ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

On the Phase Diagram of the Multiscale Mean-Field Spin-Glass

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Camilli, F., Contucci, P., Mingione, E., Tantari, D. (2026). On the Phase Diagram of the Multiscale Mean-Field Spin-Glass. ANNALES HENRI POINCARE', 27(7), 2649-2684 [10.1007/s00023-025-01584-w].

Availability:
This version is available at: https://hdl.handle.net/11585/1028447 since: 2025-11-09

Published:
DOI: http://doi.org/10.1007/s00023-025-01584-w

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

16 September 2026

On the phase diagram of the multiscale mean-field spin-glass

Francesco Camilli[1*], Pierluigi Contucci[2], Emanuele Mingione[2], Daniele Tantari[2†]

[1]Quantitative Life Sciences, International Centre for Theoretical Physics, Str. Costiera 11, Trieste, 34151, Italy.
[2]Dipartimento di Matematica, Alma Mater Studiorum - Università di Bologna, Via Zamboni 33, Bologna, 40126, Italy.

*Corresponding author(s). E-mail(s): fcamilli@ictp.it;
Contributing authors: pierluigi.contucci@unibo.it;
emanuele.mingione2@unibo.it; daniele.tantari@unibo.it;
†Authors are ordered in alphabetical order and they all contributed equally.

Abstract

In this paper we study the phase diagram of a Sherrington-Kirkpatrick (SK) model where the couplings are forced to thermalize at different time scales. Besides being a challenging generalization of the SK model, such settings may arise naturally in physics whenever part of the many degrees of freedom of a system relaxes to equilibrium considerably faster than the others. For this model we compute the asymptotic value of the second moment of the overlap distribution. Furthermore, we provide a rigorous sufficient condition for an annealed solution to hold, identifying a high temperature, or weak coupling, region. In addition, we also prove that for sufficiently strong couplings the solution must present a number of replica symmetry breaking levels at least equal to the number of time scales already present in the multiscale model. Finally, we give a sufficient condition for the existence of gaps in the support of the functional order parameters.

Keywords: Multiscale Spin-Glasses, Replica Symmetry Breaking, Disordered Systems, Statistical Mechanics

1 Introduction

The Sherrington-Kirkpatrick (SK) model [1–6] is a spin system with random interactions characterized by two main features. Firstly the interaction is of mean field nature. Secondly the randomness is *quenched*, i.e. the interaction couplings are random parameters and not thermodynamic degrees of freedom. This means that the thermodynamics is described by the so called *quenched measure*: given a realization of the disorder the spins thermalize according to the Boltzmann-Gibbs distribution, and the interaction randomness is treated with a successive statistical average.

In the *multiscale* SK model [7] part of the random interactions become thermodynamic degrees of freedom and the quenched measure is replaced by a *multiscale measure*, obtained by the following procedure. The interactions are divided in a finite number of families equilibrating in a hierarchical succession. Recursively, each family thermalizes at increasing temperature according to a Boltzmann-Gibbs distribution, where the effective Hamiltonian is given by the free energy associated to the equilibrium of the previous family. The above recursive construction is deeply related to Derrida-Ruelle probability cascades [8–11]. A formal definition of the multiscale measure will be given in Section 2.

It is worth to stress that a multiscale measure does not describe a standard thermodynamic equilibrium but rather an out of equilibrium scenario. Indeed, it can be viewed as the stationary measure of a dynamical system in which different degrees of freedom, say $(x_\ell)_{1\le\ell\le r}$, are coupled to thermal baths at different temperatures and evolve on widely separated timescales, $\tau_1 \gg \ldots \gg \tau_r$. At a given time $t \sim \tau_\ell$, the dynamics can be effectively described on the variable x_ℓ, where the faster degrees of freedom $(x_{\ell'})_{\ell'>\ell}$ have already equilibrated, and the slower ones $(x_{\ell'})_{\ell'<\ell}$ are effectively frozen. The stationary measure of this effective dynamics is then obtained through the same recursive procedure outlined above. A concrete example of such dynamics is discussed in Remark 4. For a thorough discussion of the physical aspects, we refer to [12–16], and for a mathematical analysis to [17]. In a broader sense the multiscale measure can be used to describe out of equilibrium systems in the limit of small entropy production [18]: here the different temperatures are defined as the ratio between correlation and response functions, generalizing the classical fluctuation-dissipation relation [19–21].

The multiscale measure spontaneously emerges in the *Replica Symmetry Breaking* (RSB) solution of SK model. More specifically, the Parisi formula [6] can be obtained from the free energy of a system with non-interacting spins subject to multiscale Gaussian random external fields with a special correlation structure [5, 22, 23]. The Parisi formula implies that at low temperature RSB occurs [24–26] and the physical behavior of the model becomes, to some extent, close to that of out of equilibrium systems [2, 27]. We also mention that mean-field models where the external field is governed by a multiscale measure have been recently considered in a series of works [28–31] leading to a formulation of the Parisi solution in terms of Hamilton-Jacobi equations.

Multiscale models also arise in high dimensional inference when multiple steps of inference procedures are concatenated. We mention as examples: expectation-maximization approaches, reconstruction tasks where a part of the hidden signal

has been revealed [32, 33], or matrix inference models [34]. Similar structures also emerge in the teacher-student analysis where the multibath property is related to a possible mismatch between inference and generating temperatures [35–38]. Multiscale multibath hierarchical procedures are also common in the context of machine learning, especially in the area of transfer learning and fine tuning techniques analysis [39–41].

In [7] the multiscale SK model was shown to obey a Parisi-like variational principle. In this work we study its solution and the related thermodynamic properties. In particular we describe how the multiscale structure affects the distribution of the order parameter. Contrary to what happens in the SK model, here the system's behavior can be characterized through the overlap distribution at all the different time scales. In physical terms the overlap at the ℓ-th scale is obtained by measuring the scalar product between the spin configurations of two system's replicas evolved up to the ℓ-th equilibrating time from the same initialization. In this situation part of the interactions has evolved while the rest is still frozen.

Based on the above physical picture the overlap is expected to decrease at increasing equilibrating time, since the greater the differences in interactions between the two replicas, the more distinctly their spin configurations will evolve. Furthermore, according to the conventional ultrametric picture of the states, at low temperature different timescales are expected to correspond to different ranges of overlap values, which are related to the corresponding different levels of the ultrametric tree [2, 5, 42]. It is also reasonable to expect that, at sufficiently high temperature, all the ℓ-averages of the overlap vanish and the system is completely annealed. Interestingly, this model can include a partial annealing mechanism in which the overlap is different from zero up to a certain timescale, beyond which it then becomes zero.

This heuristic picture is confirmed and specified by the results of this work. We start by showing how the ℓ-th scale average of the overlap is related to the minimizer of the Parisi formula by the synchronization mechanism [43–45] and in Theorem 3 we compute it. Following that, in Theorem 4 we characterize the high-temperature region of the model where in absence of external magnetic fields the free energy is yielded by an annealed computation, namely by considering all interactions as thermodynamic degrees of freedom at the same time scale of the spins. In Theorem 5 we give a sufficient low temperature type condition to have at least as many RSB levels as the time scales originally present in the model. Finally in Theorem 6 we characterize the overlap distribution across the various scales providing a sufficient condition for the existence of gaps in its support.

The paper is organized as follows. In Section 2 the SK multiscale measure is introduced and the Parisi like variational principle for the multiscale pressure is recalled. In Section 3 the main results are presented and proved in Section 4. In Section 5 we present a synthetic picture of our finding together with conclusions and perspectives.

2 Definitions

Let us first define the pressure of the multiscale Sherrington-Kirkpatrick (SK) model [7]. Consider an integer $r \geq 1$ and two sequences $\zeta = (\zeta_\ell)_{\ell \leq r}$ and $\gamma = (\gamma_\ell)_{\ell \leq r}$ such that

$$0 = \zeta_{-1} < \zeta_0 < \cdots < \zeta_r = 1 \tag{1}$$

$$0 = \gamma_0 < \gamma_1 < \cdots < \gamma_r < \infty, \tag{2}$$

and a set of N spins $\sigma \equiv (\sigma_i)_{i \leq N} \in \{-1,1\}^N$ interacting via a random Hamiltonian

$$H_N(\sigma) = \sum_{\ell=1}^{r} H_N^{(\ell)}(\sigma). \tag{3}$$

Analogously to the standard SK model, we take each of the Hamiltonians $H_N^{(\ell)}$ to be a 2^N-dimensional Gaussian process indexed by the spin configurations, identified by the following covariance

$$\mathbb{E} H_N^{(\ell)}(\sigma) H_N^{(\ell')}(\tau) = N \delta_{\ell\ell'} (\gamma_\ell^2 - \gamma_{\ell-1}^2) q_N^2(\sigma, \tau) \tag{4}$$

for any $\sigma, \tau \in \{-1,1\}^N$, where we have introduced the overlap

$$q_N(\sigma, \tau) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i \tau_i \,. \tag{5}$$

The same model can be equivalently represented in terms of independent and identically distributed standard Gaussians $g_{ij}^{(\ell)}$:

$$H_N^{(\ell)}(\sigma) = \sqrt{\gamma_\ell^2 - \gamma_{\ell-1}^2} \sum_{i,j=1}^{N} \frac{g_{ij}^{(\ell)}}{\sqrt{N}} \sigma_i \sigma_j \,. \tag{6}$$

The thermodynamic pressure of the multiscale SK model is defined by recursive integration of the randomness in the different Hamiltonians $H_N^{(\ell)}$ at different scales. More specifically, define the backwards recursion

$$Z_{\ell-1,N}^{\zeta_{\ell-1}} = \mathbb{E}_{\ell-1} Z_{\ell,N}^{\zeta_{\ell-1}}, \quad \mathbb{E}_{\ell-1} = \mathbb{E}_{(g_{ij}^{(\ell)})_{i,j \leq N}} \tag{7}$$

for any $\ell = 1, \ldots, r$, with starting point

$$Z_{r,N} = \sum_{\sigma \in \{-1,1\}^N} e^{-H_N(\sigma) - \sum_{i=1}^N h_i \sigma_i}, \tag{8}$$

where $h_i \sim P_h$ are i.i.d. copies of a compactly supported quenched random variable. We stress that even if (4) looks like a sum of independent Gaussians, the recursion (7) induces non trivial dependencies, as later illustrated.

Given a realization of h, $Z_{r,N}$ depends on all the randomness, i.e. $(g_{ij}^{(\ell)})_{i,j\le N}^{\ell\le r}$, while $Z_{\ell,N}$ depends only on the randomness up to level ℓ, that is $(g_{ij}^{(p)})_{i,j\le N}^{p\le \ell}$. We are now ready to define the main object under investigation:

Definition 1 (Pressure per particle)**.** The pressure per particle of the multiscale SK model is

$$p_N = \frac{1}{N}\mathbb{E}_h \log Z_{0,N}\,. \tag{9}$$

Remark 1. For $r = 1$ and $\zeta_0 > 0$ the model was already studied in [46, 47], while for $\zeta_0 \to 0$ the quantity p_N reduces to the quenched pressure of an SK model at inverse temperature $\beta = \gamma_r$.

Remark 2. One can also define the ℓ-level pressure $P_{\ell,N} = \log Z_{\ell,N}$, then the recursion (7) rewrites as

$$e^{\zeta_{\ell-1} P_{\ell-1,N}} = \mathbb{E}_{\ell-1} e^{\zeta_{\ell-1} P_{\ell,N}}. \tag{10}$$

The above relation is common in renormalization group approach to field theory (see [48]).

Remark 3. The pressure (9) can be also written as the quenched pressure of an auxiliary system with configuration space $\{1,-1\}^N \times \mathbb{N}^r$

$$p_N = \frac{1}{N}\,\mathbb{E} \log \sum_{\alpha\in\mathbb{N}^r} \nu_\alpha \sum_{\sigma\in\{-1,1\}^N} e^{H_N(\sigma,\alpha)} \tag{11}$$

where $H_N(\sigma,\alpha)$ is a suitable centered Gaussian process $(\nu_\alpha)_{\alpha\in\mathbb{N}^r}$ are random probabilistic weights associated to a Ruelle Probability Cascade [5, 7].

2.1 The multiscale measure

The pressure (9) is the generating functional of the Hamiltonians $(H_N^{(\ell)})_{\ell\le r}$ with respect to the *multiscale measure*. In particular, the recursion (7) implies that the average w.r.t. the multiscale measure is obtained by a sequence of Boltzmann-Gibbs averages of the different degrees of freedom, each of them performed at a proper temperature and effective potential. In the current setting the degrees of freedom are the spins $\sigma \in \{-1,1\}^N$ and the collection of couplings $(g^{(\ell)})_{\ell\le r}$ where $g^{(\ell)} = (g_{ij}^{(\ell)})_{i,j\le N}$. The spins are the fastest variables (i.e. first to thermalize) and if $\ell > \ell'$ $g^{(\ell)}$ is faster than $g^{(\ell')}$. Finally the external field $(h_i)_{i\le N}$ is completely quenched, namely one has to average it at the end of the thermalization procedure.

