial_l\{h(y,l,s(l))\}\big|_{\hat{l}-\varepsilon_1}<0,$$

for all $y\in\mathcal{Y}(\varepsilon_2), l\in\mathcal{L}(\varepsilon_1)$ (see (6.6) and (6.7)), and

$$\lambda_{\max}\left(\nabla_y^2 h(y,l,s(l))-\frac{\partial_l\nabla_y h(y,l,s(l))(\partial_l\nabla_y h(y,l,s(l)))^T}{\partial_{ll}h(y,l,s(l))}\right)<0, \tag{6.59}$$

for all $y\in\mathcal{Y}(\varepsilon_2), l\in\mathcal{L}(\varepsilon_1)$ (see (6.8)), and since $\hat{y}$ is the unique maximum (see (6.5))

$$\partial_{(\hat{y}-y)}\{h(y,l,s(l))\}>0 \text{ for all } y \text{ s.t. } |y-\hat{y}|=\varepsilon_2.$$

By Lemma 4.5 and (6.4) the same holds with $s_{\lambda,u}(l)$ in place of $s(l)$, on an event with probability tending to one. Therefore by applying Lemma 6.10 to $(a,b) \to h(\hat{y}+a, \hat{l}+b, s_{\lambda,u}(\hat{l}+b))$ on this event one obtains that

$$\lim_{N\to\infty} \mathbb{P}\left(\begin{array}{c} (y,l) \to h(y,l,s_{\lambda,u}(l)) \text{ has a unique critical point } (y^*,l^*) \text{ in } \mathcal{Y}(\varepsilon_2)\times\mathcal{L}(\varepsilon_1) \\ \text{and } \sup_{y\in\mathcal{Y}(\varepsilon_2)} \inf_{l\in\mathcal{L}(\varepsilon_1)} h(y,l,s_{\lambda,u}(l)) \text{ is achieved at } (y^*,l^*) \end{array} \right) = 1. \tag{6.60}$$

By the Schur complement formula and (6.59) it holds that $\nabla^2_{y,l}\{h(y,l,s(l))\}$ is non-degenerate, and $\nabla_{y,l}\{h(\hat{y},\hat{l},s(\hat{l}))\} = h_{\{y,l\}} = 0$ as stated in (6.38), so for $\varepsilon_2, \varepsilon_1 > 0$ small enough there is a constant c such that

$$|\nabla_{y,l}\{h(y,l,s(l))\}| \geq c(|y-\hat{y}|^2 + |l-\hat{l}|^2) \text{ for all } (y,l)\in\mathcal{Y}(\varepsilon_2)\times\mathcal{L}(\varepsilon_1).$$

Since $|\nabla h(y,l,s(l)) - \nabla h(y,l,s_{\lambda,u}(l))| = O_{\mathbb{P}}(N^{-1/2})$ by (6.26) it follows that

$$\lim_{N\to\infty} \mathbb{P}\left(\begin{array}{c} h(y,l,s_{\lambda,u}(l)) \text{ has no critical point in } \mathcal{Y}(\varepsilon_2)\times\mathcal{L}(\varepsilon_1) \\ \text{with } |y-\hat{y}| \geq \frac{\log N}{N^{1/2}}, |l-\hat{l}| \geq \frac{(\log N)^2}{N^{1/2}} \end{array} \right) = 1. \tag{6.61}$$

Since we can pick $\varepsilon_2 > 0$ and then $\varepsilon_1 > 0$ small enough so that (6.58), (6.60), (6.61) hold simultaneously the claim follows. □

We can now prove a version of Lemma 6.8 for the actual function $h(y,l,s_{\lambda,u}(l))$ rather than its quadratic expansion.

Proposition 6.12. *For any h satisfying (6.3)-(6.8) it holds that*

$$\sup_{y\in\mathcal{Y}} \inf_{l\in\mathcal{L}} h(y,l,s_{\lambda,u}(l)) = E_1 + \frac{1}{\sqrt{N}} E_2 W_N + \frac{1}{N} F_N + o_{\mathbb{P}}\left(\frac{1}{N}\right), \tag{6.62}$$

for E_1, E_2, E_3, F_N as in (6.43)-(6.44).

Proof. By Lemma 6.7 and Lemma 6.11

$$\sup_{y\in\mathcal{Y}} \inf_{l\in\mathcal{L}} h(y,l,s_{\lambda,u}(l)) = \sup_{y\in\mathcal{Y}(N^{-1/2}\log N)} \inf_{l\in\mathcal{L}(N^{-1/2}(\log N)^2)} p_N(\Delta_y,\Delta_l) + o_{\mathbb{P}}(N^{-1}).$$

Recall from the proof of Lemma 6.8 that $\hat{\Delta}_l$ from (6.50) is the minimizer of $\inf_{l\in\mathbb{R}} p_N(\Delta_y,\Delta_l)$. Note that for all $y\in\mathcal{Y}(N^{-1/2}\log N)$ it holds that $\mathbb{P}(|\hat{\Delta}_l| \leq N^{-1/2}(\log N)^2) \to 1$, so

$$\lim_{N\to\infty} \mathbb{P}\left(\inf_{l\in\mathcal{L}(N^{-1/2}(\log N)^2)} p_N(\Delta_y,\Delta_l) = p_N(\Delta_y,\hat{\Delta}_l) \right) = 1.$$

Similarly recall that $\hat{\Delta}_y$ from (6.52) is the maximizer of $\sup_{y\in\mathbb{R}^n}\inf_{l\in\mathbb{R}} p_N(\Delta_y,\Delta_l)$ and note that $\mathbb{P}(|\hat{\Delta}_y| \leq N^{-1/2}\log N) \to 1$ so that furthermore

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{y\in\mathcal{Y}(N^{-1/2}\log N)} \inf_{l\in\mathcal{L}(N^{-1/2}(\log N)^2)} p_N(\Delta_y,\Delta_l) = p_N(\hat{\Delta}_y,\hat{\Delta}_l) \right) = 1.$$

Thus the claim follows from (6.42). □

The next lemma computes the distributional limit of $(W_N(l), W_N'(l))$.

Lemma 6.13. *For all l it holds that*

$$(W_N(l), W_N'(l)) \xrightarrow{d} (U(l), U'(l)), \tag{6.63}$$

where $(U(l), U'(l))$ is a centered Gaussian vector with covariance matrix

$$\Sigma = \begin{pmatrix} -2s'(l) - 2s(l)^2 & -s''(l) - 2s(l)s'(l) \\ -s''(l) - 2s(l)s'(l) & -\frac{1}{3}s'''(l) - 2s'(l)^2 \end{pmatrix}.$$

Proof. Define

$$\tilde{W}_N^{(k)}(l) = \frac{1}{\sqrt{N}} \sum_{i=1}^N (N\tilde{u}_i^2 - 1)\left(\frac{k!(-1)^k}{(l-\theta_{i/N})^{k+1}} - s_\theta^{(k)}(l)\right).$$

Then

$$\begin{aligned}\mathbb{E}\left[\tilde{W}_N^{(k)}(l)\tilde{W}_N^{(k')}(l)\right] &= \frac{2}{N}\sum_{i=1}^N \left(\frac{k!(-1)^k}{(l-\theta_{i/N})^{k+1}} - s_\theta^{(k)}(l)\right)\left(\frac{k'!(-1)^{k'}}{(l-\theta_{i/N})^{k'+1}} - s_\theta^{(k')}(l)\right) \\ &= 2\left(\frac{1}{N}\sum_{i=1}^N \frac{k!k'!(-1)^{k+k'}}{(l-\theta_{i/N})^{k+k'+2}} - s_\theta^{(k)}(l)s_\theta^{(k')}(l)\right) \\ &\xrightarrow{(2.4)} -2\frac{k!k'!}{(k+k'+1)!}s^{(k+k'+1)}(l) - 2s^{(k)}(l)s^{(k')}(l).\end{aligned} \tag{6.64}$$

Note that for all $t = (t_1, t_2) \in \mathbb{R}^2$

$$\mathbb{E}\left[t_1\tilde{W}_N(l) + t_2\tilde{W}'_N(l)\right] \longrightarrow 0,$$

and by (6.64)

$$\mathbb{E}\left[\left(t_1\tilde{W}_N(l) + t_2\tilde{W}'_N(l)\right)^2\right] \longrightarrow t^T\Sigma t.$$

Therefore

$$t_1\tilde{W}_N(l) + t_2\tilde{W}'_N(l) \xrightarrow{d} t_1U(l) + t_2U'(l) \sim \mathcal{N}\left(0, t^T\Sigma t\right)$$

by Lyapunov's CLT (see Lindeberg's theorem [ADD99, Theorem 7.3.1 and Lyapunov's condition p. 307-309]; note that $\sum_{i=1}^N E[|(N\tilde{u}_i^2 - 1)(t_1/(l-\theta_{i/N}) - t_2/(l-\theta_{i/N}))|^3] = O(N)$ while $\mathrm{Var}(t_1\tilde{W}_N(l) + t_2\tilde{W}'_N(l))^{3/2} = (\sum_{i=1}^N 2(t_1/(l-\theta_{i/N}) - t_2/(l-\theta_{i/N}))^2)^{3/2} = O(N^{3/2})$, so Lyapunov's condition is satisfied).
By (6.21) and Slutzky's theorem thus also

$$t_1W_N(l) + t_2W'_N(l) \xrightarrow{d} t_1U(l) + t_2U'(l) \sim \mathcal{N}\left(0, t^T\Sigma t\right)$$

for all $t \in \mathbb{R}^3$. By the Cramér-Wold theorem [Kal21, Corollary 6.5] one obtains the joint convergence (6.63). ☐

We also compute the distributional limit of Λ_N.

Lemma 6.14. *For any $l > \sqrt{2}$*

$$\Lambda_N(l) \xrightarrow{d} \mathcal{N}\left(\frac{l - \sqrt{l^2-2}}{2(l^2-2)}, \frac{1}{(l^2-2)^2}\right)$$

as $N \to \infty$.

Proof. By Lemma 2.3 the random variable Λ_N converges in law to a normal distribution with mean

$$m(w) = \frac{\frac{1}{l-\sqrt{2}} + \frac{1}{l+\sqrt{2}}}{4} - \frac{1}{2\pi}\int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{(l-x)\sqrt{2-x^2}}dx \tag{6.65}$$

and variance

$$v(w) = \frac{1}{2\pi^2}\int_{-\sqrt{2}}^{\sqrt{2}}\int_{-\sqrt{2}}^{\sqrt{2}} \left(\frac{\frac{1}{l-x} - \frac{1}{l-y}}{x-y}\right)^2 \frac{2-xy}{\sqrt{2-x^2}\sqrt{2-y^2}}dxdy \tag{6.66}$$

with $w(x) = \frac{1}{l-x}$. It only remains to compute the integrals in (6.65)-(6.66).

First, note that for any $k \in \mathbb{N}$ by integration by parts

$$\int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{(l-x)^k}\frac{x}{\pi\sqrt{2-x^2}}dx = \int_{-\sqrt{2}}^{\sqrt{2}} \frac{k}{(l-x)^{k+1}}\mu_{\mathrm{sc}}(dx) = \frac{(-1)^k}{(k-1)!}s^{(k)}(l) \tag{6.67}$$

and also

$$\begin{aligned} \int_{-\sqrt{2}}^{\sqrt{2}} \tfrac{1}{l-x} \tfrac{1}{\pi\sqrt{2-x^2}} dx &= \tfrac{1}{l} \int_{-\sqrt{2}}^{\sqrt{2}} \left(\tfrac{x}{l-x}+1\right) \tfrac{1}{\pi} \tfrac{1}{\sqrt{2-x^2}} dx \\ &\overset{(6.67)}{=} \tfrac{1}{l}\left(-s^{(1)}(l)+1\right) \\ &\overset{(4.5)}{=} \tfrac{1}{\sqrt{l^2-2}} \end{aligned} \tag{6.68}$$

as well as

$$\begin{aligned} \int_{-\sqrt{2}}^{\sqrt{2}} \tfrac{1}{(l-x)^2} \tfrac{1}{\pi\sqrt{2-x^2}} dx &= \tfrac{1}{l} \int_{-\sqrt{2}}^{\sqrt{2}} \tfrac{1}{l-x} \left(\tfrac{x}{l-x}+1\right) \tfrac{1}{\pi\sqrt{2-x^2}} dx \\ &\overset{(6.67),(6.68)}{=} \tfrac{1}{l}\left(s^{(2)}(l)+\tfrac{1}{\sqrt{l^2-2}}\right) \\ &\overset{(4.5)}{=} \tfrac{l}{(l^2-2)^{\frac{3}{2}}}. \end{aligned} \tag{6.69}$$

Therefore the expectation of the limiting distribution is

$$m(w) \overset{(6.68)}{=} \frac{l}{2(l^2-2)} - \frac{1}{2\sqrt{l^2-2}} = \frac{l-\sqrt{l^2-2}}{2(l^2-2)}. \tag{6.70}$$

The variance on the other hand is given by

$$\begin{aligned} &\frac{1}{2\pi^2} \int_{-\sqrt{2}}^{\sqrt{2}} \int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{(l-x)^2(l-y)^2} \frac{2-xy}{\sqrt{2-x^2}\sqrt{2-y^2}} dxdy \\ =&\frac{1}{2\pi} \int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{(l-y)^2\sqrt{2-y^2}} \int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{(l-x)^2} \frac{2-xy}{\pi\sqrt{2-x^2}} dxdy, \end{aligned} \tag{6.71}$$

where the inner integral is by (6.67) and (6.69)

$$\frac{2l}{(l^2-2)^{\frac{3}{2}}} - 2ys^{(2)}(l) \overset{(4.5)}{=} \frac{2(l-y)}{(l^2-2)^{\frac{3}{2}}}. \tag{6.72}$$

Therefore the variance is

$$\frac{1}{\pi} \int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{(l-y)^2\sqrt{2-y^2}} \frac{l-y}{(l^2-2)^{\frac{3}{2}}} dy = \frac{1}{(l^2-2)^{\frac{3}{2}}} \int_{-\sqrt{2}}^{\sqrt{2}} \frac{1}{l-y} \frac{1}{\pi\sqrt{2-y^2}} dy \overset{(6.68)}{=} \frac{1}{(l^2-2)^2}.$$

□

6.4. Derivation of main fluctuation results

Now we are ready to prove Theorem 1.1 (b) and Theorem 1.2 (b). Before giving the proof, we state the following simplified versions of (4.5) using (4.19):

$$s^{(k)}(\hat{l}(\alpha)) = \begin{cases} \hat{l}(\alpha) - \sqrt{\hat{l}(\alpha)^2-2} &= \sqrt{2(1-\alpha^2)} & \text{for } k=0, \\ -\frac{\hat{l}-\sqrt{\hat{l}(\alpha)^2-2}}{\sqrt{\hat{l}(\alpha)^2-2}} &= -\frac{2(1-\alpha^2)}{\alpha^2} & \text{for } k=1, \\ \frac{2}{(\hat{l}(\alpha)^2-2)^{\frac{3}{2}}} &= \frac{2(2(1-\alpha^2))^{\frac{3}{2}}}{\alpha^6} & \text{for } k=2, \\ -\frac{6\hat{l}(\alpha)}{(\hat{l}(\alpha)^2-2)^{\frac{5}{2}}} &= -\frac{24(2-\alpha^2)(1-\alpha^2)^2}{\alpha^{10}} & \text{for } k=3. \end{cases} \tag{6.73}$$

Using this with $\alpha = \hat{\alpha}$

$$s(\hat{l}) = \hat{z}, \quad s'(\hat{l}) = -\frac{\hat{z}^2}{\hat{\alpha}^2}, \quad s''(\hat{l}) = 2\frac{\hat{z}^3}{\hat{\alpha}^6}, \quad s'''(\hat{l}) = -\frac{6(2-\alpha^2)\hat{z}^4}{\hat{\alpha}^{10}}, \quad \text{where } \hat{z} = \sqrt{2\,(1-\hat{\alpha}^2)}. \tag{6.74}$$

Proof of Theorem 1.1 (b). Applying Lemma 6.1 and Proposition 6.12 with

$$h(\alpha, l, g) = f(\alpha) + \beta\left(l - \frac{\alpha^2}{g}\right), \tag{6.75}$$

we obtain

$$\frac{1}{N} L_N = E_1 + \frac{1}{\sqrt{N}} E_2 W_N + \frac{1}{N} F_N + o_{\mathbb{P}}\left(\frac{1}{N}\right). \tag{6.76}$$

Note that

$$U_N^{(k)} - W_N^{(k)} \overset{(6.17)}{=} \frac{1}{\sqrt{N}} R_N^{(k)}(l) = o_{\mathbb{P}}(N^{-1/2})$$

by Lemma 6.4. It follows that (U_N, U'_N) and (W_N, W'_N) have the same limit, and that we can swap all W_N for U_N and W'_N for U'_N in (6.76) at the cost of a negligible error.

The remainder of the proof will revolve around computing $E_1, E_2, E_3, J, L, W, K, G$ of Proposition 6.12. Note first that

$$E_1 = \mathcal{B}(\hat{\alpha}) \quad \text{and} \quad J = \frac{1}{\mathcal{B}''(\hat{\alpha})}.$$

Furthermore for the h in (6.75) we obtain with $\hat{z} = \sqrt{2(1-\hat{\alpha}^2)}$

$$E_2 = h_g = \frac{\beta\hat{\alpha}^2}{s(\hat{l})^2} = \frac{\beta\hat{\alpha}^2}{\hat{z}^2} = \kappa, \quad E_3 = h_{gg} = -\frac{2\beta\hat{\alpha}^2}{s(\hat{l})^3} = -\frac{2\beta\hat{\alpha}^2}{\hat{z}^3}$$

as well as

$$\begin{aligned} h_{l,\alpha} &= \frac{2\beta\hat{\alpha}s'(\hat{l})}{s(\hat{l})^2} = -\frac{2\beta}{\hat{\alpha}}, & h_{ll} &= \beta\hat{\alpha}^2\left(\frac{s''(\hat{l})}{s(\hat{l})^2} - \frac{2s'(\hat{l})^2}{s(\hat{l})^3}\right) = \beta\frac{\hat{z}^3}{\hat{\alpha}^4}, \\ h_{l,g} &= -\frac{2\beta\hat{\alpha}^2 s'(\hat{l})}{s(\hat{l})^3} = \frac{2\beta}{\hat{z}}, & h_{\alpha,g} &= \frac{2\beta\hat{\alpha}}{s(\hat{l})^2} = \frac{2\beta\hat{\alpha}}{\hat{z}^2}, \end{aligned}$$

which gives

$$L = \begin{pmatrix} h_{\alpha,g} & 0 \end{pmatrix} = \beta\begin{pmatrix} \frac{2\hat{\alpha}}{\hat{z}^2} & 0 \end{pmatrix} \in \mathbb{R}^{1\times 2}, \quad w = \begin{pmatrix} h_{l,g} \\ E_2 \end{pmatrix} = \begin{pmatrix} h_{l,g} \\ h_g \end{pmatrix} = \frac{\beta}{\hat{z}}\begin{pmatrix} 2 \\ \frac{\hat{\alpha}^2}{\hat{z}} \end{pmatrix} \in \mathbb{R}^2,$$

$$K = L - \frac{h_{l,\alpha} w^T}{h_{ll}} = \beta\begin{pmatrix} \frac{2\hat{\alpha}}{\hat{z}^2} & 0 \end{pmatrix} + \frac{2\beta\hat{\alpha}^3}{\hat{z}^4}\begin{pmatrix} 2 & \frac{\hat{\alpha}^2}{\hat{z}} \end{pmatrix} = \frac{2\beta\hat{\alpha}}{\hat{z}^4}\begin{pmatrix} 2 & \frac{\hat{\alpha}^4}{\hat{z}} \end{pmatrix} \in \mathbb{R}^{1\times 2}$$

$$H = \frac{K^T K}{\mathcal{B}''(\hat{\alpha})} = \frac{8\beta^2\hat{\alpha}^2}{\hat{z}^8\mathcal{B}''(\hat{\alpha})}\begin{pmatrix} 2 & \frac{\hat{\alpha}^4}{\hat{z}} \\ \frac{\hat{\alpha}^4}{\hat{z}} & \frac{\hat{\alpha}^8}{2\hat{z}^2} \end{pmatrix} \in \mathbb{R}^{2\times 2},$$

$$G = H - \begin{pmatrix} E_3 & 0 \\ 0 & 0 \end{pmatrix} = \frac{8\beta^2\hat{\alpha}^2}{\hat{z}^8\mathcal{B}''(\hat{\alpha})}\begin{pmatrix} 2 & \frac{\hat{\alpha}^4}{\hat{z}} \\ \frac{\hat{\alpha}^4}{\hat{z}} & \frac{\hat{\alpha}^8}{2\hat{z}^2} \end{pmatrix} + \frac{2\beta\hat{\alpha}^2}{\hat{z}^3}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and finally

$$F_N = E_2\Lambda_N - \frac{1}{2}\begin{pmatrix} U_N \\ U'_N \end{pmatrix}^T G \begin{pmatrix} U_N \\ U'_N \end{pmatrix}.$$

This proves (1.5).

The joint convergence in law of W_N, W'_N, Λ_N follows from Lemma 6.13, Lemma 6.14 and since Λ_N is independent from (W_N, W'_N) for all N. Note that using (6.73) the matrix Σ can be simplified to

$$\Sigma = \begin{pmatrix} \frac{4(1-\hat{\alpha}^2)^2}{\hat{\alpha}^2} & -\frac{4\sqrt{2}\sqrt{1-\hat{\alpha}^2}^5(1+\hat{\alpha}^2)}{\hat{\alpha}^6} \\ -\frac{4\sqrt{2}\sqrt{1-\hat{\alpha}^2}^5(1+\hat{\alpha}^2)}{\hat{\alpha}^6} & \frac{8(1-\hat{\alpha}^2)^3(2+\hat{\alpha}^2+\hat{\alpha}^4)}{\hat{\alpha}^{10}} \end{pmatrix},$$

while the limiting distribution of Λ_N is given by

$$\mathcal{N}\left(\frac{\hat{\imath}(\hat{\alpha}) - \sqrt{\hat{\imath}(\hat{\alpha})^2 - 2}}{2(\hat{\imath}(\hat{\alpha})^2 - 2)}, \frac{1}{(\hat{\imath}(\hat{\alpha})^2 - 2)^2}\right) \overset{(4.19)}{=} \mathcal{N}\left(2\frac{\sqrt{2}(1-\hat{\alpha}^2)^{\frac{3}{2}}}{\hat{\alpha}^4}, \frac{4(1-\hat{\alpha}^2)^2}{\hat{\alpha}^8}\right). \tag{6.77}$$

□

Proof of Theorem 1.2 (b). As in the previous proof we apply Lemma 6.1 and Proposition 6.12, this time with

$$h((\alpha,r),l,g)=f(\alpha r)+g(r)+\beta r^2\left(l-\frac{\alpha^2}{g}\right), \tag{6.78}$$

and also use that $U_N^{(k)}-W_N^{(k)}=o_{\mathbb{P}}(N^{-1/2})$ to exchange W_N,W_N' for U_N,U_N', yielding

$$\frac{1}{N}\tilde{L}_N=E_1+\frac{1}{\sqrt{N}}E_2W_N+\frac{1}{N}F_N+o_{\mathbb{P}}\left(\frac{1}{N}\right).$$

Now

$$E_1=\tilde{\mathcal{B}}(\hat{\alpha},\hat{r})\quad\text{and}\quad J=\left(\nabla^2\tilde{\mathcal{B}}((\hat{\alpha},\hat{r}))\right)^{-1},$$

and for the h in (6.78) with $\hat{z}=\sqrt{2(1-\hat{\alpha}^2)}$

$$E_2=h_g=\beta\frac{\hat{r}^2\hat{\alpha}^2}{s(\hat{l})^2}=\beta\frac{\hat{r}^2\hat{\alpha}^2}{\hat{z}^2}=\tilde{\kappa},\quad E_3=h_{gg}=-2\beta\frac{\hat{r}^2\hat{\alpha}^2}{s(\hat{l})^3}=-2\beta\frac{\hat{r}^2\hat{\alpha}^2}{\hat{z}^3},$$

$$h_{l,y}=\begin{pmatrix}2\beta\hat{r}^2\frac{\hat{\alpha}}{s(l)^2}s'(l)\\ 2\beta\hat{r}\left(1+\frac{\hat{\alpha}^2}{s(l)^2}s'(l)\right)\end{pmatrix}=\begin{pmatrix}2\beta\hat{r}^2\frac{\hat{\alpha}}{\hat{z}^2}s'(l)\\ 2\beta\hat{r}\left(1+\frac{\hat{\alpha}^2}{\hat{z}^2}s'(l)\right)\end{pmatrix}=\begin{pmatrix}-2\beta\hat{r}^2\frac{\hat{\alpha}}{\hat{z}^2}\frac{\hat{z}^2}{\hat{\alpha}^2}\\ 2\beta\hat{r}\left(1-\frac{\hat{\alpha}^2}{\hat{z}^2}\frac{\hat{z}^2}{\hat{\alpha}^2}\right)\end{pmatrix}=\begin{pmatrix}-\frac{2\beta\hat{r}^2}{\hat{\alpha}}\\ 0\end{pmatrix},$$

$$h_{ll}=-2\beta\hat{r}^2\frac{\hat{\alpha}^2}{s(l)^3}s'(l)^2+\beta\hat{r}^2\frac{\hat{\alpha}^2}{s(l)^2}s''(l)=\beta\hat{r}^2\left(\frac{2\hat{z}}{\hat{\alpha}^4}-\frac{2\hat{z}}{\hat{\alpha}^2}\right)=\beta\frac{\hat{r}^2\hat{z}^3}{\hat{\alpha}^4},$$

$$h_{l,g}=-\frac{2\beta\hat{\alpha}^2\hat{r}^2s'(\hat{l})}{s(\hat{l})^3}=\frac{2\beta\hat{r}^2}{\hat{z}},\quad h_{\alpha,g}=\frac{2\beta\hat{\alpha}\hat{r}^2}{s(\hat{l})^2}=\frac{2\beta\hat{\alpha}\hat{r}^2}{\hat{z}^2},\quad h_{r,g}=\frac{2\beta\hat{\alpha}^2\hat{r}}{s(\hat{l})^2}=\frac{2\beta\hat{\alpha}^2\hat{r}}{\hat{z}^2},$$

which gives

$$L=\begin{pmatrix}h_{y,g} & 0\end{pmatrix}=\frac{2\beta\hat{\alpha}\hat{r}}{\hat{z}^2}\begin{pmatrix}\hat{r} & 0\\ \hat{\alpha} & 0\end{pmatrix}\in\mathbb{R}^{2\times2},\quad w=\begin{pmatrix}h_{l,g}\\ h_g\end{pmatrix}=\frac{\beta\hat{r}^2}{\hat{z}}\begin{pmatrix}2\\ \frac{\hat{\alpha}^2}{\hat{z}}\end{pmatrix}\in\mathbb{R}^2$$

$$K=L-\frac{h_{l,y}w^T}{h_{ll}}=L-\frac{\hat{\alpha}^4}{2\beta\hat{r}^2\hat{z}^3}\frac{-2\beta^2\hat{r}^4}{\hat{z}}\begin{pmatrix}\frac{2}{\hat{\alpha}} & \frac{\hat{\alpha}}{\hat{z}}\\ 0 & 0\end{pmatrix}=\frac{2\beta\hat{r}\hat{\alpha}}{\hat{z}^2}\begin{pmatrix}\frac{2\hat{r}}{\hat{z}^2} & \frac{\hat{r}\hat{\alpha}^4}{\hat{z}^3}\\ \hat{\alpha} & 0\end{pmatrix}\in\mathbb{R}^{2\times2}.$$

Furthermore we obtain

$$H=K^T\left(\nabla^2\mathcal{B}((\hat{\alpha},\hat{r}))\right)^{-1}K,\quad G=H-\begin{pmatrix}E_3 & 0\\ 0 & 0\end{pmatrix}=K^T\left(\nabla^2\mathcal{B}((\hat{\alpha},\hat{r}))\right)^{-1}K+\begin{pmatrix}2\beta\frac{\hat{r}^2\hat{\alpha}^2}{\hat{z}^3} & 0\\ 0 & 0\end{pmatrix},$$

and finally

$$F_N=E_2\Lambda_N-\frac{1}{2}\begin{pmatrix}W_N\\ W_N'\end{pmatrix}^T G\begin{pmatrix}W_N\\ W_N'\end{pmatrix}.$$

This proves (1.11). □

Remark 6.15. *Theorem 1.1 (b) and Theorem 1.2 (b) were stated in terms of the sums U_N,U_N' over the random vector u with weakly dependent but not independent entries. It may be more natural to write the result instead in terms of sums of truly independent summands. This can be done if one constructs u from i.i.d. $\tilde{u}_1,\dots,\tilde{u}_N$ as we did in the proofs of Lemma 6.5 and Lemma 6.13. If we define*

$$X_N^{(k)}=\frac{1}{\sqrt{N}}\sum_{i=1}^N(N\tilde{u}_i^2-1)\left(\frac{k!(-1)^k}{(\hat{l}-\theta_{i/N})^{k+1}}-s^{(k)}(l)\right) \tag{6.79}$$

$$Y_N=\frac{1}{\sqrt{N}}\sum_{i=1}^N(N\tilde{u}_i^2-1)$$

one can verify that

$$W_N^{(k)} = X_N^{(k)} - \frac{1}{\sqrt{N}} X_N^{(k)} Y_N + o_{\mathbb{P}}(N^{-1/2}). \tag{6.80}$$

Theorem 1.1 (b) can then be reformulated as

$$L_N - N\mathcal{B}(\hat\alpha) - \sqrt{N}\kappa X_N - \left(\kappa\Lambda_N - \kappa X_N Y_N - \frac{1}{2}\begin{pmatrix} X_N \\ X_N' \end{pmatrix}^T G \begin{pmatrix} X_N \\ X_N' \end{pmatrix}\right) \xrightarrow{\mathbb{P}} 0, \tag{6.81}$$

where the random variables satisfy

$$(X_N, X_N', Y_N, \Lambda_N) \xrightarrow{d} (X, X', Y, \Lambda),$$

where with $\hat z = \sqrt{2(1-\hat\alpha^2)}$

$$X \sim \mathcal{N}\left(0, \frac{\hat z^4}{\hat\alpha^2}\right), \qquad X' \sim \mathcal{N}\left(0, \frac{\hat z^6(2+\hat\alpha^2+\hat\alpha^4)}{\hat\alpha^{10}}\right),$$
$$Y \sim \mathcal{N}(0,2), \qquad \Lambda \sim \mathcal{N}\left(\frac{\hat z^3}{2\hat\alpha^4}, \frac{\hat z^4}{\hat\alpha^8}\right), \tag{6.82}$$

with (X, X')*,* Y *and* Λ *mutually independent and*

$$\mathrm{Cov}(X, X') = -\frac{\hat z^5(1+\hat\alpha^2)}{\hat\alpha^6}.$$

The constant κ *and matrix* G *are the same as before. Comparing the estimate (1.5) in terms of* U_N, U_N' *and (6.81) one sees that the extra term* $\kappa X_N Y_N$ *of order one appears, which arises from the* $N^{-1/2}$ *correction in (6.80). Note furthermore that (6.81) would remain true if one defined* $X_N^{(k)}$ *with the random eigenvalues* λ_i *instead of deterministic classical locations* $\theta_{i/N}$ *in (6.79).*

Similarly Theorem 1.2 (b) can be formulated as

$$\tilde L_N - \tilde{\mathcal{B}}(\hat\alpha, \hat r) - \sqrt{N}\kappa X_N - \left(\kappa\Lambda_N - \kappa X_N Y_N - \frac{1}{2}\begin{pmatrix} X_N \\ X_N' \end{pmatrix}^T \tilde G \begin{pmatrix} X_N \\ X_N' \end{pmatrix}\right) \xrightarrow{\mathbb{P}} 0, \tag{6.83}$$

where X_N, X_N, Y_N, Λ_N *are as in (6.81).*

7. Examples: Subleading order

We showed in Section 5 that for $f(x) = hx^k$ and $\beta < \beta_c(k,h)$ (see (5.1)) the function $\mathcal{B}(\alpha)$ has a unique maximizer in $[0,1]$. Theorem 1.1 (b) requires also that $\mathcal{B}''(\hat\alpha) < 0$, which the next lemma shows is always satisfied.

Lemma 7.1. *Let* $h \in \mathbb{R}^+$*,* $k \in \mathbb{N}$*,* $f(x) = hx^k$*. If* $\beta < \beta_c(k,h)$ *then all global maximizers* $\hat\alpha \in \arg\max_{\alpha\in(-1,1)} \mathcal{B}(\alpha)$ *satisfy* $\mathcal{B}''(\hat\alpha) < 0$*.*

Proof. By Lemma 5.1 there is a unique maximizer $\hat\alpha \in (0,1)$ for $\beta < \beta_c(k,h)$. Note that we must have $\mathcal{B}''(\hat\alpha) \le 0$, so we only have to prove that $\mathcal{B}''(\hat\alpha) \neq 0$. In the case $k = 1$ we have $\mathcal{B}''(\alpha) = -\frac{\sqrt{2}\beta}{(1-\alpha^2)^{\frac{3}{2}}} < 0$ for all $\alpha \in (-1,1)$. In the case $k = 2$ we have by Lemma 5.1 that $\hat\alpha^2 = 1 - \frac{\beta^2}{2h^2}$ and thus

$$\mathcal{B}''(\hat\alpha) = 2h - \frac{\sqrt{2}\beta}{\left(1 - \left(1 - \frac{\beta^2}{2h^2}\right)\right)^{\frac{3}{2}}} = \frac{2h}{\beta^2}(\beta^2 - 2h^2) < 0$$

for all $\beta < \beta_c(2) = \sqrt{2}h$. In the case $k \ge 3$ note that for any critical $\alpha \in (0,1)$

$$\mathcal{B}''(\alpha) \overset{(5.3)}{=} (k-1)\frac{\sqrt{2}\beta}{\sqrt{1-\alpha^2}} - \frac{\sqrt{2}\beta}{(1-\alpha^2)^{\frac{3}{2}}} = \frac{\sqrt{2}\beta}{\sqrt{1-\alpha^2}}\left(k - 1 - \frac{1}{1-\alpha^2}\right), \tag{7.1}$$

which can only be equal to zero if $\alpha^2 = \frac{k-2}{k-1}$. Thus, it remains to show that $\alpha = \sqrt{\frac{k-2}{k-1}}$ is not the global maximizer of $\mathcal{B}$. Now suppose we have

$$\mathcal{B}'\left(\sqrt{\tfrac{k-2}{k-1}}\right) = \mathcal{B}''\left(\sqrt{\tfrac{k-2}{k-1}}\right) = 0, \tag{7.2}$$

and note that for any critical $\alpha \in (0,1)$ the third derivative is

$$\begin{aligned} \mathcal{B}'''(\alpha) &= hk(k-1)(k-2)\alpha^{k-3} - 3\sqrt{2}\beta\tfrac{\alpha}{(1-\alpha^2)^{\frac{5}{2}}} \\ &\overset{(5.3)}{=} \tfrac{\sqrt{2}\beta\alpha}{\sqrt{1-\alpha^2}}\left(\tfrac{(k-1)(k-2)}{\alpha^2} - \tfrac{3}{(1-\alpha^2)^2}\right). \end{aligned} \tag{7.3}$$

Then for $\alpha = \sqrt{\frac{k-2}{k-1}}$

$$\mathcal{B}'''\left(\sqrt{\tfrac{k-2}{k-1}}\right) = \sqrt{2(k-2)}\beta\left((k-1)^2 - 3(k-1)^2\right) = -2\sqrt{2(k-2)}\beta(k-1)^2 < 0, \tag{7.4}$$

which means that a critical $\alpha = \sqrt{\frac{k-2}{k-1}}$ is a saddle point and not a maximizer. $\square$

From Lemma 5.1 and Lemma 7.1 it follows that one can apply Theorem 1.1 (b) for all $f(x) = hx^k$ whenever $\beta < \beta_c(k)$. In the linear and quadratic case one can obtain the following more explicit results.

