arXiv:2209.08985v2 [math.PR] 20 Sep 2022

ON CONCAVITY OF TAP FREE ENERGY IN THE SK MODEL.

STEPHAN GUFLER, ADRIEN SCHERTZER, AND MARIUS A. SCHMIDT

Abstract. We analyse the Hessian of the Thouless-Anderson-Palmer (TAP) free energy for the Sherrington-Kirkpatrick model, below the de Almeida-Thouless line, evaluated in Bolthausen's approximate solutions of the TAP equations. We show that while its empirical spectrum weakly converges to a measure with negative support, positive outlier eigenvalues occur for some (β, h) below the AT line. In this sense, TAP free energy may lose concavity in the order parameter of the theory, i.e. the random spin-magnetisations, even below the AT line. Possible interpretations of these findings within Plefka's expansion of the Gibbs potential are not definitive and include the following: *i)* either higher order terms shall not be neglected even if Plefka's first convergence criterion (yielding, in infinite volume, the AT line) is satisfied, *ii)* Plefka's first convergence criterion (hence the AT line) is necessary yet hardly sufficient, or *iii)* Bolthausen's magnetizations do not approximate the TAP solutions sufficiently well up to the AT line.

Contents

Date: September 21, 2022.
We are grateful to Nicola Kistler for suggesting the problem of concavity of the TAP free energy and for helpful discussions. We also thank Christian Brennecke and Véronique Gayrard for valuable discussions, and Jan Lukas Igelbrink for help with numerical simulations. We are also grateful to Mireille Capitaine for explaining to us how [11] can be extended to the GOE case, and for telling us about the reference [15]. This work was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - GZ 2047/1, Projekt-ID 390685813 and GZ 2151 - Project-ID 390873048, through Project-ID 211504053 - SFB 1060, and by DFG research grant contract number 2337/1-1, project 432176920.

1. Introduction

We consider the standard Sherrington-Kirkpatrick (SK for short) model with an external field. In its random Hamiltonian

$$H_{\beta,h}(\sigma) := \frac{\beta}{\sqrt{2N}} \sum_{i,j=1}^{N} g_{ij}\sigma_i\sigma_j + h\sum_{i=1}^{N}\sigma_i \tag{1.1}$$

for $N \in \mathbb{N}$ spins $\sigma = (\sigma_i) \in \Sigma_N := \{-1,1\}^N$, the disorder is modeled by i.i.d. centered Gaussians g_{ij} with variance 1 on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The parameters $\beta > 0$ and $h \in \mathbb{R}$ are called inverse temperature and external field. The partition function is given by

$$Z_N(\beta,h) := 2^{-N}\sum_{\sigma} \exp H_{\beta,h}(\sigma), \tag{1.2}$$

and the free energy by

$$f_N(\beta,h) := \frac{1}{N}\log Z_N(\beta,h). \tag{1.3}$$

A well-known consequence of Gaussian concentration of measure is that the free energy is self-averaging in the sense that

$$f(\beta,h) := \lim_{N\to\infty}\frac{1}{N}\log Z_N(\beta,h) = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\log Z_N(\beta,h) \text{ almost surely.} \tag{1.4}$$

The existence of the limit on the right-hand side was established in a celebrated paper by Guerra and Toninelli [17]. The limit is given by the Parisi variational formula (see [29, 23]). In addition, in high temperature (β small), $f(\beta,h)$ is given by the replica-symmetric formula, originally proposed by Sherrington and Kirkpatrick [26]:

Theorem 1.1 ([26, 8, 10])**.** *There exists $\beta_0 > 0$ such that for all h, β with $0 < \beta \le \beta_0$,*

$$f(\beta,h) = RS(\beta,h) := \inf_{q\ge 0}\left\{\mathsf{E}\log\cosh(h+\beta\sqrt{q}Z) + \frac{\beta^2(1-q)^2}{4}\right\}, \tag{1.5}$$

where Z is a standard Gaussian.

Talagrand [28, Proposition 1.3.8] proved that for $h \neq 0$, the infimum is uniquely attained at $q = q(\beta,h)$ which satisfies

$$q = \mathsf{E}\tanh^2(h+\beta\sqrt{q}Z), \tag{1.6}$$

which is unique for $h \neq 0$, and for $h = 0$ if $\beta \le 1$. Here and in the following, Z (under a probability P with associated expectation E) always denotes a standard Gaussian. For $\beta > 1$ (and $h = 0$), there are two solutions, one being 0, the other one being positive. The latter is the relevant solution for the minimization and will be denoted by q. A proof of Theorem 1.1 based on an approach of Thouless-Anderson-Palmer (TAP for short) [32] can be found in [8]. The critical temperature β_0 in Theorem 1.1 has then been improved in [10] using the same approach. Actually, $f(\beta,h) = RS(\beta,h)$ is believed to hold under the de Almeida-Thouless condition (AT for short), i.e. for (β,h) with

$$\beta^2\mathsf{E}\frac{1}{\cosh^4(h+\beta\sqrt{q}Z)} \le 1, \tag{1.7}$$

but this problem is still open (however, Toninelli [33] proved that when (1.7) is not satisfied, then the assertion of Theorem 1.1 does not hold anymore). De Almeida and Thouless found the condition (1.7) in 1978 in the context of an instability in the replica procedure [2] which is hard to make rigorous. We also mention that Chen [12] recently established the de Almeida-Thouless line as the transition curve between the replica symmetric and the replica symmetry breaking phases in a SK model with centered Gaussian external field.

To state our results, we first introduce the TAP free energy of the SK model. Analysis of the SK model in terms of the TAP equations was first given by [32]: shortening $\bar{g}_{ij} = \frac{1}{\sqrt{2}}(g_{ij} + g_{ji})$, and for $\boldsymbol{m} = (m_i) \in [-1,1]^N$, this is given by

$$\mathrm{TAP}_N(\boldsymbol{m}) = \frac{\beta}{\sqrt{N}} \sum_{i<j\le N} \bar{g}_{ij} m_i m_j + h \sum_{i=1}^N m_i + \frac{\beta^2}{4} N \left(1 - \frac{1}{N}\sum_{i=1}^N m_i^2\right)^2 - \sum_{i=1}^N I(m_i), \quad (1.8)$$

where for $x \in [-1,1]$,

$$I(x) = \frac{1+x}{2}\log(1+x) + \frac{1-x}{2}\log(1-x) = x\tanh^{-1}(x) - \log\cosh\tanh^{-1}(x)\,. \quad (1.9)$$

The TAP free energy can be related to the free energy by a variational principle: Chen and Pancheko [13, Theorem 1] show that

$$f(\beta,h) = \lim_{\epsilon\downarrow 0} \lim_{N\to\infty} \mathbb{E} \max N^{-1}\mathrm{TAP}_N(\boldsymbol{m}), \quad (1.10)$$

where the maximum is over all $\boldsymbol{m} \in [-1,1]^N$ with $N^{-1}\sum_{i=1}^N m_i^2 \in [q_P - \epsilon, q_P + \epsilon]$, q_P denoting the right edge of the support of the Parisi measure. We also mention that an upper bound of the free energy in terms of the TAP free energy has recently been given by Belius [4]. For the SK model with spherical spins, a variational principle for the TAP free energy has been proved in [5].

The TAP free energy can also be interpreted in terms of the power expansion up to second order of the Gibbs potential of the SK model [25] (see Appendix C and also [19] for further discussion). A necessary condition of Plefka [25] for the convergence of the infinite expansions is that the magnetizations are in

$$P_N^1 := \{\boldsymbol{m} \in [-1,1]^N, \frac{\beta^2}{N}\sum_{i=1}^N \left(1 - m_i^2\right)^2 < 1\}. \quad (1.11)$$

P_N^1 is the set of magnetizations satisfying the so-called first Plefka condition. Before Plefka, this condition was also noted by Bray and Moore [9] who investigated the Hessian matrix of the TAP free energy and affirmed that a stable solution must be a local extremum. For the stability of a diagrammatic expansion of the free energy, Sommers [27] also obtained condition (1.11). There is no rigorous justification whether the first Plefka condition suffices for neglecting the higher-order terms (cf. also the discussion in [22]).

From Plefka's expansion, it is reasonable to expect concavity of TAP_N in $\boldsymbol{m}$ if the Hessian of the third and higher order terms in (C.13) can be neglected. However, the TAP functional is not necessarily concave: let us consider the Hessian

$$\boldsymbol{H}(\boldsymbol{m}) := \frac{\partial^2}{\partial m_i\, \partial m_j}\mathrm{TAP}_N(\boldsymbol{m}) \quad (1.12)$$

at arbitary magnetizations $\boldsymbol{m} \in [-1,1]^N$. Then we have:

Theorem 1.2. *There exists $\beta_0 \in (0,1)$ such that for all $\beta > \beta_0$, $h \neq 0$, there exist $\epsilon > 0$ and random $\boldsymbol{m}_N \in P_N^1$ such that*

$$\lim_{N\to\infty} \mathbb{P}\left(\lambda_1\left(\boldsymbol{H}(\boldsymbol{m}_N)\right) > \epsilon\right) = 1. \tag{1.13}$$

This observation is proved in Section 8. Now the following questions arise: *i)* whether a non-concavity also arises in the vicinity of the maximizer of TAP_N, as the magnetization $\boldsymbol{m}_N$ for which Theorem 1.2 can be proved is somewhat arbitrary (see (8.1)) and probably not in the domain over which the maximum is taken in the variational principle (1.10). *ii)* whether the con-concavity arises from a single outlier eigenvalue, or whether a positive proportion of the eigenvalues are positive.

The fixed point of the TAP equations [32]

$$m_i = \tanh\left(h + \frac{\beta}{\sqrt{N}}\sum_{j\neq i}\bar{g}_{ij}m_j - \beta^2\left(1 - \frac{1}{N}\sum_{i=1}^N m_i^2\right)m_i\right) \tag{1.14}$$

are the critical points of the TAP free energy TAP_N. As we are not able to control these fixed points, we base our analysis on Bolthausen's algorithm [7, 8] which yields a sequence $\boldsymbol{m}^{(k)} \in [-1,1]^N$ of magnetizations which are considered as an approximation of the solutions of (1.14) In [7], the magnetizations $\boldsymbol{m}^{(k)}$ are constructed by a two-step Banach algorithm: $m_i^{(0)} := 0$, $m_i^{(1)} := \sqrt{q}$, and then iteratively

$$m_i^{(k+1)} = \tanh\left(h + \frac{\beta}{\sqrt{N}}\sum_{j\neq i}\bar{g}_{ij}m_j^{(k)} - \beta^2(1-q)m_i^{(k-1)}\right), \tag{1.15}$$

for $k \geq 1$. In the present paper, we use the similar algorithm from [7] whose precise definition we recall in Section 2, and which satisfies the two-step recursion (1.15) approximately (see Remark 3.2). Bolthausen [7, 8] proves that such sequence of magnetisations converges up to the AT-line. Precisely, by means of a sophisticated conditioning procedure which will be recalled in Section 2, Bolthausen shows that the iterates satisfy

$$\lim_{k,l\to\infty}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\left(m_i^{(k)} - m_i^{(l)}\right)^2\right] = 0, \tag{1.16}$$

provided (β, h) satisfy the AT-condition. For our purposes, it is crucial to emphasize that due to the somewhat "unreasonable" limit $N \uparrow +\infty$ first, and only in a second step $k, l \uparrow +\infty$, it is not even clear "to what exactly" Bolthausen's approximate solutions converge. Notwithstanding, the following suggests that Bolthausen's magnetizations are *good enough* when it comes to computing the limiting free energy within the TAP-approximation:

Theorem 1.3. *For $\beta > 0$, $h \neq 0$ satisfying* (1.7)*, the TAP free energy evaluated at the Bolthausen approximate fixed points converges to the replica symmetric functional,*

$$\lim_{k\to\infty}\lim_{N\to\infty} N^{-1}\mathrm{TAP}_N(\boldsymbol{m}^{(k)}) = RS(\beta,h) \text{ in } L_1(\mathbb{P}). \tag{1.17}$$

By the above, we shall henceforth refer to Bolthausen's magnetisations as *approximate solutions* of the TAP-equations. We have not found the proof of Theorem 1.3 in the literature and the proof will be given in Section 3. [1] Under the AT condition (1.7), Bolthausen's magnetizations $\boldsymbol{m}^{(k)}$ actually satisfy Plefka's first condition (1.11) with high probability as $N \to \infty$: indeed, it follows from Lemma 2.1 below that

$$\lim_{N\to\infty} \frac{\beta^2}{N} \sum_{i=1}^{N} \left(1 - m_i^{(k)2}\right)^2 = \beta^2 \mathsf{E} \frac{1}{\cosh^4(h + \beta\sqrt{q}Z)} \quad \text{in } L_1(\mathbb{P})\,. \tag{1.18}$$

As a consequence, if the AT condition (1.7) holds with strict inequality, then with probability tending to 1 as $N \to \infty$, Bolthausen's approximate solution satisfies the first Plefka condition, $\boldsymbol{m}^{(k)} \in P_N^1$. That the AT condition and the first Plefka condition are related for suitable magnetizations was clear to Plefka [25].

In Theorem 1.5 below, we find $\beta > 0, h \neq 0$ satisfying the AT condition for which the TAP free energy is not concave in the Bolthausen magnetizations $\boldsymbol{m}^{(k)}$ for large N, k. It remains unclear to us whether this finding can be interpreted within Plefka's expansion of the Gibbs potential, as Bolthausen's $\boldsymbol{m}^{(k)}$ might not provide a sufficient approximation of the TAP solutions for our purposes, in particular as the non-concavity results only from one single outlier eigenvalue. If one may consider here the Bolthausen magnetizations as TAP solutions, however, then the non-concavity of TAP free energy in $\boldsymbol{m}^{(k)}$ might suggest that even when $\boldsymbol{m}^{(k)} \in P_N^1$, the higher-order terms *cannot* be neglected, and additional conditions besides the first Plefka condition are necessary. As we have in Theorem 1.5 a positive eigenvalue of the Hessian of TAP_N in $\boldsymbol{m}^{(k)}$, and as $N^{-1}\mathrm{TAP}_N(\boldsymbol{m}^{(k)})$ converges to $RS(\beta, h)$, a natural question which we do not address in the present paper also arises, namely whether a Taylor expansion around $\boldsymbol{m}^{(k)}$ yields TAP free energies larger than $RS(\beta, h)$ in the $N \to \infty$, $k \to \infty$ limit, or whether this is prevented by higher order terms in the Taylor expansion.

