ON A NONHIERARCHICAL GENERALIZATION OF THE PERCEPTRON GREM

NICOLA KISTLER AND GIULIA SEBASTIANI

Abstract. We introduce a nonlinear, nonhierarchical generalization of Derrida's GREM and establish through a Sanov-type large deviation analysis both a Boltzmann-Gibbs principle as well as a Parisi formula for the limiting free energy. In line with the predictions of the Parisi theory, the free energy is given by the minimal value over all Parisi functionals/hierarchical structures in which the original model can be coarse-grained.

1. Introduction

The study of mean field spin glasses, started by physicists in the 1970's and culminated in the Parisi theory [6], has over time revealed a rich and challenging mathematical structure. Despite the enormous progress over the last decades, the whole Parisi picture and its presumed universality for mean field models still remains not fully understood. Indeed, for the prototypical models of the theory such as the Sherrington-Kirkpatrick [SK] model [15], a rigorous treatment via large deviations techniques within the classical Gibbs-Boltzmann formalism is still lacking, and one has to resort to sophisticated machinery, in particular Guerra's interpolations [8, 16] and the Ghirlanda-Guerra identities [7]. These tools are remarkably efficient but cannot provide, by their own nature, insights for given realization of the disorder. We refer the reader to [9] and references therein for a comprehensive account of the state of the art for the SK-model[1].

One of the most puzzling features of the Parisi theory is the *ultrametricity*, namely the emergence in the infinite volume limit of hierarchical structures closely related to the generalized random energy models, the GREM, introduced by Derrida in the 1980's [4, 5]. For a large class of SK-type Hamiltonians, the ultrametricity has meanwhile been put on rigorous ground by Panchenko [10]. The analysis relies crucially on the Ghirlanda-Guerra identities.

In order to shed some light on the *origin* of ultrametricity, simplified models for spin glasses which are amenable to a rigorous solution via large deviation techniques have been introduced by Bolthausen and the first author in [1, 2, 3]: the first paper deals with a

Date: June 14, 2021.

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

Key words and phrases. Mean Field Spin Glasses, Large Deviations, Gibbs-Boltzmann and Parisi Variational Principles.

This work has been supported by a DFG research grant, contract number 2337/1-1.

[1]A treatment of spin glasses which is somewhat complementary to the Parisi theory goes through the TAP-Plefka framework [17, 11]. The rigorous studies which bypass the interpolations/GG-identities employ the Kac-Rice formula and some ensuing large deviations estimates for the complexity/number of critical points of Gaussian random fields. This type of analysis is to date however limited to spherical models with REM-like ultrametric structure, see [13, 14] and references.

nonhierarchical yet linear superposition of Derrida's GREMs, whereas the second paper analyses a non-linear version of the GREM which is closely related to the perceptron model from neural networks [6] and which we refer to henceforth as *Perceptron GREM*.

In the present paper we introduce a random Hamiltonian which is neither hierarchical from the start nor given by linear superpositions of GREM-like models. For such nonhierarchical, nonlinear generalization of the random energy models we establish a quenched, Sanov-type large deviation principle for the empirical distribution of the random energies which in turn allows to derive a Boltzmann-Gibbs principle, as well as a Parisi formula for the free energy in the thermodynamical limit. The limiting free energy turns out to be given by the minimal value over *multiple* nonlinear Parisi functionals / hierarchical structures.

What is perhaps more, our analysis suggests that the onset of ultrametricity is intimately related to certain monotonicity properties of relative entropies/Kullback-Leibler functionals, and shows how these monotonicities lead to the optimal balance in the energy-entropy competition.

We emphasize that the spin glass we investigate here is *nonlinear* insofar the sub-energies interact via a perceptron-like function, see (2.2) below for the definition of the model, but the Hamiltonian still is a linear functional of the empirical measures of the random energies. The case of nonlinear functionals, which give rise to models which better approximate realistic spin glasses such as the SK-model, will be addressed in a forthcoming work.

2. Definition of the model and main results

We start with some notation: for $n \in \mathbb{N}$, we set $I \stackrel{\text{def}}{=} \{1, 2, ..., n\}$ for the set of *species*. For $J \subseteq I$ we denote by

$$\mathcal{P}_J \stackrel{\text{def}}{=} \{A \subseteq J : A \neq \emptyset\}$$

the power set of J (without empty set). For $N \in \mathbb{N}$, which will play the role of the volume, and $j \in I$, we denote by

$$\alpha_j \in \left\{1, \cdots, 2^{\frac{N}{n}}\right\}$$

the *spins* associated to the jth-species. For $J \in \mathcal{P}_I$ we identify the $|J|$-tuple $(\alpha_j)_{j\in J}$ with

$$\alpha_J \stackrel{\text{def}}{=} (\alpha_j)_{j\in J}, \qquad \alpha_J \in \left\{1, \cdots, 2^{\frac{|J|N}{n}}\right\}.$$

By a slight abuse of notation, if $J = I$ we shorten $\mathcal{P}_n$ for $\mathcal{P}_I$, and write α for a *configuration*, i.e. the whole vector of spins $(\alpha_1, \cdots, \alpha_n)$.

Remark that we have altogether 2^N configurations, and since our configuration space is in one-to-one correspondence with the set $\{1, \ldots, 2^N\}$, to lighten notations, we will refer to the latter in the counting procedures which we implement below.

Consider independent random variables

$$\left\{X^J_{\alpha_J,i}\right\}_{J\in\mathcal{P}_n, i\in\mathbb{N}}$$

with values in a given Polish space S equipped with the Borel σ-field $\mathcal{S}$, and defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We assume that for $J \in \mathcal{P}_n$ each variable $X^J_{\alpha_J,i}$ is distributed

according to the same probability distribution μ_J. In this case, the product measure $\mu \stackrel{\text{def}}{=} \bigotimes_{J\in\mathcal{P}_n} \mu_J$ gives the joint distribution of the random vector

$$X_{\alpha,i} \stackrel{\text{def}}{=} \left(X^J_{\alpha_J,i}\right)_{J\in\mathcal{P}_n}. \tag{2.1}$$

Since $|\mathcal{P}_n| = 2^n - 1$, the joint law μ is an element of $\mathcal{M}_1^+(S^{2^n-1})$, the space of all probability distributions on the product space S^{2^n-1} which, endowed with the topology of weak convergence, is Polish itself.

For $d \in \mathbb{N}$, $\nu \in \mathcal{M}_1^+(S^d)$, we denote by $B_r(\nu)$ the open ball centered in ν with radius $r > 0$ in one of the standard metrics (e.g. Prokhorov's metric); a *neighborhood* or an *open subset* are conceived in the corresponding topology. This turns $\mathcal{M}_1^+(S^d)$ into a complete, separable metric space.

For $d \in \mathbb{N}$, $\nu \in \mathcal{M}_1^+(S^d)$ and B a subset of indeces, $B \subseteq \{1, \cdots, d\}$, we write $\nu^{(B)} \in \mathcal{M}_1^+\left(S^{|B|}\right)$ for the marginal of ν restricted to the coordinates corresponding to indices in B; with $\pi_B : S^d \to S^{|B|}$ the associated natural projection, it thus holds that

$$\nu^{(B)} \stackrel{\text{def}}{=} (\pi_B)_* (\nu) = \nu \circ \pi_B^{-1}.$$

Given a function $\phi : S^{2^n-1} \to \mathbb{R}$, our spin glass Hamiltonian is then

$$H_N(\alpha) = \sum_{i\le N} \phi\left(X_{\alpha,i}\right) = N \int \phi(x) L_{N,\alpha}(dx), \tag{2.2}$$

where

$$L_{N,\alpha} = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_{\alpha,i}}, \tag{2.3}$$

is the empirical measure, a random element in $\mathcal{M}_1^+\left(S^{2^n-1}\right)$. The partition function and free energy associated to (2.2) are defined as

$$Z_N(\phi) = 2^{-N}\sum_{\alpha=1}^{2^N} e^{H_N(\alpha)}, \qquad \text{resp.} \quad F_N(\phi) = \frac{1}{N}\log Z_N(\phi). \tag{2.4}$$

This spin glass model is thus a generalization of the Perceptron GREM discussed in [3] but which, unlike the latter, has no in-built ultrametric structure.

For $d \in \mathbb{N}$, $\nu, \nu' \in \mathcal{M}_1^+(S^d)$, we denote by

$$H(\,\nu\,|\,\nu'\,) \stackrel{\text{def}}{=} \begin{cases} \int \log \frac{d\nu}{d\nu'} d\nu & \text{if} \quad \nu \ll \nu' \\ \infty & \text{otherwise.} \end{cases}$$

the usual relative entropy of ν with respect to ν'. Finally, for $J \in \mathcal{P}_n$, we set

$$C_J \stackrel{\text{def}}{=} \left\{\nu \in \mathcal{M}_1^+(S^{2^n-1}) : H(\nu^{(\mathcal{P}_J)}\,|\,\mu^{(\mathcal{P}_J)}) \le \frac{|J|}{n}\log 2\right\}. \tag{2.5}$$

Our first result states that the quenched infinite volume free energy exists, and is given by the solution of a Boltzmann-Gibbs variational principle. The upshot is analogous to the control of the thermodynamical limit established in [1] through second moment estimates on the energy levels for the case of linear ϕ, and with disorder given by Gaussian random variables.

Theorem 2.1. *Let $\phi : S^{2^n-1} \to \mathbb{R}$ be continuous and bounded. Then $f(\phi) \stackrel{def}{=} \lim_{N\to\infty} F_N(\phi)$ exists almost surely, and is given by*

$$f(\phi) = \sup_{\nu\in\mathcal{M}_1^+(S^{2^n-1})} \left\{ \int \phi d\nu - \mathfrak{J}(\nu) \right\} \tag{2.6}$$

where the quenched rate function $\mathfrak{J} : \mathcal{M}_1^+(S^{2^n-1}) \to \overline{\mathbb{R}}$ *is given by*

$$\mathfrak{J}(\nu) \stackrel{def}{=} \begin{cases} H(\nu|\mu) & \text{if} \quad \nu \in \bigcap_{J\in\mathcal{P}_n} C_J \\ \infty & \text{otherwise.} \end{cases} \tag{2.7}$$

Since the set $\mathcal{P}_n$ is partially ordered through the inclusion relation, we can consider its totally ordered subsets, which are referred to, with the terminology from [1], as *chains*. We set

$$\mathcal{C}_n \stackrel{\text{def}}{=} \left\{ T = \left\{A_k^T\right\}_{k=0}^n : A_0^T = \emptyset, \forall k \in I \setminus \{n\} : \ A_k^T \subset I, \ A_k^T \subset A_{k+1}^T, \ |A_k^T| = k, \ A_n^T = I \right\}. \tag{2.8}$$