Corollary 7.2. *Let $h \in \mathbb{R} \setminus \{0\}$ and $f(x) = hx$. Then*

$$L_N - N\sqrt{h^2 + 2\beta^2} - \sqrt{N}\kappa U_N - \left(\kappa\Lambda_N - \frac{1}{2}\begin{pmatrix} U_N \\ U_N' \end{pmatrix}^T G \begin{pmatrix} U_N \\ U_N' \end{pmatrix}\right) \xrightarrow{\mathbb{P}} 0$$

with κ and G given by

$$\kappa = \frac{h^2}{4\beta}, \qquad G_{11} = -\frac{h^4\sqrt{h^2+2\beta^2}}{8\beta^4},$$

$$G_{12} = G_{21} = -\frac{h^6}{32\beta^5}, \qquad G_{22} = \frac{h^{10}}{2^7\beta^6(h^2+2\beta^2)^{\frac{3}{2}}},$$

and the joint convergence

$$U_N \to U \sim \mathcal{N}\left(0, \tfrac{16\beta^4}{h^2(h^2+2\beta^2)}\right), \qquad U_N' \to U' \sim \mathcal{N}\left(0, \tfrac{2^7\beta^6(4\beta^4+5\beta^2h^2+2h^4)}{h^{10}}\right),$$

$$\Lambda_N \to \Lambda \sim \mathcal{N}\left(\tfrac{4\beta^3\sqrt{h^2+2\beta^2}}{h^4}, \tfrac{16\beta^4(h^2+2\beta^2)^2}{h^8}\right),$$

in distribution, where (U, U') and Λ are independent and

$$Cov(U, U') = -\frac{2^6\beta^5(h^2+\beta^2)}{h^6\sqrt{h^2+2\beta^2}}. \tag{7.5}$$

Remark 7.3. *Note that it follows from Corollary 7.2 that*

$$\frac{1}{\sqrt{N}}\left(L_N - \sqrt{h^2+2\beta^2}N\right) \longrightarrow \mathcal{N}\left(0, \frac{\beta^2h^2}{h^2+2\beta^2}\right), \tag{7.6}$$

which coincides with the results from [CS17]. To see this let $\gamma_2 = \beta$, $\gamma_p = 0$ for $p > 2$ and

$$\xi(s) = \beta^2 s^2 \tag{7.7}$$

in [CS17, Theorem 5]. By [CS17, Proposition 1] we then have

$$L_0 = \frac{1}{\sqrt{\xi'(1)+h^2}} = \frac{1}{\sqrt{2\beta^2+h^2}}, \tag{7.8}$$

and by [CS17, Theorem 3] the function $u_t : (0,1) \to \mathbb{R}$ is the solution of

$$L_0^2(t\xi'(u_t)+h^2) = u_t \quad \Leftrightarrow \quad u_t(2\beta^2+h^2) = t2\beta^2 u_t + h^2, \tag{7.9}$$

which is

$$u_t = \frac{h^2}{2\beta^2(1-t)+h^2}. \tag{7.10}$$

This in turn gives us by [CS17, Theorem 5] that

$$\sqrt{N}\left(L_N - \sqrt{2\beta^2+h^2}\right) \longrightarrow \mathcal{N}_N(0,\chi) \tag{7.11}$$

with

$$\chi = \int_0^1 \xi(u_t)dt = \int_0^1 2\beta^2\left(\frac{h^2}{2\beta^2(1-t)+h^2}\right)^2 dt = \frac{\beta^2 h^2}{2\beta^2+h^2}. \tag{7.12}$$

Corollary 7.4. *Let $h \in \mathbb{R}^+$ and $f(x) = hx^2$. If $\beta < \frac{h}{\sqrt{2}}$ then*

$$L_N - \frac{2h^2+\beta^2}{2h} - \sqrt{N}\kappa U_N - \left(\kappa\Lambda_N - \frac{1}{2}\begin{pmatrix} U_N \\ U_N' \end{pmatrix}^T G \begin{pmatrix} U_N \\ U_N' \end{pmatrix}\right) \xrightarrow{\mathbb{P}} 0$$

with constants

$$\tilde{\kappa} = \frac{2h^2-\beta^2}{2\beta}, \qquad G_{11} = -\frac{h(4h^4-2h^2\beta^2+\beta^4)}{\beta^4},$$

$$G_{12} = G_{21} = -\frac{h^2(2h^2-\beta^2)}{2\beta^5}, \qquad G_{22} = -\frac{(2h^2-\beta^2)^4}{16h\beta^6},$$

and the joint convergence in law

$$U_N \to U \sim \mathcal{N}\left(0, \tfrac{2\beta^4}{h^2(2h^2-\beta^2)}\right), \qquad U_N' \to U' \sim \mathcal{N}\left(0, \tfrac{8\beta^6(16h^4-6\beta^2h^2+\beta^4)}{(2h^2-\beta^2)^5}\right),$$

$$\Lambda_N \to \Lambda \sim \mathcal{N}\left(\tfrac{2h\beta^3}{(2h^2-\beta^2)^2}, \tfrac{16h^4\beta^4}{(2h^2-\beta^2)^4}\right),$$

where (U,U') and Λ are independent and

$$Cov(U,U') = -\frac{4\beta^5(4h^2-\beta^2)}{h(2h^2-\beta^2)^3}. \tag{7.13}$$

Note that it follows from Corollary 7.4 that

$$\frac{1}{\sqrt{N}}\left(L_N - \frac{2h^2+\beta^2}{2h}N\right) \longrightarrow \mathcal{N}\left(0, \frac{\beta^2(2h^2-\beta^2)}{2h^2}\right). \tag{7.14}$$

Recall

$$\tilde{\mathcal{B}}(\alpha, r) = f(r\alpha) + \sqrt{2}\beta r^2\sqrt{1-\alpha^2}. \tag{7.15}$$

The next lemma will show that the remaining requirements for Theorem 1.2 (b) are also satisfied for monomial f with $h > h_c(k,\beta)$.

Lemma 7.5. *Let $k \in \mathbb{N}$, $\beta > 0$, $h > h_c(k,\beta)$. Then there is a unique maximizer $(\hat\alpha,\hat r) \in (-1,1) \times \mathrm{Plef}(\beta)^{\mathrm{o}}$ of $\tilde{\mathcal{B}}(\alpha,r) + g(r)$ and*

$$\nabla^2 \tilde{\mathcal{B}}(\hat\alpha,\hat r) \quad \textit{is negative definite.} \tag{7.16}$$

Proof. We know from Lemma 5.3, Lemma 5.5 and Lemma 5.7 that there is a unique $(\hat\alpha,\hat r) \in (0,1) \times \mathrm{Plef}(\beta)^{\mathrm{o}}$. We will show the negative definiteness of $\nabla^2\tilde{\mathcal{B}}(\hat\alpha,\hat r)$ by showing that the determinant is positive while the trace is negative.

Trace: Let us first look at $\nabla^2\tilde{\mathcal{B}}(\hat\alpha,\hat r) = \tilde{\mathcal{B}}''(\hat\alpha,\hat r)$. Since $(\hat\alpha,\hat r)$ is a maximizer it must hold $\tilde{\mathcal{B}}''(\hat\alpha,\hat r) \le 0$. We have

$$\partial_{\alpha\alpha}\tilde{\mathcal{B}}(\alpha,r) = r^2 f''(r\alpha) - \frac{\sqrt{2}\beta r^2}{(1-\alpha^2)^{\frac{3}{2}}},$$

which is negative for $k=1$ for all (α,r), while for $k \ge 2$ the critical point equation implies

$$hk(\hat r\hat\alpha)^{k-2} = \frac{\sqrt{2}\beta}{\sqrt{1-\hat\alpha^2}} \tag{7.17}$$

and thus

$$\begin{aligned}
\partial_{\alpha\alpha}\tilde{\mathcal{B}}(\hat\alpha,\hat r) &= hk(k-1)\hat r^k\hat\alpha^{k-2} - \tfrac{\sqrt{2}\beta\hat r^2}{(1-\hat\alpha^2)^{\frac{3}{2}}}\\
&\overset{(7.17)}{=} (k-1)\hat r^2\tfrac{\sqrt{2}\beta}{\sqrt{1-\hat\alpha^2}} - \tfrac{\sqrt{2}\beta\hat r^2}{(1-\hat\alpha^2)^{\frac{3}{2}}}\\
&= \tfrac{\sqrt{2}\beta\hat r^2}{(1-\hat\alpha^2)^{\frac{3}{2}}}\left((k-1)(1-\hat\alpha^2)-1\right).
\end{aligned}$$

This is obviously negative for $k=2$, while for $k=3$ it can be zero if $\hat\alpha^2 = \frac{k-2}{k-1}$. But if we had $\hat\alpha^2 = \frac{k-2}{k-1}$ we would obtain

$$\begin{aligned}
\tilde{\mathcal{B}}'''(\hat\alpha,\hat r) &= hk(k-1)(k-2)\hat r^k\hat\alpha^{k-3} - \tfrac{3\sqrt{2}\beta\hat r^2\hat\alpha}{(1-\hat\alpha^2)^{\frac{5}{2}}}\\
&\overset{(7.17)}{=} (k-1)(k-2)\hat r^2\hat\alpha^{-1}\tfrac{\sqrt{2}\beta}{\sqrt{1-\hat\alpha^2}} - \tfrac{3\sqrt{2}\beta\hat r^2\hat\alpha}{(1-\hat\alpha^2)^{\frac{5}{2}}}\\
&= \tfrac{\sqrt{2}\beta\hat r^2}{\hat\alpha(1-\hat\alpha^2)^{\frac{5}{2}}}\left((k-1)(k-2)(1-\hat\alpha^2)^2 - 3\hat\alpha^2\right)\\
&= -\tfrac{2(k-2)}{k-1} < 0,
\end{aligned}$$

which would make this a saddle point and not a maximum, so it must hold that $\partial_{\alpha\alpha}\tilde{\mathcal{B}}(\alpha,r) < 0$. Since we also have $\partial_{rr}\tilde{\mathcal{B}}(\alpha,r) \le 0$ the trace is negative.

Determinant: The Hessian of $\tilde{\mathcal{B}}$ is given by

$$\nabla^2\tilde{\mathcal{B}}(\alpha,r) = \begin{pmatrix} hk(k-1)r^k\alpha^{k-2} - \sqrt{2}\beta\frac{r^2}{(1-\alpha^2)^{\frac{3}{2}}} & hk^2(r\alpha)^{k-1} - 2\sqrt{2}\beta\frac{r\alpha}{\sqrt{1-\alpha^2}}\\ hk^2(r\alpha)^{k-1} - 2\sqrt{2}\beta\frac{r\alpha}{\sqrt{1-\alpha^2}} & hk(k-1)r^{k-2}\alpha^k + 2\sqrt{2}\beta\sqrt{1-\alpha^2} + g''(r)\end{pmatrix}.$$

Using (7.17) it follows that if α, r are critical points that

$$\begin{aligned}
\nabla^2\tilde{\mathcal{B}}(\alpha,r) &= \begin{pmatrix} (k-1)r^2\frac{\sqrt{2}\beta}{\sqrt{1-\alpha^2}} - \sqrt{2}\beta\frac{r^2}{(1-\alpha^2)^{\frac{3}{2}}} & k(r\alpha)\frac{\sqrt{2}\beta}{\sqrt{1-\alpha^2}} - 2\sqrt{2}\beta\frac{r\alpha}{\sqrt{1-\alpha^2}}\\ k(r\alpha)\frac{\sqrt{2}\beta}{\sqrt{1-\alpha^2}} - 2\sqrt{2}\beta\frac{r\alpha}{\sqrt{1-\alpha^2}} & (k-1)\alpha^2\frac{\sqrt{2}\beta}{\sqrt{1-\alpha^2}} + 2\sqrt{2}\beta\sqrt{1-\alpha^2} + g''(r)\end{pmatrix},\\
&= \frac{\sqrt{2}\beta}{\sqrt{1-\alpha^2}}\begin{pmatrix} r^2\left((k-1) - \frac{1}{1-\alpha^2}\right) & (k-2)r\alpha\\ (k-2)r\alpha & (k-1)\alpha^2 + 2(1-\alpha^2) + \frac{g''(r)\sqrt{1-\alpha^2}}{\sqrt{2}\beta}\end{pmatrix}.
\end{aligned}$$

For $k=2$

$$\det\nabla^2\tilde{\mathcal{B}}(\alpha,r) = \tfrac{2\beta^2}{1-\alpha^2}\left(\left(\alpha^2 + 2(1-\alpha^2) + \tfrac{g''(r)\sqrt{1-\alpha^2}}{\sqrt{2}\beta}\right)\left(r^2\left(1-\tfrac{1}{1-\alpha^2}\right)\right)\right)$$

Since $\beta < \sqrt{2}h$ and we have $\hat\alpha^2 = 1 - \frac{\beta^2}{2h^2}$ (by Lemma 5.5) it holds that

$$r^2\left(1 - \frac{1}{1-\alpha^2}\right) = r^2\left(1 - \frac{2h^2}{\beta^2}\right) < 0,$$

and since also $h > \frac{1}{2}$ as well as $\hat{r}^2 = 1 - \frac{1}{2h}$

$$\alpha^2 + 2(1-\alpha^2) + \frac{g''(r)\sqrt{1-\alpha^2}}{\sqrt{2}\beta} = \frac{(2h-1)(\beta^2 - 2h^2)}{h^2} < 0.$$

Therefore $\det \nabla^2 \tilde{\mathcal{B}}(\alpha, r) > 0$.
For $k \geq 3$ using (5.33) we can write the Hessian of $\tilde{\mathcal{B}}$ at critical points (α, r) as

$$\nabla^2 \tilde{\mathcal{B}}(\alpha, r) = \frac{1}{1-r^2} \begin{pmatrix} r^2\left(k - 1 - \frac{1}{2\beta^2(1-r^2)^2}\right) & (k-2)r\sqrt{1-2\beta^2(1-r^2)^2} \\ (k-2)r\sqrt{1-2\beta^2(1-r^2)^2} & \frac{(k(1-r^2)-2)(1-2\beta^2(1-r^2)^2)}{1-r^2} \end{pmatrix}, \tag{7.18}$$

where the determinant is given by

$$\begin{aligned} \det\left(\nabla^2 \tilde{\mathcal{B}}(\alpha, r)\right) &= \frac{1}{(1-r^2)^2} \frac{-r^2(1-2\beta^2(1-r^2)^2(-2+k(1-r^2)+2\beta^2(1-r^2)^2(2-4r^2+k(3r^2-1))))}{2\beta^2(1-r^2)^3} \\ &= \underbrace{-\frac{r^2(1-2\beta^2(1-r^2)^2)}{2\beta^2(1-r^2)^5}}_{<0} \zeta_{k,\beta}(r^2), \end{aligned} \tag{7.19}$$

where

$$\zeta_{k,\beta}(q) = -2 + k(1-q) + 2\beta^2(1-q)^2(2 - 4q + k(3q-1)).$$

Recall $T(q)$ from (5.36), which is a non-negative function with $T(q_P) = T(1) = 0$. We showed in Lemma 5.7 that $T(q)$ has exactly one critical point, and that $T(q) = \frac{1}{hk}$ has two solutions $q_1 < q_2$, where $\hat{q}_2 = \hat{r}^2$. Thus we have $T'(q_1) > 0$ and $T'(q_2) = T'(\hat{r}^2) < 0$. Since

$$T'(q) = \underbrace{\frac{\left((1-2b^2(1-q)^2)q\right)^{\frac{k-4}{2}}}{2}}_{>0} \underbrace{\left(-2 + k(1-q) + 2\beta^2(1-q)^2(2-4q+k(3q-1))\right)}_{=\zeta_{k,\beta}(q)},$$

this must mean that $\zeta_{k,\beta}(\hat{r}^2) < 0$ and thus

$$\det\left(\nabla^2 \tilde{\mathcal{B}}(\hat{\alpha}, \hat{r})\right) \overset{(7.19)}{>} 0.$$

For $k = 1$ let us substitute q for r^2, i.e. instead of $\tilde{\mathcal{B}}$ consider

$$\mathscr{B}(\alpha, q) = h\sqrt{q}\alpha + \sqrt{2}\beta q\sqrt{1-\alpha^2} + \frac{\beta^2}{2}(1-q)^2 + \frac{1}{2}\log(1-q),$$

where the Hessian is

$$\nabla^2 \mathscr{B}(\alpha, q) = \begin{pmatrix} -\frac{\sqrt{2}\beta q}{(1-\alpha^2)^{\frac{3}{2}}} & -\frac{\sqrt{2}\beta\alpha}{\sqrt{1-\alpha^2}} + \frac{h}{2\sqrt{q}} \\ -\frac{\sqrt{2}\beta\alpha}{\sqrt{1-\alpha^2}} + \frac{h}{2\sqrt{q}} & \beta^2 - \frac{1}{2(1-q)^2} - \frac{h\alpha}{4q^{\frac{3}{2}}} \end{pmatrix}. \tag{7.20}$$

Since for fixed q the maximizing $\alpha(q)$ is $\frac{h}{\sqrt{h^2+2\beta^2 q}}$ the determinant of $\nabla^2 \mathscr{L}(q, \alpha(q))$ at the maximizer is given by

$$\det \nabla^2 \mathscr{B}(\alpha(q), q) = \frac{2\sqrt{q}(h^2 + 2\beta^2 q)^{\frac{3}{2}}\left(\frac{1}{2(1-q)^2} - \beta^2\right) + \frac{h^4}{2q}}{4q\beta^2}.$$

Since $q \geq 1 - \frac{1}{\sqrt{2}\beta}$ we have that $\det \nabla^2 \mathscr{B}(q, \alpha(q)) > 0$, and therefore $\det \nabla^2 \tilde{\mathcal{B}}(\hat{\alpha}, \hat{r}) > 0$. □

Lemmas 5.3 - 5.7 together with Lemma 7.5 show that we can apply Theorem 1.2 (b) for monomials $f(x) = hx^k$ and $\beta > 0$ whenever $h > h_c(k, \beta)$.

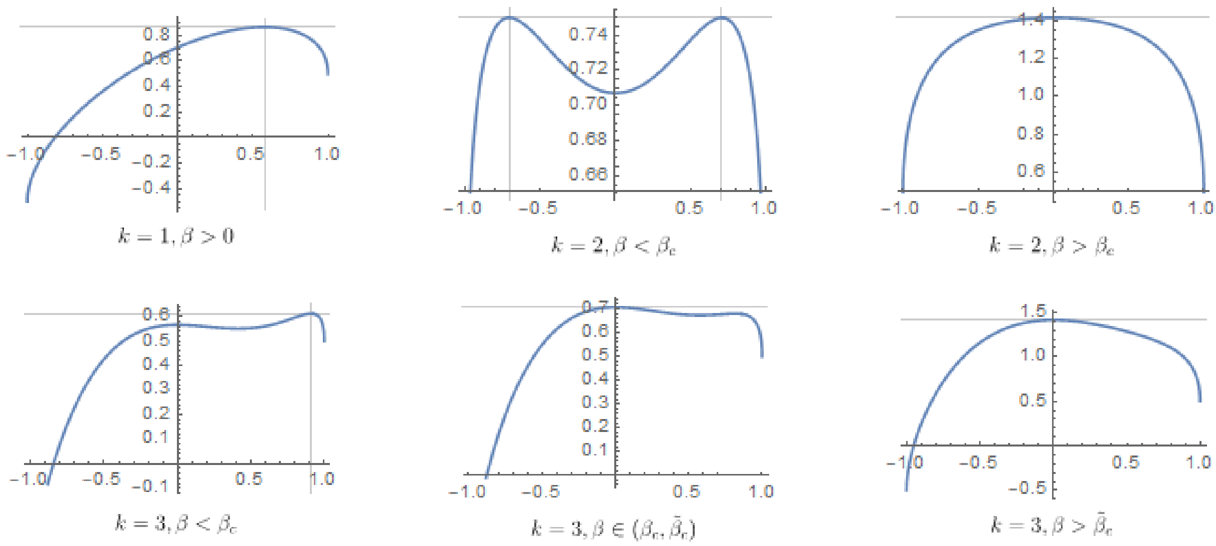

Figure III.1: Plot of $\mathcal{B}(\alpha)$ for $\alpha \in [-1, 1]$

CHAPTER IV

TAP variational principle for the constrained multiple spherical SK model

TAP variational principle for the constrained overlap multiple spherical Sherrington-Kirkpatrick model

David Belius, Leon Fröber, Justin Ko

Abstract. Spin glass models involving multiple replicas with constrained overlaps have been studied in [FPV92, PT07, Pan18a]. For the spherical versions of these models [Ko19, Ko20] showed that the limiting free energy is given by a Parisi type minimization. In this work we show that for Sherrington-Kirkpatrick (i.e. 2-spin) interactions, it can also be expressed in terms of a Thouless-Andersson-Palmer (TAP) variational principle. This is only the second model where a mathematically rigorous TAP computation of the free energy at all temperatures and external fields has been achieved. The variational formula we derive here also confirms that the model is replica symmetric, a fact which is natural but not obviously deducible from the Parisi formula for the model.

1. Introduction

We study the free energy of the constrained multiple replica spin glass model of [FPV92, PT07, Pan18a], also called the vector spin model. In physics this free energy is known as the Franz–Parisi potential [FPV92]. The model involves multiple replicas with constrained overlaps and was originally introduced to study metastable states of standard one replica spin glasses [FPV92], and has since been used to study several other of properties of one replica models [Pan16, CP17, BAJ18, Jag19, AJ21, AK18, FR20, JLM20].

We introduce a new approach to studying the model by adapting the Thouless-Andersson-Palmer (TAP) approach of [BK19] to the model's spherical Sherrington-Kirkpatrick (SK; i.e. 2-spin) version. We prove a variational formula for its free energy in terms of a TAP free energy, and compute a formula for the maximal TAP free energy, thus yielding a concrete formula for the original free energy. After [BK19] this represents only the second setting where the free energy of a spin glass model has been computed at all temperatures and external fields using a mathematically rigorous TAP approach.

We now formally introduce the model. The 2-spin SK Hamiltonian is a Gaussian process of the form

$$H_N(\sigma) = \sqrt{N} \sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j, \tag{1.1}$$

where J_{ij} are i.i.d. standard Gaussian random variables, which is indexed by $\sigma \in \mathbb{R}^N$. For $n \geq 1$, we consider the multiple spin configuration of n replicas denoted by the matrix

$$\boldsymbol{\sigma} = (\sigma^1, \dots, \sigma^n) \in \mathbb{R}^{n\times N},$$

where each $\sigma^k \in \mathbb{R}^N$ denotes the k-th row of $\boldsymbol{\sigma}$ and σ_i^k the entry in the i-th column and k-th row. Let

$$\mathcal{S}_{N-1} = \{\sigma \in \mathbb{R}^N : \sigma_1^2 + ... + \sigma_N^2 = 1\}$$

denote the unit sphere in $\mathbb{R}^N$. Let $\boldsymbol{h} = (h^1, \dots, h^n) \in \mathbb{R}^{n\times N}$ and $\beta = (\beta_1, \dots, \beta_n)$ denote the external fields and inverse temperatures of each replica. Furthermore, assume that $|h^k| = h_k \in \mathbb{R}$ and let $h = (h_1, \dots, h_n) \in \mathbb{R}^n$. Let $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ be a positive semi-definite matrix with 1's along the diagonal giving a constraint on the overlaps of the replicas. For a matrix $\boldsymbol{A}$ let $\|\boldsymbol{A}\|_\infty$ denote the sup-norm $\max_{k,l}|A_{k,l}|$, and for $\varepsilon > 0$ let

$$\boldsymbol{Q}_\varepsilon = \{\boldsymbol{\sigma} : \|\boldsymbol{\sigma}\boldsymbol{\sigma}^\mathsf{T} - \boldsymbol{Q}\|_\infty \leq \varepsilon\}, \tag{1.2}$$

denote the set of replicas with overlaps close to $\boldsymbol{Q}$.

Our goal is to compute the limit of the replica constrained free energy

$$F_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) = \frac{1}{N}\log\int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k)+Nh^k\cdot\sigma^k}\,d\boldsymbol{\sigma} \tag{1.3}$$

for fixed model parameters $\boldsymbol{Q},\boldsymbol{h},\beta$, where $d\boldsymbol{\sigma} = (d\sigma)^{\otimes n}$ is the product of uniform measures $d\sigma$ on the sphere $\mathcal{S}_{N-1}$. Note that the integral can not be trivially reduced to a one replica integral using Fubini's theorem because the replica overlaps are constrained to the corresponding values of $\boldsymbol{Q}$. Note further that each replica shares the same disorder J_{ij} but can be subject to different inverse temperatures β_k and external fields h^k.

The TAP free energy we derive for this model is given by

$$F_{\mathrm{TAP}}(\boldsymbol{m}) = \frac{N}{2}\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T}| + \sum_{k=1}^n \beta_k H_N(m^k) + N\sum_{k=1}^n h^k\cdot m^k + \frac{N}{2}\beta^\mathsf{T}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T})^{\odot 2}\beta, \tag{1.4}$$

where $\boldsymbol{m} = (m^1,\ldots,m^n)\in\mathbb{R}^{n\times N}$ are magnetization vectors, $|\cdot|$ denotes the determinant, and $\boldsymbol{A}^{\odot 2} = \boldsymbol{A}\odot\boldsymbol{A} = (A_{k,l}^2)_{k,l=1,\ldots,n}$ denotes the Hadamard square of the entries of $\boldsymbol{A}$. We further introduce a Plefka condition [TAP77, Ple82b] for the vector spin model given by $\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta)$ for

$$\begin{aligned}\mathrm{Plef}_n(\boldsymbol{Q},\beta) &= \left\{\tilde{\boldsymbol{Q}}\in[-1,1]^{n\times n} : \mathbf{0}\le\tilde{\boldsymbol{Q}}<\boldsymbol{Q},\ \|\boldsymbol{\beta}^{\frac12}(\boldsymbol{Q}-\tilde{\boldsymbol{Q}})\boldsymbol{\beta}^{\frac12}\|_2\le\tfrac{1}{\sqrt2}\right\}\\ \mathrm{Plef}_N(\boldsymbol{Q},\beta) &= \left\{\boldsymbol{m}\in\mathbb{R}^{n\times N} : \boldsymbol{m}\boldsymbol{m}^\mathsf{T}\in\mathrm{Plef}_n(\boldsymbol{Q},\beta)\right\}\end{aligned} \tag{1.5}$$

where $\boldsymbol{\beta} = \mathrm{diag}(\beta)\in\mathbb{R}^{n\times n}$, $\|\cdot\|_2$ denotes the spectral norm (largest eigenvalue for symmetric positive semi-definite matrices) and $\le$ is the Loewner partial order on matrices (so that $\boldsymbol{A}\ge 0$ for $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ means that $\boldsymbol{A}$ is positive semi-definite).

Our main theorem is a TAP variational principle giving the limiting free energy of the model as a supremum over $\boldsymbol{m}\in\mathrm{Plef}_n(\boldsymbol{Q},\beta)$.

Theorem 1.1 (TAP Variational Principle). *Let $n\ge 1$ and $\boldsymbol{Q}\in[-1,1]^{n\times n}$ be positive definite with $Q_{k,k}=1$ for $k=1,\ldots,n$. It holds that*

$$\lim_{\varepsilon\to0}\limsup_{N\to\infty}\left|F_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) - \sup_{\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta)}\frac{1}{N}F_{\mathrm{TAP}}(\boldsymbol{m})\right| = 0, \tag{1.6}$$

where the limits are in probability.

If $\boldsymbol{Q}$ is not positive definite then $\lim_{\varepsilon\to0}\lim_{N\to\infty}F_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) = -\infty$ (see (3.12)).

We also compute the supremum in (1.6) when $h^1,\ldots,h^n$ are multiples of a single vector. To this end we let for any β, $h\in\mathbb{R}^n$ and positive definite $n\times n$ constraint matrix $\tilde{\boldsymbol{Q}}$

$$\mathrm{GSE}(\beta,h,\tilde{\boldsymbol{Q}}) = \sqrt{2}\mathrm{Tr}\left(\sqrt{\left(\frac12 hh^\mathsf{T}+\boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\right)^{\frac12}\tilde{\boldsymbol{Q}}\left(\frac12 hh^\mathsf{T}+\boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\right)^{\frac12}}\right). \tag{1.7}$$

Note that the trace on the right-hand side is the sum of the singular values of $(\frac12 hh^\mathsf{T}+\boldsymbol{\beta}\tilde{\boldsymbol{Q}})^{1/2}\tilde{\boldsymbol{Q}}^{1/2}$. The ground state of the energy over magnetizations $\boldsymbol{m}$ with constrained overlaps converges to this limit:

Theorem 1.2 (Ground state energy). *Assume that $h^i = h_i u$ for a sequence of unit vectors $u\in\mathbb{R}^N$ for $i=1,\ldots,n$, For all β,h and positive definite $\tilde{\boldsymbol{Q}}$ we have that*

$$\lim_{N\to\infty}\sup_{\boldsymbol{m}\boldsymbol{m}^\mathsf{T}=\tilde{\boldsymbol{Q}}}\left(\sum_{k=1}^n\frac{1}{N}\beta_k H_N(m^k) + \sum_{k=1}^n m^k\cdot h^k\right) = \mathrm{GSE}(\beta,h,\tilde{\boldsymbol{Q}}), \tag{1.8}$$

where the limit is in probability.

The next theorem expresses the limiting maximum TAP free energy as a lower dimensional optimizion, namely as one of $n\times n$ (so bounded in N) rather than $n\times N$ dimensions. It follows immediately from (1.4) and Theorem 1.2.

Corollary 1.3 (Low Dimensional Variational Principle)**.** *Assume that $h^i = h_i u$ for a sequence of unit vectors $u \in \mathbb{R}^N$ for $i = 1, \dots, n$. For all β, h and positive definite $\boldsymbol{Q}$ it holds that*

$$\lim_{N\to\infty} \sup_{\boldsymbol{m}\in \mathrm{Plef}_N(\boldsymbol{Q},\beta)} \frac{1}{N} F_{\mathrm{TAP}}(\boldsymbol{m}) = \sup_{\tilde{\boldsymbol{Q}}\in \mathrm{Plef}_n(\boldsymbol{Q},\beta)} \left(\mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}}) + \frac{1}{2} \log |\boldsymbol{Q} - \tilde{\boldsymbol{Q}}| + \frac{1}{2} \beta^{\mathsf{T}} (\boldsymbol{Q} - \tilde{\boldsymbol{Q}})^{\odot 2} \beta \right), \tag{1.9}$$

where the limit is in probability.

It follows immediately from Theorems 1.1 and Corollary 1.3 that also the limiting free energy is given by the same low dimensional optimization problem.

Corollary 1.4. *Assume that $h^i = h_i u$ for a sequence of unit vectors $u \in \mathbb{R}^N$ for $i = 1, \dots, n$. For all β, h and positive definite $\boldsymbol{Q}$ the limit of the free energy is*

$$\lim_{\varepsilon\to 0} \lim_{N\to\infty} F_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) = \sup_{\tilde{\boldsymbol{Q}}\in \mathrm{Plef}_n(\boldsymbol{Q},\beta)} \left(\mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}}) + \frac{1}{2} \log |\boldsymbol{Q} - \tilde{\boldsymbol{Q}}| + \frac{1}{2} \beta^{\mathsf{T}} (\boldsymbol{Q} - \tilde{\boldsymbol{Q}})^{\odot 2} \beta \right), \tag{1.10}$$

where the limits are in probability.

Remark 1.5. *When $n = 1$ we recover the results of [BK19]. Indeed the only valid constraint is $\boldsymbol{Q} = 1$, and with this constraint F_{TAP} coincides with H_{TAP} of [BK19], and Theorem 1.1 coincides with [BK19, Theorem 1]. The functional (1.7) is*

$$\mathrm{GSE}(\beta, h, \tilde{q}) = \sqrt{2\beta^2 \tilde{q}^2 + h^2 \tilde{q}}, \tag{1.11}$$

cf. [BK19, (1.6) and Lemma 20]. Corollary 1.3 says

$$\lim_{N\to\infty} F_N^{\varepsilon}(\beta, h, 1) = \sup_{m:\beta(1-\tilde{q})\le \frac{1}{\sqrt{2}}} \left(\sqrt{2\beta^2 \tilde{q}^2 + h^2 \tilde{q}} + \frac{\beta^2}{2} (1 - \tilde{q})^2 + \frac{1}{2} \log |1 - \tilde{q}| \right),$$

for all $\varepsilon > 0$, cf. [BK19, Lemma 2].

1.1. Discussion

The most important result about one replica ($n = 1$) spin glass models [SK75][1] is the Parisi formula [Par80, Par79, MPV87] for the limiting free energy which has been proved rigorously using the methods of Guerra, Aizenman–Sims–Starr, Talagrand and Panchenko [Gue03, ASS03, Che13, Tal06b, Tal06a, Pan13a, Pan14]. The TAP approach is an attractive proposal [TAP77] of an alternative framework to compute the free energy which is under active investigation, with at least three projects underway to implement it mathematically rigorously ([Bol14, Bol19, BY22], [Sub17b, Sub18, CPS22, Sub21], [BK19, Bel22]).

Concerning constrained multiple spin glass models ($n \geq 1$; [FPV92, PT07, Pan18a]) an upper bound for the free energy of spherical models was proved in [PT07] using the Guerra interpolation scheme. The matching lower bound for this model was proved in [Ko20, Ko19] by adapting the synchronization property derived for constrained multiple spin models with respect to product measures by Panchenko in [Pan18a, Pan18b] and the Aizenman–Sims–Starr scheme. In this article, we investigate the 2-spin constrained multiple spherical spin model using the TAP approach of [BK19] (see also [Bel22]) and derive the new variational expression (1.10) for the limiting free energy. The variational formula is expressed as the maximum of a functional defined on $n \times n$ matrices. It is much simpler than the 2-spin version of the Parisi variational formula from [Ko19] defined in terms of matrix paths [Ko19, Theorem 1 and Theorem 3]. After [BK19] our results represents only the second setting where the free energy of a spin glass model has been computed at all temperatures and external fields using a mathematically rigorous TAP approach ([Sub21] uses a different version of the TAP approach to compute the free energy for pure p-spin spherical spin glasses without external field at all temperatures). We

[1] See [KTJ76, Der80, GM84b, CS92, Tal00, CL04, Tal06a] for the various generalizations of the original Ising type 2-spin SK model.

hope that in the future a further improvement of the present TAP approach can be extended to a wider class of spin glass models.

A well-known property of the classical ($n = 1$) spherical 2-spin model is that it is replica symmetric at any inverse temperature and external field, as can be verified by studying the Parisi formula for the model [Tal06a, Section 2]. For the constrained multiple spin model, [AZ22] gives a zero temperature Parisi formula for the ground state and shows that the minimizer is replica symmetric in the case of 2-spin interaction [AZ22, Proposition 7]. A similar computation at positive temperature seems infeasible, so presently one can not deduce that the free energy of the constrained multiple 2-spin model is replica symmetric from its Parisi formula. Since we use the TAP approach we do not directly study the Parisi formula for the model, instead obtaining the different formula (1.10). However the formula (1.10) expresses that the free energy is replica symmetric, since the maximization is over only one matrix $\tilde{\boldsymbol{Q}}$.

1.2. Outline of proof

The starting point of the proof is the computation of the free energy at high temperature in the absence of external field. When $n = 1$ (with the unique possibility $\boldsymbol{Q} = 1$ as the constraint) the annealed free energy is $\frac{1}{2}\beta^2$, and this is also the quenched free energy if the Hamiltonian is at high temperature, which is the case if $\beta \leq \frac{1}{\sqrt{2}}$. When $n \geq 2$ with a constraint $\boldsymbol{Q}$ the annealed free energy turns out to be $\frac{1}{2}\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta$ (after subtracting the normalizing factor $\frac{1}{2}\log|\boldsymbol{Q}|$ corresponding to log-scale volume of spin vectors that satisfy the constraint; see Lemmas 3.2, 3.3). Similarly this is also the quenched free energy if the Hamiltonian is at high temperature, which turns out to be the case if $\|\boldsymbol{\beta}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{\beta}^{\frac{1}{2}}\|_2 \leq \frac{1}{\sqrt{2}}$. As is well-known, these properties of the model with $n = 1$ can be verified using a second moment method [T+03, Section 2.2]. In this paper we find that a second moment computation also gives the aforementioned properties of the model with $n \geq 2$, though the second moment computation is more challenging (see Lemmas 3.3, 3.4 and Propositions 3.6, 3.10). As an aside, note that the aforementioned claim about the quenched free energy is the special case $h = 0$ and $\|\boldsymbol{\beta}^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{\beta}^{\frac{1}{2}}\|_2 \leq \frac{1}{\sqrt{2}}$ of (1.10), in which it can be seen that the maximizer is $\tilde{\boldsymbol{Q}} = 0$.

Armed with this knowledge of the high temperature phase, the proof of Theorem 1.1 splits into a lower and an upper bound for $F_N^\varepsilon = F_N^\varepsilon(\beta, \boldsymbol{h}, \boldsymbol{Q})$, both of which proceed by estimating the partition function integral restricted to certain subsets of $\mathcal{S}_{N-1}^n$ that are neighborhoods of a magnetization vector $\boldsymbol{m}$. That is, for each such subset $A(\boldsymbol{m}) \subset \mathcal{S}_{N-1}^n$ we estimate $\int_{A(\boldsymbol{m})\cap\boldsymbol{Q}_\varepsilon} e^{f(\boldsymbol{m})} d\boldsymbol{\sigma}$ where $f(\boldsymbol{\sigma}) = \sum_{k=1}^n \left(\beta_k H_N\left(\sigma^k\right) + Nh^k \cdot \sigma^k\right)$. We normalize the integral, subtract the centering term $f(\boldsymbol{m})$ and take the log to obtain

$$\text{(I)} \qquad \overbrace{\log \int_{A(\boldsymbol{m})\cap \boldsymbol{Q}_\varepsilon} e^{f(\sigma)} d\boldsymbol{\sigma} = \log \int_{A(\boldsymbol{m})\cap \boldsymbol{Q}_\varepsilon} 1 d\boldsymbol{\sigma} \quad + \quad f(\boldsymbol{m}) \quad + \quad \log \frac{\int_{A(\boldsymbol{m})\cap \boldsymbol{Q}_\varepsilon} e^{f(\boldsymbol{\sigma})-f(\boldsymbol{m})} d\sigma}{\int_{A(\boldsymbol{m})\cap \boldsymbol{Q}_\varepsilon} 1 d\boldsymbol{\sigma}}}$$

These terms each give rise to one of the terms of $F_{\text{TAP}}(\boldsymbol{m})$, through the approximations

$$\underbrace{\overset{\wr\wr}{\frac{N}{2} \log |\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^T|} \quad + \quad \overset{||}{\sum_{k=1}^n \left(\beta_k H_N\left(m^k\right) + N h^k \cdot m^k\right)} \quad + \quad \overset{\wr\wr}{\frac{N}{2} \beta^\top \left(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^T\right)^{\odot 2} \beta}} \qquad \text{(II)} \quad = F_{\text{TAP}}(\boldsymbol{m})$$

Furthermore each term has the natural interpretation

Entropy + Local mean energy + Local free energy (Onsager term)

as we now explain.

Indeed the first term in (I) is precisely the log-volume of $\boldsymbol{\sigma}$ that lie in $A(\boldsymbol{m})$ and satisfy the constraint given by $\boldsymbol{Q}$, and is thus an entropy. The neighborhood $A(\boldsymbol{m})$ is chosen essentially as a subset of the "slice" passing through $\boldsymbol{m}$, i.e. the hyperplane with normal $\boldsymbol{m}$ passing through $\boldsymbol{m}$ intersected with $\mathcal{S}_{N-1}^n$. Such a slice turns out to have log-volume approximately given by $\frac{N}{2}\log|\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\top|$ (i.e. by the first term in (II)), and the subset we choose retains enough of the volume of the slice to have approximately the same log-volume.

The centering term $f(\boldsymbol{m}) = \sum_{k=1}^n \left(\beta H_N\left(m^k\right) + N h^k \cdot m^k\right)$ of (I), (II) represents the "local" mean energy on $A(\boldsymbol{m})$.