The above picture is formally defined as follows. The spins thermalize according to the standard Boltzmann-Gibbs measure given a realization of $g = (g^{(\ell)})_{\ell\le r}$ and h,

namely for any function $A(\sigma)$ we set

$$\langle A \rangle_N = \sum_{\sigma \in \{-1,1\}^N} \frac{e^{-H_N(\sigma)+\sum_{i\le N} h_i\sigma_i}}{Z_{r,N}} A(\sigma). \tag{12}$$

The average of the remaining degrees of freedom is taken using a suitable collection of probability weights. For any $\ell \le r$ we set

$$f_{\ell,N} := \frac{Z_{\ell,N}^{\zeta_{\ell-1}}}{\mathbb{E}_{\ell-1} Z_{\ell,N}^{\zeta_{\ell-1}}}, \quad 1 \le \ell \le r. \tag{13}$$

Notice that $(f_\ell)_{\ell\le r}$ are random probabilistic weights, in particular f_ℓ depends on the families $(g^{(\ell')})_{\ell'\le\ell}$ and h.

Definition 2. Given $\ell \in \{0,\dots,r\}$ and a measurable function $A(\sigma,(g^{(\ell')})_{\ell<\ell'\le r})$ we define its *ℓ-th scale average* as

$$\langle A \rangle_N^{(\ell)} = \mathbb{E}_\ell \mathbb{E}_{\ell+1} \dots \mathbb{E}_{r-1} f_{\ell+1,N} \dots f_{r,N} \langle A \rangle_N \tag{14}$$

with the convention $\langle \cdot \rangle_N^{(r)} \equiv \langle \cdot \rangle_N$.

The above definition can be equivalently rewritten in terms of the recursion

$$\langle \cdot \rangle_N^{(\ell-1)} = \mathbb{E}_{\ell-1} f_{\ell,N} \langle \cdot \rangle_N^{(\ell)}. \tag{15}$$

$\langle A \rangle_N^{(\ell)}$ is random through $(g^{(\ell')})_{\ell'\le\ell}$ and h; from a probabilistic point of view it can be viewed as the conditional expectation given $(g^{(\ell')})_{\ell'\le\ell}$ and h.

Remark 4. The recursion (15) can be interpreted as defining a multiscale measure induced by a generic Hamiltonian H of a system with $r+1$ families of degrees of freedom. Let us denote these families by $x = (x_\ell)_{\ell\le r+1}$ and the average w.r.t. the randomness of x_ℓ by $\mathbb{E}_{\ell-1}$. The multiscale measure induced by H is obtained by (15) starting from

$$\langle \cdot \rangle^{(r)} = \frac{1}{\mathbb{E}_r e^{-\beta H(x)}} \mathbb{E}_r e^{\beta H(x)}(\cdot). \tag{16}$$

As mentioned in the introduction, there is a dynamical interpretation of the above construction. Let us assume to deal with continuous degrees of freedom and consider the following system of Langevin equations:

$$\tau_\ell \, \mathrm{d}x_\ell = -\partial_\ell H \, \mathrm{d}t + \sqrt{\frac{2\tau_\ell}{\beta_\ell}} \, \mathrm{d}W_\ell, \quad \ell = 1,\dots,r+1 \tag{17}$$

where $\tau_\ell, \beta_\ell > 0$, and $\{W_\ell\}_{\ell\le r+1}$ are independent Wiener processes. Then, under the asymptotic regime $\frac{\tau_\ell}{\tau_{\ell+1}} \to \infty$, one can show (see [17, 49]) that the stationary measure

of the system converges to the recursive multiscale structure given by (15), with the identifications $\beta_{r+1} = \beta$ and $\zeta_\ell = \beta_{\ell+1}/\beta$.

The order parameter of the model turns out to be [7] the distribution of overlap (5), which is a function of two spin configurations that in physical jargon are called *replicas*. Clearly one has to specify from which distribution spin configurations are sampled. In fact, in the multiscale setting there are $r+1$ different ways to sample them.

Definition 3. Given $\ell \in \{0,\dots,r\}$ we denote by $\mu_N^{(\ell)}$ the measure in (14). Let $A_{1,\dots,n} \equiv A(\sigma^{(1)},\dots,\sigma^{(n)})$ be a function of n spin configurations that are sampled independently from $\mu_N^{(\ell)}$, the *ℓ-th level replicated average* is defined as

$$\langle A_{1,\dots,n} \rangle_N^{(\ell)} = \int \prod_{a=1}^{n} \mu_N^{(\ell)}(d\sigma^{(a)}) A(\sigma^{(1)},\dots,\sigma^{(n)}). \tag{18}$$

Some remarks are in order. Recall that, as for the single replica average (14), $\langle A \rangle_N^{(\ell)}$ is random through $(g^{(\ell')})_{\ell' \le \ell}$ and h. Second, each replica $\sigma^{(a)}$ shares the same "outer" disorder $(g^{(\ell')})_{\ell' \le \ell}$, but they all come with their own replicas of the couplings for the levels $(g^{(\ell')})_{\ell' > \ell}$. In this sense, the multiscale model presents a fundamental difference with the standard SK.

We are mostly interested in the case $n = 2$ and $A_{12} \equiv q_N(\sigma^1,\sigma^2)$ where q_N is overlap (5). Hence, for any $\ell \in \{0,\dots,r\}$ and bounded function ψ, we have the expectation

$$\mathbb{E}\langle\langle \psi(q_{1,2})\rangle_N^{(\ell)}\rangle^{(0)} = \mathbb{E}_h \prod_{p=1}^{\ell} f_{p,N} \int \mu_N^{(\ell)}(d\sigma^1)\mu_N^{(\ell)}(d\sigma^2)\psi(q_N(\sigma^1,\sigma^2)). \tag{19}$$

The average (19) can be divided into three steps. Given $(g^{(\ell')})_{\ell' \le \ell}$ we first take the average value of the overlap between two spin configuration sampled from $\mu_N^{(\ell)}$. The result is a random function of $(g^{(\ell')})_{\ell' \le \ell}$ which is integrated using the tilted measure $\langle\cdot\rangle_N^{(0)}$. Finally we take the expectation $\mathbb{E}_h$ to average out the quenched external field.

The connection between the ℓ-scale average (14) and the multiscale pressure is essentially due to the recursion (7) (see [22]). As an example one can compute the various contributions to the *internal energy* coming from $(H_N^{(\ell)})_{\ell \le r}$. Setting $\beta_\ell = \sqrt{\gamma_\ell^2 - \gamma_{\ell-1}^2}$ one obtains

$$\frac{\partial p_N}{\partial \beta_\ell} = -\mathbb{E}\prod_{s=1}^{r} f_{s,N} \left\langle \frac{H_N^{(\ell)}}{\beta_\ell N} \right\rangle_N = -\mathbb{E}_h \left\langle \frac{H_N^{(\ell)}}{\beta_\ell N} \right\rangle_N^{(0)}. \tag{20}$$

Since $H_N^{(\ell)}$ is a Gaussian one can apply integration by parts to rewrite the r.h.s. of (20) as

$$\frac{\partial p_N}{\partial \beta_\ell} = \beta_\ell \Big(1 - \sum_{p=\ell}^{r} (\zeta_p - \zeta_{p-1}) \, \mathbb{E}_h \Big\langle \langle q_{12}^2 \rangle_N^{(p)} \Big\rangle_N^{(0)} \Big) \tag{21}$$

where $\langle \cdot \rangle_N^{(p)}$ is the p-level replicated average defined in (18). Moreover the quantity $\frac{\partial p_N}{\partial \gamma_\ell}$ gives the contribution to the internal energy due to the ℓ-average. From (21) one obtains

$$\frac{\partial p_N}{\partial \gamma_\ell} = \begin{cases} -\gamma_\ell (\zeta_\ell - \zeta_{\ell-1}) \, \mathbb{E}_h \Big\langle \langle q_{12}^2 \rangle_N^{(\ell)} \Big\rangle_N^{(0)}, & \text{if} \quad \ell = 1, \dots, r-1 \\ \gamma_r \Big(1 - (1 - \zeta_{r-1}) \, \mathbb{E}_h \Big\langle \langle q_{12}^2 \rangle_N^{(r)} \Big\rangle_N^{(0)} \Big) & \text{if} \quad \ell = r. \end{cases} \tag{22}$$

2.2 The variational formula

The limiting value as $N \to \infty$ of the pressure (9) exists and it can be represented as the solution of an infinite dimensional variational problem [7]. Let us introduce it first. Consider an integer $k \geq r$ and two sequences

$$0 = \xi_{-1} \leq \xi_0 \leq \cdots \leq \xi_k = \xi_{k+1} = 1 \tag{23}$$

$$0 = x_0 \leq x_1 \leq \cdots \leq x_k \leq x_{k+1} = 1 \,, \tag{24}$$

where $\xi = (\xi_j)_{j=0,\dots,k+1} \supseteq \zeta = (\zeta_\ell)_{\ell=0,\dots,r}$. Given $(\gamma_\ell)_{\ell \leq r}$ in (2) we define

$$\tilde{\gamma}_j = \gamma_\ell \quad \text{if} \quad j \in K_\ell = \{ j : \zeta_{\ell-1} < \xi_j \leq \zeta_\ell \} \quad \text{for some } \ell \in \{0, \dots, r\} \tag{25}$$

and the one-body Hamiltonian

$$-\tilde{H}(\sigma) = \sigma \Big(\sqrt{2} \sum_{j=1}^{k+1} \eta_j \sqrt{\tilde{\gamma}_j^2 x_j - \tilde{\gamma}_{j-1}^2 x_{j-1}} + h \Big) = \sigma \Big(\sqrt{2} \sum_{\ell=0}^{r} \gamma_\ell \sum_{j \in K_\ell} \eta_j \sqrt{x_j - x_{j-1}} + h \Big), \tag{26}$$

where $\sigma = \pm 1$, $\eta_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, and $h \sim P_h$. Starting from

$$\tilde{Z}_{k+1} := \sum_{\sigma = \pm 1} e^{-\tilde{H}(\sigma)} = 2 \cosh \Big(\sqrt{2} \sum_{j=1}^{k+1} \eta_j \sqrt{\tilde{\gamma}_j^2 x_j - \tilde{\gamma}_{j-1}^2 x_{j-1}} + h \Big) \tag{27}$$

we define again the backwards recursion

$$\tilde{Z}_{j-1}^{\xi_{j-1}} = \mathbb{E}_{j-1} \tilde{Z}_j^{\xi_{j-1}}, \quad \mathbb{E}_{j-1} = \mathbb{E}_{\eta_j}, \tag{28}$$

for $j = 1,\dots,k+1$. This finally allows to introduce the *Parisi functional* for the multiscale model as

$$\mathcal{P}(x,\xi) = \log \tilde{Z}_0 - \frac{1}{2}\sum_{j=0}^{k} \xi_j\Big((\tilde{\gamma}_{j+1}x_{j+1})^2 - (\tilde{\gamma}_j x_j)^2\Big) = \log \tilde{Z}_0 - \frac{1}{2}\sum_{\ell=0}^{r}\gamma_\ell^2 \sum_{j\in K_\ell} \xi_j(x_{j+1}^2 - x_j). \tag{29}$$

Out of convenience we introduce the following sets of allowed sequences

$$\mathcal{M} = \bigcup_{k\geq r} \mathcal{M}_k, \quad \mathcal{M}_k := \{(x,\xi) \text{ verifying (23)-(24)}, \xi \supseteq \zeta, \text{card}(x) = k+2\}. \tag{30}$$

We can now state the main result of [7], as it plays a central role in our analysis.

Theorem 1. *The thermodynamic limit of the quenched pressure density of the Multiscale SK (MSK) model (9) exists and is given by the infinite dimensional variational principle*

$$\lim_{N\to\infty} p_N = \inf_{(x,\xi)\in\mathcal{M}} \mathcal{P}(x,\xi). \tag{31}$$

The proof of this theorem requires a non-trivial extension of the Guerra replica symmetry breaking upper bound [22] and the use of the synchronization property [43] for Ruelle Probability Cascades (RPCs) for the lower bound. We refer the reader to [7] for the details. Here we aim at assessing some fundamental properties of the optimization over $\mathcal{M}$. The problem of uniqueness of the solution is not addressed here. However, since Theorem 1 has many analogies with the Parisi formula for the SK model where uniqueness holds [24, 50], it is reasonable to expect the same here.

3 Results

3.1 The order parameter and synchronization

In this section we analyze the continuity properties of $\mathcal{P}$ seen as a function over $\mathcal{M}$. We shall see that, contrary to the plain SK model, in this multiscale version $\mathcal{P}$ is not only a function of the probability distribution associated to the sequences (x,ξ), but it preserves some memory of the scales originally present, identified by the sequences (γ,ζ), in a sense that will be rigorously specified.