Each element in $\mathcal{C}_n$ corresponds to a specific hierarchical structure among all those in which the original model can be *coarse-grained*, i.e. to each chain $T \in \mathcal{C}_n$ we can associate a Hamiltonian H_N^T which gives rise to a Perceptron GREM. To see how this goes, for $T \in \mathcal{C}_n$, we define

$$T_k \stackrel{\text{def}}{=} \{J \subset I : J \subseteq A_k^T, J \not\subseteq A_{k-1}^T\} \subset \mathcal{P}_n\,, \quad k \in I. \tag{2.9}$$

The sets $T_1, \cdots, T_n$ describe explicitely how a specific chain T assembles the n-levels into the corresponding hierarchical structure: if

$$T = \{\, \emptyset \subset \{a_1\} \subset \{a_1, a_2\} \subset \cdots \subset \{a_1, \cdots, a_{n-1}\} \subset \{1, \cdots, n\} \,\} \tag{2.10}$$

then the α_{a_1}'s play the role of indeces for the first level, $(\alpha_{a_1}, \alpha_{a_2})$'s for the second etc.. Since $\mathcal{P}_n$ is the disjoint union of the $\{T_k\}_{k\in I}$ for any $J \in \mathcal{P}_n$ there is a unique

$$k_J = k_J(T) \in I \text{ for which } J \in T_{k_J}\,. \tag{2.11}$$

Given these indeces we shorten[2]

$$Y_{\alpha,i}^T = \left(Y_{\alpha_{a_1},\cdots\alpha_{a_{k_J}},i}^J \right)_{J\in\mathcal{P}_n}, \tag{2.12}$$

where the $Y_{\cdot,\cdot}^J$-s are μ_J-distributed and independent both over the α's as well as over the i's. Hence, the coarse grained Hamiltonian corresponding to the chain $T \in \mathcal{C}_n$ is defined as

$$H_N^T(\alpha) \stackrel{\text{def}}{=} \sum_{i=1}^N \phi(Y_{\alpha,i}^T)\,, \tag{2.13}$$

[2]We stress that the order in which the random sub-energies appear in the vectors (2.12) is the same as the one of the original vector (2.1): if $\mathcal{P}_n = \{J_1, \cdots, J_{2^n-1}\}$, then the c-th variable is distributed according to μ_{J_c}, $\forall\, c \in \{1, \cdots, 2^n - 1\}$.

with associated partition function and free energy in finite volume denoted by

$$Z_N^T(\phi) = 2^{-N} \sum_{\alpha=1}^{2^N} e^{H_N^T(\alpha)}, \quad F_N^T(\phi) = \frac{1}{N} \log Z_N^T(\phi). \tag{2.14}$$

The Hamiltonian (2.13) is in the form of a Perceptron GREM[3] introduced and solved in [3]: both a Gibbs variational principle, as well as a Parisi-like formula, for the limiting free energy are available.

For later purposes, we need to recall how the Parisi variational principle for the Perceptron GREM is constructed [3]: consider the standard simplex

$$\Delta \stackrel{\text{def}}{=} \{\boldsymbol{m} = (m_1, \dots, m_n) : \ 0 < m_1 \le m_2 \le \cdots \le m_n \le 1\} \ , \tag{2.15}$$

and recursively construct functions

$$\begin{aligned} \phi_n^T &= \phi, \\ \phi_{k-1}^T \left(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(k-1)}\right) &= \frac{1}{m_k} \log \int \ \exp\left[m_k \phi_k^T \left(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(k-1)}, \boldsymbol{y}\right)\right] \mu^{(T_k)}(d\boldsymbol{y}) \,, \end{aligned} \tag{2.16}$$

for $k \in I$, shortening also $\boldsymbol{x}^{(j)} = (x_J)_{J \in T_j}$ and $\mu^{(T_k)}(d\boldsymbol{y}) = \prod_{J \in \mathcal{P}_k^T \setminus \mathcal{P}_{k-1}^T} \mu_J(dy_J)$, and omitting the dependence of the ϕ-functions on the parameter $\boldsymbol{m}$ to lighten notations.

We then set

$$P^T(\boldsymbol{m}) \stackrel{\text{def}}{=} \frac{\log 2}{n} \sum_{k=1}^{n} \frac{1}{m_k} + \phi_0^T(\boldsymbol{m}) \,. \tag{2.17}$$

It is proved in [3, Theorem 2.3-2.5] that for continuous and bounded ϕ, the limit $f^T(\phi) \stackrel{\text{def}}{=} \lim_{N \to \infty} F_N^T$ exists almost surely and is given by the solution of the Boltzmann-Gibbs variational principle

$$f^T(\phi) = \sup_{\nu \in \mathcal{M}_1^+(S^{2^n-1})} \left\{ \int \phi(x) d\nu - \mathfrak{J}^T(\nu) \right\} \ , \tag{2.18}$$

with quenched rate function $\mathfrak{J}^T : \mathcal{M}_1^+(S^{2^n-1}) \to \overline{\mathbb{R}}$ given by

$$\mathfrak{J}^T(\nu) \stackrel{\text{def}}{=} \begin{cases} H(\nu|\mu) & \text{if} \quad \nu \in \bigcap_{k=1}^{n} C_{A_k^T} \\ \infty & \text{otherwise,} \end{cases} \tag{2.19}$$

[3]The coarse-graining requires some burdening notation, but the procedure is quickly understood through a concrete example which the reader may want to keep in mind throughout: let $n = 2$, and consider $X_{\alpha,i} = (X_{\alpha_1,i}^{\{1\}}, \ X_{\alpha_2,i}^{\{2\}}, \ X_{(\alpha_1,\alpha_2),i}^{\{1,2\}}) \sim \mu_{\{1\}} \otimes \mu_{\{2\}} \otimes \mu_{\{1,2\}}$, in which case $\phi : S^3 \to \mathbb{R}$ and $H_N(\alpha) = \sum_i \phi(X_{\alpha_1,i}^{\{1\}}, \ X_{\alpha_2,i}^{\{2\}}, \ X_{\alpha_1,\alpha_2,i}^{\{1,2\}})$. There are two possible chains: $R \stackrel{\text{def}}{=} \{\{1\}, \{1,2\}\}$ or $T \stackrel{\text{def}}{=} \{\{2\}, \{1,2\}\}$. By way of example we focus on the latter, in which case the sets (2.9) are given by $T_1 = \{\{2\}\}, T_2 = \{\{1\}, \{1,2\}\}$, and the new sub-energies are given by $Y_{\alpha_2,i}^{\{2\}} \sim \mu_{\{2\}}$ for the "first level", and $(Y_{(\alpha_1,\alpha_2),i}^{\{1\}}, Y_{(\alpha_1,\alpha_2),i}^{\{1,2\}}) \sim \mu_{\{1\}} \otimes \mu_{\{1,2\}}$ for the second. The Hamiltonian of the coarse-grained model is then $H_N^T(\alpha) = \sum_i \phi(Y_{\alpha_1,\alpha_2,i}^{\{1\}}, Y_{\alpha_2,i}^{\{2\}}, Y_{\alpha_1,\alpha_2,i}^{\{1,2\}})$. The key aspect of the procedure is thus to replace the $X_{\alpha_1,i}^{\{1\}}$ with $Y_{\alpha_1,\alpha_2,i}^{\{1\}}$ the latter being independent over the $\alpha_2' s$: this eventually induces a hierarchical correlation structure as in the Perceptron GREM.

and the sets $C_{A_k^T}$ as in (2.5). Moreover, the solution can be given in terms of the following Parisi variational principle

$$f^T(\phi) = \min_{\boldsymbol{m}\in\Delta} P^T(\boldsymbol{m})\,. \tag{2.20}$$

Our contribution here is to extend the above to the *nonhierarchical* Hamiltonian (2.2).

Theorem 2.2. *Assume $\phi : S^{2^n-1} \to \mathbb{R}$ is continuous and bounded. Then*

$$f(\phi) = \min_{T\in\mathcal{C}_n} f^T(\phi) = \min_{T\in\mathcal{C}_n}\left\{\min_{\boldsymbol{m}\in\Delta} P^T(\boldsymbol{m})\right\} \tag{2.21}$$

where $f(\phi)$ is the limiting free energy (2.6)*, and $f^T(\phi)$ the limiting free energy* (2.18) *associated to the chain $T \in \mathcal{C}_n$.*

In other words, the limiting free energy of the nonhierarchical Perceptron GREM is given by the minimum among the free energies associated to all chains in which the original model can be coarse-grained. The novelty compared to [3] is thus the minimization in (2.21) over all $T \in \mathcal{C}_n$, i.e. over all possible hierarchical structures: this feature is in full agreement with the Parisi theory for the more sophisticated mean field spin glasses like the SK-model, where the corresponding minimization is indeed over *all q-functions* [6].

3. The Gibbs variational principle

In this section we give a proof of Theorem 2.1. For this we will closely follow the large deviation analysis of [3] of the Perceptron GREM, which is flexible, and carries over to this more general setting quite swiftly. Lest the reader is overwhelmed with innumerous pointers and references, and given that the treatment is reasonably short, we will work out out all steps in detail.

We first need some infrastructure: for an open subset $U \subset \mathcal{M}_1^+(S^{2^n-1})$ we will denote by $M_N(U)$ the random variable counting the number of α-s for which the corresponding empirical distribution (2.3) lies in U, to wit:

$$M_N(U) = \#\{\alpha = 1, \dots, 2^N : L_{N,\alpha} \in U\}.$$

The following two lemmata state some well known properties of relative entropies. The short and elementary proofs are given for completeness.

Lemma 3.1. *Let S be a Polish space, $d \in \mathbb{N}$, $\nu, \mu \in \mathcal{M}_1^+\left(S^d\right)$ with $\nu \ll \mu$, $\mu = \bigotimes_{k=1}^d \mu_k$ for some $\mu_k \in \mathcal{M}_1^+(S)$, $k \in \{1, \cdots, d\}$. Then for every nonempty $B \subseteq \{1, \cdots, d\}$,*

$$H(\nu\,|\,\mu) = H\left(\nu^{(B)}\,|\,\mu^{(B)}\right) + H\left(\nu\,|\,\nu^{(B)} \otimes \mu^{(B^c)}\right) \tag{3.1}$$

$$H\left(\nu^{(B)}\,|\,\mu^{(B)}\right) \le H\left(\nu\,|\,\mu\right) \tag{3.2}$$

where $B^c = \{1, \cdots, d\} \setminus B$.

