Commun. Math. Phys. 367, 991–1017 (2019)
Digital Object Identifier (DOI) https://doi.org/10.1007/s00220-019-03304-y

Communications in Mathematical Physics

The TAP–Plefka Variational Principle for the Spherical SK Model

David Belius[1], **Nicola Kistler**[2]

[1] University of Basel, Basel, Switzerland. E-mail: david.belius@cantab.net
[2] University of Frankfurt, Frankfurt, Germany. E-mail: Kistler@math.uni-frankfurt.de

Received: 11 March 2018 / Accepted: 7 October 2018
Published online: 6 February 2019 – © Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract: We reinterpret the Thouless–Anderson–Palmer approach to mean field spin glass models as a variational principle in the spirit of the Gibbs variational principle and the Bragg–Williams approximation. We prove this TAP–Plefka variational principle rigorously in the case of the spherical Sherrington–Kirkpatrick model.

1. Introduction

There are several approaches in theoretical physics and mathematics to study the Sherrington–Kirkpatrick (SK) mean field spin glass model [21] and its variants. The most successful in physics is the replica approach, which with Parisi's replica symmetry breaking Ansatz led him to his celebrated formula for the free energy [16]. The mathematically rigorous proofs of the formula due to Guerra, Talagrand and Panchenko are based on a subtle combination of interpolation, recursion, the Ghirlanda–Guerra identities and an invariance property for the limiting Gibbs measure [14,17,18,25]. A further approach in the physics literature is the one due to Thouless, Anderson and Palmer (TAP) and Plefka. It originates in [27] as a diagrammatic expansion of the partition function of the Ising SK model relating the free energy to the so called TAP free energy, which is a disorder-dependent function defined on the space of magnetizations of the spins. It claims that the free energy equals the TAP free energy at magnetizations that solve a set of mean field equations and satisfy certain convergence conditions, which have not been completely clarified. *Plefka's condition* [19,20] is believed to be necessary, but it is not clear if it is also sufficient. The high temperature analysis of [27] has been made rigorous in [1]. The physicist's TAP approach has been adapted to spherical models in [11].

In this paper we reinterpret the TAP approach as a variational principle for the free energy, which states that the free energy equals the maximum of the TAP free energy taken over magnetizations satisfying appropriate conditions. We make this rigorous in the case of the spherical Sherrington–Kirkpatrick model, and show that for this model Plefka's condition is the only condition needed to formulate the variational principle.

Let $H_N(\sigma), \sigma \in \mathbb{R}^N$, be the 2-spin spherical SK Hamiltonian which is a centered Gaussian process on $\mathbb{R}^N$ with covariance

$$\mathbb{E}\left[H_N(\sigma) H_N\left(\sigma'\right)\right] = N\left(\sigma \cdot \sigma'\right)^2, \tag{1.1}$$

which can be constructed by setting

$$H_N(\sigma) = \sqrt{N} \sum_{i,j=1}^{N} J_{ij} \sigma_i \sigma_j \tag{1.2}$$

for iid standard Gaussian random variables J_{ij} and $\sigma \in \mathbb{R}^N$. Let E be the uniform measure on the unit sphere in $\mathbb{R}^N$ and let

$$Z_N(\beta, h_N) = E\left[e^{\beta H_N(\sigma) + N h_N \cdot \sigma}\right] \text{ and } F_N(\beta, h_N) = \frac{1}{N} \log Z_N(\beta, h_N) \tag{1.3}$$

be the partition function and free energy in the presence of an external field $h_N \in \mathbb{R}^N$. The TAP free energy for this model is given by [11,27]

$$H_{TAP}(m) = \beta H_N(m) + N m \cdot h_N + \frac{N}{2} \log\left(1 - |m|^2\right) + N \frac{\beta^2}{2} \left(1 - |m|^2\right)^2$$

for $m \in \mathbb{R}^N$ with $|m| < 1$, and Plefka's condition [19,20] reads

$$\beta(m) \le \frac{1}{\sqrt{2}},$$

where

$$\beta(m) = \beta\left(1 - |m|^2\right).$$

We refer to the approximation

$$F_N(\beta, h_N) \approx \frac{1}{N} \sup_{m \in \mathbb{R}^N : |m| < 1, \beta(m) \le \frac{1}{\sqrt{2}}} H_{TAP}(m) \tag{1.4}$$

as the TAP–Plefka variational principle and prove it in the following form.

Theorem 1. *For any $\beta > 0$, $h \ge 0$ and any sequence $h_1, h_2, \ldots$ with $|h_N| = h$ one has*

$$\left| F_N(\beta, h_N) - \frac{1}{N} \sup_{m \in \mathbb{R}^N : |m| < 1, \beta(m) \le \frac{1}{\sqrt{2}}} H_{TAP}(m) \right| \to 0 \textit{ in probability.} \tag{1.5}$$

We also include a solution of the TAP–Plefka variational problem that reduces it from a random N-dimensional optimization problem to one which is deterministic and one dimensional.

Lemma 2. *For any $\beta, h, h_1, h_2, \ldots$, as in Theorem 1 one has*

$$\left| \frac{1}{N} \sup_{m \in \mathbb{R}^N : |m| < 1, \beta(m) \le \frac{1}{\sqrt{2}}} H_{TAP}(m) - \sup_{q \in [0,1] : \beta(1-q) \le \frac{1}{\sqrt{2}}} \mathcal{B}(q) \right| \to 0 \text{ in probability},$$

where

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

Together, Theorem 1 and Lemma 2 show that

$$F_N(\beta, h_N) \to \sup_{q \in [0,1] : \beta(1-q) \le \frac{1}{\sqrt{2}}} \mathcal{B}(q) . \tag{1.6}$$

For comparison, the Parisi formula in this context [12,24] states that

$$F_N(\beta, h_N) \to \inf_{q \in [0,1]} \mathcal{P}(q) , \tag{1.7}$$

where

$$\mathcal{P}(q) = \frac{1}{2} h^2 (1-q) + \frac{1}{2} \frac{q}{1-q} + \frac{1}{2} \log(1-q) + \frac{1}{2} \beta^2 \left(1 - q^2\right) .$$

1.1. Discussion.

1.1.1. The TAP–Plefka variational principle The TAP–Plefka variational principle (1.4) should be compared to the classical Gibbs variational principle which states that

$$F_N(\beta, h_N) = \frac{1}{N} \sup_{\mathcal{G}} \{ \mathcal{G}(\beta H_N(\sigma) + N\sigma \cdot h_N) - H(\mathcal{G} || E) \} , \tag{1.8}$$

where the supremum is over all probability measures which are absolutely continuous with respect to E, and $H(\mathcal{G}||E)$ is the relative entropy of $\mathcal{G}$ with respect to E. The first term is the internal energy and the second is the entropy.

In the classical Bragg–Williams approximation [8,28, Section 4.1.2] in non-disordered statistical physics one restricts this sup to simple measures $\mathcal{G}$ that are parameterized by a mean magnetization $m \in \mathbb{R}^N$; in the case of ± 1 spins one considers measures under which the spins σ_i are independent with mean m_i. For any m the corresponding measure gives a lower bound for the free energy, because of the Gibbs variational principle. If the Bragg–Williams approximation is successful, maximizing over m yields the true free energy (at least to leading order). If applied to approximate the free energy of the Curie-Weiss Hamiltonian $\frac{\beta}{N} \sum_{i,j} \sigma_i \sigma_j + h \sum_{i=1}^N \sigma_i$ one obtains a variational problem over $m \in \mathbb{R}^N$ that is equivalent to

$$\frac{1}{N} \sup_{\bar{m} \in [-1,1]} \left\{ \beta \bar{m}^2 + h\bar{m} - \frac{1+\bar{m}}{2} \log\left(\frac{1-\bar{m}}{2}\right) - \frac{1-\bar{m}}{2} \log\left(\frac{1+\bar{m}}{2}\right) \right\} ,$$

which also appears in the classical solution of the model via the large deviation rate function of the binomial distribution, and is thus indeed an accurate approximation.

In the spherical setting a product measure on the spins is not absolutely continuous with respect to E, but a natural family of measures is provided by exponential tilts of the uniform distribution given by $e^{\lambda\sigma\cdot m} dE$ appropriately normalized, for $\lambda = \lambda(m)$ chosen so that the mean magnetization is m. For such a measure the internal energy will be close to $\beta H_N(m) + Nm \cdot h_N$ and the entropy will be close to $\frac{N}{2}\log\left(1 - |m|^2\right)$. Thus the Bragg–Williams approximation of the free energy is

$$\frac{1}{N} \sup_{m\in\mathbb{R}^N:|m|<1} \left\{\beta H_N(m) + Nm \cdot h_N + \frac{N}{2}\log\left(1 - |m|^2\right)\right\},$$

which is in fact inaccurate, in light of (1.4). However, the TAP–Plefka variational principle can be seen as the appropriate modification of the Bragg–Williams approximation to obtain an accurate approximation for this disordered system, by adding the Onsager correction term $\frac{N}{2}\beta^2\left(1 - |m|^2\right)^2$ and restricting the sup to m-s satisfying Plefka's condition.

1.1.2. The 2*-spin model* The 2-spin spherical SK model, which is the model we consider in this paper, is a much simpler model than the other Ising and spherical SK variants. It is always replica symmetric, for all inverse temperatures β and external field strengths h, and the Parisi formula can be written as a one parameter variational principle (see (1.7)). If the external field vanishes ($h = 0$) an explicit closed form (non-variational) formula for $\lim_{N\to\infty} F_N$ exists, even in low temperature.

Furthermore, the Hamiltonian can be written as $H_N(\sigma) = \sqrt{N}\sigma^T S_N \sigma$ for a random matrix S_N from the Gaussian orthogonal ensemble, and by the rotational invariance of the sphere we can work in the diagonalizing basis of S_N, in which case $H_N(\sigma) = \sqrt{N}\sum_{i=1}^N \lambda_i \sigma_i^2$ where λ_i are the eigenvalues of S_N. Because of this the free energy can be computed by a random matrix approach, without using the Parisi formula [4,13,15]. Part of our analysis also relies on random matrix considerations. The resulting formulas (1.5) and (1.6) are not related to previously obtained formulas for the free energy. Our proof is the first rigorous derivation of a TAP variational principle based on a microcanonical analysis that yields bounds valid for finite N, and where Plefka's condition appears naturally.

1.1.3. Previous work in the mathematical literature Recently in [10] Chen and Panchenko used the Parisi formula to verify a TAP variational principle for mixed Ising SK models in the thermodynamic limit, that is an equality after taking the limit $N \to \infty$, with a different condition replacing Plefka's condition. In [22] Subag constructs for very low temperatures a decomposition of the Gibbs measure of pure p-spin spherical models into pure states in a microcanonical fashion, and notices that the log of the weight of each pure state coincides with its TAP free energy.