For the last term of (I) we use the knowledge of the high temperature phase of the first paragraph of this subsection. The identity

$$H_N(\sigma) = H_N(m) + \nabla H_N(m)\cdot(\sigma - m) + H_N(\sigma - m)$$

valid for all $m, \sigma \in \mathbb{R}^N$ implies that

$$f(\boldsymbol{\sigma}) - f(\boldsymbol{m}) = \sum_{k=1}^N \underbrace{\left(\beta_k \nabla H_N(m^k) + h^k\right)}_{\text{effective external fields}} \cdot (\sigma^k - m^k) + \sum_{k=1}^N \underbrace{\beta_k H_N(\sigma^k - m^k)}_{\text{effective Hamiltonian}}.$$

From this one sees that the last term in (I) can be interpreted as the free energy of an effective Hamiltonian on the spin space $A(\boldsymbol{m}) \cap \boldsymbol{Q}_\varepsilon$ subject to effective external fields. In the proof we construct the sets $A(\boldsymbol{m})$ so that the effective external field term vanishes for $\sigma \in A(\boldsymbol{m})$ (in the easiest case, simply by intersecting the slice with a hyperplane with normal given by the effective external field). Furthermore after normalizing $\sigma^k - m^k$ it turns out that the recentered Hamiltonian is essentially the original Hamiltonian with an effective constraint $\hat{Q}(\boldsymbol{m})_{ij} = (\boldsymbol{Q} - (\boldsymbol{m}\boldsymbol{m}^\top))_{ij}/(\sqrt{1-|m^i|^2}\sqrt{1-|m^j|^2})$ subject to an effective temperature $\boldsymbol{\beta_m} = (\beta_1(1-|m^1|^2), ..., \beta_n(1-|m^n|^2))$. Therefore applying the approximations for the high temperature free energy in the first paragraph of the subsection one obtains that if $\|\boldsymbol{\beta_m}\hat{\boldsymbol{Q}}(\boldsymbol{m})\boldsymbol{\beta_m}\|_2 \le \frac{1}{\sqrt{2}}$ then the third term of (I) can be approximated by $\frac{N}{2}\beta_{\boldsymbol{m}}^\top \hat{\boldsymbol{Q}}(\boldsymbol{m})^{\odot 2}\beta_{\boldsymbol{m}}$. Since $\|\boldsymbol{\beta_m}\hat{\boldsymbol{Q}}(\boldsymbol{m})\boldsymbol{\beta_m}\|_2 = \|\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q} - \tilde{\boldsymbol{Q}})\boldsymbol{\beta}^{\frac{1}{2}}\|_2$ the former condition is precisely Plefka's condition, and since $\frac{N}{2}\beta_{\boldsymbol{m}}^\top \hat{\boldsymbol{Q}}(\boldsymbol{m})^{\odot 2}\beta_{\boldsymbol{m}} = \frac{N}{2}\beta^\top\left(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^T\right)^{\odot 2}\beta$ the latter is precisely the approximation of the last term of (I) by the Onsager term in (II).

This justifies the approximation $\log \int_{A(\boldsymbol{m})\cap Q_\varepsilon} e^{f(\boldsymbol{\sigma})} d\boldsymbol{\sigma} \approx F_{\text{TAP}}(\boldsymbol{m})$ provided Plefka's condition holds for $\boldsymbol{m}$.

Finally, it turns out that only $\boldsymbol{m}$ satisfying Plefka's condition are relevant. Indeed, in Section 3 we prove the lower bound for F_N^ε by simply only considering $\boldsymbol{m}$ that satisfy Plefka's condition, and deduce that F_N^ε is lower bounded by $F_{\mathrm{TAP}}(\boldsymbol{m})$ for any $\boldsymbol{m}$ that satisfies the condition.

The central difficulty in proving the upper bound for F_N^ε in Section 4 is that we cannot a priori ignore $\boldsymbol{m}$ that do not satisfy Plefka's condition. Instead we approximate the recentered Hamiltonian by one that is in some sense always at high temperature, even when Plefka's condition is not satisfied. This gives rise to an upper bound of F_N^ε in terms of a modified TAP free energy which has a different Onsager term. We then show that any maximizer of this modifed TAP free energy in fact must satisfy Plefka's condition, and that in this case its Onsager term is close to the usual Onsager term.

The above constitutes a multidimensional ($n \geq 2$) adaption of the method (for $n = 1$) in [BK19] (elements of the above ideas are also used by TAP [TAP77], Bolthausen [Bol14, Bol19] and Subag and collaborators [Sub17b, Sub18, CPS22, Sub21]).

Lastly in Section 5, we express the ground state of the Hamiltonian as a finite dimensional variational problem over positive semi-definite matrices using the method of Lagrange multipliers. The resulting variational problem can be solved explicitly yielding the closed form representation in Theorem 1.2.

2. Preliminaries

We denote constants, whose value may change from line to line or even in the same expression, by c. They may depend on the number of replicas n, but are independent of all other parameters unless otherwise stated.

At certain points in the proof we will use the standard fact that

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{m\in\mathbb{R}^N:|m|\leq 1} |H_N(m)| \leq cN, \sup_{m\in\mathbb{R}^N:|m|\leq 1} |\nabla H_N(m)| \leq cN \right) = 1. \tag{2.1}$$

This follows for instance by writing $\boldsymbol{J} = (J_{i,j})_{i,j=1,\dots,N}$ so that $\frac{\boldsymbol{J}+\boldsymbol{J}^\intercal}{2}$ is a GOE random matrix we have $H_N(m) = m^\intercal \frac{\boldsymbol{J}+\boldsymbol{J}^\intercal}{2} m$, and noting that $\nabla H_N(m) = (\boldsymbol{J}+\boldsymbol{J}^\intercal)m$ and

$$\lim_{N\to\infty} \mathbb{P}\left(\|\boldsymbol{J}+\boldsymbol{J}^\intercal\|_2 \leq cN \right) = 1. \tag{2.2}$$

We will also use that writing $N\lambda_1 < \ldots < N\lambda_N$ for the eigenvalues of $\frac{\boldsymbol{J}+\boldsymbol{J}^\intercal}{2}$ we have

$$\max_{i=1,\dots,N-1} |\lambda_{i+1} - \lambda_i| \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty. \tag{2.3}$$

Finally for the upper bound we will use that

$$\max_{i=1,\dots,N} |\lambda_i - \theta_{i/N}| \xrightarrow{\mathbb{P}} 0 \text{ as } N \to \infty, \tag{2.4}$$

(see [EYY12, Theorem 2.2]) where $\theta_{i/N}$ are the classical locations

$$\theta_{i/N} = \inf\left\{ \theta : \int_{-\sqrt{2}}^{\theta} d\mu_{\mathrm{sc}}(x) = \frac{i}{N} \right\}, \tag{2.5}$$

defined in terms of the semi-circle distribution

$$d\mu_{\mathrm{sc}}(x) = \frac{1}{\pi}\sqrt{2-x^2}\,1_{[-\sqrt{2},\sqrt{2}]}(x)dx. \tag{2.6}$$

It follows from (2.5) that

$$\lim_{N\to\infty} \lim_{\varepsilon\to 0} \sup_{|i-j|\leq \varepsilon N} \left|\theta_{i/N} - \theta_{j/N}\right| = 0. \tag{2.7}$$

Note that (2.2) and (2.3) are consequences of (2.4)-(2.7).

3. Lower bound

In this section, we will prove the following lower bound of the free energy.

Proposition 3.1 (TAP lower bound)**.** *Let $n \geq 1$ and $\boldsymbol{Q} \in [-1,1]^{n\times n}$ be positive definite with $Q_{k,k} = 1$ for $k = 1,\ldots,n$. Let $h_1,\ldots,h_n \in [0,\infty)$ and $h^1,\ldots,h^n$ a sequence of vectors with $h^k \in \mathbb{R}^N$ and $|h^k| = h_k$. Then there exists a $c = c(n,h_1,\ldots,h_n,\boldsymbol{Q}) > 0$ such that for all $\varepsilon > (0,c^{-1})$*

$$\lim_{N\to\infty} \mathbb{P}\left(F_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) \geq \frac{1}{N}\sup_{\boldsymbol{m}\in\mathrm{Plef}_N(\boldsymbol{Q},\beta)} F_{\mathrm{TAP}}(\boldsymbol{m}) - c\sqrt{\varepsilon}\right) = 1. \tag{3.1}$$

To prove this we first compute the free energy at high temperature in the absence of external field using the second moment method in Subsection 3.1. Then in Subsection 3.2 we consider the model with external field at arbitrary temperature, and as described in Subsection 1.2 proceed by fixing a $\boldsymbol{m}$ that satisfies Plefka's condition, constructing a set $A(\boldsymbol{m})$ (see (3.51)) that is "centered around" $\boldsymbol{m}$, recentering the Hamiltonian around this $\boldsymbol{m}$ (see (3.48)) and estimating the free energy of the recentered Hamilontian on the set $A(\boldsymbol{m})$ (see (3.66)).

3.1. Free energy without external field

Let us define

$$Z_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) = \int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k) + N h^k \sigma^k}\, d\boldsymbol{\sigma} \overset{(1.3)}{=} e^{N F_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q})}. \tag{3.2}$$

The goal of this subsection is to use the second moment method on $Z_N^\varepsilon(\beta,0,\boldsymbol{Q})$ to show that it concentrates as $N\to\infty$. In the first lemma of this section we show that the volume of the $\boldsymbol{Q}$-constrained n-fold product of spheres is approximately $\frac{1}{2}\log|\boldsymbol{Q}|$ at exponential scale, with which we can calculate the moments of $Z_N^\varepsilon(\beta,0,\boldsymbol{Q})$.

Recall that $\|\boldsymbol{A}\|_2$ denotes the spectral norm of $\boldsymbol{A}$.

Lemma 3.2 (Constrained volume)**.** *Let $n \geq 1$. There is a constant $c > 0$ such that for for all symmetric positive semi-definite $\boldsymbol{Q} \in [-1,1]^{n\times n}$ with 1's on the diagonal and all $\varepsilon \in (0,c^{-1})$ and $N \geq c\varepsilon^{-1}$ we have*

$$\left|\frac{1}{N}\log\int 1_{\{\|\boldsymbol{\sigma}\boldsymbol{\sigma}^\top - \boldsymbol{Q}\|_\infty \leq \varepsilon\}}\, d\boldsymbol{\sigma} - \frac{1}{2}\log|\boldsymbol{Q}|\right| \leq c(1+\|\boldsymbol{Q}^{-1}\|_2)\varepsilon, \tag{3.3}$$

and

$$\frac{1}{N}\log\int 1_{\{\|\boldsymbol{\sigma}\boldsymbol{\sigma}^\top - \boldsymbol{Q}\|_\infty \leq \varepsilon\}}\, d\boldsymbol{\sigma} \leq \frac{1}{2}\log|\varepsilon\boldsymbol{I} + \boldsymbol{Q}| + c. \tag{3.4}$$

Proof. Let $u_{i,k}$, $k = 1,\ldots,n$, $i = 1,\ldots,N$ be i.i.d. standard normal random variables, and let $u^k = (u_{1,k},\ldots,u_{n,k}) \in \mathbb{R}^N$ and $u_i = (u_{i,1},\ldots,u_{i,n}) \in \mathbb{R}^n$. Note that the conditional law of $u^k/|u^k|$ on the event $|\frac{|u^k|}{\sqrt{N}} - 1| < \varepsilon$ is the same as the law of σ_k for $k = 1,\ldots,n$, so

$$\begin{aligned}
&\frac{1}{N}\log\int 1_{\{\|\boldsymbol{\sigma}\boldsymbol{\sigma}^\top - \boldsymbol{Q}\|_\infty \leq \varepsilon\}}\, d\boldsymbol{\sigma} \\
&= \frac{1}{N}\log\mathbb{P}\left(\max_{k,l}\left|\frac{u^k\cdot u^l}{|u^k|\,|u^l|} - Q_{k,l}\right| \leq \varepsilon \,\middle|\, \sup_k\left|\frac{|u^k|}{\sqrt{N}} - 1\right| < \varepsilon\right) \\
&= \frac{1}{N}\log\mathbb{P}\left(\max_{k,l}\left|\frac{u^k\cdot u^l}{|u^k|\,|u^l|} - Q_{k,l}\right| \leq \varepsilon, \sup_k\left|\frac{|u^k|}{\sqrt{N}} - 1\right| < \varepsilon\right) - \frac{1}{N}\log\mathbb{P}\left(\sup_k\left|\frac{|u^k|}{\sqrt{N}} - 1\right| < \varepsilon\right).
\end{aligned}$$

By the Chebyshev inequality $\mathbb{P}\left(\left|\frac{|u^k|}{\sqrt{N}} - 1\right| > \varepsilon\right) \leq \frac{c}{\varepsilon^2 N^2}$, which implies the last term is bounded by $c\varepsilon$ if c is large enough and $N \geq c\varepsilon^{-1}$, so it suffices to control the probability of the event

$$A = \left\{\max_{k,l}\left|\frac{u^k\cdot u^l}{|u^k|\,|u^l|} - Q_{k,l}\right| \leq \varepsilon, \sup_k\left|\frac{|u^k|}{\sqrt{N}} - 1\right| < \varepsilon\right\}.$$

We apply the standard proof of Cramér's theorem to the i.i.d. vectors $u_1,\ldots,u_N$, taking care to obtain a bound that is uniform in $\boldsymbol{Q}$. There exists constants c_1, c_2 such that for all ε smaller than some constant and all $\boldsymbol{Q}$ with $\|\boldsymbol{Q}\|_\infty \le 1$

$$\left\{\max_{k,l} \left|u^k \cdot u^l - NQ_{k,l}\right| \le c_1\varepsilon N\right\} \subseteq A \subseteq \left\{\max_{k,l} \left|u^k \cdot u^l - NQ_{k,l}\right| \le c_2\varepsilon N\right\}.$$

We begin with the upper bound. For all symmetric $n \times n$ matrices $\boldsymbol{\Lambda}$

$$\mathbb{P}\left(\max_{k,l} \left|u^k \cdot u^l - NQ_{k,l}\right| \le c_2\varepsilon N\right) \le \mathbb{E}\left[\exp\left(\sum_{k,l=1}^n \Lambda_{k,l} u^k \cdot u^l\right)\right] e^{-N\sum_{k,l=1}^n \Lambda_{k,l}Q_{k,l} + c_2\varepsilon n^2 \|\boldsymbol{\Lambda}\|_\infty N}.$$

If $2\boldsymbol{\Lambda} < \boldsymbol{I}$, it holds that

$$\mathbb{E}\left[\exp\left(\sum_{k,l=1}^n \Lambda_{k,l} u^k \cdot u^l\right)\right] = \mathbb{E}\left[\exp\left(\sum_{i=1}^N (u_i)^\top \boldsymbol{\Lambda} u_i\right)\right] = \left(\mathbb{E}\left[\exp\left((u_1)^\top \boldsymbol{\Lambda} u_1\right)\right]\right)^N = |\boldsymbol{I} - 2\boldsymbol{\Lambda}|^{-\frac{N}{2}}.$$

Thus for all such $\boldsymbol{\Lambda}$

$$\mathbb{P}\left(\max_{k,l} \left|u^k \cdot u^l - NQ_{k,l}\right| \le c_2\varepsilon N\right) \le \exp\left(-\frac{N}{2}\log|\boldsymbol{I} - 2\boldsymbol{\Lambda}| - N\sum_{kl}\Lambda_{k,l}Q_{k,l} + c_2\varepsilon n^2\|\boldsymbol{\Lambda}\|_\infty N\right). \tag{3.5}$$

The non-error terms on the r.h.s are minimized by choosing $\boldsymbol{\Lambda} = \frac{\boldsymbol{I}-\boldsymbol{Q}^{-1}}{2}$, for which $-\frac{1}{2}\log|\boldsymbol{I} - 2\boldsymbol{\Lambda}| = \frac{1}{2}\log|\boldsymbol{Q}|$ and

$$\sum_{kl}\Lambda_{k,l}Q_{k,l} = \mathrm{Tr}\,(\boldsymbol{\Lambda}\boldsymbol{Q}) = \mathrm{Tr}\left(\frac{\boldsymbol{Q}-\boldsymbol{I}}{2}\right) = 0. \tag{3.6}$$

Thus we have that

$$\mathbb{P}\left(\max_{k,l} \left|u^k \cdot u^l - NQ_{k,l}\right| \le c_2\varepsilon N\right) \le \exp\left(\frac{N}{2}\log|\boldsymbol{Q}| + c(1 + \|\boldsymbol{Q}^{-1}\|_2)\varepsilon N\right),$$

since $c_2 n^2\|\boldsymbol{\Lambda}\|_\infty \le c(1 + \|\boldsymbol{Q}^{-1}\|_2)$ for a large enough c depending only on n. This proves the upper bound of (3.3).

To obtain (3.4) let $\boldsymbol{\Lambda} = \frac{\boldsymbol{I}-(\boldsymbol{Q}+\varepsilon\boldsymbol{I})^{-1}}{2}$ and note that then $\sum_{kl}\Lambda_{k,l}Q_{k,l} = \mathrm{Tr}\,(\boldsymbol{\Lambda}\boldsymbol{Q}) = \mathrm{Tr}\left(\frac{\boldsymbol{Q}-(\boldsymbol{Q}+\varepsilon\boldsymbol{I})^{-1}\boldsymbol{Q}}{2}\right) \ge -\frac{n}{2}$ and $\|\boldsymbol{\Lambda}\|_\infty \le c\varepsilon^{-1}$.

For the lower bound of (3.3) we use the change of measure

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \prod_{i=1}^N \frac{\exp\left(\sum_{kl}\Lambda_{k,l}u^k \cdot u^l\right)}{|\boldsymbol{I} - 2\boldsymbol{\Lambda}|^{-1/2}} = \prod_{i=1}^N \frac{\exp\left(\sum_i (u_i)^\top \boldsymbol{\Lambda} u_i\right)}{|\boldsymbol{I} - 2\boldsymbol{\Lambda}|^{-1/2}}$$

for $\boldsymbol{\Lambda} = \frac{\boldsymbol{I}-\boldsymbol{Q}^{-1}}{2}$. Under the measure $\mathbb{Q}$ the u_i are i.i.d. centered Gaussian vectors in $\mathbb{R}^n$ with covariance $\boldsymbol{Q}$. We have

$$\begin{aligned}
&\mathbb{P}\left(\left|u^k \cdot u^l - NQ_{k,l}\right| \le c_1\varepsilon N\right)\\
&= \mathbb{Q}\left(1_{\{|u^k\cdot u^l - NQ_{k,l}| \le c_1\varepsilon N\}}\frac{d\mathbb{P}}{d\mathbb{Q}}\right)\\
&\ge \mathbb{Q}\left(\max_{k,l}\left|u^k \cdot u^l - NQ_{k,l}\right| \le c_1\varepsilon N\right)\frac{\exp\left(-\sum_{kl}\Lambda_{k,l}Q_{k,l} - c_1 n^2\varepsilon\|\boldsymbol{\Lambda}\|_\infty N\right)}{|\boldsymbol{I}-2\boldsymbol{\Lambda}|^{N/2}}\\
&\overset{(3.6)}{\ge} \mathbb{Q}\left(\max_{k,l}\left|\tfrac{1}{N}\textstyle\sum_i u_i^k u_i^l - Q_{k,l}\right| \le c_1\varepsilon\right)\exp\left(\tfrac{N}{2}\log|\boldsymbol{Q}| - c(1 + \|\boldsymbol{Q}^{-1}\|_2)\varepsilon N\right).
\end{aligned} \tag{3.7}$$

Using a union bound and the Chebyshev inequality (recall $\mathbb{Q}\left(u_i^k u_i^l\right) = Q_{k,l}$) we obtain

$$\begin{aligned}
\mathbb{Q}\left(\left\{\max_{k,l}\left|\tfrac{1}{N}\textstyle\sum_i u_i^\top u_i - Q_{k,l}\right| \le c_1\varepsilon\right\}^c\right) &\le \textstyle\sum_{k,l}\mathbb{Q}\left(\left|\tfrac{1}{N}\sum_i u_i^k u_i^l - Q_{k,l}\right| \ge c_1\varepsilon\right)\\
&\le \textstyle\sum_{k,l}\frac{\mathrm{Var}_{\mathbb{Q}}\left(\frac{1}{N}\sum_i u_i^k u_i^l\right)}{c_1 2\varepsilon^2}\\
&= \textstyle\sum_{k,l}\frac{\mathrm{Var}_{\mathbb{Q}}\left(u_1^k u_1^l\right)}{c_1^2\varepsilon^2 N}.
\end{aligned} \tag{3.8}$$

Now crudely bounding

$$\mathrm{Var}_{\mathbb{Q}}\left(u_1^k u_1^l\right) \le \mathbb{Q}\left(\left(u_1^k u_1^l\right)^2\right) \le \sqrt{\mathbb{Q}\left(\left(u_1^k\right)^4\right)}\sqrt{\mathbb{Q}\left(\left(u_1^l\right)^4\right)} \le c,$$

where the last constant is independent of $\boldsymbol{Q}$ since u_1^k is Gaussian with variance $Q_{kk} = 1$ under $\mathbb{Q}$, for all k. Thus provided ε is smaller than some constant depending only on n the r.h.s. of (3.8) is at most $\frac{1}{2}$, and so from (3.7) it follows that

$$\mathbb{P}\left(\left|u^k \cdot u^l - NQ_{k,l}\right| \le c_1\varepsilon N\right) \ge \exp\left(\frac{N}{2}\log|\boldsymbol{Q}| - c(1+\|\boldsymbol{Q}^{-1}\|_2)\varepsilon N - c\right),$$

giving the lower bound of (3.3). □

We now compute the first moment, or equivalently annealed free energy. Recall that $\boldsymbol{A} \ge \delta\boldsymbol{I}$ means that all eigenvalues of $\boldsymbol{A}$ are greater than δ. We also use the notation $\boldsymbol{A}^{\odot 2} = \boldsymbol{A}\odot\boldsymbol{A} = (A_{k,l}^2)_{k,l=1,\dots,n}$ to denote the Hadamard square of the entries of $\boldsymbol{A}$.

Lemma 3.3 (First moment; Annealed free energy in absence of external field)**.** *Let $n \ge 1$. For all $\delta \in (0,1), C > 0$ there exists a constant $c = c(\delta, C) > 0$ so that for all ε less than a universal constant, $|\beta| \le C$ and $N \ge c(\delta,\varepsilon)$ we have*

$$\sup_{\boldsymbol{Q}\ge\delta\boldsymbol{I}}\left|\frac{1}{N}\log\mathbb{E}[Z_N^\varepsilon(\beta,0,\boldsymbol{Q})] - \left(\frac{1}{2}\beta^\mathsf{T}\boldsymbol{Q}^{\odot 2}\beta + \frac{1}{2}\log|\boldsymbol{Q}|\right)\right| \le c\varepsilon, \tag{3.9}$$

where the supremum is taken over all symmetric $\boldsymbol{Q} \in [-1,1]^{n\times n}$ with 1's on the diagonal.

Proof. We have

$$\mathbb{E}[Z_N^\varepsilon(\beta,0,\boldsymbol{Q})] \overset{(3.2)}{=} \mathbb{E}\int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k)} d\boldsymbol{\sigma} = \int_{\boldsymbol{Q}_\varepsilon}\mathbb{E}\left[\exp\left(\sum_{k=1}^n \beta_k H_N(\sigma^k)\right)\right] d\boldsymbol{\sigma}. \tag{3.10}$$

Since the Hamiltonian is a sum of Gaussians for fixed $\boldsymbol{\sigma}$ we have for $\boldsymbol{\sigma}\in\boldsymbol{Q}_\varepsilon$

$$\mathbb{E}\left[\exp\left(\sum_{k=1}^n \beta_k H_N(\sigma^k)\right)\right] = \exp\left(\frac{1}{2}\mathrm{Var}\left(\sum_{k=1}^n \beta_k H_N(\sigma^k)\right)\right). \tag{3.11}$$

Since $\mathbb{E}[H_N(\sigma)H_N(\sigma')] = N(\sigma\cdot\sigma')^2$ we have

$$\mathrm{Var}\left(\sum_{k=1}^n \beta_k H_N(\sigma^k)\right) = \sum_{k,\ell=1}^n \beta_k\beta_\ell(\sigma_k\cdot\sigma_l)^2. \tag{3.12}$$

For $\boldsymbol{\sigma}\in Q_\varepsilon$ we have $|\sum_{k,\ell=1}^n \beta_k\beta_\ell(\sigma_k\cdot\sigma_l)^2 - \sum_{k,\ell=1}^n \beta_k\beta_\ell Q_{k,l}^2| \le c\varepsilon N$ for a constant c depending only on C and n, and $\sum_{k,\ell=1}^n \beta_k\beta_\ell Q_{k,l}^2 = \beta^\mathsf{T}\boldsymbol{Q}^{\odot 2}\beta$. Therefore for all $\boldsymbol{\sigma}\in\boldsymbol{Q}_\varepsilon$

$$\left|\log\mathbb{E}\left[\exp\left(\sum_{k=1}^n \beta_k H_N(\sigma^k)\right)\right] - \frac{1}{2}\beta^\mathsf{T}\boldsymbol{Q}^{\odot 2}\beta\right| \le c\varepsilon N.$$

Lemma 3.2 implies that for $\boldsymbol{Q}\ge\delta\boldsymbol{I}$

$$\left|\log\int_{\boldsymbol{Q}_\varepsilon} 1 d\boldsymbol{\sigma} - \frac{N}{2}\log|\boldsymbol{Q}|\right| \le c\varepsilon N,$$

which completes the proof. □

Next we compute the second moment.

Lemma 3.4 (Second moment)**.** *Let $n \geq 1$. For all $\delta, C > 0$ there exists a constant $c = c(\delta, C) > 0$ so that for all $\varepsilon \in (0, c^{-1}), |\beta| \leq C$ and all $N \geq c(\varepsilon, \delta)$ we have*

$$\sup_{\boldsymbol{Q} \geq \delta \boldsymbol{I}} \left(\frac{1}{N} \log \mathbb{E}[Z_N^\varepsilon(\beta, 0, \boldsymbol{Q})^2] - \left(\beta^\mathsf{T} \boldsymbol{Q}^{\odot 2} \beta + \sup_{\boldsymbol{A}} V(\boldsymbol{A}) \right) \right) \leq c\varepsilon, \tag{3.13}$$

where the supremum is taken over all symmetric $\boldsymbol{Q} \in [-1,1]^{n \times n}$ with 1's on the diagonal and

$$V(\boldsymbol{A}) = \beta^\mathsf{T} \boldsymbol{A}^{\odot 2} \beta + \frac{1}{2} \log \begin{vmatrix} \boldsymbol{Q} & \boldsymbol{A} \\ \boldsymbol{A}^\mathsf{T} & \boldsymbol{Q} \end{vmatrix}. \tag{3.14}$$

Proof. We have

$$\begin{aligned} \mathbb{E}[Z_N^\varepsilon(\beta, 0, \boldsymbol{Q})^2] &= \mathbb{E}\left[\int_{\boldsymbol{Q}_\varepsilon} \int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k)} e^{\sum_{k=1}^n \beta_\ell H_N(\tau^\ell)} d\boldsymbol{\sigma} d\boldsymbol{\tau} \right] \\ &= \int_{\boldsymbol{Q}_\varepsilon} \int_{\boldsymbol{Q}_\varepsilon} \mathbb{E}\left[e^{\sum_{k=1}^n \beta_k H_N(\sigma^k) + \sum_{\ell=1}^n \beta_\ell H_N(\tau^\ell)} \right] d\boldsymbol{\sigma} d\boldsymbol{\tau}. \end{aligned}$$

Similarly to in the proof of the previous lemma the inner expectation is a Gaussian exponential moment that satisfies

$$\left| \log \mathbb{E}\left[e^{\sum_{k=1}^n \beta_k H_N(\sigma^k) + \sum_{\ell=1}^n \beta_\ell H_N(\tau^\ell)} \right] - N \beta^\mathsf{T} \boldsymbol{Q}^{\odot 2} \beta - N \sum_{k,\ell=1}^n \beta_k \beta_\ell \left(\sigma^k \cdot \tau^\ell \right)^2 \right| \leq N c \varepsilon,$$

for all $\sigma, \tau \in \boldsymbol{Q}_\varepsilon$, where c depends only on C, n.

It remains to prove that

$$\int_{\boldsymbol{Q}_\varepsilon} \int_{\boldsymbol{Q}_\varepsilon} e^{N \sum_{k,\ell=1}^n \beta_k \beta_\ell \left(\sigma^k \cdot \tau^\ell \right)^2} d\boldsymbol{\sigma} d\boldsymbol{\tau} \leq \exp\left(N \sup_{\boldsymbol{A}} V(\boldsymbol{A}) + N c \varepsilon \right). \tag{3.15}$$

By partitioning the space $[-1,1]^{n \times n}$ into at most $\lceil \frac{1}{2\varepsilon} \rceil^{n^2}$ subsets of diameter of order ε one obtains that for all $\varepsilon > 0$

$$\begin{aligned} &\int_{\boldsymbol{Q}_\varepsilon} \int_{\boldsymbol{Q}_\varepsilon} e^{N \sum_{k,\ell=1}^n \beta_k \beta_\ell \left(\sigma^k \cdot \tau^\ell \right)^2} d\boldsymbol{\sigma} d\boldsymbol{\tau} \\ &\leq \exp\left(N \sup_{A \in [-1,1]^{n \times n}} \left\{ \beta^\mathsf{T} \boldsymbol{A}^{\odot 2} \beta + \tfrac{1}{N} \log \int_{\boldsymbol{Q}_\varepsilon} \int_{\boldsymbol{Q}_\varepsilon} 1_{\{\|\boldsymbol{\sigma} \boldsymbol{\tau}^\mathsf{T} - \boldsymbol{A}\|_\infty \leq \varepsilon\}} d\boldsymbol{\sigma} d\boldsymbol{\tau} \right\} + N c \varepsilon \right), \end{aligned} \tag{3.16}$$

for a c depending on C, for all $N \geq c(\varepsilon)$. Note that the second term in the supremum equals

$$\frac{1}{N} \log \int 1_{\left\{ \left\| \boldsymbol{\nu} \boldsymbol{\nu}^\mathsf{T} - \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A} \\ \boldsymbol{A}^\mathsf{T} & \boldsymbol{Q} \end{pmatrix} \right\|_\infty \leq \varepsilon \right\}} d\boldsymbol{\nu},$$

where the integral is over $\boldsymbol{\nu} \in \mathcal{S}_{N-1}^{2n}$.

Note that for any matrix $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ such that $\|\boldsymbol{B}\|_\infty \leq 1$, we have

$$\|\boldsymbol{B}\|_2 \leq n \text{ and } \delta^n \leq |\boldsymbol{B}| \leq \delta n^{n-1} \text{ if } \delta \geq 0 \text{ is } \boldsymbol{B}\text{'s smallest eigenvalue.} \tag{3.17}$$

Fix a $\tilde{\delta} \in \left(0, \frac{1}{2}\right)$ small enough depending only on δ, n, C such that

$$\frac{1}{4} \log(2\tilde{\delta}) + \frac{1}{2} \log (4n)^{2n-1} + n^2 \max_i \beta_i^2 \leq n \log \delta \overset{(3.17)}{\leq} V(0).$$

Then if $\boldsymbol{A}$ is s.t. $\begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A} \\ \boldsymbol{A}^\mathsf{T} & \boldsymbol{Q} \end{pmatrix}$ has an eigenvalue smaller or equal to $\tilde{\delta}$ and $\varepsilon \in (0, \tilde{\delta})$ then by (3.4) and (3.17)

$$\frac{1}{N} \log \int 1_{\left\{ \left\| \boldsymbol{\nu} \boldsymbol{\nu}^\mathsf{T} - \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A} \\ \boldsymbol{A}^\mathsf{T} & \boldsymbol{Q} \end{pmatrix} \right\|_\infty \leq \varepsilon \right\}} d\boldsymbol{\nu} \leq V(0) - N \beta^\mathsf{T} \boldsymbol{A}^{\odot 2} \beta. \tag{3.18}$$

(after possible decreasing $\tilde{\delta}$ further depending on the constant in (3.4)). If on the other hand $\begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A} \\ \boldsymbol{A}^\intercal & \boldsymbol{Q} \end{pmatrix} > \delta' \boldsymbol{I}$ then (3.3) implies that

$$\frac{1}{N} \log \int 1_{\left\{ \left\| \boldsymbol{\nu}\boldsymbol{\nu}^\intercal - \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{A} \\ \boldsymbol{A}^\intercal & \boldsymbol{Q} \end{pmatrix} \right\|_\infty \leq \varepsilon \right\}} d\boldsymbol{\nu} \leq \frac{1}{2} \log \begin{vmatrix} \boldsymbol{Q} & \boldsymbol{A} \\ \boldsymbol{A}^\intercal & \boldsymbol{Q} \end{vmatrix} + c\tilde{\delta}^{-1}\varepsilon. \tag{3.19}$$

The bounds (3.16), (3.18) and (3.19) imply (3.15). □

We will show that $V(\boldsymbol{A})$ is maximized at zero for β that lie in

$$\mathrm{HT}(\boldsymbol{Q}) := \left\{ \beta \in \mathbb{R}^n : \|\boldsymbol{\beta}^{\frac{1}{2}} \boldsymbol{Q} \boldsymbol{\beta}^{\frac{1}{2}}\|_2 \leq \frac{1}{\sqrt{2}} \right\} \tag{3.20}$$

which is the high temperature region of the model (recall that $\boldsymbol{\beta} = \mathrm{diag}(\beta_1, \ldots, \beta_n)$). Together with Lemma 3.4 this will give us

$$\frac{1}{N} \log \mathbb{E}[Z_N^\varepsilon(\beta, 0, \boldsymbol{Q})^2] - \frac{1}{N} \log \mathbb{E}[Z_N^\varepsilon(\beta, 0, \boldsymbol{Q})]^2 \leq c\varepsilon$$

for $\beta \in \mathrm{HT}(\boldsymbol{Q})$, with which we can use a second moment method to prove concentration of $Z_N^\varepsilon(\beta, 0, \boldsymbol{Q})$ for such β.

In the computation showing that $V(\boldsymbol{A})$ is maximized at zero a different form of the high temperature condition naturally appears. The next lemma shows that this form is equivalent to the condition in (3.20).

Lemma 3.5 (Equivalence of the two forms of high temperature condition)**.** *For any positive definite symmetric matrix $\boldsymbol{Q}$,*

$$\mathrm{HT}(\boldsymbol{Q}) = \left\{ \beta \in \mathbb{R}^n : \sup_{\|\boldsymbol{B}\|_F = 1} \|\boldsymbol{\beta}^{\frac{1}{2}} \boldsymbol{Q}^{\frac{1}{2}} \boldsymbol{B} \boldsymbol{Q}^{\frac{1}{2}} \boldsymbol{\beta}^{\frac{1}{2}}\|_F \leq \frac{1}{\sqrt{2}} \right\}, \tag{3.21}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Proof. The claim follows once we have shown that

$$\sup_{\|\boldsymbol{B}\|_F^2 = 1} \|\boldsymbol{\beta}^{1/2} \boldsymbol{Q}^{1/2} \boldsymbol{B} \boldsymbol{Q}^{1/2} \boldsymbol{\beta}^{1/2}\|_F = \|\boldsymbol{\beta}^{\frac{1}{2}} \boldsymbol{Q} \boldsymbol{\beta}^{\frac{1}{2}}\|_2. \tag{3.22}$$

To this end note that

$$\begin{aligned} \|\boldsymbol{\beta}^{1/2} \boldsymbol{Q}^{1/2} \boldsymbol{B} \boldsymbol{Q}^{1/2} \boldsymbol{\beta}^{1/2}\|_F^2 &= \mathrm{Tr}\left(\boldsymbol{\beta}^{1/2} \boldsymbol{Q}^{1/2} \boldsymbol{B} \boldsymbol{Q}^{1/2} \boldsymbol{\beta}^{1/2} \left(\boldsymbol{\beta}^{1/2} \boldsymbol{Q}^{1/2} \boldsymbol{B} \boldsymbol{Q}^{1/2} \boldsymbol{\beta}^{1/2} \right)^\intercal \right) \\ &= \mathrm{Tr}\left(\boldsymbol{B} \boldsymbol{Q}^{1/2} \boldsymbol{\beta} \boldsymbol{Q}^{1/2} \boldsymbol{B}^\intercal \boldsymbol{Q}^{1/2} \boldsymbol{\beta} \boldsymbol{Q}^{1/2} \right) \\ &= \|\boldsymbol{B} \boldsymbol{Q}^{1/2} \boldsymbol{\beta} \boldsymbol{Q}^{1/2}\|_F^2. \end{aligned}$$

Let $\tilde{\boldsymbol{\beta}}$ be the diagonal matrix of eigenvalues of $\boldsymbol{Q}^{1/2} \boldsymbol{\beta} \boldsymbol{Q}^{1/2}$, and let $\tilde{\boldsymbol{B}}$ denote $\boldsymbol{B}$ in the (orthogonal) diagonalizing basis of $\boldsymbol{Q}^{1/2} \boldsymbol{\beta} \boldsymbol{Q}^{1/2}$. Then $\|\tilde{\boldsymbol{B}}\|_F^2 = \|\boldsymbol{B}\|_F^2$ and $\|\boldsymbol{B} \boldsymbol{Q}^{1/2} \boldsymbol{\beta} \boldsymbol{Q}^{1/2}\|_F^2 = \|\tilde{\boldsymbol{B}} \tilde{\boldsymbol{\beta}}\|_F^2$, so

$$\sup_{\|\boldsymbol{B}\|_F^2 = 1} \|\boldsymbol{\beta}^{1/2} \boldsymbol{Q}^{1/2} \boldsymbol{B} \boldsymbol{Q}^{1/2} \boldsymbol{\beta}^{1/2}\|_F^2 = \sup_{\|\tilde{\boldsymbol{B}}\|_F^2 = 1} \|\tilde{\boldsymbol{B}} \tilde{\boldsymbol{\beta}}\|_F^2.$$

Since $\|\tilde{\boldsymbol{B}} \tilde{\boldsymbol{\beta}}\|_F^2 = \sum_i \left(\sum_j \tilde{B}_{i,j}^2 \right) \tilde{\beta}_{i,i}^2$ the r.h.s. clearly equals $\max_i \tilde{\beta}_{i,i}^2$. Since AB and BA have the same eigenvalues for any square matrices A, B, also

$$\boldsymbol{Q}^{1/2} \boldsymbol{\beta} \boldsymbol{Q}^{1/2} \quad \text{and} \quad \boldsymbol{\beta}^{1/2} \boldsymbol{Q} \boldsymbol{\beta}^{1/2} \quad \text{have the same eigenvalues,} \tag{3.23}$$

and this proves (3.22). □

We are now ready to show that $V(\boldsymbol{A})$ is maxmized for $A = 0$ when $\beta \in \mathrm{HT}(\boldsymbol{Q})$.