Proof. The second claim (3.2) is an immediate consequence of (3.1), together with the positivity of relative entropies. In order to see the first claim, let $\{S_k\}_{k=1}^d$ be d copies of S and $B = \{b_1, \cdots, b_M\}$, $B^c = \{a_1, \cdots, a_{d-M}\}$. Since the product space of Polish spaces is Polish, specifically Radon,the disintegration Theorem holds for every $\nu \in \mathcal{M}_1^+\left(S^d\right)$ and we can define the conditional distribution $\nu^{(B^c|B)}$ of ν on $S_{a_1} \times \cdots \times S_{a_{d-M}}$ given the

projection on $S_{b_1} \times \cdots \times S_{b_M}$. Expressing $\nu = \nu^{(B)} \otimes \nu^{(B^c|B)}$ as a semi-direct product, it holds

$$\begin{aligned} H(\nu \,|\, \mu) = \int \log\left(\frac{d\nu}{d\mu}\right) d\nu &= \int \log\left(\frac{d\nu^{(B)}}{d\mu^{(B)}}\right) d\nu + \int \log\left(\frac{d\nu^{(B^c|B)}}{d\mu^{(B^c)}}\right) d\nu \\ &= H\left(\nu^{(B)} \,|\, \mu^{(B)}\right) + \int \log\left(\frac{d\nu^{(B^c|B)}}{d\mu^{(B^c)}}\right) d\nu\,. \end{aligned} \tag{3.3}$$

We now focus on the second term on the r.h.s. above: by well-known properties of Radon-Nikodym derivatives[4] we may write

$$\frac{d\nu^{(B^c|B)}}{d\mu^{(B^c)}} = \frac{d\nu^{(B)}}{d\nu^{(B)}} \cdot \frac{d\nu^{(B^c|B)}}{d\mu^{(B^c)}} = \frac{d\nu}{d\left(\nu^{(B)} \otimes \mu^{(B^c)}\right)}\,, \tag{3.4}$$

and therefore

$$\int \log\left(\frac{d\nu^{(B^c|B)}}{d\mu^{(B^c)}}\right) d\nu = \int \log\left(\frac{d\nu}{d\left(\nu^{(B)} \otimes \mu^{(B^c)}\right)}\right) d\nu = H\left(\nu \,|\, \nu^{(B)} \otimes \mu^{(B^c)}\right). \tag{3.5}$$

Plugging this in (3.3) then yields (3.1).

□

Lemma 3.2. *(Semicontinuity of relative entropies) With the notation of the previous Lemma, the functions $\nu \to H(\nu \,|\, \mu)$, $\nu \to H\left(\nu \,|\, \nu^{(B)} \otimes \mu^{(B^c)}\right)$ are lower semicontinuous in the weak topology.*

Proof. It is well known that for relative entropies the following representations hold:

$$H(\nu \,|\, \mu) = \sup_{u \in \mathcal{U}} \left[\int u d\nu - \log \int e^u d\mu\right], \tag{3.6}$$

where $\mathcal{U}$ is the set of all bounded continous functions $S^d \to \mathbb{R}$, and analogously for

$$H\left(\nu \,|\, \nu^{(B)} \otimes \mu^{(B^c)}\right) = \sup_{u \in \mathcal{U}} \left[\int u d\nu - \log \int e^u d\left(\nu^{(B)} \otimes \mu^{(B^c)}\right)\right]. \tag{3.7}$$

But for $u \in \mathcal{U}$, the functionals

$$\nu \to \int u d\nu, \quad \nu \to \log \int e^u d\mu, \quad \nu \to \log \int e^u d\left(\nu^{(B)} \otimes \mu^{(B^c)}\right)$$

are all continuous: the claim of the Lemma thus steadily follows from this observation and the representations (3.6) and (3.7). □

The next technical result is a generalization of [3, Lemma 3.1].

Lemma 3.3. *(Relative Sanov Theorem) Let $\alpha = (\alpha_1, \cdots, \alpha_n)$, $\alpha' = (\alpha'_1, \cdots, \alpha'_n)$ be two configurations and $L_{N,\alpha}, L_{N,\alpha'}$ the associated empirical distributions defined in* (2.3)*. Assume that for $A \in \mathcal{P}_n$,*

$$\alpha_j = \alpha'_j,\ j \in A \qquad \text{and} \qquad \alpha_j \neq \alpha'_j,\ j \in I \setminus A.$$

Then the couple $(L_{N,\alpha}, L_{N,\alpha'})$ satisfies a LDP with good rate function

$$\mathfrak{J}_A(\nu, \theta) = H(\nu^{(\mathcal{P}_A)} \,|\, \mu^{(\mathcal{P}_A)}) + H\left(\nu \,|\, \nu^{(\mathcal{P}_A)} \otimes \mu^{(\mathcal{P}_A^c)}\right) + H\left(\theta \,|\, \theta^{(\mathcal{P}_A)} \otimes \mu^{(\mathcal{P}_A^c)}\right), \tag{3.8}$$

[4]This step is somewhat informal, but see e.g. [12].

if $\nu^{(\mathcal{P}_A)} = \theta^{(\mathcal{P}_A)}$ *(and* $= \infty$ *otherwise), where* $\mathcal{P}_A^c = \mathcal{P}_n \setminus \mathcal{P}_A$.

Proof. Let S be a Polish space, $d \in \mathbb{N}$,

$$\mathcal{P}' = \{b_1, \cdots, b_M\} \subseteq \{1, \cdots, d\}, \mathcal{P}' \neq \emptyset,$$

and $\{\boldsymbol{X}_i\}$, $\{\boldsymbol{Y}_i\}$ two independent families of indendent random variables with

$$\boldsymbol{X}_i \sim \mu \in \bigotimes_{k=1}^{d} \mathcal{M}_1^+(S_k), \qquad \boldsymbol{Y}_i \sim \mu' \in \bigotimes_{l=1}^{M} \mathcal{M}_1^+(S'_{b_l}),$$

where each S_k, S'_{b_l} is a copy of S.
We consider the empirical distribution

$$M_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{(\boldsymbol{X}_i, \boldsymbol{Y}_i)} \tag{3.9}$$

associated to vectors $(\boldsymbol{X}_i, \boldsymbol{Y}_i) \in S_1 \times \cdots \times S_d \times S'_{b_1} \times \cdots \times S'_{b_M}$ and the natural projections

$$\begin{aligned} \pi &: S_1 \times \cdots \times S_d \times S'_{b_1} \times \cdots \times S'_{b_M} \longrightarrow S_1 \times \cdots \times S_d, \\ \pi' &: S_1 \times \cdots \times S_d \times S'_{b_1} \times \cdots \times S'_{b_M} \longrightarrow \tilde{S}_1 \times \cdots \times \tilde{S}_d, \end{aligned} \tag{3.10}$$

where

$$\tilde{S}_k \stackrel{\text{def}}{=} \begin{cases} S_k & \text{if} \quad k \in \mathcal{P} \\ S'_k & \text{if} \quad k \notin \mathcal{P} \end{cases}, \qquad \mathcal{P} \stackrel{\text{def}}{=} \{1, \cdots, d\} \setminus \mathcal{P}' = \{c_1, c_2, \cdots, c_{d-M}\}.$$

We define

$$(L_N, R_N) \stackrel{\text{def}}{=} M_N \circ (\pi, \pi')^{-1},$$

which is a random element in $\mathcal{M}_1^+(S_1 \times \cdots \times S_d) \times \mathcal{M}_1^+(\tilde{S}_1 \times \cdots \times \tilde{S}_d)$, and

$$\mathfrak{J}(\nu, \theta) \stackrel{\text{def}}{=} \begin{cases} H(\nu^{(\mathcal{P})} \,|\, \mu^{(\mathcal{P})}) + H\left(\nu \,|\, \nu^{(\mathcal{P})} \otimes \mu^{(\mathcal{P}')}\right) + H\left(\theta \,|\, \theta^{(\mathcal{P})} \otimes \mu'\right) & \text{if} \quad \nu^{(\mathcal{P})} = \theta^{(\mathcal{P})} \\ \infty & \text{otherwise.} \end{cases} \tag{3.11}$$

Applying Sanov's Theorem to (3.9), and by continuous projection, we obtain that the couple (L_N, R_N) satisfies a LDP with good rate function

$$\mathfrak{J}'(\nu, \theta) = \inf_{\rho} \left\{ H\left(\rho \,|\, \mu \otimes \mu'\right) : \quad \rho\pi^{-1} = \nu, \quad \rho\pi'^{-1} = \theta \right\}.$$

Since the projections (3.10) coincide on coordinates indexed in $\mathcal{P}$, for all $\rho \in \mathcal{M}_1^+\left(S_1 \times \cdots \times S_d \times S'_{b_1} \times \cdots \times S'_{b_M}\right)$ it holds $(\rho\pi^{-1})^{(\mathcal{P})} = (\rho\pi'^{-1})^{(\mathcal{P})}$ and the following implication is immediate

$$\nu^{(\mathcal{P})} \neq \theta^{(\mathcal{P})} \Longrightarrow \mathfrak{J}'(\nu, \theta) = \infty.$$

We may thus henceforth assume $\nu^{(\mathcal{P})} = \theta^{(\mathcal{P})}$, and consider $\tilde{\rho} = \tilde{\rho}(\nu, \theta)$ with marginal $\nu^{(\mathcal{P})} = \theta^{(\mathcal{P})}$ on $S_{c_1} \times \cdots \times S_{c_{d-M}}$, and conditional distribution on $S_{b_1} \times \cdots \times S_{b_M} \times S'_{b_1} \times \cdots \times S'_{b_M}$ given the projection on $S_{c_1} \times \cdots \times S_{c_{d-M}}$ as the product of the conditional distribution of ν on $S_{b_1} \times \cdots \times S_{b_M}$ given the projection on $S_{c_1} \times \cdots \times S_{c_{d-M}}$, and of the conditional

distribution θ on $S'_{b_1} \times \cdots \times S'_{b_M}$ given the projection on $S_{c_1} \times \cdots \times S_{c_{d-M}}$. In which case, applying (3.1), we easily obtain

$$H\left(\tilde{\rho}\,|\,\mu\otimes\mu'\right) = H\left(\nu^{(\mathcal{P})}\,|\,\mu^{(\mathcal{P})}\right) + H\left(\nu\,|\,\nu^{(\mathcal{P})}\otimes\mu^{(\mathcal{P}')}\right) + H\left(\theta\,|\,\theta^{(\mathcal{P})}\otimes\mu'\right)$$

and therefore $\mathfrak{J} \geq \mathfrak{J}'$. Conversely, for any ρ satisfying $\rho\pi^{-1} = \nu$, $\rho\pi'^{-1} = \theta$, we claim that

$$\mathfrak{J}(\nu,\theta) \leq H(\rho\,|\,\mu\otimes\mu')\,. \tag{3.12}$$

(Remark that we can assume that the right hand side is finite). To see this, we first write

$$H\left(\rho\,|\,\mu\otimes\mu'\right) = H\left(\rho\,|\,\tilde{\rho}(\nu,\theta)\right) + \int d\rho \log\frac{d\tilde{\rho}(\nu,\theta)}{d(\mu\otimes\mu')}.$$

The first term is non-negative, while the second equals

$$\int d\tilde{\rho}(\nu,\theta)\log\frac{d\tilde{\rho}(\nu,\theta)}{d(\mu\otimes\mu')} = \mathfrak{J}(\nu,\theta),$$

and (3.12) follows. The statement of the Lemma corresponds then to the case

$$d = 2^n - 1, \qquad \mathcal{P}' = \mathcal{P}_A^c.$$

□

Proposition 3.4. *Let $\nu \in \mathcal{M}_1^+\left(S^{2^n-1}\right)$ be such that $H(\nu\,|\,\mu) < \infty$, $\epsilon > 0$ and V an open neighborhood of ν. Then there exists an open neighborhood U of ν, $U \subset V$, and $\delta > 0$ for which*

$$\mathbb{P}\left[M_N(U) \geq e^{N(\log 2 - H(\nu\,|\,\mu)+\epsilon)}\right] \leq e^{-\delta N}.$$

