ON THE REM APPROXIMATION OF TAP FREE ENERGIES.

NICOLA KISTLER, MARIUS A. SCHMIDT, AND GIULIA SEBASTIANI

Abstract. The free energy of TAP-solutions for the SK-model of mean field spin glasses can be expressed as a nonlinear functional of local terms: we exploit this feature in order to contrive abstract REM-like models which we then solve by a classical large deviations treatment. This allows to identify the origin of the physically unsettling quadratic (in the inverse of temperature) correction to the Parisi free energy for the SK-model, and formalizes the *true* cavity dynamics which acts on TAP-space, i.e. on the space of TAP-solutions. From a non-spin glass point of view, this work is the first in a series of refinements which addresses the stability of hierarchical structures in models of evolving populations.

1. Introduction

The Generalized Random Energy Models, GREM for short, are toy models for mean field spin glasses introduced by Derrida in the 1980's [11], which have played a key role in our understanding of certain aspects of the Parisi theory [18]. Notwithstanding, the deeper relation between the GREMs and more realistic spin glasses such as the prototypical Sherrington-Kirkpatrick model [22], SK for short, hasn't yet been identified: the goal of this paper is to fill this gap.

Precisely, we relate the simplest of Derrida's models, the REM [10], and the Thouless-Anderson-Palmer free energies [25], TAP for short; this seamlessly leads to abstract, and what is crucial: *highly nonlinear*, REM-like Hamiltonians involving only the alleged geometrical properties of the (relevant) TAP solutions, which we then solve within Boltzmann formalism by means of a classical, Sanov-type large deviation analysis.

For these abstract models we furthermore derive a dual, Parisi-like formula for the free energy, establish their convergence to the Derrida-Ruelle cascades [21], and show that the abstract overlap concentrates on two possible values only – in complete agreement with the Parisi theory for models within the 1-step replica symmetry breaking (1RSB) approximation. The inherent nonlinearities also shed new light on the nature of the Parisi formula for the SK-model, see Section 3.1 below.

What is perhaps more, our findings *i)* considerably improve and clarify the cavity approach [18] to mean field spin glasses put forward by Aizenman, Ruzmaikina and Arguin [2, 1], as well as by Bolthausen and the first author [6, 7]; *ii)* provide a first[1] answer to the question raised in [5, p.109] concerning the link between the Bolthausen-Sznitman

Date: December 23, 2022.

2000 *Mathematics Subject Classification.* 60J80, 60G70, 82B44.

Key words and phrases. Mean Field Spin Glasses, Large Deviations, Gibbs-Boltzmann and Parisi Variational Principles, Random Energy Models and Derrida-Ruelle cascades.

This work has been supported by a DFG research grant, contract number 2337/1-1. NK wishes to express his gratitude to Anton Bovier for constant support and interest in this line of research.

[1]partial, due to the REM-assumption made here.

abstract cavity set-up [4] and the SK-model. In both cases, headway is made by enhancing the framework of these papers with the missing ingredient "TAP-free energy" : this simple, yet far-reaching insight is arguably the main contribution of this work.

This paper is organised as follows: in Section 2 we recall some of the main aspects from the picture canvassed in [25]. This will motivate and justify our abstract REM-like models which are introduced in Section 3, where the main results are also presented. The proofs are given in the fourth section, with some useful (technical) facts being recalled in the Appendix for the reader's convenience.

2. SK, TAP, and Plefka.

The SK-model is the archetypical mean field spin glass: for $N \in \mathbb{N}$, consider centered Gaussians $(g_{ij})_{1\leq i<j\leq N}$ issued on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. These Gaussians, *the disorder*, are assumed to be all independent and with variance $1/N$. The SK-Hamiltonian, defined on the Ising configuration space is then

$$\sigma \in \Sigma_N \stackrel{\text{def}}{=} \{\pm 1\}^N \mapsto H_N(\sigma) = \sum_{1\leq i<j\leq N} g_{ij}\sigma_i\sigma_j \,. \tag{1}$$

The quenched SK-free energy to inverse temperature $\beta > 0$ and external field $h \in \mathbb{R}$ is

$$N f_N(\beta, h) \equiv \log \sum_{\sigma\in\Sigma_N} \exp\left(\beta H_N(\sigma) + h\sum_{i=1}^N \sigma_i\right) . \tag{2}$$

In order to 'solve' the model, Thouless, Anderson and Palmer [25] play the delicate (and debatable) card of the spin magnetisation $m_i \equiv \langle\sigma_i\rangle_{\beta,h,N}\,, i = 1\ldots N$ as order parameter of the theory, with $\langle\rangle_{\beta,h,N}$ denoting average with respect to the quenched Gibbs measure. By means of a nonrigorous (and troublesome) diagrammatic expansion, Thouless *et. al.* suggest that the following approximation holds true with overwhelming probability:

$$N f_N(\beta, h) = \max_{\boldsymbol{m}\in\Delta} f_{\text{TAP}}(\boldsymbol{m}) + o(N) \quad (N \uparrow \infty), \tag{3}$$

where

i) the *TAP-free energy* is given by

$$\begin{aligned} N f_{TAP}(\boldsymbol{m}) \equiv \beta &\sum_{1\leq i<j\leq N} g_{ij} m_i m_j + h\sum_{i=1}^N m_i \\ &+ \frac{\beta^2}{4} N\left[1 - \frac{1}{N}\sum_{i=1}^N m_i^2\right]^2 - \sum_{i=1}^N I(m_i)\,; \end{aligned} \tag{4}$$

ii) for $m \in [-1, 1]$

$$I(m) \equiv \frac{1+m}{2}\log(1+m) + \frac{1-m}{2}\log(1-m)\,;$$

is the classical coin tossing rate function; and

iii) $\Delta \subset [-1, 1]^N$ an unspecified set of restrictions on the quenched magnetisations $\boldsymbol{m}$.

Remark 2.1. *Plefka has shown* [19] *that the TAP-approximation* (3) *neatly emerges from a high temperature expansion of the Gibbs potential. As any (finite volume) Gibbs potential, the map* $\boldsymbol{m} \mapsto f_{TAP}(\boldsymbol{m})$ *must necessarily be* concave*: in* [19] *it is claimed that this should indeed be the case provided that* $\boldsymbol{m}$ *satisfy*

$$\text{Plefka's criterium:} \qquad \frac{\beta^2}{N}\sum_{i=1}^{N}\left(1-m_i^2\right)^2 < 1\,.$$

This condition is widely accepted within the theoretical physics literature (in other words: this restriction should definitely appear in the definition of the Δ*-set) but there seems to be divergent opinions if this suffices for the validity of the high temperature expansions and thus of the TAP-approximation* (4)*. For a mathematical analysis of the TAP-Plefka approximation within Guerra's interpolation scheme* [14]*, the reader may check* [8] *and references therein. For an in-depth study of Plefka's convergence criteria for the SK-model, see* [15]*.*

Assuming the validity of the TAP-approximation (4), we therefore see that extremal "states" must necessarily be critical points of the TAP-free energy: taking the gradient, and rearranging, this leads to the TAP-equations

$$\nabla f_{\text{TAP}}(\boldsymbol{m}) = 0 \Longleftrightarrow m_i = \tanh\left(h + \beta\sum_{j\neq i} g_{ij}m_j - \beta^2(1-q_N(\boldsymbol{m}))m_i\right)\;,\; i = 1\ldots N, \tag{5}$$

where $q_N(\boldsymbol{m}) \overset{\text{def}}{=} (1/N)\sum_{j=1}^{N} m_j^2$.

In the theoretical physics literature it is claimed that, for large enough β, the TAP-equations admit exponentially many solutions $\boldsymbol{m}^\alpha, \alpha = 1\ldots 2^{\Theta N}$, where $\Theta = \Theta(\beta, h)$ is the currently unknown complexity.

Let us now assume to be given a TAP-solution $\boldsymbol{m}^\alpha$: using (5) we may express

$$\beta\sum_{j\neq i} g_{ij}m_j^\alpha = \tanh^{-1}(m_i^\alpha) - h + \beta^2\left\{1 - q_N(\boldsymbol{m}^\alpha)\right\}m_i^\alpha. \tag{6}$$

Plugging this into (4), and performing some straightforward algebraic manipulations, we obtain a representation of the TAP-FE as a sum of N local terms, as anticipated in the abstract: omitting the elementary details, the upshot reads (by a slight abuse of notation)

$$f_{\text{TAP}}(\alpha) = \frac{1}{N}\sum_i\left[\frac{1}{2}m_i^\alpha\tanh^{-1}(m_i^\alpha) + \frac{h}{2}m_i^\alpha - I(m_i^\alpha)\right] + \frac{\beta^2}{4}\left\{1 - q_N(\mathbf{m}^\alpha)^2\right\}. \tag{7}$$

What is crucial for our considerations is the nonlinear[2] term in the curly brackets above : this nonlinearity, and only this, will mark the point of departure from the abstract models studied in [6, 7]. Indeed, introducing the fields

$$h_i^\alpha \equiv \sum_{j\neq i} g_{ij}m_j^\alpha - \beta\left(1 - q_N(\mathbf{m}^\alpha)\right), \tag{8}$$

[2]as a matter of fact, quadratic: an analogous expression for the TAP-FE of any p-spin model is also available, in which case the quadratic term turns into a polynomial of degree $p \geq 3$, see e.g. [9].

and the associated *empirical measures*

$$l_{N,\alpha} \equiv \frac{1}{N}\sum_{i=1}^N \delta_{h_i^\alpha}, \tag{9}$$

we obtain, through the identity $I(y) = y\tanh^{-1}(y) - \log\cosh\tanh^{-1}(y)$, the following representation:

$$f_{\text{TAP}}(\alpha) = \Phi\left(\int g(x)^2 l_{N,\alpha}(dx)\right) + \int f_1(x) l_{N,\alpha}(dx), \tag{10}$$

where Φ, g, f_1 are real valued functions given by, respectively:

SK1) $x \ni \mathbb{R} \mapsto \Phi(x) \equiv \frac{\beta^2}{4}(1-x^2)$;
SK2) $x \ni \mathbb{R} \mapsto f_1(x) \equiv -\frac{1}{2}\log\left(1-\tanh^2(h+\beta x)\right) - \frac{\beta}{2}x\tanh(h+\beta x)$;
SK3) $x \ni \mathbb{R} \mapsto g(x) \equiv \tanh(h+\beta x)$.

The "nonlinear randomness" thus stems from the fluctuations of the *self-overlap*

$$q_{\text{EA}}(\alpha) \stackrel{\text{def}}{=} \int g(x)^2 l_{N,\alpha}(dx) = \int \tanh(h+\beta x)^2 l_{N,\alpha}(dx) \tag{11}$$

the *Edwards-Anderson order parameter*, indeed as claimed on [18, p. 69]. (The abstract version of the EA-order parameter will play a key role in our abstract models).

In order to contrive tractable models we shall perform a *REM-approximation*: we replace the local fields by a collection of *independent* standard Gaussians $h_i^\alpha \hookrightarrow g_{\alpha,i}$, where $i = 1\dots N$ and $\alpha = 1\dots 2^N$ (the complexity of the relevant TAP-solutions being currently unknown we simply set, here and henceforth, $\Theta = 1$). Denoting by

$$r_{N,\alpha} \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{g_{\alpha,i}} \tag{12}$$

the empirical measure, we thus consider the REM-approximation of the TAP free energy

$$f_{\text{REM-TAP}}(\alpha) = \Phi\left(\int g(x)^2\, r_{N,\alpha}(dx)\right) + \int f_1(x) r_{N,\alpha}(dx). \tag{13}$$

This leads to an approximation of the SK-model which only relies on the alleged geometrical organisation of the relevant[3] TAP-solutions: remark in fact that for the *abstract overlap* it holds

$$\begin{aligned} q_N(\alpha,\alpha') &\stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \tanh(h+\beta g_{\alpha,i})\tanh(h+\beta g_{\alpha',i}) \\ &\approx \mathbb{E}\left[\tanh(h+\beta g_{1,1})^2\right]\mathbf{1}_{\{\alpha=\alpha'\}} + \mathbb{E}\left[\tanh(h+\beta g_{1,1})\right]^2\mathbf{1}_{\{\alpha\neq\alpha'\}}, \end{aligned} \tag{14}$$

for large enough N, by the law of large numbers; this is indeed the "black or white dichotomy" of the REM [10], or, which is the same, the perpendicularity of TAP-solutions within a 1RSB Ansatz [18].

The above begs the following, natural questions:

Q1. what is the law of the overlap $q_N(\alpha,\alpha')$ under the Gibbs sampling (13) ?

[3]Again emphasizing that, at the time of writing, the meaning of "relevant" still isn't settled.

Q2. How does the abstract overlap transform under the *extensive* cavity dynamics [18]? This amount to studying, for $\varepsilon > 0$, the impact of an ε-perturbation of the Hamiltonian (13), i.e. to study the limiting Gibbs measure under transformations of the type

$$f_{\text{REM-TAP}}(\alpha) \hookrightarrow f^{(\varepsilon)}_{\text{REM-TAP}}(\alpha) \equiv f_{\text{REM-TAP}}(\alpha) + \varepsilon \int \log\cosh(x)\bar{r}_{N,\alpha}(dx), \tag{15}$$

where

$$\bar{r}_{N,\alpha}(dx) \equiv \frac{1}{N}\sum_{i=1}^{N} \delta_{\bar{g}_{\alpha,i}}, \tag{16}$$

and with $\{\bar{g}_{\alpha,i}\}_{\alpha,i}$ being some *fresh* disorder, i.e. a random field of centered, independent Gaussians which are also independent of the *reservoir* $\{g_{\alpha,i}\}_{\alpha,i}$. The limit we are interested in is, of course, the double limit $N \uparrow \infty$, followed by $\varepsilon \downarrow 0$.

The questions Q1 & Q2 are addressed below, in general setting.

3. The REM in TAP: definition, and main results.

We start with some notation: $(S, \mathcal{S})$ denotes a Polish space and $C(S), C_b(S)$ the spaces of all real valued continuous, resp. continous bounded functions on S.
For $d \in \mathbb{N}$, we denote by $\mathcal{M}_1^+\left(S^d\right)$ the space of Borel probability measures on S^d, endowed with the topology of weak convergence of measures. Notice that $\mathcal{M}_1^+\left(S^d\right)$ is Polish itself and we can consider one of the standard metrics (e.g. Prokhorov) that makes it a complete, separable metric space. Given a measure $\boldsymbol{\nu} \in \mathcal{M}_1^+\left(S^d\right)$ and $r > 0$ we indicate with $B_{\boldsymbol{\nu},r}$, resp. $\overline{B_{\boldsymbol{\nu},r}}$, the open, resp. closed ball in the metric space $\mathcal{M}_1^+\left(S^d\right)$ with center $\boldsymbol{\nu}$ and ray r. Our abstract Hamiltonian, which parallels (15), is defined through a continous functional $\boldsymbol{\Phi} : \mathcal{M}_1^+\left(S^2\right) \to \mathbb{R}$ of the form

$$\boldsymbol{\Phi}[\boldsymbol{\rho}] = \Phi_1\left[\rho_1\right] + \mathsf{E}_{\rho_2}(f_2) \tag{17}$$

where $\rho_1, \rho_2 \in \mathcal{M}_1^+(S)$ are the marginals of $\boldsymbol{\rho} \in \mathcal{M}_1^+\left(S^2\right)$ on the first, resp. second coordinate, $\Phi_1 : \mathcal{M}_1^+(S) \to \mathbb{R}$ is a continous functional and $f_2 \in C_b(S)$. To lighten notation, we shorten $\mathsf{E}_{\boldsymbol{\rho}}(u) \equiv \int u(\boldsymbol{x})\boldsymbol{\rho}(d\boldsymbol{x})$ for the expectation of a function $u : S^d \to \mathbb{R}$ w.r.t. a given $\boldsymbol{\rho} \in \mathcal{M}_1^+\left(S^d\right)$ and $\operatorname{var}_{\boldsymbol{\rho}}(u) \equiv \mathsf{E}_{\boldsymbol{\rho}}(u^2) - \mathsf{E}_{\boldsymbol{\rho}}(u)^2$ for the variance; we also write $\boldsymbol{\rho} \circ u^{-1}$ for the push-forward of $\boldsymbol{\rho}$ along a $\boldsymbol{\rho}$-measurable u. Finally we indicate with $\pi_1, \pi_2 : S^2 \to S$ the natural projections on the first, resp. second coordinate.