Let us denote by Pr the space of probability measures supported on $[0,1]$. Define Pr_k, for k integer, the set of probability measures supported on $k+1$ points on $[0,1)$. Any $\mu \in \text{Pr}_k$ can be identified with a pair of strictly increasing sequences (y,m) such that

$$0 = m_{-1} < m_0 < \cdots < m_k = 1 \tag{32}$$

$$0 = y_0 < y_1 < \ldots\ldots < y_k < 1\,. \tag{33}$$

The mapping to such distribution is realized by the following

$$\mu(y_i) = m_i - m_{i-1}\,, \quad i = 0,\dots,k\,. \tag{34}$$

Definition 4 (Quantile)**.** Let ρ be a probability measure supported on $[0,a]$ for some $a > 0$. The quantile function associated to ρ is

$$\rho^{-1}(p) := \inf\{s \in [0,a] : \rho([0,s]) \geq p\}\,, \quad p \in [0,1]. \tag{35}$$

Consider a k' integer and $\mu \equiv (y,m) \in \mathrm{Pr}_{k'}$. Set $k := |m \cup \zeta| \geq k'$. Note that the sequence m can already contain some of the ζ's, or even all of them. We define a pair $(x^\mu, \xi^\mu) \in \mathcal{M}_k$ associated to μ by

$$\begin{cases} \xi^\mu & \text{take } m \cup \zeta \text{ increasingly ordered and set } \xi^\mu_{k+1} = 1 \\ x^\mu_j & = \mu^{-1}(\xi^\mu_j), \quad \text{for } j = 0\dots,k \text{ and set } x^\mu_{k+1} = 1\,. \end{cases} \tag{36}$$

The fact that x^μ is defined through the quantile function μ^{-1} implies that the

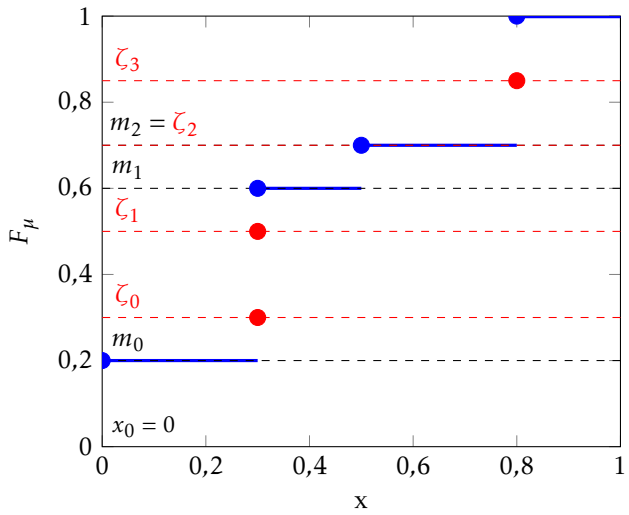

Figure 1 Example of construction of the sequence (x^μ, ξ^μ) associated to a probabilty measure μ with CDF F_μ plotted in blue. In this example we have $k = 7$, $m_0 < \zeta_0, \zeta_1 < m_1$, $\zeta_2 = m_2$, and $m_2 < \zeta_3 < 1$. It results in the sequence $x^\mu = \{0, y_1, y_1, y_1, y_2, y_3, y_3\}$, that corresponds to the horizontal coordinates of the colored circles in the figure. Blue circles are due to the elements already present in the couple of sequences (m,y), whereas red circles come from the newly introduced elements which form ξ. As apparent from this figure, repetitions occur all the times we are adding a ζ to the sequence m which was not already contained in the latter.

elements of the sequence x^μ are those of the sequence y with possibly some repetitions. There are no repetitions in the x^μ's iff $m \supseteq \zeta$. On the other hand, if m does

not contain, say, $\zeta_\ell = \xi^\mu_{j_\ell}$ for a given $\ell = 0,\dots,r$, then by definition of the quantile $\mu^{-1}(\xi^\mu_{j_\ell}) = x^\mu_{j_\ell} = x^\mu_{j_\ell+1}$. In other words, if there are repetitions, they occur in correspondence to those ζ_ℓ's missing from the sequence m (see Fig.1 for an example).

This reasoning is necessary precisely because the Parisi functional $\mathcal{P}$ does not depend only on the probability distribution identified by the pair (x,ξ), but also on which of the x's repeat. As we shall clarify later, if x_{j+1} and x_j collapse on one another, but $\xi_j \notin \zeta$, then the Parisi functional becomes independent on ξ_j. It is by all means as if we had reduced the sequences by one element. This however is no longer true if $\xi_j \in \zeta$. Following definition (29) it is indeed not difficult to see that all the ζ's appear anyway in the functional, as they should.

That being said, (36) defines a map $\bigcup_{k\geq 1}\mathrm{Pr}_k \ni \mu \to (x^\mu,\xi^\mu) \in \mathcal{M}$. We set

$$\bar{\mathcal{P}}(\mu) := \mathcal{P}(x^\mu,\xi^\mu)\,, \qquad \mu \in \bigcup_{k\geq 1}\mathrm{Pr}_k\,. \tag{37}$$

We can now state a proposition concerning the Lipschitz continuity of $\bar{\mathcal{P}}$, and relating the latter to the Parisi functional. Its proof is deferred to Section 4.2.

Proposition 2. *Given $\mu_1 \in \mathrm{Pr}_{k_1}$ and $\mu_2 \in \mathrm{Pr}_{k_2}$ one has*

$$|\,\bar{\mathcal{P}}(\mu_1) - \bar{\mathcal{P}}(\mu_2)\,| \leq L W_1(\mu_1,\mu_2) \tag{38}$$

for some constant L independent from μ_1 and μ_2 and

$$W_1(\mu_1,\mu_2) := \int_0^1 |\,\mu_1^{-1}(p) - \mu_2^{-1}(p)\,|\,dp\,. \tag{39}$$

Therefore one can extend continuously $\bar{\mathcal{P}}$ to all $\mu \in \mathrm{Pr}$ and

$$\inf_{(x,\xi)\in\mathcal{M}} \mathcal{P}(x,\xi) = \inf_{\mu\in\mathrm{Pr}} \bar{\mathcal{P}}(\mu). \tag{40}$$

Remark 5. Notice also that by definition (25) one has $\tilde{\gamma}_j = 0$ for all j such that $0 < \xi_j \leq \zeta_0$, hence these ξ_j don't play any role in $\mathcal{P}(x,\xi)$ and then, without loss of generality, one can assume that $\xi_0 = \zeta_0$. In other words one say that

$$\inf_{\mu\in\mathrm{Pr}} \bar{\mathcal{P}}(\mu) = \inf_{\mu\in\mathrm{Pr}^{(0)}} \bar{\mathcal{P}}(\mu) \quad \text{where} \quad \mathrm{Pr}^{(0)} = \{\mu \in \mathrm{Pr} : \lim_{x\to 0^+} \mu([0,x]) \geq \zeta_0\}. \tag{41}$$

The sequences (x^μ,ξ^μ) and the associated $\tilde{\gamma}$ defined in (25) can be related to a pair of *synchronized* random variables using a construction similar to [29, 51]. We start identifying the sequences (γ,ζ) in (1) and (2) with a r.v. Γ setting

$$\mu_\Gamma(\gamma_\ell) := \mathbb{P}(\Gamma = \gamma_\ell) = \zeta_\ell - \zeta_{\ell-1}\,,\ \forall\ 0 \leq \ell \leq r.$$

Definition 5. Let $U \sim \mathrm{Unif}[0,1]$. Given a random variable Y supported on $[0,1)$ with law μ, we denote by ρ_μ the probability measure $[0,1]\times[0,\gamma_r]$ identified by

$$\rho_\mu = \mathrm{Law}(C) \quad \text{where} \quad C \equiv \Big(\mu^{-1}(U), \mu_\Gamma^{-1}(U)\Big) \tag{42}$$

or equivalently in terms of CDF

$$F_C(c_1,c_2) := \min\Big(\mu([0,c_1]), \mu_\Gamma([0,c_2])\Big), \quad (c_1,c_2)\in[0,1]\times[0,\gamma_r]. \tag{43}$$

For any $\ell \le r$ we denote by $\rho_{\mu,\ell}$ the conditional probability

$$\rho_{\mu,\ell}(A) := \int_A \rho_\mu(dx \,|\, \Gamma = \gamma_\ell) = \frac{1}{\zeta_\ell - \zeta_{\ell-1}} \int_A \rho_\mu(dx, \Gamma = \gamma_\ell). \tag{44}$$

From the above definition, follows that the marginals of C are Γ and Y, but its components are strongly correlated random variables: they are deterministic functions of the same uniform random variable U. This construction is known as *synchronization* [43] or monotone coupling [51]. One can check that if $\mu \in \mathrm{Pr}_{k'}$ is a discrete probability measure, then the map (36) and the associate $\tilde\gamma$ coincide with the following definition

$$\tilde\gamma_j = \mu_\Gamma^{-1}(\xi_j^\mu) \quad \text{and} \quad x_j^\mu = \mu^{-1}(\xi_j^\mu). \tag{45}$$

Furthermore, averages w.r.t. $\rho_{\mu,\ell}$ rewrite in terms of ξ and x as follows

$$\begin{aligned}\int_0^1 f(x)\rho_\mu(dx \,|\, \Gamma = \gamma_\ell) &= \frac{1}{\zeta_\ell - \zeta_{\ell-1}} \int_0^1 f(x)\rho_\mu(dx, \Gamma = \gamma_\ell) \\ &= \frac{1}{\zeta_\ell - \zeta_{\ell-1}} \mathbb{E} f(\mu^{-1}(U)) \mathbb{1}(\mu_\Gamma^{-1}(U) = \gamma_\ell).\end{aligned} \tag{46}$$

Notice that $\mathbb{1}(\mu_\Gamma^{-1}(U) = \gamma_\ell)$ is non-zero only if $U \in (\zeta_{\ell-1}, \zeta_\ell]$. Said interval can be decomposed as the union of $(\xi_{j-1}, \xi_j]$ for $j \in K_\ell$, over which $\mu^{-1}(U)$ is constant and equal to x_j. Hence the above turns into

$$\int_0^1 f(x)\rho_\mu(dx \,|\, \Gamma = \gamma_\ell) = \frac{1}{\zeta_\ell - \zeta_{\ell-1}} \sum_{j\in K_\ell} (\xi_j - \xi_{j-1}) f(x_j). \tag{47}$$

3.2 Overlap moments

By Proposition 2 we know that a solution of the variational problem (31) is a probability measure on $[0,1]$. Here we link the above solution to the distribution of the overlap (5) w.r.t. the large N limit of the multiscale measure (18). Let us recall that in the SK model the following holds (see [5, Theorem 3.7], and [52]): the optimizer

of the Parisi formula can be used to obtain the second moment of the overlap in the large N limit of the quenched measure, which is in turn connected to the internal energy of the system. Our next result contains an analogous statement for the multiscale SK model. We stress that in the multiscale setting the picture is more involved since, as observed (see Section 2.1), there are several possible measures for the overlap. One can consider for instance all the ℓ-th scale averages $\mathbb{E}_h\left\langle \langle q_N^2\rangle_N^{(\ell)}\right\rangle_N^{(0)}$, $\ell = 1,\dots,r$.

Theorem 3 (Internal energy and overlap moments)**.** *For any $\ell = 1,\dots,r$*

$$\lim_{N\to\infty} \mathbb{E}_h\left\langle \langle q_N^2\rangle_N^{(\ell)}\right\rangle_N^{(0)} = \int x^2\, \rho_{\mu^*}(dx \,|\, \Gamma = \gamma_\ell). \tag{48}$$

where μ^* is some Parisi measure solving (40).

There is a natural ordering in the moments of the overlap induced by the recursion (15) and the fact that $f_{\ell,N}$ are probability tilts (namely $\mathbb{E}_{\ell-1} f_{\ell,N} = 1$). More precisely, we have

$$\begin{aligned}\mathbb{E}_h\left\langle \langle q_N^2\rangle_N^{(\ell)}\right\rangle_N^{(0)} &= \frac{1}{N} + \frac{N-1}{N}\mathbb{E}_h\left\langle (\langle\sigma_1\sigma_2\rangle_N^{(\ell)})^2\right\rangle_N^{(0)} = \frac{1}{N} + \frac{N-1}{N}\mathbb{E}_h\left\langle (\mathbb{E}_\ell f_{\ell+1}\langle\sigma_1\sigma_2\rangle_N^{(\ell+1)})^2\right\rangle_N^{(0)} \\ &\le \frac{1}{N} + \frac{N-1}{N}\mathbb{E}_h\left\langle (\langle\sigma_1\sigma_2\rangle_N^{(\ell+1)})^2\right\rangle_N^{(0)} = \mathbb{E}_h\left\langle \langle q_N^2\rangle_N^{(\ell+1)}\right\rangle_N^{(0)}\end{aligned} \tag{49}$$

where in the last step we have reabsorbed $\mathbb{E}_\ell f_{\ell+1}$ in $\langle\cdot\rangle_N^{(0)}$ and used Jensen's inequality. Hence one should expect that the r.h.s. of (48) respects the same ordering. This is indeed the case thanks to the fact that ρ_μ is generated through a monotone coupling. Hence if the r.v. Γ takes higher values, so does the corresponding moment of the overlap.

Note that, whereas the points of the supports of $\rho_{\mu^*}(x \mid \Gamma = \gamma_\ell)$ are disjoint for different ℓ's because of the monotone coupling, this does not imply that the finite size multiscale measures $\mathbb{E}_h\langle\langle\cdot\rangle_N^{(\ell)}\rangle_N^{(0)}$ also have disjoint and ordered supports. That being said, if we stick to the interpretation that $\rho_{\mu^*}(x \mid \Gamma = \gamma_\ell)$ is the asymptotic conditional average for the ℓ-th scale, Theorem 3 is telling us that the overlaps become synchronized with Γ and the supports of the asymptotic ℓ-th scale measures separate.

3.3 The annealed regime

Our next result concerns the annealed solution of the model. Indeed, we prove that in absence of external magnetic fields an annealed solution holds iff a high-temperature condition is fulfilled:

Theorem 4 (Annealed region)**.** *The annealed solution is exact whenever $h = 0$ almost surely and $\gamma_r^2 \le \frac{1}{2}$, namely*

$$\lim_{N\to\infty} p_N = \log 2 + \frac{\gamma_r^2}{2} \text{ if and only if } \gamma_r^2 \le \frac{1}{2} \text{ and } h = 0 \text{ a.s.} \tag{50}$$

As we shall see, a consequence of Theorem 3 and 4 is that in the annealed region

$$\lim_{N\to\infty} \mathbb{E}_h \left\langle \langle q_N^2 \rangle_N^{(\ell)} \right\rangle_N^{(0)} = 0, \tag{51}$$

for all $\ell \in \{1,\dots,r\}$. The annealed solution is an obvious upper bound for the pressure per particle thanks to Jensen's inequality. The proof of Theorem 4 is reported in Section 4.4. However, proving it is exact requires a direct use of the high-temperature result for the standard SK model an absence of external magnetic field [3].