Further mathematical results concern the TAP equations. These are a system of nonlinear equations for the quenched mean magnetization which have been interpreted as a self-consistency property; within our framework it is natural to view the TAP equation as the critical point equations of the TAP free energy. Bolthausen has developed an iterative scheme for solving the TAP equations for the Ising SK model [7] that converges in the whole conjectured high temperature regime. Talagrand [23] and Chatterjee [9] showed that in high enough temperature the mean magnetization of the Ising SK Gibbs measure satisfies the TAP equations. Auffinger and Jagganath have used the Parisi formula to prove that solutions of the TAP equations describe the magnetization inside appropriately defined pure states of generic mixed Ising models for all temperatures [3]. Auffinger, Ben Arous & Cerny have studied the (annealed) complexity of TAP solutions for pure p-spin spherical Hamiltonians [2].

1.2. A word on the proof. The proof of Theorem 1 splits into a proof of a lower bound and a proof of an upper bound for the partition function $Z_N(\beta, h_N)$. Both are based on recentering the Hamiltonian around magnetizations m of potential pure states (a similar recentering has been used by TAP [27], Bolthausen [6] and Subag [22]). In general, recentering around a given m gives rise to an effective external field for the recentered Hamiltonian.

The lower bound is presented in Sect. 3 and is proved by considering a recentering around any magnetization m that satisfies Plefka's condition. We then restrict the integral in $Z_N(\beta, h_N)$ to a subset of the sphere which is "centered at m", namely the intersection of the sphere with a plane that contains m and is perpendicular to both m and the effective external field. The mean energy (value of Hamiltonian and external field) on this subset is $\beta H_N(m) + Nm \cdot h_N$, cf. the first two terms of $H_{TAP}(m)$. The log of the measure of the subset is approximately $\frac{N}{2} \log\left(1 - |m|^2\right)$, cf. the third term of $H_{TAP}(m)$. Finally, the recentered Hamiltonian on this subset turns out to be a 2-spin Hamiltonian on a lower dimensional sphere without external field at inverse temperature $\beta(m) = \beta\left(1 - |m|^2\right)$. If Plefka's condition is satisfied this is less than the critical inverse temperature $\beta_c = \frac{1}{\sqrt{2}}$, and it is therefore natural that using the uniform measure on the subset as a reference measure the free energy of the recentered Hamiltonian is $\frac{1}{2}\beta(m)^2 = \frac{1}{2}\beta^2(1 - |m|)^2$, cf. the last term of $H_{TAP}(m)$ (the Onsager correction). In this way we show that the subset contributes approximately $\exp(H_{TAP}(m))$ to $Z_N(\beta, h)$. This shows that $H_{TAP}(m)$ is a lower bound of the free energy for any m satisfying Plefka's condition. Note that it also provides a natural interpretation of the terms in $H_{TAP}(m)$, and of Plefka's condition as the condition that a pure state should effectively be in high temperature.

The upper bound is significantly harder and is proved in Sect. 4. It involves the construction of a low-dimensional subspace of magnetizations $\mathcal{M}_N$ with the property that after recentering around any $m \in \mathcal{M}_N$, the effective external field is again almost completely contained in $\mathcal{M}_N$. We write the integral in $Z_N(\beta, h_N)$ as a double integral first over $\mathcal{M}_N$ and then over the perpendicular space $\mathcal{M}_N^{\perp}$. For a fixed $m \in \mathcal{M}_N$ the integral over the perpendicular space $\mathcal{M}_N^{\perp}$ is seen to be related to a partition function without external field at a higher effective temperature, and is shown to be close to the exponential of a modified TAP energy, with the Onsager correction $\frac{\beta^2}{2}\left(1 - |m|^2\right)^2$ replaced by a different, not entirely explicit, expression. The integral in $Z_N(\beta, h_N)$ thus reduces to an integral of the exponential of the modified TAP energy over the low-dimensional space $\mathcal{M}_N$, and by the Laplace method the log of the integral turns into the supremum over the modified TAP energy over *all* m. We then show that if the Hessian at a critical point of the modified TAP energy is negative semi-definite, as it must be at any local maximum, then m satisfies Plefka's condition and furthermore the modified TAP energy and the original TAP energy $H_{TAP}(m)$ are close. From this the upper bound on $Z_N(\beta, h_N)$ is seen to follow.

In Sect. 5 we prove Lemma 2. In the next section we fix notation and recall some basic facts.

2. Notation and Basic Facts

The letter c denotes a constant that does not depend on N, possibly a different one each time it is used.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with random variables $J_{ij}, i, j \geq 1$ that are iid standard Gaussians. Define

$$H_N(\sigma) = \sqrt{N} \sum_{i,j=1}^{N} J_{i,j}\sigma_i\sigma_j \text{ for } \sigma \in \mathbb{R}^N.$$

For any $\lambda \in \mathbb{R}$ and $\sigma \in \mathbb{R}^N$ we have

$$H_N(\lambda\sigma) = \lambda^2 H_N(\sigma). \tag{2.1}$$

Let S_N be the $N \times N$ matrix given by

$$(S_N)_{ij} = \frac{J_{ij} + J_{ji}}{2}.$$

Note that

$$H_N(\sigma) = \sqrt{N}\sigma^T S_N \sigma,$$

and

$$\nabla H_N(\sigma) = 2\sqrt{N} S_N \sigma.$$

For this reason the 2-spin Hamiltonian gradient is linear, i.e.

$$\nabla H_N(\sigma_1 + \sigma_2) = \nabla H_N(\sigma_1) + \nabla H_N(\sigma_2) \quad \text{for all } \sigma_1, \sigma_2 \in \mathbb{R}^N. \tag{2.2}$$

We will use, especially in the upper bound, that the empirical spectral distribution of S_N converges to the semi-circle law. Let $\sqrt{N}\theta_1^N < \cdots < \sqrt{N}\theta_N^N$ be the eigenvalues of the matrix S_N. We have that

$$\frac{1}{N}\sum_{i=1}^{N} \delta_{\theta_i^N} \to \mu(x)\,dx \text{ in distribution, } \mathbb{P} - a.s.,$$

where

$$\mu(x) = \frac{1}{\pi}\sqrt{2 - x^2}\,1_{\left[-\sqrt{2},\sqrt{2}\right]}. \tag{2.3}$$

In addition if we let

$$\theta_u = \inf\left\{\theta : \int_{-\sqrt{2}}^{\theta} \mu(x)\,dx = u\right\}, \tag{2.4}$$

then

$$\theta_i^N = \theta_{\frac{i}{N}} + o(1) \text{ for } i = 1, \ldots, N, \tag{2.5}$$

where the $o(1)$ terms tend to zero $\mathbb{P}$-a.s. uniformly in i (see e.g. Theorem 2.9 [5]).

For instance from the fact the eigenvalue of largest magnitude is of order $\sqrt{N}$ one can deduce that

$$\sup_{\sigma \in \mathbb{R}^N : |\sigma| \leq 1} |H_N(\sigma)| \leq cN \quad \text{and} \quad \sup_{\sigma \in \mathbb{R}^N : |\sigma| \leq 1} |\nabla H_N(\sigma)| \leq cN, \tag{2.6}$$

for all N large enough, almost surely. The latter implies that

$$\left|H_N\left(\sigma^1\right) - H_N\left(\sigma^2\right)\right| \leq cN\left|\sigma^1 - \sigma^2\right| \quad \text{for all } \sigma^i \in \mathbb{R}^N, \left|\sigma^i\right| \leq 1, i = 1, 2. \quad (2.7)$$

We let E^M denote the uniform measure on the unit sphere of $\mathbb{R}^M$. When $M = N$ we drop the superscript and write E. If $\mathcal{U}$ is a linear subspace of $\mathbb{R}^N$ we let $E^{\mathcal{U}}$ denote the uniform measure on the unit sphere of $\mathbb{R}^N$ intersected with $\mathcal{U}$.

The surface area of the N-dimensional sphere of radius r is $\frac{2\pi^{\frac{N}{2}}}{\Gamma\left(\frac{N}{2}\right)} r^{N-1}$, and for any unit vector v the inner product $\sigma \cdot v$ has a density under E given by

$$E\left[\sigma \cdot v = dx\right] = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)} \left(1 - x^2\right)^{\frac{N-3}{2}} dx. \quad (2.8)$$

More generally, for any linear subspace $\mathcal{U} \subset \mathbb{R}^N$ of dimension M the the projection $\tilde{\sigma}$ of σ onto $\mathcal{U}$ has density

$$E\left[d\tilde{\sigma}\right] = \frac{1}{\pi^{\frac{M}{2}}} \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-M}{2}\right)} \left(1 - |\tilde{\sigma}|^2\right)^{\frac{N-M-2}{2}} d\tilde{\sigma}, \quad (2.9)$$

with respect to the standard Lebesgue measure on $\mathbb{R}^N$ restricted to $\mathcal{U}$.

3. Lower Bound

In this section we show the following lower bound for the free energy.

Proposition 3. *For $\beta, h, h_1, h_2, \ldots$ as in Theorem 1 one has*

$$F_N\left(\beta, h_N\right) \geq \frac{1}{N} \sup_{m \in \mathbb{R}^N : |m| < 1, \beta(m) \leq \frac{1}{\sqrt{2}}} H_{TAP}(m) + o(1), \quad (3.1)$$

where the $o(1)$ term tends to zero $\mathbb{P}$-a.s.

We prove this by noting that the partition function is certainly larger than the integral of $e^{\beta H_N(\sigma) + N\sigma \cdot h_N}$ over a slice $\left\{\sigma : \left|\sigma \cdot m - |m|^2\right| < \varepsilon\right\}$ for any m inside the unit ball and $\varepsilon > 0$. On this slice we recenter the spins

$$\hat{\sigma} = \sigma - m,$$

and use the decomposition

$$H_N(\sigma) = H_N(m) + \nabla H_N(m) \cdot \hat{\sigma} + H_N\left(\hat{\sigma}\right), \quad (3.2)$$

which holds deterministically, to note that the integral over the slice is essentially the partition function of a 2-spin Hamiltonian on an $N - 1$-dimensional sphere of radius $1 - |m|^2$ with mean $\beta H_N(m)$ and external field $\beta \nabla H_N(m) + N h_N$. By further restricting the integral to a subspace where the external field vanishes the Onsager correction term $\frac{1}{2}\beta^2\left(1 - |m|^2\right)^2$ of the TAP free energy arises as the free energy of the partition function of this recentered Hamiltonian without external field. Plefka's condition arises as the condition that the recentered Hamiltonian is in high temperature.

By the second moment method and concentration of measure one can show the following.

Lemma 4. *It holds that*

$$\sup_{\beta\in\left[0,\frac{1}{\sqrt{2}}\right]}\left|\frac{1}{N}\log E\left[\exp\left(\beta H_N\left(\sigma\right)\right)\right]-\frac{\beta^2}{2}\right|\to 0,\quad \mathbb{P}-a.s. \tag{3.3}$$

It will be important to consider the partition function restricted to the intersection of the unit sphere with a hyperplane of dimension $N-2$ (or $N-1$). The next lemma shows that (3.3) remains true uniformly over all such restrictions. Recall that $E^{\langle u,v\rangle^\perp}$ denotes the uniform measure on the unit sphere in the subspace $\langle u,v\rangle^\perp$ perpendicular to u and v.