More specifically:

Definition 3.1. $\boldsymbol{\Phi} : \mathcal{M}_1^+\left(S^2\right) \to \mathbb{R}$ *denotes a functional of the form* (17) *with*

$$\Phi_1[\rho] \overset{def}{=} \Phi\left(\mathsf{E}_\rho(g^2)\right) + \mathsf{E}_\rho(f_1) \qquad \forall \rho \in \mathcal{M}_1^+(S)$$

where

H1) $\Phi : \mathbb{R} \to \mathbb{R}$ *is a twice differentiable concave function with* $\Phi''(x) < 0$ *for every* $x \in \mathbb{R}$*;*

H2) $g, f_1, f_2 \in C_b(S)$ *with* $\sup g = 1, \inf g = -1$.

The random Hamiltonian of our abstract model in a configuration α of the configuration space $\{1,\ldots,2^N\}$ for a finite volume $N\in\mathbb{N}$ is then defined as

$$H_N(\alpha) \stackrel{def}{=} N\boldsymbol{\Phi}\left[\boldsymbol{L}_{N,\alpha}\right] \tag{18}$$

where for every $\alpha\in\{1,\cdots,2^N\}$

$$\boldsymbol{L}_{N,\alpha} \stackrel{def}{=} \frac{1}{N}\sum_{i=1}^{N}\delta_{(X_{\alpha,i},Y_{\alpha,i})} \tag{19}$$

are the empirical measures associated to independent sequences of S^2-valued, i.i.d. random vectors $\{(X_{\alpha,i},Y_{\alpha,i})\}_{i=1}^N$ defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ with common distribution $\mu\otimes\gamma$, for some product measure $\mu\otimes\gamma\in\mathcal{M}_1^+(S^2)$.

Some words on the assumptions given in the above definition: concavity is imposed on Φ to warrant thermodynamical stability, somewhat in line with Plekfa's convergence criterium recalled in Remark 2.1; the boundedness of g is technically convenient, but also tailor-suited for our applications such as the SK-model. The variables $X_{\alpha,\cdot}$ correspond to the random energies associated to the configuration α, while the $Y_{\alpha,\cdot}$ encode the disorder necessary to perturbate the Hamiltonian with an extensive cavity dynamics.

To shorten notation, we define

$$f(x,y) \stackrel{\text{def}}{=} f_1(x)+f_2(y) \tag{20}$$

so that the nonlinear functional $\boldsymbol{\Phi}:\mathcal{M}_1^+\left(S^2\right)\to\mathbb{R}$ in the definition 3.1 is

$$\boldsymbol{\Phi}[\boldsymbol{\rho}] \stackrel{\text{def}}{=} \Phi\left(\mathsf{E}_{\rho_1}(g^2)\right)+\mathsf{E}_{\boldsymbol{\rho}}(f). \tag{21}$$

Notice that the continuity of the real valued function Φ and the boundedness of the three functions g,f_1,f_2, imply that the functional $\boldsymbol{\Phi}$ is continous on $\mathcal{M}_1^+\left(S^2\right)$ by duality and continous projection.

We consider the partition function, free energy and Gibbs measure associated to the Hamiltonian (18): for finite volume $N\in\mathbb{N}$, these are defined as usual as

$$Z_N \stackrel{\text{def}}{=} \sum_{\alpha=1}^{2^N}\exp H_N(\alpha), \qquad F_N \stackrel{\text{def}}{=} \frac{1}{N}\log Z_N, \tag{22}$$

and for $\alpha\in\{1,\ldots,2^N\}$,

$$\mathcal{G}_N(\alpha) \stackrel{\text{def}}{=} Z_N^{-1}\exp H_N(\alpha). \tag{23}$$

Finally, given two configurations α,α', we define their abstract overlap as

$$q_N(\alpha,\alpha') \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} g\left(X_{\alpha,i}\right) g\left(X_{\alpha',i}\right). \tag{24}$$

Our first result concerns the limiting free energy, and requires some notation: set

$$\mathcal{K} \stackrel{\text{def}}{=} \left\{\boldsymbol{\nu}\in\mathcal{M}_1^+\left(S^2\right): \quad H(\boldsymbol{\nu}\mid\mu\otimes\gamma)\le\log 2\right\}, \tag{25}$$

with $H(\boldsymbol{\nu}\mid\mu\otimes\gamma) \stackrel{\text{def}}{=} \mathsf{E}_{\boldsymbol{\nu}}\log\left(d\boldsymbol{\nu}/d(\mu\otimes\gamma)\right)$ being the usual relative entropy (see Section A for some relevant properties).

Theorem 3.2. *(Boltzmann-Gibbs principle). The infinite volume limit of the free energy* (22) *exists* $\mathbb{P}$*-almost surely, is non-random, and given by*

$$\lim_{N\to\infty} F_N = \sup_{\boldsymbol{\nu}\in\mathcal{K}} \boldsymbol{\Phi}[\boldsymbol{\nu}] - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) + \log 2\,. \tag{26}$$

A complete solution of our abstract models thus requires a discussion of the Boltzmann-Gibbs variational principle (26). This will be achieved by relating it to a simpler, Parisi-like variational principle in finite dimensions. To see how this comes about, we recall from [6] that for all $f \in \mathcal{L}\left(S^2, \mu\otimes\gamma\right)$ with

$$\mathcal{L}\left(S^2, \mu\otimes\gamma\right) \stackrel{\text{def}}{=} \left\{u \in C\left(S^2\right) : \quad \mathsf{E}_{\mu\otimes\gamma}\left(e^{\lambda u}\right) < \infty \ \forall\lambda\in\mathbb{R}\right\}, \tag{27}$$

almost surely it holds

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N}\log\sum_{\alpha=1}^{2^N} \exp N\mathsf{E}_{\boldsymbol{L}_{N,\alpha}}(f) &= \sup_{\boldsymbol{\nu}\in K}\left\{\mathsf{E}_{\mu\otimes\gamma}(f) - H(\boldsymbol{\nu} \mid \mu\otimes\gamma)\right\} \\ &= \inf_{0\le m\le 1}\left\{\frac{1}{m}\log\mathsf{E}_{\mu\otimes\gamma}(\exp\, mf) + \frac{\log 2}{m}\right\}. \end{aligned} \tag{28}$$

The first equality is the Boltzmann-Gibbs principle given in Theorem 3.2 for a Hamiltonian of the form (18) without nonlinear term. The second equality, in full agreement with the Parisi theory [18], establishes a duality between the Gibbs principle and a finite-dimensional minimization problem. We shall call the target function in the minimization problem a *Parisi function*; its (unique, cfr. [6]) minimizer $\overline{m}$ on $[0,1]$ gives the limiting Gibbs measure as the one whose Radon-Nikodym derivative with respect to $\mu\otimes\gamma$ is given by the Boltzmann factor $\exp\overline{m}f \equiv \exp\overline{m}(f_1\circ\pi_1 + f_2\circ\pi_2)$.
Our second main result provides the analogous duality principle for the nonlinear Hamiltonian (18). Specifically, defining for every $(q,m)\in[0,1]^2$

$$\begin{aligned} Z^{q,m} &\stackrel{\text{def}}{=} \mathsf{E}_{\mu\otimes\gamma}\left(\exp m\left[\Phi'(q)g^2\circ\pi_1 + f\right]\right) \\ &= \int \exp m\left[f_1(x) + \Phi'(q)g^2(x)\right]\mu(dx)\cdot\int \exp mf_2(y)\,\gamma(dy) \end{aligned} \tag{29}$$

and

$$P(q,m) \stackrel{\text{def}}{=} \Phi(q) - q\Phi'(q) + \frac{1}{m}\left(\log Z^{q,m} + \log 2\right), \tag{30}$$

the following holds.

Theorem 3.3. *(Parisi principle). It holds:*

$$\lim_{N\to\infty} F_N = \inf_{(q,m)\in[0,1]^2} \mathrm{P}(q,m) + \log 2\,,$$

$\mathbb{P}$*-almost surely.*

When comparing the *Parisi function* (30) with its counterpart (28) for the linear models one observes, in particular, the appearance of the term $\Phi(q) - q\Phi'(q)$. We will see in the course of the proof that such corrections, which are constituent parts of the Parisi free energy for the SK-model, play the role of *Lagrange multipliers* accounting for the in-built nonlinearities.

We now present our main results concerning the Gibbs measure. First of all, we note that the family of *generalized Gibbs measures*

$$\{\boldsymbol{\nu}^{q,m}\}_{(q,m)\in[0,1]^2} \subset \mathcal{M}_1^+\left(S^2\right)$$

defined through their Radon-Nikodym derivatives with respect to $\mu\otimes\gamma$:

$$\frac{d\boldsymbol{\nu}^{q,m}}{d(\mu\otimes\gamma)}(x,y) \stackrel{\text{def}}{=} \frac{\exp m\left[\Phi'(q)\, g^2(x) + f(x,y)\right]}{Z^{q,m}}, \tag{31}$$

satisfies

$$\begin{aligned} \log Z^{q,m} &= \mathsf{E}_{\boldsymbol{\nu}^{q,m}}\left(m\left[\Phi'(q)\, g^2\circ\pi_1 + f\right]\right) - H(\boldsymbol{\nu}^{q,m} \mid \mu\otimes\gamma) \\ &= m\left[\Phi'(q)\mathsf{E}_{\nu_1^{q,m}}\left(g^2\right) + \mathsf{E}_{\boldsymbol{\nu}^{q,m}}(f)\right] - H(\boldsymbol{\nu}^{q,m} \mid \mu\otimes\gamma) \end{aligned} \tag{32}$$

being $\nu_1^{q,m} \in \mathcal{M}_1^+(S)$ the first marginal of $\boldsymbol{\nu}^{q,m}$. Moreover, one can easily compute the partial derivatives of P to see that

$$\begin{aligned} \partial_m P(q,m) &= \frac{1}{m^2}\left[H(\boldsymbol{\nu}^{q,m} \mid \mu\otimes\gamma) - \log 2\right], \\ \partial_q P(q,m) &= \Phi''(q)\left[\mathsf{E}_{\nu_1^{q,m}}\left(g^2\right) - q\right]. \end{aligned} \tag{33}$$

Through equations (32), (33) we will relate the Boltzmann-Gibbs principle (26) with the Parisi function (30), showing that the solution of the latter is given by $\boldsymbol{\nu}^{\bar{q},\bar{m}}$ where $(\bar{q},\bar{m})$ minimizes $P\colon[0,1]^2\to\mathbb{R}$.

It is furthermore well-known that the Poisson-Dirichlet law for the *pure states* appears naturally as the weak limit of the Gibbs measure associated to a classical REM in low temperature. Our third main result, Theorem 3.4 below, aligns with this alleged universality.

In order to formulate the statement, we point out that for our abstract models, *low temperature* corresponds to the situation where the parameter $\bar{m}$ which achieves the minimum in Parisi principle is such that $\bar{m}<1$. As will become clear below, if the minimum point $(\bar{q},\bar{m})\in[0,1]^2$ of a Parisi function (30) is such that $\bar{m}<1$, then the measure

$$\bar{\boldsymbol{\nu}} \equiv \boldsymbol{\nu}^{\bar{q},\bar{m}} \in \mathcal{M}_1^+(S)$$

is the optimal measure for the Boltzmann-Gibbs principle and satisfies

$$\begin{cases} H(\bar{\boldsymbol{\nu}} \mid \mu\otimes\gamma) = \bar{m}\left[\Phi'(\bar{q})\bar{q} + \mathsf{E}_{\bar{\boldsymbol{\nu}}}(f)\right] - \log Z^{\bar{q},\bar{m}} = \log 2, \\ \mathsf{E}_{\bar{\nu}_1}(g^2) = \bar{q}, \end{cases} \tag{34}$$

where $\bar{\nu}_1 \equiv \nu_1^{\bar{q},\bar{m}}$ is the first marginal of $\bar{\boldsymbol{\nu}}$; in particular, in low temperature the side constraint on the relative entropy is saturated.

Theorem 3.4. *(Onset of Derrida-Ruelle cascades). Assume that at least one of the measures $\bar{\boldsymbol{\nu}}\circ(g\circ\pi_1-\bar{q})^{-1}$ or $\bar{\boldsymbol{\nu}}\circ(f-\mathsf{E}_{\bar{\boldsymbol{\nu}}}(f))^{-1}$ has a density w.r.t. the Lebesgue measure on $\mathbb{R}$. Then, for a system in low temperature, the point process $\{\mathcal{G}_N(\alpha)\}_{\alpha\le 2^N}$ associated to the Gibbs measure* (23) *converges weakly as $N\to\infty$ to a Poisson-Dirichlet point process with parameter $\overline{m}$.*

Let us furthermore denote by

$$\langle O_N(\alpha,\alpha')\rangle_N \stackrel{\text{def}}{=} Z_N^{-2}\sum_{\alpha,\alpha'} O_N(\alpha,\alpha')\mathcal{G}_N(\alpha)\mathcal{G}_N(\alpha')$$

the average with respect to the *replicated Gibbs measure* of a quantity $O_N(\alpha, \alpha')$ depending on two configurations α, α'; the following then holds for the limiting law of the abstract overlap (24).

Proposition 3.5. *(Overlap concentration). Let*

$$\bar{q} = \int g(x)^2 \bar{\nu}_1(dx), \qquad \bar{q}_0 \stackrel{def}{=} \left[\int g(x)\bar{\nu}_1(dx)\right]^2. \tag{35}$$

Then, under the assumptions of Theorem 3.4, it holds

$$\lim_{N\to\infty} \mathbb{E}\left\langle \delta_{\alpha=\alpha'}(q_N(\alpha,\alpha') - \bar{q})^2 \right\rangle_N = 0, \tag{36}$$

$$\lim_{N\to\infty} \mathbb{E}\left\langle \delta_{\alpha\neq\alpha'}\left(q_N(\alpha,\alpha') - \bar{q}_0\right)^2 \right\rangle_N = 0. \tag{37}$$

We thus see that our abstract models correctly recover *all* of the main features of the Parisi theory under a 1RSB approximation, in a generic setting. For the readers convenience, we shall conclude by briefly dwelling on the upshot of our analysis for the concrete case of the SK-model.