We now investigate the concavity of the $\mathrm{TAP}_N(\boldsymbol{m})$ functional in the Bolthausen magnetizations, that is, we study the Hessian $\boldsymbol{H}^{(k)}$ of the TAP free energy evaluated in $\boldsymbol{m}^{(k)}$,

$$\boldsymbol{H}^{(k)} := \frac{\partial^2}{\partial m_i\, \partial m_j} \mathrm{TAP}_N(\boldsymbol{m})\Big|_{\boldsymbol{m}=\boldsymbol{m}^{(k)}}. \tag{1.19}$$

First we consider the weak limit of the empirical distribution of the eigenvalues $\lambda_i(\boldsymbol{H}^{(k)})$ $(i = 1, \ldots, N)$ of $\boldsymbol{H}^{(k)}$,

$$\mu_{\boldsymbol{H}^{(k)}} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(\boldsymbol{H}^{(k)})}, \tag{1.20}$$

which we show to be concentrated strictly below 0 if the AT condition (1.7) holds with strict inequality, and to touch zero if (1.7) holds with equality.

Theorem 1.4. *For all $\beta > 0$, $h \neq 0$ satisfying condition* (1.7)*, the empirical spectral distribution $\mu_{\boldsymbol{H}^{(k)}}$ converges weakly in distribution as $N \to \infty$ followed by $k \to \infty$ to a deterministic limiting measure μ. If* (1.7) *holds with strict inequality, then $\mu(t, \infty) = 0$ for some $t < 0$. If* (1.7) *holds with equality, then* $\sup\{t \in \mathbb{R} : \mu(t, \infty) > 0\} = 0$.

[1] Véronique Gayrard (private communication) is going to publish the almost sure convergence in Theorem 1.3. She also pointed out that this result is close to Theorem 2 of [13] but with an other kind of magnetization as those given by Bolthausen.

For the proof of Theorem 1.4, we use in Section 5 the explicit control of the weak independence between the disorder (g_{ij}) and the approximate magnetizations $(m_i^{(k)})$, which is given by Bolthausen's algorithm, and we conclude in Section 6 using results from free probability which we recall in Section 4. We remark, however, that Theorem 1.4 and its proof also pass through for magnetizations (m_i) that are assumed to be independent (or sufficiently weakly dependent) of (g_{ij}). Theorem 1.4 ensures that under the AT condition, no positive proportion of the eigenvalues of $\boldsymbol{H}^{(k)}$ becomes positive, in this sense, $\boldsymbol{H}^{(k)}$ does not lose concavity "on a macroscopic scale". If and only if the AT condition holds with *strict* inequality, the right edge of the support of the weak limit of the spectrum is strictly smaller than zero, ensuring that *strict* concavity is not lost "on a macroscopic scale". Spectral interpretations of the AT line in this vein are given non-rigorously in [1] (in the $N \to \infty$ limit), and contained implicitly in [25] (through relation (1.18)). However, outlier eigenvalues, which are too few to have positive mass and thus are not visible in the weak limit, may still lead to a loss of concavity "on a microscopic scale" for large N, k, as we see in the following.

Besides the weak limit of the spectrum, which is related to Plefka's first condition, we are also interested in the question whether the limiting maximal eigenvalue of $\boldsymbol{H}^{(k)}$ can be positive under the AT condition. The positive answer is the main result of this paper and given by the following theorem.

Theorem 1.5. *There exist $\beta > 0$, $h \neq 0$ satisfying the AT condition* (1.7) *with strict inequality, such that with probability tending to* 1 *as $N \to \infty$ followed by $k \to \infty$, the largest eigenvalue of the Hessian $\boldsymbol{H}^{(k)}$ evaluated in the Bolthausen magnetization $\boldsymbol{m}^{(k)}$ is positive and bounded away from zero.*

The proof of this theorem is given in Section 7, where in particular the range of (β, h) for which we obtain positive eigenvalues is depicted in Figure 1. Gayrard (personal communication) will actually prove a complementary result, namely that the strict concavity of the Hessian (i.e., the maximum eigenvalue is negative) for a specific region of the $\boldsymbol{m}$'s which comprises the Bolthausen approximations, and for (β, h) in a region that does not correspond to the AT condition. Her result is also related to the study of the Hessian. The loss of concavity at a microscopic scale was already subject of non-rigorous argumentation by Plefka [25] for finite N. Indeed, Plefka [25] stated another condition

$$P_N^2 := \{m \in [-1,1]^N, \frac{2\beta^2}{N}\sum_{i=1}^{N}(m_i^2 - m_i^4) < 1\} \tag{1.21}$$

which, as Owen [22] pointed out, was based on the assumption that the disorder (g_{ij}) and the magnetizations (m_i) are independent. Other than for the weak limit of the spectrum in Theorem 1.4, we expect that weak dependence (as present in Bolthausen's "approximate solutions") between (g_{ij}) and (m_i) does change the condition for the limiting largest eigenvalue to be positive. Still, Theorem 1.5 suggests that some additional condition besides $\boldsymbol{m}^{(k)} \in P_N^1$ is necessary if we want to ensure the negativity of all eigenvalues of $\boldsymbol{H}^{(k)}$ for large N, k.

2. Bolthausen's iterative procedure

We now recall the algorithm from [8] which we will use throughout the paper. Also throughout the paper, we will assume that $\beta > 0$ and $h \neq 0$. A scalar product on $\mathbb{R}^N$ is given by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := N^{-1} \sum_{i=1}^N x_i y_i$ with associated norm $\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$. Furthermore, $\boldsymbol{x} \otimes \boldsymbol{y} := N^{-1}(x_i y_j)_{ij}$, and for a matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$, we denote its symmetrization by $\bar{\boldsymbol{A}} := \frac{1}{\sqrt{2}}(\boldsymbol{A} + \boldsymbol{A}^T)$.

Let $\boldsymbol{g} = (g_{ij})_{i,j=1,\ldots,N}$ be an array of independent centered Gaussians with variance 1. The interaction matrix will be its symmetrization $\bar{\boldsymbol{g}}$, normalized by $N^{-1/2}$. Let $\psi : [0,q] \to [0,q]$ be defined by

$$\psi(t) = \mathsf{E} \tanh\left(h + \beta\sqrt{t}Z + \beta\sqrt{q-t}Z'\right) \tanh\left(h + \beta\sqrt{t}Z + \beta\sqrt{q-t}Z''\right), \tag{2.1}$$

where Z, Z', Z'' are independent standard Gaussians. Then set

$$\gamma_1 := \mathsf{E} \tanh\left(h + \beta\sqrt{q}Z\right), \quad \rho_1 := \sqrt{q}\gamma_1 \tag{2.2}$$

and

$$\rho_k := \psi(\rho_{k-1}), \quad \gamma_k := \frac{\rho_k - \sum_{j=1}^{k-1} \gamma_j^2}{\sqrt{q - \sum_{j=1}^{k-1} \gamma_j^2}}. \tag{2.3}$$

Let $\boldsymbol{g}^{(1)} := \boldsymbol{g}$, $\phi^{(1)} = \mathbf{1}$, $\boldsymbol{m}^{(1)} = \sqrt{q}\mathbf{1}$. With the shorthand $\Gamma_k^2 := \sum_{j=1}^k \gamma_j^2$, we set recursively for $k \in \mathbb{N}$

$$\xi^{(k)} = \tfrac{1}{\sqrt{N}} \boldsymbol{g}^{(k)} \phi^{(k)}, \quad \eta^{(k)} = \tfrac{1}{\sqrt{N}} {\boldsymbol{g}^{(k)}}^T \phi^{(k)}, \quad \zeta^{(k)} = \frac{1}{\sqrt{2}}\left(\xi^{(k)} + \eta^{(k)}\right), \tag{2.4}$$

$$\boldsymbol{h}^{(k+1)} = h\mathbf{1} + \beta \sum_{s=1}^{k-1} \gamma_s \zeta^{(s)} + \beta\sqrt{q - \Gamma_{k-1}^2}\zeta^{(k)}, \tag{2.5}$$

$$\boldsymbol{m}^{(k+1)} = \tanh(\boldsymbol{h}^{(k+1)}), \tag{2.6}$$

moreover $\{\phi^{(1)}, \ldots, \phi^{(k+1)}\}$ as the Gram-Schmidt orthonormalization of $\{\boldsymbol{m}^{(1)}, \ldots, \boldsymbol{m}^{(k+1)}\}$,

$$\phi^{(k+1)} = \frac{\boldsymbol{m}^{(k+1)} - \sum_{s=1}^k \langle \phi^{(s)}, \boldsymbol{m}^{(k+1)} \rangle \phi^{(s)}}{\left\|\boldsymbol{m}^{(k+1)} - \sum_{s=1}^k \langle \phi^{(s)}, \boldsymbol{m}^{(k+1)} \rangle \phi^{(s)}\right\|}, \tag{2.7}$$

and the modifications of the interaction matrices

$$\boldsymbol{g}^{(k+1)} = \boldsymbol{g}^{(k)} - \sqrt{N}\rho^{(k)}, \tag{2.8}$$

where

$$\rho^{(k)} = \xi^{(k)} \otimes \phi^{(k)} + \phi^{(k)} \otimes \eta^{(k)} - \left\langle \phi^{(k)}, \xi^{(k)} \right\rangle \left(\phi^{(k)} \otimes \phi^{(k)}\right). \tag{2.9}$$

By Lemma 2b of [8], we have

$$\sum_{s=1}^{\infty} \gamma_s^2 = q. \tag{2.10}$$

Noting that $\{\phi^{(s)}\}_{s \le k}$ are orthonormal with respect to $\langle \cdot, \cdot \rangle$, we define

$$P_{ij}^{(k)} = \frac{1}{N} \sum_{s=1}^k \phi_i^{(s)} \phi_j^{(s)}, \tag{2.11}$$

and one readily checks that $\boldsymbol{P}^{(k)}$ is an orthogonal projection. Furthermore, let

$$\mathcal{G}_k = \sigma\left(\xi^{(m)}, \eta^{(m)} : m \leq k\right). \tag{2.12}$$

Then $\zeta^{(k)}$ is $\mathcal{G}_k$-measurable and $\boldsymbol{m}^{(k)}$ is $\mathcal{G}_{k-1}$-measurable. Moreover, by Proposition 4 of [8], $\boldsymbol{g}^{(k)}$ is centered Gaussian under $\mathbb{P}(\cdot \mid \mathcal{G}_{k-1})$ with covariances given by

$$V^{(k)}_{ij,st} := \mathbb{E}\Big(g^{(k)}_{ij} g^{(k)}_{st} \,\Big|\, \mathcal{G}_{k-1}\Big) = Q^{(k-1)}_{is} Q^{(k-1)}_{jt}, \tag{2.13}$$

where $\boldsymbol{Q}^{(k)} = (Q^{(k)}_{ij})_{ij\leq N} = \mathbf{1} - \boldsymbol{P}^{(k)}$. As we show in Lemma 5.1 below, this covariance matrix itself is a projection.[2]

If X_N, Y_N are two sequences of random variables, we write

$$X_N \simeq Y_N, \tag{2.14}$$

if there exists a constant $C > 0$, depending possibly on other parameters, but not on N, with

$$\mathbb{P}\big(|X_N - Y_N| > t\big) \leq C\mathrm{e}^{-t^2 N/C}. \tag{2.15}$$

$X_N \simeq Y_N$ in particular implies $X_N - Y_N \to 0$ in $L_p(\mathbb{P})$ for every $p > 0$ as $N \to \infty$. By Proposition 6 of [8], we have

$$\|\boldsymbol{m}^{(k)}\| \simeq q \tag{2.16}$$

for each $k \in \mathbb{N}$.

We will prove in the appendix the following version of the law of large numbers given in Lemma 14 of [8]:

Lemma 2.1. *(Law of large numbers) For any continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ with $|f(x)| \leq Ce^{C|x|}$ for some constant $C < \infty$, and any $k \geq 2$, we have for $\beta > 0$, $h \neq 0$ satisfying* (1.7) *that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f\left(h^{(k)}_i\right) = \mathsf{E} f(h + \beta\sqrt{q} Z), \tag{2.17}$$

in $L_1(\mathbb{P})$. Moreover, there exists a coupling of $\{\boldsymbol{h}^{(k)}\}_N$ such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{h^{(k)}_i} = \mathcal{L}(h + \beta\sqrt{q} Z), \tag{2.18}$$

a. s. with respect to the weak topology on $\mathbb{R}$.

3. Replica symmetric formula for the TAP free energy

To prove Theorem 1.3, we will use the following lemma which will be proved in the appendix.

[2]As we want this covariance to be a projection, we define the entries of $\boldsymbol{g}$ with unit variance, while [8] defines them with variance $1/N$. As a consequence, we have to carry along the scaling factor $N^{-1/2}$ when $\boldsymbol{g}$ is used.

Lemma 3.1. *Let*

$$\Delta^{(k)} = \tanh^{-1}(\boldsymbol{m}^{(k)}) - h\mathbf{1} - \frac{\beta}{\sqrt{N}}\bar{\boldsymbol{g}}\boldsymbol{m}^{(k)} + \beta^2(1-q)\boldsymbol{m}^{(k)}, \tag{3.1}$$

and assume that $\beta > 0$, $h \neq 0$ satisfies the AT condition (1.7)*. Then,*

$$\|\Delta^{(k)}\| \to 0 \quad \textit{in } L_2(\mathbb{P}) \quad \textit{as } N \to \infty \textit{ followed by } k \to \infty\,. \tag{3.2}$$

Remark 3.2. *A variant of Lemma 3.1 bridges the papers* [7] *and* [8] *of Bolthausen: in* [7]*, "approximate solutions" similar to those used in the present paper are constructed using a two-step Banach algorithm (see also* [18] *for a recent analysis of such algorithms). The iterative procedure from* [8] *that we use in the present paper is close to the one from* [7] *as it approximately satisfies this two-step recursion under the AT condition* (1.7)*, namely*

$$\left\| \tanh^{-1}(\boldsymbol{m}^{(k+1)}) - h\mathbf{1} - \frac{\beta}{\sqrt{N}}\Big(\sum_{j\neq i}\bar{g}_{ij}m_j^{(k)}\Big)_i + \beta^2(1-q)\boldsymbol{m}^{(k-1)}\right\| \to 0 \tag{3.3}$$

in $L_2(\mathbb{P})$ as $N \to \infty$ followed by $k \to \infty$. The relation (3.3) *is proved along the lines of the proof of Lemma 3.1.*

We are now ready to prove the convergence of the TAP functional to the replica-symmetric free energy:

Proof of Theorem 1.3. To see how this goes, we first reformulate (1.8) with the help of (1.9) with the right scaling,

$$\begin{aligned} N^{-1}\mathrm{TAP}_N(\boldsymbol{m}^{(k)}) = \tfrac{\beta}{2}N^{-3/2}\sum_{i\neq j}^{N}\bar{g}_{ij}m_i^{(k)}m_j^{(k)} + hN^{-1}\sum_{i=1}^{N}m_i^{(k)} - N^{-1}\sum_{i=1}^{N}\tanh^{-1}(m_i^{(k)})m_i^{(k)} \\ + N^{-1}\sum_{i=1}^{N}\log\cosh\tanh^{-1}(m_i^{(k)}) + \frac{\beta^2}{4}\left(1 - \frac{1}{N}\sum_{i=1}^{N}{m_i^{(k)}}^2\right)^2. \end{aligned} \tag{3.4}$$

By Lemma 2.1, the terms in the second line of the latter converge in $L_1(\mathbb{P})$ to the r.h.s. of (1.17) as $N \to \infty$.