3.4 Strong coupling regime

The annealed condition (50), namely when all the couplings $(\gamma_\ell)_{\ell\le r}$ are weak, implies that all the ℓ-averages of the overlap are zero (51), and thus all the mass of the Parisi measure is concentrated at 0. Conversely, if the interactions are strong enough, this is no longer true. In particular we provide a sufficient strong coupling condition such that the distribution μ^*, where the infimum of $\mathcal{P}$ is attained, has at least r distinguished atoms. Specifically, the (finite or infinite) sequence x^*, associated to μ^*, must have at least r distinguished values. In this sense, the model must have at least r levels of replica symmetry breaking.

Theorem 5. *Assume that*

$$1 - 2\gamma_1^2 \zeta_1^2 < 0 \tag{52}$$

or $\mathbb{E}_h h^2 > 0$. *Let* $\mu^* \in Pr$ *be a Parisi measure solving* (40). *If* μ^* *is supported on a finite number of points then* $\zeta \subseteq \xi^*$ *and*

$$\lim_{N\to\infty} \mathbb{E}_h \left\langle \langle q_N^2 \rangle_N^{(\ell)} \right\rangle_N^{(0)} > 0 \quad \forall 1 \le \ell \le r. \tag{53}$$

Condition (52) is actually the strongest among the set of conditions $\left(\gamma_\ell^2 > \frac{1}{2\zeta_\ell^2}\right)_{\ell\le r}$ involved in Lemma 12 later proved in Section 4.5. Informally, what Lemma 12 is saying is that if $\gamma_\ell^2 > \frac{1}{2\zeta_\ell^2}$ (in absence of an external magnetic field) and the measure μ^* is finitely supported, then it must have an atom detached from the origin, and it is the one in correspondence of the element ζ_ℓ fo the sequence ζ. This fact will be also crucial to assert that $\zeta_\ell \in \xi^*$.

By comparing (51) and (53), one may wonder if there exists a region of the parameters where only a subset of the ℓ-averages are non-zero, indicating a mechanism of *partial annealing*. Theorem 5, and in particular Lemma 12 are indeed leaving this possibility open. Strictly speaking, if $\gamma_\ell^2 > \frac{1}{2\zeta_\ell^2}$ the only thing we can conclude is that $\lim_{N\to\infty} \mathbb{E}_h \left\langle \langle q_N^2 \rangle_N^{(\ell)} \right\rangle_N^{(0)} > 0$, but we were not able to prove also the converse statement, as is instead possible for the total annealing dealt with in Theorem 4. We

finally notice that Theorem 5 does not rule out the possibility of full RSB, namely that μ^* is supported on non-discrete set.

3.5 Lower bound on gaps sizes

The final result is a sufficient condition for which, at fixed ℓ, the CDF associated to μ^* exhibits a plateau at height ζ_ℓ. This can be identified as a discontinuity in the associated quantile function μ^{*-1}. Recalling that the quantile is a non-decreasing and left-continuous function, we define for any $\nu \in Pr$ the quantity

$$\Delta_\ell(\nu) := \nu^{-1}(\zeta_\ell^+) - \nu^{-1}(\zeta_\ell) \tag{54}$$

where $\nu^{-1}(\zeta_\ell^+) := \lim_{p\to\zeta_\ell^+} \nu^{-1}(p)$. If $\Delta_\ell(\nu) \geq C$ for some positive $C > 0$ then the CDF of ν has a plateau at height ζ_ℓ of length at least C and hence a gap in its support. This possibility is stated in the following theorem, where μ^* is assumed to be the W_1-limit of a sequence of k-stationary pairs, introduced in Definition 6 below.

Theorem 6. *Given $\ell = 1,\dots,r-1$, if $\zeta_\ell\,\gamma_{\ell+1}^2 < 1/2$ then*

$$\Delta_\ell(\mu^*) \geq 2\left(\gamma_{\ell+1}^2 - \gamma_\ell^2\right)\mu^{*-1}(\zeta_\ell)\left(\int_{\mu^{*-1}(\zeta_\ell^+)}^1 \mu^*([0,x])\,dx\right)^2. \tag{55}$$

Theorem 6 can be interpreted as a result for the support of the conditional overlap distribution $\rho_{\mu^*,\ell}$. In fact it implies that there is a gap between the support of $\rho_{\mu^*,\ell}$ and that of $\rho_{\mu^*,\ell-1}$. Note that the result is informative only outside the annealed region, provided that $\mu^{*-1}(\zeta_\ell) > 0$ and $\mu^{*-1}(\zeta_\ell^+) < 1$. Notice also that if $\zeta_\ell\,\gamma_{\ell+1}^2 < 1/2$ for some ℓ then $\zeta_{\ell'}\,\gamma_{\ell'+1}^2 < 1/2$ for all $\ell' < \ell$.

4 Proofs

4.1 Preliminaries

Let us start with some basic properties of the Parisi functional (29).

Definition 6. The pair of sequences $(\bar{x},\bar{\xi}) \in \mathcal{M}_k$ is called k-stationary pair if

$$\inf_{(x,\xi)\in\mathcal{M}_k} \mathcal{P}(x,\xi) = \mathcal{P}(\bar{x},\bar{\xi}). \tag{56}$$

In analogy with (13) we introduce the random probability weights

$$f_j := \frac{\tilde{Z}_j^{\xi_{j-1}}}{\mathbb{E}_{j-1}\tilde{Z}_j^{\xi_{j-1}}}, \quad 1 \leq j \leq k+1, \tag{57}$$

where the $\tilde{Z}_j$'s are defined in (28). Recall that f_j depends on the remaining randomness in $\eta_1,\dots,\eta_j$. We denote the j-th scale average as

$$\langle A\rangle^{(j)} = \mathbb{E}_j\mathbb{E}_{j+1}\dots\mathbb{E}_k f_{j+1}\dots f_{k+1}\langle A\rangle, \quad 0\le j\le k+1 \tag{58}$$

where

$$\langle A\rangle = \frac{1}{\tilde{Z}_{k+1}}\sum_{\sigma=\pm1} A(\sigma,(g^{(\ell')})_{\ell<\ell'\le r}))e^{-\tilde{H}(\sigma)}. \tag{59}$$

Hence $\langle\cdot\rangle^{(j)}$ is still random through $\eta_1,\dots,\eta_j$, but $\eta_{j+1},\dots,\eta_{k+1}$ have been averaged out. By definition (58) one has

$$\langle\cdot\rangle^{(j-1)} = \mathbb{E}_{j-1}f_j\langle\cdot\rangle^{(j)}. \tag{60}$$

Moreover the recursion (28) implies the following

Proposition 7. *Denote by ∂ a generic derivative w.r.t. a variable in $\tilde{Z}_{k+1}$, be it $\tilde{\gamma}$ or x, and by $\mathbb{E}$ the expectation over all the disorder $\eta_1,\dots,\eta_{k+1}$. Then*

$$\partial\log\tilde{Z}_0 = \mathbb{E}\prod_{j=1}^{k+1} f_j\frac{1}{\tilde{Z}_{k+1}}\partial\tilde{Z}_{k+1}. \tag{61}$$

Furthermore, consider $p>j$:

$$\partial_{\eta_j}f_j = \sqrt{2(\tilde{\gamma}_j^2x_j-\tilde{\gamma}_{j-1}^2x_{j-1})}f_j\langle\sigma\rangle^{(j)}\xi_{j-1} \tag{62}$$

$$\partial_{\eta_j}f_p = \sqrt{2(\tilde{\gamma}_j^2x_j-\tilde{\gamma}_{j-1}^2x_{j-1})}f_p\xi_{p-1}\Big(\langle\sigma\rangle^{(p)}-\langle\sigma\rangle^{(p-1)}\Big). \tag{63}$$

For its proof we refer to [22]. Here we prove a specialization of the above for the Parisi functional.

Lemma 8. *The gradient components of the Parisi potential w.r.t. x read:*

$$\partial_{x_j}\mathcal{P}(x,\xi) = \tilde{\gamma}_j^2(\xi_j-\xi_{j-1})\Big[x_j-\mathbb{E}\prod_{p=1}^{j} f_p\cdot(\langle\sigma\rangle^{(j)})^2\Big] \tag{64}$$

where

$$\langle\sigma\rangle^{(j)} = \mathbb{E}_j\dots\mathbb{E}_k f_{j+1}\dots f_{k+1}\tanh z(\eta), \quad z(\eta) := \sum_{j=1}^{k+1}\eta_j\sqrt{2\Big(\tilde{\gamma}_j^2x_j-\tilde{\gamma}_{j-1}^2x_{j-1}\Big)}+h. \tag{65}$$

Proof. Let us start with the second term in (29):

$$\partial_{x_j} \frac{1}{2} \sum_{0\le p\le k} \xi_p\Big((\tilde\gamma_{p+1}x_{p+1})^2-(\tilde\gamma_p x_p)^2\Big) = (\xi_{j-1}-\xi_j)\tilde\gamma_j^2 x_j. \tag{66}$$

Using Proposition 7, we can derive the functional $\log \tilde Z_0$:

$$\begin{aligned}
\partial_{x_j}\log\tilde Z_0 &= \mathbb{E}\prod_{p=1}^{k+1} f_p \tanh(z(\eta))\Big[\frac{\tilde\gamma_j^2\eta_j}{\sqrt{2\big(\tilde\gamma_j^2 x_j-\tilde\gamma_{j-1}^2x_{j-1}\big)}}-\frac{\tilde\gamma_j^2\eta_{j+1}}{\sqrt{2\big(\tilde\gamma_{j+1}^2x_{j+1}-\tilde\gamma_j^2x_j\big)}}\Big]\\
&= \tilde\gamma_j^2\mathbb{E}\prod_{p=1}^{k+1} f_p\Big[1-\tanh^2(z(\eta))\Big]-\tilde\gamma_j^2\mathbb{E}\prod_{p=1}^{k+1} f_p\Big[1-\tanh^2(z(\eta))\Big]\\
&\quad+\frac{\tilde\gamma_j^2}{\sqrt{2\big(\tilde\gamma_j^2 x_j-\tilde\gamma_{j-1}^2x_{j-1}\big)}}\sum_{p\ge j}^{k+1}\mathbb{E}f_1\dots\partial_{\eta_j}f_p\dots f_{k+1}\cdot\tanh(z(\eta))\\
&\quad-\frac{\tilde\gamma_j^2}{\sqrt{2\big(\tilde\gamma_{j+1}^2x_{j+1}-\tilde\gamma_j^2x_j\big)}}\sum_{p\ge j+1}^{k+1}\mathbb{E}f_1\dots\partial_{\eta_{j+1}}f_p\dots f_{k+1}\cdot\tanh(z(\eta))\Big].
\end{aligned} \tag{67}$$

Using the formulae for derivatives (62)-(63) we get

$$\begin{aligned}
\partial_{x_j}\log\tilde Z_0 &= \tilde\gamma_j^2\mathbb{E}f_1\dots f_{k+1}\langle\sigma\rangle\sum_{p\ge j}^{k+1}\langle\sigma\rangle^{(p)}(\xi_{p-1}-\xi_p)-\tilde\gamma_j^2\mathbb{E}f_1\dots f_{k+1}\langle\sigma\rangle\sum_{p\ge j+1}^{k+1}\langle\sigma\rangle^{(p)}(\xi_{p-1}-\xi_p)\\
&= -\tilde\gamma_j^2(\xi_j-\xi_{j-1})\mathbb{E}f_1\dots f_{k+1}\langle\sigma\rangle\langle\sigma\rangle^{(j)} = -\tilde\gamma_j^2(\xi_j-\xi_{j-1})\mathbb{E}f_1\dots f_j(\langle\sigma\rangle^{(j)})^2.
\end{aligned} \tag{68}$$

In the last step we used the fact that $\langle\sigma\rangle^{(j)}$ is independent on the r.v.'s $\eta_{j+1},\dots,\eta_{k+1}$. Putting the two contributions together the proof is complete. □

The lemma that follows characterizes stationary pairs.

Lemma 9. *Consider $(\bar x,\bar\xi)\in\mathcal{M}_k$ a k-stationary pair, such that all the $\bar\xi_j$'s are different without loss of generality. Then the following consistency equations hold*

$$\bar x_j = \mathbb{E}\prod_{p=1}^{j} f_p(\langle\sigma\rangle^{(j)})^2 \tag{69}$$

for all $j=1,\dots,k$.

Proof. The boundary of the optimization set of the x's is identified by the various possible matchings $x_j = x_{j-1}$. Select a $j=2,\dots,k$. If $\bar x_j$ realizes an infimum point, it

must satisfy the following conditions:

$$\begin{cases} \partial_{x_j}\mathcal{P}(x,\xi)\Big|_{(\bar{x},\bar{\xi}):\bar{x}_j=\bar{x}_{j-1}} \geq 0 \\ (\bar{x}_j-\bar{x}_{j-1})\partial_{x_j}\mathcal{P}(\bar{x},\bar{\xi})=0 \end{cases} . \tag{70}$$

In the hypothesis $\bar{x}_{j-1}$ has collapsed on $\bar{x}_j$ one must also add

$$\partial_{x_{j-1}}\mathcal{P}(x,\xi)\Big|_{(\bar{x},\bar{\xi}):\bar{x}_{j-1}=\bar{x}_j} \leq 0 . \tag{71}$$

If $\bar{x}_j > \bar{x}_{j-1}$ then the statement is automatically proved.