3.1. **SK vs. REM-TAP.** We recall that the 1RSB Ansatz [18] for the limiting free energy of the SK model (1) reads

$$f_{\text{1RSB-SK}}(\beta,h) \stackrel{\text{def}}{=} \inf_{0\leq q_0,\leq q_1\leq 1, m_1\in[0,1]} \left\{ \frac{\beta^2}{4}\left[(1-m_1)q_1^2 + m_1 q_0^2 - 2q_1 + 1\right] + \log 2 + \right.$$
$$\left. + \frac{1}{m_1}\int \log\left\{\int \exp m_1 \log\cosh\left[\beta(\sqrt{q_0}u + \sqrt{q_1 - q_0}v + h)\right]\varphi(dv)\right\}\varphi(du)\right\}, \tag{38}$$

where φ is the standard Gaussian measure on $\mathbb{R}$. As a matter of fact, this is a *degenerate* 2RSB-formula: indeed, the law of the pure states which hides behind (38) is that of a superposition of two Derrida-Ruelle processes [21] with parameters $0 \leq m_0 \leq m_1 \leq 1$, with the first one eventually absorbed through the limiting procedure $m_0 \downarrow 0$: this operation gives rise to the aforementioned degeneracy, to the "common trunk" captured by the second Gaussian du-integral, and stands behind the *third* order parameter q_0. For better comparison, one may in fact simply set $q_0 = 0$ in the above formula. Due to this complication, comparison with our 1RSB-setting shall be taken *cum grano*. Notwithstanding, we do get a number of important insights: for the Hamiltonian $H_N(\alpha) \equiv N f_{\text{REM-TAP}}(\alpha)$, and recalling SK1-3) above, the REM-TAP approximation leads in fact to a free energy

$$f_{\text{REM-TAP}}(\beta,h) = \inf_{0\leq q\leq 1, m\in[0,1]} \left\{ \frac{\beta^2}{4}(q^2+1) + \log 2 + \right.$$
$$\left. + \frac{1}{m}\log\int \exp m \log\cosh(h+\beta x)\cdot e^{m\left[-\frac{\beta}{2}x\tanh(h+\beta x) - \frac{\beta^2}{2}q\tanh^2(h+\beta x)\right]}\varphi(dx)\right\} \tag{39}$$

(The term $\exp m\left[-\frac{\beta}{2}x\tanh(h+\beta x) - \frac{\beta^2}{2}q\tanh^2(h+\beta x)\right]$ in (39) plays a role in the cavity dynamics only, and is completely irrelevant when it comes to the free energy: for the sake of the current discussion, such factor is an artefact which can also be immediately removed : simply set $f_1 \equiv 0$). The appearance of the *quadratic* term in the first line on the r.h.s. above is central to this work: although a constituent part of (38), such terms cannot be explained by the linear models studied in [6, 7]. As already mentioned, these corrections turn out to be Lagrange multipliers accounting for the non-linearities induced by the true REM-models hiding "within" the TAP-free energies.

Let us also spend a few words on the cavity dynamics for the SK-model, i.e. for the Hamiltonian given by the perturbed REM-TAP $H_{N;\varepsilon}(\alpha) \stackrel{\text{def}}{=} N f^{(\varepsilon)}_{\text{REM-TAP}}(\alpha)$, and with $f_2(y) \equiv f_2(\varepsilon; y) \equiv \varepsilon \log\cosh(y)$. Let us denote by $(\bar{q}_\varepsilon, \bar{m}_\varepsilon) \in [0,1]^2$ the minimum of the corresponding Parisi function, and by $\bar{\boldsymbol{\nu}}_\varepsilon = \boldsymbol{\nu}^{\bar{q}_\varepsilon, \bar{m}_\varepsilon}$ the associated generalized Gibbs measure. We furthermore denote by $F_{N,\varepsilon}$ the law of the overlap $q_N(\alpha,\alpha') = (1/N)\sum_{i=1}^N \tanh(h+\beta g_{\alpha,i})\tanh(h+\beta g_{\alpha',i})$ under the perturbed, finite volume Gibbs measure, and by F_N that of the *un*-perturbed:

$$x \in [0,1] \mapsto F_{N,\varepsilon}(x) \equiv \mathbb{E}\left\langle \mathbf{1}_{q_N(\alpha,\alpha') \le x} \right\rangle_{N;\varepsilon}, \qquad F_N(x) \equiv F_{N,0}(x). \tag{40}$$

Proposition 3.5 would imply that in low temperature ($\bar{m}_\varepsilon < 1$) the law of the overlap of two relevant TAP-solutions for the SK-model is given by

$$F_\varepsilon(x) \stackrel{\text{def}}{=} \lim_{N\to\infty} F_{N,\varepsilon}(x) = \begin{cases} 0 & \text{if} \quad x < \bar{q}_{0;\varepsilon} \\ \bar{m}_\varepsilon & \text{if} \quad \bar{q}_{0;\varepsilon} \le x < \bar{q}_\varepsilon \ ; \\ 1 & \text{otherwise} \end{cases} \tag{41}$$

where

$$\bar{q}_\varepsilon = \int \tanh^2(h+\beta x)\bar{\boldsymbol{\nu}}_\varepsilon(dx,dy), \qquad \bar{q}_{0;\varepsilon} = \left[\int \tanh(h+\beta x)\bar{\boldsymbol{\nu}}_\varepsilon(dx,dy)\right]^2.$$

A Parisi fixed point equation [18, III.63] encoding the stability of the hierarchical structure under the extensive cavity dynamics would then appear through the *imposition*

$$F(x) \stackrel{(!)}{=} \lim_{\varepsilon\downarrow 0} F_\varepsilon(x), \qquad \forall\, x \in [0,1]. \tag{42}$$

In case of a REM-approximation (1RSB), it is easily seen that the self-consistency (42) is, in fact, *automatically* satisfied and as such a void requirement: in order to avoid trivializations one needs a much more sophisticated GREM-TAP approximation, but this will be addressed in forthcoming works.

4. Proofs.

4.1. **The Boltzmann-Gibbs principle: proof of Theorem 3.2.** In this subsection we prove Theorem 3.2, i.e. the validity of the following Boltzmann-Gibbs principle

$$\lim_{N\to\infty} F_N = \lim_{N\to\infty} \frac{1}{N}\log Z_N = \sup_{\boldsymbol{\nu}\in\mathcal{K}} \Omega[\boldsymbol{\nu}] + \log 2 \quad \mathbb{P}-\text{a.s.}, \tag{43}$$

where to lighten notation we shortened

$$\begin{aligned} \Omega[\boldsymbol{\rho}] &\stackrel{\text{def}}{=} \boldsymbol{\Phi}[\boldsymbol{\rho}] - H(\boldsymbol{\rho} \mid \mu\otimes\gamma) \\ &= \Phi\left(\mathsf{E}_{\boldsymbol{\rho}_1}(g^2)\right) + \mathsf{E}_{\boldsymbol{\rho}}(f) - H(\boldsymbol{\rho} \mid \mu\otimes\gamma) \end{aligned} \tag{44}$$

being $\boldsymbol{\Phi} : \mathcal{M}_1^+(S^2) \to \mathbb{R}$ the functional defined in 3.1.

Remark 4.1. *It is a well known fact that* $\boldsymbol{\nu} \in \mathcal{M}_1^+(S^2) \to H(\boldsymbol{\nu} \mid \mu\otimes\gamma)$ *is lower semicontinous, strictly convex and with compact sub-levels. Therefore, for compact* $\mathcal{C} \subset \mathcal{M}_1^+(S^2)$ *and an upper semicontinous* Ω*, the generalized Bolzano-Weierstrass Theorem on metric spaces ensures the existence of a solution to*

$$\sup_{\boldsymbol{\nu}\in\mathcal{C}} \Omega[\boldsymbol{\nu}]. \tag{45}$$

Moreover, the assumption H1) on Φ*, together with the fact that* $g \in [-1,1]$*, suffices to get strict concavity of the functional* $\Omega : \mathcal{M}_1^+(S^2) \to \mathbb{R}$*. This implies that, for a convex* $\mathcal{C} \subseteq \mathcal{M}_1^+(S^2)$*, the existence of a solution to a problem of the form* (45) *implies its uniqueness. In other words: if* $\mathcal{C}$ *is compact, there exists a solution to* (45)*; if* $\mathcal{C}$ *is also convex, such solution is furthemore unique: since the set* $\mathcal{K}$*, i.e. the sub-level of the relative entropy appearing in the Boltzmann-Gibbs principle* (43)*, is convex and compact, the variational principle* (43) *admits a unique solution. (Recall also that, in a complete metric space like* $\mathcal{M}_1^+(S^2)$ *a set is sequentially compact if and only if it is compact, every compact set is closed and every closed subset of a compact set is compact).*

We will proceed via a second moment method on the counting variable

$$\mathcal{A}_N(E) = \#\{\alpha : \quad \boldsymbol{L}_{N,\alpha} \in E\},$$

defined for every $E \subseteq \mathcal{M}_1^+(S^2)$. The proof is analogous to the one given in [6] and relies both on the independence of the $\{\boldsymbol{L}_{N,\alpha}\}_{\alpha=1}^{2^N}$ as well as on the continuity of $\boldsymbol{\rho} \in \mathcal{M}_1^+(S^2) \to \boldsymbol{\Phi}[\boldsymbol{\rho}]$. We start with a technical Lemma, which will be useful also later, that provides the upper bound to the free energy - see equation (46). Here and henceforth we denote by $B_r \stackrel{\text{def}}{=} B_{\bar{\boldsymbol{\nu}},r} = \{\boldsymbol{\nu} \in \mathcal{M}_1^+(S^2) : \quad d(\boldsymbol{\nu}, \bar{\boldsymbol{\nu}}) < r\}$ the open ball centered in $\bar{\boldsymbol{\nu}}$, and with radius $r > 0$.

Lemma 4.2. *Let* $\bar{\boldsymbol{\nu}} \stackrel{def}{=} argmax_{\mathcal{K}}\Omega$*. Then,* $\mathbb{P}$*-almost surely :*

- *For all* $r > 0$*,*

$$\limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha:\boldsymbol{L}_{N,\alpha}\notin B_r} e^{N\boldsymbol{\Phi}[L_{N,\alpha}]} < F,$$

- *With* $F \stackrel{def}{=} \max_{\mathcal{K}} \Omega[\boldsymbol{\nu}] = \Omega[\bar{\boldsymbol{\nu}}]$*, it holds:*

$$\limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha=1}^{2^N} e^{N\boldsymbol{\Phi}[\boldsymbol{L}_{N,\alpha}]} \leq F\,. \tag{46}$$

Proof. Let $r \geq 0$. Shorten $B_0^c \equiv \mathcal{M}_1^+(S^2)$, denote by B_r^c the complement of B_r in $\mathcal{M}_1^+(S^2)$:

$$B_r^c \stackrel{\text{def}}{=} \{\boldsymbol{\rho} \in \mathcal{M}_1^+\left(S^2\right) : \quad d(\boldsymbol{\rho}, \bar{\boldsymbol{\nu}}) \geq r\}) = \mathcal{M}_1^+\left(S^2\right) \setminus B_r,$$

and set

$$F_r \stackrel{\text{def}}{=} \begin{cases} \sup\limits_{\mathcal{K}\cap B_r^c} \Omega & \text{if} \quad \mathcal{K} \cap B_r^c \neq \emptyset, \\ -\infty & \text{otherwise.} \end{cases} \tag{47}$$

Notice that if $r : \mathcal{K} \cap B_r^c = \emptyset$ it trivially holds $F_r < F_0 \equiv F$. If instead $r : \mathcal{K} \cap B_r^c \neq \emptyset$ then, by compactness (see Remark 4.1), we can find $\boldsymbol{\nu}_r \in \mathcal{K} \cap B_r^c$, $\boldsymbol{\nu}_r \neq \bar{\boldsymbol{\nu}}$ s.t.

$$F_r = \max_{\mathcal{K}\cap B_r^c} \Omega = \Omega[\boldsymbol{\nu}_r]$$

and, as the maximal measure $\bar{\boldsymbol{\nu}}$ of Ω on $\mathcal{K}$ is unique, the strict inequality

$$F_r = \Omega[\boldsymbol{\nu}_r] < \Omega[\bar{\boldsymbol{\nu}}] = F$$

holds also in this case. Hence, in order to settle the Lemma, we just need to prove the following

$$\limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha : \boldsymbol{L}_{N,\alpha}\in B_r^c} e^{N\boldsymbol{\Phi}[\boldsymbol{L}_{N,\alpha}]} \le F_r \quad \forall r \ge 0. \tag{48}$$

In order to see this, consider a decreasing sequence of positive real numbers $\{a_n\}_{n\in\mathbb{N}}$ with $a_n \in (0,1)$ for all $n \in \mathbb{N}$ and $a_n \to 0$ as $n \to \infty$. Then fix $n \in \mathbb{N}$ and consider the open set $\mathcal{K}_{a_n}^c \stackrel{\text{def}}{=} \mathcal{M}_1^+(S^2) \setminus \mathcal{K}_{a_n}$, together with its closure $\overline{\mathcal{K}_{a_n}^c}$, where for every $a \ge 0$

$$\mathcal{K}_a \stackrel{\text{def}}{=} \{\boldsymbol{\nu} \in \mathcal{M}_1^+\left(S^2\right) : \quad H\left(\boldsymbol{\nu} \mid \mu\otimes\gamma\right) \le \log 2 + a\}.$$

Sanov's Theorem implies that

$$\begin{aligned}\mathbb{P}\left(L_{N,1} \in \overline{\mathcal{K}_{a_n}^c}\right) &= \exp\big[-\inf_{\boldsymbol{\rho}\in\overline{\mathcal{K}_{a_n}^c}} H(\boldsymbol{\rho} \mid \mu\otimes\gamma)N + o(N)\big] \\ &\le 2^{-N} e^{-a_n N + o(N)}\end{aligned} \qquad (N\to\infty), \tag{49}$$

hence

$$\mathbb{P}\left(\exists \alpha \in \{1, \cdots, 2^N\} : \quad \boldsymbol{L}_{N,\alpha} \in \mathcal{K}_{a_n}^c\right) \le 2^N \mathbb{P}\left(L_{N,1} \in \overline{\mathcal{K}_{a_n}^c}\right) \le e^{-a_n N + o(N)}. \tag{50}$$

Through the Borel-Cantelli lemma, (50) implies that

$$\limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha : \boldsymbol{L}_{N,\alpha}\in B_r^c} e^{N\boldsymbol{\Phi}[\boldsymbol{L}_{N,\alpha}]} = \limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha : \boldsymbol{L}_{N,\alpha}\in \mathcal{K}_{a_n}\cap B_r^c} e^{N\boldsymbol{\Phi}(\boldsymbol{L}_{N,\alpha})} \quad \mathbb{P}-\text{a.s.} \tag{51}$$

If r, n are such that $\mathcal{K}_{a_n} \cap B_r^c = \emptyset$, the latter implies

$$\limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha : \boldsymbol{L}_{N,\alpha}\in B_\epsilon^c} e^{N\boldsymbol{\Phi}[\boldsymbol{L}_{N,\alpha}]} = -\infty \quad \mathbb{P}-\text{a.s.} \tag{52}$$

Notice that by construction it holds $\mathcal{K} = \mathcal{K}_0 \subseteq \mathcal{K}_{a_n}$, which implies that if $\mathcal{K}_{a_n} \cap B_r^c$ is empty then $F_r = -\infty$, i.e. the claim (48) is exactly (52).