It remains to show that the sum of the first three terms on the r.h.s. of (3.4) converges to 0 in $L_1(\mathbb{P})$ as $N \to \infty$, followed by $k \to \infty$. By Lemma 2.1, first note that the limit in $L_1(\mathbb{P})$ of the second and third term on the right-hand side of (3.4) is

$$\begin{aligned} \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\left(h - \tanh^{-1}(m_i^{(k)})\right)m_i^{(k)} &= -\mathsf{E}\left(\beta\sqrt{q}Z\tanh\left(h+\beta\sqrt{q}Z\right)\right) \\ &= -\beta^2 q(1-q), \end{aligned} \tag{3.5}$$

the last line by combining a simple integration by parts with (1.6). It only remains to prove that the first term on the r.h.s. (3.4) tends to $\beta^2 q(1-q)$. By Lemma 3.1, it holds

$$\frac{\beta}{\sqrt{N}}\sum_{j:\, j\neq i}\bar{g}_{ij}m_j^{(k)} = -h + \beta^2(1-q)m_i^{(k)} + \tanh^{-1}(m_i^{(k)}) - \Delta_i^{(k)} - \frac{\beta}{\sqrt{N}}\bar{g}_{ii}m_i^{(k)}, \tag{3.6}$$

Multiplying the latter by $\frac{m_i^{(k)}}{2N}$ and taking the sum over i yield

$$\tfrac{1}{2}\beta N^{-3/2}\sum_{i\neq j}^{N}\bar{g}_{ij}m_i^{(k)}m_j^{(k)} = -\tfrac{1}{2}hN^{-1}\sum_{i=1}^{N}m_i^{(k)} + \tfrac{1}{2}\beta^2(1-q)N^{-1}\sum_{i=1}^{N}{m_i^{(k)}}^2$$
$$+\tfrac{1}{2}N^{-1}\sum_{i=1}^{N}m_i^{(k)}\tanh^{-1}(m_i^{(k)}) - \tfrac{1}{2}N^{-1}\sum_{i=1}^{N}m_i^{(k)}\Delta_i^{(k)} - \frac{\beta}{2N\sqrt{N}}\sum_{i=1}^{N}\bar{g}_{ii}{m_i^{(k)}}^2. \quad (3.7)$$

The last term on the r.h.s. of (3.7) tends to 0 in $L_2(\mathbb{P})$ as $N\to\infty$ as

$$\mathbb{E}\frac{1}{N}\sum_{i=1}^{N}\left|\frac{\beta}{\sqrt{N}}\bar{g}_{ii}m_i^{(k)}\right|^2 \le \frac{\beta}{N^2}\sum_{i=1}^{N}\mathbb{E}\bar{g}_{ii}^2 = \frac{\beta}{N}. \quad (3.8)$$

Combining Cauchy-Schwarz with Lemma 3.1 and (2.16), we have that the second last term on the r.h.s. of (3.7) tends to 0 in $L_2(\mathbb{P})$ as $N\to\infty$ followed by $k\to\infty$. The sum of the remaining terms on the r.h.s. of (3.7) converges, as $N\to\infty$ in $L_1(\mathbb{P})$ by Lemma 2.1 to

$$-\tfrac{1}{2}h\mathsf{E}\left(\tanh\left(h+\beta\sqrt{q}Z\right)\right) + \tfrac{1}{2}\beta^2(1-q)\mathsf{E}\left(\tanh^2\left(h+\beta\sqrt{q}Z\right)\right)$$
$$+\tfrac{1}{2}\mathsf{E}\left(\tanh\left(h+\beta\sqrt{q}Z\right)\left(h+\beta\sqrt{q}Z\right)\right) = \beta^2q(1-q), \quad (3.9)$$

again using (3.5) and (1.6). All in all, we obtain that

$$\lim_{N\to\infty,\,k\to\infty}\tfrac{1}{2}\beta N^{-\frac{3}{2}}\sum_{i\neq j}^{N}\bar{g}_{ij}m_i^{(k)}m_j^{(k)} = \beta^2q(1-q), \text{in } L_1(\mathbb{P}). \quad (3.10)$$

We proved that the first term on the r.h.s. of (3.4) tends to $\beta^2q(1-q)$ and the assertion of Theorem 1.3 follows. □

4. Gaussian orthogonal ensemble

As a tool to study the Hessian of the TAP free energy functional, we record some known facts about the Gaussian orthogonal ensemble (GOE). A GOE with variance $\sigma^2>0$ is a real symmetric random matrix $\boldsymbol{X}$ with centered Gaussian entries of variance σ^2 off the diagonal, variance $2\sigma^2$ on the diagonal, and the entries $(X_{ij})_{1\le i\le j\le N}$ being independent. The matrix $(\beta N^{-1/2}\bar{g}_{ij})_{i,j=1,\ldots,N}$ is a GOE with variance β^2/N. Thus, by Wigner's Theorem (see e. g. Theorem 2.1.1 in [3]), its empirical spectral distribution converges weakly in probability to the semicircle law μ_β which is defined by its density

$$\frac{\mathrm{d}\mu_\beta(x)}{\mathrm{d}x} = 1_{[-2\beta,2\beta]}(x)\sqrt{4\beta^2-x^2}\,. \quad (4.1)$$

Also, the largest eigenvalue $\lambda_1(\beta N^{-1/2}\bar{\boldsymbol{g}})$ converges a. s. to 2β (see e. g. Theorem 1.13 of [31]).

For each real symmetric matrix $\boldsymbol{M}$ of size n, we denote the enumeration of its eigenvalues in non-increasing order by $\lambda_1(\boldsymbol{M})\ge\ldots\ge\lambda_n(\boldsymbol{M})$, and its empirical spectral distribution by

$$\mu_{\boldsymbol{M}} := \frac{1}{N}\sum_{i=1}^{n} \delta_{\lambda_i(\boldsymbol{M})}. \tag{4.2}$$

We recall that the Frobenius norm of a matrix $\boldsymbol{M}$ of size n is defined by $\|\boldsymbol{M}\|_{\mathrm{F}} = (\sum_{i,j=1}^{N} |M_{ij}|^2)^{1/2}$. The following standard result, for which we refer to Exercises 2.4.3 and 2.4.4 of [30], states that the limiting empirical spectral distributions of random matrices are invariant under additive perturbations in the prelimiting sequence that have either small rank or small Frobenius norm.

Lemma 4.1. *Let $\boldsymbol{M}_n$ and $\boldsymbol{N}_n$ be random Hermitian matrices of size n such that the empirical spectral distribution of $\boldsymbol{M}_n$ converges weakly a. s. to a probability measure μ. Suppose that at least one of the following conditions holds true:*

(i) $n^{-1}\|\boldsymbol{N}_n\|_{\mathrm{F}}^2 \to 0$ *a. s.* ,
(ii) $n^{-1}\mathrm{rank}(\boldsymbol{N}_n) \to 0$ *a. s.* .

Then the empirical spectral distribution of $\boldsymbol{M}_n + \boldsymbol{N}_n$ converges to μ weakly a. s. .

Similarly, for the largest eigenvalue we have:

Lemma 4.2. *Let $\boldsymbol{M}_n$ and $\boldsymbol{N}_n$ be random Hermitian matrices of size n such that the largest eigenvalue $\lambda_1(\boldsymbol{M}_n)$ of $\boldsymbol{M}_n$ converges a. s. to a limit λ_1 as $n \to \infty$. Suppose that $\|\boldsymbol{N}_n\|_{\mathrm{F}}^2 \to 0$ a. s. Then also the largest eigenvalue $\lambda_1(\boldsymbol{M}_n + \boldsymbol{N}_n)$ of $\boldsymbol{M}_n + \boldsymbol{N}_n$ converges to the same limit $\hat{\lambda}_1$ almost surely.*

Proof. This follows from the Hoffman-Wielandt inequality (below Lemma 2.4.3 of [30]). □

4.1. **Free convolution.** First we state a definition of the free convolution (see [6, 34, 21]). The Stieltjes transform of a probability measure μ on $\mathbb{R}$ is defined by

$$g_\mu(z) = \int \frac{\mathrm{d}\mu(x)}{z - x} \tag{4.3}$$

which is analytic in $\mathbb{C} \setminus \mathrm{supp}\,\mu$. It can be shown that there exists a domain D on which g_μ is univalent. Denoting by K_μ the inverse function of g_μ defined on $g_\mu(D)$, the R-transform of μ is defined on $g_\mu(D)$ by

$$R_\mu(z) = K_\mu(z) - \frac{1}{z}\,. \tag{4.4}$$

Free probability theory shows that for probability measures μ, λ on $\mathbb{R}$, there exists a unique probability measure κ with

$$R_\kappa = R_\lambda + R_\mu \tag{4.5}$$

on a domain on which these three functionals are defined. The measure κ is denoted by $\lambda \boxplus \mu$ and called the free (additive) convolution of λ and μ.

The following result ensures that limiting spectral distribution of a sum of a GOE and a deterministic matrix whose spectral distribution weakly converges is given by a free additive convolution with the semicircle law. The support of this free convolution is analyzed in Lemma 6.1 below.

Lemma 4.3. *For $n \in \mathbb{N}$, let $\boldsymbol{X}_n$ be a GOE with unit variance, and let $\boldsymbol{A}_n$ be a deterministic real and symmetric matrix, each of size n, such that the empirical spectral distribution $\mu_{\boldsymbol{A}_n}$ converges weakly to some probability measure ν on $\mathbb{R}$ as $n \to \infty$. Then, for each $\sigma > 0$, the empirical spectral distribution of $\sigma n^{-1/2}\boldsymbol{X}_n + \boldsymbol{A}_n$ converges weakly almost surely to $\mu_\sigma \boxplus \nu$.*

Proof. This is a standard result from free probability theory, see for example Theorem 5.4.5 in [3]. Also, Pastur [20] (p.12) gives a functional equation solved by the Stieltjes transform of the limiting spectral distribution of $\sigma n^{-1/2}\boldsymbol{X}_n + \boldsymbol{A}_n$. The Stieltjes transform of the limiting distribution solves a functional equation, which has a unique solution [24], (p.69). We conclude with the fact that the Stieltjes transform of $\mu_\sigma \boxplus \nu$ solves the same functional equation (see e.g. Proposition 2.1 in [11]). □

We will also use the following version of a result of Capitaine et al. [11] for the largest eigenvalue. For $\sigma > 0$ and a probability measure ν on $\mathbb{R}$, let

$$H_{\sigma,\nu}(z) := z + \sigma^2 g_\nu(z) \tag{4.6}$$

and

$$\mathcal{O}_{\sigma,\nu} := \{u \in \mathbb{R} \setminus \operatorname{supp}\nu \ : H'_{\sigma,\nu}(u) > 0\} \tag{4.7}$$

where g_ν denotes the Stieltjes transform defined as in (4.3).

Lemma 4.4 (cf. [11], Theorem 8.1)**.** *Let $\sigma > 0$, let $\boldsymbol{X}_N$ be a GOE with unit variance, and let $\boldsymbol{A}_N$ be a deterministic real and symmetric matrix. Assume that the empirical spectral distribution $\mu_{\boldsymbol{A}_N}$ converges weakly to a probability measure ν on $\mathbb{R}$ as $N \to \infty$, and that there exists $d \in \mathbb{R}$ with $\nu(d,\infty) = 0$. Also, suppose that there exist an integer $r \geq 2$ and $\theta \in \mathcal{O}_{\sigma,\nu}$ with $\lim_{N\to\infty} \lambda_1(\boldsymbol{A}_N) = \theta$ and*

$$\max_{j=r,\dots,N} \mathrm{d}(\lambda_j(\boldsymbol{A}_N), \operatorname{supp}(\nu)) \overset{N\to\infty}{\longrightarrow} 0. \tag{4.8}$$

Then $\lim_{N\to\infty} \lambda_1(\sigma N^{-1/2}\boldsymbol{X}_N + \boldsymbol{A}_N) = H_{\sigma,\nu}(\theta)$ almost surely.

The proof of this lemma is discussed in Appendix B.

5. Conditional Hessian

To analyze the spectral behavior of the Hessian $\boldsymbol{H}^{(k)}$ from (1.19), it is useful to condition on the σ-algebra $\mathcal{G}_{k-1}$ with respect to which the magnetization $\boldsymbol{m}^{(k)}$ is measurable. Under this conditioning, $\boldsymbol{g}^{(k)}$ remains centered Gaussian with covariances given by (2.13). In the present section, we show that up to a negligible additive error, $\bar{\boldsymbol{g}}^{(k)}$ can be considered as a GOE also under the conditioning on $\mathcal{G}_{k-1}$. Thus we obtain a representation of $\boldsymbol{H}^{(k)}$ as the sum of a GOE and independent $\mathcal{G}_{k-1}$-measurable terms.

First we give some properties of the covariance matrix $\boldsymbol{V}^{(k)} = (V^{(k)}_{ij,st})_{i,j,s,t\leq N}$ of $\boldsymbol{g}^{(k)}$ under $\mathbb{P}(\cdot \mid \mathcal{G}_{k-1})$ which follow from its definition (2.13) in terms of the projection $\boldsymbol{Q}^{(k)}$. In the following, we will denote by $\mathbb{P}_{k-1} := \mathbb{P}(\cdot \mid \mathcal{G}_{k-1})$ with associated expectation $\mathbb{E}_{k-1}$ the conditional probability given $\mathcal{G}_{k-1}$.