Say instead $0=\bar{x}_j-\bar{x}_{j-1}$. From the third of the above equations, using (64), we infer that

$$\bar{x}_{j-1}=\bar{x}_j \leq \mathbb{E}\prod_{p=1}^{j-1} f_p(\langle\sigma\rangle^{(j-1)})^2 = \mathbb{E}\prod_{p=1}^{j-1} f_p(\mathbb{E}_{j-1}f_j\langle\sigma\rangle^{(j)})^2 \leq \mathbb{E}\prod_{p=1}^{j} f_p(\langle\sigma\rangle^{(j)})^2 \tag{72}$$

where we used (60) and Jensen inequality. The only way for the above to be compatible with the first condition in (70) is to have $\partial_{x_j}\mathcal{P}(x,\xi)\Big|_{(\bar{x},\bar{\xi}):\bar{x}_j=\bar{x}_{j-1}}=0$. Therefore, even if $\bar{x}_j$ is an extremal point on the boundary, it must still satisfy the fixed point equations. □

Lemma 10. *Let $\mu\in\mathrm{Pr}_{k'}$ and assume that its associated pair $(x^\mu,\xi^\mu)\in\mathcal{M}_k$ is k-stationary, then one has*

$$\frac{\partial}{\partial\gamma_\ell}\mathcal{P}(x^\mu,\xi^\mu)=\begin{cases} -\gamma_\ell\sum_{j\in K_\ell}(\xi^\mu_j-\xi^\mu_{j-1})(x^\mu_j)^2, & \text{if} \quad \ell=1,\dots,r-1 \\ \gamma_r\Big(1-\sum_{j\in K_r}(\xi^\mu_j-\xi^\mu_{j-1})(x^\mu_j)^2\Big), & \text{if} \quad \ell=r \end{cases} . \tag{73}$$

Equivalently denoting by ρ_μ the law of the the synchronized pair of random vector defined in (42) one has

$$\frac{\partial}{\partial\gamma_\ell}\bar{\mathcal{P}}(\mu)=\begin{cases} -\gamma_\ell(\zeta_\ell-\zeta_{\ell-1})\int x^2\rho_\mu(dx|\Gamma=\gamma_\ell), & \text{if} \quad \ell=1,\dots,r-1 \\ \gamma_r\Big(1-(1-\zeta_{r-1})\int x^2\rho_\mu(dx|\Gamma=\gamma_r)\Big), & \text{if} \quad \ell=r \end{cases} . \tag{74}$$

Proof. The proof of (73) proceeds as the one of Lemma 8, with the difference that we exploit the fact that (x^μ,ξ^μ) is a stationary pair. It is indeed easy to verify through Karush–Kuhn–Tucker conditions that even if some x's lie at the boundary, e.g. $x^\mu_{j+1}=x^\mu_j$ for some j's, they are still stationary points for the Parisi functional, namely (64) always equals zero for a stationary pair. The equivalence with (74) is proved by (47). □

4.2 Proof of Proposition 2

Let us start with a simple observation. If $(x,\xi) \in \mathcal{M}_k$ is such that $\xi_j = \xi_{j-1}$ for some $j \le k$, then $\mathcal{P}(x,\xi) \equiv \mathcal{P}(x,\xi)^-$ where $(x,\xi)^- = (x,\xi) \setminus (x_j,\xi_j) \in \mathcal{M}_{k-1}$. On the other hand assume that $x_j = x_{j+1}$ for some j, then there are two cases: if $\tilde\gamma_j = \tilde\gamma_{j+1}$ then $\mathcal{P}(x,\xi) = \mathcal{P}(x,\xi)^-$ where $(x,\xi)^-$ is again obtained dropping x_j and ξ_j.

On the other hand, if $\tilde\gamma_{j+1} > \tilde\gamma_j$, that occurs when $\xi_j = \zeta_\ell$ for some $\ell \in \{0,\dots,r\}$, and still $x_{j+1} = x_j$, then none of the ξ can be dropped. Whenever repetitions of the x's occur in correspondence of an index j such that $\xi_j \in \zeta$ the pair (x,ξ) cannot be simplified. We thus call a pair (x,ξ) *minimal* iff it has only such repetitions in the sequence x. As noticed in Section 3.1 minimal pairs are precisely the ones generated from discrete distributions $\mu \in \mathrm{Pr}'_k$.

Let us now move to the proof of the Lipschitz continuity of $\bar{\mathcal{P}}$. For $\alpha = 1,2$ take $\mu_\alpha = (y^\alpha, m^\alpha) \in \mathrm{Pr}_{k_\alpha}$ and denote by (x^α,ξ^α) the associated sequence in $\mathcal{M}$. As mentioned earlier, such sequences (x^α,ξ^α) are minimal.

Let $\xi = (\xi_j)_{j\le k^*+1}$ be the strictly increasing sequence obtained by ordering $\xi^1 \cup \xi^2$ where $k^*+1 = |\xi^1 \cup \xi^2|$. For $\alpha = 1,2$ define also an increasing sequence $\tilde x^\alpha = (\tilde x^\alpha_j)_{j\le k^*+1}$ where

$$\tilde x^\alpha_j = \mu_\alpha^{-1}(\xi_j), \quad \tilde x^\alpha_{k^*+1} = 1\,. \tag{75}$$

From the above definition, it is clear that the image of μ_α^{-1} must be contained in x^α. Therefore the above operation introduces yet again other repetitions in the $\tilde x^\alpha_j$. In particular $\tilde x^\alpha_{j+1} = \tilde x^\alpha_j$ implies $\xi_j \in \xi^\alpha$. We also denote by $\bar\gamma$ the associated sequence (25) through $\mu_\Gamma^{-1}(\xi)$.

For any $t \in [0,1]$ consider $x(t) = (x_j(t))_{j\le k^*+1}$ by

$$x_j(t) = t\,\tilde x^1_j + (1-t)\,\tilde x^2_j\,. \tag{76}$$

Therefore $(x(t),\xi) \in \mathcal{M}_{k^*}$ and we can define $\phi(t) = \mathcal{P}(x(t),\xi)$. It is not difficult to check that

$$\phi(1) = \mathcal{P}(\tilde x^1,\xi) = \mathcal{P}(x^1,\xi^1) = \bar{\mathcal{P}}(\mu_1) \tag{77}$$

$$\phi(0) = \mathcal{P}(\tilde x^2,\xi) = \mathcal{P}(x^2,\xi^2) = \bar{\mathcal{P}}(\mu_2)\,. \tag{78}$$

Using formula (64) one obtains

$$\phi'(t) = \sum_{j\le k^*+1} \tilde\gamma^2_j(\xi_j - \xi_{j-1})(\tilde x^1_j - \tilde x^2_j)C_j(t) \tag{79}$$

where $|C_j(t)| \le 2$. Recall that $\tilde x^1_j = \mu_1^{-1}(\xi_j)$ and $\tilde x^2_j = \mu_2^{-1}(\xi_j)$ one obtains

$$|\phi'(t)| \le 2\gamma_r^2 \sum_{j\le k^*+1} (\xi_j - \xi_{j-1})|\mu_1^{-1}(\xi_j) - \mu_2^{-1}(\xi_j)| = 2\gamma_r^2 \int_0^1 |\mu_1^{-1}(p) - \mu_2^{-1}(p)|dp \tag{80}$$

and hence (38) follows.

In order to prove (40) it is enough to show that

$$\left\{\bar{\mathcal{P}}(\mu) \mid \mu \in \bigcup_{k\geq 1} \mathrm{Pr}_k\right\} = \left\{\mathcal{P}(x,\xi) \mid (x,\xi) \in \mathcal{M}\right\}. \tag{81}$$

Clearly because of the map (36) we have the inclusion $\subseteq$ in (81). In order to prove the converse is enough to show that for any $(x,\xi) \in \mathcal{M}_k$ there exists $k' \leq k$ and $\mu \in \mathrm{Pr}_{k'}$ such that $\mathcal{P}(x,\xi) = \bar{\mathcal{P}}(\mu)$. Clearly one can assume without loss that (x,ξ) is minimal and then it's easy to check that the desired μ is simply the law of the random variable Y identified by $\mathbb{P}(Y = x_j) = \xi_j - \xi_{j-1}$.

4.3 Proof of Theorem 3

In this proof we will use the same idea of [52]. We want to exploit convexity properties of the pressure p_N in (9) with respect to each component of $\beta = (\beta_\ell)_{\ell\leq r}$ where $\beta_\ell = \sqrt{\gamma_\ell^2 - \gamma_{\ell-1}^2}$. Indeed β_ℓ plays the role of an inverse temperature in the definition (6) of H_N^ℓ. Convexity is inherited by the function

$$\mathcal{P}(\gamma(\beta)) := \inf_{\mu\in\mathrm{Pr}} \bar{\mathcal{P}}(\mu, \gamma(\beta)) \tag{82}$$

since it is the limit of the sequence of convex functions p_N. We are going to show that $\mathcal{P}(\gamma(\beta))$ is differentiable in each of the β_ℓ. Fix some $\ell \leq r$ and denote by $\partial_\ell \mathcal{P}$ the subdifferential of $\mathcal{P}$ w.r.t. β_ℓ. Thanks to convexity it is enough to show that $\partial_\ell \mathcal{P}$ is a singleton. Let $(x^{(k)}, \xi^{(k)})_{k\in\mathbb{N}}$ be a sequence of k-stationary pairs and denote by $\mu_k \in \mathrm{Pr}_k$ the associated probability measure, namely μ_k is the law of a random variable Y_k with $\mathbb{P}(Y_k = x_j^{(k)}) = \xi_j^{(k)} - \xi_{j-1}^{(k)}$. By Lipschitz continuity in Proposition 2 we have that $\lim_{k\to\infty} \bar{\mathcal{P}}(\mu_k) = \mathcal{P}(\gamma(\beta))$ and the limit is approached monotonically. Then there exists a sequence $\epsilon_k \to 0$ such that

$$0 \leq \bar{\mathcal{P}}(\mu_k, \gamma(\beta)) - \mathcal{P}(\gamma(\beta)) \leq \epsilon_k. \tag{83}$$

Now if $a \in \partial_\ell \mathcal{P}$ repeating the proof of [52, Theorem 1], one obtains

$$a = \frac{\partial \bar{\mathcal{P}}}{\partial \beta_\ell}(\mu_k, \gamma(\beta)) + \mathcal{O}\left(\sqrt{\varepsilon_k}\right). \tag{84}$$

The above also requires a uniform bound on the second derivative of the Parisi functional w.r.t. β_ℓ, that can be dealt with as in [52].

Recall that if μ_k is stationary then $\frac{\partial}{\partial \gamma_p}\bar{\mathcal{P}}(\mu_k, \gamma(\beta))$ is given in Lemma 10. Then using the relation

$$\frac{\partial}{\partial \beta_\ell} = \beta_\ell \sum_{p\geq \ell}^{r} \frac{1}{\gamma_p} \frac{\partial}{\partial \gamma_p} \tag{85}$$

we get

$$a = \beta_\ell \Big\{ 1 - \sum_{p\geq \ell}^{r} (\zeta_p - \zeta_{p-1}) \int x^2 \rho_{\mu_k}(dx|\Gamma = \gamma_p) \Big\} + \mathcal{O}(\sqrt{\varepsilon_k}) \,. \tag{86}$$

Now since $\bar{\mathcal{P}}$ is continuous there exists a subsequence (μ_{k_n}) of (μ_k) such that $\mu_{k_n} \xrightarrow{W_1} \mu^*$ and $\mathcal{P}(\gamma(\beta)) = \bar{\mathcal{P}}(\mu^*, \gamma(\beta))$, namely μ^* is some Parisi measure. By definition of ρ_μ

$$\begin{aligned} \Big| \int x^2 \Big(\rho_{\mu_k}(dx, \Gamma = \gamma_p) - \rho_{\mu^*}(dx, \Gamma = \gamma_p) \Big) \Big| &\leq \mathbb{E}_U \Big| (\mu_{k_n}^{-1}(U))^2 - (\mu^{*-1}(U))^2 \Big| \mathbb{1}(\mu_\Gamma^{-1}(U) = \gamma_p) \\ &\leq 2 W_1(\mu_{k_n}, \mu^*) \xrightarrow{n\to\infty} 0 \,, \end{aligned} \tag{87}$$

where we used that $\mathbb{1}(\dots) \leq 1$ and that $\mu_{k_n}^{-1}(U), \mu^{*-1}(U) \in [0,1]$. The limit along such subsequence then uniquely determines a, proving $\bar{\mathcal{P}}$ is differentiable, and

$$\frac{\partial}{\partial \beta_\ell} \mathcal{P}(\gamma(\beta)) = \beta_\ell \Big\{ 1 - \sum_{p\geq \ell}^{r} (\zeta_p - \zeta_{p-1}) \int x^2 \rho_{\mu^*}(dx|\Gamma = \gamma_p) \Big\} . \tag{88}$$

Using the chain rule we can finally write

$$\frac{\partial}{\partial \gamma_\ell} \mathcal{P}(\gamma(\beta)) = \begin{cases} -\gamma_\ell (\zeta_\ell - \zeta_{\ell-1}) \int x^2 \rho_{\mu^*}(dx \,|\, \Gamma = \gamma_\ell) \,, & \text{if } \ell = 1, \dots, r-1 \\ \gamma_r \Big(1 - (1 - \zeta_{r-1}) \int x^2 \rho_{\mu^*}(dx \,|\, \Gamma = \gamma_r) \Big) \,, & \text{if } \ell = r \end{cases} \tag{89}$$

which proves the statement when matched with the limit of (22).