We can therefore assume $r, n : \ \mathcal{K}_{a_n} \cap B_r^c \neq \emptyset$ and, for an aribitrary $\delta > 0$ cover the compact set $\mathcal{K}_{a_n} \cap B_r^c$ through open balls

$$\mathcal{K}_{a_n} \cap B_r^c \subset \bigcup_{\boldsymbol{\nu}\in\mathcal{K}_{a_n}\cap B_\epsilon^c} B_{\boldsymbol{\nu}}(r_{\boldsymbol{\nu}})$$

such that for each $\boldsymbol{\nu} \in \mathcal{K}_{a_n} \cap B_r^c$ the associated ray $r_{\boldsymbol{\nu}} > 0$ is small enough s.t.

$$\boldsymbol{\Phi}(\boldsymbol{\rho}) - \boldsymbol{\Phi}(\boldsymbol{\nu}) < \delta \quad \forall \boldsymbol{\rho} \in B_{\boldsymbol{\nu}}(r_{\boldsymbol{\nu}}), \tag{53}$$

and

$$H(\boldsymbol{\nu} \mid \mu\otimes\gamma) - \inf_{\boldsymbol{\rho}\in\overline{B_{\boldsymbol{\nu}}(r_{\boldsymbol{\nu}})}} H(\boldsymbol{\rho} \mid \mu\otimes\gamma) < \frac{\delta}{2}. \tag{54}$$

This is clearly possible by upper semicontinuity of $\boldsymbol{\Phi}$ and the fact that

$$\inf_{\boldsymbol{\rho}\in\overline{B_{\boldsymbol{\nu}}(r)}} H(\boldsymbol{\rho} \mid \mu\otimes\gamma) \to H(\boldsymbol{\nu} \mid \mu\otimes\gamma) \quad (r\to 0).$$

By compactness, we can extract a finite sub-cover: $\exists M < \infty, \{\boldsymbol{\nu}_i\}_{i=1}^M \subset \mathcal{K}_{a_n} \cap B_r^c$ s.t. $\mathcal{K}_{a_n} \cap B_r^c \subset \bigcup_{i=1}^M B_{\boldsymbol{\nu}_i}(r_i)$, where we shortened

$$r_i \stackrel{\text{def}}{=} r_{\boldsymbol{\nu}_i}, \;\; B_i \stackrel{\text{def}}{=} B_{\boldsymbol{\nu}_i}(r_i).$$

Therefore, almost surely, it holds

$$\begin{aligned} \limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha: \boldsymbol{L}_{N,\alpha} \in \mathcal{K}_{a_n} \cap B_r^c} e^{N\boldsymbol{\Phi}[\boldsymbol{L}_{N,\alpha}]} &\leq \limsup_{N\to\infty} \frac{1}{N} \log \sum_{i=1}^M \sum_{\alpha: \boldsymbol{L}_{N,\alpha} \in B_i} \exp N\boldsymbol{\Phi}(\boldsymbol{L}_{N,\alpha}) \\ &\leq \delta + \limsup_{N\to\infty} \frac{1}{N} \log \sum_{i=1}^M \exp N\boldsymbol{\Phi}(\boldsymbol{\nu}_i) \mathcal{A}_N \left(\overline{B}_i\right) \end{aligned} \tag{55}$$

(recall that $\mathcal{A}_N\left(\overline{B}_i\right)$ is the variable that counts the $\alpha-$s for which $\boldsymbol{L}_{N,\alpha} \in B_i$).

Markov's inequality implies that for each $i = 1, \cdots, M$

$$\mathbb{P}\left(\mathcal{A}_N\left(\overline{B}_i\right) \geq e^{N[\log 2 - H(\boldsymbol{\nu}_i | \mu\otimes\gamma) + \delta]}\right) \leq 2^{-N} \mathbb{E}\, \mathcal{A}_N\left(\overline{B}_i\right) \exp N[H(\boldsymbol{\nu}_i \mid \mu\otimes\gamma) - \delta] \tag{56}$$

with

$$\mathbb{E}\left(\mathcal{A}_N\left(\overline{B}_i\right)\right) = 2^N \mathbb{P}\left(L_{N,1} \in \overline{B}_i\right) = 2^N \exp\left[-\inf_{\boldsymbol{\rho}\in\overline{B}_i} H\left(\boldsymbol{\rho} \mid \mu\otimes\gamma\right) N + o(N)\right] \quad (N\to\infty), \tag{57}$$

the last step by Sanov's theorem. Plugging (57) into (56), and using (54), we find that for each $i \in \{1, \cdots, M\}$

$$\mathbb{P}\left(\mathcal{A}_N\left(\overline{B}_i\right) < e^{N[\log 2 - H(\boldsymbol{\nu}_i | \mu\otimes\gamma) + \delta]}\right) \geq 1 - \exp\left[-\frac{\delta}{2} N + o(N)\right] \quad (N\to\infty).$$

This, together with Borel-Cantelli, implies

$$\begin{aligned} &\limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha: \boldsymbol{L}_{N,\alpha} \in \mathcal{K}_{a_n} \cap \in B_r^c} e^{N\boldsymbol{\Phi}[\boldsymbol{L}_{N,\alpha}]} \\ &\qquad \leq \limsup_{N\to\infty} \frac{1}{N} \log \sum_{i=1}^M \exp N(\boldsymbol{\Phi}(\boldsymbol{\nu}_i) - H(\boldsymbol{\nu}_i \mid \mu\otimes\gamma)) + \log 2 + 2\delta \\ &\qquad \leq \sup_{\boldsymbol{\nu}\in B_\epsilon^c \cap \mathcal{K}_{a_n}} \boldsymbol{\Phi}(\boldsymbol{\nu}) - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) + \log 2 + 2\delta. \end{aligned}$$

Being δ arbitrary, the latter, together with (51) give

$$\limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha: \boldsymbol{L}_{N,\alpha} \in B_r^c} e^{N\boldsymbol{\Phi}[\boldsymbol{L}_{N,\alpha}]} \leq \sup_{\boldsymbol{\nu}\in\mathcal{K}_{a_n}\cap B_r^c} \Omega[\boldsymbol{\nu}] + \log 2.$$

Notice that, again by compactness, we can select a measure $\bar{\boldsymbol{\nu}}_n \in \mathcal{K}_{a_n} \cap B_r^c$ s.t.

$$\max_{\mathcal{K}_{a_n}\cap B_r^c} \Omega = \Omega[\bar{\boldsymbol{\nu}}_n],$$

which implies that, for every $n \in \mathbb{N}$,

$$\limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha: \boldsymbol{L}_{N,\alpha} \in B_r^c} e^{N\boldsymbol{\Phi}[\boldsymbol{L}_{N,\alpha}]} \leq \Omega[\bar{\boldsymbol{\nu}}_n] + \log 2.$$

By construction it holds $B_r^c \cap \mathcal{K}_{a_n} \subseteq B_r^c \cap \mathcal{K}_1 \ \forall n \in \mathbb{N}$, specifically the whole sequence $\{\bar{\boldsymbol{\nu}}_n\}_{n\in\mathbb{N}}$ lies in the compact set $B_r^c \cap \mathcal{K}_1$ and we can extract a convergent subsequence: $\{\bar{\boldsymbol{\nu}}_{n_k}\}_{k\in\mathbb{N}}$ s.t. $\bar{\boldsymbol{\nu}}_{n_k} \to \boldsymbol{\xi}$ weakly as $k \to \infty$ for some $\boldsymbol{\xi} \in \mathcal{K}_1 \cap B_r^c$. We get

$$\begin{aligned}\limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha: \boldsymbol{L}_{N,\alpha} \in B_r^c} e^{N\boldsymbol{\Phi}[\boldsymbol{L}_{N,\alpha}]} &\leq \limsup_{k\to\infty} \Omega[\bar{\boldsymbol{\nu}}_{n_k}] + \log 2 \\ &\leq \Omega[\boldsymbol{\xi}] + \log 2,\end{aligned} \tag{58}$$

the last step by upper semicontinuity of Ω. It is also immediately checked that

$$\boldsymbol{\xi} \in \mathcal{K} \cap B_r^c \equiv \mathcal{K}_0 \cap B_r^c, \tag{59}$$

which implies the claim (48) straightforwardly through (58).

The validity of (59) is a consequence of the fact that for every $\delta > 0$ there exists $k_\delta > 0$ s.t. $\bar{\boldsymbol{\nu}}_{n_k} \in \mathcal{K}_\delta \cap B_r^c$ for all $k \geq k_\delta$ and that by compactness of the latter it must hold $\boldsymbol{\xi} \in \mathcal{K}_\delta \cap B_r^c$. Specifically:

$$\begin{aligned}\boldsymbol{\xi} \in \bigcap_{\delta>0} K_\delta \cap B_r^c &= \bigcap_{\delta>0} \{\boldsymbol{\nu} : H(\boldsymbol{\nu} \mid \mu\otimes\gamma) \leq \log 2 + \delta\} \cap B_r^c \\ &= \{\boldsymbol{\nu} : H(\boldsymbol{\nu} \mid \mu\otimes\gamma) \leq \log 2\} \cap B_r^c = \mathcal{K} \cap B_r^c,\end{aligned}$$

and the Lemma follows. □

The proof of the lower bound requires a refinement of our moments' analysis. A key ingredient is the strong concentration of the $\mathcal{A}_N(V)$-random variable around its mean:

Lemma 4.3. *(Variance estimate) For every $\boldsymbol{\nu} \in \mathcal{K}$ and any open neigborhood U of $\boldsymbol{\nu}$, there exists $B_{\boldsymbol{\nu},r} \subset U$ and $\delta > 0$ such that for large enough N*

$$\mathbb{V}ar\mathcal{A}_N(B_{\boldsymbol{\nu},r}) \leq e^{-N\delta}(\mathbb{E}\,\mathcal{A}_N(B_{\boldsymbol{\nu},r}))^2. \tag{60}$$

Proof. For each $B_{\boldsymbol{\nu},r}$, and by independence, the second moment of $\mathcal{A}_N(B_{\boldsymbol{\nu},r})$ satisfies

$$\begin{aligned}\mathbb{E}\left[\mathcal{A}_N^2(B_{\boldsymbol{\nu},r})\right] &= \sum_\alpha \mathbb{P}(\boldsymbol{L}_{N,\alpha} \in B_{\boldsymbol{\nu},r}) + \sum_{\alpha\neq\alpha'} \mathbb{P}(\boldsymbol{L}_{N,\alpha} \in B_{\boldsymbol{\nu},r})\,\mathbb{P}(\boldsymbol{L}_{N,\alpha'} \in B_{\boldsymbol{\nu},r}) \\ &\leq 2^N \mathbb{P}\left(\boldsymbol{L}_{N,1} \in \overline{B_{\boldsymbol{\nu},r}}\right) + [\mathbb{E}\,\mathcal{A}_N(B_{\boldsymbol{\nu},r})]^2,\end{aligned} \tag{61}$$

and thus

$$\mathbb{V}ar\mathcal{A}_N(B_{\boldsymbol{\nu},r}) \leq 2^N \mathbb{P}\left(\boldsymbol{L}_{N,1} \in \overline{B_{\boldsymbol{\nu},r}}\right). \tag{62}$$

Notice that the statement of the Lemma is trivial if $\boldsymbol{\nu} = \mu\otimes\gamma$, therefore we assume $\boldsymbol{\nu} \neq \mu\otimes\gamma$; this implies that there exists $B_{\boldsymbol{\nu},r} \subset U$ and $\eta > 0$ such that $\mu\otimes\gamma \notin \overline{B_{\boldsymbol{\nu},r}}$ and

$$H(B_{\boldsymbol{\nu},r}) = H(\overline{B_{\boldsymbol{\nu},r}}) = H(\boldsymbol{\nu} \mid \mu\otimes\gamma) - \eta.$$

Together with Sanov's theorem, the latter implies that for N large

$$\begin{aligned} 2^N \mathbb{P}\left(\boldsymbol{L}_{N,1} \in \overline{B_{\boldsymbol{\nu},r}}\right) &\leq \exp N\left[\log 2 - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) + \frac{7}{6}\eta\right] \\ &\leq e^{-\frac{\eta}{2}N} \exp 2N\left[\log 2 - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) + \frac{5}{6}\eta\right] \\ &\leq e^{-\frac{\eta}{2}N}\left[2^N \mathbb{P}\left(\boldsymbol{L}_{N,1} \in B_{\boldsymbol{\nu},r}\right)\right]^2 = e^{-\frac{\eta}{2}N}\left[\mathbb{E}\,\mathcal{A}_N\left(B_{\boldsymbol{\nu},r}\right)\right]^2 \end{aligned} \tag{63}$$

where in the second line we used that, as $\boldsymbol{\nu} \in K$, $\log 2 - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) \geq 0$. The Lemma now follows from (62) with $\delta \equiv \frac{\eta}{2}$.

□

For $\boldsymbol{\nu} \in \mathcal{K}$, $\delta > 0$ let U be an open neigborhood of $\boldsymbol{\nu}$ such that

$$\boldsymbol{\Phi}(\boldsymbol{\nu}) - \Phi(\boldsymbol{\rho}) \leq \delta \qquad \forall\, \boldsymbol{\rho} \in U,$$

and let $\overline{B_{\boldsymbol{\nu},r}} \subset U$ be the ball prescribed by Lemma 4.3. Then

$$\begin{aligned} \liminf_{N\to\infty} \mathcal{F}_N &\geq \liminf_{N\to\infty} \frac{1}{N} \log \sum_{\alpha\colon \boldsymbol{L}_{N,\alpha} \in B_{\boldsymbol{\nu},r}} \exp N\boldsymbol{\Phi}(\boldsymbol{L}_{N,\alpha}) \\ &\geq \boldsymbol{\Phi}(\boldsymbol{\nu}) + \liminf_{N\to\infty} \frac{1}{N} \log \mathcal{A}_N\left(B_{\boldsymbol{\nu},r}\right) - \delta. \end{aligned}$$

Moreover, Lemma (4.3) together with Chebyshev's inequality and Borel-Cantelli, immediately implies that

$$\lim_{N\to\infty} \frac{1}{N} \log \mathcal{A}_N\left(B_{\boldsymbol{\nu},r}\right) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}\,\mathcal{A}_N\left(B_{\boldsymbol{\nu},r}\right) \qquad \mathbb{P}\text{-a.s.}$$

By Sanov's theorem, we thus get

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathcal{A}_N\left(B_{\boldsymbol{\nu},r}\right) \geq -\inf_{\boldsymbol{\rho}\in B_{\boldsymbol{\nu},r}} H(\boldsymbol{\rho} \mid \mu\otimes\gamma) + \log 2 \qquad \mathbb{P}\text{-a.s.}$$

All in all,

$$\begin{aligned} \liminf_{N\to\infty} \mathcal{F}_N &\geq \boldsymbol{\Phi}(\boldsymbol{\nu}) - \inf_{\boldsymbol{\rho}\in B_{\boldsymbol{\nu},r}} H(\boldsymbol{\rho} \mid \mu\otimes\gamma) + \log 2 - \epsilon \\ &\geq \boldsymbol{\Phi}(\boldsymbol{\nu}) - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) + \log 2 - \delta. \end{aligned}$$

Since δ is arbitrary, by taking the supremum over $\boldsymbol{\nu} \in \mathcal{K}$, we get the lower bound

$$\liminf_{N\to\infty} \mathcal{F}_N \geq \sup_{\boldsymbol{\nu}\in\mathcal{K}} \boldsymbol{\Phi}(\boldsymbol{\nu}) - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) + \log 2,$$

and Theorem 3.2 follows.