Lemma 5. *We have*

$$\sup_{\beta\in\left[0,\frac{1}{\sqrt{2}}\right],u,v\in\mathbb{R}^N}\left|\frac{1}{N}\log E^{\langle u,v\rangle^\perp}\left[\exp\left(\beta H_N\left(\sigma\right)\right)\right]-\frac{\beta^2}{2}\right|\to 0,\quad \mathbb{P}\text{-}a.s. \tag{3.4}$$

Proof. Recall that $H_N\left(\sigma\right)=\sqrt{N}\sigma^T S_N\sigma$ where S_N is a real symmetric matrix. For any $u,v\in\mathbb{R}^N$ that are linearly independent, let $w_1,\ldots,w_N$ be an orthonormal basis such that $\langle u,v\rangle=\langle w_{N-1},w_N\rangle$, and let A be the top left $(N-2)\times(N-2)$ minor of S_N when written in basis $w_1,\ldots,w_N$. For $\sigma\in\langle u,v\rangle^\perp$ we have $H_N\left(\sigma\right)=\sqrt{N}\tilde{\sigma}^T A\tilde{\sigma}$ where $\tilde{\sigma}=(\sigma_1,\ldots,\sigma_{N-2})\in\mathbb{R}^N$. Let $\sqrt{N}a_1,\ldots,\sqrt{N}a_{N-2}$ be the eigenvalues of A. Then

$$E^{\langle u,v\rangle^\perp}\left[\exp\left(\beta H_N\left(\sigma\right)\right)\right]=E^{N-2}\left[\exp\left(N\beta\sum_{i=1}^{N-2}a_i\sigma_i^2\right)\right]. \tag{3.5}$$

Let B be the top left $(N-2)\times(N-2)$ minor of S_N when written in the standard basis and let $\sqrt{N}b_1,\ldots,\sqrt{N}b_{N-2}$ be its eigenvalues. Note that $H_{N-2}\left(\sigma\right)=\sqrt{N-2}\sigma^T B\sigma$ for $\sigma\in\mathbb{R}^{N-2}$, and by (3.3) with $N-2$ in place of N we have

$$E^{N-2}\left[\exp\left(\sqrt{N}\beta\sigma^T B\sigma\right)\right]=e^{N\left(\frac{\beta^2}{2}+o(1)\right)}, \tag{3.6}$$

where the $o\left(1\right)$ term tends to zero almost surely. Also

$$E^{N-2}\left[\exp\left(\sqrt{N}\beta\sigma^T B\sigma\right)\right]=E^{N-2}\left[\exp\left(N\beta\sum_{i=1}^{N-2}b_i\sigma_i^2\right)\right]. \tag{3.7}$$

Let $\theta_1^N,\ldots,\theta_N^N$ be the eigenvalues of S_N. By the eigenvalue interlacing inequality (see e.g. Exercise 1.3.14 [26])

$$\theta_i^N\le a_i,b_i\le\theta_{i+2}^N \text{ for } i=1,\ldots,N-2,$$

so by (2.5) we have

$$\sup_{i=1,\ldots,N}|a_i-b_i|\to 0 \text{ a.s., as } N\to\infty.$$

Therefore

$$\sup_{\sigma \in S_{N-2}} \left| \beta \sum_{i=1}^{N-2} a_i \sigma_i^2 - \beta \sum_{i=1}^{N-2} b_i \sigma_i^2 \right| \to 0 \text{ a.s., as } N \to \infty,$$

so from (3.5), (3.6) and (3.7) it follows that

$$E^{\langle u,v\rangle^\perp} \left[\exp\left(\beta H_N(\sigma)\right)\right] = e^{N \frac{\beta^2}{2}(1+o(1))},$$

uniformly over all linearly independent u, v, where the $o(1)$ terms tend to zero almost surely. The above argument but with $(N-1) \times (N-1)$ minors easily extends this to u and v that are linearly dependent. This proves (3.4). □

We can now prove the lower bound Proposition 3.

Proof of Proposition 3. For any m and σ, recenter the spins σ around m by letting $\hat{\sigma} = \sigma - m$. Recentering the Hamiltonian (see (3.2)) and the external field one obtains

$$\beta H_N(\sigma) + N\sigma \cdot h_N = \beta H_N(m) + Nm \cdot h_N + Nh^m \cdot \hat{\sigma} + \beta H_N\left(\hat{\sigma}\right), \tag{3.8}$$

where

$$h^m = \frac{\beta}{N} \nabla H_N(m) + h_N, \tag{3.9}$$

is the effective external field after recentering. Note that by our assumption $|h_N| = h$ and (2.6) we have that for N large enough

$$\left|h^m\right| \le c, \tag{3.10}$$

for a constant c depending only on β and h.

Fix an $m \in \mathbb{R}^N$ with $|m| < 1$. Let v_1, v_2 be basis vectors of an arbitrary two dimensional linear subspace of $\mathbb{R}^N$ that contains m and h^m. For $\varepsilon > 0$ to be fixed later consider

$$A = \left\{ \sigma : \hat{\sigma} \cdot v_i \in (-\varepsilon, \varepsilon), i = 1, 2 \right\}. \tag{3.11}$$

Note that for $\sigma \in A$

$$\left|\hat{\sigma} \cdot m\right| \le c\varepsilon \text{ and } \left|\hat{\sigma} \cdot h^m\right| \le c\varepsilon, \tag{3.12}$$

(the latter constant depends on the one in (3.10)) and

$$\left|\hat{\sigma}\right|^2 = |\sigma|^2 - |m|^2 - 2\hat{\sigma} \cdot m = 1 - |m|^2 + O(\varepsilon). \tag{3.13}$$

Certainly we have

$$Z_N(\beta, h_N) \ge E\left[1_A \exp\left(\beta H_N(\sigma) + N\sigma \cdot h_N\right)\right].$$

Rewriting in terms of $\hat{\sigma}$ and using (3.8) and the second inequality of (3.12) the right hand-side can be bounded below by

$$\exp\left(\beta H_N(m) + Nm \cdot h_N - c\varepsilon N\right) E\left[1_A \exp\left(\beta H_N\left(\hat{\sigma}\right)\right)\right]. \tag{3.14}$$

Let $\gamma\sigma^\perp$ be the projection of $\hat{\sigma}$ onto the hyperplane $\langle v_1, v_2\rangle^\perp$, where $\sigma^\perp$ is a unit vector and $\gamma \in \mathbb{R}$ is the magnitude of the projection. From (3.11) we have for $\sigma \in A$

$$\left|\hat{\sigma} - \gamma\sigma^\perp\right| \le c\varepsilon, \tag{3.15}$$

so that by (3.13)

$$\gamma^2 = 1 - |m|^2 + O(\varepsilon). \tag{3.16}$$

Using (3.15) and (2.6), (2.7) we have

$$H_N(\hat{\sigma}) = H_N\left(\gamma\sigma^{\perp}\right) + O(\varepsilon N),$$

and by (2.1), (2.6) and (3.16)

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

This gives that (3.14) is at least

$$\exp\left(\beta H_N(m) + Nm \cdot h_N - c\varepsilon N\right) E\left[1_A \exp\left(\beta\left(1 - |m|^2\right) H_N\left(\sigma^{\perp}\right)\right)\right]. \tag{3.17}$$

Now $\sigma^{\perp}$ is independent of $\sigma \cdot m, \sigma \cdot h^m$ under E, and is uniform on the unit sphere intersected with $\langle v_1, v_2\rangle^{\perp}$. Therefore (3.17) in fact equals

$$\exp\left(\beta H_N(m) + Nm \cdot h_N - c\varepsilon N\right) E[A]\, E^{\langle v_1, v_2\rangle^{\perp}}\left[\exp\left(\beta\left(1 - |m|^2\right) H_N(\sigma)\right)\right].$$

Using (2.9) with $M = 2$ and (3.13) and it holds that

$$E[A] \geq Nc\varepsilon^2\left(1 - |m|^2 - c\varepsilon\right)^{\frac{N-4}{2}},$$

and setting e.g. $\varepsilon = \frac{1}{\sqrt{N}}$ this equals

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

Thus, Z_N is at least

$$\exp\left(\beta H_N(m) + Nm \cdot h_N + \tfrac{N}{2}\log\left(1 - |m|^2\right) + o(N)\right) \\ \times E^{\langle v_1, v_2\rangle^{\perp}}\left[\exp\left(\beta\left(1 - |m|^2\right) H_N(\sigma)\right)\right],$$

for any m with $|m| < 1$, where the error term is $o(N)$ uniformly in m, almost surely.

By Lemma 5 this is in turn at least

$$\exp\left(\beta H_N(m) + Nm \cdot h_N + \frac{N}{2}\log\left(1 - |m|^2\right) + N\frac{\beta^2}{2}\left(1 - |m|^2\right)^2 + o(N)\right), \tag{3.18}$$

provided

$$\beta\left(1 - |m|^2\right) \leq \frac{1}{\sqrt{2}}, \quad \text{i.e. } \beta(m) \leq \frac{1}{\sqrt{2}}, \tag{3.19}$$

where the error term is $o(N)$ almost surely, uniformly in m that satisfy (3.19). Since (3.18) equals $\exp\left(H_{TAP}(m) + o(N)\right)$ the claim (3.1) follows. □

4. Upper Bound

In this section we prove the following upper bound on the free energy.

Proposition 6. *For $\beta, h, h_1, h_2, \ldots$ as in Theorem 1 one has*

$$F_N(\beta, h_N) \leq \frac{1}{N} \sup_{m \in \mathbb{R}^N : |m| < 1, \beta(m) \leq \frac{1}{\sqrt{2}}} H_{TAP}(m) + o(1), \tag{4.1}$$

where the $o(1)$ term tends to zero $\mathbb{P}$-a.s.

As for the lower bound, our proof is based on considering the Hamiltonian recentered around certain m-s inside the unit ball. However, for an upper bound we are not free to simply restrict the integral in the partition function to slices around an m and ignore the complement. Neither can we further restrict the integral inside the slice to a space where the effective external field vanishes. Lastly we can not ignore slices for which Plefka's condition is not satisfied.

We get around these issues by constructing a low-dimensional subspace $\mathcal{M}_N$ of m-s, such that the recentered Hamiltonian restricted to the space of configurations perpendicular to $\mathcal{M}_N$ has almost vanishing external field for *any* $m \in \mathcal{M}_N$, without further restriction. Because the dimension of $\mathcal{M}_N$ is $o(N)$ we are able to use the Laplace method to upper bound the free energy by a sup of the free energy contribution of each of these restricted Hamiltonians. Lastly, a coarse-graining of the recentered Hamiltonian gives a sequence of approximations to the free energy of the restricted Hamiltonians in a form that allows to show that the supremum must be attained at an m that satisfies Plefka's condition.