Lemma 5.1. *The matrix $\boldsymbol{V}^{(k)}$ is a projection, that is, $\boldsymbol{V}^{(k)} = \boldsymbol{V}^{(k)2}$. Furthermore, $\boldsymbol{V}^{(k)} = \boldsymbol{1} + \boldsymbol{J}$ for a matrix $\boldsymbol{J}$, where $\boldsymbol{J}$ has eigenvalue -1 with multiplicity $(k-1)^2$, and all other eigenvalues are zero.*

Proof. By (2.11) and as $\boldsymbol{Q}^{(k-1)}$ is a projection,

$$V^{(k)}_{ij,st} = Q^{(k-1)}_{is} Q^{(k-1)}_{jt} = \left(\sum_{u=1}^{N} Q^{(k-1)}_{iu} Q^{(k-1)}_{us}\right)\left(\sum_{v=1}^{N} Q^{(k-1)}_{jv} Q^{(k-1)}_{vt}\right) = \sum_{u,v=1}^{N} V^{(k)}_{ij,uv} V^{(k)}_{uv,st}, \tag{5.1}$$

which shows that $\boldsymbol{V}^{(k)}$ is a projection.

To show the assertion on the eigenvalues of $\boldsymbol{J}$, we first note that

$$\boldsymbol{Q}^{(k-1)} = \mathbf{1} - \boldsymbol{P}^{(k-1)} = \mathbf{1} - \frac{1}{N}\sum_{s=1}^{k-1} \phi^{(s)} {\phi^{(s)}}^T = \mathbf{1} - \boldsymbol{O}\left(\sum_{s=1}^{k-1} \boldsymbol{D}^{(s)}\right)\boldsymbol{O}^T, \tag{5.2}$$

where $\boldsymbol{O}$ is an orthogonal matrix and $\boldsymbol{D}^{(s)}$ are diagonal matrices with one coordinate equal to 1, and the rest is equal to 0. The last equality is due to the fact that $\boldsymbol{P}^{(k-1)}$ is a sum of projectors of rank 1 to orthogonal subspaces: thus, these projectors are orthogonally diagonalisable in the same basis. Let

$$\boldsymbol{D} = \mathbf{1} - \sum_{s=1}^{k-1} \boldsymbol{D}^{(s)}, \tag{5.3}$$

one readily checks that $\boldsymbol{D}$ has $k-1$ entries that are equal to 0 and the rest equal to 1. Defining $\boldsymbol{J}$ by $\boldsymbol{V} = \mathbf{1} - \boldsymbol{J}$ and using the definition (2.13) of $\boldsymbol{V}$, we obtain

$$J_{ij,st} = Q^{(k-1)}_{is} Q^{(k-1)}_{jt} - \delta_{ij,st} = \sum_{u,v=1}^{N} O_{ui} D_{uu} O_{us} O_{vj} D_{vv} O_{vt} - \delta_{ij,st}. \tag{5.4}$$

Next we define $\widetilde{\boldsymbol{O}}$ and $\widetilde{\boldsymbol{D}}$ by $\widetilde{O}_{ij,st} = O_{is} O_{jt}$ and $\widetilde{D}_{ij,st} = D_{is} D_{jt}$. Then $\widetilde{\boldsymbol{O}}$ is orthogonal as

$$(\widetilde{\boldsymbol{O}}^T \widetilde{\boldsymbol{O}})_{ij,st} = \sum_{u,v=1}^{N} O_{i,u} O_{j,v} O_{u,s} O_{v,t} = (\boldsymbol{O}^T \boldsymbol{O})_{is} (\boldsymbol{O}^T \boldsymbol{O})_{jt} = \delta_{is}\delta_{jt}. \tag{5.5}$$

Hence, we get from (5.4) that $\boldsymbol{J} = \widetilde{\boldsymbol{O}}^T(\widetilde{\boldsymbol{D}} - \mathbf{1})\widetilde{\boldsymbol{O}}$. As the diagonal matrix $(\widetilde{\boldsymbol{D}} - \mathbf{1})$ has $(k-1)^2$ entries equal to -1, the other entries being zero, the assertion follows. □

As a consequence, we can approximate $\bar{\boldsymbol{g}}^{(k)}$ by a GOE:

Lemma 5.2. *Under* $\mathbb{P}(\cdot \mid \mathcal{G}_{k-1})$*, there exists a GOE* $\mathbf{X}$ *such that* $\bar{\boldsymbol{g}}^{(k)} = \boldsymbol{X} + \boldsymbol{Y}$ *and* $\|\boldsymbol{Y}\|_{\mathrm{F}}$ *is tight in* N*.*

Proof. From (2.13) and as $\boldsymbol{V}^{(k)} = {\boldsymbol{V}^{(k)}}^2$ by Lemma 5.1, there exists a vector $\boldsymbol{Z} = (Z_{ij})_{ij \le N}$ of length N^2 whose entries are iid standard Gaussians, such that $g^{(k)}_{ij} = (\boldsymbol{V}^{(k)}\boldsymbol{Z})_{ij}$ for all $i, j \le N$. Using again Lemma 5.1, we diagonalize $\boldsymbol{V}^{(k)} - \mathbf{1} = \boldsymbol{O}^T \boldsymbol{D} \boldsymbol{O}$, where $\boldsymbol{O}$ is an orthogonal matrix, and $\boldsymbol{D}$ is a deterministic diagonal matrix with $(k-1)^2$ entries equal to -1 and the rest to 0. Then we write

$$g^{(k)}_{ij} = \big((\boldsymbol{V}^{(k)} - \mathbf{1})\boldsymbol{Z}\big)_{ij} + Z_{ij} = \big(\boldsymbol{O}^T \boldsymbol{D} \boldsymbol{O} \boldsymbol{Z}\big)_{ij} + Z_{ij}. \tag{5.6}$$

We have $\bar{\boldsymbol{g}} = \boldsymbol{Y} + \boldsymbol{X}$, where

$$Y_{ij} := \frac{1}{\sqrt{2}}[\big(\boldsymbol{O}^T\boldsymbol{DOZ}\big)_{ij} + \big(\boldsymbol{O}^T\boldsymbol{DOZ}\big)_{ji}], \quad X_{ij} := \frac{1}{\sqrt{2}}[Z_{ij} + Z_{ji}], \tag{5.7}$$

where $\boldsymbol{X}, \boldsymbol{Y}$ are $N \times N$ matrices. It remains to prove that $\|\boldsymbol{Y}\|_{\mathrm{F}}$ is tight in N. By a simple convexity argument,

$$\|\boldsymbol{Y}\|_{\mathrm{F}}^2 = \sum_{i,j=1}^N Y_{ij}^2 \leq \sqrt{2} \sum_{i,j=1}^N \left[\big(\boldsymbol{O}^T\boldsymbol{DOZ}\big)_{ij}^2 + \big(\boldsymbol{O}^T\boldsymbol{DOZ}\big)_{ji}^2\right]. \tag{5.8}$$

By symmetry, it remains to consider

$$\sum_{i,j=1}^N \big(\boldsymbol{O}^T\boldsymbol{DOZ}\big)_{ij}^2 = \|\boldsymbol{O}^T\boldsymbol{DOZ}\|_{\ell_2(\mathbb{R}^{N^2})} \tag{5.9}$$

and to show that this expression is tight in N. As the ℓ_2-norm is invariant under orthogonal transformations, and as $\boldsymbol{OZ}$ is again standard Gaussian distributed, we have

$$\|\boldsymbol{O}^T\boldsymbol{DOZ}\|_{\ell_2(\mathbb{R}^{N^2})} = \|\boldsymbol{DOZ}\|_{\ell_2(\mathbb{R}^{N^2})} \stackrel{d}{=} \|\boldsymbol{DZ}\|_{\ell_2(\mathbb{R}^{N^2})}. \tag{5.10}$$

Note that $(k-1)^2$ many components of the vector $\boldsymbol{DZ}$ are $\mathcal{N}(0,1)$-distributed, the other components being 0. Therefore, $\|\boldsymbol{DZ}\|_{\ell_2(\mathbb{R}^{N^2})}$ is tight in N, which yields the assertion. □

We consider the Hessian $\boldsymbol{H}^{(k)}$ from (1.19) which reads

$$H_{ij}^{(k)} = \frac{\beta}{\sqrt{N}}\bar{g}_{ij} + \frac{2\beta^2}{N} m_i^{(k)} m_j^{(k)}, \quad i,j = 1,\ldots,N, i \neq j$$
$$H_{ii}^{(k)} = -\beta^2\left(1 - \frac{1}{N}\sum_{p=1}^N {m_p^{(k)}}^2\right) - \frac{1}{1 - {m_i^{(k)}}^2} + \frac{2\beta^2}{N}{m_i^{(k)}}^2. \tag{5.11}$$

Now we obtain the following approximation under $\mathbb{P}$:

Lemma 5.3. *Let $\boldsymbol{A}^{(k)}$ be defined by*

$$A_{ii}^{(k)} = -\frac{1}{1 - {m_i^{(k)}}^2}, i = 1,\ldots,N, \quad A_{ij}^{(k)} = 0 \text{ for } i \neq j, \tag{5.12}$$

and let

$$\boldsymbol{B}^{(k)} = 2\beta^2 \boldsymbol{m}^{(k)} \otimes \boldsymbol{m}^{(k)} + \beta \sum_{s=1}^{k-1} \big(\zeta^{(s)} \otimes \phi^{(s)} + \phi^{(s)} \otimes \zeta^{(s)}\big). \tag{5.13}$$

Then, with $\boldsymbol{X}$ from Lemma 5.2,

$$\boldsymbol{H}^{(k)} = \frac{\beta}{\sqrt{N}}\boldsymbol{X} + \boldsymbol{A}^{(k)} + \boldsymbol{B}^{(k)} - \beta^2(1-q)\mathbf{1} + \boldsymbol{R} - \epsilon\mathbf{1} \tag{5.14}$$

where, in $\mathbb{P}(\cdot \mid \mathcal{G}_{k-1})$-probability, $\|\boldsymbol{R}\|_{\mathrm{F}} \to 0$ and $\epsilon \to 0$, as $N \to \infty$.

Proof. We set $\epsilon = \beta^2(q - \|\boldsymbol{m}^{(k)}\|^2)$, then $\epsilon \to 0$ in probability as $N \to \infty$ by (2.16). Using the definitions (5.11), (2.8) and (2.9), we can then set

$$\boldsymbol{R} = -\sqrt{2}\sum_{s=1}^{k-1}\langle \phi^{(s)}, \xi^{(s)}\rangle(\phi^{(s)} \otimes \phi^{(s)}) + \beta N^{-1/2}\boldsymbol{Y} \tag{5.15}$$

with $\boldsymbol{Y}$ from Lemma 5.2, so that $N^{-1/2}\|\boldsymbol{Y}\|_F$ converges to zero in probability. For the first term on the r.h.s., we note that $\|\phi^{(s)} \otimes \phi^{(s)}\|_{\mathrm{F}}^2 = \|\phi^{(s)}\|^2 = 1$, hence it suffices to show for each s that $\langle \phi^{(s)}, \zeta^{(s)}\rangle \to 0$ in $\mathbb{P}(\cdot \mid \mathcal{G}_{k-1})$-probability as $N \to \infty$. This, however, follows from Lemma 11 of [8] which states that $\langle \phi^{(s)}, \xi^{(s)}\rangle$ is a centered Gaussian with variance $1/N$ under $\mathbb{P}$, hence it converges to 0 $\mathbb{P}$-a. s with Borel-Cantelli. □

6. Proof for weak limit of spectral distribution

The proof of Theorem 1.4 comes in two parts: first we show, using Bolthausen's algorithm, that $\boldsymbol{H}^{(k)}$ can be considered asymptotically as $N \to \infty$ followed by $k \to \infty$ as the sum of a GOE with variance β/N, a deterministic diagonal matrix $-\beta^2(1-q)\mathbf{1}$, and an independent diagonal matrix with independent entries having distribution

$$\nu := \mathcal{L}\left(-\frac{1}{1-\tanh^2(h+\beta\sqrt{q}Z)}\right), \tag{6.1}$$

Z being a standard Gaussian. The spectrum of such a sum can be characterized as a free convolution. We also set $\hat{\nu} := \nu(\cdot + \beta^2(1-q))$, then $\hat{\nu}$ is the image measure of ν under the shift $t \mapsto t - \beta^2(1-q)$.

Proof of Theorem 1.4. We can rewrite $\boldsymbol{B}^{(k)}$ as

$$\boldsymbol{B}^{(k)} = 2\beta^2\boldsymbol{m}^{(k)}\otimes\boldsymbol{m}^{(k)} + \tfrac{1}{2}\beta\sum_{s=1}^{k-1}\left[\left(\zeta^{(s)}+\phi^{(s)}\right)\otimes\left(\zeta^{(s)}+\phi^{(s)}\right) - \left(\zeta^{(s)}-\phi^{(s)}\right)\otimes\left(\zeta^{(s)}-\phi^{(s)}\right)\right] \tag{6.2}$$

which is a sum of $2k-1$ matrices of rank 1. Hence, by Lemma 4.1 (and induction over k), $\boldsymbol{B}^{(k)}$ has no influence on the limiting spectral distribution of $\boldsymbol{H}^{(k)}$ as $N \to \infty$. Thus, the empirical spectral distribution of $\boldsymbol{M} := \beta N^{-1/2}\boldsymbol{X} + \boldsymbol{A}^{(k)} - \beta^2(1-q)\mathbf{1}$ converges by Lemmas 4.1 and 5.3 and Slutzky's lemma to the same weak limit as $\mu_{\boldsymbol{H}^{(k)}}$ a. s. as $N \to \infty$ followed by $k \to \infty$.