4.2. **The Parisi principle: proof of Theorem 3.3.** Our proof of the Parisi principle crucially relies on some classic properties of the relative entropy functional $\boldsymbol{\rho} \in \mathcal{M}_1^+(S^2) \to H(\boldsymbol{\rho} \mid \mu\otimes\gamma) \in [0,\infty]$ which are recalled in the Appendix A. We start with a couple of technical and straightforward results.

Lemma 4.4. *Let* $(\overline{q}, \overline{m})$ *be a minimum point of the Parisi function* P*, defined in* (30)*, on* $[0,1]^2$*. Then*

$$\begin{aligned} \overline{m} \in (0,1) \quad &\Rightarrow \quad H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma) = \log 2, \\ \overline{m} = 1 \quad &\Rightarrow \quad H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma) \leq \log 2, \end{aligned} \tag{64}$$

$$\mathsf{E}_{\nu_1^{\bar{q},\bar{m}}}\left(g^2\right) = \overline{q} \tag{65}$$

being $\nu_1^{\bar{q},\bar{m}} \in \mathcal{M}_1^+(S)$ *the marginal of* $\boldsymbol{\nu}^{\bar{q},\bar{m}} \in \mathcal{M}_1^+(S^2)$ *on the first coordinate.*
Moreover, it holds

$$\boldsymbol{\Phi}[\boldsymbol{\nu}^{\bar{q},\bar{m}}] - H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma) = \mathrm{P}(\overline{q},\overline{m}) + \log 2. \tag{66}$$

Proof. As a lower semicontinuous function on a compact set, P must attain a minimum on $[0,1]^2$, which we denote by $(\overline{q},\overline{m}) \in [0,1]^2$. Since $P(q,m) \to +\infty$ for $m \to 0^+$, $\overline{m}$ must lie in $(0,1]$: this implies that $\partial_m P(\overline{q},\overline{m}) \leq 0$ with equality if $\overline{m} \in (0,1)$. Through this and the expression for $\partial_m P$ in (33) we get that the implications (64) hold true.
Now, assume *ad absurdum*

$$\mathsf{E}_{\nu_1^{q,m}}\left(g^2\right) \neq \overline{q}. \tag{67}$$

Since $\Phi'' < 0$ the assumption (67) implies $\overline{q} \in \{0,1\}$, as otherwise one would get from the equation for $\partial_q P$ in (33) that $\partial_q P(\overline{q},\overline{m}) \neq 0$ which is impossible if $\overline{q} \in (0,1)$. As 0 is a left border value in order for it to be a component of a minumum point it should hold $\partial_q P(\overline{q},\overline{m}) \geq 0$. But this together with (67), (33) and $\Phi'' < 0$ would imply $\mathsf{E}_{\boldsymbol{\nu}_1^{q,m}}(g^2) < \overline{q} = 0$, which contradicts the assumption $-1 \leq g \leq 1$. Using that $\sup g^2 \leq 1$, the case $\overline{q} = 1$ brings to an analogous contradiction. It thus follows that also (65) holds true.

Moreover, as (32) is valid for all couples (q,m), we find that P can be rewritten as

$$\begin{aligned} P(q,m) = \Phi(q) - q\Phi'(q) + \mathsf{E}_{\boldsymbol{\nu}^{q,m}}\left(f + \Phi'(q)g^2 \circ \pi_1\right) + \\ + \frac{1}{m}\left(\log 2 - H(\boldsymbol{\nu}^{q,m} \mid \mu\otimes\gamma)\right) - \log 2. \end{aligned} \tag{68}$$

Using (65) we get

$$\begin{aligned} P(\overline{q},\overline{m}) &= \Phi(\overline{q}) - \overline{q}\Phi'(\overline{q}) + \mathsf{E}_{\boldsymbol{\nu}^{\bar{q},\bar{m}}}(f) + \Phi'(\overline{q})\overline{q} + \frac{1}{\overline{m}}\left(\log 2 - H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma)\right) - \log 2 \\ &= \Phi\left(\mathsf{E}_{\boldsymbol{\nu}^{\bar{q},\bar{m}}}\left(g^2 \circ \pi_1\right)\right) + \mathsf{E}_{\boldsymbol{\nu}^{\bar{q},\bar{m}}}(f) + \frac{1}{\overline{m}}\left(\log 2 - H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma)\right) - \log 2. \end{aligned} \tag{69}$$

From the latter and (64) the equation (66) follows straighforwardly. This concludes the Proof of the Lemma. □

Proposition 4.5. *There exists a minimum point* $(\overline{q},\overline{m})$ *of* P *on* $[0,1]^2$ *such that the unique optimal measure for the Boltzmann-Gibbs principle* (26) *is* $\boldsymbol{\nu}^{\bar{q},\bar{m}}$*; i.e.*

$$\sup_{\nu\in\mathcal{K}} \boldsymbol{\Phi}(\boldsymbol{\nu}) - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) = \boldsymbol{\Phi}(\boldsymbol{\nu}^{\bar{q},\bar{m}}) - H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma)$$

where $\boldsymbol{\nu}^{\bar{q},\bar{m}}$ *is the generalized Gibbs measure with Radon-Nikodym derivative w.r.t.* $\mu\otimes\gamma$ *given by* (31) *for* $(q,m) = (\overline{q},\overline{m})$.

Proof. Consider $\bar{\boldsymbol{\nu}} \in \mathcal{K}$ solution of the Boltzmann-Gibbs principle (26). Then, let $(\overline{q},\overline{m})$ be a minimum point of P on $[0,1]^2$. Lemma (4.4), specifically (64), ensures that $\boldsymbol{\nu}^{\bar{q},\bar{m}} \in K$; this amounts to say that $\boldsymbol{\nu}^{\bar{q},\bar{m}}$ is a viable candidate to solve the Boltzmann-Gibbs principle.

We now use Proposition A.3 from the Appendix A, recalling the variational expression for relative entropy functionals, to show that

$$\boldsymbol{\Phi}(\boldsymbol{\nu}) - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) \leq \boldsymbol{\Phi}(\boldsymbol{\nu}^{\bar{q},\bar{m}}) - H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma) \quad \forall \boldsymbol{\nu} \in K, \tag{70}$$

which proves the Proposition. Specifically, as $g, f_1, f_2 \in C_b(S)$, we can apply A.3 to $\overline{m}(\Phi'(\overline{q})g^2 \circ \pi_1 + f) \in C_b(S^2)$ to see that

$$H(\boldsymbol{\nu} \mid \mu\otimes\gamma) \geq H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma) + \overline{m} \int \left(f + \Phi'(\overline{q})g^2 \circ \pi_1\right) d(\boldsymbol{\nu} - \boldsymbol{\nu}^{\bar{q},\bar{m}}) \tag{71}$$

for all $\boldsymbol{\nu} \in \mathcal{M}_1^+(S^2)$. Specifically

$$\begin{aligned}
\mathsf{E}_{\boldsymbol{\nu}}(f) - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) &\leq \mathsf{E}_{\boldsymbol{\nu}}(f) - H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma) - \overline{m} \int \left(f + \Phi'(\overline{q})g^2 \circ \pi_1\right) d(\boldsymbol{\nu} - \boldsymbol{\nu}^{\bar{q},\bar{m}}) \\
&= \mathsf{E}_{\boldsymbol{\nu}^{\bar{q},\bar{m}}}(f) - H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma) - \int \Phi'(\overline{q})g^2 \circ \pi_1 d(\boldsymbol{\nu} - \boldsymbol{\nu}^{\bar{q},\bar{m}}) + \\
&\quad + (1-\overline{m}) \int \left(f + \Phi'(\overline{q})g^2 \circ \pi_1\right) d(\boldsymbol{\nu} - \boldsymbol{\nu}^{\bar{q},\bar{m}}) \\
&= -\Phi'(\overline{q}) \left(\mathsf{E}_{\boldsymbol{\nu}}(g \circ \pi_1) - \bar{q} \right) + \mathsf{E}_{\boldsymbol{\nu}^{\bar{q},\bar{m}}}(f) - H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma) + \\
&\quad + (1-\overline{m}) \int \left(f + \Phi'(\overline{q})g^2 \circ \pi_1\right) d(\boldsymbol{\nu} - \boldsymbol{\nu}^{\bar{q},\bar{m}})
\end{aligned} \tag{72}$$

where in the last line we used that, as shown in Lemma 4.4, for a minimum point $(\bar{q}, \bar{m})$ of P it holds

$$\overline{q} = \mathsf{E}_{\boldsymbol{\nu}^{\bar{q},\bar{m}}}\left(g^2 \circ \pi_1\right). \tag{73}$$

Since Φ is concave and differentiable, it satisfies $\Phi(x)-\Phi'(y)(x-y) \leq \Phi(y)$ for all $x, y \in \mathbb{R}$. Through (72), and again (73), this implies

$$\begin{aligned}
\boldsymbol{\Phi}(\boldsymbol{\nu}) - H(\boldsymbol{\nu} \mid \mu\otimes\gamma) &= \Phi\left(\mathsf{E}_{\boldsymbol{\nu}}\left(g \circ \pi_1\right)\right) + \mathsf{E}_{\boldsymbol{\nu}}\left(f\right) - H(\boldsymbol{\nu} \mid \mu\otimes\gamma). \\
&\leq \boldsymbol{\Phi}(\boldsymbol{\nu}^{\bar{q},\bar{m}}) - H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma) + (1-\overline{m}) \int \left(f + \Phi'(\overline{q})g^2 \circ \pi_1\right) d(\boldsymbol{\nu} - \boldsymbol{\nu}^{\bar{q},\bar{m}}).
\end{aligned} \tag{74}$$

From the latter we see that if $\overline{m} = 1$ the claim (70) follows immediately. If instead $\overline{m} \in (0,1)$ then $H(\boldsymbol{\nu}^{\bar{q},\bar{m}} \mid \mu\otimes\gamma) = \log 2$ and Proposition A.3 implies that

$$\overline{m} \int \left(f + \Phi'(\overline{q})g^2 \circ \pi_1\right) d(\boldsymbol{\nu} - \boldsymbol{\nu}^{\bar{q},\bar{m}}) \leq H(\boldsymbol{\nu} \mid \mu\otimes\gamma) - \log 2 \leq 0$$

for all $\boldsymbol{\nu} \in K$. Since both $\overline{m}$ and $1-\overline{m}$ are non-negative this implies

$$(1-\overline{m}) \int \left(f + \Phi'(\overline{q})g^2 \circ \pi_1\right) d(\boldsymbol{\nu} - \boldsymbol{\nu}^{\bar{q},\bar{m}}) \leq 0$$

which used on (74) gives (70), which is therefore now proved also for $\overline{m} \in (0,1)$.
All in all, we have shown that if $(\overline{q}, \overline{m})$ is a minimum point of P on $[0,1]^2$ then the unique extremal measure $\bar{\boldsymbol{\nu}}$ of the Boltzmann-Gibbs principle (26) must be one of the $\boldsymbol{\nu}^{\bar{q},\bar{m}}$, namely the thesis of the Proposition is settled. □

4.3. **The limiting Gibbs measure: proof of Theorem 3.4.** Here and henceforth, we will denote by $\bar{\boldsymbol{\nu}} \in \mathcal{M}_1^+(S^2)$ the measure solving the Boltzmann-Gibbs variational principle for a system in low temperature; i.e.

$$d\bar{\boldsymbol{\nu}}(x,y) \stackrel{\text{def}}{=} \frac{\exp \bar{m}\left[\Phi'(\bar{q})\, g^2(x) + f(x,y)\right]}{Z^{\bar{q},\bar{m}}} d\mu(x) d\gamma(y)\,,$$

where $(\bar{q},\bar{m}) \stackrel{\text{def}}{=} \text{argmin}_{[0,1]^2} P$ and $\bar{m} < 1$. Some notation: let $\{(Z_i, W_i)\}_{i \leq N}$ be i.i.d. S^2-valued random vectors with common distribution $\bar{\boldsymbol{\nu}}$ and defined on a probability space $(\Omega', \mathcal{F}', \mathsf{E})$. Setting $\boldsymbol{Z} \stackrel{\text{def}}{=} (Z_1, W_1, \cdots, Z_N, W_N) \in S^{2N}$ define

$$X_N(\mathbf{Z}) \stackrel{\text{def}}{=} \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(g^2(Z_i) - \bar{q}\right), \qquad Y_N(\mathbf{Z}) \stackrel{\text{def}}{=} \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(f(Z_i, W_i) - \mathsf{E}_{\bar{\boldsymbol{\nu}}}(f)\right)$$

and consider the real valued random vectors

$$\begin{pmatrix} X_N \\ Y_N \end{pmatrix} \equiv \begin{pmatrix} X_N(\boldsymbol{Z}) \\ Y_N(\boldsymbol{Z}) \end{pmatrix}, \tag{75}$$

together with their covariance matrix, say Σ, which we assume to be invertible.

Set

$$C_{\bar{q},\bar{m}} \stackrel{\text{def}}{=} -\bar{m}\Phi''(\bar{q}) + (1, -\Phi'(\bar{q})) \cdot \Sigma^{-1} \cdot (1, -\Phi'(\bar{q}))^\top, \tag{76}$$

where $(1, -\Phi'(\bar{q}))^\top$ is the transpose of the vector $(1, -\Phi'(\bar{q})) \in \mathbb{R}^2$ and "$\cdot$" the matrix-vector product.
Notice that $\Phi''(\bar{q}) < 0$ implies $C \equiv C_{\bar{q},\bar{m}} > 0$.

Proposition 4.6. *Under the assumptions of Theorem 3.4, the point process*

$$\Xi_N \stackrel{\text{def}}{=} \sum_{\alpha=1}^{2^N} \delta_{H_N(\alpha) - N[\Phi(\bar{q}) - \mathsf{E}_{\bar{\boldsymbol{\nu}}}(f)] - \omega_N}, \tag{77}$$

where

$$\omega_N \stackrel{\text{def}}{=} -\frac{1}{\bar{m}} \log \sqrt{2\pi N |\Sigma| C}; \tag{78}$$

converges weakly to a Poisson point process with intensity measure $e^{-\bar{m}z}dz$.