4.4 Proof of Theorem 4

Recall that in this proof we assume $h = 0$ almost surely. Let us start noticing that

$$p_N \leq \log 2 + \frac{\gamma_r^2}{2} \tag{90}$$

uniformly in N. Indeed one has

$$\log 2 + \frac{\gamma_r^2}{2} \;=\; p_N|_{\zeta = \mathbf{1}} \tag{91}$$

for any integer N. Hence the inequality (90) easily follows from a repeated applications of Jensen inequality and the fact that $x \mapsto x^t$ is concave for any $t \in [0,1]$. Moreover again by Jensen inequality one has that

$$p_N \geq \frac{1}{N} \mathbb{E} \log Z_{r,N} \tag{92}$$

uniformly in N, where $\mathbb{E}$ denotes expectation w.r.t. all the disorder. Now observe that the quantity $\frac{1}{N}\mathbb{E}\log Z_{r,N}$ coincides with quenched pressure on a Sherringhton-Kirkpatrick model at inverse temperature $\beta = \sqrt{2}\gamma_r$ and then [3] one has

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\log Z_{r,N} = \log 2 + \frac{\beta^2}{4} \iff \beta \le 1\,. \tag{93}$$

This implies that if $\gamma_r^2 \le \frac{1}{2}$ then

$$\lim_{N\to\infty} p_N = \log 2 + \frac{\gamma_r^2}{2}\,. \tag{94}$$

In order to prove that $\gamma_r^2 \le \frac{1}{2}$ is also a necessary condition for the equality (94) we start by noticing that $\log 2 + \frac{\gamma_r^2}{2}$ is obtained as a limiting value of the Parisi functional $\mathcal{P}(x,\xi)$. More precisely let us define

$$f(x_1,\dots,x_r) = \mathcal{P}(x,\xi^*) \tag{95}$$

where $\boldsymbol{\xi}^* = (\xi^*)_{j\le r+1}$ is fixed trough the choice

$$\xi_0^* = \zeta_0 < \dots < \xi_{r-1}^* = \zeta_{r-1} < \xi_r^* = \xi_{r+1}^* = 1\,. \tag{96}$$

Then it's easy to check that

$$\lim_{\substack{x_j\to 0^+\\ j\le r}} f(x_1,\dots,x_r) = \log 2 + \frac{\gamma_r^2}{2}\,. \tag{97}$$

Let us consider the quantity

$$f(x_r) = \lim_{\substack{x_j\to 0^+\\ j\le r-1}} f(x_1,\dots,x_r) \tag{98}$$

$$= \log 2 + \frac{1}{\zeta_{r-1}}\log \mathbb{E}_\eta \cosh^{\zeta_{r-1}}\left(\eta\gamma_r\sqrt{2x_r}\right) + \frac{\gamma_r^2}{2}\left(1 - 2x_r + (1-\zeta_{r-1})x_r^2\right). \tag{99}$$

Recall that the RSB bound (Proposition 3.1 in [7]) implies that

$$\lim_{N\to\infty} p_N \le \inf_{x_r\in[0,1]} f(x_r). \tag{100}$$

We want to prove that if $\gamma_r^2 > \frac{1}{2}$ then $\inf_{x_r} f(x_r) < \lim_{x_r \to 0^+} f(x_r) = \log 2 + \frac{\gamma_r^2}{2}$. Equation (64) yields

$$\frac{d}{dx_r} f(x_r) = (1-\zeta_{r-1})\frac{\gamma_r^2}{2}\left(x_r - \frac{\mathbb{E}_\eta \cosh^{\zeta_{r-1}}\left(\eta\gamma_r\sqrt{2x_r}\right)\tanh^2\left(\eta\gamma_r\sqrt{2x_r}\right)}{\mathbb{E}_\eta \cosh^{\zeta_{r-1}}\left(\eta\gamma_r\sqrt{2x_r}\right)}\right). \tag{101}$$

This implies that $\lim_{x_r\to 0^+} \frac{d}{dx_r} f(x_r) = 0$ and also that

$$\lim_{x_r \to 0^+} \frac{d^2}{dx_r^2} f(x_r) = (1-\zeta_{r-1})\frac{\gamma_r^2}{2}(1-2\gamma_r^2). \tag{102}$$

Hence, if $\gamma_r^2 > \frac{1}{2}$ one has that $\lim_{x_r\to 0^+} \frac{d^2}{dx_r^2} f(x_r) < 0$. From (100) we can thus conclude

$$\lim_{N\to\infty} p_N \le \inf_{x_r \in [0,1]} f(x_r) < \lim_{x_r \to 0^+} f(x_r) = \log 2 + \frac{\gamma_r^2}{2}. \tag{103}$$

4.5 Proof of Theorem 5

The proof leverages on the fact that there are at least r different values of γ_ℓ's, and that the stationary sequences $(x_j^*)_{j\le k}$ of the Parisi functional are naturally ordered.

Suppose, by contradiction, that the infimum of the variational formula (31) is attained on a k-stationary pair (x^*, ξ^*) such that

$$0 = x_0^* < x_1^* < x_2^* < \cdots < x_{j_\ell}^* = x_{j_\ell+1}^* < x_{j_\ell+2}^* < \cdots < x_k^* < x_{k+1}^* = 1 \tag{104}$$

where we stress that $x_{j_\ell}^* = x_{j_\ell+1}^*$ collapsed. The index j_ℓ is chosen in such a way that $\xi_{j_\ell}^* = \zeta_\ell$. Hence the CDF associated to the above choice of μ^*, does not include ζ_ℓ in its image, as suggested by the limiting procedure outlined in Figure 2. Our assumption then entails

$$\inf_{(x,\xi)\in\mathcal{M}} \mathcal{P}(x,\xi) = \mathcal{P}(x^*,\xi^*). \tag{105}$$

Thanks to Lemma 9 we have that

$$\partial_{x_{j_\ell}} \mathcal{P}(x,\xi)\Big|_{(x^*,\xi^*)} = \tilde{\gamma}_{j_\ell}^2(\xi_{j_\ell}^* - \xi_{j_\ell-1}^*)\Big[x_{j_\ell}^* - \mathbb{E}\prod_{p=1}^{j_\ell} f_p\Big(\langle\sigma\rangle^{(j_\ell)}\Big)^2\Big] = 0. \tag{106}$$

Recall that $\xi_{j_\ell}^* = \xi_{j_\ell-1}^*$ can be discarded without loss of generality. In fact, in that case the inf would be attained on a distribution with less than k distinct relevant values in the sequence x. It is indeed intuitive from Figure 2, that in order to skip ζ_2, ξ_1, ξ_3 and ξ_2 (which equals ζ_2) must be all different.

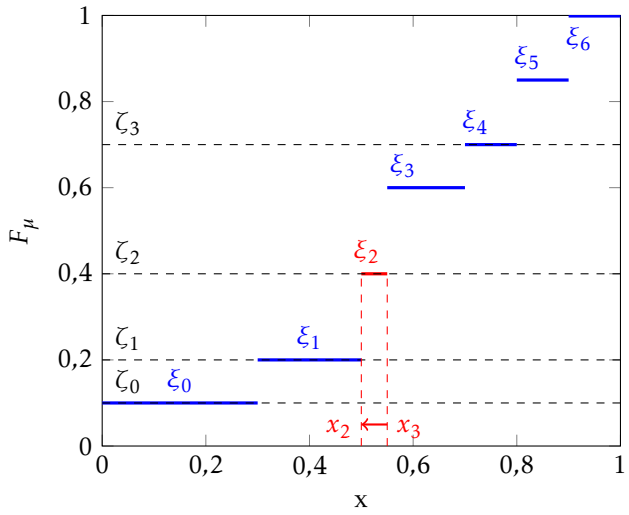

Figure 2 Typical limiting situation in which one of the ζ_ℓ's (in this case ζ_2) disappears from the cumulative distribution function. In general, ζ_ℓ is not in the final limiting distribution when $x_{j_\ell+1} \to x_{j_\ell}$ for j_ℓ s.t. $\xi_{j_\ell} = \zeta_\ell$. In this plot, $r = 4$, $k = 6$.

In order to prove Theorem 5, we show that, under the assumption that (x^*, ξ^*) is the infimum point of $\mathcal{P}$ on $\mathcal{M}_k$, we are actually able to construct a new pair $(\bar{x}, \bar{\xi})$ with finite number of atoms, whose cumulative distribution contains ζ_ℓ in the image and such that

$$\mathcal{P}(x^*, \xi^*) > \mathcal{P}(\bar{x}, \bar{\xi}). \tag{107}$$

Let us design a pair $(\bar{x}, \bar{\xi})$ with $k+1$ distinct atoms as follows. We take $\bar{\xi} = \xi^*$ and

$$0 = x_0^* = \bar{x}_0 < x_1^* = \bar{x}_1 < \cdots < x_{j_\ell}^* = \bar{x}_{j_\ell} = x_{j_\ell+1}^* < \bar{x}_{j_\ell+1} < \cdots < x_{k+1}^* = \bar{x}_{k+1} = 1\,. \tag{108}$$

In broad terms the sequence $\bar{x}$ is the sequence x^* where $x_{j_\ell}^*$ remained detatched from $x_{j_\ell+1}^*$. With these notations, following (25), one has

$$\tilde{\gamma}_{j_\ell} = \gamma_\ell < \tilde{\gamma}_{j_\ell+1} = \gamma_{\ell+1}\,. \tag{109}$$

It remains to show that there exists a choice of $\bar{x}_{j_\ell+1}$ such that (107) holds true. This is verified if we prove that

$$\partial_{x_{j_\ell+1}} \mathcal{P}(x, \xi)\Big|_{(x^*,\xi^*)} = \tilde{\gamma}_{j_\ell+1}^2 (\xi_{j_\ell+1}^* - \xi_{j_\ell}^*)\Big[x_{j_\ell+1}^* - \mathbb{E}\prod_{p=1}^{j_\ell+1} f_p\Big(\langle\sigma\rangle^{(j_\ell+1)}\Big)^2\Big] \tag{110}$$

is strictly negative. Using the fact that $x^*_{j_\ell} = x^*_{j_\ell+1}$, together with (106) and (60) one readily gets

$$\partial_{x_{j_\ell+1}} \mathcal{P}(x,\xi)\Big|_{(x^*,\xi^*)} = -\tilde\gamma^2_{j_\ell+1}(\xi^*_{j_\ell+1} - \xi^*_{j_\ell})\mathbb{E}\prod_{p=1}^{j_\ell} f_p\Big[\mathbb{E}_{j_\ell} f_{j_\ell+1}\Big(\langle\sigma\rangle^{(j_\ell+1)} - \mathbb{E}_{j_\ell} f_{j_\ell+1}\langle\sigma\rangle^{(j_\ell+1)}\Big)^2\Big] \tag{111}$$

which is non positive. Keeping in mind the example of Figure 2, it is not difficult to see that, in the case we are interested in, we can assume $\xi^*_{j_\ell+1} > \xi^*_{j_\ell}$ without any loss of generality. Hence, the sign of the previous derivative is uniquely determined by the variance in (111).

Recall that

$$\begin{aligned}\langle\sigma\rangle^{(j_\ell+1)} = \mathbb{E}_{j_\ell+1}\dots\mathbb{E}_k f_{j_\ell+2}\dots f_{k+1}\tanh\Big(&\sum_{p=1}^{j_\ell}\eta_p\sqrt{2\big(\tilde\gamma^2_p x^*_p - \tilde\gamma^2_{p-1}x^*_{p-1}\big)}\\ &+\eta_{j_\ell+1}\sqrt{2\big(\gamma^2_{\ell+1}-\gamma^2_\ell\big)x^*_{j_\ell}} + \sum_{p=j_\ell+2}^{k+1}\eta_p\sqrt{2\big(\tilde\gamma^2_p x^*_p - \tilde\gamma^2_{p-1}x^*_{p-1}\big)} + h\Big),\end{aligned} \tag{112}$$

where we have imposed $x^*_{j_\ell+1} = x^*_{j_\ell}$ and $\tilde\gamma_{j_\ell+1} = \gamma_{\ell+1} > \tilde\gamma_{j_\ell} = \gamma_\ell$. The summation of the first $j_\ell+1$ contributions in the above, and possibly the magnetic field, is what makes $\langle\sigma\rangle^{(j_\ell+1)}$ random. From the previous formula we can see that, as long as $\eta_{j_\ell+1}$ appears, $\langle\sigma\rangle^{(j_\ell+1)}$ cannot have vanishing variance in the measure $\mathbb{E}_{j_\ell} f_{j_\ell+1}(\cdot)$. This always occurs precisely because $\gamma_{\ell+1} > \gamma_\ell$, unless all the $x^*_j, \forall\, j \le j_\ell+1$ collapse to zero. This is ruled out by the two following Lemmas.