4.1. Diagonalization. To prove the upper bound Proposition 6 we are obliged to make stronger use the diagonalized Hamiltonian

$$N \sum_{i=1}^{N} \theta_i^N \sigma_i^2, \tag{4.2}$$

and the semi-circle law. Let

$$\tilde{h}_N \text{ be the vector } h_N \text{ written in the diagonalizing basis of the matrix } S_N. \tag{4.3}$$

By rotational symmetry we have

$$F_N(\beta, h_N) = \frac{1}{N} \log E\left[\exp\left(N\beta \sum_{i=1}^{N} \theta_i^N \sigma_i^2 + N\tilde{h}_N \cdot \sigma\right)\right].$$

For convenience we also replace the diagonalized Hamiltonian (4.2) by its deterministic counterpart

$$\tilde{H}_N(\sigma) = N \sum_{i=1}^{N} \theta_{i/N} \sigma_i^2,$$

where each random eigenvalue θ_i^N is replaced by its deterministic typical position $\theta_{i/N}$ (recall (2.4)). The error made is controlled by (2.5), giving

$$\lim_{N\to\infty} \frac{1}{N} \sup_{\sigma:|\sigma|=1} \left| N \sum_{i=1}^{N} \theta_i^N \sigma_i^2 - \tilde{H}_N(\sigma) \right| = 0, \quad \mathbb{P} - a.s. \tag{4.4}$$

Let

$$\tilde{F}_N(\beta, h_N) = \frac{1}{N} \log E\left[\exp\left(N\beta \sum_{i=1}^{N} \theta_{i/N}\sigma_i^2 + N\tilde{h}_N \cdot \sigma\right)\right]$$

and let

$$\tilde{H}_{TAP}(m) = \beta \tilde{H}_N(m) + Nm \cdot \tilde{h}_N + \frac{N}{2}\log\left(1 - |m|^2\right) + N\frac{\beta^2}{2}\left(1 - |m|^2\right)^2. \tag{4.5}$$

By (4.4) the upper bound Proposition 6 follows from the following deterministic bound.

Proposition 7. *For $\beta, h, h_1, h_2, \ldots$ as in Theorem 1 one has*

$$\tilde{F}_N(\beta, h_N) \le \frac{1}{N} \sup_{m\in\mathbb{R}^N:|m|<1, \beta(m)\le\frac{1}{\sqrt{2}},} \tilde{H}_{TAP}(m) + o(1). \tag{4.6}$$

The rest of this section is devoted to the proof of Proposition 7.

4.2. Free energy of coarse-grained Hamiltonian without external field. We will approximate $\tilde{H}_N(\sigma)$ by a coarse-grained Hamiltonian where the $\theta_{i/N}$ are replaced by a bounded number of distinct coefficients. For such a Hamiltonian it will be straightforward to bound the free energy using the Laplace method. To this end consider for each $K \ge 2$ equally spaced numbers $x_1, \ldots, x_K$ in $\left[-\sqrt{2}, \sqrt{2}\right]$, so that,

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

and a partition $I_1, \ldots, I_K$ of $\{1, \ldots, N\}$ given by

$$I_k = \left\{i : x_k \le \theta_{i/N} < x_{k+1}\right\}, k = 1, \ldots, K-1 \quad \text{and} \quad I_K = \left\{i : x_K \le \theta_{i/N}\right\}. \tag{4.7}$$

Let

$$\sigma_{[k]}^2 = \sum_{i\in I_k} \sigma_i^2 \quad \text{and} \quad \mu_k = \frac{|I_k|}{N}. \tag{4.8}$$

The next lemma gives the density of the vector $\left(\sigma_{[1]}^2, \ldots \sigma_{[K-1]}^2\right)$ under E.

Lemma 8. *The E-distribution of the vector $\left(\sigma_{[1]}^2, \ldots \sigma_{[K-1]}^2\right)$ has a density on $\mathbb{R}^{K-1}$ with respect to Lebesgue measure given by*

$$\Gamma\left(\frac{N}{2}\right) \prod_{k=1}^{K} \frac{\rho_k^{\frac{|I_k|-2}{2}}}{\Gamma\left(\frac{|I_k|}{2}\right)} 1_A d\rho_1 \ldots d\rho_{K-1}, \tag{4.9}$$

where we write $\rho_K = 1 - \rho_1 - \cdots - \rho_{K-1}$ and $A = \{\rho_1, \ldots, \rho_{K-1} \ge 0, \rho_1 + \cdots + \rho_{K-1} \le 1\}$.

Proof. One can sample the random variable σ with law E by sampling from the standard Gaussian distribution on $\mathbb{R}^N$ and normalizing the result. Therefore $\left(\sigma_{[1]}^2, \ldots \sigma_{[K-1]}^2\right)$ has the same law as $(R_1, \ldots, R_{K-1})$, where

$$R_i = \frac{X_i}{X_1 + \cdots + X_K}, i \leq K,$$

the X_i are independent, and X_i has the χ^2-distribution with $|I_k|$ degrees of freedom, i.e. has density $\frac{1}{2^{|I_k|/2}\Gamma(|I_k|/2)} x_i^{\frac{|I_k|-2}{2}} e^{-\frac{x_k}{2}} 1_{\{x_k \geq 0\}} dx_k$. We now let $Z = X_1 + \cdots + X_K$ and make the change of variables $x_i = z\rho_i, i = 1, \ldots, K-1$ which has Jacobian z^{k-1} to obtain that $(R_1, \ldots, R_{K-1}, Z)$ has density

$$\begin{aligned} &1_A 1_{\{z\geq 0\}} \left(\prod_{k=1}^K \frac{1}{2^{|I_k|/2}\Gamma\left(\frac{|I_k|}{2}\right)} (z\rho_i)^{\frac{|I_k|-2}{2}} e^{-\frac{z\rho_k}{2}} \right) z^{K-1} d\rho_1 \ldots d\rho_{k-1} dz \\ &= \left(1_A \prod_{k=1}^K \frac{\rho_i^{\frac{|I_k|-2}{2}}}{\Gamma\left(\frac{|I_k|}{2}\right)} \right) \left(\frac{1}{2^{N/2}} 1_{\{z\geq 0\}} z^{\frac{N-2}{2}} e^{-\frac{z}{2}} dz \right) d\rho_1 \ldots d\rho_{k-1}. \end{aligned}$$

Since

$$\int \frac{1}{2^{N/2}} z^{\frac{N-2}{2}} e^{-\frac{z}{2}} dz = \Gamma\left(\frac{N}{2}\right),$$

integrating out z to get the marginal of $(R_1, \ldots, R_{K-1})$ one obtains (4.9). □

We first show the following variational principle for the free energy of the coarse-grained Hamiltonians in the absence of an external field.

Lemma 9. *For all* $C > 0$ *we have uniformly in* $0 < \beta \leq C$, large enough K and $N \geq c(K)$ that

$$\begin{aligned} &\frac{1}{N} \log E\left[\exp\left(N\beta \sum_{k=1}^K x_k \sigma_{[k]}^2\right)\right] \\ &= \sup_{0 \leq f_k, f_1+\cdots+f_K=1} \left\{ \beta \sum_{k=1}^K x_k f_k + \frac{1}{2} \sum \mu_k \log \frac{f_k}{\mu_k} \right\} + O\left(\frac{K^3 \log N}{N}\right). \end{aligned} \tag{4.10}$$

Proof. By Lemma 8 the integral $E\left[\exp\left(N\beta \sum_{k=1}^K x_k \sigma_{[k]}^2\right)\right]$ equals

$$\begin{aligned} &\frac{\Gamma\left(\frac{N}{2}\right)}{\prod_{k=1}^K \Gamma\left(\frac{|I_k|}{2}\right)} \int_{[0,1]^{K-1}} 1_A \exp\left(N\left\{\beta \sum_{k=1}^K x_k \rho_k + \sum_{k=1}^K \frac{1}{2}\left(\mu_k - \frac{2}{N}\right) \log \rho_k\right\}\right) \\ &\times d\rho_1 \ldots d\rho_{K-1}. \end{aligned} \tag{4.11}$$

By the Laplace method the integral in (4.11) is at most

$$\exp\left(N\left\{\sup_{0 \leq f_k, f_1+\cdots+f_K=1} \left\{\beta \sum_{k=1}^K x_k f_k + \frac{1}{2} \sum \left(\mu_k - \frac{2}{N}\right) \log f_k\right\}\right\}\right). \tag{4.12}$$

To get rid of the nuisance term $\frac{2}{N}$ we use the following ad-hoc argument. For any maximizer of the sup the value must exceed $-\sqrt{2}\beta - \frac{1}{2}\log K$, since one obtains at least this by setting $f_1 = \cdots = f_K = \frac{1}{K}$. Note that $\mu_k \geq \frac{c}{K^{3/2}}$ for all k, K, by (2.3), so for $N \geq c(K)$ also $\mu_k - \frac{2}{N} \geq \frac{c}{K^{3/2}}$. Assume now that $f_k \leq e^{-K^2}$ from some k. Then $\beta \sum_{k=1}^{K} x_k f_k + \frac{1}{2}\sum\left(\mu_k - \frac{2}{N}\right)\log f_k \leq \sqrt{2}\beta - c\frac{K^2}{K^{3/2}} < -\sqrt{2}\beta - \frac{1}{2}\log K$, for K large enough. So for K large enough and $N \geq c(K)$ any maximizer in the sup above must satisfy $f_k \geq e^{-K^2}$. But for such f_k the nuisance term contributes at most $\frac{K^3}{N}$. Therefore (4.12) equals

$$\exp\left(N\left\{\sup_{0\leq f_k, f_1+\cdots+f_K=1}\left\{\beta\sum_{k=1}^{K} x_k f_k + \frac{1}{2}\sum \mu_k \log f_k\right\} + O\left(\frac{K^3}{N}\right)\right\}\right).$$

Using the bounds $\Gamma(x) \asymp \sqrt{2\pi x}\,(x/e)^x$ for $x \geq \frac{1}{2}$ and $1 \leq \prod_{k=1}^{K}|I_k| \leq N^K$, one sees that $\frac{1}{N}$ log of the factor multiplying the integral in (4.11) equals

$$\begin{aligned}
&\frac{1}{N}\log\frac{\Gamma\left(\frac{N}{2}\right)}{\prod_{k=1}^{K}\Gamma\left(\frac{|I_k|}{2}\right)}\\
&= \frac{1}{N}\log\frac{\left(\frac{N}{2}\right)^{\frac{N}{2}}}{\prod_{k=1}^{K}\left(\frac{|I_k|}{2}\right)^{\frac{|I_k|}{2}}} + \frac{1}{N}\log\frac{\sqrt{2\pi\frac{N}{2}}}{\prod_{k=1}^{K}\sqrt{2\pi\frac{|I_k|}{2}}} + O\left(\frac{K}{N}\right)\\
&= -\frac{1}{2}\sum \mu_k \log \mu_k + O\left(\frac{K\log N}{N}\right).
\end{aligned}$$

This completes the proof. □

The variational problem on the bottom line of (4.10) can be solved. To state the result let

$$g_K(\lambda) = \sum_{k=1}^{K}\frac{\mu_k}{\lambda - x_k}. \tag{4.13}$$

For all $\beta > 0$ there is a unique $\lambda_K(\beta) > x_K$ such that

$$g_K(\lambda_K(\beta)) = 2\beta. \tag{4.14}$$

Let

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

and

$$\mathcal{F}_K(\beta) = \beta\lambda_K(\beta) - \frac{1}{2} - \frac{1}{2}\log(2\beta) - \frac{1}{2}h_K(\lambda_K(\beta)). \tag{4.15}$$

The next lemma shows that $\mathcal{F}_K(\beta)$ is the supremum in the variational problem from (4.10).