By Lemma 4.3, the empirical spectral distribution of $\boldsymbol{M}$ converges a. s. in the weak topology as $N \to \infty$ to the free additive convolution $\mu_\beta \boxplus \hat{\nu}$. The assertion now follows from Lemma 6.1 below. □

The support $\operatorname{supp}\mu$ of a probability measure μ on $\mathbb{R}$ is defined by

$$\operatorname{supp}\mu := \mathbb{R} \setminus \{t \in \mathbb{R} : \exists \epsilon > 0 \text{ with } \mu(t-\epsilon, t+\epsilon) = 0\}. \tag{6.3}$$

Lemma 6.1. *The free additive convolution $\mu_\beta \boxplus \nu$ has support of the form $(-\infty, d]$ with $d < \beta^2(1-q)$ below and above the AT line (i.e. if (1.7) holds with strict inequality or if (1.7) does not hold), and $d = \beta^2(1-q)$ on the AT line (i.e. if (1.7) holds with equality).*

Proof. Let $H_{\beta,\nu}(z)$ be defined by (4.6) and $\mathcal{O}_{\beta,\nu}$ by (4.7). From the work of Biane [6], see Proposition 2.2 of [11], we have the equivalence

$$x \in \mathbb{R} \setminus \operatorname{supp} \mu_\beta \boxplus \nu \quad \Leftrightarrow \quad \exists u \in \mathcal{O}_{\beta,\nu} \text{ such that } x = H_{\beta,\nu}(u), \tag{6.4}$$

noting that the proof of Proposition 2.2 of [11] passes through even though our ν is not compactly supported. Let

$$d := \inf_{u \in \mathcal{O}_{\beta,\nu}} H_{\beta,\nu}(u). \tag{6.5}$$

We note that $\operatorname{supp} \nu = (-\infty, -1]$ and

$$H'_{\beta,\nu}(u) = 1 - \beta^2 \mathsf{E}\left(\frac{1}{(u + \frac{1}{1-\tanh^2(h+\beta\sqrt{q}Z)})^2}\right), \tag{6.6}$$

For $u = 0$, we evaluate

$$H_{\beta,\nu}(0) = \beta^2 \mathsf{E}\left(1 - \tanh^2(h + \beta\sqrt{q}Z\right) = \beta^2(1-q). \tag{6.7}$$

From (6.6) and as $1 - \tanh^2(x) = \cosh^{-2}(x)$, we can rewrite

$$H'_{\beta,\nu}(0) = 1 - \beta^2 \mathsf{E} \cosh^{-4}(h + \beta\sqrt{q}Z). \tag{6.8}$$

Hence, the AT condition (1.7) is equivalent to $H'_{\beta,\nu}(0) \geq 0$, and that (1.7) with strict inequality is equivalent to $H'_{\beta,\nu}(0) > 0$. Moreover, (6.6) shows that $H'_{\beta,\nu}(u)$ is strictly increasing in $u \in (-1, \infty)$. From (4.6) and as $\operatorname{supp} \nu = (-\infty, -1]$, we obtain that $H_{\beta,\nu}$ exists and is analytic in $(-1, \infty)$.

We first consider (β, h) that satisfy (1.7) with strict inequality. Then from $H'_{\beta,\nu}(0) > 0$, we infer that $H_{\beta,\nu}$ attains its infimum over $\mathcal{O}_{\beta,\nu}$ at some $u_* < 0$, and $d = H_{\beta,\nu}(u_*) < H_{\beta,\nu}(0) = \beta^2(1-q)$.

Next, we consider the case hat (β, h) does not satisfy (1.7). Then from $H'_{\beta,\nu}(0) < 0$, we infer that $H_{\beta,\nu}$ attains its infimum over $\mathcal{O}_{\beta,\nu}$ at some $u_* > 0$, that $H_{\beta,\nu}$ is decreasing in $(0, u_*)$, and hence $d = H_{\beta,\nu}(u_*) < H_{\beta,\nu}(0) = \beta^2(1-q)$.

For (β, h) satisfying (1.7) with equality, we have $H'_{\beta,\nu}(0) = 0$, and $H_{\beta,\nu}$ attains its infimum over $\mathcal{O}_{\beta,\nu}$ at 0, which implies $d = H_{\beta,\nu}(0) = \beta^2(1-q)$. □

7. Proof for largest eigenvalue at Bolthausen approximation

From (5.14) we recall that under $\mathbb{P}(\cdot \mid \mathcal{G}_{k-1})$, we have

$$\boldsymbol{H}^{(k)} \approx \frac{\beta}{\sqrt{N}} \boldsymbol{X} + \boldsymbol{A}^{(k)} + \boldsymbol{B}^{(k)} - \beta^2(1-q)\mathbf{1} \tag{7.1}$$

where $\boldsymbol{X}$ is a GOE independent of $\boldsymbol{A}^{(k)} + \boldsymbol{B}^{(k)}$.

In the next two lemmas, we give lower bounds for the largest eigenvalue of $\boldsymbol{A}^{(k)}$ and $\boldsymbol{B}^{(k)}$.

Lemma 7.1. *For $\beta > 0$, $h \neq 0$ satisfying* (1.7)*, we have*

$$\|\boldsymbol{m}^{(k)}\|_2^{-2} \boldsymbol{m}^{(k)T} \boldsymbol{B}^{(k)} \boldsymbol{m}^{(k)} \underset{\mathbb{P}}{\overset{N\to\infty}{\longrightarrow}} 2\beta^2 q + 2q^{-1}\beta^2(1-q)\Gamma_{k-2}^2 + 2q^{-1}\beta^2(1-q)\sqrt{q - \Gamma_{k-2}^2} \overset{k\to\infty}{\longrightarrow} 2\beta^2, \tag{7.2}$$

where the first limit is in probability.

Proof. The limit in probability on the left-hand side in (7.2) is equal to the limit of

$$2\beta^2\|\boldsymbol{m}^{(k)}\|^{-2}\frac{1}{N^2}\sum_{i,j=1}^{N} m_i^{(k)^2} m_j^{(k)^2} + 2\sum_{s=1}^{k-1}\|\boldsymbol{m}^{(k)}\|^{-2}\frac{\beta}{N^2}\sum_{i,j=1}^{N} m_i^{(k)}\zeta_i^{(s)}\phi_j^{(s)} m_j^{(k)}. \tag{7.3}$$

By Lemma 2.1, the first term of (7.3) converges in probability to $2\beta^2 q$ as $N\to\infty$. Each summand with $s\le k-2$ in the last term equals $2\beta\|\boldsymbol{m}^{(k)}\|^{-2}\langle \boldsymbol{m}^{(k)},\zeta^{(s)}\rangle\langle\phi^{(s)},\boldsymbol{m}^{(k)}\rangle$, which converges in probability as $N\to\infty$ to $2q^{-1}\beta^2\gamma_s^2(1-q)$ by Proposition 6 and Lemma 16 of [8]. The summand with $s=k-1$ converges in probability to $2q^{-1}\beta^2(1-q)\sqrt{q-\Gamma_{k-2}^2}$ as $N\to\infty$ again by Proposition 6 and Lemma 16 of [8]. Taking also $k\to\infty$ and using that $\Gamma_{k-2}^2=\sum_{s=1}^{k-2}\gamma_s^2\to q$ by Lemma 2b of [8], the limit in probability of the last term of (7.3) is $2\beta^2(1-q)$. □

Remark 7.2. *We also note that the largest eigenvalue of $\boldsymbol{B}^{(k)}$ is tight,*

$$\lim_{N\to\infty}\mathbb{P}\left(\lambda_1(\boldsymbol{B}^{(k)})\le 6\beta^2\right)=1. \tag{7.4}$$

This follows as

$$\beta\sum_{s=1}^{k-1}\left(\zeta^{(s)}\otimes\phi^{(s)}+\phi^{(s)}\otimes\zeta^{(s)}\right)=\beta N^{-1/2}\bar{\boldsymbol{g}}^{(k)}-\beta N^{-1/2}\bar{\boldsymbol{g}} \tag{7.5}$$

by (2.8), and $\lambda_1(\beta N^{-1/2}\bar{\boldsymbol{g}}^{(k)})\to 2\beta$, $\lambda_1(\beta N^{-1/2}\bar{\boldsymbol{g}})\to 2\beta$ $\mathbb{P}$-a. s. as $\beta N^{-1/2}\bar{\boldsymbol{g}}$ is a GOE with variance $\beta N^{-1/2}$ under $\mathbb{P}$, and $-\beta N^{-1/2}\bar{\boldsymbol{g}}^{(k)}$ has the same distribution under the absolutely continuous measure $\mathbb{P}(\cdot|\mathcal{G}_{k-1})$. Here we also used that $\lambda_1(\boldsymbol{M}_1+\boldsymbol{M}_2)\le\lambda_1(\boldsymbol{M}_1)+\lambda_1(\boldsymbol{M}_2)$ for any real symmetric matrices $\boldsymbol{M}_1$, $\boldsymbol{M}_2$ of the same dimension. Evaluating the summands in the definition (5.13) of $\boldsymbol{B}^{(k)}$ in this way, we obtain $\lambda_1(\boldsymbol{B}^{(k)})\le 2\beta^2q+4\beta^2<6\beta^2$ with high probability, which yields the assertion.

Lemma 7.3. *For $\beta>0$, $h\neq 0$ satisfying (1.7), we have*

$$\lim_{N\to\infty}\|\boldsymbol{m}^{(k)}\|_2^{-2}\boldsymbol{m}^{(k)^T}\boldsymbol{A}^{(k)}\boldsymbol{m}^{(k)}=-\tfrac{1}{2q}\left(\cosh(2h)\exp(2q\beta^2)-1\right), \tag{7.6}$$

in probability.

Proof. By definition of $\boldsymbol{A}^{(k)}$, the l.h.s. of (7.6) equals the l.h.s. of

$$-\lim_{N\to\infty}\frac{\boldsymbol{m}^{(k)^T}}{\|\boldsymbol{m}^{(k)}\|_2}\mathbf{diag}\left(\frac{1}{1-m_i^{(k)^2}}\right)\frac{\boldsymbol{m}^{(k)}}{\|\boldsymbol{m}^{(k)}\|_2}=-\frac{1}{q}\mathsf{E}\sinh^2(h+\beta\sqrt{q}Z)\text{ in probability,} \tag{7.7}$$

where the equality follows from Lemma 2.1 and the fact that $\frac{1}{1-m_i^{(k)^2}}=\cosh({h_i}^{(k+1)})^2$. As $\sinh^2(x)=\frac{1}{2}\left(\frac{1}{2}(\mathrm{e}^{2x}+\mathrm{e}^{-2x})-1\right)$, $\cosh(2x)=\frac{1}{2}\left(\mathrm{e}^{2x}+\mathrm{e}^{-2x}\right)$ and $\mathsf{E}\mathrm{e}^{\pm 2\beta\sqrt{q}Z}=\mathrm{e}^{2\beta^2 q}$, we obtain

$$\mathsf{E}\sinh^2(h+\beta\sqrt{q}Z)=\tfrac{1}{2}\left(\cosh(2h)\exp(2q\beta^2)-1\right). \tag{7.8}$$

□

Lemma 7.4. *For each $\beta>0$, $h\neq 0$, $k\in\mathbb{N}$, we have*

$$\max_{j=2k,\dots,N}\mathrm{d}(\lambda_j(\boldsymbol{A}^{(k)}+\boldsymbol{B}^{(k)}),\mathrm{supp}\,(\nu))=0. \tag{7.9}$$

Proof. By the representation in (6.2), $\boldsymbol{B}^{(k)}$ is the sum of $2k-1$ projections to one-dimensional subspaces of $\mathbb{R}^N$, and hence the span of the eigenspaces of the non-zero eigenvalues of $\boldsymbol{B}^{(k)}$ has dimension at most $2k-1$. All eigenvalues of $\boldsymbol{A}^{(k)}$ lie in the support of ν, which is the half-open interval $(-\infty,-1]$. The Weyl inequality (see e. g. 1.54 in [30]) now yields

$$\lambda_j(\boldsymbol{A}^{(k)}+\boldsymbol{B}^{(k)}) \le \lambda_{j-2k+1}(\boldsymbol{A}^{(k)}) + \lambda_{2k}(\boldsymbol{B}^{(k)}) \le -1 \tag{7.10}$$

for all $j = 2k, 2k+1, \dots, n$. □

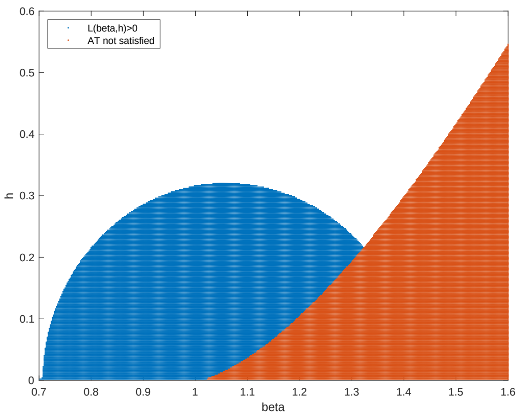

FIGURE 1. Our proof of Theorem 1.5 yields a positive eigenvalue for $\beta > 0$, $h \neq 0$ that satisfy the AT condition (1.7) and $L(\beta,h) > 0$, where $L(\beta,h) = 2\beta^2 - \frac{1}{2q}\left(\cosh(2h)\exp(2q\beta^2)-1\right)$. The diagram illustrates the sets of β, h for which $L(\beta,h) > 0$ (blue) and where AT is not satisfied (red). We expect positive eigenvalues in a larger region.

For illustration, we plotted in Figure 1 the β, h where the lower bound for $\lambda_1(\boldsymbol{A}_N^{(k)}+\boldsymbol{B}_N^{(k)})$ given by Lemmas 7.3 and 7.3 is positive. To show the existence of such β, h analytically, we also use a small-h expansion of q:

Lemma 7.5. *For $\beta < 1$, we have*

$$q = \frac{h^2}{1-\beta^2}\left(1+o(1)\right). \tag{7.11}$$

as $h \to 0$.

Proof. For $h = 0$ and $\beta < 1$, the unique solution of (1.6) is $q = 0$. By the implicit function theorem, applied to $f(q,h) = \mathsf{E}\tanh^2\left(h+\beta\sqrt{q}Z\right) - q = 0$, the solution $q(h)$ of (1.6) is continuous in 0, hence $h \to 0$ implies $q(h) \to 0$. By linearizing the tanh function in (1.6), we thus obtain, as $h \to 0$,

$$q(h) = \mathsf{E}\left(\left(h+\beta\sqrt{q(h)}Z\right)^2\right)(1+o(1)) = \left(h^2+\beta^2 q(h)\right)(1+o(1)). \tag{7.12}$$

Solving for $q(h)$ yields the assertion. □

We are now ready for the proof of the main result:

Proof of Theorem 1.5. For $\beta \in (0,1)$, Lemma 7.5 gives

$$\tfrac{1}{2q}\left(\cosh(2h)\exp(2q\beta^2)-1\right) = \frac{1-\beta^2}{2h^2}\left(2h^2 + 2\frac{h^2}{1-\beta^2}\beta^2\right)(1+o(1)) = 1 + o_h(1) \tag{7.13}$$

as $h \to 0$. We abbreviate $u_k := 2\beta^2 q + 2q^{-1}\beta^2(1-q)\Gamma_{k-2}^2 + 2q^{-1}\beta^2(1-q)\sqrt{q-\Gamma_{k-2}^2}$ for the middle term in (7.2), which converges to $2\beta^2$ as $k \to \infty$. Let $\beta \in (1/\sqrt{2}, 1)$ and $\epsilon \in (0, \beta^2 - \frac{1}{2})$. We find $h > 0$ such that the l.h.s. of (7.13) is bounded from above by $1+\epsilon$. Moreover, let k_0 be sufficiently large such that for all $k \geq k_0$, we have $u_k > 2\beta^2 - \epsilon$. Then,

$$\lim_{N\to\infty} \mathbb{P}\left(\lambda_1\left(\boldsymbol{A}^{(k)} + \boldsymbol{B}^{(k)}\right) \in \left[2\beta^2 - 1 - 2\epsilon, 6\beta^2\right]\right) = 1, \tag{7.14}$$

where we used Lemmas 7.1 and 7.3, that

$$\lambda_1\left(\boldsymbol{A}^{(k)} + \boldsymbol{B}^{(k)}\right) \geq \|\boldsymbol{m}^{(k)}\|_2^{-2} \boldsymbol{m}^{(k)T}\left(\boldsymbol{A}^{(k)} + \boldsymbol{B}^{(k)}\right)\boldsymbol{m}^{(k)}, \tag{7.15}$$

and Remark 7.2. We fix these β, h and note that they satisfy the AT condition (1.7) with strict inequality.