Proposition 3.6. *For any positive definite $\boldsymbol{Q}$ and $\beta \in \mathrm{HT}(\boldsymbol{Q})$ it holds that*

$$\sup_{\boldsymbol{A}} V(A) = V(0). \tag{3.24}$$

Proof. Using the Schur complement formula

$$\begin{vmatrix} \boldsymbol{Q} & \boldsymbol{A} \\ \boldsymbol{A}^\top & \boldsymbol{Q} \end{vmatrix} = |\boldsymbol{Q}||\boldsymbol{Q} - \boldsymbol{A}^\top \boldsymbol{Q}^{-1}\boldsymbol{A}|.$$

We have

$$|\boldsymbol{Q} - \boldsymbol{A}^\top \boldsymbol{Q}^{-1}\boldsymbol{A}| = |\boldsymbol{Q} - \boldsymbol{A}^\top \boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}| = |\boldsymbol{Q} - (\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A})^\top(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A})|.$$

By the matrix determinant lemma this equals

$$|\boldsymbol{Q}||\boldsymbol{I} - (\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A})\boldsymbol{Q}^{-1}(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A})^\top| = |\boldsymbol{Q}||\boldsymbol{I} - (\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{Q}^{-\frac{1}{2}})(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}^\top\boldsymbol{Q}^{-\frac{1}{2}})^\top|. \tag{3.25}$$

Thus

$$V(\boldsymbol{A}) = \beta^\top \boldsymbol{A}^{\odot 2}\beta + \log|\boldsymbol{Q}| + \frac{1}{2}\log|\boldsymbol{I} - (\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{Q}^{-\frac{1}{2}})(\boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}^\top\boldsymbol{Q}^{-\frac{1}{2}})^\top|.$$

Now make the change of variables $\boldsymbol{B} = \boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{A}^\top\boldsymbol{Q}^{-\frac{1}{2}} \Leftrightarrow \boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}^\top\boldsymbol{Q}^{\frac{1}{2}} = \boldsymbol{A}$ to obtain

$$V(\boldsymbol{A}) = \beta^\top(\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}^\top\boldsymbol{Q}^{\frac{1}{2}})^{\odot 2}\beta + \log|\boldsymbol{Q}| + \frac{1}{2}\log|\boldsymbol{I} - \boldsymbol{B}^\top\boldsymbol{B}|. \tag{3.26}$$

It thus suffices to show that the right-hand side is maximizes for $\boldsymbol{B} = 0$.

To this end we first optimize along rays by fixing $\boldsymbol{B}$ and considering

$$v(t) = V(\sqrt{t}\boldsymbol{B}) = t\beta^\top(\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}^\top\boldsymbol{Q}^{\frac{1}{2}})^{\odot 2}\beta + \log|\boldsymbol{Q}| + \frac{1}{2}\log|\boldsymbol{I} - t\boldsymbol{B}^\top\boldsymbol{B}|, t \geq 0.$$

The functional $v(t)$ is clearly concave in $[0, \infty)$ because the first term is linear in t, the second term is constant, and the last term is concave in t (for instance by diagonalizing $\boldsymbol{B}\boldsymbol{B}^\top$). Thus to show that $v(t)$ has a global maximum at 0, it suffices to show that $v'(0) \leq 0$.

We have

$$v'(0) = \beta^\top(\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}^\top\boldsymbol{Q}^{\frac{1}{2}})^{\odot 2}\beta + \frac{1}{2}\frac{d}{dt}\log|\boldsymbol{I} - t\boldsymbol{B}^\top\boldsymbol{B}|\Big|_{t=0}.$$

Since

$$\frac{d}{dt}\log|\boldsymbol{I} - t\boldsymbol{B}^\top\boldsymbol{B}|\Big|_{t=0} = -\mathrm{Tr}(\boldsymbol{I} - t\boldsymbol{B}^\top\boldsymbol{B})^{-1}\boldsymbol{B}^\top\boldsymbol{B})\Big|_{t=0} = -\mathrm{Tr}(\boldsymbol{B}^\top\boldsymbol{B}) = -\|\boldsymbol{B}\|_F^2,$$

and $w^\top\boldsymbol{A}^{\odot 2}w = \|\mathrm{diag}(w)^{\frac{1}{2}}\boldsymbol{A}\mathrm{diag}(w)^{\frac{1}{2}}\|_F^2$ for any vector w and matrix $\boldsymbol{A}$ we obtain

$$v'(0) = \|\boldsymbol{\beta}^{1/2}\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{\beta}^{1/2}\|_F^2 - \frac{1}{2}\|\boldsymbol{B}\|_F^2 = \|\boldsymbol{B}\|_F^2\left(\|\boldsymbol{\beta}^{1/2}\boldsymbol{Q}^{\frac{1}{2}}\hat{\boldsymbol{B}}\boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{\beta}^{1/2}\|_F^2 - \frac{1}{2}\right),$$

where $\hat{\boldsymbol{B}} = \boldsymbol{B}/\|\boldsymbol{B}\|_F$. If β satisfies the high temperature condition (3.21) then the r.h.s. is non-negative, so $v'(0) \leq 0$ and indeed $v(t), t \in [0, \infty)$ is maximized at $t = 0$.

But since this holds for any $\boldsymbol{B}$, it must be that the r.h.s. of (3.26) is maximized when $\boldsymbol{B} = \boldsymbol{0}$, so $\boldsymbol{A} = \boldsymbol{Q}^{\frac{1}{2}}\boldsymbol{B}^\top\boldsymbol{Q}^{\frac{1}{2}} = \boldsymbol{0}$ is the global maximizer of the functional $V(\boldsymbol{A})$. □

Remark 3.7. *Note that for v in the previous proof $v'(0) \leq 0$ for all $\boldsymbol{B}$* only if *the condition* (3.20) *is satisfied, so the reverse of the implication of the Proposition also holds (though we do not need this fact).*

For the lower bound of the free energy with the second moment method one needs the standard exponential concentration inequality.

Lemma 3.8 (Exponential concentration for free energy)**.** *Let $n \geq 1$ and $C > 0$. There exists a $c = c(C) > 0$ such that for all $\varepsilon > 0$, $|\beta| \leq C$ and $\boldsymbol{Q} > 0$*

$$\mathbb{P}\left(\left|F_N^\varepsilon(\beta, 0, \boldsymbol{Q}) - \mathbb{M}\left[F_N^\varepsilon(\beta, 0, \boldsymbol{Q})\right]\right| \geq t\right) \leq \exp\left(-ct^2 N\right), \tag{3.27}$$

where $\mathbb{M}$ denotes the median.

Proof. This follows by Gaussian concentration [BLM13, Theorem 10.17], since for all $i, j, \beta, \boldsymbol{Q}$

$$\partial_{J_{ij}} F_N^\varepsilon(\beta, 0, \boldsymbol{Q}) \overset{(1.3)}{=} \frac{1}{\sqrt{N}} \sum_{k=1}^n \beta_k \left\langle \sigma_i^k \sigma_j^k \right\rangle,$$

where $\langle \cdot \rangle = \int_{\boldsymbol{Q}_\varepsilon} \cdot \, e^{\sum_{k=1}^n \beta_k H_N(\sigma^k)} d\boldsymbol{\sigma} / \int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k)} d\boldsymbol{\sigma}$ denotes the expectation over the Gibbs measure, so that when $|\beta| \leq C$

$$\left|\nabla_J F_N^\varepsilon(\beta, 0, \boldsymbol{Q})\right|^2 \leq \frac{C^2}{N} \sum_{i,j} \sum_{k=1}^n \left\langle \sigma_i^k \sigma_j^k \right\rangle^2 \leq \frac{C^2}{N} \sum_{i,j} \sum_{k=1}^n \left\langle \left(\sigma_i^k\right)^2 \left(\sigma_j^k\right)^2 \right\rangle = \frac{C^2 n}{N},$$

implying that the map $J \to F_N^\varepsilon(\beta, 0, \boldsymbol{Q})$ is Lipschitz with Lipschitz constant $N^{-1/2} C n^{1/2}$. □

To obtain an estimate for the free energy uniformly over β and $\boldsymbol{Q}$ (see (3.29)) we will use the next result.

Lemma 3.9 (Lipschitz property of the free energy)**.** *Let $n \geq 1, C > 0, \varepsilon > 0$. There exists a $L = L(C) > 0$ such that*

$$\lim_{N \to \infty} \mathbb{P}\left(\forall \boldsymbol{Q} > 0 \text{ and } |\beta^1|, |\beta^2| \leq C: \ \left|F_N^\varepsilon(\beta^1, 0, \boldsymbol{Q}) - F_N^\varepsilon(\beta^2, 0, \boldsymbol{Q})\right| \leq L\left|\beta^1 - \beta^2\right|\right) = 1.$$

Proof. We have for any $k \in \{1, ..., n\}$ that

$$\frac{\partial}{\partial \beta^k} F_N^\varepsilon(\beta, 0, \boldsymbol{Q}) \overset{(1.3)}{=} \langle \tfrac{1}{N} \beta_k H_N(\sigma^k) \rangle \tag{3.28}$$

where $\langle \cdot \rangle$ denotes the expectation over the Gibbs measure as in the previous lemma. Thus by (2.1) we have for $L = L(C)$ large enough

$$\mathbb{P}\left(\sup_{\boldsymbol{Q}, |\beta| \leq C} \left|\frac{\partial}{\partial \beta^k} F_N^\varepsilon(\beta, 0, \boldsymbol{Q})\right| \leq L\right) \to 1.$$

This implies that $F_N^\varepsilon(\beta, 0, \boldsymbol{Q})$ is Lipschitz continuous in β with probability tending to one. □

The next Proposition will now combine all previous arguments to show that F_N^ε concentrates as $N \to \infty$ if the external field $\boldsymbol{h}$ is zero and β lies in the high temperature region.

Proposition 3.10 (Free energy at high temperature)**.** *Let $n \geq 1$. Let $\delta, C > 0$ be some constants. There exists a $c = c(\delta, C) > 0$ such that for all $\varepsilon \in (0, c^{-1})$*

$$\lim_{N \to \infty} \mathbb{P}\left(\sup_{\boldsymbol{Q} \geq \delta \boldsymbol{I}} \sup_{\substack{\beta \in \mathrm{HT}(\boldsymbol{Q}) \\ |\beta| \leq C}} \left|F_N^\varepsilon(\beta, 0, \boldsymbol{Q}) - \frac{1}{2}(\beta^\mathsf{T} \boldsymbol{Q}^{\odot 2} \beta + \log|\boldsymbol{Q}|)\right| \leq c\sqrt{\varepsilon}\right) = 1, \tag{3.29}$$

where the supremum is taken over all symmetric $\boldsymbol{Q} \in [-1, 1]^{n \times n}$ with 1's on the diagonal.

Proof. Let $\boldsymbol{A}^i \in [-1, 1]^{n \times n}, i = 1, \ldots, M$, with $M \leq \lceil \frac{1}{2\varepsilon} \rceil^{n^2}$ such that for all $\boldsymbol{Q} \in [-1, 1]^{n \times n}$ there exists an i such that

$$\|\boldsymbol{A}^i - \boldsymbol{Q}\|_\infty \leq \varepsilon, \quad \|(\boldsymbol{A}^i)^{\odot 2} - \boldsymbol{Q}^{\odot 2}\|_\infty \leq 2\varepsilon, \quad \left|\left|\boldsymbol{A}^i\right| - |\boldsymbol{Q}|\right| \leq n^{n+1}\varepsilon. \tag{3.30}$$

For this i we have $\boldsymbol{A}^i{}_{\frac{1}{2}\varepsilon} \subset \boldsymbol{Q}_\varepsilon \subset \boldsymbol{A}^i{}_{2\varepsilon}$ and thus for all β

$$F_N^{\frac{1}{2}\varepsilon}(\beta, 0, \boldsymbol{A}^i) \;\le\; F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) \;\le\; F_N^{2\varepsilon}(\beta, 0, \boldsymbol{A}^i). \tag{3.31}$$

We also construct a finite sequence $\beta^1, ..., \beta^L$ with $L \le \lceil \frac{C}{\varepsilon} \rceil^n$ such that for each $\beta \in \{b \in \mathbb{R}^n : |b| \le C\}$ there is a $j \le L$ with $|\beta - \beta^j| \le \varepsilon$. Then with probability tending to one by Lemma 3.9 there is for each β with $|\beta| \le C$ some j such that

$$\left|F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) - F_N^{\varepsilon}(\beta^j, 0, \boldsymbol{Q})\right| \le c(C)\varepsilon \;\text{ for all } \boldsymbol{Q} \tag{3.32}$$

and for each $\boldsymbol{Q}$ with $\boldsymbol{Q} > \delta \boldsymbol{I}$ and the i such that (3.31) holds

$$\left|\beta^\top \boldsymbol{Q}^{\odot 2}\beta + \log|\boldsymbol{Q}| - (\beta^j)^\top \boldsymbol{A}^{i\odot 2}\beta^j - \log\left|\boldsymbol{A}^i\right|\right| \le c\,(\delta, n, C)\,\varepsilon \tag{3.33}$$

provided ε is small enough depending on δ.

Upper bound: This implies that for if the constant c is chosen large enough depending on δ, C, n then

$$\begin{aligned}&\mathbb{P}\left(\forall \boldsymbol{Q} \ge \delta \boldsymbol{I},\ |\beta| \le C:\ Z_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) > \exp\left(\frac{N}{2}\left(\beta^\top \boldsymbol{Q}^{\odot 2}\beta + \log|\boldsymbol{Q}| + c\varepsilon\right)\right)\right)\\ \le&\mathbb{P}\left(\exists i = 1, ..., M,\ j = 1, ..., L : Z_N^{2\varepsilon}(\beta^j, 0, \boldsymbol{A}^i) > \exp\left(\frac{N}{2}\left((\beta^j)^\top(\boldsymbol{A}^i)^{\odot 2}\beta^j + \log|\boldsymbol{A}^i| + \frac{c}{2}\varepsilon\right)\right)\right),\end{aligned}$$

where by Markov's inequality and Lemma 3.3 the r.h.s. is bounded by

$$\sum_{i=1}^{M}\sum_{j=1}^{L}\frac{\mathbb{E}\left[Z_N^{2\varepsilon}(\beta^j, 0, \boldsymbol{A}^i)\right]}{\exp\left(\frac{N}{2}\left((\beta^j)^\top(\boldsymbol{A}^i)^{\odot 2}\beta^j + \log|\boldsymbol{A}^i| + \frac{c}{2}\varepsilon\right)\right)} \le \exp\left(-N\frac{c}{4}\varepsilon\right)$$

and thus

$$\mathbb{P}\left(\forall \boldsymbol{Q} \ge \delta \boldsymbol{I},\ |\beta| \le C:\ F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) \le \frac{1}{2}(\beta^\top \boldsymbol{Q}^{\odot 2}\beta) + \frac{1}{2}\log|\boldsymbol{Q}| + c\varepsilon\right) \to 1.$$

Lower bound: By the Paley-Zygmund inequality and Lemmas 3.3 and 3.4 we have for any large enough c depending on δ, C, n that for all $\varepsilon \in (0, c^{-1})$ and $N \ge c\,(\varepsilon, \delta)$ and i, j

$$\begin{aligned}&\mathbb{P}\left(F_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i) > \tfrac{1}{2}(\beta^j)^\top(\boldsymbol{A}^i)^{\odot 2}\beta^j + \tfrac{1}{2}\log|\boldsymbol{A}^i| - \tfrac{c}{8}\varepsilon\right)\\ \ge\;&\mathbb{P}\left(Z_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i) > \tfrac{1}{2}\mathbb{E}[Z_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i)]\right)\\ \ge\;&\tfrac{1}{4}\tfrac{\mathbb{E}[Z_N^{\frac{\varepsilon}{2}}(\beta^j,0,\boldsymbol{A}^i)]^2}{\mathbb{E}[Z_N^{\frac{\varepsilon}{2}}(\beta^j,0,\boldsymbol{A}^i)^2]}\\ \ge\;&\tfrac{1}{4}e^{-c\varepsilon N}.\end{aligned} \tag{3.34}$$

Since otherwise there is a contradiction by Lemma 3.8 (after possibly enlarging c) this implies that $\mathbb{M}\left(F_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i)\right) \ge \frac{1}{2}(\beta^j)^\top(\boldsymbol{A}^i)^{\odot 2}\beta^j + \frac{1}{2}\log|\boldsymbol{A}^i| - \frac{c}{4}\sqrt{\varepsilon}$ for all $\varepsilon \in (0, c^{-1})$ and $N \ge N\,(\varepsilon, \delta)$, and then another use of Lemma 3.8 implies that

$$\lim_{N\to\infty}\mathbb{P}\left(F_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i) < \frac{1}{2}(\beta^j)^\top(\boldsymbol{A}^i)^{\odot 2}\beta^j + \frac{1}{2}\log|\boldsymbol{A}^i| - \frac{c}{2}\sqrt{\varepsilon}\right) = 0 \text{ for all } \varepsilon \in (0, c^{-1}), i, j.$$

Then (possibly enlarging c again) we have

$$\mathbb{P}\left(\exists \boldsymbol{Q} \ge \delta \boldsymbol{I},\ |\beta| \le C:\ F_N^{\varepsilon}(\beta, 0, \boldsymbol{Q}) < \frac{1}{2}(\beta^\top \boldsymbol{Q}^{\odot 2}\beta) + \frac{1}{2}\log|\boldsymbol{Q}| - c\sqrt{\varepsilon}\right) \tag{3.35}$$

$$\le \sum_{i=1}^{M}\sum_{j=1}^{L}\mathbb{P}\left(F_N^{\frac{\varepsilon}{2}}(\beta^j, 0, \boldsymbol{A}^i) < \exp\left(\frac{1}{2}(\beta^j)^\top(\boldsymbol{A}^i)^{\odot 2}\beta^j + \frac{1}{2}\log|\boldsymbol{A}^i| - \frac{c}{2}\sqrt{\varepsilon}\right)\right) \to 0 \tag{3.36}$$

for all $\varepsilon \in (0, c^{-1})$, which gives the lower bound.

$\square$

3.2. With external field

We now prove the lower bound at all temperatures in the presence of an external field. We will follow the proof of [BK19, Lemma 5]. We start by showing that Lemma 3.4 still holds if we restrict the integral in the parition function to the intersection of the product of unit spheres with hyperplanes of high dimension.

In the following it will be convenient to denote the integral $\int \cdot d\sigma$ over the sphere $\mathcal{S}_{N-1}$ and the integral $\int \cdot d\boldsymbol{\sigma}$ over $\mathcal{S}_{N-1}^n$ by $E[\cdot]$. For a subspace $U \subset \mathbb{R}^{n\times N}$ let us write E^U to denote the expectation/integral with respect to $\boldsymbol{\sigma}$ conditioned on $\boldsymbol{\sigma} \in U$.

Lemma 3.11. *For all $\delta > 0$ it holds that*

$$\mathbb{P}\left(\sup_{\boldsymbol{Q}\geq\delta\boldsymbol{I}}\sup_{\beta\in\mathrm{HT}(\boldsymbol{Q})}\sup_{U}\left|\frac{1}{N}\log E^{U^\perp}\left[1_{\boldsymbol{Q}_\varepsilon}e^{\sum_{k=1}^n\beta_k H_N(\sigma^k)}\right]-\left(\frac{\beta^\mathsf{T}\boldsymbol{Q}^{\odot 2}\beta}{2}+\frac{1}{2}\log|\boldsymbol{Q}|\right)\right|\leq c\sqrt{\varepsilon}\right)\to 1, \tag{3.37}$$

as $N \to \infty$, where the innermost supremum is over all subspaces of dimension $N-2n$ and the outermost supremum is taken over all symmetric $\boldsymbol{Q} \in [-1,1]^{n\times n}$ with 1's on the diagonal.

Proof. Define an orthonormal basis $w_1, ..., w_N$ of $\mathbb{R}^N$ such that

$$U = \langle w_{N-2n}, \dots, w_N\rangle.$$

Let $\boldsymbol{A}$ be the top left $(N-2n)\times(N-2n)$-minor of $\frac{\boldsymbol{J}+\boldsymbol{J}^\mathsf{T}}{2}$ when written the basis $w_1, ..., w_N$. For $\sigma \in U^\perp$ we have $H_N(\sigma) = \sum_{i,j=1}^{N-2n}\tilde{\sigma}_i\tilde{\sigma}_j A_{ij} = N\sum_{i=1}^{N-2n}a_i(\sigma_i)^2$ where $\tilde{\sigma}$ is σ in the basis $w_1, \dots, w_N$, and $Na_1 < ... < Na_{N-2n}$ are the eigenvalues of $\boldsymbol{A}$. Thus

$$E^{U^\perp}\left[1_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\beta_k H_N(\sigma^k)\right)\right] = E^{N-2n}\left[1_{\boldsymbol{Q}_\varepsilon}\exp\left(N\sum_{k=1}^n\beta_k\sum_{i=1}^{N-2n}a_i(\sigma_i^k)^2\right)\right], \tag{3.38}$$

where E^{N-2n} is the expectation over $\boldsymbol{\sigma}$ uniform on $\mathcal{S}_{N-2n-1}^n$. Let $\boldsymbol{B}$ be the top left $(N-2n)\times(N-2n)$-minor of $\frac{\boldsymbol{J}+\boldsymbol{J}^\mathsf{T}}{2}$ when written in standard basis and let $Nb_1 < ... < Nb_{N-2n}$ be its eigenvalues. Recalling that $N\lambda_1 < \dots < N\lambda_N$ are the eigenvalues of $\frac{\boldsymbol{J}+\boldsymbol{J}^\mathsf{T}}{2}$ so by Cauchy's eigenvalue interlacing inequality (see [Par98, Theorem 10.1.1]) we have $\lambda_i < a_i, b_i < \lambda_{i+2n}$. Thus (2.3) implies that

$$\sup_U\sup_{\boldsymbol{\sigma}\in\mathcal{S}_{N-2n-1}^n}\left|\sum_{k=1}^n\beta_k\sum_{i=1}^{N-2n}a_i(\sigma_i^k)^2-\sum_{k=1}^n\beta_k\sum_{i=1}^{N-2n}\frac{\sqrt{N-2n}}{\sqrt{N}}b_i(\sigma_i^k)^2\right|\xrightarrow{\mathbb{P}}0 \text{ as } N\to\infty, \tag{3.39}$$

and so

$$E^{U^\perp}\left[1_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\beta_k H_N(\sigma^k)\right)\right] = E^{N-2n}\left[1_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\beta_k(\sigma^k)^\mathsf{T}\frac{\sqrt{N-2n}}{\sqrt{N}}\boldsymbol{B}\sigma^k\right)\right]e^{o(N)}, \tag{3.40}$$

uniformly in U. By applying Proposition 3.10 with $N-2n$ in place of N the r.h.s. equals

$$e^{\frac{N-2n}{2}(\beta^\mathsf{T}\boldsymbol{Q}^{\odot 2}\beta+\log|\boldsymbol{Q}|)(1+\mathcal{O}(\sqrt{\varepsilon}))}, \tag{3.41}$$

for all $\boldsymbol{Q} \geq \delta\boldsymbol{I}$ and $\beta \in \mathrm{HT}(\boldsymbol{Q})$ (note that $\mathrm{HT}(\boldsymbol{Q})$ is a bounded set), proving the claim □

Let us define the matrix $\hat{\boldsymbol{Q}}(\boldsymbol{m}) \in \mathbb{R}^{n\times n}$ given by

$$\hat{Q}_{k,\ell} = \frac{Q_{k,\ell}-m^k\cdot m^\ell}{\sqrt{1-|m^k|^2}\sqrt{1-|m^\ell|^2}}, \tag{3.42}$$

which is thus a function of $\boldsymbol{Q}$ and $\boldsymbol{m}\boldsymbol{m}^\mathsf{T}$.

The next lemma will be used in the proof of Proposition 3.1 to exclude $\boldsymbol{m}$ such that $\hat{\boldsymbol{Q}}(\boldsymbol{m})$ has a small eigenvalue or $\boldsymbol{m}\boldsymbol{m}^\mathsf{T}$ has an entry close to 1.

Lemma 3.12. *For any $\beta, (h_1, \dots, h_n)$ and positive symmetric $\boldsymbol{Q} \in [-1,1]^{n\times n}$ with 1's on the diagonal, there is a $\delta \in (0,1)$ such that*

$$\mathbb{P}\left(\sup_{\substack{\boldsymbol{m}\in \mathrm{Plef}_N(\boldsymbol{Q},\beta):\hat{\boldsymbol{Q}}(\boldsymbol{m})>0,\hat{\boldsymbol{Q}}(\boldsymbol{m}) \text{ has eval. } <\delta \\ \text{or } \|\boldsymbol{m}\boldsymbol{m}^\intercal\|_\infty > 1-\delta}} F_{\mathrm{TAP}}(\boldsymbol{m}) \le \sup_{\boldsymbol{m}\in \mathrm{Plef}_N(\boldsymbol{Q},\beta):\hat{\boldsymbol{Q}}(\boldsymbol{m})\ge\delta \boldsymbol{I}, \|\boldsymbol{m}\boldsymbol{m}^\intercal\|_\infty \le 1-\delta} F_{\mathrm{TAP}}(\boldsymbol{m}) \right) \to 1.$$

Proof. Choose $\eta \in (0,1)$ satisfying $\eta \le (\sqrt{2}\|\boldsymbol{\beta}^{\frac12}\boldsymbol{Q}\boldsymbol{\beta}^{\frac12}\|_2)^{-1}$ and let $\tilde{\boldsymbol{Q}} = (1-\eta)\boldsymbol{Q}$, so that $\tilde{\boldsymbol{Q}} \in \mathrm{Plef}_n(\boldsymbol{Q},\beta)$ by (1.5). Let $\tilde{m}_i^k = (\tilde{\boldsymbol{Q}}^{1/2})_{k,i}$ for $k, i = 1,\dots,n$ and $\tilde{m}_i^k = 0$ otherwise and $\tilde{\boldsymbol{m}} = (\tilde{m}^1, \dots, \tilde{m}^n)$, so that $\tilde{\boldsymbol{m}}\tilde{\boldsymbol{m}}^\intercal = \tilde{\boldsymbol{Q}}$ and so $\tilde{\boldsymbol{m}} \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$. Then $\|\tilde{\boldsymbol{m}}\tilde{\boldsymbol{m}}^\intercal\|_\infty = 1-\eta$. Let $\tilde{\delta}$ be the minimum of $\eta > 0$ and the smallest eigenvalue of $\hat{\boldsymbol{Q}}(\tilde{\boldsymbol{m}})$. We then have

$$\begin{aligned} &\frac{1}{N} \sup_{\boldsymbol{m}\in \mathrm{Plef}_N(\boldsymbol{Q},\beta):\hat{\boldsymbol{Q}}(\boldsymbol{m})\ge\tilde{\delta} \boldsymbol{I}, \|\boldsymbol{m}\boldsymbol{m}^\intercal\|_\infty \le 1-\tilde{\delta}} F_{\mathrm{TAP}}(\boldsymbol{m}) \\ &\overset{(1.4),(2.1)}{\ge} -cn \max_i(|\beta_i| + h_i) + \tfrac{\eta}{2}\beta^\intercal \boldsymbol{Q}^{\odot 2}\beta + \tfrac{n}{2}\log\eta + \tfrac12 \log|\boldsymbol{Q}| \end{aligned} \tag{3.43}$$

with probability going to 1.

On the other hand assume that $\delta \in (0,1)$ and $\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$ and $\|\boldsymbol{m}\boldsymbol{m}^\intercal\|_\infty \ge 1-\delta$. Then $(\boldsymbol{m}\boldsymbol{m}^\intercal)_{k,k} \ge 1-\delta$ for some k and because $|\boldsymbol{A}| \le \prod_k A_{k,k}$ for any positive semi-definite $\boldsymbol{A}$ we have $|\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\intercal| \le \delta$ and

$$\frac{1}{N} F_{\mathrm{TAP}}(\boldsymbol{m}) \overset{(1.4),(2.1)}{\le} cn \max_i(|\beta_i| + h_i) + n^2 \max_i \beta_i^2 + \frac12 \log\delta. \tag{3.44}$$

Assume now instead that $\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$ and the smallest eigenvalue of $\hat{\boldsymbol{Q}}(\boldsymbol{m})$ is less than δ. Let $\boldsymbol{S} = \mathrm{Diag}((1-|m^1|^2)^{-1/2}, ..., (1-|m^n|^2)^{-1/2})$. Since $\hat{\boldsymbol{Q}}(\boldsymbol{m}) = \boldsymbol{S}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\intercal)\boldsymbol{S}$ and $\boldsymbol{S}_{ii} > 1$, the smallest eigenvalue of $\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\intercal$ is bounded above by the smallest eigenvalue of $\hat{\boldsymbol{Q}}(\boldsymbol{m})$. Thus if $\boldsymbol{m}$ is such that the smallest eigenvalue of $\hat{\boldsymbol{Q}}(\boldsymbol{m})$ is less than δ, then the smallest eigenvalue of $\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\intercal$ is also less than δ, so with probability tending to 1 all such $\boldsymbol{m}$ satisfy

$$\frac{1}{N} F_{\mathrm{TAP}}(\boldsymbol{m}) \overset{(1.4),(2.1),(3.17)}{\le} cn \max_i(|\beta_i| + h_i) + n^2 \max_i \beta_i^2 + \frac12 (n-1)\log n + \frac12 \log\delta. \tag{3.45}$$

If δ is picked small enough depending on $n, \beta, \boldsymbol{Q}, (h_1, \dots, h_n)$ then the r.h.s of both (3.44) and (3.45) are less than the bottom line of (3.43), and if we also ensure that $\delta \le \tilde{\delta}$ this proves the claim. □

We are now ready to prove the TAP lower bound Proposition 3.1. The idea of the proof will revolve around recentering the $\boldsymbol{\sigma}$ around some vector $\boldsymbol{m}$ and then restricting our integral to a set where the contribution of the external field is negligible, which enables us to use the results of the previous subsection about the free energy without an external field.

Proof of Proposition 3.1. By Lemma 3.12 there is a $\delta \in (0,1)$ such that (3.1) follows once we have shown that

$$\lim_{N\to\infty} \mathbb{P}\left(F_N^\varepsilon(\beta, \boldsymbol{h}, \boldsymbol{Q}) \ge \frac{1}{N} \sup_{\boldsymbol{m}\in \mathrm{Plef}_N(\boldsymbol{Q},\beta):\hat{\boldsymbol{Q}}(\boldsymbol{m})\ge\delta \boldsymbol{I}, \|\boldsymbol{m}\boldsymbol{m}^\intercal\|_\infty \le 1-\delta} F_{\mathrm{TAP}}(\boldsymbol{m}) - c\sqrt{\varepsilon} \right) = 1, \tag{3.46}$$

for all $\varepsilon \in (0, c^{-1})$.

Fix some $m^1, ..., m^n \in \mathbb{R}^N$ with $\|\boldsymbol{m}\boldsymbol{m}^\intercal\|_\infty \le 1-\delta$, $\hat{\boldsymbol{Q}}(\boldsymbol{m}) \ge \delta \boldsymbol{I}$ and $\boldsymbol{m}\boldsymbol{m}^\intercal \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$. By definition (1.1) it follows that for all $\sigma, m \in \mathbb{R}^N$

$$H_N(\sigma) = H_N(m) + \nabla H_N(m)\cdot(\sigma - m) + H_N(\sigma - m), \tag{3.47}$$

so that for all k

$$\beta_k H_N(\sigma^k) + N h^k \cdot \sigma^k = \beta_k H_N(m^k) + N h^k \cdot m^k + N h^{m,k}\cdot(\sigma^k - m^k) + \beta_k H_N(\sigma^k - m^k), \tag{3.48}$$

where

$$h^{m,k} = \frac{\beta_k}{N}\nabla H_N(m^k) + h^k,\ k = 1, \dots, n, \tag{3.49}$$

is the *effective external field*. Using this we obtain

$$Z_N^\varepsilon(\beta, \boldsymbol{h}, \boldsymbol{Q}) = e^{\sum_{k=1}^n \left(\beta_k H_N(m^k) + N h^k \cdot m^k\right)}\ E\left[1_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \left(\beta_k H_N(\sigma^k - m^k) + N h^{m,k}\cdot(\sigma^k - m^k)\right)}\right]. \tag{3.50}$$

Let U be a $2n$-dimensional space whose span includes $m^k, h^{m,k}$ for $k = 1, \dots, n$. We will now bound the expectation on the r.h.s. from below by inserting another indicator 1_A given by

$$A := \left\{\boldsymbol{\sigma} : |P^U(\sigma^k - m^k)| \le \frac{\varepsilon}{4} \text{ for } k = 1, \dots, n\right\}. \tag{3.51}$$

Note that on the event (2.1) for $\boldsymbol{\sigma} \in A$

$$|(\sigma^k - m^k)\cdot m^\ell| \le \frac{\varepsilon}{4} \quad \text{and} \quad |(\sigma^k - m^k)\cdot h^{m,\ell}| \overset{(2.1)}{\le} c\varepsilon, \tag{3.52}$$

for all $k, \ell \in \{1, ..., n\}$. Therefore we obtain

$$E\left[1_{\boldsymbol{Q}_\varepsilon \cap A} e^{\sum_{k=1}^n \left(\beta_k H_N(\sigma^k - m^k) + N h^{m,k}\cdot(\sigma^k - m^k)\right)}\right] \ge e^{-c\varepsilon N} E\left[1_{\boldsymbol{Q}_\varepsilon \cap A} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k - m^k)}\right]. \tag{3.53}$$

Let us define the normalised projection of $\sigma^k - m^k$ onto $U^\perp$ by

$$\hat\sigma^k = \frac{P^{U^\perp}\left(\sigma^k - m^k\right)}{\left|P^{U^\perp}\left(\sigma^k - m^k\right)\right|}, k = 1, \dots, n.$$

For $\boldsymbol{\sigma} \in A$ and k it holds

$$|\sigma - m^k|^2 = |\sigma|^2 - |m^k|^2 - 2(\sigma^k - m^k)\cdot m^k = 1 - |m^k|^2 + \mathcal{O}(\varepsilon) \tag{3.54}$$

and so

$$\left| |P^{U^\perp}\left(\sigma - m^k\right)|^2 - (1 - \left|m^k\right|^2)\right| \le c\varepsilon.$$

Thus using that H_N is 2-homogeneous we obtain that on the event (2.1)

$$H_N\left(\sigma - m^k\right) \ge (1 - \left|m^k\right|^2) H_N\left(\hat\sigma\right) - c\varepsilon,$$

for all k, and

$$E\left[1_{\boldsymbol{Q}_\varepsilon \cap A} e^{\sum_{k=1}^n \beta_k H_N(\sigma^k - m^k)}\right] \ge e^{-c\varepsilon N} E\left[1_{\boldsymbol{Q}_\varepsilon \cap A} e^{\sum_{k=1}^n \beta_k (1 - |m^k|^2) H_N(\hat\sigma^k)}\right]. \tag{3.55}$$

To replace $1_{\hat{\boldsymbol{Q}}_\varepsilon}$ with an indicator that is a function only of the $\hat{\boldsymbol{\sigma}} = (\hat\sigma^1, ..., \hat\sigma^n)$, note that if

$$\left|\hat\sigma^k \cdot \hat\sigma^\ell - \hat Q_{k,\ell}\right| \overset{(3.42)}{=} \left|\hat\sigma^k \cdot \hat\sigma^\ell - \frac{Q_{k,\ell} - m^k \cdot m^\ell}{\sqrt{1 - |m^k|^2}\sqrt{1 - |m^\ell|^2}}\right| \le \frac{\varepsilon}{2}, \quad \forall k, \ell = 1, ..., n$$

then

$$\begin{aligned} |\sigma^k \cdot \sigma^\ell - Q_{k,\ell}| &\le |(\sigma^k - m^k)\cdot(\sigma^\ell - m^\ell) - (Q_{k,\ell} - m^k \cdot m^\ell)| + |m^k \cdot (\sigma^\ell - m^\ell) + m^\ell \cdot (\sigma^k - m^k)| \\ &\overset{(3.52)}{\le} \sqrt{1 - |m^k|^2}\sqrt{1 - |m^\ell|^2}\tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} \le \varepsilon. \end{aligned} \tag{3.56}$$

Thus we obtain

$$\{\boldsymbol{\sigma} : \hat{\boldsymbol{\sigma}} \in \hat{\boldsymbol{Q}}(\boldsymbol{m})_{\frac{\varepsilon}{2}}\} \subset \{\boldsymbol{\sigma} : \boldsymbol{\sigma} \in \boldsymbol{Q}_\varepsilon\}$$

and

$$E\left[1_{\boldsymbol{Q}_\varepsilon \cap A} e^{\sum_{k=1}^n \beta_k (1 - |m^k|^2) H_N(\hat\sigma^k)}\right] \ge E\left[1_A 1_{\{\hat{\boldsymbol{\sigma}} \in \hat{\boldsymbol{Q}}(\boldsymbol{m})_{\frac{\varepsilon}{2}}\}} e^{\sum_{k=1}^n \beta_k (1 - |m^k|^2) H_N(\hat\sigma^k)}\right]. \tag{3.57}$$

Let $\mathcal{A}$ be the σ-algebra generated by $P^U\sigma^k$ for $k=1,\dots,n$. Note that the $\hat\sigma^k$ are independent and uniform on $\mathcal{S}_{N-1}\cap U^\perp$ under $P[\cdot|\mathcal{A}]$. Thus

$$\begin{aligned}
&E\left[1_A 1_{\{\hat{\boldsymbol\sigma}\in\hat{\boldsymbol Q}(\boldsymbol m)_{\frac{\varepsilon}{2}}\}} e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)H_N(\hat\sigma^k)}\right]\\
&=E\left[1_A E\left[1_{\{\hat{\boldsymbol\sigma}\in\hat{\boldsymbol Q}(\boldsymbol m)_{\frac{\varepsilon}{2}}\}} e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)H_N(\hat\sigma^k)}\middle|\mathcal{A}\right]\right]\\
&=E\left[1_A E^{U^\perp}\left[1_{\hat{\boldsymbol Q}(\boldsymbol m)_{\frac{\varepsilon}{2}}} e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)H_N(\sigma)}\right]\right]
\end{aligned}\tag{3.58}$$