Lemma 10. *For each K and $\beta > 0$ we have*

$$\sup_{0\leq f_k, f_1+\cdots+f_K=1}\left\{\beta\sum_{k=1}^{K}x_k f_k + \frac{1}{2}\sum\mu_k\log\frac{f_k}{\mu_k}\right\} = \mathcal{F}_K(\beta). \tag{4.16}$$

Proof. Since the quantity being maximized tends to $-\infty$ if $f_k \to 0$ for some k there must be a global maximum satisfying $f_k > 0$ for all k. Using Lagrange multipliers to solve the constrained optimization problem one considers

$$\mathcal{L}(f_1, \ldots, f_k, \lambda) = \beta \sum_{k=1}^{K} x_k f_k + \frac{1}{2} \sum \mu_k \log \frac{f_k}{\mu_k} + \tilde{\lambda} \left(\sum_{k=1}^{K} f_k - 1 \right).$$

If $(f_1, \ldots, f_K)$ is a global maximum then there must be a $\tilde{\lambda}$ such that $\left(f_1, \ldots, f_K, \tilde{\lambda}\right)$ is a critical point of $\mathcal{L}$. The critical point equations of $\mathcal{L}$ read

$$\beta x_k - \frac{1}{2} \frac{1}{f_k} + \tilde{\lambda} = 0, k = 1, \ldots, K, \quad \text{and} \quad \sum_{k=1}^{K} f_k - 1 = 0.$$

The first K equations are equivalent to

$$f_k = \frac{1}{2\beta} \frac{\mu_k}{\lambda - x_k}, k = 1, \ldots, K, \tag{4.17}$$

where we reparameterized $\lambda = -\beta\tilde{\lambda}$. Therefore for some $\lambda > x_K$ it holds that

$$\sum_{k=1}^{K} f_k = 1,$$

for the f_k in (4.17) and these f_k maximize (4.16). Inspection of (4.13)–(4.14) reveal that $\lambda = \lambda_K(\beta)$ is the unique such λ. When f_k take the form in (4.17) then

$$\sum_{k=1}^{K} x_k f_k = \sum_{k=1}^{K} x_k \frac{1}{2\beta} \frac{\mu_k}{\lambda - x_k} = \frac{1}{2\beta} \left(\lambda \sum_{k=1}^{K} \frac{\mu_k}{\lambda - x_k} - 1 \right) = \lambda - \frac{1}{2\beta}, \tag{4.18}$$

and

$$\frac{1}{2} \sum \mu_k \log \frac{f_k}{\mu_k} = \frac{1}{2} \sum \mu_k \log \frac{1}{2\beta} \frac{1}{\lambda - x_k} = -\frac{1}{2} \log(2\beta) - \frac{1}{2} h_K(\lambda).$$

Therefore, the value of the quantity being maximized at the unique maximizer is the right-hand side of (4.16). □

Note that Lemmas 9 and 10 show that the free energy of the coarse-grained Hamiltonians has no phase transition for any finite K. Also those lemmas and the bound

$$\sum_{i=1}^{N} \theta_{i/N} \sigma_i^2 = \sum_{k=1}^{K} x_k \sigma_{[k]}^2 + O\left(K^{-1}\right), \tag{4.19}$$

imply that $\mathcal{F}_K(\beta)$ is an approximation of the free energy of the Hamiltonian $N \sum_{i=1}^{N} \theta_{i/N} \sigma_i^2$.

Lemma 11. *For all $C > 0$ and $K \geq 2$ we have*

$$\limsup_{N \to \infty} \sup_{\beta \in [0,C]} \left| \frac{1}{N} \log E^N \left[\exp\left(\beta N \sum_{i=1}^{N} \theta_{i/N} \sigma_i^2 \right) \right] - \mathcal{F}_K(\beta) \right| \leq \frac{c}{K}. \tag{4.20}$$

We now investigate the behavior of $\mathcal{F}_K(\beta)$ as $K \to \infty$. Let

$$g(\lambda) = \int_{-\sqrt{2}}^{\sqrt{2}} \frac{\mu(x)}{\lambda - x} dx \text{ for } \lambda \geq \sqrt{2}. \tag{4.21}$$

By standard estimates for Riemann sums

$$\lim_{K \to \infty} g_K(\lambda) = g(\lambda) \text{ for } \lambda > \sqrt{2}. \tag{4.22}$$

The integral can be computed explicitly, and in fact

$$g(\lambda) = \lambda - \sqrt{\lambda^2 - 2}.$$

Note that $g\left(\sqrt{2}\right) = \sqrt{2}$. If $\beta \leq \frac{1}{\sqrt{2}}$ there is a unique $\lambda(\beta) \geq \sqrt{2}$ such that $g(\lambda(\beta)) = 2\beta$. In fact

$$\lambda(\beta) = \frac{1}{\sqrt{2}} \left(\sqrt{2}\beta + \frac{1}{\sqrt{2}\beta} \right) \quad \text{for} \quad \beta \leq \frac{1}{\sqrt{2}}. \tag{4.23}$$

The convergence (4.22) implies that

$$\lim_{K \to \infty} \lambda_K(\beta) = \lambda(\beta) \quad \text{for} \quad \beta < \frac{1}{\sqrt{2}}. \tag{4.24}$$

Also define

$$h(\lambda) = \int_{-\sqrt{2}}^{\sqrt{2}} \mu(x) \log(\lambda - x)\, dx \text{ for } \lambda \geq \sqrt{2},$$

which can be computed explicitly as

$$h(\lambda) = \frac{\lambda^2}{2} - \frac{1}{2} - \frac{\lambda\sqrt{\lambda^2 - 2}}{2} + \log\left(\frac{\lambda + \sqrt{\lambda^2 - 2}}{2} \right). \tag{4.25}$$

By the convergence of the Riemann sum

$$\lim_{K \to \infty} h_K(\lambda) = h(\lambda) \text{ for } \lambda > \sqrt{2}. \tag{4.26}$$

Define

$$\mathcal{F}(\beta) = \beta\lambda(\beta) - \frac{1}{2} - \frac{1}{2}\log(2\beta) - \frac{1}{2} h(\lambda(\beta)), \beta \in \left[0, \frac{1}{\sqrt{2}}\right]. \tag{4.27}$$

Using the identities (4.23) and (4.25), this expression simplifies to

$$\mathcal{F}(\beta) = \frac{\beta^2}{2} \quad \text{for } \beta \in \left[0, \frac{1}{\sqrt{2}}\right]. \tag{4.28}$$

Also it follows from (4.24) and (4.26) and the monotonicity of $h_K(\lambda)$ that

$$\lim_{K \to \infty} \mathcal{F}_K(\beta) = \mathcal{F}(\beta) \text{ if } \beta < \frac{1}{\sqrt{2}}. \tag{4.29}$$

A posteriori it is clear that for $\beta > \frac{1}{\sqrt{2}}$ the function $\mathcal{F}_K(\beta)$ converges to the low-temperature free energy of the Hamiltonian $H_N(\sigma)$ without external field, but this is not a step in the proof of our main results, but rather a consequence.

In the proof of Proposition 7 at the end of the next section we will use the two lemmas that now follow to rule out m that do not satisfy Plefka's condition. First, note that $g_K(\lambda)$, $\lambda_K(\beta)$, $h_K(\lambda)$ and thus $\mathcal{F}_K(\beta)$ are all continuous and differentiable. We have the following identity.

Lemma 12. *For all* $\beta > 0$

$$\mathcal{F}_K^{'}(\beta) = \lambda_K(\beta) - \frac{1}{2\beta}. \tag{4.30}$$

Proof. This follows from the definition (4.15) and the equalities $h_K^{'} = g_K$ and $g_K(\lambda(\beta)) = 2\beta$. □

Lemma 13. *For all* $K \geq 2$ *there is an* $\varepsilon \in \left(0, \frac{2\sqrt{2}}{K}\right)$ *such that*

$$\lambda_K(\beta) \geq \sqrt{2} - \varepsilon \implies \beta \leq \frac{1}{\sqrt{2}}.$$

Proof. We may set $\sqrt{2} - \varepsilon = \sqrt{2} - \lambda_K\left(\frac{1}{\sqrt{2}}\right)$ since

$$\lambda_K(\beta) \geq \lambda_K\left(\frac{1}{\sqrt{2}}\right) \implies \beta \leq \frac{1}{\sqrt{2}},$$

and

$$x_K < \lambda_K\left(\frac{1}{\sqrt{2}}\right) < \lambda\left(\frac{1}{\sqrt{2}}\right) = \sqrt{2},$$

where the second inequality follows because $g_K(\lambda) < g(\lambda)$ for $\lambda \geq \sqrt{2}$ (see (4.7), (4.8), (4.21), (4.13)) and $g_K(\lambda)$ is decreasing in λ, implying that the solution to $g_K(\lambda) = \sqrt{2}$ must occur for $\lambda < \sqrt{2}$. □

Lemma 12 also allows us to strengthen the pointwise convergence (4.29) to uniform convergence.

Lemma 14. *We have*

$$\lim_{K\to\infty} \sup_{\beta\in\left[0, \frac{1}{\sqrt{2}}\right]} |\mathcal{F}_K(\beta) - \mathcal{F}(\beta)| = 0. \tag{4.31}$$

Proof. The $\mathcal{F}_K(\beta)$ are increasing in β (because the left-hand side of (4.16) is) and $\mathcal{F}(\beta)$ is increasing in $\beta \in \left[0, \frac{1}{\sqrt{2}}\right]$ and uniformly continuous (recall (4.28)). This implies that the pointwise convergence (4.29) can be strengthened to uniform convergence on $\left[0, \frac{1}{\sqrt{2}} - \delta\right]$ for any $\delta > 0$, i.e.