In the remainder of the proof, we write $\boldsymbol{A}_N^{(k)} := \boldsymbol{A}^{(k)}$, $\boldsymbol{B}_N^{(k)} := \boldsymbol{B}^{(k)}$, $\boldsymbol{X}_N = \boldsymbol{X}$ to explicitly record the N-dependence of these quantities. From the proof of Theorem 1.4, we recall that $H'_{\beta,\nu}(0) \geq 0$ under the AT condition (1.7). From (6.6), we recall that $H'_{\beta,\nu}$ is increasing on $(-1, \infty)$ as $\operatorname{supp}\nu = (-\infty, -1]$. Hence, $\theta \in \mathcal{O}_{\beta,\nu}$ for all $\theta > 0$. From (6.7) and as $H'_{\beta,\nu}$ is positive in $(0,\infty)$, it also follows that $H_{\beta,\nu}(\theta) - \beta^2(1-q) > 0$ for all $\theta > 0$.

For any fixed $k \geq k_0$ and any sequence along which N tends to infinity, there exists by (7.14) and compactness of $[2\beta^2 - 1 - 2\epsilon, 6\beta^2]$ a. s. a subsequence (N_i) tending to infinity along which $\lambda_1(\boldsymbol{A}_{N_i}^{(k)} + \boldsymbol{B}_{N_i}^{(k)})$ converges to some $\widetilde{\theta} \in [2\beta^2 - 1 - 2\epsilon, 6\beta^2]$. From the above, we also have $\theta := \inf \widetilde{\theta} > 0$ a. s., where the infimum is over all such sequences and subsequences. Noting that $\widetilde{\theta} \in \mathcal{O}_{\beta,\nu}$ and that assumption (4.8) is satisfied by Lemma 7.4, we can now apply Lemma 4.4 along the subsequence (N_i) to obtain

$$\lim_{i\to\infty} \lambda_1\left(\frac{\beta}{\sqrt{N_i}}\boldsymbol{X}_{N_i} + \boldsymbol{A}_{N_i}^{(k)} + \boldsymbol{B}_{N_i}^{(k)}\right) = H_{\beta,\nu}(\widetilde{\theta}) \quad \text{a. s.} \tag{7.16}$$

By monotonicity of $H_{\beta,\nu}$, as θ is a lower bound of $\widetilde{\theta}$ for all such subsequences (N_i), and as $k \geq k_0$ was arbitrary, it follows that

$$\lim_{k\to\infty} \liminf_{N\to\infty} \mathbb{P}\left(\lambda_1\left(\frac{\beta}{\sqrt{N}}\boldsymbol{X}_N + \boldsymbol{A}_N^{(k)} + \boldsymbol{B}_N^{(k)}\right) \geq H_{\beta,\nu}(\theta)\right) = 1. \tag{7.17}$$

As $H_{\beta,\nu}(\theta) - \beta^2(1-q) > 0$, Lemmas 5.3 and 4.2 now yield the assertion. □

8. Proof for largest eigenvalue at another magnetization

In Theorem 1.2, we rely on a specific magnetization $\boldsymbol{m}_N$ at which we evaluate the Hessian of the TAP functional: for $N \in \mathbb{N}$, let $\boldsymbol{v}$ be an eigenvector to the largest eigenvalue of $\beta N^{-1/2}\bar{\boldsymbol{g}}$ with $\|\boldsymbol{v}\|_2 = 1$, then we recall that $\beta N^{-1/2}\boldsymbol{v}^T\bar{\boldsymbol{g}}\boldsymbol{v} \to 2\beta$ a. s. For $\alpha \in [0,1]$ to be chosen later, we define the magnetization $\boldsymbol{m}_N^\alpha$ by

$$m_{N,i}^\alpha = \alpha \operatorname{sign}(v_i), \quad i = 1, \dots, N. \tag{8.1}$$

First we note that for $\beta > 0$ and $\alpha^2 > 1 - 1/\beta$,

$$\frac{\beta^2}{N}\sum_{i=1}^N \left(1 - {m_{N,i}^\alpha}^2\right)^2 = \beta^2\left(1-\alpha^2\right)^2 < 1, \tag{8.2}$$

and thus $\boldsymbol{m}_N^\alpha \in P_N^1$.

Proof of Theorem 1.2. Let $\boldsymbol{m}_N = \boldsymbol{m}_N^\alpha$ and $\boldsymbol{v}$ be defined by (8.1). As in (5.11), we evaluate $\boldsymbol{H} = \boldsymbol{H}(\boldsymbol{m}_N)$ as follows:

$$H_{ij} = \frac{\beta}{\sqrt{N}}\bar{g}_{ij} + \frac{2\beta^2\alpha^2}{N}\operatorname{sign}(v_i)\operatorname{sign}(v_j), \quad i,j = 1,\dots,N, i \neq j$$

$$H_{ii} = -\beta^2\left(1-\alpha^2\right) - \frac{1}{1-\alpha^2} + \frac{2\beta^2\alpha^2}{N}. \tag{8.3}$$

We now estimate $\boldsymbol{v}^T\boldsymbol{H}\boldsymbol{v}$ which is a lower bound for $\lambda_1(\boldsymbol{H})$. First, recall that $\boldsymbol{v}^T\frac{\beta}{\sqrt{N}}\bar{\boldsymbol{g}}\boldsymbol{v} \to 2\beta$ a.s. as $N \to \infty$. The random vector $\boldsymbol{v}$ is distributed as the first column of a Haar distributed random matrix on the orthogonal group on $\mathbb{R}^N$ (see e.g. Corollary 2.5.4 in [3]). Hence, by Lemma 8.1 below,

$$\frac{2\beta^2\alpha^2}{N}\sum_{i,j=1}^N v_i\operatorname{sign}(v_i)\operatorname{sign}(v_j)\,v_j \to \frac{4\beta^2\alpha^2}{\pi} \tag{8.4}$$

in probability as $N \to \infty$. It follows that

$$\boldsymbol{v}^T\boldsymbol{H}\boldsymbol{v} \to 2\beta - \beta^2\left(1-\alpha^2\right) - \frac{1}{1-\alpha^2} + \frac{4\beta^2\alpha^2}{\pi} \tag{8.5}$$

in probability as $N \to \infty$. For fixed $\beta > 0$, the expression on the r.h.s. attains its maximum at $\alpha^2 = 1-\beta^{-1}(1+4/\pi)^{-1/2}$ which is larger than $1-\beta^{-1}$ and hence $\boldsymbol{m}^\alpha \in P_N^1$ by (8.2). The value of the maximum of the r.h.s. of (8.5) is strictly positive for $\beta > \frac{\pi}{2}\left(\sqrt{1+4/\pi}-1\right) =: \beta_0 \approx 0.798$. □

Lemma 8.1. *Let $\boldsymbol{v}$ be distributed as the first column of a Haar distributed random matrix on the orthogonal group on $\mathbb{R}^N$. Then,*

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N |v_i| \to \sqrt{2/\pi} \tag{8.6}$$

in probability as $N \to \infty$.

Proof. First we consider the expectation

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}\,|v_i| = \mathbb{E}\sqrt{N}\,|v_1|\,, \tag{8.7}$$

which converges to $\mathbb{E}|Z| = \sqrt{2/\pi}$ by (1) of Diaconis and Freedman [14], noting that convergence in total variation implies convergence of (absolute) moments. Likewise, for the second moment, we have

$$\frac{1}{N}\sum_{i,j=1}^N \mathbb{E}\,|v_i|\,|v_j| = (N-1)\mathbb{E}\,|v_1|\,|v_2| + \mathbb{E}v_1^2. \tag{8.8}$$

Here the second term on the r.h.s. converges to zero, and the first term to $(\mathbb{E}|Z|)^2$ again by (1) of [14]. This shows that the variance of the expression on the l.h.s. of (8.6) converges to zero, so that the convergence of the expectation implies the assertion. □

Appendix A. Proof of Lemmas 3.1 and 2.1

Proof of Lemma 3.1. Let us write $X \overset{N,k}{\sim} Y$ if $\|X-Y\| \to 0$ in $L_2(\mathbb{P})$ as $N \to \infty$ followed by $k \to \infty$, and $\overset{N}{\sim}$ if the norm vanishes already as $N \to \infty$. From (2.5), we have

$$\beta^{-1}\Delta^{(k)} = \sum_{s=1}^{k-2} \gamma_s \zeta^{(s)} + \sqrt{q-\Gamma_{k-2}^2}\zeta^{(k)} - \frac{1}{\sqrt{N}}\bar{\boldsymbol{g}}\boldsymbol{m}^{(k)} + \beta(1-q)\boldsymbol{m}^{(k)}\,. \tag{A.1}$$

We note that the second term on the right-hand side of (A.1) is $\overset{N,k}{\sim} 0$ by Lemmas 2 and 15a of [8]. Thus it holds

$$\beta^{-1}\Delta^{(k)} \overset{N,k}{\sim} \sum_{s=1}^{k-2} \gamma_s \zeta^{(s)} - \frac{1}{\sqrt{N}}\bar{\boldsymbol{g}}\boldsymbol{m}^{(k)} + \beta(1-q)\boldsymbol{m}^{(k)}\,. \tag{A.2}$$

We will now rewrite the second term of (A.2). From (2.9), we obtain

$$\bar{\boldsymbol{g}} = \bar{\boldsymbol{g}}^{(k)} + \sqrt{\tfrac{N}{2}} \sum_{s=1}^{k-1} \left(\rho^{(s)} + {\rho^{(s)}}^T\right), \tag{A.3}$$

which we use together with (2.9) in

$$\tfrac{1}{\sqrt{N}}\bar{\boldsymbol{g}}\boldsymbol{m}^{(k)} = \tfrac{1}{\sqrt{N}}\bar{\boldsymbol{g}}^{(k)}\boldsymbol{m}^{(k)} + \sum_{s=1}^{k-1}\left[\zeta^{(s)}\langle\phi^{(s)},\boldsymbol{m}^{(k)}\rangle + \phi^{(s)}\langle\zeta^{(s)},\boldsymbol{m}^{(k)}\rangle - \sqrt{2}\langle\phi^{(s)},\xi^{(s)}\rangle\phi^{(s)}\langle\phi^{(s)},\boldsymbol{m}^{(k)}\rangle\right]. \tag{A.4}$$

The expression on the right-hand side of the last display is

$$\overset{N,k}{\sim} \sum_{s=1}^{k-1}\gamma_s\zeta^{(s)} + \sum_{s=1}^{k-2}\phi^{(s)}\beta\gamma_s(1-q) + \phi^{(k-1)}\beta(1-q)\sqrt{q-\Gamma_{k-2}^2} - \sqrt{2}\sum_{s=1}^{k-1}\langle\phi^{(s)},\xi^{(s)}\rangle\phi^{(s)}\gamma_s, \tag{A.5}$$

by Proposition 6, Lemmas 13 and 16 of [8]. As $\|\phi^{(s)}\| = 1$, the last term in (A.5) is $\overset{N}{\sim} 0$ by Lemma 11 of [8].

Replacing $\frac{1}{\sqrt{N}}\bar{\boldsymbol{g}}\boldsymbol{m}^{(k)}$ in (A.2) by (A.5), after cancellations we obtain

$$\beta^{-1}\Delta^{(k)} \overset{N,k}{\sim} -\gamma_{k-1}\zeta^{(k-1)} + \beta(1-q)\left(\boldsymbol{m}^{(k)} - \sum_{s=1}^{k-2}\phi^{(s)}\gamma_s - \phi^{(k-1)}\sqrt{q-\Gamma_{k-2}^2}\right). \tag{A.6}$$

By Lemmas 2 and 15a of [8], the first term on the r.h.s. of (A.6) vanishes and we obtain

$$\beta^{-1}\Delta^{(k)} \overset{N,k}{\sim} \beta(1-q)\left(\boldsymbol{m}^{(k)} - \sum_{s=1}^{k-2}\phi^{(s)}\gamma_s - \phi^{(k-1)}\sqrt{q-\Gamma_{k-2}^2}\right), \tag{A.7}$$

where the $\|\cdot\|$ norm of the r.h.s. is bounded by

$$\beta(1-q)\left(\left\|\boldsymbol{m}^{(k)} - \sum_{s=1}^{k-2}\phi^{(s)}\gamma_s\right\| + \sqrt{q-\Gamma_{k-2}^2}\left\|\phi^{(k-1)}\right\|\right). \tag{A.8}$$

By (2.10), we have that $\lim_{k\to\infty}\sqrt{q-\Gamma_{k-2}^2}=0$, recalling that $\|\phi^{(k-1)}\|=1$, the last term in the brackets on the r.h.s. of (A.8) vanishes. As for the first term in the brackets, using the fact that $\{\phi^{(s)}\}$ is an orthonormal basis, it holds

$$\left\|\boldsymbol{m}^{(k)}-\sum_{s=1}^{k-2}\phi^{(s)}\gamma_s\right\|^2=\left\|\boldsymbol{m}^{(k)}\right\|^2+\sum_{s=1}^{k-2}\gamma_s^2-2\sum_{s=1}^{k-2}\gamma_s\langle\boldsymbol{m}^{(k)},\boldsymbol{\phi}^{(s)}\rangle, \tag{A.9}$$

By Proposition 6 of [8] together with (2.10) implies that the $\lim_{k\to\infty}\lim_{N\to\infty}$ of r.h.s. of the latter is equal to 0. □