Proof. By the semicontinuity property of the relative entropy, given $B_r(\nu)$ a family of open balls with radius $r > 0$, one has

$$\inf_{\rho\in B_r(\nu)} H(\rho\,|\,\mu) \to H(\nu\,|\,\mu) \qquad \text{as} \quad r \to 0.$$

Specifically, we can extract a sequence $\{r_k\}_{k\in\mathbb{N}}$ with $r_k \to 0$ as $k \to \infty$ so that

$$\inf_{\rho\in B_r(\nu)} H(\rho\,|\,\mu) = \inf_{\rho\in\overline{B_r}(\nu)} H(\rho\,|\,\mu) \to H(\nu\,|\,\mu).$$

Therefore for every $\epsilon > 0$, V we may find $k \in \mathbb{N}$ for which

$$\overline{B_{r_k}}(\nu) \subset V \quad \text{and} \quad \inf_{\rho\in B_{r_k}(\nu)} H(\rho\,|\,\mu) = \inf_{\rho\in\overline{B_{r_k}}(\nu)} H(\rho\,|\,\mu) \geq H(\nu\,|\,\mu) - \frac{\epsilon}{4}\,. \tag{3.13}$$

All in all,

$$\mathbb{E}\left[M_N(B_{r_k}(\nu))\right] = \sum_{\alpha=1}^{2^N}\mathbb{P}\left[L_{N,\alpha}\in B_{r_k}(\nu)\right] \leq 2^N\exp\left[-N\left(H(\nu\,|\,\mu) - \frac{\epsilon}{2}\right)\right],$$

the last step by Sanov's theorem together with (3.13).

The claim of the Proposition steadily follows from Markov's inequality, and $\delta = \epsilon/4$. □

Proposition 3.5. *Let $\nu \in \mathcal{M}_1^+\left(S^{2^n-1}\right)$ be such that for some $J \in \mathcal{P}_n$, $H\left(\nu^{(\mathcal{P}_J)} \,|\, \mu^{(\mathcal{P}_J)}\right) > |J| \log 2/n$ and V an open neighborhood of ν. Then there exists an open neighborhood U of ν, $U \subset V$, and $\delta > 0$ such that for N large enough*

$$\mathbb{P}\left[M_N(U) \neq 0\right] \leq e^{-\delta N}.$$

Proof. Let $J \in \mathcal{P}_n$ satisfy $H\left(\nu^{(\mathcal{P}_J)} \,|\, \mu^{(\mathcal{P}_J)}\right) > |J| \log 2/n$. Similarly to the proof of Proposition 3.4 we can select a neighborhood W of $\nu^{(\mathcal{P}_J)} \in \mathcal{M}_1^+\left(S^{|\mathcal{P}_J|}\right)$ with

$$\inf_{\rho \in \overline{W}} H\left(\rho \mid \mu^{(\mathcal{P}_J)}\right) = \inf_{\rho \in W} H\left(\rho \mid \mu^{(\mathcal{P}_J)}\right) > \frac{|J|}{n} \log 2 + 2\delta,$$

for some $\delta > 0$. If we now define

$$U \stackrel{\text{def}}{=} \left\{\nu \in \mathcal{M}_1^+(S^{2^n-1}) : \ \nu \in V, \nu^{(\mathcal{P}_J)} \in W\right\},$$

it follows from union bounds/subadditivity that

$$\begin{aligned}
\mathbb{P}\left[M_N(U) \neq 0\right] &= \mathbb{P}\left[\exists \alpha : \ L_{N,\alpha} \in U\right] \leq \mathbb{P}\left[\exists \alpha : \ (L_{N,\alpha})^{(\mathcal{P}_J)} \in W\right] \\
&\leq \sum_{\alpha_J=1}^{2^{|J|N/n}} \mathbb{P}\left[\,(L_{N,\alpha})^{(\mathcal{P}_J)} \in W\right] \qquad (3.14) \\
&= 2^{\frac{|J|}{n}N}\, \mathbb{P}\left[\,(L_{N,\alpha})^{(\mathcal{P}_J)} \in W\right] ,
\end{aligned}$$

the last line using that the empirical measures are identically distributed. Since the r.h.s. above is, by Sanov's theorem, *at most*

$$2^{\frac{|J|}{n}N} \exp\left[-N\left(\inf_{\rho \in \overline{W}} H\left(\rho \mid \mu^{(\mathcal{P}_J)}\right) - \delta\right)\right] \leq 2^{\frac{|J|}{n}N} 2^{-\frac{|J|}{n}N} e^{-\delta N} = e^{-\delta N}, \qquad (3.15)$$

the proof of the Proposition is concluded. □

Proposition 3.6. *Assume that $\nu \in \mathcal{M}_1^+(S^{2^n-1})$ satisfies the C-conditions* (2.5) strictly, *namely that $H\left(\nu^{(\mathcal{P}_J)} | \mu^{(\mathcal{P}_J)}\right) < |J|/n \log 2$, $J \in \mathcal{P}_n$. Let V be an open neighborhood of ν, and $\epsilon > 0$. Then there exists an open neighborhood U of ν with $U \subset V$ and $\delta > 0$ such that for large enough N*

$$\mathbb{P}\left[M_N(U) \leq e^{N(\log 2 - H(\nu \,|\, \mu) - \epsilon)}\right] \leq e^{-\delta N}.$$

Proof. We claim that for all $\epsilon > 0$ there exists $\delta > 0$ and $U \subset V$ an open neighborhood of ν for which

$$\mathbb{V}\mathrm{ar}\, M_N(U) \leq e^{-2\delta N} \left[\mathbb{E}\, M_N(U)\right]^2 . \qquad (3.16)$$

Assuming this for the time being, the proposition follows steadily as an elementary application of Chebycheff's inequality: indeed, by Sanov's theorem,

$$\mathbb{E}\, M_N(U) = 2^N\, \mathbb{P}\left(L_{N,\alpha} \in U\right) \geq 2^N \exp\left(-NH(\nu \,|\, \mu) - N\xi\right) , \qquad (3.17)$$

for $\xi > 0$. Elementary manipulations then yield

$$\begin{aligned}
&\mathbb{P}\left[M_N(U) \le 2^N \exp\left(-NH(\nu\,|\,\mu) - 2N\xi\right)\right] \\
&\qquad = \mathbb{P}\left[M_N(U) - \mathbb{E}\,M_N(U) \le e^{-N\xi}\,2^N \exp\left(-NH(\nu\,|\,\mu) - N\xi\right) - \mathbb{E}\,M_N(U)\right] \\
&\qquad \overset{(3.17)}{\le} \mathbb{P}\left[M_N(U) - \mathbb{E}\,M_N(U) \le \left(e^{-N\xi} - 1\right)\mathbb{E}\,M_N(U)\right] \\
&\qquad \le \mathbb{P}\left[M_N(U) - \mathbb{E}\,M_N(U) \le -\frac{1}{2}\,\mathbb{E}\,M_N(U)\right],
\end{aligned} \tag{3.18}$$

for N large enough. But the r.h.s. of (3.18) is *at most*

$$\mathbb{P}\left[|M_N(U) - \mathbb{E}\,M_N(U)| \ge \frac{1}{2}\,\mathbb{E}\,M_N(U)\right] \le 4\frac{\mathbb{V}\mathrm{ar}\,M_N(U)}{[\mathbb{E}\,M_N(U)]^2} \overset{(3.16)}{\le} 4e^{-2N\delta} \le e^{-\delta N}, \tag{3.19}$$

settling the claim of the proposition.

It thus remains to prove (3.16). To see how this goes, we first observe that, under the assumptions of the proposition, there exists $\eta > 0$ such that

$$H\left(\nu^{(\mathcal{P}_J)}\,|\,\mu^{(\mathcal{P}_J)}\right) < \frac{|J|}{n}\log 2 - \eta. \tag{3.20}$$

Moreover, we claim that for $J \in \mathcal{P}_n$,

$$\begin{aligned}
\lim_{r\to 0} \inf_{\substack{\rho,\theta\in\overline{B}_r(\nu)\\ \rho^{(\mathcal{P}_J)}=\theta^{(\mathcal{P}_J)}}} &\left\{H(\rho\,|\,\mu) + H\left(\theta\,|\,\theta^{(\mathcal{P}_J)} \otimes \mu^{(\mathcal{P}_J^c)}\right)\right\} = \\
&= H(\nu\,|\,\mu) + H\left(\nu\,|\,\nu^{(\mathcal{P}_J)} \otimes \mu^{(\mathcal{P}_J^c)}\right).
\end{aligned} \tag{3.21}$$

This can be checked through approriate upper- and lower bounds. The bound "$\le$" is straightforward: simply plug in $\rho = \theta = \nu$ on the l.h.s. above. The bound "$\ge$" is a consequence of the semicontinuity properties of the relative entropies: in fact, by Lemma 3.2, one gets that for a sequence (ρ_m, θ_m) with $\rho_m^{(\mathcal{P}_J)} = \theta_m^{(\mathcal{P}_J)}$ and $\rho_m, \theta_m \to \nu$ it holds

$$\begin{aligned}
\liminf_{m\to\infty} H(\rho_m\,|\,\mu) &\ge H(\nu\,|\,\mu), \\
\liminf_{m\to\infty} H\left(\theta_m\,|\,\theta_m^{(\mathcal{P}_J)} \otimes \mu^{(\mathcal{P}_J^c)}\right) &\ge H\left(\nu\,|\,\nu^{(\mathcal{P}_J)} \otimes \mu^{(\mathcal{P}_J^c)}\right),
\end{aligned} \tag{3.22}$$

which settles the "$\ge$"-case, and thus (3.21).