Lemma 11. *If $\mathbb{E}h^2 > 0$ then $x^*_j > 0$ for all $j = 1,\dots,k$.*

Proof. To prove the result, we show that at $x^*_j = 0$ for some j, $\partial_{x_j}\mathcal{P}$ is strictly negative. Recall that

$$\langle\sigma\rangle^{(j)} = \mathbb{E}_j\dots\mathbb{E}_k f_{j+1}\dots f_{k+1}\tanh\Big(\sum_{p=1}^{k+1}\eta_p\sqrt{2\big(\tilde\gamma^2_p x^*_p - \tilde\gamma^2_{p-1}x^*_{p-1}\big)} + h\Big). \tag{113}$$

Using a reasoning similar to that used to prove (62)-(63), one can show

$$\partial_h f_j = \xi_{j-1}(\langle\sigma\rangle^{(j)} - \langle\sigma\rangle^{(j-1)})f_j\,. \tag{114}$$

Then we can finally compute the derivative

$$\partial_h\langle\sigma\rangle^{(j)} = \mathbb{E}_j\dots\mathbb{E}_k\sum_{s=j+1}^{k+1} f_{j+1}\dots\partial_h f_s\dots f_{k+1}\langle\sigma\rangle + 1 - \mathbb{E}_j\dots\mathbb{E}_k f_{j+1}\dots f_{k+1}\langle\sigma\rangle^2 =$$

$$= 1 - \mathbb{E}_j \dots \mathbb{E}_k f_{j+1} \dots f_{k+1} \langle\sigma\rangle^2 + \sum_{s=j+1}^{k+1} \mathbb{E}_j \dots \mathbb{E}_k f_{j+1} \dots f_{k+1} \xi_{s-1} (\langle\sigma\rangle^{(s)} - \langle\sigma\rangle^{(s-1)}) \langle\sigma\rangle . \tag{115}$$

The last set of terms in the above line can be rewritten using the recursive relation (60), and the first one is strictly positive. Hence:

$$\partial_h \langle\sigma\rangle^{(j)} > \sum_{s=j+1}^{k+1} \xi_{s-1} \mathbb{E}_j \dots \mathbb{E}_{s-1} f_{j+1} \dots f_s [(\langle\sigma\rangle^{(s)})^2 - (\mathbb{E}_{s-1} f_s \langle\sigma\rangle^{(s)})^2] \geq 0 . \tag{116}$$

The above is actually true for any value of the variational parameters ξ, x. Suppose now, by contradiction, that $x_j^* = 0$, which also entails $x_s^* = 0$ for all $s \leq j$ by inequality constraints automatically satisfied by stationary points, see Lemma 9. In that case one would also have $\langle\sigma\rangle^{(j)}|_{h=0} = 0$ by symmetry of the Gaussian law of the η's, and of the weights f (recall $x_{s\leq j}^* = 0$). This then means that as soon as the magnetic field is non zero $\langle\sigma\rangle^{(j)}|_{h\neq 0} \neq 0$. This would then make it impossible for the derivative (64) to vanish when evaluated at $x_j = 0$. This contradiction is born from the assumption $x_j^* = 0$ which is thus falsified concluding the proof. □

Lemma 12. *Consider the case* $\mathbb{E}h^2 = 0$. *If* $\gamma_\ell^2 > \frac{1}{2\zeta_\ell^2}$ *then* $x_{j_\ell}^* > 0$.

Proof. The strategy is to show that $x_{j\leq j_\ell}^* = 0$ is an unstable stationary point under the hypothesis. In order to do it, we need an expansion of $\langle\sigma\rangle^{(j_\ell)}$ around $x_{j_\ell} \to 0$, with $x_{j_\ell - 1} = \dots = x_1 = 0$. Assuming the first $j_\ell - 1$ coordinates are 0 is licit as for stationary points x^* they are naturally ordered, as per Lemma 9.

In this selected limiting direction (which is compatible with the inequality constraints) we can write:

$$\langle\sigma\rangle^{(j_\ell)} = \mathbb{E}_{j_\ell} \dots \mathbb{E}_k f_{j_\ell+1} \dots f_{k+1} \tanh\Big(\sqrt{2x_{j_\ell}}\eta_{j_\ell}\gamma_\ell + \sum_{p=j_\ell+1}^{k+1} \eta_p \sqrt{2(\tilde{\gamma}_p^2 x_p - \tilde{\gamma}_{p-1}^2 x_{p-1})}\Big), \tag{117}$$

where we have kept the remaining x's generic, and $\tilde{\gamma}_{j_\ell} = \gamma_\ell$. Even though it was not specified, a similar separation can be carried out inside the f's. The term $\sqrt{2x_{j_\ell}}\eta_{j_\ell}\gamma_\ell$ behaves as a random magnetic field. Hence we can expand around it using the formula for the derivative (115):

$$\langle\sigma\rangle^{(j_\ell)} = \sqrt{2x_{j_\ell}}\eta_{j_\ell}\gamma_\ell \mathbb{E}_{j_\ell} \dots \mathbb{E}_k f_{j_\ell+1}^0 \dots f_{k+1}^0 \Big(1 - \langle\sigma\rangle^2\Big)$$

$$+\sqrt{2x_{j_\ell}}\eta_{j_\ell}\gamma_\ell\sum_{p=j_\ell+1}^{k+1}\xi_{p-1}\mathbb{E}_{j_\ell}\dots\mathbb{E}_k f^0_{j_\ell+1}\dots f^0_{k+1}\Big(\langle\sigma\rangle^{(p)}-\langle\sigma\rangle^{(p-1)}\Big)\langle\sigma\rangle+\eta^2_{j_\ell}O(x_{j_\ell}),\tag{118}$$

where the superscript ^{0} signals that $x_{j_\ell}=0$ in those f's, and for the sake of presentation we have omitted that, inside all brackets $\langle\sigma\rangle^{(\dots)}$, $x_{j\le j_\ell}$ have been set to 0 too. The $O(x_{j_\ell})$ comes from the fact that the second derivative w.r.t. an external field is uniformly bounded, thanks to (114), and the fact that derivatives of tanh are uniformly bounded. The $\eta^2_{j_\ell}$ (and its powers) will be later averaged under a Gaussian standard measure yielding a controllable constant. Notice also that $f^0_j=1$ for all $j\le j_\ell$. Therefore

$$\begin{aligned}\mathbb{E}\prod_{s=1}^{j_\ell}f_s(\langle\sigma\rangle^{(j_\ell)})^2 &= 2x_{j_\ell}\gamma_\ell^2\Big[\mathbb{E}_{j_\ell}\dots\mathbb{E}_k f^0_{j_\ell+1}\dots f^0_{k+1}\Big(1-\langle\sigma\rangle^2\Big)\\ &\quad+\sum_{p=j_\ell+1}^{k+1}\xi_{p-1}\mathbb{E}_{j_\ell}\dots\mathbb{E}_k f^0_{j_\ell+1}\dots f^0_{k+1}\Big(\langle\sigma\rangle^{(p)}-\langle\sigma\rangle^{(p-1)}\Big)\langle\sigma\rangle\Big]^2+O(x_{j_\ell}^{3/2}).\end{aligned}\tag{119}$$

For all the p's in the summation one has $\xi_{p-1}\ge\xi_{j_\ell}=\zeta_\ell$. Furthermore, as we have already observed for (115), the last summation $\sum_{p=j_\ell+1}^{k+1}\dots$ contains only non-negative terms. Following these considerations the previous equation turns into

$$\dots=2x_{j_\ell}\gamma_\ell^2\Big[\mathbb{E}_{j_\ell}\dots\mathbb{E}_k f^0_{j_\ell+1}\dots f^0_{k+1}\Big(1-(1-\zeta_\ell)\langle\sigma\rangle^2+\zeta_\ell\langle\sigma\rangle^{(j_\ell)}\langle\sigma\rangle\Big)\Big]^2+O(x_{j_\ell}^{3/2}),\tag{120}$$

where we have solved the last telescopic sum. We stress again that $x_{j\le j_\ell}=0$ in all the above brackets on the r.h.s. This implies that $\langle\sigma\rangle^{(j_\ell)}=0$ again by symmetry of the laws of the η's and their tilts f. Hence, finally

$$\mathbb{E}\prod_{s=1}^{j_\ell}f_s(\langle\sigma\rangle^{(j_\ell)})^2\ge 2x_{j_\ell}\gamma_\ell^2\zeta_\ell^2(1+o(1)),\tag{121}$$

where we just used $\langle\sigma\rangle^2\le 1$ and $o(1)$ vanishes as $x_{j_\ell}\to 0$. This can be used to bound the derivative of the Parisi functional when $x_{j_\ell}\to 0$:

$$\partial_{x_{j_\ell}}\mathcal{P}\le\gamma_\ell^2(\zeta_\ell-\xi_{j_\ell-1})x_{j_\ell}\Big[1-2\gamma_\ell^2\zeta_\ell^2+o(1)\Big].\tag{122}$$

The above also yields a bound on the second derivative at $x_{j\le j_\ell}=0$ in direction x_{j_ℓ}:

$$\partial^2_{x_{j_\ell}}\mathcal{P}|_{x_{j\le j_\ell}=0}\le\gamma_\ell^2(\zeta_\ell-\xi_{j_\ell-1})\Big[1-2\gamma_\ell^2\zeta_\ell^2\Big],\tag{123}$$

which is negative whenever $\gamma_\ell^2 > \frac{1}{2\zeta_\ell^2}$. Therefore the stationary point $x_{j\le j_\ell} = 0$ is unstable precisely in direction x_{j_ℓ} proving the claim.

□

To summarize, we have shown that, under the hypothesis (52) or $h \neq 0$ with positive probability

$$\mathcal{P}(x^*, \xi^*) > \mathcal{P}(\bar{x}, \bar{\xi}). \tag{124}$$

This concludes the proof by absurd.

4.6 Proof of Theorem 6

In this section we denote by μ a solution of the variational problem (40) which is is the W_1-limit of a sequence (μ_n) associated to some k_n-stationary pair. We start proving the following

Proposition 13. *If $\zeta_\ell \gamma_{\ell+1}^2 < 1/2$ then*

$$\Delta_\ell(\mu) \ge 2\left(\gamma_{\ell+1}^2 - \gamma_\ell^2\right) \limsup_{n\to\infty} \mu_n^{-1}(\zeta_\ell^+)\Big(\int_{\mu^{-1}(\zeta_\ell^+)}^1 \mu([0,x])\,dx\Big)^2. \tag{125}$$

The above inequality still contains a limit. The latter can be removed at the expense of a slightly weaker inequality, that however offers the advantage of involving the limiting distribution μ only. Indeed one can define the following smoothing of the quantile function:

$$\nu_{\delta,\ell}^{-1} := \frac{1}{\delta}\int_{\zeta_\ell}^{\zeta_\ell+\delta} \nu^{-1}(p)dp\,, \tag{126}$$

for any $\delta > 0$. As an immediate consequence of its definition we have that

$$\begin{aligned} \nu^{-1}(\zeta_\ell^+) \le \nu_{\delta,\ell}^{-1} \le \nu^{-1}(\zeta_\ell + \delta) \\ \nu^{-1}(\zeta_\ell - \delta) \le \nu_{-\delta,\ell}^{-1} \le \nu^{-1}(\zeta_\ell). \end{aligned} \tag{127}$$

Thanks to the above equalities we will be able to exploit the W_1 convergence of μ_n to μ. For instance, we can prove the following

Lemma 14.

$$\mu_n \xrightarrow{W_1} \mu \Rightarrow \Delta_\ell(\mu) \ge \limsup_{n\to\infty} \Delta_\ell(\mu_n). \tag{128}$$

Proof. For any non empty interval $A \subseteq [0,1]$ one has

$$\mu_n \xrightarrow{W_1} \mu \Rightarrow \lim_{n\to\infty} \Big|\int_A \left(\mu^{-1}(p) - \mu_n^{-1}(p)\right) dp\Big| = 0. \tag{129}$$

The above implies

$$\mu^{-1}_{\delta,\ell} = (\mu_n^{-1})_{\delta,\ell} + o_n(1), \quad \mu^{-1}_{-\delta,\ell} = (\mu_n^{-1})_{-\delta,\ell} + o_n(1) \tag{130}$$

for any positive δ. Therefore, using (127), we get

$$\mu^{-1}(\zeta_\ell+\delta)-\mu^{-1}(\zeta_\ell-\delta) \geq \mu^{-1}_{\delta,\ell}-\mu^{-1}_{-\delta,\ell} = (\mu_n^{-1})_{\delta,\ell}-(\mu_n^{-1})_{-\delta,\ell}+o_n(1) \geq \mu_n^{-1}(\zeta_\ell^+)-\mu_n^{-1}(\zeta_\ell)+o_n(1). \tag{131}$$

Take first $n \to \infty$ and then $\delta \to 0^+$ one get (128). □

From (127) we observe that

$$\mu^{-1}(\zeta_\ell + \delta) \geq \mu^{-1}_{\delta,\ell} = (\mu_n^{-1})_{\delta,\ell} + o_n(1) \geq \mu_n^{-1}(\zeta_\ell^+) + o_n(1). \tag{132}$$

Hence, by sending $n \to \infty$ and then $\delta \to 0$ we obtain

$$\mu^{-1}(\zeta_\ell^+) \geq \limsup_{n\to\infty} \mu_n^{-1}(\zeta_\ell^+). \tag{133}$$

On the other hand, if one is willing to sacrifice the right limit,

$$\limsup_{n\to\infty} \mu_n^{-1}(\zeta_\ell^+) \geq \limsup_{n\to\infty} (\mu_n^{-1})_{-\delta,\ell} = \mu^{-1}_{-\delta,\ell} \tag{134}$$

for any $\delta > 0$. By sending $\delta \to 0$, and left continuity of the quantile, we get

$$\limsup_{n\to\infty} \mu_n^{-1}(\zeta_\ell^+) \geq \mu^{-1}(\zeta_\ell). \tag{135}$$

Used in (125) the above leads to (55). Let's go back to the proof of Proposition (13). We will prove at first the following finite k version of it:

Proposition 15. *Given $k \in \mathbb{N}$, let μ_k be a measure associated to a k-stationary pair. If $\zeta_\ell\, \gamma^2_{\ell+1} < 1/2$ then*

$$\mu_k^{-1}(\zeta_\ell^+) - \mu_k^{-1}(\zeta_\ell) \geq 2\, \mu_k^{-1}(\zeta_\ell^+)\Big(\gamma^2_{\ell+1} - \gamma^2_\ell\Big)\left(\int_{\mu_k^{-1}(\zeta_\ell^+)}^1 \mu_k([0,x])\, dx\right)^2. \tag{136}$$

Assume for the moment that the above proposition holds, then (125) also holds. In fact, it suffices to use (136) and (128) together with the continuity of the integral w.r.t. its integration extremes and (133).