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

For any $\delta > 0$ we have

$$\sup_{\beta \in \left[\frac{1}{\sqrt{2}} - \delta, \frac{1}{\sqrt{2}}\right]} |\mathcal{F}(\beta) - \mathcal{F}(\beta - \delta)| \leq c\delta,$$

and for any $\delta \in \left(0, \frac{1}{2\sqrt{2}}\right)$ (say) we have uniformly in K that

$$\sup_{\beta \in \left[\frac{1}{\sqrt{2}} - \delta, \frac{1}{\sqrt{2}}\right]} |\mathcal{F}_K(\beta) - \mathcal{F}_K(\beta - \delta)| \leq c\delta,$$

by (4.30) ($\lambda_K(\beta)$ is decreasing in β and $\lambda_K\left(\frac{1}{2}\right)$ is bounded by (4.24)). Thus for such δ also

$$\sup_{\beta \in \left[\frac{1}{\sqrt{2}} - \delta, \frac{1}{\sqrt{2}}\right]} |\mathcal{F}_K(\beta) - \mathcal{F}(\beta)| \leq c\delta.$$

Thus

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

for all $\delta \in \left(0, \frac{1}{2\sqrt{2}}\right)$, so the claim (4.31) follows. □

4.3. Making the external field after recentering vanish. As for the lower bound, an important step in the proof of the upper bound is to recenter the Hamiltonian around an $m \in \mathbb{R}^N$ which yields an effective external field $\beta \frac{1}{N} \nabla \tilde{H}_N(m) + \tilde{h}_N$ [cf. (3.8)–(3.9) and (4.44)]. In this section the main goal is Lemma 17, which constructs a low-dimensional subspace $\mathcal{M}_N \subset \mathbb{R}^N$, such that if we recenter around any $m \in \mathcal{M}_N$ the effective external field is again (almost) contained in $\mathcal{M}_N$ (so that if we restrict to the space perpendicular to $\mathcal{M}_N$, the effective external field after recentering almost vanishes). Its use will be an important step in the proof of the upper bound in the next subsection.

The construction in the proof of Lemma 17 will involve taking the span of a vector (close to) $\tilde{h}_N$ iterated under the the map $\frac{1}{N} \nabla \tilde{H}_N$. For this the next lemma will be needed, whose claim (4.32) says says that after applying the map $\frac{1}{N} \nabla \tilde{H}_N$ to a vector $v \in \mathbb{R}^N$ a large number of times, the resulting vector will be almost completely contained in the space spanned by the eigenvectors associated to the eigenvalues of largest magnitude. Let Π^A denote the projection onto a subspace $A \subset \mathbb{R}^N$.

Lemma 15. *For any $\varepsilon > 0$, $N \geq 1$, $v \in \mathbb{R}^N$ with $v_N \neq 0$ and $k \geq 1$ it holds that*

$$\frac{\left|\Pi^{\left\langle e_i : |\theta_{i/N}| < \sqrt{2} - \varepsilon \right\rangle} \left(\frac{1}{N} \nabla \tilde{H}_N\right)^k v\right|}{\left|\left(\frac{1}{N} \nabla \tilde{H}_N\right)^k v\right|} \leq \sqrt{N}\, |v|\, v_N^{-1} e^{-c\varepsilon k}. \tag{4.32}$$

Proof. Denote the matrix $\frac{1}{N}\nabla\tilde{H}_N$ by D. Note that D is diagonal and $D_{ii}=2\theta_{i/N}$. Thus for any $v\in\mathbb{R}^N$ we have

$$\left(D^k v\right)_i=\left(2\theta_{i/N}\right)^k v_i.$$

Now for v such that $v_N\neq 0$ and i such that $\left|\theta_{i/N}\right|<\sqrt{2}-\varepsilon$ we have

$$\frac{\left|\left(D^k v\right)_i\right|}{\left|D^k v\right|}=\frac{\left|\left(2\theta_{i/N}\right)^k v_i\right|}{\sqrt{\sum_{i=1}^N\left(2\theta_{i/N}\right)^{2k}v_i^2}}\leq\frac{\left|\theta_{i/N}\right|^k|v|}{\sqrt{2}^k v_N}\leq v_N^{-1}|v|(1-c\varepsilon)^k\leq|v|\,v_N^{-1}e^{-c\varepsilon k}.$$

By taking the square and summing over the at most N indices i such that $\left|\theta_{i/N}\right|<\sqrt{2}-\varepsilon$ the claim (4.32) follows. □

The next lemma is a weak bound on the proportion of all eigenvalues have magnitude close to the maximal magnitude.

Lemma 16. *For all $N\geq 1$ and $\varepsilon>0$*

$$\left|\left\{i:\left|\theta_{i/N}\right|\geq\sqrt{2}-\varepsilon\right\}\right|\leq c\varepsilon N. \tag{4.33}$$

Proof. This follows for instance by noting that $\theta_{(i+1)/N}-\theta_{i/N}\geq cN^{-1}$, which is a consequence of the definition (2.4) of $\theta_{i/N}$ and the bound $\int_{\sqrt{2}-\varepsilon}^{\sqrt{2}}\mu(x)\,dx\leq\varepsilon\sup_x\mu(x)\leq c\varepsilon$. □

We now construct the subspaces $\mathcal{M}_N$. Recall that $\frac{1}{N}\nabla\tilde{H}_N$ is a linear map (cf. (2.2)) and that the standard basis vectors e_i are its eigenvectors, so the span of any set of basis vectors in invariant under $\frac{1}{N}\nabla\tilde{H}_N$.

Lemma 17. *Let $\beta, h, h_1, h_2, \ldots$ be as in Theorem 1. There exists a sequence of linear spaces $\mathcal{M}_1, \mathcal{M}_2, \ldots$ such that $\mathcal{M}_N\subset\mathbb{R}^N$,*

$$\dim\left(\mathcal{M}_N\right)=\lfloor N^{3/4}\rfloor, \tag{4.34}$$

and $\mathcal{M}_N$ is approximately invariant under the map $m\to\beta\frac{1}{N}\nabla\tilde{H}_N(m)+\tilde{h}_N$ in the sense that

$$\lim_{N\to\infty}\sup_{m\in\mathcal{M}_N,|m|\leq 1}\left|\Pi^{\mathcal{M}_N^{\perp}}\left(\beta\frac{1}{N}\nabla\tilde{H}_N(m)+\tilde{h}_N\right)\right|=0. \tag{4.35}$$

Proof. We will construct $\mathcal{M}_N$ so that it contains a vector $\bar{h}_N$ close to $\tilde{h}_N$ and is approximately invariant under the map $\frac{1}{N}\nabla\tilde{H}_N(m)$. More precisely let

$$\bar{h}_{N,i}=\tilde{h}_{N,i}\quad\text{for}\quad i\leq N-1\quad\text{and}\quad\bar{h}_{N,N}=\begin{cases}\tilde{h}_{N,N} & \text{if } \left|\tilde{h}_{N,N}\right|\geq\frac{1}{N},\\ \frac{1}{N} & \text{if } \left|\tilde{h}_{N,N}\right|<\frac{1}{N},\end{cases} \tag{4.36}$$

so that

$$\left|\bar{h}_N-\tilde{h}_N\right|\leq\frac{1}{N}\quad\text{and}\quad\left|\bar{h}_{N,N}\right|\geq\frac{1}{N}.$$

We construct $\mathcal{M}_N$ so that

$$\bar{h}_N \in \mathcal{M}_N, \tag{4.37}$$

and $\mathcal{M}_N$ is almost invariant under $\frac{1}{N}\nabla\tilde{H}_N$ in the sense that

$$\lim_{N\to\infty} \sup_{m\in\mathcal{M}_N, |m|\le 1} \left|\Pi^{\mathcal{M}_N^\perp}\left(\frac{1}{N}\nabla\tilde{H}_N(m)\right)\right| = 0. \tag{4.38}$$

Since $\left|\Pi^{\mathcal{M}_N^\perp}\tilde{h}_N\right| \le \left|\Pi^{\mathcal{M}_N^\perp}\bar{h}_N\right| + \frac{1}{N} = \frac{1}{N}$ and

$$\left|\Pi^{\mathcal{M}_N^\perp}\left(\beta\frac{1}{N}\nabla\tilde{H}_N(m) + \tilde{h}_N\right)\right| \le \beta\left|\Pi^{\mathcal{M}_N^\perp}\left(\frac{1}{N}\nabla\tilde{H}_N(m)\right)\right| + \left|\Pi^{\mathcal{M}_N^\perp}\tilde{h}_N\right|,$$

this implies (4.35). Furthermore, it suffices to construct $\mathcal{M}_N$ so that

$$\dim \mathcal{M}_N \le \lfloor N^{3/4} \rfloor, \tag{4.39}$$

since by adding arbitrary basis vectors e_i (which are invariant under $\frac{1}{N}\nabla\tilde{H}_N$) to the span of $\mathcal{M}_N$ one can ensure $\dim \mathcal{M}_N = \lfloor N^{3/4} \rfloor$ while maintaining (4.37) and (4.38).

To ensure (4.38) we will let $\mathcal{M}_N$ contain the span of a sufficient number of vectors

$$\bar{h}_N^k = \left(\frac{1}{N}\nabla\tilde{H}_N\right)^k \bar{h}_N, k \ge 0,$$

and basis vectors e_i belonging to the eigenvalues $\theta_{i/N}$ of largest magnitude. Let

$$\hat{h}_N^k = \frac{\bar{h}_N^k}{\left|\bar{h}_N^k\right|},$$

be normalized vectors and construct

$$\mathcal{M}_N = \left\langle \hat{h}_N^0, \ldots, \hat{h}_N^{V-1}, e_j : j \in J \right\rangle,$$

for

$$V = \sqrt{N}(\log N)^2,$$

and

$$J = \left\{ j : \left|\theta_{j/N}\right| \ge \sqrt{2} - N^{-1/2} \right\}.$$

Clearly, (4.37) holds since $\mathcal{M}_N$ contains $\hat{h}_N^0 = \bar{h}_N / \left|\bar{h}_N\right|$. Using Lemma 16 it also holds that

$$\dim \mathcal{M}_N \le V + |J| \le \sqrt{N}(\log N)^2 + c\sqrt{N},$$

which implies (4.39).