By (3.21), for r small enough such that $\overline{B}_r(\nu) \subset V, U = B_r(\nu)$ we have that for $J \in \mathcal{P}_n$, it holds:

$$\begin{aligned}
&\inf_{\substack{\rho,\theta\in\overline{B}_r(\nu)\\ \rho^{(\mathcal{P}_J)}=\theta^{(\mathcal{P}_J)}}} \left\{H(\rho\,|\,\mu) + H\left(\theta\,|\,\theta^{(\mathcal{P}_J)} \otimes \mu^{(\mathcal{P}_J^c)}\right)\right\} \\
&\qquad \ge H(\nu\,|\,\mu) + H\left(\nu\,|\,\nu^{(\mathcal{P}_J)} \otimes \mu^{(\mathcal{P}_J^c)}\right) - \frac{\eta}{2} \\
&\qquad = H(\nu\,|\,\mu) + H(\nu\,|\,\mu) - H\left(\nu^{(\mathcal{P}_J)}\,|\,\mu^{(\mathcal{P}_J)}\right) - \frac{\eta}{2},
\end{aligned} \tag{3.23}$$

the last step by the chain rule (3.1). The r.h.s. above equals

$$2H(\nu\,|\,\mu) - H\left(\nu^{(\mathcal{P}_J)}\,|\,\mu^{(\mathcal{P}_J)}\right) - \frac{\eta}{2} \geq 2H(\nu\,|\,\mu) - \frac{|J|}{n}\log 2 + \frac{\eta}{2}, \tag{3.24}$$

the last step using (3.20). All in all, we have established that

$$\inf_{\substack{\rho,\theta\in\overline{U}\\ \rho^{(\mathcal{P}_J)}=\theta^{(\mathcal{P}_J)}}} \left\{H(\rho\,|\,\mu) + H\left(\theta\,|\,\theta^{(\mathcal{P}_J)}\otimes\mu^{(\mathcal{P}_J^c)}\right)\right\} \geq 2H(\nu\,|\,\mu) - \frac{|J|}{n}\log 2 + \frac{\eta}{2}\,. \tag{3.25}$$

This inequality will play a crucial role in the control of the second moment of the M_N-counting randon variable which stands behind (3.16): we begin by writing

$$\begin{aligned}\mathbb{E}\,M_N^2(U) &= \sum_{\alpha,\alpha'=1}^{2^N} \mathbb{P}\left[L_{\alpha,N}\in U, L_{\alpha',N}\in U\right]\\ &\leq \sum_{\alpha_j\neq\alpha'_j,\ \forall j\in I} p_{\alpha,\alpha'} + \sum_{J\in\mathcal{P}_n}\ \sum_{\substack{\alpha_j=\alpha'_j\ \forall j\in J\\ \alpha_j\neq\alpha'_j\ \forall j\in I\setminus J}} p_{\alpha,\alpha'},\end{aligned} \tag{3.26}$$

where we have shortened, to lighten notations,

$$p_{\alpha,\alpha'} \stackrel{\text{def}}{=} \mathbb{P}\left[L_{\alpha,N}\in\overline{U}, L_{\alpha',N}\in\overline{U}\right].$$

Now, if $\alpha_j\neq\alpha'_j$ for all $j\in I$, then the associated empirical measures $L_{\alpha,N}$ and $L_{\alpha',N}$ are independent, and we may bound the corresponding contribution in (3.26) with the square of the first moment to obtain

$$\mathbb{E}\,M_N^2(U) \leq \left[\mathbb{E}\,M_N(\overline{U})\right]^2 + \sum_{J\in\mathcal{P}_n}\ \sum_{\substack{\alpha_j=\alpha'_j\ \forall j\in J\\ \alpha_j\neq\alpha'_j\ \forall j\in I\setminus J}} p_{\alpha,\alpha'}. \tag{3.27}$$

In order to control the probabilities in the sum on the r.h.s. above, we consider the generic situation where for some $J\in\mathcal{P}_n$, $\alpha_j=\alpha'_j\quad\forall j\in J$, $\alpha_j\neq\alpha'_j\quad\forall j\in I\setminus J$. In this case, by the relative Sanov principle from Lemma 3.3, and for N large enough, it holds:

$$p_{\alpha,\alpha'} = \mathbb{P}\left[L_{\alpha,N}\in\overline{U}, L_{\alpha',N}\in\overline{U}\right] \leq \exp\left[-N\inf_{\substack{\rho,\theta\in\overline{U}\\ \rho^{(\mathcal{P}_J)}=\theta^{(\mathcal{P}_J)}}}\left\{\mathfrak{J}_J(\rho,\theta)\right\} + \frac{N\eta}{4}\right], \tag{3.28}$$

where the quenched rate function with prescribed marginals as above satisfies

$$\begin{aligned}\mathfrak{J}_J(\rho,\theta) &= H(\rho^{(\mathcal{P}_J)}\,|\,\mu^{(\mathcal{P}_J)}) + H\left(\rho\,|\,\rho^{(\mathcal{P}_J)}\otimes\mu^{(\mathcal{P}_J^c)}\right) + H\left(\theta\,|\,\theta^{(\mathcal{P}_J)}\otimes\mu^{(\mathcal{P}_J^c)}\right)\\ &= H(\rho\,|\,\mu) + H\left(\theta\,|\,\theta^{(\mathcal{P}_J)}\otimes\mu^{(\mathcal{P}_J^c)}\right),\end{aligned} \tag{3.29}$$

the first step by definition, and the second by the chain rule from Lemma 3.1. Plugging this into (3.28) we thus get

$$\begin{aligned} p_{\alpha,\alpha'} &\leq \exp\left[-N \inf_{\substack{\rho,\theta\in\overline{U}\\ \rho^{(\mathcal{P}_J)}=\theta^{(\mathcal{P}_J)}}} \left\{H(\rho\,|\,\mu) + H\left(\theta\,|\,\theta^{(\mathcal{P}_J)}\otimes\mu^{(\mathcal{P}_J^c)}\right)\right\} + \frac{N\eta}{4}\right] \\ &\leq 2^{\frac{N|J|}{n}} \exp\left[-2NH(\nu\,|\,\mu) - \frac{N\eta}{4}\right], \end{aligned} \tag{3.30}$$

the last inequality by the key bound (3.25), and therefore

$$\begin{aligned} \sum_{\substack{\alpha_j=\alpha'_j \quad \forall j\in J\\ \alpha_j\neq\alpha'_j \quad \forall j\in I\setminus J}} p_{\alpha,\alpha'} &\leq 2^{(n-|J|)\frac{N}{n}} 2^{\frac{N|J|}{n}} \left(2^{\frac{N}{n}}-1\right)^{(n-|J|)} 2^{\frac{N|J|}{n}} \exp\left[-2NH(\nu\,|\,\mu) - \frac{N\eta}{4}\right] \\ &\leq 2^{2N} \exp\left[-2NH(\nu\,|\,\mu) - \frac{N\eta}{4}\right]. \end{aligned} \tag{3.31}$$

Plugging this in (3.27), and recalling (3.17) with $\xi = \frac{\eta}{16}$, we thus get that for N large enough

$$\mathbb{V}\mathrm{ar}\, M_N(U) \leq 2^{2N}(2^n-1)\exp\left[-2NH(\nu\,|\,\mu) - \frac{N\eta}{4}\right] \leq e^{-\frac{N\eta}{8}} \left[\mathbb{E}\, M_N(U)\right]^2, \tag{3.32}$$

which settles the claim (3.16) with $\delta = \frac{\eta}{16}$, say.

□

Proof of Theorem 2.1. We begin with a couple of elementary observations.

- First recall that for $J \in \mathcal{P}_n$,

$$C_J = \left\{\nu \in \mathcal{M}_1^+(S^{2^n-1}) : H(\nu^{(\mathcal{P}_J)}\,|\,\mu^{(\mathcal{P}_J)}) \leq \frac{|J|}{n}\log 2\right\}. \tag{3.33}$$

  These sets are compact thanks to the semicontinuity of $\nu \to H(\nu\,|\,\mu)$ from Lemma 3.2 . It thus follows that the intersection $\bigcap_{J\in\mathcal{P}_n} C_J \subset \mathcal{M}_1^+(S^{2^n-1})$ is itself compact, and in particular that there exists $\nu' \in \bigcap_{J\in\mathcal{P}_n} C_J$ such that

$$\sup_{\nu\in\bigcap\limits_{J\in\mathcal{P}_n} C_J} \left\{\int \phi d\nu - H(\nu\,|\,\mu)\right\} = \int \phi d\nu' - H(\nu'\,|\,\mu).$$

- It is well known that $\nu \to H(\nu\,|\,\mu)$ is convex.
- It clearly holds that $\mu \in \bigcap_{J\in\mathcal{P}_n} C_J$.

Given a convex combination $\nu_\lambda \stackrel{\text{def}}{=} \lambda\mu + (1-\lambda)\nu'$, $0<\lambda<1$ we thus have

$$H(\nu_\lambda^{(\mathcal{P}_J)}\,|\,\mu^{(\mathcal{P}_J)}) < \frac{|J|}{n}\log 2 \qquad \forall J \in \mathcal{P}_n.$$

We also know that as $\lambda \to 0$, $\nu_\lambda \to \nu'$ weakly and $\int \phi d\nu_\lambda \to \int \phi d\nu'$, $H(\nu_\lambda \,|\, \mu) \to H(\nu' \,|\, \mu)$. In particular, to $\epsilon > 0$ we can find $\lambda > 0$ with

$$\int \phi d\nu_\lambda - H(\nu_\lambda \,|\, \mu) \geq \int \phi d\nu' - H(\nu' \,|\, \mu) - \epsilon. \tag{3.34}$$

Moreover, since the map $\nu \to \int \phi(x)\nu(dx)$ is continuous, by Proposition 3.6 we can find a neighbourhood U of ν_λ, and $\delta > 0$ such that

$$\int \phi(x)\nu_\lambda(dx) - \int \phi(x)\rho(dx) \leq \epsilon \quad \forall \rho \in U,$$

and with the neighbourhood U such that

$$\mathbb{P}\left[M_N(U) \leq 2^N e^{-H(\nu_\lambda \,|\, \mu)N - \epsilon N)}\right] \leq e^{-\delta N}.$$

With this choice, we thus have for the partition function of our spin glass (2.4) that

$$\begin{aligned} Z_N(\phi) &= 2^{-N} \sum_{\alpha=1}^{2^N} \exp\left[N \int \phi(x) L_{N,\alpha}(dx)\right] \\ &\geq \exp\left[N \int \phi(x)\nu_\lambda(dx) - \epsilon N\right] \exp\left[-H(\nu_\lambda \,|\, \mu)N - \epsilon N\right] \\ &\geq \exp\left[N \sup_{\nu \in \bigcap_{J \in \mathcal{P}_n} C_J} \left\{\int \phi(x) d\nu - H(\nu \,|\, \mu)\right\} - 3\epsilon N\right]. \end{aligned} \tag{3.35}$$

As ϵ is arbitrary, it follows from Borel-Cantelli that

$$\liminf_{N\to\infty} F_N(\phi) = \liminf_{N\to\infty} \frac{1}{N} \log Z_N(\phi) \geq \sup_{\nu \in \bigcap_{J \in \mathcal{P}_n} C_J} \left\{\int \phi(x) d\nu - H(\nu \,|\, \mu)\right\}, \tag{3.36}$$

almost surely. This settles the "first half of the theorem", i.e. the lower bound to the free energy.