Theorem 3.4 follows from

i) Proposition 4.6 together with
ii) the exponential transform

$$H_N(\alpha) - N\left[\Phi(\bar{q}) - \mathsf{E}_{\bar{\boldsymbol{\nu}}}(f)\right] - \omega_N \mapsto \exp\left(H_N(\alpha) - N\left[\Phi(\bar{q}) - \mathsf{E}_{\bar{\boldsymbol{\nu}}}(f)\right] - \omega_N\right),$$

which maps the Poisson point process with intensity measure $e^{-\bar{m}z}dz$ to a Poisson point process on the positive line with intensity measure $t^{-\bar{m}-1}dt$;
iii) the fact that infinite volume limit and the Gibbs-normalization commute.

Items *ii-iii)* are fairly standard in the literature: their proof is omitted, but we refer the reader to, say, [6] for details.

For the sake of simplicity, we shall prove Proposition 4.6 assuming that Theorem B.1 (see Appendix B and [3, Theorem 19.5]) holds for the normalized vectors (75); this is equivalent to the assumption that at least one measure among $\bar{\boldsymbol{\nu}} \circ (g \circ \pi_1 - \bar{q})^{-1}$ and $\bar{\boldsymbol{\nu}} \circ (f - \mathsf{E}_{\bar{\boldsymbol{\nu}}}(f))^{-1}$ has a density w.r.t. the Lebesgue measure on $\mathbb{R}$.

Lemma 4.7. *In low temperature, for any $(\mu\otimes\gamma)^{\otimes N}-$measurable function $F : S^{2N} \to \mathbb{R}$*

$$2^N \mathbb{E}\left(F\left(X_{1,1}, Y_{1,1}, \ldots, X_{1,N}, Y_{1,N}\right)\right) = \mathsf{E}\left(e^{-\bar{m}\sqrt{N}(\Phi'(\bar{q})X_N + Y_N)}\, F(\boldsymbol{Z})\right). \tag{79}$$

Proof. By definition, for any $\mathbf{y} = (y_1, \cdots, y_N) \in S^{2N}, y_i \in S^2$, it holds:

$$\frac{d(\mu\otimes\gamma)^{\otimes N}}{d\bar{\boldsymbol{\nu}}^{\otimes N}}(\mathbf{y}) = \prod_{i=1}^{N} \frac{d(\mu\otimes\gamma)}{d\bar{\boldsymbol{\nu}}}(y_i) = \exp\left(N \log Z^{\bar{q},\bar{m}} - \bar{m}\sum_{i=1}^{N} \Phi'(\bar{q})\, g^2 \circ \pi_1(y_i) + f(y_i)\right), \tag{80}$$

hence

$$2^N \mathbb{E}\left(F\left(X_{1,1}, Y_{1,1}, \ldots, X_{1,N}, Y_{1,N}\right)\right) = e^{N(\log Z^{\bar{q},\bar{m}} + \log 2)} \int e^{-\bar{m}\sum_{i=1}^{N} \Phi'(\bar{q}) g^2 \circ \pi_1(y_i) + f(y_i)} F(\mathbf{y}) d\bar{\boldsymbol{\nu}}^{\otimes N}(\mathbf{y}). \tag{81}$$

By the entropy condition (34),

$$\log Z^{\bar{q},\bar{m}} + \log 2 = \bar{m}\left[\Phi'(\bar{q})\bar{q} + \mathsf{E}_{\bar{\boldsymbol{\nu}}}(f)\right]. \tag{82}$$

Plugging this in (81), and remembering the definition of the vectors (75), the Lemma follows straightforwardly. □

Proof of Proposition 4.6. We will show that for any compact $K \subset \mathbb{R}$

$$\lim_{N\to\infty} \mathbb{E}\left(\Xi_N(K)\right) = \int_K e^{-\bar{m}z} dz\,. \tag{83}$$

Due to the complete independence over the α-s, this suffices to prove the Lemma by Kallenberg's theorem [16, Theorem 4.15]. To this aim, recall that, by definition,

$$H_N(1) = N\Phi\left(\frac{1}{N}\sum_{i=1}^{N} g^2(X_{1,i})\right) + \sum_{i=1}^{N} f(X_{1,i}, Y_{1,i}), \tag{84}$$

so that

$$\begin{aligned} \mathbb{E}\left[\Xi_N(K)\right] &= \mathbb{E}\left[\sum_{\alpha=1}^{2^N} \delta_{H_N(\alpha) - N(\Phi(\bar{q}) + \mathsf{E}_{\bar{\boldsymbol{\nu}}}(f)) - \omega_N}(K)\right] \\ &= 2^N \mathbb{P}\left[H_N(1) - N(\Phi(\bar{q}) + \mathsf{E}_{\bar{\boldsymbol{\nu}}}(f)) - \omega_N \in K\right] \\ &= 2^N \mathbb{E}\left[\mathbf{1}_{H_N(1) - N(\Phi(\bar{q}) + \mathsf{E}_{\bar{\boldsymbol{\nu}}}(f)) - \omega_N \in K}\right]. \end{aligned} \tag{85}$$

By Lemma 4.7,

$$\mathbb{E}\left[\Xi_N(K)\right] = \mathsf{E}\left[e^{-\bar{m}\sqrt{N}(\Phi'(\bar{q})X_N + Y_N)} \mathbf{1}_{N\left[\Phi\left(\frac{X_N}{\sqrt{N}} + \bar{q}\right) - \Phi(\bar{q})\right] + \sqrt{N}Y_N - \omega_N \in K}\right]. \tag{86}$$

As Φ is twice differentiable, for any fixed $N \in \mathbb{N}$ we can consider a map

$$x \in I_N \stackrel{\text{def}}{=} \left[-\sqrt{N},\ \sqrt{N}\right] \to \xi_N(x) \in I \stackrel{\text{def}}{=} [\bar{q}-1,\ \bar{q}+1], \tag{87}$$

that to any $x \in I_N$ associates a point $\xi_N(x) \in I$ such that

$$\begin{aligned} \Phi\left(\frac{x}{\sqrt{N}}+\bar{q}\right)-\Phi(\bar{q})-\Phi'(\bar{q})\frac{x}{\sqrt{N}} &= \frac{1}{2}\Phi''(\xi_N(x))\frac{x^2}{N} \\ &= R_N(x)\frac{x^2}{N}, \end{aligned} \tag{88}$$

where to lighten notation we defined

$$R_N(x) \stackrel{\text{def}}{=} \frac{1}{2}\Phi''(\xi_N(x)). \tag{89}$$

As $\bar{q}, g^2(x) \in [0,1]$ $\forall x \in S$, it holds

$$X_N \in I_N \text{ for any realization of } X_N; \tag{90}$$

specifically we can write

$$N\left[\Phi\left(\frac{X_N}{\sqrt{N}}+\bar{q}\right)-\Phi(\bar{q})\right] \equiv \sqrt{N}\Phi'(\bar{q})X_N + R_N(X_N)X_N^2. \tag{91}$$

Notice also (recalling that, by assumption, $\Phi''(a) < 0$ $\forall a \in \mathbb{R}$):

$$R_N(x) < 0 \quad \forall x \in I_N, \tag{92}$$

$$\lim_{N\to\infty} R_N(x) = \frac{1}{2}\Phi''(\bar{q}) < 0 \quad \text{uniformly for } x \in o(\sqrt{N}), \tag{93}$$

$$\text{if} \qquad R \stackrel{\text{def}}{=} \frac{1}{2}\inf_I |\Phi''| > 0 \qquad \text{then} \qquad |R_N(x)| \geq R \quad \forall x \in I_N. \tag{94}$$

Going back to (86) we rewrite it as

$$\mathbb{E}\left[\Xi_N(K)\right] = \int_{I_N\times\mathbb{R}} e^{-\bar{m}\sqrt{N}(\Phi'(\bar{q})x+y)}\Psi_N(x,y)dQ_N(x,y) \tag{95}$$

where $Q_N \in \mathcal{M}_1^+(\mathbb{R}^2)$ is the distribution of (X_N, Y_N) and

$$\Psi_N(x,y) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if} \quad \sqrt{N}(\Phi'(\bar{q})x+y)+R_N(x)x^2-\omega_N \in K, \\ 0 & \text{otherwise.} \end{cases} \tag{96}$$

It is easily seen that

$$\frac{1}{\sqrt{N}}\int_{I_N\times\mathbb{R}} e^{-\bar{m}\sqrt{N}(\Phi'(\bar{q})x+y)}\Psi_N(x,y)\ dxdy \tag{97}$$

vanishes as $N \to \infty$. Indeed, integrating by substitution according to the change of variables

$$(\tilde{x},\ \tilde{y}) = \left(x,\ \sqrt{N}(\Phi'(\bar{q})x+y)+R_N(x)x^2-\omega_N\right) \tag{98}$$

the function $\Psi_N(x,y)$ becomes $\mathbf{1}_K(\tilde{y})$ so that through (92), (93) we easily obtain

$$(97) \leq \frac{e^{-\bar{m}\omega_N}}{N}\int_K e^{-\bar{m}y'}dy'\int_{\mathbb{R}} e^{-\bar{m}Rx'^2}dx' = O\left(\frac{1}{\sqrt{N}}\right) \qquad (N\uparrow\infty), \tag{99}$$

where we also used the fact that $z\to e^{-\bar{m}z}$ is non-increasing and that the definition of ω_N (78) implies $e^{-\bar{m}\omega_N}/N = O\left(N^{-1/2}\right)$. But this, through (95) and Theorem B.1, yields

$$\lim_{N\to\infty}\mathbb{E}\left(\Xi_N(K)\right) = \lim_{N\to\infty}\int_{I_N\times\mathbb{R}} e^{-\bar{m}\sqrt{N}(\Phi'(\bar{q})x+y)}\Psi(x,y)\,\varphi_\Sigma(x,y)dxdy \tag{100}$$

where φ_Σ is the Gaussian bivariate density with mean $(0,0)$ and covariance matrix Σ.

We nof focus on the right hand side of (100) and write it as

$$\int_{I_N\times\mathbb{R}} e^{-\bar{m}\sqrt{N}(\Phi'(\bar{q})x+y)}\Psi_N(x,y)\,\varphi_\Sigma(x,y)dxdy = \mathcal{J}_N^1+\mathcal{J}_N^2, \tag{101}$$

where

$$\mathcal{J}_N^1 \stackrel{\text{def}}{=} \int\int_{-\log N}^{\log N} e^{-\bar{m}\sqrt{N}(\Phi'(\bar{q})x+y)}\Psi_N(x,y)\,\varphi_\Sigma(x,y)\,dx\,dy, \tag{102}$$

and

$$\mathcal{J}_N^2 \stackrel{\text{def}}{=} \int\int_{[-\log N,\log N]^c\cap I_N} e^{-\bar{m}\sqrt{N}(\Phi'(\bar{q})x+y)}\Psi_N(x,y)\,\varphi_\Sigma(x,y)\,dx\,dy. \tag{103}$$

We begin by showing that

$$\lim_{N\to\infty}\mathcal{J}_N^2 = 0\,. \tag{104}$$

Indeed, by K-compactness it holds that $K\subset[t,+\infty)$ for some $t\in\mathbb{R}$, so that for every $(x,y):\ \Psi_N(x,y)\neq 0$ the exponential term in (103) is bounded above by $e^{-\bar{m}\left(-R_N(x)x^2+\omega_N+t\right)}$. Specifically, again by virtue of (92), (94), we obtain

$$\begin{aligned}\mathcal{J}_N^2 &\leq e^{-\bar{m}(\omega_N+t)}\int_{[-\log N,\log N]^c} e^{\bar{m}R_N(x)x^2}\varphi_\Sigma^1(x)dx\\ &\leq e^{-\bar{m}\left(R\log^2 N+\omega_N+t\right)} \in o(1) \quad \text{as} \quad N\to\infty,\end{aligned} \tag{105}$$

where $x\to\varphi_\Sigma^1(x)$ is the Gaussian univariate density of the first marginal of a bivariate Gaussian with density $(x,y)\to\varphi_\Sigma(x,y)$. This settles the claim (104).

As for $\mathcal{J}_N^1$, we have

$$\mathcal{J}_N^1 = \frac{1}{2\pi\sqrt{|\Sigma|}}\int_{-\log N}^{\log N}\int e^{-m\sqrt{N}(\Phi'(\bar{q})x+y)-\frac{1}{2}\bar{\boldsymbol{v}}\cdot\Sigma^{-1}\cdot\bar{\boldsymbol{v}}^\top}\Psi_N(\bar{\boldsymbol{v}})\;dy\,dx, \tag{106}$$

with $\bar{\boldsymbol{v}}^\top$ denoting transpose. Note that with $(\tilde{x},\,\tilde{y})$ the variables as in (98), then

$$\begin{aligned}\bar{\boldsymbol{v}} &\equiv \left(\tilde{x},\,\frac{\tilde{y}-R_N(\tilde{x})\tilde{x}^2+\omega_N}{\sqrt{N}}-\Phi'(\bar{q})\tilde{x}\right)\\ &= \tilde{x}\cdot(1,\,-\Phi'(\bar{q})) + \frac{\tilde{y}-R_N(\tilde{x})\tilde{x}^2+\omega_N}{\sqrt{N}}\cdot(0,1)\end{aligned}$$

and that for any scalars $a, b \in \mathbb{R}$ and vectors $\bar{\boldsymbol{v}}_1, \bar{\boldsymbol{v}}_2 \in \mathbb{R}^2$ it trivially holds

$$-\frac{1}{2}(a\bar{\boldsymbol{v}}_1 + b\bar{\boldsymbol{v}}_2) \cdot \Sigma^{-1} \cdot \left(a\bar{\boldsymbol{v}}_1^\top + b\bar{\boldsymbol{v}}_2^\top\right) = -\frac{1}{2}a^2\bar{\boldsymbol{v}}_1 \cdot \Sigma^{-1} \cdot \bar{\boldsymbol{v}}_1^\top - ba\bar{\boldsymbol{v}}_1 \cdot \Sigma^{-1} \cdot \bar{\boldsymbol{v}}_2^\top + \\ -\frac{1}{2}b^2\bar{\boldsymbol{v}}_2 \cdot \Sigma^{-1} \cdot \bar{\boldsymbol{v}}_2^\top. \tag{107}$$

Specifically, as

$$\tilde{y} \in K, \ |\tilde{x}| \leq \log N \quad \Rightarrow \quad \frac{\tilde{y} - R_N(\tilde{x})\tilde{x}^2 + \omega_N}{\sqrt{N}} \in O\left(\frac{\log^2 N}{\sqrt{N}}\right) \quad (N \to \infty) \tag{108}$$

uniformly, and in light of (107), we get that the quadratic exponent of the Gaussian density in the new variables $(\tilde{x}, \ \tilde{y})$ equals

$$-\frac{1}{2}\tilde{x}^2 \left(1, \, -\Phi'(\bar{q})\right) \cdot \Sigma^{-1} \cdot \left(1, \, -\Phi'(\bar{q})\right)^\top + O\left(\frac{\log^3 N}{\sqrt{N}}\right) \quad (N \to \infty). \tag{109}$$

Therefore

$$\mathcal{J}_N^1 = \frac{e^{-\bar{m}\omega_N}}{2\pi\sqrt{|\Sigma| N}} \int_K e^{-\bar{m}\tilde{y}} \int_{-\log N}^{\log N} e^{-\tilde{x}^2\left\{-\bar{m}R_N(\tilde{x}) + \frac{1}{2}(1, -\Phi'(\bar{q}))\cdot\Sigma^{-1}\cdot(1, -\Phi'(\bar{q}))^\top\right\} + o(1)} \, d\tilde{x} \, d\tilde{y}. \tag{110}$$

As (94) holds, we have

$$\lim_{N\to\infty} \sup_{x\in\mathbb{R}: \ |x|\leq \log N} R_N(x) = \frac{1}{2}\Phi''(\bar{q}), \tag{111}$$

and by definition

$$\frac{e^{-\bar{m}\omega_N}}{2\pi\sqrt{|\Sigma| N}} = \sqrt{\frac{C}{2\pi}}. \tag{112}$$

All in all, we get

$$\lim_{N\to\infty} \mathcal{J}_N^1 = \sqrt{\frac{C}{2\pi}} \int_K e^{-\bar{m}\tilde{y}} \int e^{-\frac{C}{2}\tilde{x}^2} \, d\tilde{x} \, d\tilde{y} = \int_K e^{-\bar{m}\tilde{y}} d\tilde{y} \tag{113}$$

which ends the proof.