Note that by letting

$$(\beta_{\boldsymbol m})_k=\beta_k(1-|m^k|^2), k=1,...,n \quad\text{and}\quad \boldsymbol\beta_m=\operatorname{diag}(\beta_{\boldsymbol m})\in\mathbb{R}^{n\times n},\tag{3.59}$$

we have

$$\|\boldsymbol\beta^{\frac12}\boldsymbol Q\boldsymbol\beta^{\frac12}\|_2=\|\boldsymbol\beta_{\boldsymbol m}^{\frac12}\hat{\boldsymbol Q}(\boldsymbol m)\boldsymbol\beta_{\boldsymbol m}^{\frac12}\|_2,$$

so that

$$\begin{aligned}
\boldsymbol m\boldsymbol m^\mathsf{T}\in\operatorname{Plef}_n(\boldsymbol Q,\beta)\quad\overset{(1.5)}{\Leftrightarrow}\quad\|\boldsymbol\beta^{\frac12}\boldsymbol Q\boldsymbol\beta^{\frac12}\|_2\le\tfrac{1}{\sqrt2}\quad&\Leftrightarrow\quad\|\boldsymbol\beta_{\boldsymbol m}^{\frac12}\hat{\boldsymbol Q}(\boldsymbol m)\boldsymbol\beta_{\boldsymbol m}^{\frac12}\|_2\le\tfrac{1}{\sqrt2}\\
&\overset{(3.20)}{\Leftrightarrow}\quad\beta_{\boldsymbol m}\in\operatorname{HT}(\hat{\boldsymbol Q}(\boldsymbol m)).
\end{aligned}\tag{3.60}$$

Therefore Lemma 3.11 implies that on the event in (3.37) the quantity (3.58) is equal to

$$\exp\left(\frac{\beta_{\boldsymbol m}{}^\mathsf{T}\hat{\boldsymbol Q}(\boldsymbol m)^{\odot2}\beta_{\boldsymbol m}}{2}+\frac12\log|\hat{\boldsymbol Q}(\boldsymbol m)|+O(\sqrt\varepsilon)\right).\tag{3.61}$$

Thus on that event (3.58) is bounded below by

$$E\left[1_A\right]\,\exp\left(N\left(\frac{\beta_{\boldsymbol m}{}^\mathsf{T}\hat{\boldsymbol Q}(\boldsymbol m)^{\odot2}\beta_{\boldsymbol m}}{2}+\frac12\log|\hat{\boldsymbol Q}(\boldsymbol m)|\right)-Nc\sqrt\varepsilon\right).\tag{3.62}$$

We also have

$$E\left[1_A\right]\ge\prod_{k=1}^n\left(c\varepsilon^{2n}(1-|m^k|^2-c\varepsilon)^{\frac{N-2n-2}{2}}\right)\ge\exp\left(\frac N2\sum_{k=1}^n\log(1-|m^k|^2)-c\varepsilon N\right),\tag{3.63}$$

for a constant c depending on δ and $N\ge c$, since $|m^k|^2=(\boldsymbol m\boldsymbol m^\mathsf{T})_{k,k}\le1-\delta$ (see [BK19, (2.9)]). Combining (3.53), (3.55), (3.57), (3.58), (3.62), (3.63) we obtain that

$$\begin{aligned}
&E\left[1_{\boldsymbol Q_\varepsilon\cap A}e^{\sum_{k=1}^n\left(\beta_kH_N(\sigma^k-m^k)+Nh^{m,k}\cdot(\sigma^k-m^k)\right)}\right]\\
&\ge\exp\left(\textstyle\sum_{k=1}^n\frac N2\log\left(1-|m^k|^2\right)+N\left(\frac{\beta_{\boldsymbol m}{}^\mathsf{T}\hat{\boldsymbol Q}(\boldsymbol m)^{\odot2}\beta_{\boldsymbol m}}{2}+\frac12\log|\hat{\boldsymbol Q}(\boldsymbol m)|\right)-Nc\sqrt\varepsilon\right)
\end{aligned}$$

Recall (3.42) and (3.59), which imply that

$$\sum_{k=1}^n\log\left(1-|m^k|^2\right)+\log|\hat{\boldsymbol Q}(\boldsymbol m)|=\log\left(|\hat{\boldsymbol Q}(\boldsymbol m)|\prod_{k=1}^n\left(1-|m^k|^2\right)\right)=\log|\boldsymbol Q-\boldsymbol m\boldsymbol m^\mathsf{T}|,\tag{3.64}$$

and

$$\beta_{\boldsymbol m}{}^\mathsf{T}\hat{\boldsymbol Q}(\boldsymbol m)^{\odot2}\beta_{\boldsymbol m}=\beta^\mathsf{T}\left(\boldsymbol Q-\boldsymbol m\boldsymbol m^\mathsf{T}\right)^{\odot2}\beta.\tag{3.65}$$

This implies that

$$\begin{aligned}
&E\left[1_{\boldsymbol Q_\varepsilon\cap A}e^{\sum_{k=1}^n\left(\beta_kH_N(\sigma^k-m^k)+Nh^{m,k}\cdot(\sigma^k-m^k)\right)}\right]\\
&\ge\exp\left(\tfrac N2\beta^\mathsf{T}(\boldsymbol Q-\boldsymbol m\boldsymbol m^\mathsf{T})^{\odot2}\beta+\tfrac N2\log|\boldsymbol Q-\boldsymbol m\boldsymbol m^\mathsf{T}|-Nc\sqrt\varepsilon\right).
\end{aligned}\tag{3.66}$$

Combining this with (3.50) we obtain that

$$Z_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) \geq e^{\sum_{k=1}^{n}\left(\beta_k H_N(m^k) + N h^k \cdot m^k\right) + \frac{N}{2}\beta^{\mathsf{T}}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})^{\odot 2}\beta + \frac{N}{2}\log|\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}| - Nc\sqrt{\varepsilon}}, \tag{3.67}$$

for all $\boldsymbol{m}$ with $\|\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\| \leq 1 - \delta$, $\hat{\boldsymbol{Q}}(\boldsymbol{m}) \geq \delta \boldsymbol{I}$ and $\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} \in \mathrm{Plef}_N(\boldsymbol{Q}, \beta)$, with probability tending to one. Recalling (1.4), we see that (3.67) is equivalent to

$$Z_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) \geq e^{F_{\mathrm{TAP}}(\boldsymbol{m}) - Nc\sqrt{\varepsilon}}.$$

This proves (3.46), so completes the proof of Proposition 3.1.

□

4. Upper bound

In this section we prove the following upper bound on the free energy.

Proposition 4.1 (The TAP Upper Bound)**.** *Let $\boldsymbol{Q}, \beta, h$ be as in Proposition 3.1. For any $\eta > 0$ there is a $c = c(\beta, \boldsymbol{h}, \boldsymbol{Q}, \eta)$ such that for all $\varepsilon \in (0, c)$*

$$\lim_{N\to\infty} \mathbb{P}\left(F_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) \leq \frac{1}{N} \sup_{\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q}, \beta)} F_{\mathrm{TAP}}(\boldsymbol{m}) + \eta\right) = 1. \tag{4.1}$$

The proof involves constructing, in Subsection 4.2, a low-dimensional subspace of magnetizations $\mathcal{M}_N^n$, with the property that after recentering around $\boldsymbol{m} \in \mathcal{M}_N^n$ the effective external field is again almost completely contained in $\mathcal{M}_N^n$. The set $A(\boldsymbol{m})$ described in Subsection 1.2 is here essentially the hyperplane $\boldsymbol{m} + (\mathcal{M}_N^n)^{\perp}$ intersected with the cartesian product $\mathcal{S}_{N-1}^n$, where $(\mathcal{M}_N^n)^{\perp}$ is the perpendicular space. We write the integral in $F_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q})$ using Fubini's theorem as a double integral first over $\mathcal{M}_N^n$ and then over the perpendicular space $(\mathcal{M}_N^n)^{\perp}$, so that the inner integral is an integral of the recentered Hamiltonian over the sets $A(\boldsymbol{m})$. The latter lacks external field and has a higher effective temperature than the original model (as long as $\boldsymbol{m} \neq 0$). However, as opposed to in the proof of the lower bound, for some $\boldsymbol{m}$ Plefka's condition may not be satisfied, which means that this recentered Hamiltonian is not at high temperature.

Therefore we replace the effective Hamiltonian by an approximation whose partition function is essentially a *low rank* Harish-Chandra-Itzykson-Zuber (HCIZ) integral, and is in some sense always at high temperature. In Subsection 4.1 we estimate such integrals. Using those estimates in Subsection 4.2 we integrate out the inner integral so that the remaining outer integral is now the integral of a modified TAP free energy, in which the Onsager term $\frac{N}{2}\beta^{\mathsf{T}}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})^{\odot 2}\beta$ is replaced by the asymptotics of the HCIZ integral. The integral in $F_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q})$ thus reduces to an integral of the exponential of N times the modified TAP free energy over the low-dimensional space $\mathcal{M}_N^n$, and by the Laplace method the log of the integral turns into the maximizer of the modified TAP free energy over all $\boldsymbol{m}$.

In Subsection 4.3 we then show that if the Hessian of the modified TAP free energy at a critical point is negative semi-definite, as it must be at the maximizer, then $\boldsymbol{m}$ satisfies Plefka's condition. Furthermore we show that the Onsager terms of the modified TAP free energy and the original TAP free energy $F_{\mathrm{TAP}}(\boldsymbol{m})$ are close, so that the upper bound on the free energy $F_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q})$ in terms of the modified TAP free energy implies the needed upper bound in terms of the original TAP free energy.

To implement the above strategy we will have to rely more heavily on random matrix calculations than in Section 3. Define a deterministic version of the Hamiltonian by

$$\tilde{H}_N(\sigma) = N \sum_{i=1}^{N} \theta_{i/N} \sigma_i^2, \tag{4.2}$$

where $\theta_{i/N}$ are the classical locations from (2.4). If $\boldsymbol{U}$ is the change of basis matrix that diagonalizes $\boldsymbol{J}+\boldsymbol{J}^\mathsf{T}$ then by (2.4)

$$\lim_{N\to\infty}\sup_{\sigma\in\mathcal{S}_{N-1}}\Big|H_N(\sigma)-\tilde{H}_N(\boldsymbol{U}\sigma)\Big| = 0 \text{ in probability,} \tag{4.3}$$

so it suffices to prove the upper bound of the free energy for the deterministic Hamiltonian

$$\tilde{F}_N^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) = \frac{1}{N}\log\int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k \tilde{H}_N(\sigma^k)+\tilde{h}^k\cdot\sigma^k}\, d\boldsymbol{\sigma}, \tag{4.4}$$

where $\tilde{h}^k$ is the external field h^k in the diagonalizing basis of the disorder matrix $\boldsymbol{J}+\boldsymbol{J}^\mathsf{T}$. The upper bound will be in terms of a corresponding TAP free energy

$$\tilde{F}_{\mathrm{TAP}}(\boldsymbol{m}) = \sum_{k=1}^n \beta_k\tilde{H}_N(m^k) + N\sum_{k=1}^n \tilde{h}^k\cdot m^k + \frac{N}{2}\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T}| + \frac{N}{2}\beta^\mathsf{T}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T})^{\odot 2}\beta, \tag{4.5}$$

which is simply F_{TAP} from (1.4) with the original Hamiltonian and external field replaced with the deterministic diagonal Hamiltonian and rotated external field.

We further discretize the deterministic Hamiltonian $\tilde{H}_N(\sigma)$. Given $K\geq 2$, we consider K equally spaced numbers in $[-\sqrt{2},\sqrt{2}]$

$$-\sqrt{2} = x_1 < x_2 < \cdots < x_K = \sqrt{2}-\frac{2\sqrt{2}}{K} \quad\text{and}\quad x_{k+1}-x_k = \frac{2\sqrt{2}}{K} \tag{4.6}$$

and the corresponding partition $I_1,\dots,I_K$ of $\{1,\dots,N\}$ given by

$$I_k = \{i : x_k\leq\theta_{i/N}<x_{k+1}\} \quad\text{and}\quad I_K = \{i : x_K\leq\theta_{i/N}\}. \tag{4.7}$$

Consider the "binned" Hamiltonian

$$\tilde{H}_N^K(\sigma) = N\sum_{k=1}^K\sum_{i\in I_k} x_k\sigma_i^2, \tag{4.8}$$

where the eigenvalues $\theta_{i/N}$ are replaced with the left end point of the "bin" it belongs to. We will compute an upper bound for the free energy of the binned Hamiltonian

$$\tilde{F}_{N,K}^\varepsilon(\beta,\boldsymbol{h},\boldsymbol{Q}) = \frac{1}{N}\log\int_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k \tilde{H}_N^K(\sigma^k)+\tilde{h}^k\cdot\sigma^k}\, d\boldsymbol{\sigma}, \tag{4.9}$$

and by taking $K\to\infty$ obtain an upper bound for (4.4).

We first prove an upper bound of the free energy $\tilde{F}_{N,K}$ in the absence of an external field, i.e. $\boldsymbol{h}=0$. For this we use a result from [GH21] about the asymptotics of HCIZ [HC56, IZ80] integral of rank n (or n dimensional spherical integrals in the terminology of [GH21, GM05, HK22]).

4.1. Binned Hamiltonian without external field

In this subsection we compute the free energy of the binned Hamiltonian without external field.

We begin by using [GH21] to compute the free energy with identity constraint $\boldsymbol{Q}=\boldsymbol{I}$, which is essentially an HCIZ integral of rank n. We now recall the limiting formula of [GH21] (which are simplified due to the absence of outlier eigenvalues here). Given any measure ν let G_ν denote its Stieltjes transform defined on $\mathbb{C}\setminus\mathrm{supp}(\nu)$,

$$G_\nu(z) = \int (z-x)^{-1}\, d\nu(x), \tag{4.10}$$

and if λ^* is the rightmost point in the support of ν, we define as in [GH21, Proposition 1] the function

$$J_\nu(z) = \lambda^* z + (v_\nu(z)-\lambda^*)G_\nu(v_\nu(z)) - \log z - \int \log|v_\nu(z)-x|\, d\nu(x) - 1 \text{ for } z>0$$

where

$$v_\nu(z) = \begin{cases} \lambda^* & \text{if } G_\nu(\lambda^*) \le z \\ G_\nu^{-1}(z), & \text{if } G_\nu(\lambda^*) > z. \end{cases} \tag{4.11}$$

Let E_{Haar} denote the probability measure where $(\sigma^1, \dots, \sigma^n)$ are uniformly sampled orthonormal vectors (i.e. the top k rows of a Haar distributed orthogonal random matrix). Also, suppose that $\boldsymbol{X}_N$ is a matrix with empirical spectral distribution ν and suppose that the extremal eigenvalues of $\boldsymbol{X}_N$ converge to the corresponding smallest and largest points in the support of ν. The result [GH21, Proposition 1] implies that

$$\lim_{N\to\infty} \frac{1}{N} \log E_{\text{Haar}} \left[\exp\left(N \sum_{k=1}^n \beta_k (\sigma^k)^\mathsf{T} \boldsymbol{X}_N \sigma^k \right) \right] = \frac{1}{2} \sum_{k=1}^n J_\nu(2\beta_k). \tag{4.12}$$

The Hamiltonian in (1.1) can be written as

$$\sum_{k=1}^n \beta_k \tilde{H}_N^K(\sigma^k) = N \sum_{k=1}^n \beta_k (\sigma^k)^\mathsf{T} \boldsymbol{X}_K \sigma^k, \tag{4.13}$$

for

$$\boldsymbol{X}_K = \operatorname{diag}\Big(\underbrace{x_1, \dots, x_1}_{|I_1|}, \underbrace{x_2, \dots, x_2}_{|I_2|}, \dots, \underbrace{x_K, \dots, x_K}_{|I_K|} \Big), \tag{4.14}$$

and that the limiting spectral distribution of $\boldsymbol{X}_K$ of is equal to

$$\mu_K = \sum_{k=1}^K \rho_k \delta_{x_k} \quad \text{where} \quad \rho_k = \lim_{N\to\infty} \frac{|I_k|}{N} = \int_{x_k}^{x_{k+1}} d\mu_{\text{sc}}(x), \tag{4.15}$$

(recall (4.6)-(4.7)). Thus defining

$$\mathcal{F}_K(\beta) = \frac{1}{2} J_{\mu_K}(2\beta) \text{ for } \beta > 0, \tag{4.16}$$

it follows from (4.12) that

$$\lim_{N\to\infty} \frac{1}{N} \log E_{\text{Haar}} \left[\exp\left(N \sum_{k=1}^n \beta_k (\sigma^k)^\mathsf{T} \boldsymbol{X}_K \sigma^k \right) \right] = \sum_{k=1}^n \mathcal{F}_K(\beta_k). \tag{4.17}$$

The upper bound for the free energy (4.9) of the binned Hamiltonian will be given in terms of a modified TAP free energy where the β_k in the right-hand side of (4.12) are replaced by the eigenvalues $\tilde{\beta}_1(\boldsymbol{m}), \dots, \tilde{\beta}_n(\boldsymbol{m})$ of $\boldsymbol{\beta}^{1/2}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T})\boldsymbol{\beta}^{1/2}$, namely

$$\tilde{F}_{\text{TAP}}^K(\boldsymbol{m}) = \sum_{k=1}^n \left(\beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k \right) + \frac{N}{2} \log |\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T}| + N \sum_{k=1}^n \mathcal{F}_K(\tilde{\beta}_k(\boldsymbol{m})), \tag{4.18}$$

for $\boldsymbol{m} = (m^1, \dots, m^n) \in \mathbb{R}^{n\times N}$, cf. (1.4), (4.5).

For $\nu = \mu_K$ we have $\lambda^* = \sqrt{2}$ and $G_{\mu_K}(\lambda^*) = \infty$ so that

$$v_{\mu_K}(z) = G_{\mu_K}^{-1}(z), \tag{4.19}$$

which is smooth on $(0, \infty)$. Therefore J_{μ_K} is a smooth function on $(0, \infty)$, and so is $\mathcal{F}_K$. Note that

$$\mathcal{F}_K \text{ coincides with the } \mathcal{F}_K \text{ from [BK19, (4.15)]}, \tag{4.20}$$

which can be verified by comparing [BK19, (4.13)-(4.15)] and (4.10)-(4.16), where by (4.19) the $\lambda_K(\beta)$ of [BK19] is the same as $v_{\mu_K}(2\beta)$. The representation of [BK19, Lemma 10], or the case $n = 1$ of (4.22) below, implies that $\mathcal{F}_K(\beta) \to 0$ as $\beta \to 0$, so setting $\mathcal{F}_K(0) = 0$ gives a continuous extension of the function to $[0, \infty)$ (in fact,

$\mathcal{F}_K$ is smooth on $[0,\infty)$, but we refrain from proving or using this fact). Furthermore the representation of [BK19, Lemma 10] (or the case $n=1$ of (4.22)) implies that $\mathcal{F}_K$ is convex, so that

$$\mathcal{F}_K(x) \text{ is Lipschitz on compact subsets of } [0,\infty). \tag{4.21}$$

We now use an approximation argument to derive a formula for the limiting free energy

$$\tilde{F}^{\varepsilon}_{N,K}(\beta,0,\boldsymbol{I}) = \frac{1}{N}\log\int_{\boldsymbol{I}_\varepsilon}\exp\left(\sum_{k=1}^n \beta_k \tilde{H}^K_N(\sigma^k)\right)$$

of the binned Hamiltonian with an identity constraint from (4.17). Since the r.h.s. of (4.22) is smooth we see that for all finite K the partition function of the right-hand side is at high temperature for all β.

Lemma 4.2. *For any $\beta\in[0,\infty)^n$ and $K\geq 1$*

$$\lim_{\varepsilon\to 0}\lim_{N\to\infty}\frac{1}{N}\log\int_{\boldsymbol{I}_\varepsilon}\exp\left(\sum_{k=1}^n \beta_k \tilde{H}^K_N(\sigma^k)\right)d\boldsymbol{\sigma} = \sum_{k=1}^n \mathcal{F}_K(\beta_k), \tag{4.22}$$

where the region of integration $\boldsymbol{I}_\varepsilon$ is the neighborhood of the identity matrix $\boldsymbol{I}$ as defined in (1.2).

Proof. We approximate the integral of the l.h.s. of (4.22) by the expectation in (4.17).

To this end let E denote the measure under which $(\sigma^1,\ldots,\sigma^n)$ are independent uniform unit vectors on the sphere. Define $\tilde{\boldsymbol{\sigma}} = (\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal)^{-\frac12}\boldsymbol{\sigma}$, which exists E-almost surely. By construction the matrix $\tilde{\boldsymbol{\sigma}}$ has orthogonal rows. Furthermore, if $\boldsymbol{O}$ is an arbitrary $N\times N$ orthogonal matrix then $\boldsymbol{\sigma}\overset{d}{=}\boldsymbol{\sigma}\boldsymbol{O}$ under E by rotational symmetry of the uniform measures on the product of spheres. It follows that under E

$$\tilde{\boldsymbol{\sigma}} \overset{d}{=} (\boldsymbol{\sigma}\boldsymbol{O}(\boldsymbol{\sigma}\boldsymbol{O})^\intercal)^{-\frac12}\boldsymbol{\sigma}\boldsymbol{O} = (\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal)^{-\frac12}\boldsymbol{\sigma}\boldsymbol{O} = \tilde{\boldsymbol{\sigma}}\boldsymbol{O},$$

so that the E-law of $\tilde{\boldsymbol{\sigma}}$ is E_{Haar}. Since $(\boldsymbol{\sigma}\boldsymbol{O})(\boldsymbol{\sigma}\boldsymbol{O})^\intercal = \boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal$ for any orthogonal $\boldsymbol{O}$ so that $P(\boldsymbol{\sigma}\in A,\boldsymbol{\sigma}\in\boldsymbol{I}_\varepsilon) = P(\boldsymbol{\sigma}\boldsymbol{O}\in A,\boldsymbol{\sigma}\boldsymbol{O}\in\boldsymbol{I}_\varepsilon) = P(\boldsymbol{\sigma}\boldsymbol{O}\in A,\boldsymbol{\sigma}\in\boldsymbol{I}_\varepsilon)$ for any measurable set A also the $E[\cdot|\boldsymbol{I}_\varepsilon]$-law of $\tilde{\sigma}$ is E_{Haar}. Lemma 3.2 implies that $\lim_{N\to\infty}\frac{1}{N}\log P(\boldsymbol{I}_\varepsilon) = \frac12\log|\boldsymbol{I}| = 0$ for all $\varepsilon>0$. Thus for all $\varepsilon>0$ it follows from (4.17) that

$$\lim_{N\to\infty}\frac{1}{N}\log E\left[1_{\boldsymbol{I}_\varepsilon}\exp\left(N\sum_{k=1}^n\beta_k(\tilde{\sigma}^k)^\intercal\boldsymbol{X}_K\tilde{\sigma}^k\right)\right] = \sum_{k=1}^n\mathcal{F}_K(\beta_k). \tag{4.23}$$

Since $(\boldsymbol{I}+\boldsymbol{A})^{-1/2} = I + O(\|\boldsymbol{A}\|_2)$ if $\|\boldsymbol{A}\|_2\leq\frac12$ we have for $\boldsymbol{\sigma}\in\boldsymbol{I}_\varepsilon$ that

$$\|\tilde{\boldsymbol{\sigma}}-\boldsymbol{\sigma}\|_F = \|\left(\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal\right)^{-1/2}\boldsymbol{\sigma}-\boldsymbol{\sigma}\|_F \leq \sqrt{n}\|\left(\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal\right)^{-1/2}-\boldsymbol{I}\|_2 \leq c\|\boldsymbol{I}-\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal\|_2 \leq c\|\boldsymbol{I}-\boldsymbol{\sigma}\boldsymbol{\sigma}^\intercal\|_\infty \leq c\varepsilon,$$

and so for $\boldsymbol{\sigma}\in\boldsymbol{I}_\varepsilon$

$$\left|\left(\tilde{\sigma}^k\right)^\intercal\boldsymbol{X}_k\tilde{\sigma}^k - \left(\sigma^k\right)^\intercal\boldsymbol{X}_k\sigma^k\right| \leq c\varepsilon.$$

Therefore (4.22) follows from (4.23) and (4.13).

□

We will extend this formula to general positive definite constraints $\boldsymbol{Q}>0$. For this we will need the next lemma which uses a change of variables to estimates integrals with a general constraint in terms of integrals with an identity constraint.

Lemma 4.3. *For any $\delta\in(0,1)$ there is a constant $c=c(\delta)$ such that for any symmetric $\boldsymbol{Q}\in[-1,1]^{n\times n}$ with 1's on the diagonal and $\boldsymbol{Q}>\delta\boldsymbol{I}$, any $\varepsilon>(0,c^{-1})$, and any Lipschitz $f:(\mathcal{S}_{N-1}(1+c^{-1}\varepsilon))^n\to\mathbb{R}$ with Lipschitz constant L we have for $N\geq c(\varepsilon)$*

$$\begin{aligned}
&\tfrac{1}{N}\log\textstyle\int 1_{\boldsymbol{I}_{c^{-1}\varepsilon}}\exp\left(f\left(\boldsymbol{Q}^{1/2}\boldsymbol{\sigma}\right)\right)d\boldsymbol{\sigma} + \tfrac12\log|\boldsymbol{Q}| - c\left(1+\tfrac{L}{N}\right)\varepsilon\\
&\leq \tfrac{1}{N}\log\textstyle\int 1_{\boldsymbol{Q}_\varepsilon}\exp\left(f\left(\boldsymbol{\sigma}\right)\right)d\boldsymbol{\sigma}\\
&\leq \tfrac{1}{N}\log\textstyle\int 1_{\boldsymbol{I}_{c\varepsilon}}\exp\left(f\left(\boldsymbol{Q}^{1/2}\boldsymbol{\sigma}\right)\right)d\boldsymbol{\sigma} + \tfrac12\log|\boldsymbol{Q}| + c(1+\tfrac{L}{N})\varepsilon.
\end{aligned} \tag{4.24}$$

Proof. Fix $\boldsymbol{Q}$ with $\boldsymbol{Q} > \delta \boldsymbol{I}$. Let $\mathbb{Q}$ be the measure under which $u_1, \ldots, u_N$ are independent Gaussian vectors in $\mathbb{R}^n$ with covariance $N\boldsymbol{Q}$, and let $u^k = (u_{1,k}, \ldots, u_{N,k})$ for $k = 1, \ldots, n$, as in the proof of Lemma 3.2. Writing also $\boldsymbol{u} = \left(u^1, \ldots, u^n\right) \in \mathbb{R}^{n \times N}$ we have using the same change of measure as in (3.7) in that lemma that

$$\begin{aligned} &\tfrac{1}{N} \log \mathbb{Q}\left(\exp\left(f\left(\tfrac{u^1}{|u^1|}, \ldots, \tfrac{u^n}{|u^n|}\right)\right) 1_{\boldsymbol{u}\boldsymbol{u}^\intercal \in N\boldsymbol{Q}_{\varepsilon c_1}}\right) + \tfrac{1}{2} \log |\boldsymbol{Q}| - c(\delta)\,\varepsilon \\ &\le \tfrac{1}{N} \log \textstyle\int 1_{\boldsymbol{Q}_\varepsilon} \exp\left(f(\boldsymbol{\sigma})\right) d\boldsymbol{\sigma} \\ &\le \tfrac{1}{N} \log \mathbb{Q}\left(\exp\left(f\left(\tfrac{u^1}{|u^1|}, \ldots, \tfrac{u^n}{|u^n|}\right)\right) 1_{\boldsymbol{u}\boldsymbol{u}^\intercal \in N\boldsymbol{Q}_{\varepsilon c_2}}\right) + \tfrac{1}{2} \log |\boldsymbol{Q}| + c(\delta)\,\varepsilon, \end{aligned} \tag{4.25}$$

for all N. Furthermore letting $\mathbb{E}$ be the law of i.i.d. independent Gaussian vectors we have that the $\mathbb{E}$-law of $\boldsymbol{Q}^{1/2}\boldsymbol{u}$ is the $\mathbb{Q}$-law of $\boldsymbol{u}$, so that for $l = 1, 2$

$$\begin{aligned} &\mathbb{Q}\left(\exp\left(f\left(\tfrac{u^1}{|u^1|}, \ldots, \tfrac{u^n}{|u^n|}\right)\right) 1_{\boldsymbol{u} \in \boldsymbol{Q}_{\varepsilon c_l}}\right) \\ &= \mathbb{E}\left(\exp\left(f\left(\frac{\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^1}{\left|\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^1\right|}, \ldots, \frac{\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^n}{\left|\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^n\right|}\right)\right) 1_{\boldsymbol{Q}^{1/2}\boldsymbol{u}\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^\intercal \in N\boldsymbol{Q}_{\varepsilon c_l}}\right). \end{aligned} \tag{4.26}$$

Writing $a \asymp b$ if there is constant c depending only on n such that $c^{-1} \le \frac{a}{b} \le c$, and writing $a \asymp_\delta b$ if the constant is allowed to depend also on δ, we have

$$\|\boldsymbol{Q}^{1/2}\boldsymbol{u}\boldsymbol{Q}^{1/2}\boldsymbol{u}^\intercal - N\boldsymbol{Q}\|_\infty \asymp \|\boldsymbol{Q}^{1/2}\boldsymbol{u}\boldsymbol{Q}^{1/2}\boldsymbol{u}^\intercal - N\boldsymbol{Q}\|_2 \asymp_\delta \|\boldsymbol{u}\boldsymbol{u}^\intercal - N\boldsymbol{I}\|_2 \asymp \|\boldsymbol{u}\boldsymbol{u}^\intercal - N\boldsymbol{I}\|_\infty. \tag{4.27}$$

Let $\tilde{\sigma}^i = \frac{u^i}{|u^i|}$ so that under $\mathbb{E}$ the $\tilde{\sigma}^1, \ldots, \tilde{\sigma}^n$ are i.i.d. uniform on the unit sphere. The inequalities (4.27) imply that on the event in the indicator of (4.26) we have $\left|\frac{\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^i}{\left|\left(\boldsymbol{Q}^{1/2}\boldsymbol{u}\right)^i\right|} - \boldsymbol{Q}^{1/2}\tilde{\sigma}^i\right| \le c(\delta)\,\varepsilon$, and that the bottom line of (4.26) is bounded below by

$$\mathbb{E}\left(\exp\left(f\left(\boldsymbol{Q}^{1/2}\tilde{\sigma}^1, \ldots, \boldsymbol{Q}^{1/2}\tilde{\sigma}^n\right)\right) 1_{\boldsymbol{u}\boldsymbol{u}^\intercal \in N\boldsymbol{I}_{c(\delta)^{-1}\varepsilon}}\right) e^{-Lc(\delta)\varepsilon}. \tag{4.28}$$

We have

$$\left\{\tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}^\intercal \in \boldsymbol{I}_{c\varepsilon}, \max_{i=1}^n \left|\left|u^i\right| - N\right| \le \varepsilon N\right\} \subset \left\{\boldsymbol{u}\boldsymbol{u}^\intercal \in N\boldsymbol{I}_\varepsilon\right\}, \tag{4.29}$$

for a small enough c and all $\varepsilon \in (0, c)$. Also $\tilde{\boldsymbol{\sigma}}$ is independent of $\left|u^i\right|$, and assuming $\varepsilon \le \delta$ (as we may) we have $\mathbb{P}\left(\max_{i=1}^n \left|\left|u^i\right| - 1\right| \le \varepsilon\right) \to 1$ as $N \to \infty$, and we obtain from (4.29) with $c(\delta)^{-1}\varepsilon$ in place of ε that (4.28) is at least

$$\frac{1}{2}\mathbb{E}\left(\exp\left(f\left(\boldsymbol{Q}^{1/2}\tilde{\sigma}^1, \ldots, \boldsymbol{Q}^{1/2}\tilde{\sigma}^n\right)\right) \tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}^\intercal \in \boldsymbol{I}_{c(\delta)^{-1}\varepsilon}\right) e^{-Lc(\delta)\varepsilon},$$

for $N \ge c(\varepsilon)$. This implies the lower bound of (4.24). The upper bound of (4.24) follows similarly, with the simplification that (4.29) is replaced by the simpler $\left\{\boldsymbol{u}\boldsymbol{u}^\intercal \in N\boldsymbol{I}_\varepsilon\right\} \subset \left\{\tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}^\intercal \in \boldsymbol{I}_{c\varepsilon}\right\}$ for a large enough c and all $\varepsilon \in (0, c^{-1})$, so that the independence of $\tilde{\boldsymbol{\sigma}}$ of $\left|u^i\right|$ need not be invoked. □

To extend Lemma 4.2 to non-identity constraints the next lemma will also be needed. Let $\tilde{\beta}_1, ..., \tilde{\beta}_n$ denote the eigenvalues of $\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}$.

Lemma 4.4. *For any $\delta, C > 0$ and $K \in \mathbb{N}$ there exists a constant $L = c(\delta, C, K)$ such that $(\beta, \boldsymbol{Q}) \to \mathcal{F}_K(\tilde{\beta}_k) + \frac{1}{2}\log|\boldsymbol{Q}|$ is L-Lipschitz continuous for $\boldsymbol{Q} \ge \delta\boldsymbol{I}$ with $\|\boldsymbol{Q}\|_\infty \le 1$ and $|\beta| \le C$.*

Proof. The eigenvalues of $\boldsymbol{Q}$ are Lipschitz continuous in the entries of $\boldsymbol{Q}$ with Lipschitz constant depending only on n. Since $\boldsymbol{Q} \ge \boldsymbol{I}\delta$ implies that all eigenvalues lie in $(\delta, n]$ it follows that $\frac{1}{2}\log|\boldsymbol{Q}|$ is Lipschitz in the entries of $\boldsymbol{Q}$ for such $\boldsymbol{Q}$. Furthermore the $\tilde{\beta}_1, \ldots, \tilde{\beta}_n$ are also the eigenvalues of $\boldsymbol{\beta}\boldsymbol{Q}$, so they are Lipschitz as functions of the entries of $\boldsymbol{Q}$ and β with Lipschitz constant depending on n and C, and they are bounded in terms of C. Since $\mathcal{F}_K$ is Lipschitz on compact intervals (recall (4.21)) the claim follows.

□

We can now compute the limiting free energy

$$\tilde{F}_{N,K}^{\varepsilon}(\beta,0,\boldsymbol{Q})=\frac{1}{N}\log\int_{\boldsymbol{Q}_{\varepsilon}}\exp\left(\sum_{k=1}^{n}\beta_k\tilde{H}_N^K(\sigma^k)\right)d\boldsymbol{\sigma},\tag{4.30}$$

of the binned model without external field and with general constraint $\boldsymbol{Q}$. Similarily to in (4.22) the smoothness of $\mathcal{F}_K$ means that for all finite K the partition function in (4.30) is at high temperature for all β.