In order to establish the upper bound, we distinguish two cases:

i) For $\nu \in \bigcap_{J \in \mathcal{P}_n} C_J$, we choose a radius $r_\nu > 0$ such that

$$\mathbb{P}\left[M_N(B_{r_\nu}(\nu)) \geq e^{N(\log 2 - H(\nu \,|\, \mu) + \epsilon)}\right] \leq e^{-\delta_\nu N},$$

for large enough N and some $\delta_\nu > 0$; this is clearly possible in virtue of Proposition 3.4. The radius r_ν is chosen however small enough such that

$$\left|\int \phi(x)\rho(dx) - \int \phi(x)\nu(dx)\right| \leq \epsilon, \qquad \forall \rho \in B_{r_\nu}(\nu),$$

(which is possible thanks to the aforementioned continuity of the map $\rho \mapsto \int \phi(x)\rho(dx)$).

ii) For $\nu \in C_I \setminus \left(\bigcap_{J \in \mathcal{P}_n} C_J\right)$ we still have $H(\nu \,|\, \mu) \leq \log 2$ and we can resort to Proposition 3.5 to find $r_\nu > 0$ such that

$$\mathbb{P}\left[M_N(B_{r_\nu}(\nu)) \neq 0\right] \leq e^{-\delta_\nu N}, \tag{3.37}$$

for large enough N, and some $\delta_\nu > 0$.

Using that $C_I \subset \bigcup_{\nu\in C_I} B_{r_\nu}(\nu)$, by compactness we may extract a finite sequence $\{\nu_j\}_{j=1}^m \subset C_I$ for which covering still holds, to wit

$$C_I \subset \bigcup_{j=1}^m B_{r_j}(\nu_j),$$

where we have also shortened $r_j \stackrel{\text{def}}{=} r_{\nu_j}$. Moreover, if we set

$$U \stackrel{\text{def}}{=} \bigcup_{j=1}^m B_{r_j}(\nu_j), \qquad \delta \stackrel{\text{def}}{=} \min_{j\in\{1,\cdots,m\}} \delta_{\nu_j},$$

it follows from Sanov theorem that

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left[L_{N,\alpha} \notin U\right] \leq -\inf_{\nu\notin U} H(\nu\,|\,\mu) < -\log 2. \tag{3.38}$$

Splitting the contributions to the partition function then steadily yields the upper bound

$$\begin{aligned} Z_N(\phi) \leq 2^{-N} \sum_{l=1}^m \sum_{\alpha: L_{N,\alpha}\in B_{r_l}(\nu_l)} \exp\left[N\int \phi(x) L_{N,\alpha}(dx)\right] + \\ + 2^{-N} \sum_{\alpha: L_{N,\alpha}\notin U} \exp\left[\int \phi(x) L_{N,\alpha}(dx)\right]. \end{aligned} \tag{3.39}$$

By (3.38), the second summand yields no contribution in the large N-limit. Analogously, (3.37) establishes that the same applies to those terms in the first summand for which $\nu_l \notin \bigcap_{J\in\mathcal{P}_n} C_J$. All in all we have

$$\begin{aligned} Z_N &\leq 2^{-N} \sum_{l:\, \nu_l\in \bigcap_{J\in\mathcal{P}_n} C_J} \sum_{\alpha: L_{N,\alpha}\in B_{r_l}(\nu_l)} \exp\left[N\int \phi(x) L_{N,\alpha}(dx)\right] \\ &\leq e^{\epsilon N} \sum_{l:\, \nu_l\in \bigcap_{J\in\mathcal{P}_n} C_J} \exp\left[N\int \phi d\nu_l\right] M_N\left(B_{r_l}(\nu_l)\right) \\ &\leq e^{2\epsilon N} \sum_{l:\, \nu_l\in \bigcap_{J\in\mathcal{P}_n} C_J} \exp\left[N\int \phi d\nu_l - NH(\nu_l\,|\,\mu)\right] \\ &\leq e^{2\epsilon N} m \exp\left[N \sup_{\nu\in \bigcap_{J\in\mathcal{P}_n} C_J} \left\{\int \phi d\nu - H(\nu\,|\,\mu)\right\}\right] \end{aligned} \tag{3.40}$$

almost surely, for large enough N. Since ϵ is arbitrary we thus have

$$\limsup_{N\to\infty} F_N(\phi) = \limsup_{N\to\infty} \frac{1}{N} \log Z_N(\phi) \leq \sup_{\nu\in \bigcap_{J\in\mathcal{P}_n} C_J} \left\{\int \phi d\nu - H(\nu\,|\,\mu)\right\},$$

almost surely. This being the sought upperbound, Theorem 2.1 follows. □

4. The Parisi variational principle

In this section we prove Theorem 2.2. The following upper bound

$$f(\phi) \leq \min_{T \in \mathcal{C}_n} f^T(\phi), \tag{4.1}$$

is immediate: since for any $T \in \mathcal{C}_n$,

$$\bigcap_{J \in \mathcal{P}_n} C_J \subset \bigcap_{k=1}^{n} C_{A_k^T},$$

the Boltzmann-Gibbs variational principle (2.6) is analogous to the ones in (2.18) with the only difference that the supremum in the former is taken on a *smaller* set. This proves that $f(\phi) \leq f^T(\phi)$ for any $T \in \mathcal{C}_n$, and thus settles (4.1).

On the other hand, the Parisi variational principle (2.20) established in [3], states that for the Perceptron GREM

$$f^T(\phi) = \min_{\boldsymbol{m} \in \Delta} P^T(\boldsymbol{m}). \tag{4.2}$$

Combining (4.1) and (4.2) we thus have

$$f(\phi) \leq \min_{T \in \mathcal{C}_n} \left\{ \min_{m \in \Delta} P^T(\boldsymbol{m}) \right\}, \tag{4.3}$$

which is "half of the theorem".

In order to prove the "second half", namely the lower bound, some infrastructure is needed. Following [3], for each $T \in \mathcal{C}_n$ we define a probability distribution $G^T = G^T(\boldsymbol{m})$ on S^{2^n-1} which depends on the parameter $\boldsymbol{m} \in \Delta$. This *generalized Gibbs measure* is associated to the Parisi functional P^T, recall (2.17) above, and is described through

- a measure γ^T on S,
- and, for $2 \leq j \leq n$, Markov kernels K_j^T with source $S^{|\mathcal{P}_{j-1}^T|} = S^{2^{j-1}-1}$ and target $S^{|T_j|} = S^{2^{j-1}}$.

Precisely, and recalling the functions $\left\{\phi_k^T\right\}_{k=1}^n$ from (2.16), the generalized Gibbs measure G^T is given by the semi-direct product

$$G^T = \gamma^T \otimes K_2^T \otimes \cdots \otimes K_n^T. \tag{4.4}$$

$$\gamma^T(dx) \stackrel{\text{def}}{=} \frac{\exp\left[m_1 \phi_1^T(x)\right] \mu_{A_1^T}(dx)}{\exp\left[m_1 \phi_0^T(x)\right]},$$

$$K_j^T(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(j-1)}, d\boldsymbol{x}^{(j)}) \stackrel{\text{def}}{=} \frac{\exp\left[m_j \phi_j^T(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(j)})\right] \mu^{(T_j)}(d\boldsymbol{x}^{(j)})}{\exp\left[m_j \phi_{j-1}^T(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(j-1)})\right]}$$

where[5] $\boldsymbol{x}^{(j)} = (x_J)_{J \in T_j} \in S^{2^{j-1}}$ We also write

$$G_{j-1}^T \stackrel{\text{def}}{=} \left(G^T\right)^{(\mathcal{P}_{j-1}^T)} \equiv \gamma^T \otimes K_2^T \otimes \cdots \otimes K_{j-1}^T \tag{4.5}$$

for the marginal of G^T on coordinates corresponding to elements of $\mathcal{P}_{j-1}^T$.

[5]Notice that with this notation it holds in particular that $(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(j-1)}) = (x_J)_{J \in \mathcal{P}_{j-1}^T} \in S^{2^{j-1}-1}$.

In virtue of [3], the generalized Gibbs measure G^T is the unique solution to the Boltzmann-Gibbs variational principle to given chain: precisely, with

$$\boldsymbol{m} \stackrel{\text{def}}{=} \arg\min_{\boldsymbol{n}\in\Delta} P^T(\boldsymbol{n}) \tag{4.6}$$

the optimal parameter in Parisi variational principle associated to a T-chain, and recalling the Boltzmann-Gibbs variational principle (2.18), it follows from [3, Theorem 2.3-2.5] that

$$G^T(\boldsymbol{m}) = \arg \sup_{\nu\in\mathcal{M}_1^+(S^{2^n-1})} \left\{\Phi(\nu) - \mathfrak{J}^T(\nu)\right\} . \tag{4.7}$$

In particular, the measure $G^T(\boldsymbol{m})$ satisfies all constraints for the empirical measures in the associated Boltzmann-Gibbs principle for the chain T, i.e.

$$G^T(\boldsymbol{m}) \in \bigcap_{k=1}^{n} C_{A_k^T}. \tag{4.8}$$

Without loss of generality[6], we shall assume here that the *minimal chain*

$$\mathcal{T} \stackrel{\text{def}}{=} \arg\min_{T\in\mathcal{C}_n} f^T(\phi)\,, \tag{4.9}$$

i.e. the one which minimizes *all* Parisi variational principles, is given by

$$\mathcal{T} = \left\{A_k^{\mathcal{T}}\right\}_{k=0}^{n} \quad \text{with} \quad A_k^{\mathcal{T}} \stackrel{\text{def}}{=} \{1, 2, \cdots, k-1, k\}. \tag{4.10}$$

We now claim that the generalized Gibbs measure $G^{\mathcal{T}}_{\boldsymbol{m}'}$ associated to the minimal chain satisfies *all* constraints (2.5) or, which is the same, that

$$\boldsymbol{m}' \stackrel{\text{def}}{=} \arg\min_{\boldsymbol{m}\in\Delta} P^{\mathcal{T}}(\boldsymbol{m}) \Longrightarrow G^{\mathcal{T}}_{\boldsymbol{m}'} \in \bigcap_{J\in\mathcal{P}_n} C_J. \tag{4.11}$$

This amounts to saying that the measure $G^{\mathcal{T}}_{\boldsymbol{m}'}$ is a viable candidate for the Boltzmann-Gibbs variational principle for the nonhierarchical Perceptron GREM: Theorem 2.1 would then imply that

$$\begin{aligned} f(\phi) &\geq \int \phi dG^{\mathcal{T}}(\boldsymbol{m}') - H\left(G^{\mathcal{T}}(\boldsymbol{m}') \mid \mu\right) \\ &= \min_{T\in\mathcal{C}_n}\left\{\min_{m\in\Delta} P_T(\boldsymbol{m})\right\}, \end{aligned} \tag{4.12}$$

with the second line by definition/construction of $\mathcal{T}, \boldsymbol{m}'$, and Theorem 2.2 would follow.