□

4.4. **The limiting law of the overlap: proof of Proposition 3.5.** In this subsection we prove Proposition 3.5, i.e. we show that if the system is in low temperature and $\bar{\boldsymbol{\nu}}$ is the solution of the Boltzmann-Gibbs principle (26), then the limits (36), (37) hold.
First, we prove (36). We start with a technical Lemma, which is a direct consequence of Lemma 4.2 and Theorem 3.2.

Lemma 4.8. *For every $r > 0$ and $O_N = O_N(\alpha, \alpha')$ s.t. for some $M > 0$: $|O_N(\alpha, \alpha')| \leq M$ for every $\alpha, \alpha' \in \{1, \cdots, 2^N\}$, $N \in \mathbb{N}$; it holds*

$$\lim_{N\to\infty} \left\langle \mathbf{1}_{B_r^c}[\boldsymbol{L}_{N,\alpha}] O_N(\alpha, \alpha') \right\rangle_N = 0 \quad \mathbb{P} - a.s.$$

where $B_r^c = \{\boldsymbol{\rho} \in \mathcal{M}_1^+(S^2) : \quad d(\boldsymbol{\rho}, \bar{\boldsymbol{\nu}}) \geq r\} = \mathcal{M}_1^+(S^2) \setminus B_{\bar{\boldsymbol{\nu}},r}$.

Proof. Fix $r > 0$, then

$$\begin{aligned}\left\langle \mathbf{1}_{B_r^c}[\boldsymbol{L}_{N,\alpha}] \, |O_N(\alpha,\alpha')| \right\rangle_N &\leq M \left\langle \mathbf{1}_{B_r^c}[\boldsymbol{L}_{N,\alpha}] \right\rangle_N \\ &= M \sum_{\alpha:\boldsymbol{L}_{N,\alpha}\notin B_r} \mathcal{G}_N(\alpha) \sum_{\alpha'\leq 2^N} \mathcal{G}_N(\alpha') \\ &= M \sum_{\alpha:\boldsymbol{L}_{N,\alpha}\notin B_r} \mathcal{G}_N(\alpha).\end{aligned}$$

Therefore, showing that $\mathbb{P}-$a.s.

$$\lim_{N\to\infty} \sum_{\alpha:\boldsymbol{L}_{N,\alpha}\notin B_r} \mathcal{G}_N(\alpha) = \lim_{N\to\infty} \frac{\sum\limits_{\alpha:\boldsymbol{L}_{N,\alpha}\notin B_r} e^{N\boldsymbol{\Phi}[L_{N,\alpha}]}}{Z_N} = 0, \tag{114}$$

would prove the Lemma. We therefore claim (114).
Lemma 4.2 and Theorem 3.2 imply that $\mathbb{P}-$a.s. it holds

$$F_r \stackrel{\text{def}}{=} \limsup_{N\to\infty} \frac{1}{N} \log \sum_{\alpha:\boldsymbol{L}_{N,\alpha}\notin B_r} e^{N\boldsymbol{\Phi}[L_{N,\alpha}]} < \lim_{N\to\infty} \frac{1}{N} \log Z_N = F.$$

Specifically, for $\eta \stackrel{\text{def}}{=} F - F_r > 0$, it holds

$$Z_N > e^{N\left(F-\frac{\eta}{4}\right)}, \qquad \sum_{\alpha:\boldsymbol{L}_{N,\alpha}\notin B_r} e^{N\boldsymbol{\Phi}[L_{N,\alpha}]} < e^{N\left(F_r+\frac{\eta}{4}\right)},$$

$\mathbb{P}-$a.s and provided that N is large. As the latter implies

$$\sum_{\alpha:\boldsymbol{L}_{N,\alpha}\notin B_r} \mathcal{G}_N(\alpha) = \frac{\sum\limits_{\alpha:\boldsymbol{L}_{N,\alpha}\notin B_r} e^{N\boldsymbol{\Phi}[L_{N,\alpha}]}}{Z_N} \leq e^{-\frac{\eta}{2}N},$$

the claim (114) is settled and the Lemma follows. □

Notice that the non-negative operator $T : \mathcal{M}_1^+(S^2) \to \mathbb{R}$ defined as

$$T[\boldsymbol{\nu}] \stackrel{\text{def}}{=} \left(\int g^2(x)\nu_1(dx) - \bar{q} \right)^2$$

is continous, bounded and such that (recall $\bar{q} = \int g^2(x)\bar{\nu}_1$)

$$T\left[\boldsymbol{L}_{N,\alpha}\right] = \left(q_N(\alpha,\alpha) - \bar{q}\right)^2, \quad T\left[\bar{\boldsymbol{\nu}}\right] = 0.$$

Specifically, given an arbitrary $\delta > 0$ we can find $r > 0$ such that if $\boldsymbol{L}_{N,\alpha} \in B_r \equiv B_{\boldsymbol{\nu},r}$, then $T[\boldsymbol{L}_{N,\alpha}] < \delta$. This implies

$$\begin{aligned}\mathbb{E}\left\langle \mathbf{1}_{B_\epsilon}[\boldsymbol{L}_{N,\alpha}] \left(q_N(\alpha,\alpha') - \bar{q}\right)^2 \delta_{\alpha=\alpha'} \right\rangle_N &= \mathbb{E}\left\langle \mathbf{1}_{B_\epsilon}[\boldsymbol{L}_{N,\alpha}] T\left[\boldsymbol{L}_{N,\alpha}\right] \delta_{\alpha=\alpha'} \right\rangle_N \\ &= \mathbb{E} \sum_{\alpha:\boldsymbol{L}_{N,\alpha}\in B_\epsilon} T[\boldsymbol{L}_{N,\alpha}] \mathcal{G}_N(\alpha)^2 \leq \delta\end{aligned} \tag{115}$$

where in the last line we used that in low temperature Theorem 3.4 guarantees that the process $\{\mathcal{G}_N(\alpha)\}_{\alpha\leq 2^N}$ converges to a Poisson-Dirichlet point process with parameter

$\bar{m} \in (0,1)$, and this implies $\mathbb{E}\sum_{\alpha\leq 2^N}\mathcal{G}_N(\alpha)^2 \sim 1-\bar{m} \in (0,1)$. Specifically, applying Lemma 4.8 we get

$$\lim_{N\to\infty}\mathbb{E}\left\langle\, (q_N(\alpha,\alpha')-\bar{q})^2\,\delta_{\alpha=\alpha'}\,\right\rangle_N = \mathbb{E}\left\langle\, \mathbf{1}_{B_\epsilon}[\boldsymbol{L}_{N,\alpha}]\,(q_N(\alpha,\alpha')-\bar{q})^2\,\delta_{\alpha=\alpha'}\,\right\rangle_N \leq \delta,$$

which, being δ arbitrarily small, proves the first claim (36) of Proposition 3.5.

In order to conclude the proof of the Proposition, we only need to show (37).

To this aim, define

$$\tilde{\boldsymbol{L}}_{N,\alpha} \stackrel{\text{def}}{=} \frac{1}{N-2}\sum_{i=3}^{N}\delta_{(X_{\alpha,i},Y_{\alpha,i})},$$

and let $\tilde{L}^1_{N,\alpha}, \tilde{L}^2_{N,\alpha}$ be its marginals. Then, similarly to the proof of 4.6, for any fixed $N\in\mathbb{N}$ consider a map $\alpha\to\zeta_N^\alpha$ that to any configuration α associates a point $\zeta_N^\alpha\in[0,1]$ (specifically in the interval between $\mathsf{E}_{L^1_{N,\alpha}}(g^2)$ and $\mathsf{E}_{\tilde{L}^1_{N,\alpha}}(g^2)$) such that

$$\Phi\left(\mathsf{E}_{L^1_{N,\alpha}}(g^2)\right) = \Phi\left(\mathsf{E}_{\tilde{L}^1_{N,\alpha}}(g^2)\right) + \Phi'\left(\zeta_N^\alpha\right)\left[\mathsf{E}_{L^1_{N,\alpha}}(g^2) - \mathsf{E}_{\tilde{L}^1_{N,\alpha}}(g^2)\right]. \tag{116}$$

By construction, one has

$$H_N(\alpha) = N\boldsymbol{\Phi}\left[\boldsymbol{L}_{N,\alpha}\right] = (N-2)\boldsymbol{\Phi}\left[\tilde{\boldsymbol{L}}_{N,\alpha}\right] + W_N(\alpha) + R_N(\alpha) \tag{117}$$

where

$$\begin{aligned} W_N(\alpha) &\stackrel{\text{def}}{=} \Phi'(\zeta_N^\alpha)\left[g^2(X_{\alpha,1}) + g^2(X_{\alpha,2})\right] + f(X_{\alpha,1},Y_{\alpha,1}) + f(X_{\alpha,2},Y_{\alpha,2})\,; \\ R_N(\alpha) &\stackrel{\text{def}}{=} 2\Phi\left(\mathsf{E}_{L^1_{N,\alpha}}(g^2)\right) - 2\Phi'\left(\zeta_N^\alpha\right)\mathsf{E}_{L^1_{N,\alpha}}(g^2)\,. \end{aligned}$$

This implies that for every map $(\alpha,\alpha')\to O_N(\alpha,\alpha')$

$$\mathbb{E}\langle O_N(\alpha,\alpha')\rangle_N = \mathbb{E}\frac{\langle\langle O_N(\alpha,\alpha')e^{W_N(\alpha)+W_N(\alpha')+R_N(\alpha)+R_N(\alpha')}\rangle\rangle_N}{\langle\langle e^{W_N(\alpha)+W_N(\alpha')+R_N(\alpha)+R_N(\alpha')}\rangle\rangle_N} \tag{118}$$

where we defined $\langle\langle\cdot\rangle\rangle_N \stackrel{\text{def}}{=} \tilde{Z}_N^{-2}\sum_{\alpha,\alpha'}\cdot\,\tilde{\mathcal{G}}_N(\alpha)\tilde{\mathcal{G}}_N(\alpha')$ and

$$\tilde{\mathcal{G}}_N(\alpha) \stackrel{\text{def}}{=} \frac{\exp{(N-2)\boldsymbol{\Phi}[\tilde{\boldsymbol{L}}_{N,\alpha}]}}{\tilde{Z}_N}, \quad \tilde{Z}_N \stackrel{\text{def}}{=} \sum_{\alpha=1}^{2^N} e^{(N-2)\boldsymbol{\Phi}[\tilde{\boldsymbol{L}}_{N,\alpha}]}.$$

As Φ is twice differentiable, for every $\delta>0$ there exists $\delta'>0$ such that, being $L^1_{N,\alpha}$ the first marginal of $\boldsymbol{L}_{N,\alpha}$, if

$$\left|\mathsf{E}_{L^1_{N,\alpha}}(g^2) - \bar{q}\right| \leq \delta', \quad |\zeta_N^\alpha - \bar{q}| < \delta' \tag{119}$$

then

$$\begin{cases} |R_N(\alpha) - 2\Phi(\bar{q}) + 2\Phi'(\bar{q})\bar{q}| < \delta, \\ |W_N(\alpha) - \Phi'(\bar{q})\left[g^2(X_{\alpha,1}) + g^2(X_{\alpha,2})\right] - f(X_{\alpha,1},Y_{\alpha,1}) - f(X_{\alpha,2},Y_{\alpha,2})| < \delta. \end{cases}$$

Moreover, by duality and continous projection, it is clear that we can choose r small enough s.t.

$$\tilde{\boldsymbol{L}}_{N,\alpha} \in B_{\bar{\boldsymbol{\nu}},r} \;\Rightarrow\; \mathsf{E}_{\tilde{L}^1_{N,\alpha}}(g^2) \in \left[\mathsf{E}_{\bar{\nu}_1}(g^2) - \delta,\, \mathsf{E}_{\bar{\nu}_1}(g^2) + \delta\right] \equiv \left[\bar{q} - \frac{\delta'}{2}, \bar{q} + \frac{\delta'}{2}\right]. \tag{120}$$

Notice also that

$$\mathsf{E}_{L^1_{N,\alpha}}(g^2) = \frac{N-2}{N}\left[\frac{g^2(X_{\alpha,1}) + g^2(X_{\alpha,2})}{N-2} + \mathsf{E}_{\tilde{L}^1_{N,\alpha}}(g^2)\right] \tag{121}$$

which implies that as $N \to \infty$, $\mathsf{E}_{L^1_{N,\alpha}}(g^2) = \zeta_N^\alpha = \mathsf{E}_{\tilde{L}^1_{N,\alpha}}(g^2) + O(N^{-1})$ uniformly.

It is readily checked that Lemma 4.2 and Theorem 3.2 still hold if we substitute the original Hamiltonian $N\boldsymbol{\Phi}[\boldsymbol{L}_{N,\alpha}]$ with $(N-2)\boldsymbol{\Phi}[\tilde{\boldsymbol{L}}_{N,\alpha}]$; this implies that Lemma 4.8 works also for the average $\langle\langle\cdot\rangle\rangle_N$. Specifically, for every $r > 0$, the couples α, α' contributing to the sums corresponding to the averages in the right hand side of (118) are the ones for which $\tilde{\boldsymbol{L}}_{N,\alpha}, \tilde{\boldsymbol{L}}_{N,\alpha'} \in B_{\bar{\boldsymbol{\nu}},r}$.