To check (4.38), note that for any $m \in \mathcal{M}_N$ with $|m| \le 1$ we may decompose m as

$$m = \sum_{k=0}^{V-1} \alpha_k \hat{h}_N^k + \sum_{i\in J} \gamma_i e_i, \tag{4.40}$$

for some $\alpha_0, \ldots, \alpha_{V-1} \in \mathbb{R}$, $\gamma_i \in \mathbb{R}$, $i \in J$, where we first set set $\alpha_{V-1} = m \cdot \hat{h}_N^{V-1}$, to ensure that

$$|\alpha_{V-1}| \leq 1, \tag{4.41}$$

before picking the other coefficients in the decomposition. Thus

$$\begin{aligned}\tfrac{1}{N}\nabla \tilde{H}_N(m) &= \textstyle\sum_{k=0}^{V-1} \alpha_k \tfrac{1}{N}\nabla \tilde{H}_N\left(\hat{h}_N^k\right) + \sum_{i\in J}\gamma_i \tfrac{1}{N}\nabla \tilde{H}_N(e_i)\\ &= \textstyle\sum_{k=0}^{V-1} \alpha_k \frac{|\bar{h}_N^{k+1}|}{|\bar{h}_N^k|}\hat{h}_N^{k+1} + \sum_{i\in J}\gamma_i 2\theta_{i/N} e_i.\end{aligned}$$

Therefore

$$\Pi^{\mathcal{M}_N^\perp}\left(\frac{1}{N}\nabla \tilde{H}_N(m)\right) = \alpha_{V-1}\frac{\left|\bar{h}_N^V\right|}{\left|\bar{h}_N^{V-1}\right|}\Pi^{\mathcal{M}_N^\perp}\left(\hat{h}_N^V\right).$$

Note that since $\|\frac{1}{N}\nabla \tilde{H}_N\| = 2\theta_0 = 2\theta_1 = 2\sqrt{2}$ we have $\left|\bar{h}_N^V\right| \leq 2\sqrt{2}\left|\bar{h}_N^{V-1}\right|$ and so using also (4.41)

$$\left|\Pi^{\mathcal{M}_N^\perp}\frac{1}{N}\nabla \tilde{H}_N(m)\right| \leq c\left|\Pi^{\mathcal{M}_N^\perp}\hat{h}_N^V\right|.$$

The point of (4.36) was to ensure that $\left|\bar{h}_{N,N}\right| \geq \frac{1}{N}$, so that Lemma 15 applies to $\bar{h}_N$. With $\varepsilon = N^{-1/2}$ it gives that

$$\left|\Pi^{\mathcal{M}_N^\perp}\hat{h}_N^V\right| = \frac{\left|\Pi^{\mathcal{M}_N^\perp}\bar{h}_N^V\right|}{\left|\bar{h}_N^V\right|} \leq \frac{\left|\Pi^{\left\langle e_i:\theta_{i/N}<\sqrt{2}-\varepsilon\right\rangle}\bar{h}_N^V\right|}{\left|\bar{h}_N^V\right|} \leq c\sqrt{N}\left|\bar{h}_N\right|\bar{h}_{N,N}^{-1}e^{-c(\log N)^2} = o(1), \tag{4.42}$$

so (4.38) follows. Since we have constructed $\mathcal{M}_N$ satisfying (4.37), (4.38) and (4.39) the proof is complete. □

We will need a version of Lemma 11 where we integrate over the subspace perpendicular to $\mathcal{M}_N$.

Lemma 18. *For any $C > 0$ and $K > 0$*

$$\limsup_{N\to\infty}\sup_{\beta\in[0,C]}\left|\frac{1}{N}\log E^{\mathcal{U}_N}\left[\exp\left(\beta N\sum_{i=1}^N \theta_{i/N}\sigma_i^2\right)\right] - \mathcal{F}_K(\beta)\right| \leq \frac{c}{K},$$

where $\mathcal{U}_N = \mathcal{M}_N^\perp$ and $\mathcal{M}_N$, $N \geq 1$, is the sequence of subspaces from Lemma 17.

Proof. This follows from Lemma 11 similarly to how Lemma 5 follows from Lemma 4. Let $M = \lfloor N^{3/4}\rfloor$. Consider an orthonormal basis of $\mathbb{R}^{N-M}$ such that the space $\mathcal{U}_N$ is spanned by the first $N-M$ basis vectors and let A be the $(N-M)\times(N-M)$ minor of the matrix D which in the standard basis is diagonal with $D_{ii} = \theta_{i/N}$. The eigenvalues $a_1, \ldots, a_{N-M}$ of A satisfy $a_i = \theta_{i/N} + o(1) = \theta_{i/(N-M)} + o(1)$ by the eigenvalue interlacing inequality, so that an estimate for $E^{\mathcal{U}_N}[\cdot]$ follows from Lemma 11 with $N-M$ in place of N. □

4.4. Proof of upper bound. We are now ready to complete the proof of the upper bound. Define a modified TAP free energy by replacing the Onsager correction $\frac{1}{2}\beta^2\left(1-|m|^2\right)^2$ by $\mathcal{F}_K\left(\beta\left(1-|m|^2\right)\right)$ to obtain

$$\tilde{H}^K_{TAP}(m) = \beta\tilde{H}_N(m) + Nm\cdot\tilde{h}_N + \frac{N}{2}\log\left(1-|m|^2\right) + N\mathcal{F}_K\left(\beta\left(1-|m|^2\right)\right).$$

We have the following version of the upper bound Proposition 7 with $\tilde{H}^K_{TAP}(m)$ in place of $\tilde{H}_{TAP}(m)$ and without a Plefka condition.

Proposition 19. *For all $K\geq 2$ and $\beta, h, h_1, h_2, \ldots$ as in Theorem 1 we have*

$$\tilde{F}_N(\beta, h_N) \leq \frac{1}{N}\sup_{m\in\mathbb{R}^N:|m|<1}\tilde{H}^K_{TAP}(m) + \frac{c}{K}, \tag{4.43}$$

for large enough N.

Proof. Let $\mathcal{M}_N$ be the space from Lemma 17 and let

$$\mathcal{U}_N = \mathcal{M}_N^{\perp}.$$

Let $M = \lfloor N^{3/4}\rfloor$. For any $\sigma\in\mathbb{R}^N$ let m be the projection of σ onto $\mathcal{M}_N$ and $\hat{\sigma} = \sigma - m \in \mathcal{U}_N$. Recentering the Hamiltonian around m [cf. (3.8)–(3.9)] we have that

$$\begin{aligned} &E\left[\exp\left(\beta\tilde{H}_N(\sigma) + N\tilde{h}_N\cdot\sigma\right)\right] \\ &\quad = E\left[\exp\left(N\beta\tilde{H}_N(m) + N\tilde{h}_N\cdot m + N\left(\beta\tfrac{1}{N}\nabla\tilde{H}_N(m) + \tilde{h}_N\right)\cdot\hat{\sigma} + \beta\tilde{H}_N(\hat{\sigma})\right)\right]. \end{aligned} \tag{4.44}$$

Lemma 17 implies that

$$\lim_{N\to\infty}\sup_{m\in\mathcal{M}_N}\sup_{\hat{\sigma}\in\mathcal{M}_N^{\perp},|\hat{\sigma}|\leq 1}\left(\beta\frac{1}{N}\nabla\tilde{H}_N(m) + \tilde{h}_N\right)\cdot\hat{\sigma} = 0,$$

so the effective external field vanishes and (4.44) is at most

$$e^{o(N)}E\left[\exp\left(N\beta\tilde{H}_N(m) + N\tilde{h}_N\cdot m + \beta\tilde{H}_N(\hat{\sigma})\right)\right]. \tag{4.45}$$

Note that the the $E[\cdot|m]$-law of $\hat{\sigma}$ is the uniform distribution on sphere in the subspace $\mathcal{U}_N$ of radius $\sqrt{1-|m|^2}$. Thus using also (2.1) this equals

$$E\left[\exp\left(N\beta\tilde{H}_N(m) + N\tilde{h}_N\cdot m\right)E^{\mathcal{U}_N}\left[\beta\left(1-|m|^2\right)\tilde{H}_N(\sigma)\right]\right]. \tag{4.46}$$

By Lemma 18 this is at most

$$E\left[\exp\left(N\beta\tilde{H}_N(m) + N\tilde{h}_N\cdot m + \mathcal{F}_K\left(\beta\left(1-|m|^2\right)\right)\right)\right]e^{o(N)+\frac{c}{K}}. \tag{4.47}$$

Using (2.9) the E-integral equals

$$\begin{aligned} a_N\int_{m:|m|<1}&\left(1-|m|^2\right)^{\frac{N-M-2}{2}}\exp\left(N\beta H_N(m) + N\tilde{h}_N\cdot m\right. \\ &\left.+ N\mathcal{F}_K\left(\beta\left(1-|m|^2\right)\right)\right)dm, \end{aligned} \tag{4.48}$$

where $a_N = \frac{1}{\pi^{\frac{N-M}{2}}} \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{M}{2}\right)}$ and the integral is M-dimensional against Lebesgue measure on $\mathcal{M}_N$. This equals

$$a_N \int_{m:|m|<1} \exp\left(\tilde{H}^K_{TAP}(m) + (M+2)\left|\log\left(1-|m|^2\right)\right|\right) dm,$$

and by the Laplace method is bounded above by

$$a_N \exp\left(\sup_{m:|m|<1}\left\{\tilde{H}^K_{TAP}(m) + (M+2)\left|\log\left(1-|m|^2\right)\right|\right\}\right)\int_{m:|m|<1} dm.$$

The M-dimensional Lebesgue integral $\int_{m:|m|<1} dm$ is the volume of the unit ball in dimension M which equals $\frac{\pi^{\frac{M}{2}}}{\Gamma\left(\frac{M}{2}+1\right)} = O(1)$, and $\log a_N = o(N)$, so this is at most

$$\exp\left(o(N) + \sup_{m:|m|<1}\left\{\tilde{H}^K_{TAP}(m) + (M+2)\left|\log\left(1-|m|^2\right)\right|\right\}\right). \tag{4.49}$$

To get rid of the nuisance term involving $M+2$, note that there is a δ depending only on β and h such that the supremum is always achieved for $|m| < 1-\delta$, since all terms in the supremum not involving log are bounded by cN. Thus (4.49) is at most

$$\exp\left(o(N) + \sup_{m:|m|<1}\tilde{H}^K_{TAP}(m) + cM\right). \tag{4.50}$$

This is then also an upper bound for (4.48), which shows that (4.47) and therefore $\tilde{F}_N(\beta, h_N)$ is bounded by $\exp\left(\sup_{m:|m|<1}\tilde{H}^K_{TAP}(m) + o(N) + \frac{c}{K}\right)$. This implies (4.43). □

We can now prove the upper bound Proposition 7 for free energy of the diagonal and deterministic Hamiltonian $\tilde{H}_N(\sigma)$, by showing that the sup in (4.43) is bounded above by that in (4.6).