Lemma 4.5 (Limiting free energy of the binned model)**.** *For every $\delta>0$ and $C>0$ we have*

$$\lim_{\varepsilon\to0}\lim_{N\to\infty}\sup_{\boldsymbol{Q}\geq\delta\boldsymbol{I}}\sup_{|\beta|\leq C}\left|\frac{1}{N}\log\int_{\boldsymbol{Q}_{\varepsilon}}\exp\left(\sum_{k=1}^{n}\beta_k\tilde{H}_N^K(\sigma^k)\right)d\boldsymbol{\sigma}-\left(\sum_{k=1}^{n}\mathcal{F}_K(\tilde{\beta}_k)+\frac{1}{2}\log|\boldsymbol{Q}|\right)\right|=0,\tag{4.31}$$

where the outermost sup *is over symmetric $\boldsymbol{Q}\in[-1,1]^n$ with 1s on the diagonal and $\tilde{\beta}_1,\ldots,\tilde{\beta}_n$ are the eigenvalues of $\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}$.*

Proof. We use

$$f(\boldsymbol{\sigma})=\sum_{k=1}^{n}\beta_k\tilde{H}_N^K(\sigma^k),$$

in Lemma 4.3. From (4.8) and using $|\beta|\leq C$ this f has Lipschitz constant at most $c(C)N$, and we obtain

$$\begin{aligned}&\tfrac{1}{N}\log\int 1_{\boldsymbol{I}_{c^{-1}\varepsilon}}\exp\left(\textstyle\sum_{k=1}^{n}\beta_k\tilde{H}_N^K((\boldsymbol{Q}^{1/2}\boldsymbol{\sigma})^k)\right)d\sigma+\tfrac{1}{2}\log|\boldsymbol{Q}|-c\varepsilon\\&\leq\tfrac{1}{N}\log\int 1_{\boldsymbol{Q}_{\varepsilon}}\exp\left(\textstyle\sum_{k=1}^{n}\beta_k\tilde{H}_N^K(\sigma^k)\right)\\&\leq\tfrac{1}{N}\log\int 1_{\boldsymbol{I}_{c\varepsilon}}\exp\left(\textstyle\sum_{k=1}^{n}\beta_k\tilde{H}_N^K((\boldsymbol{Q}^{1/2}\boldsymbol{\sigma})^k)\right)d\sigma+\tfrac{1}{2}\log|\boldsymbol{Q}|+c\varepsilon,\end{aligned}\tag{4.32}$$

for any $\boldsymbol{Q}$ as in the statement of the lemma. Next writing $\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2}=\boldsymbol{O}^{\mathsf{T}}\tilde{\boldsymbol{\beta}}\boldsymbol{O}$ for an $n\times n$ orthogonal matrix and $\tilde{\boldsymbol{\beta}}$ the diagonal matrix of eigenvalues of $\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}$ (recall (3.23)) we have using (4.13) that

$$\begin{aligned}\textstyle\sum_{k=1}^{n}\beta_k\tilde{H}_N^K((\boldsymbol{Q}^{1/2}\boldsymbol{\sigma})^k)&=N\mathrm{Tr}\left(\boldsymbol{\beta}(\boldsymbol{Q}^{1/2}\boldsymbol{\sigma})\boldsymbol{X}_K(\boldsymbol{Q}^{1/2}\boldsymbol{\sigma})^{\mathsf{T}}\right)\\&=N\mathrm{Tr}\left(\boldsymbol{Q}^{1/2}\boldsymbol{\beta}\boldsymbol{Q}^{1/2}\boldsymbol{\sigma}\boldsymbol{X}_K\boldsymbol{\sigma}^{\mathsf{T}}\right)\\&=N\mathrm{Tr}\left(\boldsymbol{O}^{\mathsf{T}}\tilde{\boldsymbol{\beta}}\boldsymbol{O}\boldsymbol{\sigma}\boldsymbol{X}_K\boldsymbol{\sigma}^{\mathsf{T}}\right)\\&=N\mathrm{Tr}\left(\tilde{\boldsymbol{\beta}}\boldsymbol{O}\boldsymbol{\sigma}\boldsymbol{X}_K(\boldsymbol{O}\boldsymbol{\sigma})^{\mathsf{T}}\right)\\&=\textstyle\sum_{k=1}^{n}\tilde{\beta}_k\tilde{H}_N^K((\boldsymbol{O}\boldsymbol{\sigma})^k).\end{aligned}$$

We have $\int g(\boldsymbol{O}\boldsymbol{\sigma})d\boldsymbol{\sigma}=\int g(\boldsymbol{\sigma})d\boldsymbol{\sigma}$ for any measurable g by symmetry so we can use Lemma 4.2 to estimate the first and last line of (4.32) we obtain that for any fixed $\boldsymbol{Q}$ and β

$$\lim_{\varepsilon\to0}\lim_{N\to\infty}\left|\frac{1}{N}\log\int 1_{\boldsymbol{Q}_{\varepsilon}}\exp\left(\sum_{k=1}^{n}\beta_k\tilde{H}_N^K(\sigma^k)\right)d\boldsymbol{\sigma}-\left(\sum_{k=1}^{n}\mathcal{F}_K(\tilde{\beta}_k)+\frac{1}{2}\log|\boldsymbol{Q}|\right)\right|=0.\tag{4.33}$$

As in Proposition 3.10 we can then deduce the uniformity in $\boldsymbol{Q}$ and β in (4.31) by making two lattices of finitely many $\boldsymbol{A}^1,...,\boldsymbol{A}^M\in[-1,1]^{n\times n}$ and $b^1,...,b^L\in(0,C]^n$ and then use Lipschitz continuity (see Lemma 4.4). More precisely we can choose two lattices such that for all $\boldsymbol{Q}$ we have

$$|\boldsymbol{Q}-\boldsymbol{A}^i|\leq\varepsilon\quad\text{ and }\quad\boldsymbol{A}^i_{\frac{\varepsilon}{2}}\subset\boldsymbol{Q}_{\varepsilon}\subset\boldsymbol{A}^i_{2\varepsilon}$$

for some $i\in\{1,...,M\}$ (cf. (3.30)-(3.31)), as well as for all β

$$\max_k|\beta_k-b^{j_k}|\leq\varepsilon\quad\text{ and }\quad\left|\sum_{k=1}^{n}\beta_k\tilde{H}_N^K(\sigma^k)-\sum_{k=1}^{n}b^{j_k}\tilde{H}_N^K(\sigma^k)\right|\leq c(C)\varepsilon N$$

for some $j_1,\ldots,j_k\in\{1,...,L\}$ (cf. (3.32)). Then using Lemma 4.4 completes the proof (cf. (3.33)).

□

To recover the free energy $\tilde{F}_N^\varepsilon$ of (4.4) from the binned version $\tilde{F}_{N,K}^\varepsilon$, we will need to send the number of bins $K \to \infty$. The next lemma shows that if β is in the high temperature region $\mathrm{HT}(\boldsymbol{Q})$ (recall (3.20)) the sum $\sum_{k=1}^n \mathcal{F}_K(\tilde{\beta}_k)$ from (4.31) converges to the simple expression that appears in the annealed free energy (recall (3.9), (3.29)).

Lemma 4.6. *It holds that*

$$\lim_{K\to\infty} \sup_{\beta\in\mathrm{HT}(\boldsymbol{Q})} \left| \sum_{k=1}^n \mathcal{F}_K(\tilde{\beta}_k) - \frac{1}{2}\beta^\mathsf{T}\boldsymbol{Q}^{\odot 2}\beta \right| = 0, \tag{4.34}$$

uniformly over symmetric positive definite matrices $\boldsymbol{Q}$, where $\tilde{\beta}_1, \dots, \tilde{\beta}_n$ are the eigenvalues of $\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}$.

Proof. Recall (4.20). By [BK19, Lemma 14 + (4.28)] we get that

$$\lim_{K\to\infty} \sup_{\tilde{\beta}\in[0,\frac{1}{\sqrt{2}}]} \left| \mathcal{F}_K(\tilde{\beta}) - \frac{\tilde{\beta}^2}{2} \right| = 0. \tag{4.35}$$

If $\beta \in \mathrm{HT}(\boldsymbol{Q})$ then $\tilde{\beta}_k \le \frac{1}{\sqrt{2}}$ for all $k \in \{1, \dots, n\}$ by the definition (3.20). Thus by (4.35) we have that

$$\lim_{K\to\infty} \sup_{\beta\in\mathrm{HT}(\boldsymbol{Q})} \left| \sum_{k=1}^n \mathcal{F}_K(\tilde{\beta}_k) - \sum_{k=1}^n \frac{\tilde{\beta}_k^2}{2} \right| = 0.$$

We can now write $\tilde{\beta}$ back in terms of β and $\boldsymbol{Q}$ using

$$\sum_{k=1}^n \tilde{\beta}_k^2 = \mathrm{Tr}\left(\left(\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}\right)^2\right) = \mathrm{Tr}\left(\boldsymbol{Q}\boldsymbol{\beta}\boldsymbol{Q}\boldsymbol{\beta}\right) = \sum_{ij} \left(\boldsymbol{Q}\boldsymbol{\beta}\right)_{ij} \left(\boldsymbol{Q}\boldsymbol{\beta}\right)_{ji} = \beta^\mathsf{T}\boldsymbol{Q}^{\odot 2}\beta.$$

□

4.2. Upper bound in terms of modified TAP free energy

In this subsection, we prove an upper bound of the free energy in the presence of external fields in terms of a modified TAP free energy.

The main idea is to divide each of the n spheres into two parts: A subspace $\mathcal{M}_N$ of dimension much smaller than N, where most of the effect of the external fields is felt, and the complementary space $\mathcal{M}_N^\perp$ which is almost orthogonal to all the external fields (as in [BK19, Section 4]). We write the partition function integral as a double integral over first the lower dimensional $\mathcal{M}_N$ and then the higher dimensional $\mathcal{M}_N$, where the inner integral is the partition function of the recentered the Hamiltonian. The inner integral is essentially a partition function without external field, so it can be estimated using the results of the previous subsection. In this way we obtain an estimate for the partition function where the remaining outer integral is now the integral of N times the exponential of a modified TAP free energy whose Onsager term is an expression involving $\mathcal{F}_K$ rather than $\frac{N}{2}\beta^\mathsf{T}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T})^{\odot 2}\beta$. Since the dimension of the outer integral is much smaller than N we can then estimate it in terms of the maximum of the modified TAP free energy using the Laplace method.

The following lemma constructs the spaces $\mathcal{M}_N$. Recall that the external fields are denoted by $\boldsymbol{h} \in \mathbb{R}^{n\times N}$ and satisfy $|h^k| = h_k$ for each $k \in \{1, ..., n\}$ for fixed values $h_1, ..., h_n \ge 0$, and that the external fields in the diagonalizing basis of the Hamiltonian is denoted by $\tilde{\boldsymbol{h}} = (\tilde{h}^1, ..., \tilde{h}^n)$.

Lemma 4.7. *Let $N \ge 1$. For any $\beta_1, \dots, \beta_k$ and $h^1, \dots, h^n \in \mathbb{R}^N$, there exists a sequence of linear subspaces $\mathcal{M}_1, \mathcal{M}_2, ...$ such that $\mathcal{M}_N \subset \mathbb{R}^N$,*

$$\dim(\mathcal{M}_N) \le nN^{3/4}$$

and $\mathcal{M}_N^n = (\mathcal{M}_N)^n$ is approximately invariant under the map

$$\boldsymbol{m} = (m^1, ..., m^n) \to \left(\beta_1 \tfrac{1}{N}\nabla\tilde{H}_N(m^1) + \tilde{h}^1, \dots, \beta_n \tfrac{1}{N}\nabla\tilde{H}_N(m^n) + \tilde{h}^n\right)$$

in the sense that

$$\lim_{N\to\infty} \sup_{\substack{m\in\mathcal{M}_N^n \\ |m^1|,\dots,|m^n|\le 1}} \max_{k=1,\dots,n} \left| P^{\mathcal{M}_N^\perp}\left(\frac{\beta_k}{N}\nabla\tilde{H}_N(m^k)+\tilde{h}^k\right)\right| = 0. \tag{4.36}$$

Proof. By [BK19, Lemma 17] with $\beta = 1$ there exists for each k a subspace $\mathcal{M}_{N,k} \subset \mathbb{R}^N$ such that

$$\lim_{N\to\infty} \sup_{\substack{m\in\mathcal{M}_{N,k} \\ |m^k|\le 1}} \left| P^{\mathcal{M}_{N,k}^\perp}\left(\frac{1}{N}\nabla\tilde{H}_N(m^k)+\tilde{h}^k\right)\right| = 0. \tag{4.37}$$

Letting $\mathcal{M}_N := \mathcal{M}_{N,1}+\ldots+\mathcal{M}_{N,n}$ we have that $\dim(\mathcal{M}_N) \le nN^{\frac{3}{4}}$, and for any k and $m^k \in \mathcal{M}_N$ with $|m^k| < 1$ one can decompose

$$m^k = v^1 + \dots + v^n \tag{4.38}$$

for some $v^l \in \mathcal{M}_{N,l}$, $|v^k| < 1, l = 1, \dots, n$. Therefore (using that $\nabla\tilde{H}_N(m)$ is linear in m and $\mathcal{M}_{N,l} \subset \mathcal{M}_N$ for all l)

$$\begin{aligned}
&\sup\nolimits_{m^k\in\mathcal{M}_N,|m^k|<1} \left| P^{\mathcal{M}_N^\perp}\left(\tfrac{\beta_k}{N}\nabla\tilde{H}_N(m^k)+\tilde{h}^k\right)\right| \\
&= \sup\nolimits_{\forall l: v^l\in\mathcal{M}_{N,l},|m^l|<1} \left| P^{\mathcal{M}_N^\perp}\left(\tfrac{\beta_k}{N}\textstyle\sum_{l=1}^n \nabla\tilde{H}_N(v^l)+\tilde{h}^k\right)\right| \\
&\le \beta_k \textstyle\sum_{l=1}^n \sup_{v\in\mathcal{M}_{N,l},|v|<1} \left| P^{\mathcal{M}_{N,l}^\perp}\left(\tfrac{1}{N}\nabla\tilde{H}_N(v^l)\right)\right| + \left|P^{\mathcal{M}_{N,k}^\perp}\tilde{h}^k\right| \\
&\le \beta_k \textstyle\sum_{l=1}^n \sup_{v\in\mathcal{M}_{N,l},|v|<1} \left| P^{\mathcal{M}_{N,l}^\perp}\left(\tfrac{1}{N}\nabla\tilde{H}_N(v^l)+\tilde{h}^l\right)\right| + c\,(\beta)\max_{l=1,\dots,n}\left|P^{\mathcal{M}_{N,l}^\perp}\tilde{h}^l\right|
\end{aligned}$$

and thus by (4.37) we get (4.36). ☐

The next lemma shows that in the absence of external fields, the partition function restricted to the complements of the previously constructed subsets satisfies the same approximation as the unrestricted partition function. Recall from the beginning of Subsection 3.2 that E^U denotes the expectation with respect to $\boldsymbol{\sigma} \in \mathcal{S}_{N-1}^n$ conditioned on $\boldsymbol{\sigma} \in U$ for some set U.

Lemma 4.8. *For any $C > 0, K > 0, \delta > 0$*

$$\lim_{\varepsilon\to 0} \limsup_{N\to\infty} \sup_{\boldsymbol{Q}\ge\delta\boldsymbol{I}} \sup_{|\beta|\le C} \left| \frac{1}{N}\log E^{(\mathcal{M}_N^n)^\perp}\left[1_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \beta_k\tilde{H}_N(\sigma^k)}\right] - \sum_{k=1}^n \mathcal{F}_K(\tilde{\beta}_k) - \frac{1}{2}\log|\boldsymbol{Q}| \right| \le \frac{c}{K}$$

where $(\mathcal{M}_N^n)_{N\ge1}$ is the sequence of subspaces from Lemma 4.7 and $\tilde{\beta}_1, \dots, \tilde{\beta}_n$ are the eigenvalues of $\boldsymbol{\beta}^{1/2}\boldsymbol{Q}\boldsymbol{\beta}^{1/2}$.

Proof. Recall $N' = \dim(\mathcal{M}_N^n) \le nN^{\frac{3}{4}}$. Similarly to in the proof of Lemma 3.11, let $w_1, ..., w_N$ be an orthonormal basis of $\mathbb{R}^N$ such that the space $\mathcal{M}_N$ is spanned by the last $N - N'$ of these vectors. Let $\boldsymbol{D}$ be the diagonal matrix with $D_{jj} = N\theta_{j/N}$ so that $\tilde{H}_N(\sigma) = \sigma^\mathsf{T}\boldsymbol{D}\sigma$. Let $\boldsymbol{A}$ be the $(N-N')\times(N-N')$ minor of $\boldsymbol{D}$ when written in the basis $w_1, \dots, w_N$. By the eigenvalue interlacing inequality and (2.7) the eigenvalues $Na_1, ..., Na_{N-N'}$ of $\boldsymbol{A}$ satisfy $Na_j = N\theta_{j/N} + o(1) = (N-N')\theta_{j/(N-N')} + o(N)$. We have

$$\begin{aligned}
&E^{(\mathcal{M}_N^n)^\perp}\left[1_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n \beta_k\tilde{H}_N(\sigma^k)\right)\right] \\
=&E^{N-N'}\left[1_{\boldsymbol{Q}_\varepsilon}\exp\left((N-N')\sum_{k=1}^n\beta_k\sum_{j=1}^{N-N'}\theta_{j/N-N'}(\sigma_j^k)^2\right)\right]e^{o(N)}.
\end{aligned}$$

Also

$$|\tilde{H}_N^K(\sigma^i) - \tilde{H}_N(\sigma^i)| = \left|N\sum_{k=1}^K\sum_{j\in I_k}(x_k - \theta_{j/N})(\sigma_j^i)^2\right| \le \frac{2\sqrt{2}}{K}N\sum_{k=1}^K\sum_{j\in I_k}(\sigma_j^i)^2 = N\frac{2\sqrt{2}}{K},$$

so we get for bounded β

$$\begin{aligned}&E^{N-N'}\left[1_{\boldsymbol{Q}_\varepsilon}\exp\left((N-N')\sum_{k=1}^n\beta_k\sum_{j=1}^{N-N'}\theta_{j/(N-N')}(\sigma_j^k)^2\right)\right]\\ =&E^{N-N'}\left[1_{\boldsymbol{Q}_\varepsilon}\exp\left(\sum_{k=1}^n\beta_k\tilde{H}^K_{N-N'}(\sigma^k)\right)\right]e^{\mathcal{O}(\frac{N}{K})}.\end{aligned}$$

The claim follows from Lemma 4.5. □

The next lemma will be used to show that $\boldsymbol{m}$ with some $|m^k|$ close to 1 have a negligible contribution to the partition function.

Lemma 4.9. *Let $\mathcal{U}_N\subset\mathbb{R}^N$ be a sequence of linear subspaces of dimension $N'=o\left(\frac{N}{\log N}\right)$. For all $\eta\in(0,1)$ it holds that*

$$\limsup_{N\to\infty}\frac{1}{N}\log\int_{\mathcal{S}^n_{N-1}}1_{\left\{\boldsymbol{\sigma}:\ \exists j\in\{1,\dots,n\}:\left|P^{\mathcal{U}_N}(\sigma^j)\right|^2>1-\eta\right\}}d\boldsymbol{\sigma}<\frac{1}{2}\log\eta.$$

Proof. First note that

$$\begin{aligned}\int_{\mathcal{S}^n_{N-1}}1_{\left\{\boldsymbol{\sigma}:\ \exists j\in\{1,\dots,n\}:\left|P^{\mathcal{U}_N}(\sigma^j)\right|^2>1-\eta\right\}}d\boldsymbol{\sigma}\leq&\sum_{j=1}^n\int_{\mathcal{S}^n_{N-1}}1_{\left\{\boldsymbol{\sigma}:\left|P^{\mathcal{U}_N}(\sigma^j)\right|^2>1-\eta\right\}}d\boldsymbol{\sigma}\\ =&\sum_{j=1}^n\int_{\mathcal{S}_{N-1}}1_{\left\{\sigma^j:\left|P^{\mathcal{U}_N}(\sigma^j)\right|^2>1-\eta\right\}}d\sigma^j.\end{aligned}\tag{4.39}$$

By [BK19, (2.9)]

$$\int_{\mathcal{S}_{N-1}}1_{\left\{\sigma^j:\left|P^{\mathcal{U}_N}(\sigma^j)\right|^2>1-\eta\right\}}d\sigma^j=\frac{\Gamma\left(\frac{N}{2}\right)}{\pi^{\frac{N'}{2}}\Gamma\left(\frac{N-N'}{2}\right)}\int_{\mathcal{B}_{N'}}1_{\{m:|m|^2>1-\eta\}}(1-|m|^2)^{\frac{N-N'-2}{2}}dm,\tag{4.40}$$

where dm denotes Lebesgue measure on $\mathcal{B}_{N'}=\{m\in\mathbb{R}^{N'}:|m|<1\}$. Since

$$\frac{1}{N}\log\left(\frac{\Gamma\left(\frac{N}{2}\right)}{\pi^{\frac{N'}{2}}\Gamma\left(\frac{N-N'}{2}\right)}\right)=o(1)$$

and

$$\begin{aligned}\frac{1}{N}\log\int_{\mathcal{B}_{N'}}1_{\{m:|m|^2>1-\eta\}}(1-|m|^2)^{\frac{N-N'-2}{2}}dm&=\frac{1}{2}\log\eta+o(1)+\frac{1}{N}\log\int_{\mathcal{B}_{N'}}dm\\ &=\frac{1}{2}\log\eta+o(1)+\frac{1}{N}\underbrace{\log\left(\frac{\pi^{\frac{N'}{2}}}{\Gamma(\frac{N'}{2}+1)}\right)}_{=o(N)},\end{aligned}$$

the claim follows from (4.39) and (4.40). □

We now prove that the free energy (4.2) of the deterministic Hamiltonian (4.8) is bounded above by the corresponding modified TAP free energy from (4.18).

Proposition 4.10. *For $K\geq2$ there is a $C=C(\beta,\boldsymbol{h},\boldsymbol{Q},K)$, such that for $\varepsilon\in(0,C)$, N large enough and $c=c(\beta)$*

$$\tilde{F}^\varepsilon_N(\beta,\boldsymbol{h},\boldsymbol{Q})\leq\frac{1}{N}\sup_{\boldsymbol{m}:\boldsymbol{m}\boldsymbol{m}^\intercal<\boldsymbol{Q}}\tilde{F}^K_{TAP}(\boldsymbol{m})+\frac{c}{K}.\tag{4.41}$$

Proof. Let $\mathcal{M}_N^n = \mathcal{M}_N \times ... \times \mathcal{M}_N$ be the space from Lemma 4.7 with each of the n components having dimension $N' \leq nN^{\frac{3}{4}}$. For any $\boldsymbol{\sigma} = (\sigma^1, ..., \sigma^n) \in \mathbb{R}^{n\times N}$ let $\boldsymbol{m}$ be the projection onto $\mathcal{M}_N^n$, i.e. $\forall i \in \{1, ..., n\}$, $m^i := P^{\mathcal{M}_N}\sigma^i$ and $\boldsymbol{m} = (m^1, ..., m^n)$.
By recentering $\tilde{H}_N(\sigma^k)$ around the m^k as in (3.50) we get

$$\begin{aligned}&\int_{\boldsymbol{Q}_\varepsilon} \exp\left(\sum_{k=1}^n \left(\beta_k \tilde{H}_N(\sigma^k) + Nh^k \cdot \sigma^k\right)\right) d\boldsymbol{\sigma}\\ &= E\left[1_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \left(\beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k\right)} e^{\sum_{k=1}^n \left(N\left(\frac{\beta_k}{N}\nabla \tilde{H}_N(m^k) + \tilde{h}^k\right)\cdot(\sigma^k - m^k) + \beta_k \tilde{H}_N(\sigma^k - m^k)\right)}\right]. \qquad (4.42)\end{aligned}$$

Since Lemma 4.7 implies that

$$\lim_{N\to\infty} \sup_{m\in\mathcal{M}_N^n} \sup_{\substack{\hat{\boldsymbol{\sigma}}\in(\mathcal{M}_N^n)^\perp \\ |\sigma^i|\leq 1, \forall i \in\{1,...,n\}}} \left|\sum_{k=1}^n \left(\frac{\beta_k}{N}\nabla \tilde{H}_N(m^k) + \tilde{h}^k\right)\cdot(\sigma^k - m^k)\right| = 0,$$

the effective external field vanishes and (4.42) is at most

$$E\left[1_{\boldsymbol{Q}_\varepsilon} e^{\sum_{k=1}^n \left(\beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k\right)} \exp\left(\sum_{k=1}^n \beta_k \tilde{H}_N(\sigma^k - m^k)\right)\right] e^{o(N)}. \qquad (4.43)$$

The expectation equals

$$E\left[e^{\sum_{k=1}^n \left(\beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k\right)} E\left[1_{\boldsymbol{Q}_\varepsilon} \exp\left(\sum_{k=1}^n \beta_k \tilde{H}_N(\sigma^k - m^k)\right)\middle|\boldsymbol{m}\right]\right] \qquad (4.44)$$

where the $E[\cdot|\boldsymbol{m}]$-law of $\boldsymbol{\sigma} - \boldsymbol{m}$ is the uniform distribution on the cartesian product of the n spheres $\mathcal{M}_N^\perp \cap \mathcal{S}_{N-1}(\sqrt{1-|m^k|^2})$ for $k \in \{1, ..., n\}$.

Note that for all $k, \ell \in \{1, ..., n\}$

$$\begin{aligned}&(\sigma^k - m^k)\cdot(\sigma^\ell - m^\ell) - (Q_{k,\ell} - (\boldsymbol{m}\boldsymbol{m}^\mathsf{T})_{k,\ell})\\ =&\sigma^k\cdot\sigma^\ell - Q_{k,\ell} - \underbrace{(\sigma^k - m^k)\cdot m^\ell}_{=0} - \underbrace{(\sigma - m)^\ell\cdot m^k}_{=0} - \underbrace{(m^k\cdot m^\ell - (\boldsymbol{m}\boldsymbol{m}^\mathsf{T})_{k,\ell})}_{=0},\end{aligned}$$

since $m^k = P^{\mathcal{M}_N}\sigma^k \in \mathcal{M}_N$ and $\sigma^k - m^k \in \mathcal{M}_N^\perp$, so

$$\{\boldsymbol{\sigma} : \boldsymbol{\sigma} \in \boldsymbol{Q}_\varepsilon\} = \{\boldsymbol{\sigma} : \boldsymbol{\sigma} - \boldsymbol{m} \in (\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T})_\varepsilon\}. \qquad (4.45)$$

Let $\hat{\sigma}^k = \frac{\sigma^k - m^k}{\sqrt{1-|m^k|^2}}$. Using also that $\tilde{H}_N$ is 2-homogeneous (recall (4.2)) the expression in (4.44) equals

$$E\left[e^{\sum_{k=1}^n \left(\beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k\right)} E\left[1_{\{\boldsymbol{\sigma}-\boldsymbol{m}\in(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T})_\varepsilon\}} e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)\tilde{H}_N(\hat{\sigma}^k)}\middle|\boldsymbol{m}\right]\right]. \qquad (4.46)$$

Let $\eta > 0$ and define $W_j(\eta) = \{\boldsymbol{\sigma} : |m^j|^2 \leq 1 - \eta\}$ and $W(\eta) = \bigcap_{j=1}^n W_j(\eta)$. Using that $N^{-1}\tilde{H}_N(\sigma), \beta_k, |\tilde{h}^k|$ are all bounded and Lemma 4.9 we obtain

$$\begin{aligned}&E\left[1_{W_j(\eta)^c} e^{\sum_{k=1}^n \left(\beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k\right)} E\left[1_{\{\boldsymbol{\sigma}-\boldsymbol{m}\in(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T})_\varepsilon\}} e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)\tilde{H}_N(\hat{\sigma}^k)}\middle|\boldsymbol{m}\right]\right]\\ &\leq e^{cN} E\left[1_{W_j(\eta)^c}\right] \leq e^{N(c+\log\eta)} \leq \exp\left(N\tilde{F}_{\mathrm{TAP}}^K(0)\right),\end{aligned} \qquad (4.47)$$

if η is picked small enough depending on $\boldsymbol{Q}, \beta$, and N is large enough. To conclude (4.41) it thus suffices to bound

$$E\left[1_{W(\eta)} e^{\sum_{k=1}^n \left(\beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k\right)} E\left[1_{\{\boldsymbol{\sigma}-\boldsymbol{m}\in(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T})_\varepsilon\}} e^{\sum_{k=1}^n \beta_k(1-|m^k|^2)\tilde{H}_N(\hat{\sigma}^k)}\middle|\boldsymbol{m}\right]\right].$$

Recall the matrix $\hat{\boldsymbol{Q}}(\boldsymbol{m})$ given by

$$\hat{\boldsymbol{Q}}(\boldsymbol{m})_{ij} = \frac{Q_{ij} - m^i \cdot m^j}{\sqrt{1-|m^i|^2}\sqrt{1-|m^j|^2}}.$$

Let $\varepsilon' = \varepsilon\eta^{-1}$. For $\boldsymbol{\sigma} \in W(\eta)$

$$\{\boldsymbol{\sigma} : \boldsymbol{\sigma} - \boldsymbol{m} \in (\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T})_\varepsilon\} \subset \{\boldsymbol{\sigma} : \hat{\boldsymbol{\sigma}} \in \hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}\}. \tag{4.48}$$

Using this we can bound (4.47) from above by

$$E\left[1_{W(\eta)} e^{\sum_{k=1}^n \beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k} E\left[1_{\{\hat{\boldsymbol{\sigma}} \in \hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}\}} \exp\left(\sum_{k=1}^n \beta_k (1-|m^k|^2) \tilde{H}_N\left(\hat{\sigma}^k\right)\right) \middle| \boldsymbol{m}\right]\right]. \tag{4.49}$$

Because $\hat{\sigma}^k$ is distributed uniformly on $\mathcal{S}^n_{N-1} \cap (\mathcal{M}^n_N)^\perp$ under $E[\cdot|\boldsymbol{m}]$ we can also write this as

$$E\left[1_{W(\eta)} e^{\sum_{k=1}^n \beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k} E^{(\mathcal{M}^n_N)^\perp}\left[1_{\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}} \exp\left(\sum_{k=1}^n \beta_k (1-|m^k|^2) \tilde{H}_N\left(\hat{\sigma}^k\right)\right)\right]\right]. \tag{4.50}$$

Note that (4.50) is bounded from above by

$$\begin{aligned} &E\left[1_{W(\eta)} 1_{\{\boldsymbol{m}: \hat{\boldsymbol{Q}}(\boldsymbol{m}) > \delta \boldsymbol{I}\}} e^{\sum_{k=1}^n \beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k} E^{(\mathcal{M}^n_N)^\perp}\left[1_{\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}} e^{\sum_{k=1}^n \beta_k (1-|m^k|^2) \tilde{H}_N\left(\hat{\sigma}^k\right)}\right]\right] \\ &+ e^{cN} E\left[1_{\{\boldsymbol{m}: \hat{\boldsymbol{Q}}(\boldsymbol{m}) > \delta \boldsymbol{I}\}^c} E^{(\mathcal{M}^n_N)^\perp}\left[1_{\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}}\right]\right], \end{aligned} \tag{4.51}$$

for any $\delta > 0$, where we have crudely bounded all terms in exp by cN to arrive at the second term. Since under $E^{(\mathcal{M}^n_N)^\mathsf{T}}$ the $\hat{\sigma}^k$ are i.i.d. uniformly distributed on a sphere of radius 1 in the subspace $\mathcal{M}_N^\perp$ of dimension $N - N'$ we have by (3.4) (with $N - N'$ in place of N) and (3.17) that $E^{(\mathcal{M}^n_N)^\perp}[1_{\hat{\boldsymbol{Q}}(\boldsymbol{m})_{\varepsilon'}}] \le e^{\frac{1}{2}\log(2\delta n^{n-1})(N-N')}$ for $\varepsilon' \le \delta$, so there is $\delta > 0$ such that the second term of (4.51) is at most $\exp(NF^K_{\text{TAP}}(0))$. It thus suffices to bound the first term of (4.51) to prove (4.41) (cf. (4.47)).

Now we can apply Lemma 4.8 with $(\beta_{\boldsymbol{m}})_k = \beta_k(1-|m^k|^2)$ in place of β_k, $\boldsymbol{\beta}_{\boldsymbol{m}} = \operatorname{diag} \beta_{\boldsymbol{m}} \in \mathbb{R}^{n\times n}$ in place of $\boldsymbol{\beta}$ and $\hat{\boldsymbol{Q}}$ in place of $\boldsymbol{Q}$ to bound the first term of (4.51) by

$$\begin{aligned} E\Bigg[1_{W(\eta)} 1_{\{\boldsymbol{m}: \hat{\boldsymbol{Q}}(\boldsymbol{m}) > \delta \boldsymbol{I}\}} \exp\Bigg(&\sum_{k=1}^n \beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k \\ &+ N\sum_{k=1}^n \mathcal{F}_K(\tilde{\beta}_k(\boldsymbol{m})) + \frac{N}{2}\log|\hat{\boldsymbol{Q}}(\boldsymbol{m})|\Bigg)\Bigg] e^{o(N) + \frac{cN}{K}}, \end{aligned} \tag{4.52}$$

recalling from (4.18) that $\tilde{\beta}_k(\boldsymbol{m})$ are the eigenvalues of the symmetric positive semi-definite matrix

$$\boldsymbol{\beta}^{1/2}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T})\boldsymbol{\beta}^{1/2} = \boldsymbol{\beta}_{\boldsymbol{m}}^{\frac{1}{2}} \hat{\boldsymbol{Q}}(\boldsymbol{m}) \boldsymbol{\beta}_{\boldsymbol{m}}^{\frac{1}{2}}.$$

Since each m^k is a projection onto $\mathcal{M}_N$ we can use [BK19, (2.9)] to write the expectation as

$$\begin{aligned} &\left(\frac{1}{\pi^{\frac{N'}{2}}} \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-N'}{2}\right)}\right)^n \int_{(\mathcal{B}_N(\sqrt{1-\eta}) \cap \mathcal{M}_N)^n} 1_{\{\boldsymbol{m}: \hat{\boldsymbol{Q}}(\boldsymbol{m}) > \delta \boldsymbol{I}\}} \prod_{k=1}^n (1-|m^k|^2)^{\frac{N-N'-2}{2}} \\ &\times \exp\left(\sum_{k=1}^n \left(\beta_k \tilde{H}_N(m^k) + N\tilde{h}^k \cdot m^k\right) + N\sum_{k=1}^n \mathcal{F}_K(\tilde{\beta}_k(\boldsymbol{m})) + \frac{N}{2}\log|\hat{\boldsymbol{Q}}(\boldsymbol{m})|\right) d\boldsymbol{m}, \end{aligned} \tag{4.53}$$

where $\mathcal{B}_N(r)$ denotes the ball of radius n in $\mathbb{R}^N$ and $d\boldsymbol{m}$ is the nN'-dimensional Lebesgue measure on $\mathcal{M}^n_N$. We have

$$\log|\hat{\boldsymbol{Q}}(\boldsymbol{m})| + \sum_{k=1}^n \log(1-|m^k|^2) \overset{(3.64)}{=} \log|\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T}|.$$

Therefore recalling (4.18) we have that (4.53) equals

$$\left(\frac{1}{\pi^{\frac{N'}{2}}}\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-N'}{2}\right)}\right)^n \int_{(\mathcal{B}_N(\sqrt{1-\eta})\cap\mathcal{M}_N)^n} 1_{\{\boldsymbol{m}:\hat{\boldsymbol{Q}}(\boldsymbol{m})>\delta \boldsymbol{I}\}} \exp\left(\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}) - \tfrac{N'+2}{2}\sum_{k=1}^n \log(1-|m^k|^2)\right) d\boldsymbol{m}. \quad (4.54)$$

Since the prefactor in (4.54) is at most $e^{o(N)}$ and $\int_{\mathcal{M}_N^n} d\boldsymbol{m} \le \int_{B_{nN'}(n)} d\boldsymbol{m} = \frac{\pi^{\frac{nN'}{2}}}{\Gamma(\frac{nN'}{2}+1)} n^{nN'} = e^{o(N)}$ the expectation in (4.52) is bounded from above by

$$\exp\left(\sup_{\boldsymbol{m}\in\mathcal{M}_N^n,\hat{\boldsymbol{Q}}(\boldsymbol{m})>\delta\boldsymbol{I}} \tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}) + \frac{cN}{K}\right), \quad (4.55)$$

for N large enough. As $\boldsymbol{D}^\intercal \boldsymbol{A}\boldsymbol{D} > 0$ and $\boldsymbol{D}$ invertible implies that $\boldsymbol{A} > 0$ we have that $\hat{\boldsymbol{Q}} > \delta\boldsymbol{I}$ implies $\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\intercal > 0$, so the claim (4.41) follows. □

4.3. Location of the maximizer

In this subsection we will derive Proposition 4.1 for the free energy in terms of the TAP free energy $F^K_{\mathrm{TAP}}(\boldsymbol{m})$ from the upper bound Proposition 4.10 for the free energy in terms of the modified TAP free energy $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$. To do so we will show that the maximum of $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$, is attained at some $\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$. Similarly to (3.60) we have

$$\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q},\beta) \overset{(3.20)}{\Leftrightarrow} \beta \in \mathrm{HT}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal) \Leftrightarrow \tilde{\beta}(\boldsymbol{m}) \in \mathrm{HT}(\boldsymbol{Q}), \quad (4.56)$$

i.e. m satisfies the Plefka condition if and only if the "effective temperature after recentering" $\tilde{\beta}(\boldsymbol{m})$ lies in the high temperature region $\mathrm{HT}(\boldsymbol{Q})$. Therefore once we have proven that the maximizer of $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$ satisfies $\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q},\beta)$ we will be able to derive the upper bound Proposition 4.1 for the free energy F^ε_N from Proposition 4.10 and Lemma 4.6 by taking the limit $K\to\infty$.

To obtain nice formulas for the derivatives of $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$, we will interpret the terms

$$\sum_{k=1}^n \mathcal{F}_K(\tilde{\beta}_k(\boldsymbol{m})) \quad \text{and} \quad \log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal|$$

of (4.18) as traces of primary matrix functions [Hig08, Chapter 1].

Definition 4.11. *Given a scalar function f and a real symmetric matrix $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^\intercal \in \mathbb{R}^{n\times n}$ we define the primary matrix function $f(\boldsymbol{A})$ associated with f by*

$$f(\boldsymbol{A}) := \boldsymbol{U} f(\boldsymbol{D})\boldsymbol{U}^\intercal \qquad \textit{where} \qquad f(\boldsymbol{D}) = \mathrm{diag}(f(\lambda_1),\ldots,f(\lambda_n)).$$

These matrix valued functions are well-defined if $f(\lambda_i)$ is well-defined for all $i \le n$.

Remark 4.12. *The primary matrix functions of [Hig08, Chapter 1.2] are defined more generally in terms of the Jordan canonical form. However, in this work, we only deal with diagonalizable matrices, so the definition simplifies.*

It follows that

$$\sum_{k=1}^n \mathcal{F}_K(\tilde{\beta}_k(\boldsymbol{m})) = \mathrm{Tr}\big(\mathcal{F}_K(\boldsymbol{\beta}^{\frac12}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal)\boldsymbol{\beta}^{\frac12})\big)$$

where $\mathcal{F}_K(\boldsymbol{A})$ is the primary matrix function associated with $\mathcal{F}_K(x)$, and (for $\boldsymbol{m} < \boldsymbol{Q}$)

$$\log|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal| = \mathrm{Tr}\big(\log(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\intercal)\big),$$

where $\log(\boldsymbol{A})$ is the primary matrix functions associated with $\log(x)$. Replacing the corresponding terms of (4.18) we arrive at the matrix form of $\tilde{F}^K_{\text{TAP}}$

$$\tilde{F}^K_{\text{TAP}}(\boldsymbol{m}) = \sum_{k=1}^{n}\left(\beta_k \tilde{H}_N(m^k) + Nm^k\cdot\tilde{h}^k\right) + N\mathrm{Tr}(\mathcal{F}_K(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T})\boldsymbol{\beta}^{\frac{1}{2}})) + \frac{N}{2}\mathrm{Tr}(\log(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T})). \tag{4.57}$$

We want to study the critical point condition of the maximizers of this function, which will require a formula to differentiate primary matrix functions.