It therefore remains to prove the key claim (4.11). This requires some preliminary observations. Recalling that ϕ_l^T depends on m_j only for $l \in \{1, \cdots, j-1\}$ a simple computation yields

$$\partial_j P^T(\boldsymbol{m}) = \frac{1}{m_j^2}\left[\int H\left(K_j^T(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(j-1)}, \cdot) \,|\, \mu^{(T_j)}\right) G_{j-1}^T(d\boldsymbol{x}^{(1)}, \cdots, d\boldsymbol{x}^{(j-1)}) - \frac{\log 2}{n}\right], \tag{4.13}$$

[6]Such form of the minimal chain can always be achieved by re-labelling.

cfr. also [3, eq. 3.13]. Using that $G^T_{j-1} \otimes K^T_j \equiv G^T_j$, the integral in the above can be written as

$$\begin{aligned}\int H\left(K^T_j(\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(j-1)},\cdot)\,|\,\mu^{(T_j)}\right)\,G^T_{j-1}(d\boldsymbol{x}^{(1)},\cdots,d\boldsymbol{x}^{(j-1)})\\ =\int \log\left(f(\boldsymbol{x}^{(j)}|\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(j-1)})\right)G^T_j(d\boldsymbol{x}^{(1)},\cdots,d\boldsymbol{x}^{(j)})\end{aligned} \tag{4.14}$$

where we shortened $f(\cdot|\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(j-1)})$ for the Radon-Nikodym derivative of the measure $K^T_j(\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(j-1)},\cdot)$ with respect to $\mu^{(T_j)}$. Following the steps (3.4)-(3.5) with $\mu^{(B^c)} := \mu^{(T_j)}$, $\nu^{(B^c|B)} := K^T_j$ (in which case $\nu^{(B)} := G^T_{j-1}$, $\nu := G^T_j$) we get

$$\int \log\left(f(\boldsymbol{x}^{(j)}|\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(j-1)})\right)G^T_j(d\boldsymbol{x}^{(1)},\cdots,d\boldsymbol{x}^{(j)}) = H\left(G^T_j\,|\,G^T_{j-1}\otimes\mu^{(T_j)}\right),$$

which implies

$$\partial_j P^T(\boldsymbol{m}) = \frac{1}{m_j^2}\left[H\left(G^T_j\,|\,G^T_{j-1}\otimes\mu^{(T_j)}\right) - \frac{\log 2}{n}\right]. \tag{4.15}$$

Coming back to our main task, a proof that the minimal chain satisfies all side constraints in the Boltzmann-Gibbs principle, let

$$J = \{j_1 < j_2 < \cdots < j_{M-1} < j_M\} \in \mathcal{P}_n \tag{4.16}$$

be a generic non-empty subset of I. The goal is thus to show that

$$H\left(\left(G^{\mathcal{T}}\right)^{(\mathcal{P}_J)}\,|\,\mu^{(\mathcal{P}_J)}\right) \le \frac{M}{n}\log 2. \tag{4.17}$$

Recall the definition (2.11) of the index k_J with the property that $J \in \mathcal{T}_{k_J}$: under (4.10) it holds that $\mathcal{T}_k = \mathcal{P}^{\mathcal{T}}_k \setminus \mathcal{P}^{\mathcal{T}}_{k-1} \equiv \mathcal{P}_{(1,\cdots,k)} \setminus \mathcal{P}_{(1,\cdots,k-1)}, k \in I$, hence

$$k_J = \max\{j:\, j \in J\} = j_M \;\ge\; M.$$

In light of (4.15), which shows how the conditions on the relative entropies relate to the partial derivatives of the Parisi functions, it is natural to distinguish two cases, depending on the latter being negative or positive.

First case: $\partial_{k_J} P^{\mathcal{T}}(\boldsymbol{m}') \le 0$. It then follows from (4.15)

$$H\left(G^{\mathcal{T}}_{k_J}\,|\,G^{\mathcal{T}}_{k_J-1}\otimes\mu^{(\mathcal{T}_{k_J})}\right) \le \frac{\log 2}{n}. \tag{4.18}$$

We distinguish two sub-cases:

i) When $M = 1$, namely when $J = \{j_1\}$ and $k_J = j_1$, we have $\mathcal{P}_J = \{J\} \subseteq \mathcal{T}_{k_J}$. In particular, the marginal of $G^{\mathcal{T}}_{k_J-1}\otimes\mu^{(\mathcal{T}_{k_J})}$ on $\mathcal{P}_J$ is $\mu^{(\mathcal{P}_J)} = \mu_J$ hence

$$\begin{aligned}H\left((G^{\mathcal{T}})^{(\mathcal{P}_J)}\,|\,\mu^{(\mathcal{P}_J)}\right) &= H\left((G^{\mathcal{T}})^{(\{J\})}\,|\,\mu_J\right)\\ &\overset{(3.2)}{\le} H\left(G^{\mathcal{T}}_{k_J}\,|\,G^{\mathcal{T}}_{k_J-1}\otimes\mu^{(\mathcal{T}_{k_J})}\right) \overset{(4.18)}{\le} \frac{\log 2}{n},\end{aligned} \tag{4.19}$$

which establishes (4.17) for J-s of cardinality one.

ii) For sets of cardinality $M = 2, \dots, n$ we proceed inductively: with

$$\mathfrak{O} \stackrel{\text{def}}{=} J \setminus \{k_J\} = \{j_1, j_2, \dots, j_{M-1}\},$$

we assume that

$$H\left((G^{\mathcal{T}})^{(\mathcal{P}_{\mathfrak{O}})} \,|\, \mu^{(\mathcal{P}_{\mathfrak{O}})}\right) \leq \frac{M-1}{n} \log 2 \,. \tag{4.20}$$

and claim that

$$H\left((G^{\mathcal{T}})^{(\mathcal{P}_J)} \,|\, \mu^{(\mathcal{P}_J)}\right) \leq \frac{M}{n} \log 2. \tag{4.21}$$

To see this we first observe that, evidently, $|\mathfrak{O}| = M - 1$, and

$$\mathcal{P}_{\mathfrak{O}} \subseteq \mathcal{P}^{\mathcal{T}}_{k_J - 1}, \quad \mathcal{P}_J \subseteq \mathcal{P}_{\mathfrak{O}} \cup \mathcal{T}_{k_J}, \qquad \mathcal{P}_{\mathfrak{O}} \cap \mathcal{T}_{k_J} = \emptyset,$$

hence by the monotonicity property (3.2) it holds

$$\begin{aligned} &H\left((G^{\mathcal{T}})^{(\mathcal{P}_J)} \,|\, \mu^{(\mathcal{P}_J)}\right) \leq H\left((G^{\mathcal{T}})^{(\mathcal{P}_{\mathfrak{O}} \cup \mathcal{T}_{k_J})} \,|\, \mu^{(\mathcal{P}_{\mathfrak{O}} \cup \mathcal{T}_{k_J})}\right) \\ &\stackrel{(3.1)}{=} H\left((G^{\mathcal{T}})^{(\mathcal{P}_{\mathfrak{O}})} \,|\, \mu^{(\mathcal{P}_{\mathfrak{O}})}\right) + H\left((G^{\mathcal{T}})^{(\mathcal{P}_{\mathfrak{O}} \cup \mathcal{T}_{k_J})} \,|\, (G^{\mathcal{T}})^{(\mathcal{P}_{\mathfrak{O}})} \otimes \mu^{(\mathcal{T}_{k_J})}\right) \\ &\stackrel{(4.20)}{\leq} \frac{M-1}{n} \log 2 + H\left((G^{\mathcal{T}})^{(\mathcal{P}_{\mathfrak{O}} \cup \mathcal{T}_{k_J})} \,|\, (G^{\mathcal{T}})^{(\mathcal{P}_{\mathfrak{O}})} \otimes \mu^{(\mathcal{T}_{k_J})}\right) . \end{aligned} \tag{4.22}$$

For the second summand above we write

$$\begin{aligned} H\left((G^{\mathcal{T}})^{(\mathcal{P}_{\mathfrak{O}} \cup \mathcal{T}_{k_J})} \,|\, (G^{\mathcal{T}})^{(\mathcal{P}_{\mathfrak{O}})} \otimes \mu^{(\mathcal{T}_{k_J})}\right) &\stackrel{(3.2)}{\leq} H\left(G^{\mathcal{T}}_{k_J} \,|\, G^{\mathcal{T}}_{k_J - 1} \otimes \mu^{(\mathcal{T}_{k_J})}\right) \\ &\stackrel{(4.18)}{\leq} \frac{\log 2}{n}, \end{aligned} \tag{4.23}$$

Plugging (4.23) into (4.22) therefore yields

$$H\left((G^{\mathcal{T}})^{(\mathcal{P}_J)} \,|\, \mu^{(\mathcal{P}_J)}\right) \leq \frac{M-1}{n} \log 2 + \frac{\log 2}{n} = \frac{M}{n} \log 2, \tag{4.24}$$

settling the induction step (4.21).

Second case: $\partial_{k_J} P^{\mathcal{T}}(\boldsymbol{m}') > 0$. It will become clear in the treatment that this is a somewhat degenerate situation where the minimal chain is not unique. As a matter of fact, we will use such lack of uniqueness to our advantage, in sofar we will deduce the validity of the side constraints from the properties of the Parisi functional evaluated in yet a second, well chosen (and also: minimal) chain. To see how this goes, we first claim that for $k \in \{2, \cdots, n\}$ the following implication holds true:

$$\partial_k P^{\mathcal{T}}(\boldsymbol{m}') > 0 \Longrightarrow \forall\, j = 1, \cdots, k-1 : \, m'_j = m'_{j+1} \,. \tag{4.25}$$

We will prove the following, fully equivalent implication: with

$$l = l(\mathcal{T}, k) \stackrel{\text{def}}{=} \min\left\{ \, j : j \in \{1, 2, \cdots, k-2, k-1\}; \;\; m'_j = m'_{j+1} \, \right\},$$

we claim that

$$\partial_k P^{\mathcal{T}}(\boldsymbol{m}') > 0 \Longrightarrow l = 1. \tag{4.26}$$

Proceeding by contradiction, let us assume that $l > 1$, in which case $l - 1 \geq 1$ and

$$m'_{l-1} < m'_l = m'_{l+1} = \cdots = m'_{k-1} = m'_k \leq m'_{k+1}$$

(with $m'_{n+1} = 1$). Since $\boldsymbol{m}' = \arg\min_{\boldsymbol{m}\in\Delta} P^{\mathcal{T}}(\boldsymbol{m})$, the function

$$m_k \in [m'_{l-1}, m'_{k+1}] \to P^{\mathcal{T}}(m'_1, \cdots, m'_{k-1}, m_k, m'_{k+1}, \cdots, m'_n)$$

assumes its minimum in $m_k = m'_k \in (m'_{l-1}, m'_{k+1}]$ and this implies $\partial_k P^{\mathcal{T}}(\boldsymbol{m}') \leq 0$, which contradicts our working assumption *unless* $l = 1$, hence

$$m'_1 = m'_2 = \cdots = m'_k, \tag{4.27}$$

and (4.26) is settled.