Proof of Lemma 2.1. Assertion (2.17) is a slight modification of Lemma 14 in [8]. Indeed, Lemma 14 in [8] is for a function $f:\mathbb{R}\to\mathbb{R}$, which is Lipschitz continuous with $|f(x)|\le C(1+|x|)$ for some $C<\infty$. Thus we have to change a little bit this result to prove Lemma 2.1. For any $M>0$, we can use a truncation to split the term as follows,

$$\frac{1}{N}\sum_{i=1}^N f\left(h_i^{(k)}\right)=\frac{1}{N}\sum_{i=1}^N f\left(h_i^{(k)}\right)\chi\left(h_i^{(k)}\right)+\frac{1}{N}\sum_{i=1}^N f\left(h_i^{(k)}\right)\left(1-\chi\left(h_i^{(k)}\right)\right), \tag{A.10}$$

where $\chi:\mathbb{R}\to[0,1]$ is a mollified indicator function of $[-M,M]$, that is, a continuously differentiable function with $\chi(x)=1$ for $x\in[-M+1,M-1]$, $\chi(x)=0$ for $|x|\ge M$, and χ monotone in $[M-1,M]$ and in $[-M-1,-M]$. We can now use the law of large number of Lemma 14 in [8] for the first term on the r.h.s. of (A.10) to wit

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N f\left(h_i^{(k)}\right)\chi\left(h_i^{(k)}\right)=\mathsf{E}\left(f(h+\beta\sqrt{q}Z)\chi\left(h+\beta\sqrt{q}Z\right)\right)\text{ in }L_1(\mathbb{P}). \tag{A.11}$$

As for the second term on the r.h.s. of (A.10), by using Cauchy–Schwarz two times, we obtain

$$\begin{aligned}\mathbb{E}\left(\left|\frac{1}{N}\sum_{i=1}^N f\left(h_i^{(k)}\right)\left(1-\chi\left(h_i^{(k)}\right)\right)\right|\right)&\le\frac{1}{N}\sum_{i=1}^N\mathbb{E}\left(\left|f\left(h_i^{(k)}\right)\right|\left(1-\chi\left(h_i^{(k)}\right)\right)\right)\\&\le\frac{1}{N}\sum_{i=1}^N\sqrt{\mathbb{E}\left(f\left(h_i^{(k)}\right)^2\right)}\sqrt{\mathbb{E}\left(1-\chi\left(h_i^{(k)}\right)\right)}\\&\le\sqrt{\frac{1}{N}\sum_{i=1}^N\mathbb{E}\left(f\left(h_i^{(k)}\right)^2\right)}\sqrt{\frac{1}{N}\sum_{i=1}^N\mathbb{E}\left(1-\chi\left(h_i^{(k)}\right)\right)}.\end{aligned} \tag{A.12}$$

For the second term on the r.h.s. of (A.12), Lemma 14 in [8] gives

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\mathbb{E}\left(1-\chi\left(h_i^{(k)}\right)\right)=\mathsf{E}\left(1-\chi\left(h+\beta\sqrt{q}Z\right)\right). \tag{A.13}$$

Moreover, we claim that there exists L_k independent of N such that

$$\frac{1}{N}\sum_{i=1}^N\mathbb{E}\left(f\left(h_i^{(k)}\right)^2\right)\le L_k. \tag{A.14}$$

Before proving this claim, we combine (A.13) and (A.14), which yields

$$\mathbb{E}\left(\left|\frac{1}{N}\sum_{i=1}^{N} f\left(h_i^{(k)}\right)\left(1-\chi\left(h_i^{(k)}\right)\right)\right|\right) \leq \sqrt{L_k \mathsf{E}\left(1-\chi\left(h+\beta\sqrt{q}Z\right)\right)(1+o_N(1))}, \quad (A.15)$$

where $o_N(1) \to 0$ in $L_1(\mathbb{P})$ as $N \to \infty$. As $N \to \infty$ and then $M \to \infty$, the second term on the r.h.s. of (A.10) vanishes a.s. and assertion (2.17) follows by (A.11) and the theorem of dominated convergence.

It remains to prove (A.14). By the growth assumption on f, there exists $C < \infty$, such that

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(f\left(h_i^{(k)}\right)^2\right) \leq \frac{C^2}{N}\sum_{i=1}^{N}\mathbb{E}\left(e^{2C|h_i^{(k)}|}\right) \leq 2\frac{C^2}{N}\sum_{i=1}^{N}\mathbb{E}\left(\cosh\left(2Ch_i^{(k)}\right)\right). \quad (A.16)$$

Let $r \in \{-2C, 2C\}$ be a constant, to prove (A.14), we only have to bound the following term in N:

$$\begin{aligned}\mathbb{E}\left(\exp\left(rh_i^{(k)}\right)\right) &= e^{rh}\mathbb{E}\left(\exp r\left(\beta\sum_{s=1}^{k-2}\gamma_s\zeta_i^{(s)} + \beta\sqrt{q-\Gamma_{k-2}^2}\zeta_i^{(k-1)}\right)\right) \\ &= e^{rh}\mathbb{E}\left(\mathbb{E}\left(\exp r\left(\beta\sum_{s=1}^{k-2}\gamma_s\zeta_i^{(s)} + \beta\sqrt{q-\Gamma_{k-2}^2}\zeta_i^{(k-1)}\right)\right)|\mathcal{G}_{k-2}\right) \\ &= e^{rh}\mathbb{E}\left(\exp\left(r\beta\sum_{s=1}^{k-2}\gamma_s\zeta_i^{(s)}\right)\mathbb{E}\left(\exp\left(r\beta\sqrt{q-\Gamma_{k-2}^2}\zeta_i^{(k-1)}\right)|\mathcal{G}_{k-2}\right)\right), \quad (A.17)\end{aligned}$$

where we used (2.5) and in the last line that $\zeta_i^{(s)}$ is $\mathcal{G}_s$-measurable (for $s = 1, \ldots, k-1$). By Proposition 4 of [8], we have that the variables $\zeta_i^{(k-1)}$ are conditionally Gaussian, given $\mathcal{G}_{k-2}$, with

$$\mathbb{E}\left(\zeta_i^{(k-1)^2}|\mathcal{G}_{k-2}\right) = 1 + \frac{1}{N}\phi_i^{(k-1)^2} - \sum_{s=1}^{k-2}\frac{1}{N}\phi_i^{(s)^2} \leq 1 + \frac{1}{N}\sum_{i=1}^{N}\phi_i^{(k-1)^2} = 2, \quad (A.18)$$

and by Jensen inequality, we deduce that

$$\left|\mathbb{E}\left(\zeta_i^{(k-1)}|\mathcal{G}_{k-2}\right)\right| \leq 2. \quad (A.19)$$

Combining (A.18) and (A.19) yields

$$\mathbb{E}\left(\exp\left(rh_i^{(k)}\right)\right) \leq e^{rh}\mathbb{E}\left(\exp\left(r\beta\sum_{s=1}^{k-2}\gamma_s\zeta_i^{(s)}\right)\right)\exp\left(2r\beta\sqrt{q-\Gamma_{k-2}^2} + r^2\beta^2(q-\Gamma_{k-2}^2)\right). \quad (A.20)$$

By repeating the previous procedure, by induction, we see that

$$\mathbb{E}\left(\exp\left(rh_i^{(k)}\right)\right) \leq \exp\left(rh + 2r\beta\sum_{s=1}^{k-2}\gamma_s + r^2\beta^2\sum_{s=1}^{k-2}\gamma_s^2 + 2r\beta\sqrt{q-\Gamma_{k-2}^2} + r^2\beta^2(q-\Gamma_{k-2}^2)\right). \tag{A.21}$$

This concludes the proof of the claim (A.14) and thus (2.17) follows.
We note that (2.17) also hods for complex-valued f and under Skorohod coupling, the convergence also holds a. s. From this, assertion (2.18) can be deduced by setting $f(x) = \mathrm{e}^{i\lambda x}$ for any $\lambda \in \mathbb{Q}$ and using standard results on the convergence of random measures. □

Appendix B. Proof of Lemma 4.4

We abbreviate $\boldsymbol{M}_N := \sigma N^{-1/2}\boldsymbol{X}_N + \boldsymbol{A}_N$. Consider an orthogonal diagonalization $\boldsymbol{A}_N = \boldsymbol{O}_N^T\boldsymbol{D}_N\boldsymbol{O}_N$ of $\boldsymbol{A}_N$. As $\boldsymbol{O}_N\boldsymbol{X}_N\boldsymbol{O}_N^T$ is again distributed as a GOE, and as $\boldsymbol{M}_N$ has the same eigenvalues as $\boldsymbol{O}^T\boldsymbol{M}_N\boldsymbol{O}_N$, we henceforth assume w. l. o. g. that $\boldsymbol{A}_N$ is diagonal.

We now infer the assertion from Theorem 8.1 1) of [11] in the case that ν has compact support. First, the proof of Theorem 8.1 1) of [11] passes through for GOE (in place of GUE) when Theorem 5.1 of [11] is replaced with Theorem 4.2 of [15]. We write $\gamma_1 := \theta$. We assume w. l. o. g. that r is the minimal integer satisfying assumption (4.8). For any subsequence of N tending to infinity, we find a subsubsequence (N_i) tending to infinity along which $\lambda_j(\boldsymbol{A}_{N_i})$ converges to some $\gamma_j \in (d, \theta]$ for all $j = 2, \dots, r$, using compactness of the interval $[d, \theta]$ and minimality of r. Hence, there exists a diagonal matrix $\widetilde{\boldsymbol{A}}_{N_i}$ with eigenvalues $\lambda_j(\widetilde{\boldsymbol{A}}_{N_i}) = \gamma_j$ whose difference to $\boldsymbol{A}_{N_i}$ vanishes in the Frobenius norm

$$\|\widetilde{\boldsymbol{A}}_{N_i} - \boldsymbol{A}_{N_i}\|_{\mathrm{F}} \overset{i\to\infty}{\longrightarrow} 0. \tag{B.1}$$

From (B.1), it follows that $\boldsymbol{A}_{N_i}$ can be replaced with $\widetilde{\boldsymbol{A}}_{N_i}$ in the definition of $\boldsymbol{M}_{N_i}$ without changing the limiting largest eigenvalue by Lemma 4.2. Then $\boldsymbol{A}_{N_i}$ satisfies the assumptions of Theorem 8.1 of [11], which yields

$$\lim_{i\to\infty}\lambda_1(\boldsymbol{M}_{N_i}) = H_{\beta,\nu}(\theta) \quad \text{a. s.} \tag{B.2}$$

As the limit in (B.2) does not depend on the choice of the subsequence of N, it also holds for the original sequence along which $N \to \infty$.

It remains to consider the case of the more general ν in the assertion. For this, we use truncation arguments for matching upper and lower bounds.

Lower bound. For $m \in \mathbb{R}_+$, let

$$V_m := \{i = 1, \dots, N : \ A_{ii} \geq -m\} \tag{B.3}$$

be the subset of the coordinates in which the corresponding diagonal elements of $\boldsymbol{A}_N$ have a value at least $-m$. The number of those coordinates will be denoted by $N_m = \#V_m$, and we set $r_{m,N} := \sqrt{N_m/N}$. Now,

$$\begin{aligned}\lambda_1(\boldsymbol{M}_N) &= \sup\left\{\boldsymbol{v}^T\boldsymbol{M}_N\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^N, \|\boldsymbol{v}\|_2 = 1\right\}\\ &\geq \sup\left\{\boldsymbol{v}^T\boldsymbol{M}_N\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^N, \|\boldsymbol{v}\|_2 = 1, \max_{i\notin V_m}|v_i| = 0\right\}\\ &= r_{m,N}\sup\left\{\boldsymbol{v}^T\boldsymbol{M}_N^{(m)}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^{N_m}, \|\boldsymbol{v}\|_2 = 1\right\} = r_{m,N}\lambda_1\left(\boldsymbol{M}_N^{(m)}\right)\end{aligned} \tag{B.4}$$

where

$$\boldsymbol{M}_N^{(m)} = \sigma N_m^{-1/2}\boldsymbol{X}_N^{(m)} + \boldsymbol{A}_N^{(m)}, \quad \boldsymbol{X}_N^{(m)} = (X_{f(i),f(j)})_{i,j\leq N_m}, \quad \boldsymbol{A}_N^{(m)} = {r_{m,N}}^{-1}(A_{f(i),f(j)})_{i,j\leq N_m} \tag{B.5}$$

and $f(i)$ denotes the i-th largest integer in V_m. Note that $\boldsymbol{X}_N^{(m)}$ is again a GOE of size N_m, that $\lim_{N\to\infty} r_{m,N}^2 = \nu([-m,d]) =: r_m^2$ for all but countably many m, and that $\mu_{\boldsymbol{A}_N^{(m)}}$ weakly converges to the probability measure ν_m as $N \to \infty$, where ν_m is defined as the image measure of $\nu(\cdot \cap [-m,d])/\nu([-m,d])$ under the dilation $t \mapsto r_m^{-1}t$. Also $r_m \to 1$ as $m \to \infty$ by definition of r_m. For $m \geq -2\theta$ and all N, we have $r_{m,N}\lambda_1(\boldsymbol{A}_N^{(m)}) = \theta$, and hence $\lambda_1(\boldsymbol{A}_N^{(m)}) \to r_m^{-1}\theta$ as $N \to \infty$. Moreover, we note that $\theta \notin \operatorname{supp}\nu_m$ for sufficiently large m, and from (4.6), we obtain $\lim_{m\to\infty} H_{\sigma,\nu_m}(\theta) = H_{\sigma,\nu}(\theta)$. By differentiating (4.6) and using the definition (4.3) of the Stieltjes transform, we also obtain

$$H'_{\beta,\nu_m}(\theta) = 1 - \beta^2\int\frac{\nu_m(\mathrm{d}x)}{(\theta-x)^2} = 1 - \frac{\beta^2}{\nu[-m,d]}\int_{[-m,d]}\frac{\nu(\mathrm{d}x)}{(\theta - r_m^{-1}x)^2}, \tag{B.6}$$

which converges to $H'_{\beta,\nu}(\theta)$ as $m \to \infty$. Hence, $H'_{\sigma,\nu_m}(\theta) > 0$ for sufficiently large m. As ν_m is compactly supported, the first part of the proof yields $\lim_{N\to\infty}\lambda_1(\boldsymbol{M}_N^{(m)}) = r_m H_{\sigma,\nu_m}(\theta)$ a. s. Using (B.4) and taking $m \to \infty$ yields $\liminf_{N\to\infty}\lambda_1(\boldsymbol{M}_N) \geq H_{\sigma,\nu}(\theta)$ almost surely.