Proof of Proposition 7. Fix $K \geq 2$. For any $N \geq 1$, consider the variational problem

$$\sup_{m\in\mathbb{R}^N:|m|<1}\tilde{H}^K_{TAP}(m).$$

Any local maximum m of $\tilde{H}^K_{TAP}(m)$ must satisfy

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

and

$$\nabla^2\tilde{H}^K_{TAP}(m) \text{ is negative semi-definite.} \tag{4.51}$$

The gradient of $\tilde{H}^K_{TAP}$ is

$$\nabla\tilde{H}^K_{TAP}(m) = \beta\nabla\tilde{H}_N(m) + N\tilde{h}_N - Nm\left(\frac{1}{1-|m|^2} + 2\beta\mathcal{F}'_K\left(\beta\left(1-|m|^2\right)\right)\right).$$

By Lemma 12 we have for all m that

$$\nabla \tilde{H}_{TAP}^{K}(m) = \beta \nabla \tilde{H}_N(m) + N \tilde{h}_N - N 2 \beta m \lambda_K\left(\beta\left(1-|m|^2\right)\right).$$

Thus, the Hessian $\nabla^2 \tilde{H}_N^K(m)$ equals

$$\beta \nabla^2 \tilde{H}_N(m) - N 2 \beta I \lambda_K\left(\beta\left(1-|m|^2\right)\right) + 4\beta^2 N m m^T \lambda_K^{'}\left(\beta\left(1-|m|^2\right)\right).$$

For any local maximum m let

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

and

$$B = 2m(m)^T \lambda_K^{'}\left(\beta\left(1-|m|^2\right)\right).$$

Since B is of rank one, the second largest eigenvalue a_{N-1} of A is bounded above by the largest eigenvalue of $A+B$. The latter matrix is the Hessian at m multiplied by a positive scalar, so all its eigenvalues are non-positive. Thus $a_{N-1} \leq 0$. Furthermore, $\frac{1}{2N}\nabla^2 \tilde{H}_N(m) = \frac{1}{N} D$ where D is the diagonal matrix with $D_{ii} = \theta_{i/N}$, so the eigenvalues of A are $\theta_{i/N} - \lambda_K\left(\beta\left(1-|m|^2\right)\right)$. This shows that

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

at m which are local maxima. Since

$$\theta_{1-1/N} = \sqrt{2} + o(1),$$

it follows from Lemma 13 that we must have for such m

$$\beta\left(1-|m|^2\right) \leq \frac{1}{\sqrt{2}}, \text{ that is } \beta(m) \leq \frac{1}{\sqrt{2}},$$

(provided N large enough depending on K), and by Lemma 14

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

where $\lim_{K\to\infty} \varepsilon_K = 0$. Thus from (4.43) it holds for such N that

$$\tilde{F}_N(\beta, h_N) \leq \frac{1}{N} \sup_{m \in \mathbb{R}^N : |m|<1, \beta(m) \leq \frac{1}{\sqrt{2}}} \tilde{H}_{TAP}(m) + \varepsilon_K + \frac{c}{K}.$$

We have shown that

$$\limsup_{N\to\infty} \left\{ \frac{1}{N} \sup_{m \in \mathbb{R}^N : |m|<1, \beta(m) \leq \frac{1}{\sqrt{2}}} \tilde{H}_{TAP}(m) - \tilde{F}_N(\beta, h_N) \right\} \leq \varepsilon_K + \frac{c}{K},$$

for all $K \geq 2$. Since the left-hand side is independent of K, it is in fact at most 0. This implies (4.6). □

This also completes the proof of the main upper bound Proposition 6. Together with the lower bound Proposition 3 this proves our main result Theorem 1.

5. Solution of the TAP–Plefka Variational Problem

In this section we prove Lemma 2. By (2.1) it follows from a result for the maximum of the Hamiltonian with external field on the unit sphere which we now state.

Lemma 20. *For $h, h_1, h_2, \ldots$ as in Theorem 1 we have*

$$\sup_{\sigma:|\sigma|=1}\left\{\beta\frac{1}{N}H_N(\sigma)+h_N\cdot\sigma\right\}\to\sqrt{h^2+2\beta^2}, \tag{5.1}$$

in probability.

Proof. We work in the diagonalizing basis of S_N and note that the left-hand side of (5.1) equals

$$\sup_{\sigma:|\sigma|=1}\left\{\beta\sum_{i=1}^{N}\theta_{i/N}\sigma_i^2+\tilde{h}_N\cdot\sigma\right\}+o(1), \tag{5.2}$$

where, as in Sect. 4.1, $\tilde{h}_N$ is the vector h_N written in the diagonalizing basis and we have used (2.5). The case $h=0$ then follows trivially since $\theta_1=\sqrt{2}$, so we assume in the sequel that $h>0$. For any $\lambda>\sqrt{2}$ let

$$\sigma_i(\lambda)=\frac{1}{2\beta}\frac{\left(\tilde{h}_N\right)_i}{\lambda-\theta_{i/N}}.$$

Using Lagrange multipliers the maximizer of (5.2) can be shown to be $\sigma_i=\sigma_i(\lambda_N)$ where $\lambda_N>\sqrt{2}$ is the number such that $\sum_{i=1}^N\sigma_i^2(\lambda_N)=1$. By rotational symmetry the $\mathbb{P}$-law of $\left(\tilde{h}_N\right)_i$ is that of a uniform random vector on $\left\{x\in\mathbb{R}^N:|x|=h\right\}$. Using this one can show that for any $\lambda>\sqrt{2}$

$$\left|\sum_{i=1}^{N}\sigma_i(\lambda)^2-\frac{h^2}{2\beta}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\left(\lambda-\theta_{i/N}\right)^2}\right|\to 0,\ \text{in probability.}$$

Also for $\lambda>\sqrt{2}$

$$\sum_{i=1}^{N}\frac{1}{\left(\lambda-\theta_{i/N}\right)^2}\to\int_{-\sqrt{2}}^{\sqrt{2}}\frac{\mu(x)}{(\lambda-x)^2}dx=\frac{\lambda}{\sqrt{\lambda^2-2}}-1,$$

and since for

$$\tilde{\lambda}=\sqrt{\frac{2}{1-\left(1+\frac{4\beta^2}{h^2}\right)^{-2}}},$$

and $\lambda=\tilde{\lambda}$ we have $\lambda/\sqrt{\lambda^2-2}-1=2\beta/h^2$, it follows that

$$\lambda_N\to\tilde{\lambda},\ \text{in probability.}$$

Similarly, for any $\lambda > \sqrt{2}$ we have that

$$\sum_{i=1}^{N} \theta_{i/N} \sigma_i (\lambda)^2 \to \frac{h^2}{2\beta} \int_{-\sqrt{2}}^{\sqrt{2}} \frac{x}{(\lambda - x)^2} \mu(x)\, dx = \frac{h^2}{2\beta^2} \left(\frac{\lambda^2 - 1}{\sqrt{\lambda^2 - 2}} - \lambda \right),$$

and

$$\sum_{i=1}^{N} \left(\tilde{h}_N \right)_i \sigma_i (\lambda) \to \frac{h^2}{2\beta} \int_{-\sqrt{2}}^{\sqrt{2}} \frac{\mu(x)}{\lambda - x} dx = \frac{h^2}{2\beta} \left(\lambda - \sqrt{\lambda^2 - 2} \right),$$

both in probability. This shows that

$$\beta \sum_{i=1}^{N} \theta_{i/N} \sigma_i (\lambda_N)^2 + \tilde{h}_N \cdot \sigma(\lambda_N) \to \frac{h^2}{2\beta} \left(\frac{\tilde{\lambda}^2 - 1}{\sqrt{\tilde{\lambda}^2 - 2}} - \tilde{\lambda} \right) + \frac{h^2}{2\beta} \left(\tilde{\lambda} - \sqrt{\tilde{\lambda}^2 - 2} \right),$$

in probability, and the right-hand side simplifies to $\sqrt{h^2 + 2\beta^2}$. □

Acknowledgement. The first author thanks Erwin Bolthausen and Giuseppe Genovese for valuable discussions on a draft of this article. The second author wishes to express his gratitude to Markus Petermann for a long-standing discussion on spin glasses, and to Anton Wakolbinger for encouragement.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Aizenman, M., Lebowitz, J.L., Ruelle, D.: Some rigorous results on the Sherrington–Kirkpatrick spin glass model. Commun. Math. Phys. **112**(1), 3–20 (1987)
2. Auffinger, A., Arous, G.B., Černý, J.: Random matrices and complexity of spin glasses. Commun. Pure Appl. Math., **66**(2), 165–201 (2013)
3. Auffinger, A., Jagannath, A.: Thouless–Anderson–Palmer equations for conditional Gibbs measures in the generic p-spin glass model (2016). arXiv preprint arXiv:1612.06359
4. Baik, J., Lee, J.O.: Fluctuations of the free energy of the spherical Sherrington–Kirkpatrick model. J. Stat. Phys. **165**(2), 185–224 (2016)
5. Benaych-Georges, F., Knowles, A.: Lectures on the local semicircle law for Wigner matrices (2016). arXiv preprint arXiv:1601.04055
6. Bolthausen, E.: Private communication
7. Bolthausen, E.: An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. Commun. Math. Phys. **325**(1), 333–366 (2014)
8. Bragg, W.L., Williams, E.J: The effect of thermal agitation on atomic arrangement in alloys. Proc. R. Soc. Lond. Ser. A **145**(855), 699–730 (1934). **(Containing Papers of a Mathematical and Physical Character)**
9. Chatterjee, S.: Spin glasses and Stein's method. Probab. Theory Relat. Fields **148**(3), 567–600 (2010)
10. Chen, W.-K., Panchenko, D.: On the TAP free energy in the mixed *p*-spin models (2017). arXiv preprint arXiv:1709.03468
11. Crisanti, A., Sommers, H.-J.: Thouless–Anderson–Palmer approach to the spherical p-spin spin glass model. J. Phys. I **5**(7), 805–813 (1995)
12. Crisanti, A., Sommers, H.-J.: The spherical p-spin interaction spin glass model: the statics. Z. Phys. B Condens. Matter **87**(3), 341–354 (1992)
13. Genovese, G., Tantari, D.: Legendre duality of spherical and Gaussian spin glasses. Math. Phys. Anal. Geom. **18**(1), 10 (2015)
14. Guerra, F.: Broken replica symmetry bounds in the mean field spin glass model. Commun. Math. Phys. **233**(1), 1–12 (2003)

15. Kosterlitz, J.M., Thouless, D.J., Jones, R.C.: Spherical model of a spin-glass. Phys. Rev. Lett. **36**(20), 1217 (1976)
16. Mézard, M., Parisi, G., Virasoro, M.A.: Spin Glass Theory and Beyond, Volume 9 of World Scientific Lecture Notes in Physics. World Scientific Publishing Co., Inc., Teaneck (1987)
17. Panchenko, D.: The Parisi ultrametricity conjecture. Ann. Math. (2) **177**(1), 383–393 (2013)
18. Panchenko, D.: The Sherrington–Kirkpatrick Model. Springer, Berlin (2013)
19. Plefka, T.: A lower bound for the spin glass order parameter of the infinite-ranged Ising spin glass model. J. Phys. A Math. Gen. **15**(5), L251 (1982)
20. Plefka, T.: Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. J. Phys. A Math. Gen. **15**(6), 1971 (1982)
21. Sherrington, D., Kirkpatrick, S.: Solvable model of a spin-glass. Phys. Rev. Lett. **35**(26), 1792 (1975)
22. Subag, E.: The geometry of the Gibbs measure of pure spherical spin glasses. Invent. Math. **210**(1), 135–209 (2017)
23. Talagrand, M.: Spin Glasses: A Challenge for Mathematicians: Cavity and Mean Field Models, vol. 46. Springer, Berlin (2003)
24. Talagrand, M.: Free energy of the spherical mean field model. Probab. Theory Relat. Fields **134**(3), 339–382 (2006)
25. Talagrand, M.: The Parisi formula. Ann. Math. (2) **163**(1), 221–263 (2006)
26. Tao, T.: Topics in Random Matrix Theory, vol. 132. American Mathematical Soc., Providence (2012)
27. Thouless, D.J., Anderson, P.W., Palmer, R.G.: Solution of 'solvable model of a spin glass'. Philos. Mag. **35**(3), 593–601 (1977)
28. Vilfan, I.: Lecture Notes in Statistical Mechanics. http://www-f1.ijs.si/~vilfan/SM/. Accessed 21 Jan 2019

Communicated by H. Spohn