The next, and final, step is to show - as anticipated - that if $l(\mathcal{T}, k_J) = 1$ we may find yet another chain $\mathcal{R}$ with the same "Parisi free energy" as the chain $\mathcal{T}$, but from which we may also deduce the validity of the side constraints (4.17), namely that for $J = \{j_1 < j_2 < \cdots < j_M\}$,

$$H\left(\left(G^{\mathcal{T}}\right)^{(\mathcal{P}_J)} \mid \mu^{(\mathcal{P}_J)}\right) \leq \frac{M}{n}\log 2. \tag{4.28}$$

Towards this goal, define

$$\{b_1 < \cdots < b_{j_M - M}\} \stackrel{\text{def}}{=} A^{\mathcal{T}}_{j_M} \setminus J \equiv \{1, 2, \cdots, j_M - 1, j_M\} \setminus \{j_1, \cdots, j_M\},$$

denote by $\mathcal{R} = \left\{A^{\mathcal{R}}_k\right\}_{k=0}^n \in \mathcal{C}_n$ the chain with

$$A^{\mathcal{R}}_k \stackrel{\text{def}}{=} \begin{cases} (j_1, j_2, \cdots, j_{k-1}, j_k) & \text{if} \quad k \in \{1, \cdots, M\} \\ J \cup \{b_1, \cdots, b_{k-M}\} & \text{if} \quad k \in \{M+1, \cdots, j_M - 1\} \\ A^{\mathcal{T}}_k & \text{if} \quad k \in \{j_M, \cdots, n\} \end{cases}, \tag{4.29}$$

and consider the corresponding extremal parameter

$$\boldsymbol{m}'' = \text{argmin}_{\boldsymbol{m}\in\Delta} P^{\mathcal{R}}(\boldsymbol{m}).$$

In light of (4.27), with $k = k_J$ and since $k_J = j_M$ we have for the coordinates of the extremal parameter $\boldsymbol{m}'$ associated to $\mathcal{T}$ that $m'_1 = m'_2 = \cdots = m'_{j_M}$ and therefore

$$\frac{m'_l}{m'_{l+1}} = 1, \qquad l = 1, \cdots, j_M - 1.$$

Due to the recursive construction (2.16), and shortening $\boldsymbol{x}^{(j)} = (x_J)_{J\in\mathcal{T}_j}$, it follows that

$$\begin{aligned} \phi^{\mathcal{T}}_0(\boldsymbol{m}') &= \frac{1}{m'_1}\log\int e^{m'_1\phi^{\mathcal{T}}_1(\boldsymbol{x}^{(1)})}\mu^{(\mathcal{T}_1)}(d\boldsymbol{x}^{(1)}) \\ &= \frac{1}{m'_1}\log\int \exp\left[\frac{m'_1}{m'_2}\log\int e^{m'_2\phi^{\mathcal{T}}_2(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)})}\mu^{(\mathcal{T}_2)}(d\boldsymbol{x}^{(2)})\right]\mu^{(\mathcal{T}_1)}(d\boldsymbol{x}^{(1)}) \\ &= \frac{1}{m'_1}\log\int \exp\left[m'_2\phi^{\mathcal{T}}_2(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)})\right]\mu^{(\mathcal{T}_1)}\otimes\mu^{(\mathcal{T}_2)}(d\boldsymbol{x}^{(1)}, d\boldsymbol{x}^{(2)}). \end{aligned} \tag{4.30}$$

Again since $m'_1 = m'_2$, the r.h.s. of (4.30) equals

$$\begin{aligned} &\frac{1}{m'_1}\log\int \exp\left[m'_1\phi^{\mathcal{T}}_2(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)})\right]\mu^{(\mathcal{T}_1)}\otimes\mu^{(\mathcal{T}_2)}(d\boldsymbol{x}^{(1)}, d\boldsymbol{x}^{(2)}) \\ &\quad = \frac{1}{m'_1}\log\int \exp\left[m'_1\phi^{\mathcal{T}}_{j_M}(\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)},\cdots,\boldsymbol{x}^{(j_M)})\right]\bigotimes_{l=1}^{j_M}\mu^{(\mathcal{T}_l)}(d\boldsymbol{x}^{(l)}), \end{aligned} \tag{4.31}$$

the last equality by iteration. All in all,

$$\phi_0^{\mathcal{T}}(\boldsymbol{m}') = \frac{1}{m_1'} \log \int \exp\left[m_1' \phi_{j_M}^{\mathcal{T}}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(j_M)})\right] \bigotimes_{l=1}^{j_M} \mu^{(\mathcal{T}_l)}(d\boldsymbol{x}^{(l)}), \tag{4.32}$$

and in complete analogy for the $\mathcal{R}$-chain but this time with $\boldsymbol{x}^{(j)} \stackrel{\text{def}}{=} (x_J)_{J \in \mathcal{R}_j}$

$$\phi_0^{\mathcal{R}}(\boldsymbol{m}') = \frac{1}{m_1'} \log \int \exp\left[m_1' \phi_{j_M}^{\mathcal{R}}(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \cdots, \boldsymbol{x}^{(j_M)})\right] \bigotimes_{l=1}^{j_M} \mu^{(\mathcal{R}_l)}(d\boldsymbol{x}^{(l)}). \tag{4.33}$$

Since $\bigcup_{l=1}^{j_M} \mathcal{T}_l = \mathcal{P}_{j_M}^{\mathcal{T}} \equiv \mathcal{P}_{j_M}^{\mathcal{R}} = \bigcup_{l=1}^{j_M} \mathcal{R}_l$, the product measures in (4.32) and (4.33) are the same, to wit

$$\bigotimes_{l=1}^{j_M} \mu^{(\mathcal{R}_l)} = \bigotimes_{l=1}^{j_M} \mu^{(\mathcal{T}_l)} = \mu^{(\mathcal{P}_{j_M}^{\mathcal{T}})} = \bigotimes_{J \in \mathcal{P}_{\{1,\cdots,j_M\}}} \mu_J(dx_J)\,. \tag{4.34}$$

Furthermore, the form taken by the functions $\phi_{j_M}^{\mathcal{R}}(\boldsymbol{m}'), \cdots, \phi_n^{\mathcal{R}}(\boldsymbol{m}')$ and respectively $\phi_{j_M}^{\mathcal{T}}(\boldsymbol{m}'), \cdots, \phi_n^{\mathcal{T}}(\boldsymbol{m}')$ depends on the sets $\{A_k^{\mathcal{T}}\}_{k=J_M+1}^n$ only, respectively $\{A_k^{\mathcal{R}}\}_{k=J_M+1}^n$: as these sets coincide, it follows, in particular, that

$$\phi_{j_M}^{\mathcal{R}}(\boldsymbol{m}') \equiv \phi_{j_M}^{\mathcal{T}}(\boldsymbol{m}'). \tag{4.35}$$

Combining (4.32)-(4.35), we therefore see that

$$P^{\mathcal{T}}(\boldsymbol{m}') = \frac{\log 2}{n} \sum_{k=1}^n \frac{1}{m_k'} + \phi_0^{\mathcal{T}}(\boldsymbol{m}') = \frac{\log 2}{n} \sum_{k=1}^n \frac{1}{m_k'} + \phi_0^{\mathcal{R}}(\boldsymbol{m}') = P^{\mathcal{R}}(\boldsymbol{m}')\,, \tag{4.36}$$

and by similar reasoning,

$$G^{\mathcal{T}}(\boldsymbol{m}') = G^{\mathcal{R}}(\boldsymbol{m}'). \tag{4.37}$$

From (4.36) and minimality of the chain $\mathcal{T}$, i.e.

$$P^{\mathcal{T}}(\boldsymbol{m}') = \min_{\boldsymbol{m} \in \Delta} P^{\mathcal{T}}(\boldsymbol{m}) \leq \min_{\boldsymbol{m} \in \Delta} P^{\mathcal{R}}(\boldsymbol{m}) = P^{\mathcal{R}}(\boldsymbol{m}''),$$

if follows in particular that

$$P^{\mathcal{R}}(\boldsymbol{m}') = P^{\mathcal{R}}(\boldsymbol{m}''). \tag{4.38}$$

By uniqueness of the minimum of the Parisi principle, see [3, Prop. 3.6], it thus follows that $\boldsymbol{m}' = \boldsymbol{m}''$, which together with (4.37) implies

$$G^{\mathcal{T}}(\boldsymbol{m}') = G^{\mathcal{R}}(\boldsymbol{m}') = G^{\mathcal{R}}(\boldsymbol{m}'')\,. \tag{4.39}$$

Since $A_M^{\mathcal{R}} = J$, we have

$$\begin{aligned} H\left((G^{\mathcal{T}})^{(\mathcal{P}_J)} \,|\, \mu^{(\mathcal{P}_J)} \right) &= H\left([G^{\mathcal{T}}(\boldsymbol{m}')]^{(\mathcal{P}_J)} \,|\, \mu^{(\mathcal{P}_J)} \right) \\ &= H\left([G^{\mathcal{R}}(\boldsymbol{m}'')]^{(\mathcal{P}_M^{\mathcal{R}})} \,|\, \mu^{(\mathcal{P}_M^{\mathcal{R}})} \right) \leq \frac{M}{n} \log 2, \end{aligned} \tag{4.40}$$

the last inequality since $G^{\mathcal{R}}(\boldsymbol{m}'') \in C_{A_M^{\mathcal{R}}}$, i.e. the measure $G^{\mathcal{R}}(\boldsymbol{m}'')$ satisfies the side constraints on the empirical measures. This concludes the proof of Theorem 2.2.

□

References

[1] Bolthausen, Erwin, and Nicola Kistler. *On a nonhierarchical version of the generalized random energy model.* The Annals of Applied Probability 16.1 (2006): 1-14.

[2] Bolthausen, Erwin, and Nicola Kistler. *On a nonhierarchical version of the generalized random energy model. II. Ultrametricity.* Stoch. Proc. Appl. Vol. 119 (2009): 2357-2386.

[3] 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 PROMS 11 (2012): 425-442.

[4] Derrida, Bernard. *Random-energy model: An exactly solvable model of disordered systems.* Physical Review B 24.5 (1981): 2613.

[5] Derrida, Bernard. *A generalization of the random energy model that includes correlations between the energies.* J. Phys. Lett. 46 (1985): 401-407.

[6] Mézard, Marc, Giorgio Parisi, and Miguel A. Virasoro. *Spin glass theory and beyond.* World Scientific, Singapore (1987).

[7] Ghirlanda, Stefano, and Francesco Guerra. *General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity.* Journal of Physics A: Mathematical and General 31.46 (1998): 9149.

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

[9] Panchenko, Dmitry. *The Sherrington-Kirkpatrick model.* Springer Science & Business Media, 2013.

[10] Panchenko, Dmitry. *The Parisi ultrametricity conjecture.* Annals of Mathematics (2013): 383-393.

[11] 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.

[12] Polyanskiy, Yuri, and Yihong Wu. *Lecture notes on Information Theory*, MIT University Press (2017)

[13] Subag, Eliran. *The geometry of the Gibbs measure of pure spherical spin glasses.* Invent. Math., 210(1):135–209, 2017

[14] Subag, Eliran. *The free energy of spherical pure p-spin models – computation from the TAP approach*, arXiv preprint: 2101.04352

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

[16] Talagrand, Michel *The Parisi formula.* Annals of Mathematics Second Series, Vol. 163, No. 1 (2006): 221-263.

[17] 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`

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