Lemma 4.13. *Let f be a smooth scalar function which is continuously differentiable on its domain, and let $\boldsymbol{A}(\alpha)$ be a smooth map from a subset of $\mathbb{R}$ into the subset of $\mathbb{R}^{n\times n}$ on which $f(\boldsymbol{A})$ is well-defined. Then $f_{ab}(\boldsymbol{A}(\alpha)), a, b = 1, \dots, n$ is continuously differentiable smooth*

$$\partial_\alpha \mathrm{Tr}(f(\boldsymbol{A}(\alpha))) = \mathrm{Tr}(f'(\boldsymbol{A}(\alpha))\,\partial_\alpha\boldsymbol{A}(\alpha)). \tag{4.58}$$

In particular, for positive definite $\boldsymbol{A}(\alpha)$

$$\partial_\alpha \mathrm{Tr}(\log(\boldsymbol{A}(\alpha))) = \mathrm{Tr}(\boldsymbol{A}(\alpha)^{-1}\partial_\alpha\boldsymbol{A}(\alpha)) \tag{4.59}$$

and

$$\partial_\alpha \mathrm{Tr}(\mathcal{F}_K(\boldsymbol{A}(\alpha))) = \mathrm{Tr}(v_{\mu_K}(2\boldsymbol{A}(\alpha))\partial_\alpha\boldsymbol{A}(\alpha)) - \frac{1}{2}\mathrm{Tr}(\boldsymbol{A}(\alpha)^{-1}\partial_\alpha\boldsymbol{A}(\alpha)). \tag{4.60}$$

Proof. By linearity we have

$$\partial_\alpha \mathrm{Tr}(f(\boldsymbol{A}(\alpha))) = \mathrm{Tr}(\partial_\alpha f(\boldsymbol{A}(\alpha))). \tag{4.61}$$

To manipulate the right-hand side we use the concepts of [Hig08, Chapter 3.2]. Let $L(\boldsymbol{A},\boldsymbol{C})$ denote the Fréchet derivative of $f(\boldsymbol{A})$ in the direction $\boldsymbol{C}$ defined in [Hig08, (3.6)]. Then

$$\partial_\alpha f(\boldsymbol{A}(\alpha)) = L(\boldsymbol{A}(\alpha), \partial_\alpha\boldsymbol{A}(\alpha)). \tag{4.62}$$

We write $\boldsymbol{A}(\alpha) = \boldsymbol{U}(\alpha)\boldsymbol{D}(\alpha)\boldsymbol{U}(\alpha)^\mathsf{T}$ in its eigendecomposition where $\boldsymbol{D}(\alpha) = \mathrm{diag}(\lambda_1(\alpha),\dots,\lambda_n(\alpha))$ are the eigenvalues of $\boldsymbol{A}$. Let $\odot$ denote the Hadamard product. By [Hig08, Corollary 3.12 (see also the top of p. 61 and the remark before equation (3.13))] we have

$$L(\boldsymbol{A}(\alpha), \partial_\alpha\boldsymbol{A}(\alpha)) = \boldsymbol{U}(\alpha)(\boldsymbol{\Delta}(\alpha)\odot\boldsymbol{U}(\alpha)^\mathsf{T}\partial_\alpha\boldsymbol{A}(\alpha)\boldsymbol{U}(\alpha))\boldsymbol{U}(\alpha)^\mathsf{T}$$

where $\boldsymbol{\Delta}$ is given by

$$\boldsymbol{\Delta} = \boldsymbol{\Delta}_{f(\boldsymbol{A})} = [\Delta f(\lambda_i,\lambda_j)]_{i,j\le n} \quad\text{and}\quad \Delta f(\lambda,\lambda') = \begin{cases} \frac{f(\lambda)-f(\lambda')}{\lambda-\lambda'} & \lambda\neq\lambda' \\ f'(\lambda) & \lambda=\lambda'. \end{cases} \tag{4.63}$$

Thus using the invariance of the trace under cyclic permutations

$$\mathrm{Tr}(\partial_\alpha f(\boldsymbol{A}(\alpha))) = \mathrm{Tr}(\boldsymbol{U}(\boldsymbol{\Delta}\odot\boldsymbol{U}^\mathsf{T}\partial_\alpha\boldsymbol{A}\boldsymbol{U})\boldsymbol{U}^\mathsf{T}) = \mathrm{Tr}(\boldsymbol{I}(\boldsymbol{\Delta}\odot\boldsymbol{U}^\mathsf{T}\partial_\alpha\boldsymbol{A}\boldsymbol{U})) = \mathrm{Tr}(\boldsymbol{U}(\boldsymbol{I}\odot\boldsymbol{\Delta})\boldsymbol{U}^\mathsf{T}\partial_\alpha\boldsymbol{A}),$$

where the last inequality follows since for any symmetric matrices $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$

$$\mathrm{Tr}(\boldsymbol{A}(\boldsymbol{B}\odot\boldsymbol{C})) = \mathrm{Tr}((\boldsymbol{A}\odot\boldsymbol{B})\boldsymbol{C}))$$

The claim then follows since $\boldsymbol{I}\odot\boldsymbol{\Delta} = \mathrm{diag}(f'(\lambda_1),\dots,f'(\lambda_n))$ so

$$\boldsymbol{U}(\boldsymbol{I}\odot\Delta)\boldsymbol{U}^\mathsf{T} = f'(\boldsymbol{A}).$$

This proves (4.58). From (4.58) we now derive (4.59)-(4.60). Recall that both the scalar functions $\log(x)$ and $\mathcal{F}_K(x)$ are smooth on $(0,\infty)$. To prove the first formula, we have $\frac{d}{dx}\log(x) = \frac{1}{x}$, so (4.58) implies

$$\partial_\alpha \mathrm{Tr}(\log(\boldsymbol{A}(\alpha))) = \mathrm{Tr}(\boldsymbol{A}(\alpha)^{-1}\partial_\alpha \boldsymbol{A}(\alpha)),$$

where $\boldsymbol{A}^{-1}$ is the primary matrix function arising from $f(x) = x^{-1}$ applied to $\boldsymbol{A}$, which coincides with the usual matrix inverse of $\boldsymbol{A}$.

By [BK19, Lemma 12] or [GM05, Theorem 6] it holds that

$$\mathcal{F}_K'(\beta) = v_{\mu_K}(2\beta) - \frac{1}{2\beta} \text{ for all } z > 0, \tag{4.64}$$

so (4.58) implies

$$\partial_\alpha \mathrm{Tr}(\mathcal{F}_K(\boldsymbol{A}(\alpha))) = \mathrm{Tr}(v_{\mu_K}(2\boldsymbol{A}(\alpha))\partial_\alpha \boldsymbol{A}(\alpha)) - \frac{1}{2}\mathrm{Tr}(\boldsymbol{A}(\alpha)^{-1}\partial_\alpha \boldsymbol{A}(\alpha)),$$

where again we can interpret $\boldsymbol{A}^{-1}$ arising from the primary matrix function x^{-1} as the usual inverse. □

We now study the maximizers of $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$ defined in (4.57). First note that the set

$$\{\boldsymbol{m} : \boldsymbol{m}\boldsymbol{m}^\mathsf{T} < \boldsymbol{Q}\} \tag{4.65}$$

is an open set. Also because $F^K_{\mathrm{TAP}}(\boldsymbol{m})$ diverges to $-\infty$ as $|\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\mathsf{T}| \to 0$ the global maximum lies in (4.65). We vectorize the matrix $\boldsymbol{m} = (m_1^1, \dots, m_1^n, \dots, m_N^1, \dots, m_N^n) \in \mathbb{R}^{Nn}$ and treat $\tilde{F}^K_{\mathrm{TAP}}$ as a function from $\mathbb{R}^{Nn} \mapsto \mathbb{R}$. With this vectorization the gradient of $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$ is a vector in $\mathbb{R}^{Nn}$ and its Hessian is an $Nn \times Nn$ block matrix which consists of $N \times N$ blocks of size $n \times n$. That is, for any sufficiently regular function $f : \mathbb{R}^{Nn} \to \mathbb{R}$,

$$\nabla f(\boldsymbol{m}) = \big(\partial_{m_1^1} f(\boldsymbol{m}), \cdots, \partial_{m_1^n} f(\boldsymbol{m}), \cdots, \partial_{m_N^1} f(\boldsymbol{m}), \cdots, \partial_{m_N^n} f(\boldsymbol{m})\big)^\mathsf{T} \in \mathbb{R}^{Nn}$$

and

$$\nabla^2 f(\boldsymbol{m}) = \begin{bmatrix} \boldsymbol{f}_{1,1} & \cdots & \boldsymbol{f}_{1,N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{f}_{N,1} & \cdots & \boldsymbol{f}_{N,N} \end{bmatrix} \in \mathbb{R}^{Nn \times Nn}, \ \boldsymbol{f}_{i,j} = \begin{bmatrix} \partial_{m_i^1}\partial_{m_j^1} f(\boldsymbol{m}) & \cdots & \partial_{m_i^1}\partial_{m_j^n} f(\boldsymbol{m}) \\ \vdots & \ddots & \vdots \\ \partial_{m_i^n}\partial_{m_j^1} f(\boldsymbol{m}) & \cdots & \partial_{m_i^n}\partial_{m_j^n} f(\boldsymbol{m}) \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Since $\tilde{F}^K_{\mathrm{TAP}}$ is smooth its local maximizers $\boldsymbol{m}^*$ satisfy

$$\nabla \tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}^*) = 0 \quad \text{and} \quad \nabla^2 \tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}^*) \leq 0.$$

Remark 4.14. *Since we formally only proved that $\mathcal{F}_K$ is smooth on $(0,\infty)$ and not on $[0,\infty)$ we can strictly speaking only claim that the term of $\tilde{F}^K_{\mathrm{TAP}}$ involving $\mathcal{F}_K$ (and hence $\tilde{F}^K_{\mathrm{TAP}}$ itself) is smooth when all entries of β are positive, so that $\boldsymbol{\beta}^{1/2}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T})\boldsymbol{\beta}^{1/2}$ is positive definite. In the proofs below we assume that β has positive entries (which also simplifies the arguments) and later remove the assumption by approximation. With additional effort one could prove that $\mathcal{F}_K$ is in fact smooth on $[0,\infty)$ and extend all the arguments below to cover β with zero components, but we refrain from this.*

The part

$$f(\boldsymbol{m}) = \sum_{\ell=1}^n \beta_\ell \tilde{H}_N(m^\ell),$$

of $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$ that depends on the Hamiltonian has a simple Hessian given by the $Nn \times Nn$ matrix

$$\nabla^2 f(\boldsymbol{m}) = 2N\mathrm{diag}(\theta_1\boldsymbol{\beta}, \theta_{\frac{N-1}{N}}\boldsymbol{\beta}, \dots, \theta_{\frac{1}{N}}\boldsymbol{\beta}) = 2N \begin{bmatrix} \theta_1\boldsymbol{\beta} & \cdots & \boldsymbol{0}_n \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}_n & \cdots & \theta_{\frac{1}{N}}\boldsymbol{\beta} \end{bmatrix} \tag{4.66}$$

(recall (4.6)-(4.8) and (4.57)).

The other part of $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$ is $Ng(\boldsymbol{m}\boldsymbol{m}^\top)$ for

$$g(\boldsymbol{A}) = \mathrm{Tr}(\mathcal{F}_K(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{A})\boldsymbol{\beta}^{\frac{1}{2}})) + \frac{1}{2}\mathrm{Tr}(\log(\boldsymbol{Q}-\boldsymbol{A})).$$

Its Hessian is given by the next lemma.

Lemma 4.15. *Assume that $\beta_k > 0$ for $k = 1,\dots,n$. Then for $\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\top > 0$, the Hessian $\nabla^2 g(\boldsymbol{m}\boldsymbol{m}^\top)$ is the $Nm \times Nm$ matrix*

$$-2\begin{bmatrix} \boldsymbol{\beta}^{\frac{1}{2}} v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}} & \cdots & \boldsymbol{0}_n \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}_n & \cdots & \boldsymbol{\beta}^{\frac{1}{2}} v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}} \end{bmatrix} + \boldsymbol{L}, \tag{4.67}$$

where

$$\boldsymbol{Q}_{\boldsymbol{m}} := \boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\top)\boldsymbol{\beta}^{\frac{1}{2}}, \tag{4.68}$$

and $\boldsymbol{L}$ is a matrix of rank at most n^4.

Proof. We have

$$\frac{1}{2}\mathrm{Tr}(\log(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\top)) = \frac{1}{2}\Big(\mathrm{Tr}(\log(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\top)\boldsymbol{\beta}^{\frac{1}{2}})) - \mathrm{Tr}(\log(\boldsymbol{\beta}))\Big).$$

By Lemma 4.13,

$$\partial_{m_i^\ell}\frac{1}{2}\mathrm{Tr}\Big(\log(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\top)\boldsymbol{\beta}^{\frac{1}{2}}\Big) = -\frac{1}{2}\mathrm{Tr}(\boldsymbol{Q}_{\boldsymbol{m}}^{-1}\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\top\boldsymbol{\beta}^{\frac{1}{2}}),$$

and

$$\partial_{m_i^\ell}\mathrm{Tr}\Big(\mathcal{F}_K(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q}-\boldsymbol{m}\boldsymbol{m}^\top)\boldsymbol{\beta}^{\frac{1}{2}})\Big) = -\mathrm{Tr}(v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\top\boldsymbol{\beta}^{\frac{1}{2}}) + \frac{1}{2}\mathrm{Tr}(\boldsymbol{Q}_{\boldsymbol{m}}^{-1}\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\top\boldsymbol{\beta}^{\frac{1}{2}}).$$

Thus the first derivatives of $g(\boldsymbol{m})$ equal

$$\partial_{m_i^\ell} g(\boldsymbol{m}\boldsymbol{m}^\top) = -\mathrm{Tr}(v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\top\boldsymbol{\beta}^{\frac{1}{2}}).$$

To obtain the second derivatives let $h : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ be given by

$$h(\boldsymbol{A}) = v_{\mu_K}(2\boldsymbol{\beta}^{1/2}(\boldsymbol{Q}-\boldsymbol{A})\boldsymbol{\beta}^{1/2}).$$

By the product and chain rules

$$\begin{aligned}\partial_{m_j^{\ell'}}\mathrm{Tr}(h(\boldsymbol{m}\boldsymbol{m}^\top)\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\top\boldsymbol{\beta}^{\frac{1}{2}}) \quad = \quad & \textstyle\sum_{ab} h_{ab}(\boldsymbol{m}\boldsymbol{m}^\top)\partial_{m_j^{\ell'}}(\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\top\boldsymbol{\beta}^{\frac{1}{2}})_{ab}) + \\ & \textstyle\sum_{abcd}\partial_{\boldsymbol{A}_{cd}}h_{ab}(\boldsymbol{m}\boldsymbol{m}^\top)\partial_{m_j^{\ell'}}(\boldsymbol{m}\boldsymbol{m}^\top)_{cd}(\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\top\boldsymbol{\beta}^{\frac{1}{2}})_{ab}).\end{aligned} \tag{4.69}$$

Therefore

$$\nabla^2 g(\boldsymbol{m}\boldsymbol{m}^\top) = \boldsymbol{W} + \boldsymbol{L}, \tag{4.70}$$

where

$$\boldsymbol{W}_{((i,\ell),(j,\ell'))} = -\sum_{ab} h_{ab}(\boldsymbol{m}\boldsymbol{m}^\top)\partial_{m_j^{\ell'}}(\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\top\boldsymbol{\beta}^{\frac{1}{2}})_{ab}),$$

and

$$\boldsymbol{L}_{((i,\ell),(j,\ell'))} = -\sum_{abcd}\partial_{\boldsymbol{A}_{cd}}h_{ab}(\boldsymbol{m}\boldsymbol{m}^\top)\partial_{m_j^{\ell'}}(\boldsymbol{m}\boldsymbol{m}^\top)_{cd}(\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\boldsymbol{m}\boldsymbol{m}^\top\boldsymbol{\beta}^{\frac{1}{2}})_{ab}). \tag{4.71}$$

We have

$$\boldsymbol{W}_{((i,\ell),(j,\ell'))} = -\mathrm{Tr}(v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}}\partial_{m_i^\ell}\partial_{m_j^{\ell'}}\boldsymbol{m}\boldsymbol{m}^\top\boldsymbol{\beta}^{\frac{1}{2}}).$$

Also

$$\partial_{m_j^{\ell'}} \boldsymbol{m}\boldsymbol{m}^\mathsf{T} = \begin{bmatrix} 0 & \cdots & m_j^1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ m_j^1 & \cdots & 2m_j^{\ell'} & \cdots & m_j^n \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & m_j^n & \cdots & 0 \end{bmatrix}$$

where only the ℓ'-th row and column is non-zero, and

$$\partial_{m_i^\ell}\partial_{m_j^{\ell'}} \boldsymbol{m}\boldsymbol{m}^\mathsf{T} = (\delta_{i=j}(\delta_{(\ell,\ell')=(a,b)} + \delta_{(\ell,\ell')=(b,a)}))_{a,b\le n} \tag{4.72}$$

which is a zero matrix if $i \neq j$ and if $i = j$ it is a zero matrix except for the entries (ℓ, ℓ') and (ℓ', ℓ) which takes values 1 if $\ell \neq \ell'$ (on the off-diagonal) and value 2 if $\ell = \ell'$ (on the diagonal). Using this and the symmetry of the matrix $v_{\mu_K}(2\boldsymbol{Q_m})$ we obtain

$$\boldsymbol{W}_{((i,\ell),(j,\ell'))} = -\delta_{i=j} 2v_{\mu_K}(2\boldsymbol{Q_m})_{\ell,\ell'}\beta_\ell^{1/2}\beta_{\ell'}^{1/2} = -2\delta_{i=j}(\boldsymbol{\beta}^{\frac12} v_{\mu_K}(2\boldsymbol{Q_m})\boldsymbol{\beta}^{\frac12})_{\ell,\ell'}.$$

This gives the first term in (4.67).

As for $\boldsymbol{L}$, we can write its entries as

$$\boldsymbol{L}_{((i,\ell),(j,\ell'))} = -\left(\sum_{a,b,c,d=1}^n d_{a,b,c,d} v^{c,d}\left(w^{a,b}\right)^\mathsf{T}\right)_{((i,\ell),(j,\ell'))}$$

where

$$d = \partial_{\boldsymbol{A}_{cd}} h_{ab}(\boldsymbol{m}\boldsymbol{m}^\mathsf{T}),$$

and $v^{c,d} \in \mathbb{R}^{Nn}$ is given by

$$v_{j,\ell'}^{cd} = \left(\partial_{m_j^{\ell'}} \boldsymbol{m}\boldsymbol{m}^\mathsf{T}\right)_{cd},$$

and $w^{a,b} \in \mathbb{R}^{Nn}$ by

$$w_{i,\ell}^{ab} = (\boldsymbol{\beta}^{\frac12} \partial_{m_i^\ell} \boldsymbol{m}\boldsymbol{m}^\mathsf{T} \boldsymbol{\beta}^{\frac12})_{ab},$$

so that $v^{c,d}\left(w^{a,b}\right)^\mathsf{T}$ is an $Nn \times Nn$ matrix. Thus $\boldsymbol{L}$ is the sum of n^4 terms of rank at most 1, so it has rank at most n^4. □

The remainder of the proof of Proposition 4.1 involves a slightly stronger version of Plefka's condition given by,

$$\mathrm{Plef}_N^\delta(\boldsymbol{Q}, \beta) = \left\{\boldsymbol{m} \in \mathbb{R}^{n\times N} : \boldsymbol{m}\boldsymbol{m}^\mathsf{T} < \boldsymbol{Q}, \|\boldsymbol{\beta}^{\frac12}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T})\boldsymbol{\beta}^{\frac12}\|_2 \le \tfrac{1}{\sqrt{2}} - \delta\right\}.$$

Note that

$$\mathrm{Plef}_N^\delta(\boldsymbol{Q}, \beta) \subset \mathrm{Plef}_N^0(\boldsymbol{Q}, \beta) = \mathrm{Plef}_N(\boldsymbol{Q}, \beta) \text{ for all } \delta \ge 0, \boldsymbol{Q}, \beta. \tag{4.73}$$

This stronger Plefka condition is a device to allow the derivation of the upper bound for all β from an upper bound for β with only non-zero entries using continuity in the proof of Proposition 4.1 below. The next lemma is a slight strengthening of [BK19, Lemma 13], and will be used below to prove that any maximizer of $\tilde{F}_{\mathrm{TAP}}^K(\boldsymbol{m})$ must satisfy the stronger Plefka condition.

Lemma 4.16. *For all $K \ge 2$ there is an $\varepsilon \in (0, \frac{2\sqrt{2}}{K})$ and an $\delta_K > 0$ such that*

$$v_{\mu_K}(\beta) \ge \sqrt{2} - \varepsilon \;\Rightarrow\; \beta \le \frac{1}{\sqrt{2}} - \delta_K$$

Proof. We may set $\varepsilon = \sqrt{2} - v_{\mu_K}(\frac{1}{\sqrt{2}} - \delta_K)$ since

$$v_{\mu_K}(\beta) \geq v_{\mu_K}\left(\tfrac{1}{\sqrt{2}} - \delta_K\right) \Rightarrow \beta \leq \tfrac{1}{\sqrt{2}} - \delta_K,$$

and

$$x_K < v_{\mu_K}\left(\tfrac{1}{\sqrt{2}} - \delta_K\right) < \sqrt{2}, \tag{4.74}$$

where the second inequality follows for some $\delta_K > 0$ small enough, because $x_K < v_{\mu_K}(\frac{1}{\sqrt{2}}) < \sqrt{2}$ by [BK19, Lemma 13] and v_{μ_K} is continuous. □

We now show that all maximizers of $\tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m})$ must satisfy the stronger Plefka condition.

Lemma 4.17 (Critical point condition)**.** *Assume that $\beta_k > 0$ for $k = 1, \ldots, n$. Let $K \geq 1$. There exists a constant $c(K)$ such that if $N \geq c(K)$ then*

$$\boldsymbol{m}\boldsymbol{m}^\mathsf{T} < \boldsymbol{Q} \text{ and } \nabla^2 \tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}) \leq \boldsymbol{0} \implies \boldsymbol{m} \in \mathrm{Plef}^{\delta_K}_N(\boldsymbol{Q}, \beta). \tag{4.75}$$

for δ_K as in Lemma 4.16.

Proof. By (4.66) and Lemma 4.15 we have

$$\nabla^2 \tilde{F}^K_{\mathrm{TAP}}(\boldsymbol{m}) = N\left(\boldsymbol{A} + \boldsymbol{L}\right) \leq \boldsymbol{0}, \tag{4.76}$$

for

$$\boldsymbol{A} = 2\begin{bmatrix} \theta_1 \boldsymbol{\beta} & \cdots & \boldsymbol{0}_n \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}_n & \cdots & \theta_{\frac{1}{N}} \boldsymbol{\beta} \end{bmatrix} - 2\begin{bmatrix} \boldsymbol{\beta}^{\frac{1}{2}} v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}} & \cdots & \boldsymbol{0}_n \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}_n & \cdots & \boldsymbol{\beta}^{\frac{1}{2}} v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}} \end{bmatrix},$$

and $\boldsymbol{L}$ has rank at most n^4 (recall that $\boldsymbol{Q}_{\boldsymbol{m}} := \boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^\mathsf{T})\boldsymbol{\beta}^{\frac{1}{2}}$).

Since $\boldsymbol{A}$ is block diagonal its eigenvalues are the eigenvalues of its blocks. By Weyl's inequality [HJ13, Theorem 4.3.1] all but $n^4 + 1$ of the eigenvalues of the matrix $\boldsymbol{A}$ are bounded above by the largest eigenvalue of the entire Hessian $\nabla^2 \tilde{F}^K_{\mathrm{TAP}}$. This means that there is a block among the last $n^4 + 2 \leq 2n^4$ that has all eigenvalues bounded by the largest eigenvalue of $\nabla^2 \tilde{F}^K_{\mathrm{TAP}}$.

Thus if $\nabla^2 \tilde{F}^K_{\mathrm{TAP}} \leq \boldsymbol{0}$ then

$$2\theta_{1-\frac{2n^4}{N}} \boldsymbol{\beta} - 2\boldsymbol{\beta}^{\frac{1}{2}} v_{\mu_K}(2\boldsymbol{Q}_{\boldsymbol{m}})\boldsymbol{\beta}^{\frac{1}{2}} \leq \boldsymbol{0}. \tag{4.77}$$

If $\boldsymbol{\beta}^{1/2}\boldsymbol{B}\boldsymbol{\beta}^{1/2} \leq \boldsymbol{0}$ for a matrix $\boldsymbol{B}$ then $\boldsymbol{B} \leq \boldsymbol{0}$, since we have assumed that $\boldsymbol{\beta}$ is diagonal with positive entries on the diagonal. Therefore (4.77) implies that

$$\theta_{1-\frac{2n^4}{N}} \leq v(\tilde{\beta}_i(\boldsymbol{m})) \qquad \forall i \leq n.$$

The properties of $v_{\mu_K}(\cdot)$ in Lemma 4.16 imply that there exists a $C(K)$ and $\delta_K > 0$ such that for all $\varepsilon \leq C(K)$,

$$v_{\mu_K}(2\beta) \geq \sqrt{2} - \varepsilon \implies \beta \leq \frac{1}{\sqrt{2}} - \delta_K.$$

Since $\theta_{1-\frac{2n^4}{N}} = \sqrt{2} + o_N(1)$, it follows that for N sufficiently large depending on K,

$$\tilde{\beta}_i(\boldsymbol{m}) \leq \frac{1}{\sqrt{2}} - \delta_K \quad \forall i \leq n$$

(recall that $\tilde{\beta}_i(\boldsymbol{m})$ are the eigenvalues of $\boldsymbol{Q}_m$, defined in (4.68)) which implies that $\boldsymbol{m} \in \mathrm{Plef}_{K,N}(\boldsymbol{Q}, \beta) \subseteq \mathrm{Plef}_N(\boldsymbol{Q}, \beta)$.

□

To conclude, we give the proof of Proposition 4.1.

Proof of Proposition 4.1. We first assume that $\beta \in (0,\infty)^n$. By (4.3), it suffices to study the free energy of the deterministic diagonalized Hamiltonian $\tilde{H}_N(\boldsymbol{\sigma})$. Starting from the upper bound Proposition 4.10 we have that for any $K \geq 2$ and $0 < \varepsilon \leq C(\beta, \boldsymbol{h}, \boldsymbol{Q}, K)$ as well as $N \geq c(\varepsilon, K)$ that

$$\tilde{F}_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) \leq \frac{1}{N} \sup_{\boldsymbol{m}: \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} < \boldsymbol{Q}} \tilde{F}_{\mathrm{TAP}}^K(\boldsymbol{m}) + \frac{c(\beta)}{K}.$$

Recall from below (4.65) that $\tilde{F}_{\mathrm{TAP}}^K$ has a global maximizer in the set (4.65). This maximizer must satisfy

$$\nabla^2 \tilde{F}_{\mathrm{TAP}}^K(\boldsymbol{m}) \leq 0.$$

Thus it follows by Lemma 4.17 that

$$\tilde{F}_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) \leq \frac{1}{N} \sup_{\boldsymbol{m} \in \mathrm{Plef}_N^{\delta_K}(\boldsymbol{Q}, \beta)} \tilde{F}_{\mathrm{TAP}}^K(\boldsymbol{m}) + \frac{c(\beta)}{K}. \tag{4.78}$$

By the definition (1.5), the equivalence in (4.56) and the uniform bound on $\mathcal{F}_K$ from Lemma 4.6 implies that for all $m \in \mathrm{Plef}_N(\boldsymbol{Q}, \beta)$

$$\left| \mathcal{F}_K(\boldsymbol{\beta}^{\frac{1}{2}}(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}})\boldsymbol{\beta}^{\frac{1}{2}}) - \frac{1}{2}\beta^{\mathsf{T}}\boldsymbol{Q}^{\odot 2}\beta \right| \leq o_K(1),$$

where the term $o_K(1)$ does not depend on any parameters and tends to zero as $K \to \infty$. This allows us to replace $\mathcal{F}_K$ of $\tilde{F}_{\mathrm{TAP}}^K(\boldsymbol{m})$ in (4.78) with the Onsager correction term of $\tilde{F}_{\mathrm{TAP}}$ (see (4.5)) in the upper bound, so that we obtain from (4.78) that

$$\tilde{F}_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) \leq \frac{1}{N} \sup_{\boldsymbol{m} \in \mathrm{Plef}_N^{\delta_K}(\boldsymbol{Q}, \beta)} \tilde{F}_{\mathrm{TAP}}(\boldsymbol{m}) + o_K(1), \tag{4.79}$$

where the term $o_K(1)$ depends on β and tends to zero as $K \to \infty$ for fixed β. This upper bound holds for all $0 < \varepsilon < C(\beta, \boldsymbol{h}, \boldsymbol{Q})$, all $K \geq 2$ and all $N \geq c(\varepsilon, K)$. Using (2.4) we can bound the difference between the normalized Hamiltonian $\frac{1}{N}H_N$ and its diagonalized and deterministic counterpart $\frac{1}{N}\tilde{H}_N$ by any $\eta > 0$ with probability going to 1. Thus, we get for F_{TAP} (recall (1.4))

$$F_N^{\varepsilon}(\beta, \boldsymbol{h}, \boldsymbol{Q}) \leq \frac{1}{N} \sup_{\boldsymbol{m} \in \mathrm{Plef}_N^{\delta_K}(\boldsymbol{Q}, \beta)} F_{\mathrm{TAP}}(\boldsymbol{m}) + o_K(1) \tag{4.80}$$

with probability going to 1. Using (4.73) and picking K large enough depending on η we arrive at (4.1). We have thus proven (4.1) provided $\beta \in (0,\infty)^n$.

To handle β with vanishing entries, note that if $\beta_1, \beta_2 \in [0,\infty)^n$ then by (1.3) and (2.1) we have

$$|F_N^{\varepsilon}(\beta_1, \boldsymbol{h}, \boldsymbol{Q}) - F_N^{\varepsilon}(\beta_2, \boldsymbol{h}, \boldsymbol{Q})| \leq c\,|\beta_1 - \beta_2|\,, \tag{4.81}$$

with probability tending to 1. Write $F_{\mathrm{TAP}}(\boldsymbol{m};\beta)$ for F_{TAP} with the dependence on β made explicit (recall (1.4)). We similarly have

$$|F_{\mathrm{TAP}}(\boldsymbol{m};\beta_1) - F_{\mathrm{TAP}}(\boldsymbol{m};\beta_2)| \leq c\,|\beta_1 - \beta_2|\,, \tag{4.82}$$

for bounded β_1, β_2, using also that

$$\boldsymbol{m} \to \beta^{\mathsf{T}}\left(\boldsymbol{Q} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}\right)^{\odot 2}\beta \text{ is Lipschitz on compact subsets of } \mathbb{R}^n \tag{4.83}$$

uniformly in $\|\boldsymbol{Q}\|_{\infty} \leq 1$ and $\boldsymbol{m}$ with $\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} \leq \boldsymbol{Q}$. Using (4.83) again we have that some small enough constant $\rho(K)$ depending only on K

$$\mathrm{Plef}_N^{\delta_K}(\boldsymbol{Q}, \beta_2) \subset \mathrm{Plef}_N^{0}(\boldsymbol{Q}, \beta_1) = \mathrm{Plef}_N(\boldsymbol{Q}, \beta_1) \text{ for } |\beta_1 - \beta_2| \leq \rho(K)\,, \tag{4.84}$$

for bounded β_1, β_2.

Therefore for any β_1 with zero entries and $\eta > 0$ we can pick K large enough depending on β_1 and η, and β_2 with all positive entries close enough to β_1, such that

$$\begin{aligned} F_N^{\varepsilon}(\beta_1, \boldsymbol{h}, \boldsymbol{Q}) \overset{(4.81)}{\leq} F_N^{\varepsilon}(\beta_2, \boldsymbol{h}, \boldsymbol{Q}) + \tfrac{\eta}{3} &\overset{(4.80)}{\leq} \sup_{\boldsymbol{m} \in \mathrm{Plef}_N^{\delta_K}(\boldsymbol{Q}, \beta_2)} F_{\mathrm{TAP}}(\boldsymbol{m}; \beta_2) + \tfrac{2\eta}{3} \\ &\overset{(4.82),(4.84)}{\leq} \sup_{\boldsymbol{m} \in \mathrm{Plef}_N(\boldsymbol{Q}, \beta_1)} F_{\mathrm{TAP}}(\boldsymbol{m}; \beta_1) + \eta \end{aligned}$$

with probability tending to one. This proves (4.1) for β with vanishing entries.

□

Combining the TAP lower bound Proposition 3.1 and the TAP upper bound Proposition 4.1 completes the proof of Theorem 1.1.

5. Ground State Energy

All that remains is to prove the ground state formula in Theorem 1.2. To avoid technical issues with the invertibility of matrices, we will first assume that β and $\boldsymbol{h}$ are non-zero, then extend to all β and external fields using continuity.

By the uniform bound (4.3), we can write Hamiltonian and external field in terms of its diagonalizing basis, so it suffices to compute the limit of

$$\sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} = \tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \beta, \boldsymbol{h}) := \sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} = \tilde{\boldsymbol{Q}}} \left(\frac{1}{N} \sum_{k=1}^{n} \beta_k \tilde{H}_N(m^k) + \sum_{k=1}^{n} m^k \cdot \tilde{h}^k \right) \tag{5.1}$$

where $\tilde{H}_N$ is the deterministic counterpart of H_N defined in (4.2) and $\tilde{h}^k$ is the vector h^k written in the diagonalizing basis of the disorder matrix J, as in the previous section. We define the following variational form of the ground state functional

$$\widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}}) = \inf_{\boldsymbol{\Lambda} - \sqrt{2}\boldsymbol{I} \geq 0} \left(\frac{1}{4} h^{\mathsf{T}} \boldsymbol{\beta}^{-1/2} \big(\boldsymbol{\Lambda} - \sqrt{\boldsymbol{\Lambda}^2 - 2\boldsymbol{I}}\big) \boldsymbol{\beta}^{-1/2} h + \mathrm{Tr}(\boldsymbol{\Lambda} \boldsymbol{\beta}^{1/2} \tilde{\boldsymbol{Q}} \boldsymbol{\beta}^{1/2}) \right). \tag{5.2}$$

We will now show that $\sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} = \tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \boldsymbol{\beta}, \boldsymbol{h})$ converges in probability to $\widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}})$.

Proposition 5.1. *For $\beta_1, \ldots, \beta_n, h_1, \ldots, h_n \neq 0$, we have*

$$\sup_{\boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} = \tilde{\boldsymbol{Q}}} \left(\frac{1}{N} \sum_{k=1}^{n} \beta_k \tilde{H}_N(m^k) + \sum_{k=1}^{n} m^k \cdot \tilde{h}^k \right) \xrightarrow{\mathbb{P}} \widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}}). \tag{5.3}$$

Proof. We use Lagrange multipliers to explicitly solve the constrained maximization problem. Consider the Lagrangian

$$\begin{aligned} f(\boldsymbol{m}, \boldsymbol{\Lambda}) &= \sum_{k=1}^{n} \beta_k \sum_{i=1}^{N} \theta_{i/N} (m_i^k)^2 + \sum_{k=1}^{n} \tilde{h}^k \cdot m^k + \mathrm{Tr}(\boldsymbol{\Lambda} \boldsymbol{\beta}^{\frac{1}{2}} (\tilde{\boldsymbol{Q}} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}}) \boldsymbol{\beta}^{\frac{1}{2}}) \\ &= \sum_{i=1}^{N} \Big((\boldsymbol{\beta}^{\frac{1}{2}} m_i)^{\mathsf{T}} (\theta_{i/N} \boldsymbol{I} - \boldsymbol{\Lambda}) (\boldsymbol{\beta}^{\frac{1}{2}} m_i) + (\boldsymbol{\beta}^{-\frac{1}{2}} \tilde{h}_i) \cdot \boldsymbol{\beta}^{\frac{1}{2}} m_i \Big) + \mathrm{Tr}(\boldsymbol{\Lambda} \boldsymbol{\beta}^{\frac{1}{2}} \tilde{\boldsymbol{Q}} \boldsymbol{\beta}^{\frac{1}{2}}). \end{aligned}$$

Note that

$$\partial_{\Lambda_{ij}} f(\boldsymbol{m}, \boldsymbol{\Lambda}) = (1 + \delta_{i \neq j}) \sqrt{\beta_i \beta_j} \left(\tilde{\boldsymbol{Q}} - \boldsymbol{m}\boldsymbol{m}^{\mathsf{T}} \right)_{ij}.$$

Thus if $\boldsymbol{\Lambda}^* > \sqrt{2}\boldsymbol{I}$ and $\boldsymbol{m}^*$ is a critical point of $f(\boldsymbol{m}, \boldsymbol{\Lambda})$ then

$$f(\boldsymbol{m}^*, \boldsymbol{\Lambda}^*) \leq \sup_{\boldsymbol{m}\boldsymbol{m}^\intercal = \tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \beta, \boldsymbol{h}).$$

Also

$$\sup_{\boldsymbol{m}\boldsymbol{m}^\intercal = \tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \beta, \boldsymbol{h}) \leq \inf_{\Lambda > \sqrt{2}\boldsymbol{I}} \sup_{\boldsymbol{m}\boldsymbol{m}^\intercal = \tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \boldsymbol{\Lambda}) \leq \inf_{\Lambda > \sqrt{2}\boldsymbol{I}} \sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda}).$$

Therefore since f is differentiable for fixed $N, \tilde{h}$, if a finite optimizer of the r.h.s. such that $\boldsymbol{\Lambda}^* > \sqrt{2}\boldsymbol{I}$ exists then

$$\sup_{\boldsymbol{m}\boldsymbol{m}^\intercal = \tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \beta, \boldsymbol{h}) = \inf_{\Lambda > \sqrt{2}\boldsymbol{I}} \sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda}). \tag{5.4}$$

Consider $\sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda})$ for fixed $\boldsymbol{\Lambda}$. We have

$$\partial_{m_i} f(\boldsymbol{m}, \boldsymbol{\Lambda}) = 2(\theta_{i/N}\boldsymbol{I} - \boldsymbol{\Lambda})\boldsymbol{\beta}^{1/2} m_i + \boldsymbol{\beta}^{-1/2}\tilde{h}_i.$$

If $\boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{I}$ then the unique critical point of $\boldsymbol{m} \to f(\boldsymbol{m}, \boldsymbol{\Lambda})$ is thus

$$m_i(\boldsymbol{\Lambda}) = \frac{1}{2}\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda} - \theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}\tilde{h}_i \tag{5.5}$$

and by concavity this critical point corresponds to a local maximizer, and thus

$$\sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda}) = \sum_{i=1}^{N} \left(\frac{1}{4}\tilde{h}_i^\intercal \boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda} - \theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}\tilde{h}_i\right) + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}). \tag{5.6}$$

Note that if $\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2} > 0$ then since $\left(\boldsymbol{U}^\intercal\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}\boldsymbol{U}\right)_{ii} = u_i^T\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}u_i > 0$ for all orthogonal $\boldsymbol{U}$ we have

$$\sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda}) \geq \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}) \to \infty \text{ if } \boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{I},\ \sup_k \lambda_k(\boldsymbol{\Lambda}) \to \infty.$$

Also if $\boldsymbol{\Lambda} \to \sqrt{2}\boldsymbol{I}$ and $\boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{I}$ then almost surely

$$\begin{aligned} \sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda}) &\geq \tfrac{1}{4}\tilde{h}_N^\intercal\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda} - \theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}\tilde{h}_N \\ &\geq \tfrac{1}{4}\frac{|\tilde{h}_N|^2}{\lambda_{\max}\left(\boldsymbol{\beta}^{1/2}(\boldsymbol{\Lambda} - \sqrt{2}\boldsymbol{I})\boldsymbol{\beta}^{1/2}\right)} \\ &\to \infty, \end{aligned}$$

since $\tilde{h}_N \neq 0$ a.s. and $\lambda_{\max}\left(\boldsymbol{\beta}^{1/2}(\boldsymbol{\Lambda} - \sqrt{2}\boldsymbol{I})\boldsymbol{\beta}^{1/2}\right) \to 0$. This shows that minimizer of

$$\inf_{\boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{I}} f(\boldsymbol{m}(\boldsymbol{\Lambda}), \boldsymbol{\Lambda}),$$

is attained at a point in $\left\{\boldsymbol{\Lambda} : \boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{\beta}\right\}$, and thus that there exists an optimizer of

$$\inf_{\boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{I}} \sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda}),$$

which is a critical point of f, so that (5.4) holds.