Upper bound. We use the truncation $\widehat{\boldsymbol{A}}_N^{(m)} := \operatorname{diag}(A_{ii} \vee (-m))_{i=1,\dots,N}$, and we set $\widehat{\boldsymbol{M}}_N^{(m)} := \sigma N^{-1/2}\boldsymbol{X}_N + \widehat{\boldsymbol{A}}_N^{(m)}$. In place of (B.4), we then have

$$\lambda_1(\boldsymbol{M}_N) = \sup\left\{\boldsymbol{v}^T\boldsymbol{M}_N\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^N, \|\boldsymbol{v}\|_2 = 1\right\} \leq \lambda_1(\widehat{\boldsymbol{M}}_N^{(m)}) \tag{B.7}$$

as

$$\boldsymbol{v}^T\boldsymbol{A}_N\boldsymbol{v} = \sum_{i=1}^N v_i^2 A_{ii} \leq \sum_{i=1}^N v_i^2\widehat{A}_{ii}^{(m)} = \boldsymbol{v}^T\widehat{\boldsymbol{A}}_N^{(m)}\boldsymbol{v}. \tag{B.8}$$

The empirical spectral distribution $\mu_{\widehat{\boldsymbol{A}}_N^{(m)}}$ weakly converges to $\nu(\cdot\cap[-m,d]) + \nu(-\infty,-m)\delta_{-m}$, and we conclude in the same way as for the lower bound.

Appendix C. Plefka's expansion

We discuss here the relation between the TAP free energy and the free energy: for finite N, the TAP free energy can be interpreted in terms of an expansion of the Gibbs potential of the SK model [25]. In the following, we give an introduction of the approach based on [19]: For $\alpha \in \mathbb{R}$, and $\boldsymbol{\varphi} = \{\varphi_i\}_{i=1,\dots,N} \in \mathbb{R}^N$, we define the Hamiltonian

$$H_{\alpha,\beta,h,\boldsymbol{\varphi}}(\sigma) = \alpha H_{\beta,0}(\sigma) + h\sum_{i\leq N}\sigma_i + \sum_{i\leq N}\varphi_i\sigma_i, \tag{C.1}$$

the partition function

$$Z_{\alpha,\beta,h,\boldsymbol{\varphi}} = 2^{-N} \sum_{\sigma \in \Sigma_N} \exp H_{\alpha,\beta,h,\boldsymbol{\varphi}} \tag{C.2}$$

and the (normalized) functional $G_N(\alpha, \boldsymbol{\varphi})$ by

$$G_N(\alpha, \boldsymbol{\varphi}) := \log Z_{\alpha,\beta,h,\boldsymbol{\varphi}}. \tag{C.3}$$

Note that by Jensen's inequality, the map $\boldsymbol{\varphi} \mapsto G_N(\alpha, \boldsymbol{\varphi})$ is, in fact, *convex*. In particular, the Legendre transform is well defined:

$$G_N^\star(\alpha, \boldsymbol{m}) = \sup_{\boldsymbol{\varphi} \in \mathbb{R}^N} \left\{ \sum_{i \leq N} \varphi_i m_i - G_N(\alpha, \boldsymbol{\varphi}) \right\} . \tag{C.4}$$

Again by convexity, the *-operation is an *involution*, i.e. with the property that $G_N = (G_N^\star)^\star$. Since by construction $N^{-1} G_N(1, \mathbf{0})$ coincides with the free energy, we therefore have that

$$F_N(\beta, h) = \frac{1}{N} \sup_{\boldsymbol{m} \in [-1,1]^N} \left\{ -G_N^\star(1, \boldsymbol{m}) \right\} . \tag{C.5}$$

Here the supremum can be taken over $\boldsymbol{m} \in [-1,1]^N$ as it is readily checked that $G_N^\star(1, \boldsymbol{m}) < \infty$ only for these $\boldsymbol{m}$. The thermodynamic variables $\boldsymbol{m} \in \mathbb{R}^N$ are dual to the magnetic fields $\boldsymbol{\varphi}$, and correspond, after closer inspection, to the *magnetization*: indeed, given a function f, we denote by

$$\langle f \rangle_\alpha := \frac{2^{-N} \sum_\sigma f(\sigma) \exp\left(H_{\alpha,\beta,h,\boldsymbol{\varphi}}(\sigma) \right)}{Z_{\alpha,\beta,h,\boldsymbol{\varphi}}}, \tag{C.6}$$

the Gibbs expectation with respect to the Hamiltonian appearing in (C.1), one immediately checks by solving the variational principle (C.4) that the fundamental relation

$$\langle \sigma_i \rangle_\alpha = m_i \tag{C.7}$$

holds. In particular, we see from the above that $m_i \in [-1, 1]$. The idea is to now proceed by Taylor expansion of the Gibbs potential,

$$- G_N^\star(\alpha, \boldsymbol{m}) = \sum_{k=0}^{\infty} \frac{\partial^k}{\partial \alpha^k} \Big(- G_N^\star(\alpha, \boldsymbol{m}) \Big) \Big|_{\alpha=0} \frac{\alpha^k}{k!} , \tag{C.8}$$

and to evaluate this in $\alpha = 1$. The calculation of the Taylor-coefficients considerably simplifies in $\alpha = 0$, as one only needs to compute "spin-correlations" under the non-interacting Hamiltonian $\sum_{i \leq N} \varphi_i \sigma_i$. First note that for $\alpha = 0$, the variational principle (C.4) is solved by $\boldsymbol{\varphi}^*$ such that

$$m_i = \langle \sigma_i \rangle_0 = \tanh\left(h + \varphi_i^* \right) . \tag{C.9}$$

One immediately checks that the 0^{th}-term of the expansion is given by

$$\begin{aligned} -G_N^\star(0,\boldsymbol{m}) &= -\sum_{i\leq N}\left(\varphi_i^* m_i - \log\cosh\left(h+\varphi_i^*\right)\right) \\ &= -\sum_{i\leq N}\tanh^{-1}(m_i)m_i + h\sum_{i\leq N} m_i + \sum_{i\leq N}\log\cosh\left(\tanh^{-1}(m_i)\right) \\ &= h\sum_{i\leq N} m_i - \sum_{i\leq N} I(m_i), \end{aligned} \tag{C.10}$$

where we used (C.9) for the second line and (1.9) for the third line.

For the first derivative in $\alpha = 0$, we have

$$\begin{aligned} -\frac{\partial}{\partial\alpha}G_N^\star(\alpha,\boldsymbol{m})\Big|_{\alpha=0} &= \left\langle \frac{\beta}{\sqrt{2N}}\sum_{i,j\leq N} g_{ij}\sigma_i\sigma_j\right\rangle_0 \\ &= \frac{\beta}{\sqrt{2N}}\sum_{i\neq j\leq N} g_{ij}\langle\sigma_i\rangle_0\langle\sigma_j\rangle_0 + \frac{\sqrt{2}\beta}{\sqrt{N}}\sum_{i\leq N} g_{ii}\left\langle\sigma_i^2\right\rangle_0 \\ &= \frac{\beta}{\sqrt{N}}\sum_{i<j\leq N}\bar{g}_{ij}m_i m_j + N\times o_N(1), \end{aligned} \tag{C.11}$$

where we used the fact that σ_i and σ_j are independent under the Gibbs measure for $\alpha = 0$. The second order term in (C.8) is left to the reader but one can check that

$$-\frac{\partial^2}{\partial\alpha^2}G_N^\star(0,m) = \frac{\beta^2}{N}\sum_{i<j}^N \bar{g}_{ij}^2(1-m_i^2)(1-m_j^2) + N\times o_N(1), \tag{C.12}$$

with $o_N(1)$ uniform in N. This computation is done in [25] by Plefka. All in all, we obtain

$$F_N(\beta,h) = \sup_{\boldsymbol{m}\in[-1,1]^N}\left\{\frac{1}{N}\mathrm{TAP}_N^*(\boldsymbol{m}) + o_N(1) + \frac{1}{N}\sum_{k=3}^\infty \frac{\partial^k}{\partial\alpha^k}\Big(-G_N(\alpha,\boldsymbol{m})^\star\Big)\Big|_{\alpha=0}\frac{1}{k!}\right\}, \tag{C.13}$$

with

$$\mathrm{TAP}_N^*(\boldsymbol{m}) = \frac{\beta}{\sqrt{N}}\sum_{i<j\leq N}\bar{g}_{ij}m_i m_j + h\sum_{i=1}^N m_i + \frac{\beta^2}{2N}\sum_{i<j}^N \bar{g}_{ij}^2(1-m_i^2)(1-m_j^2) - \sum_{i=1}^N I(m_i). \tag{C.14}$$

By replacing $\bar{g}_{ij}^2$ by one in (C.14), like Plefka does in [25], we obtain $\mathrm{TAP}_N(\boldsymbol{m})$. The problem of this approach is to justify when the $\frac{1}{N}\sum_{k=3}^\infty \frac{d^k}{d\alpha^k}\Big(-G_N^\star(\alpha,m)\Big)\Big|_{\alpha=0}\frac{1}{k!}$ term is negligible.

References

[1] A. Adhikari, C. Brennecke, P.v. Soosten and H.-T. Yau (2021). Dynamical Approach to the TAP Equations for the Sherrington-Kirkpatrick Model. *J. Stat. Phys.* **183** 35.

[2] J.R.L. de Almeida and D.J. Thouless (1978). Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *J. Phys. A: Math. Gen.* **11** 983.

[3] G. Anderson, A. Guionnet and O. Zeitouni (2009). *An introduction to random matrices.* Cambridge University Press.
[4] D. Belius (2022). High temperature TAP upper bound for the free energy of mean field spin glasses. arXiv:2204.00681.
[5] D. Belius and N. Kistler (2019). The TAP-Plefka variational principle for the spherical SK model. *Comm. Math. Phys.* **367**, 991-1017.
[6] P. Biane (1997). On the free convolution with a Semi-circular Distribution. *Indiana Univ. Math. J.* **46**, no. 3, 705 - 718.
[7] E. Bolthausen (2014). An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Commun. Math. Phys.* **325** 333–366.
[8] E. Bolthausen (2019). A Morita type proof of the replica-symmetric formula for SK. In: *Statistical Mechanics of Classical and Disordered Systems.* Springer.
[9] A.J. Bray and M. Moore (1979). Evidence for massless modes in the 'solvable model' of a spin glass. *J. Phys. C: Solid State Phys.* **12** L441.
[10] C. Brennecke and H.-T. Yau (2022). The replica symmetric formula for the SK model revisited. *J. Math. Phys.* **63**, no. 073302, 12 pp.
[11] M. Capitaine, C. Donati-Martin, D. Féral and M. Février (2011). Free convolution with a semicircular distribution and eigenvalues of spiked deformations of Wigner matrices. *Electron. J. Probab.* **16**, no. 64, 1750–1792.
[12] W.-K. Chen (2021). On the Almeida-Thouless transition line in the Sherrington-Kirkpatrick model with centered Gaussian external field. *Electron. Commun. Probab.* **26**, no. 65, 1–9.
[13] W.-K. Chen and D. Panchenko (2018). On the TAP free energy in the mixed p-spin models. *Commun. Math. Phys.* **362**, 219–252.
[14] P. Diaconis and D. Freedman (1987). A dozen de Finetti-style results in search of a theory. *nn. Inst. H. Poincaré Probab. Statist.* **23**, no. 2, suppl., 397–423.
[15] Z. Fan, Y. Sun, and Z. Wang (2021). Principal components in linear mixed models with general bulk. *Ann. Statist.* **49**, no. 3, pp. 1489–1513.
[16] F. Guerra (2002). Broken replica symmetry bounds in the mean field spin glass model. *Commun. Math. Phys.* **233**, no. 1.
[17] F. Guerra and F.L. Toninelli (2002). The thermodynamic limit in mean field spin glass models. *Commun. Math. Phys.* **230**, 71-79.
[18] S. Gufler, J.L. Igelbrink and N. Kistler (2021). TAP equations are repulsive. arXiv:2111.02134.
[19] G. Kersting, N. Kistler, A. Schertzer and M.A. Schmidt (2019). Gibbs Potentials and high temperature expansions for mean field models. In: *Statistical Mechanics of Classical and Disordered Systems*, pp 193–214, Springer.
[20] A. Lejay and L. Pastur (2003). Matrices aléatoires: statistique asymptotique des valeurs propres. In: *Séminaire de Probabilités*, XXXVI, volume 1801 of Lecture Notes in Math.: 135–164. Springer.
[21] A. Nica and R. Speicher (2006). *Lectures on the combinatorics of free probability.* Cambridge University Press.
[22] J.C. Owen (1982). Convergence of sub-extensive terms for long-range Ising spin glasses. *J. Phys. C: Solid State Phys.* **15**, L1071-L1075.
[23] D. Panchenko (2013). *The Sherrington-Kirkpatrick model.* Springer.
[24] L.A. Pastur (1972). On the spectrum of random matrices. *Teor. Math. Phys.* **10**, pp. 67–74. Berlin, 2003. MR1971583.
[25] T. Plefka (1982). Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *J. Phys. A: Math. Gen.* **15**, 1971-1978.
[26] D. Sherrington and S. Kirkpatrick (1975). Solvable model of a spin-glass. *Phys. Rev. Lett.* **35**, 1792.
[27] H.-J. Sommers (1978). Solution of the Long-Range Gaussian-Random Ising Model. *Z. Phys. B* **31**, 301-307.
[28] M. Talagrand (2011). *Mean field models for spin glasses.* Springer.
[29] M. Talagrand (2006). The Parisi formula. *Ann. of Math.* (2) **163**, no. 1: 221-263.
[30] T. Tao (2012). *Topics in random matrix theory.* AMS.

[31] T. Tao and V. Vu (2010). Random Matrices: Universality of Local Eigenvalue Statistics up to the Edge. *Commun. Math. Phys.* **298**, 549–572.

[32] D.J. Thouless, P.W. Anderson and R.G. Palmer (1977). Solution of 'solvable model of a spin glass'. *Phil. Mag.* **35**, 593-601.

[33] F.L. Toninelli (2002). About the Almeida-Thouless transition line in the Sherrington-Kirkpatrick mean-field spin glass model. *Europhys. Lett.* **60**, pp. 764–767.

[34] D.V. Voiculescu, K.J. Dykema, and A. Nica (1992). *Free random variables.* CRM Monograph Series, vol. 1, AMS.

S. Gufler, J.W. Goethe-Universität Frankfurt, Germany.
Email address: gufler@math.uni-frankfurt.de

A. Schertzer, Institut für Angewandte Mathematik, Bonn University, Germany
Email address: aschertz@uni-bonn.de

M. A. Schmidt, J.W. Goethe-Universität Frankfurt, Germany.
Email address: m.schmidt@mathematik.uni-frankfurt.de