All in all, being δ arbitrary, we immediately get

$$\lim_{N\to\infty} \mathbb{E}\langle F_{\alpha,\alpha'}\rangle_N = \lim_{N\to\infty} \mathbb{E}\frac{\langle\langle\, F_{\alpha,\alpha'}\, V_\alpha V_{\alpha'}\, \rangle\rangle_N}{\langle\langle\, V_\alpha V_{\alpha'}\, \rangle\rangle_N} \tag{122}$$

where we defined

$$V_\alpha \stackrel{\text{def}}{=} \exp\sum_{k=1}^{2} \Phi'(\bar{q}) g^2\left(X_{\alpha,k}\right) + f(X_{\alpha,k}, Y_{\alpha,k}).$$

For $O_N(\alpha,\alpha') \equiv \delta_{\alpha\neq\alpha'}\left(q_N(\alpha,\alpha') - \bar{q}_0\right)^2$, the r.h.s. of (122) equals the large-N limit of

$$\mathbb{E}\frac{\langle\langle\, q_N(\alpha,\alpha')^2 \delta_{\alpha\neq\alpha'}\, V_\alpha V_{\alpha'}\, \rangle\rangle_N}{\langle\langle\, V_\alpha V_{\alpha'}\, \rangle\rangle_N} + \bar{q}_0^2 \mathbb{E}\frac{\langle\langle\, \delta_{\alpha\neq\alpha'}\, V_\alpha V_{\alpha'}\, \rangle\rangle_N}{\langle\langle\, V_\alpha V_{\alpha'}\, \rangle\rangle_N} - 2\bar{q}_0 \mathbb{E}\frac{\langle\langle\, q_N(\alpha,\alpha')\delta_{\alpha\neq\alpha'}\, V_\alpha V_{\alpha'}\, \rangle\rangle_N}{\langle\langle\, V_\alpha V_{\alpha'}\, \rangle\rangle_N}. \tag{123}$$

Since g is bounded, and by symmetry, the first term in (123) converges in the N-limit to

$$\begin{aligned} &\frac{1}{N^2}\sum_{i\neq j} \mathbb{E}\frac{\langle\langle\, g(X_{\alpha,i})g(X_{\alpha',i})g(X_{\alpha,j})g(X_{\alpha',j})\delta_{\alpha\neq\alpha'}\, V_\alpha V_{\alpha'}\, \rangle\rangle_N}{\langle\langle\, V_\alpha V_{\alpha'}\, \rangle\rangle_N} + O\left(\frac{1}{N}\right) = \\ &= \frac{N(N-1)}{N^2}\mathbb{E}\frac{\langle\langle\, g(X_{\alpha,1})g(X_{\alpha',1})g(X_{\alpha,2})g(X_{\alpha',2})\delta_{\alpha\neq\alpha'}\, V_\alpha V_{\alpha'}\, \rangle\rangle_N}{\langle\langle\, V_\alpha V_{\alpha'}\, \rangle\rangle_N} + O\left(\frac{1}{N}\right). \end{aligned} \tag{124}$$

Notice now that Theorem 3.4 clearly also applies to the collection of weights $\tilde{\mathcal{G}}_N(\alpha), \alpha = 1\ldots 2^N$, which therefore converges weakly to a Poisson-Dirichlet point process with parameter $\bar{m}$.

Some particularly useful properties of such process are given in Theorem B.2 of the Appendix. A simple domination argument shows that one can pass to the $N \to \infty$ limit replacing the $\{\tilde{\mathcal{G}}_N(\alpha)\}_\alpha$ by the points of its weak limit so that[4] we can use the formula

[4] also using that the random vectors $(V_\alpha,\, g(X_{\alpha,1})g(X_{\alpha,2})V_\alpha)$ are independent of the process $\{\tilde{\mathcal{G}}_N(\alpha)\}_\alpha$.

(131) from Theorem B.2 to get

$$
\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}\frac{\langle\langle\, q_N(\alpha,\alpha')^2 \delta_{\alpha\neq\alpha'}\, V_\alpha V_{\alpha'}\,\rangle\rangle_N}{\langle\langle\, V_\alpha V_{\alpha'}\,\rangle\rangle_N} \\
&\quad = \lim_{N\to\infty} \mathbb{E}\,\frac{\langle\langle\, g(X_{\alpha,1})g(X_{\alpha',1})g(X_{\alpha,2})g(X_{\alpha',2})\delta_{\alpha\neq\alpha'}\, V_\alpha V_{\alpha'}\,\rangle\rangle_N}{\langle\langle\, V_\alpha V_{\alpha'}\,\rangle\rangle_N} \\
&\quad = \lim_{N\to\infty} \mathbb{E}\,\frac{\sum\limits_{\alpha\neq\alpha'} g(X_{\alpha,1})g(X_{\alpha,2})V_\alpha\; g(X_{\alpha',1})g(X_{\alpha',2})V_{\alpha'}\; \tilde{\mathcal{G}}_N(\alpha)\tilde{\mathcal{G}}_N(\alpha')}{\left[\sum\limits_{\alpha} V_\alpha \tilde{\mathcal{G}}_N(\alpha)\right]^2} \qquad (125) \\
&\quad = \lim_{N\to\infty} \bar{m}\left[\frac{\mathbb{E}g(X_{1,1})g(X_{1,2})V_1 V_1^{\bar{m}-1}}{\mathbb{E}V_1^{\bar{m}}}\right]^2 = \bar{m}\left(\int g d\boldsymbol{\bar{\nu}}\right)^4 = \bar{m}\bar{q}_0^2.
\end{aligned}
$$

Similarly we get

$$
\lim_{N\to\infty} \mathbb{E}\frac{\langle\langle\, \delta_{\alpha\neq\alpha'}\, V_\alpha V_{\alpha'}\,\rangle\rangle_N}{\langle\langle\, V_\alpha V_{\alpha'}\,\rangle\rangle_N} = \bar{m}, \quad \lim_{N\to\infty} \mathbb{E}\frac{\langle\langle\, q_N(\alpha,\alpha')\delta_{\alpha\neq\alpha'}\, V_\alpha V_{\alpha'}\,\rangle\rangle_N}{\langle\langle\, V_\alpha V_{\alpha'}\,\rangle\rangle_N} = \bar{m}\bar{q}_0,
$$

and (37) follows. This ends the proof of Proposition 3.5.

Appendix A. The relative entropy.

Definition A.1. *Given a couple ν, μ of Borel probability measures on a Polish space S, their relative entropy is defined as*

$$
H(\nu \mid \mu) = \begin{cases} \mathsf{E}_\nu\left(\log\left(\frac{d\nu}{d\mu}\right)\right) & \text{if} \quad \nu << \mu, \quad \mathsf{E}_\nu\left(\left|\log\left(\frac{d\nu}{d\mu}\right)\right|\right) < \infty, \\ \infty & \text{else.} \end{cases} \qquad (126)
$$

It is a well-known fact that the Definition A.1 is equivalent to the following (see [12], [17] for details)

$$
H(\nu \mid \mu) = \sup_{u\,\in\, C_b(S)} \left\{ \int u d\nu - \log \int e^u d\mu \right\} \qquad (127)
$$

and it is easily seen that $\nu \to H(\nu \mid \mu)$ from $\mathcal{M}_1^+(S)$ to $\mathbb{R}$ is non-negative, convex and lower semicontinous. Specifically, the sublevel K defined in (25) is compact in $\mathcal{M}_1^+(S)$.

Definition A.2. *For all $u \in C_b(\mathbb{R})$ define the measure*

$$
G_u \in \mathcal{M}_1^+(S): \quad \frac{dG_u}{d\mu}(s) = \frac{e^{u(s)}}{Z_u}.
$$

Proposition A.3. *For all $u \in C_b(S)$, $\nu \in \mathcal{M}_1^+(S)$ it holds*

$$
H(\nu \mid \mu) \geq H(G_u \mid \mu) + \int u\, d(\nu - G_u).
$$

Proof. From the variational definition of H (127) we have

$$
H(\nu \mid \mu) \geq \int u d\nu - \log Z_u \quad \forall \nu \in \mathcal{M}_1^+(S), \forall u \in C_b(S).
$$

Moreover

$$H(G_u \mid \mu) = \int \log\left(\frac{dG_u}{d\mu}\right) dG_u = \int u dG_u - \log Z_u,$$

and the thesis follows straightforwardly. □

Appendix B. Edgeworth expansions, and Poisson-Dirichlet identities.

Denoting by $\bar{\boldsymbol{v}} = (x, y) \to \varphi_\Sigma(\bar{\boldsymbol{v}})$ the bivariate Gaussian density with mean zero and covariance matrix Σ, the following normal approximation result holds:

Theorem B.1. [[3], *Theorem* 19.5] *Let* $\{\mathbf{V}_i\}_{i\geq 1}$ *be a sequence of i.i.d. random vectors with values in* $\mathbb{R}^2$*, having mean zero, a positive-definite covariance matrix* Σ *and a nonzero, absolutely continous component. Then if* $\mathbb{E}\|\mathbf{V}_1\|^3 < \infty$*, writing* Q_N *for the distribution on of* $N^{-\frac{1}{2}}(\mathbf{V}_1 + \cdots + \mathbf{V}_N)$*, one has*

$$\int \left(1 + \|x\|^3\right) d\left|Q_N - \Upsilon_N\right|(x, y) = o\left(\frac{1}{\sqrt{N}}\right) \quad (N \to \infty) \tag{128}$$

where

$$\frac{d\Upsilon_N}{d\lambda_2}(\bar{\boldsymbol{v}}) = \frac{h(\bar{\boldsymbol{v}})}{\sqrt{N}} + \varphi_\Sigma(\bar{\boldsymbol{v}}) \tag{129}$$

for a bounded smooth function $h : \mathbb{R}^2 \to \mathbb{R}$*,* λ_2 *being the Lebesgue measure on* $\mathbb{R}^2$*.*

Theorem B.2. [[24], *Theorem* 6.4.5] *Assume that* $\{v_\alpha\}_{\alpha\leq 2^N}$ *is a Poisson-Dirichlet point process with intensity measure* $e^{-mt}dt$ *for some* $m < 1$*, independent of a sequence* $\{(U_\alpha, V_\alpha)\}_{\alpha\leq 2^N}$ *of i.i.d. vectors, copies of some* (U, V)*:* $\mathbb{E}U^2 < \infty, \mathbb{E}V^2 < \infty, V \geq 1$*. Then we have the formulas*

$$\mathbb{E}\frac{\sum_\alpha v_\alpha U_\alpha}{\sum_\alpha v_\alpha V_\alpha} = \frac{\mathbb{E}UV^{m-1}}{\mathbb{E}V^m}, \tag{130}$$

$$\mathbb{E}\frac{\sum_{\alpha\neq\alpha'} v_\alpha v_\beta U_\alpha U_\beta}{\left(\sum_\alpha v_\alpha V_\alpha\right)^2} = m\left(\frac{\mathbb{E}UV^{m-1}}{\mathbb{E}V^m}\right)^2, \tag{131}$$

$$\mathbb{E}\frac{\sum_\alpha v_\alpha^2 U_\alpha^2}{\left(v_\alpha V_\alpha\right)^2} = (1 - m)\frac{\mathbb{E}U^2 V^{m-2}}{\mathbb{E}V^m}. \tag{132}$$

References

[1] Arguin, Louis-Pierre, and Michael Aizenman. *On the structure of quasi-stationary competing particle systems.* The Annals of Probability 37.3 (2009): 1080-1113.

[2] Ruzmaikina, Anastasia, and Michael Aizenman. *Characterization of invariant measures at the leading edge for competing particle systems.* The Annals of Probability 33.1 (2005): 82-113.

[3] Bhattacharya, Rabi N., and R. Ranga Rao. *Normal approximation and asymptotic expansions.* Society for Industrial and Applied Mathematics, (2010).

[4] Bolthausen, Erwin, and Alain-Sol Sznitman. *On Ruelle's probability cascades and an abstract cavity method.* Communications in mathematical physics 197.2 (1998): 247-276.

[5] Bolthausen, Erwin, and Alain-Sol Sznitman. *Ten lectures on random media.* Vol. 32. Springer Science & Business Media, (2002).

[6] Bolthausen, Erwin, and Nicola Kistler. *Universal structures in some mean field spin glasses and an application.* Journal of mathematical physics 49.12 (2008): 125205.

[7] Bolthausen, Erwin, and Nicola Kistler. *A quenched large deviation principle and a Parisi formula for a Perceptron version of the GREM.* Probability in complex physical systems. Springer, Berlin, Heidelberg, (2012). 425-442.

[8] Chen, Wei-Kuo, Dmitry Panchenko, and Eliran Subag. *The generalized TAP free energy II.* Communications in Mathematical Physics 381.1 (2021): 257-291.

[9] Crisanti, Andrea, Luca Leuzzi, and Tommaso Rizzo. *Complexity in mean-field spin-glass models: Ising p-spin.* Physical Review B 71.9 (2005): 094202.

[10] Derrida, Bernard. *Random Energy Model: An exactly solvable model of disordered systems.*, Phys. Rev. B 24 (1981)

[11] Derrida, Bernard. *A generalization of the random energy model which includes correlations between energies.* Journal de Physique Lettres 46.9 (1985): 401-407.

[12] Donsker, Monroe D., and S.R. Srinivasa Varadhan. *Asymptotic evaluation of certain Markov process expectations for large time, I.* Communications on Pure and Applied Mathematics 28.1 (1975): 1-47.

[13] Feller, William. *An introduction to probability theory and its applications.* Vols. I & II, Wiley I 968 (1971).

[14] Guerra, Francesco. *B*roken replica symmetry bounds in the mean field spin glass model. Comm. in Math. Phys. Vol. 233 (2002).

[15] Gufler, Stephan, Adrien Schertzer, and Marius A. Schmidt. *On concavity of TAP free energy in the SK model.* arXiv:2209.08985 (2022).

[16] Kallenberg, Olav. (2017). *Random measures, theory and applications.* Springer International Publishing.

[17] Lee, Se Yoon. *Gibbs sampler and coordinate ascent variational inference: A set-theoretical review.* Communications in Statistics -Theory and Methods 51.6 (2022): 1549-1568.

[18] Mezard, Marc, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications.* Vol. 9. World Scientific Publishing Company, (1987).

[19] Plefka, Timm. *Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model.* Journal of Physics A: Mathematical and general 15.6 (1982): 1971.

[20] Polyanskiy, Yury, and Yihong Wu. *Lecture notes on information theory* (2019).

[21] Ruelle, David. *A mathematical reformulation of Derrida's REM and GREM.*, Comm.Math.Phys. 108 (1987).

[22] Sherrington, David, and Scott Kirkpatrick. *Solvable model of a spin-glass.* Physical review letters 35.26 (1975): 1792.

[23] Nishimori, Hidetoshi. *Statistical Physics of Spin Glasses and Information Processing: an Introduction.* Oxford; New York: Oxford University Press (2001).

[24] Talagrand, Michel. *Spin Glasses: A Challenge for Mathematicians. Cavity and Mean Field Models*, Springer Verlag (2003).

[25] Thouless, David J., Philip W. Anderson, and Robert G. Palmer. *Solution of 'solvable model of a spin glass'.* Philosophical Magazine 35.3 (1977): 593-601.

Nicola Kistler, J.W. Goethe-Universität Frankfurt, Germany.
Email address: kistler@math.uni-frankfurt.de

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

Giulia Sebastiani, J.W. Goethe-Universität Frankfurt, Germany.
Email address: sebastia@math.uni-frankfurt.de