We now show that $f(\boldsymbol{m}(\boldsymbol{\Lambda}), \boldsymbol{\Lambda})$ converges to the limiting function of $\boldsymbol{\Lambda}$ so that

$$\inf_{\boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{I}} \sup_{\boldsymbol{m}} f(\boldsymbol{m}, \boldsymbol{\Lambda}) \to \inf_{\boldsymbol{\Lambda} > \sqrt{2}\boldsymbol{I}} \left(\frac{1}{4}h^\intercal\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda} - \sqrt{\boldsymbol{\Lambda}^2 - 2\boldsymbol{I}})\boldsymbol{\beta}^{-1/2}h + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2})\right).$$

By Proposition 5.3 and (5.4), this convergence is uniform on compact subsets of $\boldsymbol{\beta}$ and $\boldsymbol{h}$ and the limit is $\sqrt{2}$-Lipschitz because the left hand side is.

Recall (5.6). Note that since $\mathbb{E}[\tilde{h}_{i,k}\tilde{h}_{i,l}] = h_i h_l$ it holds that

$$\mathbb{E}\left[\frac{1}{4}\tilde{h}_i^\mathsf{T}\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda}-\theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}\tilde{h}_i\right] = \frac{1}{4}h^\mathsf{T}\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda}-\theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}h,$$

and also the $\tilde{h}_i$ are independent, so by the law of large numbers

$$\begin{aligned} f(\boldsymbol{\Lambda},\boldsymbol{m}) &= \sum_{i=1}^N\left(\frac{1}{4}\tilde{h}_i^\mathsf{T}\boldsymbol{\beta}^{-1/2}(\boldsymbol{\Lambda}-\theta_{i/N}\boldsymbol{I})^{-1}\boldsymbol{\beta}^{-1/2}\tilde{h}_i\right) + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}) \\ &\to \frac{1}{4}h^\mathsf{T}\boldsymbol{\beta}^{-1/2}\left(\int_{-\sqrt{2}}^{\sqrt{2}}(\boldsymbol{\Lambda}-x\boldsymbol{I})^{-1}\,d\mu_{\mathrm{sc}}(x)\right)\boldsymbol{\beta}^{-1/2}h + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}), \end{aligned}$$

in probability. For $\boldsymbol{\Lambda}$ such that $\lambda_{min}(\boldsymbol{\Lambda}) > \sqrt{2}$, we can compute the integral explicitly. Let $\boldsymbol{\Lambda} = \boldsymbol{U}\boldsymbol{D}_\lambda\boldsymbol{U}^\mathsf{T}$. We see that

$$\int_{-\sqrt{2}}^{\sqrt{2}}(\boldsymbol{\Lambda}-x\boldsymbol{I})^{-1}\mu_{\mathrm{sc}}(x)\,dx = \int_{-\sqrt{2}}^{\sqrt{2}}\boldsymbol{U}(\boldsymbol{D}_\lambda-x\boldsymbol{I})^{-1}\boldsymbol{U}^\mathsf{T}\mu_{\mathrm{sc}}(x)\,dx = \boldsymbol{U}\int_{-\sqrt{2}}^{\sqrt{2}}(\boldsymbol{D}_\lambda-x\boldsymbol{I})^{-1}\mu_{\mathrm{sc}}(x)\,dx\boldsymbol{U}^\mathsf{T}$$

and the integral on the inside is easy to compute. In fact, using the formula for the one dimensional case, we see that

$$\int_{-\sqrt{2}}^{\sqrt{2}}(\boldsymbol{\Lambda}-x\boldsymbol{I})^{-1}\mu_{\mathrm{sc}}(x)\,dx = \boldsymbol{U}\boldsymbol{D}_{\lambda-\sqrt{\lambda^2-4}}\boldsymbol{U}^\mathsf{T} = \boldsymbol{U}\boldsymbol{D}_\lambda\boldsymbol{U}^\mathsf{T} - \boldsymbol{U}\boldsymbol{D}_{\sqrt{\lambda^2-4}}\boldsymbol{U}^\mathsf{T} = \boldsymbol{\Lambda} - \boldsymbol{U}\boldsymbol{D}_{\sqrt{\lambda^2-4}}\boldsymbol{U}^\mathsf{T}.$$

Since

$$\boldsymbol{U}\boldsymbol{D}_{\sqrt{\lambda^2-2}}\boldsymbol{U}^\mathsf{T} = \sqrt{\boldsymbol{U}\boldsymbol{D}_{\lambda^2-2}\boldsymbol{U}^\mathsf{T}} = \sqrt{\boldsymbol{U}\boldsymbol{D}_{\lambda^2}\boldsymbol{U}^\mathsf{T}-2\boldsymbol{I}} = \sqrt{(\boldsymbol{U}\boldsymbol{D}_\lambda\boldsymbol{U}^\mathsf{T})^2-2\boldsymbol{I}} = \sqrt{\boldsymbol{\Lambda}^2-2\boldsymbol{I}}$$

we have

$$\int_{-\sqrt{2}}^{\sqrt{2}}(\boldsymbol{\Lambda}-x\boldsymbol{I})^{-1}\mu_{\mathrm{sc}}(x)\,dx = \boldsymbol{\Lambda} - \sqrt{\boldsymbol{\Lambda}^2-2\boldsymbol{I}}.$$

With this formula, it follows that

$$\sup_{\boldsymbol{m}} f(\boldsymbol{\Lambda},\boldsymbol{m}) \to \frac{1}{4}h^\mathsf{T}\boldsymbol{\beta}^{-1/2}\big(\boldsymbol{\Lambda}-\sqrt{\boldsymbol{\Lambda}^2-2\boldsymbol{I}}\big)\boldsymbol{\beta}^{-1/2}h + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2})$$

and that the critical point corresponds to a maximum. Notice that this formula is well defined for $\boldsymbol{\Lambda} \geq \sqrt{2}\boldsymbol{I}$. It follows from (5.4) that the maximum of (5.1) is attained at

$$\inf_{\boldsymbol{\Lambda}\geq\sqrt{2}\boldsymbol{I}}\left(\frac{1}{4}h^\mathsf{T}\boldsymbol{\beta}^{-1/2}\big(\boldsymbol{\Lambda}-\sqrt{\boldsymbol{\Lambda}^2-2\boldsymbol{I}}\big)\boldsymbol{\beta}^{-1/2}h + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2})\right) \tag{5.7}$$

in the limit. □

We now explicitly solve the optimization in $\widetilde{\mathrm{GSE}}(\beta,h,\tilde{\boldsymbol{Q}})$ to arrive at the closed form expression from (1.7).

5.1. The One Dimensional Case

We first address the case $n=1$ as a warm-up. Solving the variational problem in this case is considerably easier because we do not have to worry about the non-commutativity of the matrices. When $n=1$ the variational problem is

$$\inf_{\lambda\geq\sqrt{2}}\left(\frac{1}{4}\frac{h^2}{\beta}\big(\lambda-\sqrt{\lambda^2-2}\big)+\lambda\beta\tilde{q}\right). \tag{5.8}$$

Let $A = \frac{1}{4}\frac{h^2}{\beta}$ and $B = \beta\tilde{q}$. With the change of variables $\lambda = \frac{1}{\sqrt{2}}\left(x+\frac{1}{x}\right), x \in (0,1]$, and using that $\sqrt{\lambda^2-2} = \frac{1}{\sqrt{2}}\left(\frac{1}{x}-x\right)$ one obtains that (5.8) equals

$$\inf_{x\in(0,1]}\left(A\sqrt{2}x + B\frac{1}{\sqrt{2}}\left(\frac{1}{x}+x\right)\right) = \sqrt{2B\,(2A+B)} = \sqrt{\tilde{q}h^2+2\beta^2\tilde{q}^2}$$

which proves (1.11) and is indeed the formula from [BK19, (1.6) and Lemma 20].

5.2. The n Dimensional Case

A matrix version of this change of variables allows one to solve also the case $n > 1$, giving rise to $\mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}})$ from (1.7).

Proposition 5.2. *For $\tilde{\boldsymbol{Q}} > 0$ and $\boldsymbol{\beta} > 0$ it holds that*

$$\widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}}) = \mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}}). \tag{5.9}$$

Proof. Let

$$\boldsymbol{A} = \frac{1}{4}\boldsymbol{\beta}^{-1/2} h h^{\mathsf{T}} \boldsymbol{\beta}^{-1/2} \text{ and } \boldsymbol{B} = \boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}.$$

Writing the first term of $\mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}})$ as the trace of a 1×1 matrix and using the cyclical property of the trace we have

$$\widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}}) = \inf_{\boldsymbol{\Lambda} - \sqrt{2}\boldsymbol{I} \geq 0} \left(\frac{1}{4}\mathrm{Tr}\big(\big(\boldsymbol{\Lambda} - \sqrt{\boldsymbol{\Lambda}^2 - 2\boldsymbol{I}}\big)\boldsymbol{A}\big) + \mathrm{Tr}(\boldsymbol{\Lambda}\boldsymbol{B}) \right).$$

We use the change of variables $\boldsymbol{\Lambda} = \frac{1}{\sqrt{2}}(\boldsymbol{X} + \boldsymbol{X}^{-1})$ where $\boldsymbol{0} < \boldsymbol{X} \leq \boldsymbol{I}$ and $\boldsymbol{X}$ is symmetric (to see that $\boldsymbol{\Lambda}$ can always be written in this form recall that $\boldsymbol{\Lambda}$ is symmetric and has eigenvalues larger or equal to $\sqrt{2}$). It follows that $(\boldsymbol{\Lambda}^2 - 2\boldsymbol{I})^{\frac{1}{2}} = \frac{1}{\sqrt{2}}(\boldsymbol{X}^{-1} - \boldsymbol{X})$. Thus

$$\widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}}) = \inf_{\boldsymbol{0} < \boldsymbol{X} \leq \boldsymbol{I}} \left(\sqrt{2}\mathrm{Tr}\big(\boldsymbol{X}\boldsymbol{A}\big) + \frac{1}{\sqrt{2}}\mathrm{Tr}((\boldsymbol{X} + \boldsymbol{X}^{-1})\boldsymbol{B}) \right). \tag{5.10}$$

Consider the critical point equation for the quantity in the inf:

$$\partial_{\boldsymbol{X}} \left(\sqrt{2}\mathrm{Tr}\big(\boldsymbol{X}\boldsymbol{A}\big) + \frac{1}{\sqrt{2}}\mathrm{Tr}(\boldsymbol{B}(\boldsymbol{X} + \boldsymbol{X}^{-1})) \right) = \sqrt{2}\boldsymbol{A} + \frac{1}{\sqrt{2}}\boldsymbol{B} - \frac{1}{\sqrt{2}}\boldsymbol{X}^{-1}\boldsymbol{B}\boldsymbol{X}^{-1} = \boldsymbol{0}.$$

Using the change of variables $\boldsymbol{Y} = \boldsymbol{B}^{\frac{1}{2}}\boldsymbol{X}^{-1}$, it is equivalent to

$$\boldsymbol{Y}^{\mathsf{T}}\boldsymbol{Y} = 2\boldsymbol{A} + \boldsymbol{B}.$$

If we diagonalize $2\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{U}_1\boldsymbol{D}\boldsymbol{U}_1^{\mathsf{T}}$, it is further equivalent to $\boldsymbol{Y} = \boldsymbol{U}_2\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{U}_1^{\mathsf{T}}$ for some orthogonal $\boldsymbol{U}_2$, and therefore to $\boldsymbol{X} = \boldsymbol{U}_1\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{U}_2^{\mathsf{T}}\boldsymbol{B}^{\frac{1}{2}} = (2\boldsymbol{A} + \boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}$, where $\boldsymbol{U} = \boldsymbol{U}_1^{\mathsf{T}}\boldsymbol{U}_2^{\mathsf{T}}$ Thus if $\boldsymbol{X}$ is symmetric and

$$\boldsymbol{X} = (2\boldsymbol{A} + \boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}, \tag{5.11}$$

for some orthogonal $\boldsymbol{U}$ it is a critical point. Such a $\boldsymbol{U}$ can be found as follows. The symmetry condition is equivalent to

$$(2\boldsymbol{A} + \boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}} = \boldsymbol{B}^{\frac{1}{2}}\boldsymbol{U}^{\mathsf{T}}(2\boldsymbol{A} + \boldsymbol{B})^{-\frac{1}{2}} \iff \boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}(2\boldsymbol{A} + \boldsymbol{B})^{\frac{1}{2}} = (2\boldsymbol{A} + \boldsymbol{B})^{\frac{1}{2}}\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{U}^{\mathsf{T}}.$$

Write $\boldsymbol{B}^{\frac{1}{2}}(2\boldsymbol{A} + \boldsymbol{B})^{\frac{1}{2}} = \boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{T}^{\mathsf{T}}$ in its singular value decomposition. The condition then becomes,

$$\boldsymbol{U}\boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{T}^{\mathsf{T}} = \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{S}^{\mathsf{T}}\boldsymbol{U}^{\mathsf{T}},$$

so $\boldsymbol{U} = \boldsymbol{T}\boldsymbol{S}^{\mathsf{T}}$ is orthogonal and makes $\boldsymbol{X}$ symmetric. We have thus proven than this $\boldsymbol{X}$ is a critical point of the expression in the infiumum of (5.10).

Now note that $\boldsymbol{X} \mapsto \mathrm{Tr}(\boldsymbol{B}\boldsymbol{X}^{-1})$ is convex because $\boldsymbol{X} \mapsto \boldsymbol{X}^{-1}$ is convex and $\boldsymbol{X} \mapsto \mathrm{Tr}(\boldsymbol{B}\boldsymbol{X})$ is increasing for positive definite $\boldsymbol{B}$, see [Bha96, Corollary V.2.6]. Therefore the expression in the infiumum of (5.10) is convex in $\boldsymbol{X}$, and thus the exhibited critical point is a global minimizer.

Next we compute the value at this minimizer. Substituting it into (5.10) and using $\boldsymbol{X}^{-1} = (\boldsymbol{X}^{-1})^\intercal = (2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{-\frac{1}{2}}$ one obtains

$$
\begin{aligned}
&\widetilde{\mathrm{GSE}}(\beta, h, \tilde{\boldsymbol{Q}})\\
=&\frac{1}{\sqrt{2}}\Big(2\mathrm{Tr}\big(\boldsymbol{A}(2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}\big) + \mathrm{Tr}((2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{B}) + \mathrm{Tr}((2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}\boldsymbol{U}\boldsymbol{B}^{-\frac{1}{2}}\boldsymbol{B})\Big)\\
=&\frac{1}{\sqrt{2}}\Big(2\mathrm{Tr}\big(\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}(2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\big) + \mathrm{Tr}(\boldsymbol{B}^{\frac{3}{2}}(2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}) + \mathrm{Tr}(\boldsymbol{B}^{\frac{1}{2}}(2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}\boldsymbol{U})\Big)\\
=&\frac{1}{\sqrt{2}}\mathrm{Tr}\Big(\boldsymbol{B}^{\frac{1}{2}}\Big(2\boldsymbol{A}+\boldsymbol{B}+(2\boldsymbol{A}+\boldsymbol{B})\Big)\Big((2\boldsymbol{A}+\boldsymbol{B})^{-\frac{1}{2}}\boldsymbol{U}\Big)\Big)\\
=&\sqrt{2}\mathrm{Tr}\big(\boldsymbol{B}^{\frac{1}{2}}(2\boldsymbol{A}+\boldsymbol{B})^{\frac{1}{2}}\boldsymbol{U}\big)\\
=&\sqrt{2}\mathrm{Tr}\big(\boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{T}^\intercal\boldsymbol{T}\boldsymbol{S}^\intercal\big) = \sqrt{2}\mathrm{Tr}(\boldsymbol{\Sigma}).
\end{aligned}
$$

Note that $\boldsymbol{\Sigma}$ is the diagonal matrix of singular values of $\boldsymbol{B}^{1/2}(2\boldsymbol{A}+\boldsymbol{B})^{1/2}$, i.e. of square roots of the eigenvalues of

$$\boldsymbol{B}^{1/2}(2\boldsymbol{A}+\boldsymbol{B})^{1/2}\left(\boldsymbol{B}^{1/2}(2\boldsymbol{A}+\boldsymbol{B})^{1/2}\right)^T.$$

Using repeatedly the property that $\boldsymbol{C}\boldsymbol{D}$ and $\boldsymbol{D}\boldsymbol{C}$ have the same eigenvalues for square $\boldsymbol{C}, \boldsymbol{D}$ we get that these eigenvalues coincide with those of

$$(2\boldsymbol{A}+\boldsymbol{B})\,\boldsymbol{B} = (2\boldsymbol{A}+\boldsymbol{B})\,\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}^{1/2}.$$

Using the same property again this r.h.s. in turn has the same eigenvalues as

$$\boldsymbol{\beta}^{1/2}(2\boldsymbol{A}+\boldsymbol{B})\,\boldsymbol{\beta}^{1/2}\tilde{\boldsymbol{Q}} = \Big(\frac{1}{2}hh^\intercal + \boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\Big)\tilde{\boldsymbol{Q}},$$

which in turn has the same eigenvalues as

$$\Big(\frac{1}{2}hh^\intercal + \boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\Big)^{\frac{1}{2}}\tilde{\boldsymbol{Q}}\Big(\frac{1}{2}hh^\intercal + \boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\Big)^{\frac{1}{2}}.$$

This proves that

$$\mathrm{Tr}\,(\boldsymbol{\Sigma}) = \mathrm{Tr}\left(\sqrt{\Big(\frac{1}{2}hh^\intercal + \boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\Big)^{\frac{1}{2}}\tilde{\boldsymbol{Q}}\Big(\frac{1}{2}hh^\intercal + \boldsymbol{\beta}\tilde{\boldsymbol{Q}}\boldsymbol{\beta}\Big)^{\frac{1}{2}}}\right),$$

and recalling the definition (1.7) of GSE this completes the proof. □

Thus for β and h with only non-zero components

$$\sup_{\boldsymbol{m}\boldsymbol{m}^\intercal=\tilde{\boldsymbol{Q}}}\left(\frac{1}{N}\sum_{k=1}^{n}\beta_k\tilde{H}_N(m^k) + \sum_{k=1}^{n}m^k\cdot\tilde{h}^k\right) \xrightarrow{\mathbb{P}} \mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}}) \tag{5.12}$$

(by combining Propositions 5.1 and 5.2). The formula for $\mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}})$ is well-defined also if some entry of β or h is zero. To extend (5.12) to this case we will using a continuity argument enabled by the next lemma, which shows that (5.1) is Lipschitz in β and h.

Lemma 5.3. *If $\tilde{\boldsymbol{Q}}$ is positive definite with entries bounded by 1, then*

$$\Big|\sup_{\boldsymbol{m}\boldsymbol{m}^\intercal=\tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \beta^1, \boldsymbol{h}^1) - \sup_{\boldsymbol{m}\boldsymbol{m}^\intercal=\tilde{\boldsymbol{Q}}} f(\boldsymbol{m}, \beta^2, \boldsymbol{h}^2)\Big| \le \sqrt{2}\|\beta^1-\beta^2\|_\infty + \|h^1-h^2\|_\infty.$$

Proof. This follows since

$$
\begin{aligned}
\sup_{\boldsymbol{m}\boldsymbol{m}^\intercal=\tilde{\boldsymbol{Q}}}\big|f(\boldsymbol{m}, \beta^1, \boldsymbol{h}^1) - f(\boldsymbol{m}, \beta^2, \boldsymbol{h}^2)\big| &\le \sup_{|m^1|,\dots,|m^n|\le 1}\big|f(\boldsymbol{m}, \beta^1, \boldsymbol{h}^1) - f(\boldsymbol{m}, \beta^2, \boldsymbol{h}^2)\big|\\
&\le \sqrt{2}\|\beta^1-\beta^2\|_\infty + \sup_{k\le n}|(h^k)^1 - (h^k)^2|\\
&\le \sqrt{2}\|\beta^1-\beta^2\|_\infty + \|h^1-h^2\|_\infty,
\end{aligned}
$$

because

$$|(h^k)^1 - (h^k)^2| = |h_k^1 u - h_k^2 u|^2 = |h_k^1 - h_k^2|^2.$$

□

Note also from the formula in (1.7) that

$$\text{for all } \tilde{\boldsymbol{Q}} > 0 \text{ the map } (h, \beta) \to \mathrm{GSE}(\beta, h, \tilde{\boldsymbol{Q}}) \text{ is continuous.} \tag{5.13}$$

Theorem 1.2 is now immediate from Propositions 5.1 and 5.2 and continuity.

Proof of Theorem 1.2. The reduction above (5.1) and Propositions 5.1 and 5.2 prove the claim (1.8) when all entries of β and h are non-zero. A simple approximation argument using Lemma 5.3 and (5.13) extends this to all β, h.

□

Bibliography

[AA13] Antonio Auffinger and Gerard Ben Arous. Complexity of random smooth functions on the high-dimensional sphere. *The Annals of Probability*, 41(6):4214–4247, 2013.

[AAČ13] Antonio Auffinger, Gérard Ben Arous, and Jiří Černỳ. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2):165–201, 2013.

[ADD99] R. B. Ash and C. A. Doléans-Dade. *Probability and Measure Theory.* 2 edition, 1999.

[AJ19] Antonio Auffinger and Aukosh Jagannath. Thouless–anderson–palmer equations for generic p-spin glasses. *The Annals of Probability*, 47(4):2230–2256, 2019.

[AJ21] Gérard Ben Arous and Aukosh Jagannath. Shattering versus metastability in spin glasses, 2021.

[AK18] Ahmed El Alaoui and Florent Krzakala. Estimation in the spiked wigner model: A short proof of the replica formula. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1874–1878, 2018.

[ALR87] Michael Aizenman, Joel L Lebowitz, and David Ruelle. Some rigorous results on the sherrington-kirkpatrick spin glass model. *Communications in mathematical physics*, 112:3–20, 1987.

[AMMN19] Gerard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330, 2019.

[ASS03] Michael Aizenman, Robert Sims, and Shannon L. Starr. Extended variational principle for the sherrington-kirkpatrick spin-glass model. *Phys. Rev. B*, 68:214403, 12 2003.

[AZ22] Antonio Auffinger and Yuxin Zhou. On properties of the spherical mixed vector p-spin model. *Stochastic Process. Appl.*, 146:382–413, 2022.

[BAJ18] Gérard Ben Arous and Aukosh Jagannath. Spectral gap estimates in mean field spin glasses. *Comm. Math. Phys.*, 361(1):1–52, 2018.

[BB21] Debapratim Banerjee and David Belius. Fluctuations of the free energy of the mixed p-spin mean field spin glass model. *arXiv preprint arXiv:2108.03109*, 2021.

[BČNS22] David Belius, Jiří Černỳ, Shuta Nakajima, and Marius A Schmidt. Triviality of the geometry of mixed p-spin spherical hamiltonians with external field. *Journal of Statistical Physics*, 186(1):1–34, 2022.

[BCWDW20] Jinho Baik, Elizabeth Collins-Wildman, Pierre Le Doussal, and Hao Wu. Spherical spin glass model with external field. *arXiv:2010.06123 [cond-mat, physics:math-ph]*, October 2020. arXiv: 2010.06123.

[BCWLDW21] Jinho Baik, Elizabeth Collins-Woodfin, Pierre Le Doussal, and Hao Wu. Spherical spin glass model with external field. *Journal of Statistical Physics*, 183(2):1–79, 2021.

[Bel22] David Belius. High temperature tap upper bound for the free energy of mean field spin glasses, 2022.

[BFK23] David Belius, Leon Fröber, and Justin Ko. Tap variational principle for the constrained overlap multiple spherical sherrington-kirkpatrick model. *arXiv preprint arXiv:2304.04031*, 2023.

[BGK16] Florent Benaych-Georges and Antti Knowles. Lectures on the local semicircle law for wigner matrices. *arXiv preprint arXiv:1601.04055*, 2016.

[Bha96] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. Springer New York, 1996.

[BK19] David Belius and Nicola Kistler. The TAP-Plefka variational principle for the spherical SK model. *Comm. Math. Phys.*, 367(3):991–1017, 2019.

[BKL02] Anton Bovier, Irina Kurkova, and Matthias Löwe. Fluctuations of the free energy in the rem and the p-spin sk models. *The Annals of Probability*, 30(2):605–651, 2002.

[BL16] J. Baik and J.O. Lee. Fluctuations of the free energy of the spherical sherrington–kirkpatrick model. *Journal of Statistical Physics*, 165(2):185–224, 2016.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[BM80] Alan J Bray and Michael A Moore. Metastable states in spin glasses. *Journal of Physics C: Solid State Physics*, 13(19):L469, 1980.

[Bol14] Erwin Bolthausen. An iterative construction of solutions of the tap equations for the sherrington–kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014.

[Bol19] Erwin Bolthausen. A morita type proof of the replica-symmetric formula for sk. In Véronique Gayrard, Louis-Pierre Arguin, Nicola Kistler, and Irina Kourkova, editors, *Statistical Mechanics of Classical and Disordered Systems*, volume 293, pages 63–93. Springer International Publishing, Cham, 2019. Series Title: Springer Proceedings in Mathematics & Statistics.

[Bru67] Stephen G Brush. History of the lenz-ising model. *Reviews of modern physics*, 39(4):883, 1967.

[BS22a] David Belius and Marius A Schmidt. Phase diagram for the tap energy of the p-spin spherical mean field spin glass model. *arXiv preprint arXiv:2207.02821*, 2022.

[BS22b] Anton Bovier and Adrien Schertzer. Fluctuations of the free energy in p-spin sk models on two scales, 2022.

[BY05] Z. Bai and J. Yao. On the convergence of the spectral empirical process of wigner matrices. *Bernoulli*, 11(6):1059–1092, 2005.

[BY22] Christian Brennecke and Horng-Tzer Yau. The replica symmetric formula for the sk model revisited. *Journal of Mathematical Physics*, 63(7):073302, 2022.

[CGPM03] Andrea Cavagna, Irene Giardina, Giorgio Parisi, and Marc Mézard. On the formal equivalence of the tap and thermodynamic methods in the sk model. *Journal of Physics A: Mathematical and General*, 36(5):1175, 2003.

[Cha09] Sourav Chatterjee. Disorder chaos and multiple valleys in spin glasses. *arXiv preprint arXiv:0907.3381*, 2009.

[Cha10] Sourav Chatterjee. Spin glasses and stein's method. *Probability theory and related fields*, 148(3):567–600, 2010.

[Che13] Wei-Kuo Chen. The Aizenman-Sims-Starr scheme and Parisi formula for mixed p-spin spherical models. *Electron. J. Probab.*, 18:no. 94, 14, 2013.

[CL04] Andrea Crisanti and Luca Leuzzi. Spherical 2+ p spin-glass model: An exactly solvable model for glass to spin-glass transition. *Physical review letters*, 93(21):217203, 2004.

[Con13] Pierluigi Contucci. Dmitry panchenko: The sherrington–kirkpatrick model. *Journal of Statistical Physics*, 153(3):551–552, 2013.

[CP17] Wei-Kuo Chen and Dmitry Panchenko. Temperature chaos in some spherical mixed p-spin models. *J. Stat. Phys.*, 166(5):1151–1162, 2017.

[CP18] Wei-Kuo Chen and Dmitry Panchenko. On the tap free energy in the mixed p-spin models. *Communications in Mathematical Physics*, 362(1):219–252, 2018.

[CPS22] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. Generalized TAP Free Energy. *Communications on Pure and Applied Mathematics*, page cpa.22040, January 2022.

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

[CS95] Andrea Crisanti and H-J Sommers. Thouless-anderson-palmer approach to the spherical p-spin spin glass model. *Journal de Physique I*, 5(7):805–813, 1995.

[CS17] Wei-Kuo Chen and Arnab Sen. Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed p-spin models. *Communications in Mathematical Physics*, 350(1):129–173, 2017.

[DDY83] Cyrano De Dominicis and A Peter Young. Weighted averages and order parameters for the infinite range ising spin glass. *Journal of Physics A: Mathematical and General*, 16(9):2063, 1983.

[Der80] B. Derrida. Random-Energy Model: Limit of a Family of Disordered Models. *Physical Review Letters*, 45(2):79–82, July 1980.

[Dob68] Roland L'vovich Dobrushin. The description of the random field by its conditional distributions and its regularity conditions. *Teoriya Veroyatnostei i ee Primeneniya*, 13(2):201–229, 1968.

[DS19] Jian Ding and Nike Sun. Capacity lower bound for the ising perceptron. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 816–827, 2019.

[DSS15] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 59–68, 2015.

[DZ15] Amir Dembo and Ofer Zeitouni. Matrix optimization under random external fields. *Journal of Statistical Physics*, 159:1306–1326, 2015.

[EYY12] L. Erdos, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized wigner matrices. *Adv. Math.*, 229(3):1435–1515, 2012.

[FLD14] Yan V Fyodorov and Pierre Le Doussal. Topology trivialization and large deviations for the minimum in the simplest random optimization. *Journal of Statistical Physics*, 154(1-2):466–490, 2014.

[FPV92] S. Franz, Giorgio Parisi, and Miguel Virasoro. The replica method on and off equilibrium. *http://dx.doi.org/10.1051/jp1:1992115*, 2, 10 1992.

[FR20] Silvio Franz and Jacopo Rocchi. Large deviations of glassy effective potentials. *Journal of Physics A: Mathematical and Theoretical*, 53(48):485002, 2020.

[Ful00] William Fulton. Eigenvalues, invariant factors, highest weights, and schubert calculus. *Bulletin of the American Mathematical Society*, 37(3):209–249, 2000.

[Fyo15] Y. V. Fyodorov. High-dimensional random fields and random matrix theory. *Markov Processes and Related Fields*, 21(3, part 1):483–518, 2015. tex.fjournal: Markov Processes and Related Fields tex.mrclass: 60G60 (60B20 82D30) tex.mrnumber: 3469265.

[GH21] Alice Guionnet and Jonathan Husson. Asymptotics of k dimensional spherical integrals and applications, 2021.

[GM84a] David J Gross and Marc Mézard. The simplest spin glass. *Nuclear Physics B*, 240(4):431–452, 1984.

[GM84b] D.J. Gross and M. Mezard. The simplest spin glass. *Nuclear Physics B*, 240(4):431–452, November 1984.

[GM05] A. Guionnet and M. Maida. A Fourier view on the R-transform and related asymptotics of spherical integrals. *J. Funct. Anal.*, 222(2):435–490, 2005.

[Gue03] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in mathematical physics*, 233(1):1–12, 2003.

[HC56] Harish-Chandra. Invariant differential operators on a semisimple Lie algebra. *Proc. Nat. Acad. Sci. U.S.A.*, 42:252–253, 1956.

[Hig08] Nicholas J. Higham. *Functions of matrices*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008. Theory and computation.

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[HK22] Jonathan Husson and Justin Ko. Spherical integrals of sublinear rank, 2022.

[Isi24] Ernst Ising. *Beitrag zur Theorie des Ferro-und Paramagnetismus*. PhD thesis, Grefe und Tiedemann, 1924.

[IZ80] C. Itzykson and J. B. Zuber. The planar approximation. II. *J. Math. Phys.*, 21:411–421, 1980.

[Jag19] Aukosh Jagannath. Dynamics of mean field spin glasses on short and long timescales. *J. Math. Phys.*, 60(8):083305, 9, 2019.

[JLM20] Aukosh Jagannath, Patrick Lopatto, and Léo Miolane. Statistical thresholds for tensor PCA. *Ann. Appl. Probab.*, 30(4):1910–1933, 2020.

[JT17] Aukosh Jagannath and Ian Tobasco. Low temperature asymptotics of spherical mean field spin glasses. *Communications in Mathematical Physics*, 352(3):979–1017, 2017.

[Kal21] Olav Kallenberg. *Foundations of Modern Probability.* 3 edition, 2021.

[Ko19] Justin Ko. The crisanti–sommers formula for spherical spin glasses with vector spins. *arXiv preprint arXiv:1911.04355*, 2019.

[Ko20] Justin Ko. Free energy of multiple systems of spherical spin glasses with constrained overlaps. *Electron. J. Probab.*, 25:34 pp., 2020.

[KPV93] Jorge Kurchan, Giorgio Parisi, and Miguel Angel Virasoro. Barriers and metastable states as saddle points in the replica approach. *Journal de Physique I*, 3(8):1819–1838, 1993.

[KR98] Jeong Han Kim and James R Roche. Covering cubes by random half cubes, with applications to binary neural networks. *Journal of Computer and System Sciences*, 56(2):223–252, 1998.

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

[Lan20] Benjamin Landon. Free energy fluctuations of the \$2\$-spin spherical SK model at critical temperature. *arXiv:2010.06691 [math-ph]*, October 2020. arXiv: 2010.06691.

[LKZ17] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017.

[LM19] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3):859–929, 2019.

[LR69] OE Lanford and David Ruelle. Observables at infinity and states with short range correlations in statistical mechanics. *Communications in Mathematical Physics*, 13(3):194–215, 1969.

[LS20] Benjamin Landon and Philippe Sosoe. Fluctuations of the 2-spin SSK model with magnetic field. *arXiv:2009.12514 [math-ph]*, September 2020. arXiv: 2009.12514.

[MM09] Marc Mezard and Andrea Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

[MPV87] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[Pan13a] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Ann. of Math. (2)*, 177(1):383–393, 2013.

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

[Pan14] Dmitry Panchenko. The parisi formula for mixed p-spin models. *The Annals of Probability*, 42(3):946–958, 2014.

[Pan16] Dmitry Panchenko. Chaos in temperature in generic $2p$-spin models. *Comm. Math. Phys.*, 346(2):703–739, 2016.

[Pan18a] Dmitry Panchenko. Free energy in the mixed p-spin models with vector spins. *Ann. Probab.*, 46(2):865–896, 2018.

[Pan18b] Dmitry Panchenko. Free energy in the Potts spin glass. *Ann. Probab.*, 46(2):829–864, 2018.

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

[Par80] Giorgio Parisi. A sequence of approximated solutions to the sk model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.

[Par98] Beresford N. Parlett. *The Symmetric Eigenvalue Problem.* Society for Industrial and Applied Mathematics, 1998.

[Ple82a] T. Plefka. A lower bound for the spin glass order parameter of the infiniteranged ising spin glass model. *Journal of Physics A: Mathematical and General*, 15(5):L251, 1982.

[Ple82b] Timm Plefka. Convergence condition of the tap equation for the infinite-ranged ising spin glass model. *Journal of Physics A: Mathematical and general*, 15(6):1971, 1982.

[PT07] Dmitry Panchenko and Michel Talagrand. On the overlap in the multiple spherical SK models. *Ann. Probab.*, 35(6):2321–2355, 2007.

[RM14] Emile Richard and Andrea Montanari. A statistical model for tensor pca. *Advances in neural information processing systems*, 27, 2014.

[She93] David Sherrington. Neural networks: the spin glass approach. In *North-Holland Mathematical Library*, volume 51, pages 261–291. Elsevier, 1993.

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

[Sub17a] Eliran Subag. The complexity of spherical p-spin models - a second moment approach. *The Annals of Probability*, 45(5):3385–3450, 2017.

[Sub17b] Eliran Subag. The geometry of the gibbs measure of pure spherical spin glasses. *Inventiones mathematicae*, 210:135–209, 2017.

[Sub18] Eliran Subag. Free energy landscapes in spherical spin glasses. *arXiv preprint arXiv:1804.10576*, 2018.

[Sub21] Eliran Subag. The free energy of spherical pure p-spin models–computation from the tap approach. *arXiv preprint arXiv:2101.04352*, 2021.

[SZ17] Eliran Subag and Ofer Zeitouni. The extremal process of critical points of the pure p-spin spherical spin glass model. *Probability theory and related fields*, 168(3):773–820, 2017.

[T$^+$03] Michel Talagrand et al. *Spin glasses: a challenge for mathematicians: cavity and mean field models*, volume 46. Springer Science and Business Media, 2003.

[Tal00] Michel Talagrand. Multiple levels of symmetry breaking. *Probability theory and related fields*, 117(4):449–466, 2000.

[Tal06a] Michel Talagrand. Free energy of the spherical mean field model. *Probab. Theory Related Fields*, 134(3):339–382, 2006.

[Tal06b] Michel Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.

[Tal10] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science and Business Media, 2010.

[TAP77] David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of a solvable model of a spin glass. *Philosophical Magazine*, 35(3):593–601, 1977.

[TW96] CA Tracey and H Widom. On orthogonal and symplectic ensembles. *Commun. Math. Phys*, 177:727–54, 1996.