Let us start recalling a classical result.

Lemma 16. *(Cramer-Rao bound) Let $\phi \in C^2(\mathbb{R},\mathbb{R})$ and consider the probability density $p(x) \propto e^{-\phi(x)}$ then for h locally Lipscthiz one has*

$$\mathrm{Var}(h) \geq \big[\mathbb{E}(h')\big]^2 \big[\mathbb{E}(\phi'')\big]^{-1}. \tag{137}$$

For a proof see [53].

In order to use the above, we pick a k-stationary pair (x,ξ) and assume without loss that it is minimal. Fix some $\ell < r$ and let $j_\ell \in \{1,\dots,k\}$ be the unique index such that $\tilde\gamma_{j_\ell} = \gamma_\ell < \tilde\gamma_{j_\ell+1} = \gamma_{\ell+1}$ or equivalently $\xi_{j_\ell} = \zeta_\ell$. By Lemma 9 one has

$$x_{j_{\ell+1}} - x_{j_\ell} = \mathbb{E}\prod_{p=1}^{j_\ell} f_p \left[\mathrm{Var}_{j_\ell+1}\left(\langle\sigma\rangle^{(j_\ell+1)}\right)\right] \tag{138}$$

where, conditionally on $(\eta_j)_{j\le j_\ell}$, the quantity $\mathrm{Var}_{j_\ell+1}$ is the variance w.r.t. the average $\mathbb{E}_{j_\ell} f_{j_\ell+1}$, that is associated with the probability density

$$p(\eta_{j_\ell+1}) \propto \exp\left[-\frac{1}{2}\eta_{j_\ell+1}^2 + \xi_{j_\ell}\log\tilde Z_{j_\ell+1}\right]. \tag{139}$$

The idea is now to use Lemma 16 to bound $\mathrm{Var}_{j_\ell+1}\left(\langle\sigma\rangle^{(j_\ell+1)}\right)$ conditionally on $(\eta_j)_{j\le j_\ell}$. In order to lighten the notation we set $j_\ell + 1 \equiv i$ and $\eta_{j_\ell+1} \equiv \eta$. Therefore by (137) we have that

$$\mathrm{Var}_i(\langle\sigma\rangle^{(i)}) \ge \left[\mathbb{E}_\eta\left(\frac{\partial h}{\partial\eta}\right)\right]^2 \left[\mathbb{E}_\eta\left(\frac{\partial^2\phi}{\partial\eta^2}\right)\right]^{-1} \tag{140}$$

where

$$\begin{aligned} \eta &\mapsto \phi(\eta) = \frac{1}{2}\eta^2 - \xi_{i-1}\log\tilde Z_i(\eta) \\ \eta &\mapsto h(\eta) = \langle\sigma\rangle^{(i)} \end{aligned} \tag{141}$$

and $\mathbb{E}_\eta$ is the average w.r.t. the probability density (139). Recall that since $\xi_k = \xi_{k+1} = 1$ we have the following simplification:

$$\langle\sigma\rangle^{(i)} = \mathbb{E}_{\ge i}\prod_{p=i+1}^{k} f_p \tanh\Big(\sum_{j=1}^{k}\eta_j\sqrt{2(\tilde\gamma_j^2 x_j - \tilde\gamma_{j-1}^2 x_{j-1})}\Big) \tag{142}$$

namely f_{k+1} can be explicitly integrated away.

Lemma 17. *Let $h(\eta)$ and $\phi(\eta)$ the random functions defined in (141). If $\zeta_\ell\,\gamma_{\ell+1}^2 < 1/2$ then*

$$0 < \mathbb{E}_\eta\left(\frac{\partial^2}{\partial\eta^2}\phi\right) \le 1\,. \tag{143}$$

Moreover,

$$\mathbb{E}_\eta\left(\frac{\partial h}{\partial\eta}\right) = \sqrt{2(\gamma_{\ell+1}^2 x_i - \gamma_\ell^2 x_{i-1})}\,\mathbb{E}_\eta\left(\sum_{p=i+1}^{k+1}\xi_{p-1}(a_p - a_{p-1})\right) \tag{144}$$

where, for any $p \le k$

$$a_p := \mathbb{E}_{\ge i}^{<p}\prod_{s=i+1}^{p} f_s\left(\langle\sigma\rangle^{(p)}\right)^2, \text{ and } a_{k+1} = 1\,. \tag{145}$$

Proof. Let us start recalling that $\tilde{Z}_i(\eta)$ is obtained recursively starting from $\tilde{Z}_{k+1} = 2\cosh(z)$ where $z = \sum_{j\le k+1} \eta_j c_j$ and $c_j = \sqrt{2(\tilde{\gamma}_j^2 x_j - \tilde{\gamma}_{j-1}^2 x_{j-1})}$.

Therefore

$$\begin{aligned}\frac{\partial}{\partial\eta}\phi(\eta) &= \eta - \xi_{i-1}\frac{\partial}{\partial\eta}\log\tilde{Z}_i(\eta)\\ &= \eta - \xi_{i-1}\mathbb{E}_{\ge i}\prod_{p=i+1}^{k+1} f_p \frac{\partial}{\partial\eta}\log\tilde{Z}_{k+1}(\eta)\\ &= \eta - \xi_{i-1}c_i\langle\sigma\rangle^{(i)} = \eta - \xi_{i-1}c_i h(\eta)\end{aligned} \tag{146}$$

and

$$\frac{\partial^2}{\partial\eta^2}\phi(\eta) = 1 - \xi_{i-1}c_i\frac{\partial}{\partial\eta}h(\eta). \tag{147}$$

Now $\frac{\partial}{\partial\eta}h(\eta)$ can be computed as follows

$$\begin{aligned}\frac{\partial}{\partial\eta}h(\eta) = \mathbb{E}_{\ge i}\frac{\partial}{\partial\eta}\Big[\prod_{p=i+1}^{k} f_p \tanh\Big(\sum_{j=1}^{k}\eta_j c_j\Big)\Big] = \mathbb{E}_{\ge i}\sum_{p=i+1}^{k}\frac{\partial}{\partial\eta}[f_p]\prod_{s=i+1,s\neq p}^{k} f_s \tanh\Big(\sum_{j=1}^{k}\eta_j c_j\Big)\\ + \mathbb{E}_{\ge i}\prod_{p=i+1}^{k} f_p\frac{\partial}{\partial\eta}[\tanh\Big(\sum_{j=1}^{k}\eta_j c_j\Big)] = A + B\end{aligned} \tag{148}$$

where A and B are defined as the first and second term respectively on the left hand side of the above.

Recall that $\eta \equiv \eta_i$ then one can use relation (63) to obtain

$$A = c_i\sum_{p=i+1}^{k}\xi_{p-1}\,\mathbb{E}_{\ge i}\prod_{s=i+1}^{k} f_s \tanh\Big(\sum_{j=1}^{k}\eta_j c_j\Big)\Big(\langle\sigma\rangle^{(p)} - \langle\sigma\rangle^{(p-1)}\Big). \tag{149}$$

Keeping in mind that $\langle\sigma\rangle^{(p)} = \mathbb{E}_{\ge p}\prod_{s=p+1}^{k+1} f_s\tanh(z) = \mathbb{E}_{\ge p}\prod_{s=p+1}^{k} f_s\tanh(\sum_{j=1}^{k}\eta_j c_j)$ depends on $(\eta_j)_{j\le p}$ and f_s depends on $(\eta_j)_{j\le s}$ one can write for any p-term in the previous line

$$\mathbb{E}_{\ge i}\prod_{s=i+1}^{k} f_s \tanh\Big(\sum_{j=1}^{k}\eta_j c_j\Big)\Big(\langle\sigma\rangle^{(p)} - \langle\sigma\rangle^{(p-1)}\Big) = \mathbb{E}_{\ge i}^{<p}\prod_{s=i+1}^{p} f_s(\langle\sigma\rangle^{(p)})^2 - \mathbb{E}_{\ge i}^{<p-1}\prod_{s=i+1}^{p-1} f_s(\langle\sigma\rangle^{(p-1)})^2. \tag{150}$$

Using (150) in (149) one gets

$$A = c_i\sum_{p=i+1}^{k}\xi_{p-1}(a_p - a_{p-1}) \tag{151}$$

with a_p as in (145).

The second term in (148) is

$$B = c_i(1 - \mathbb{E}_{\geq i} \prod_{s=i+1}^{k} f_s \tanh^2\Big(\sum_{j=1}^{k} \eta_j c_j\Big) = c_i(1 - a_k). \tag{152}$$

Therefore from (148) one gets

$$\frac{\partial}{\partial \eta} h(\eta) = A + B = c_i \sum_{p=i+1}^{k+1} \xi_{p-1}(a_p - a_{p-1}) \tag{153}$$

where we used $a_{k+1} = 1$, $\xi_k = 1$. Taking the expectation on both sides we prove (144).

Therefore, since $1 \geq a_p \geq a_{p-1} \geq 0$ for all p's, one has

$$0 \leq \frac{\partial}{\partial \eta} h(\eta) \leq c_i \sum_{p=i+1}^{k+1} (a_p - a_{p-1}) = c_i(1 - a_i) \leq c_i\,. \tag{154}$$

and hence

$$1 - \xi_{i-1} c_i^2 \leq \frac{\partial^2}{\partial \eta^2} \phi(\eta) \leq 1 \tag{155}$$

which is uniform in the length k of the stationary pair (x, ξ) . In particular (155) implies that

$$1 - \xi_{i-1} c_i^2 \leq \mathbb{E}_\eta \left(\frac{\partial^2}{\partial \eta^2} \phi \right) \leq 1. \tag{156}$$

Keeping in mind that $c_i^2 = 2(\gamma_{\ell+1}^2 x_{j_{\ell+1}} - \gamma_\ell^2 x_{j_\ell})$, we have that $c_i^2 \leq 2\gamma_{\ell+1}^2$ and $1 - \xi_{i-1} c_i^2 \geq 1 - 2\xi_{i-1}\gamma_{\ell+1}^2$. Therefore the condition $1 - 2\xi_{i-1}\gamma_{\ell+1}^2 > 0$ implies that

$$0 < \mathbb{E}_\eta \left(\frac{\partial^2}{\partial \eta^2} \phi \right) \leq 1\,. \tag{157}$$

□

Proof of Proposition 15. Thanks to the previous Lemma if $1 - 2\zeta_\ell \gamma_{\ell+1}^2 > 0$ one can use (143), equation (144) and Jensen inequality in the relations (138) and (140) obtaining

$$x_i - x_{i-1} \geq 2\left(\gamma_{\ell+1}^2 x_i - \gamma_\ell^2 x_{i-1}\right) \left[\mathbb{E} \prod_{s=1}^{i-1} f_s \, \mathbb{E}_\eta \Big(\sum_{p=i+1}^{k+1} \xi_{p-1}(a_p - a_{p-1}) \Big) \right]^2. \tag{158}$$

Now by stationary conditions in Lemma 9 one has for all p's

$$\mathbb{E} \prod_{s=1}^{i-1} f_s \, \mathbb{E}_\eta \, a_p = \mathbb{E} \prod_{s=1}^{p} f_s (\langle \sigma \rangle^{(p)})^2 = x_p\,. \tag{159}$$

Keeping in mind that $x_i \equiv x_{j_{\ell+1}} = \mu_k^{-1}(\zeta_\ell^+)$ and $x_{i-1} = \mu_k^{-1}(\zeta_\ell)$ one has

$$\mu_k^{-1}(\zeta_\ell^+) - \mu_k^{-1}(\zeta_\ell) \geq 2\left(\gamma_{\ell+1}^2 \mu_k^{-1}(\zeta_\ell^+) - \gamma_\ell^2 \mu_k^{-1}(\zeta_\ell)\right)\left(\int_{\mu_k^{-1}(\zeta_\ell^+)}^1 \mu_k([0,x])\,dx\right)^2 . \tag{160}$$

Since $\mu_k^{-1}(\zeta_\ell) \leq \mu_k^{-1}(\zeta_\ell^+)$ the statement holds true. □

5 Conclusions and perspectives

In this paper we have investigated the phase diagram of the multiscale Sherrington-Kirkpatrick model and proved some new features. First of all, we clarify the underlying synchronization property of the solution of the variational problem (31). Following that, we computed the limiting overlap moments in the various multiscale measures in Theorem 3, whose ordering is shown to reflect consistently the synchronization property mentioned earlier. In Theorem 4 we exhibit a high-temperature, or weak-coupling, sufficient condition for an annealed solution to hold, which entails also the vanishing of all the overlap second moments. Theorem 5 instead pinpoints a strong coupling regime in which, even in absence of an external magnetic field, the solution of the variational principle (31) must present at least r (i.e. the number of scales originally planted in the model) distinct points in its support.

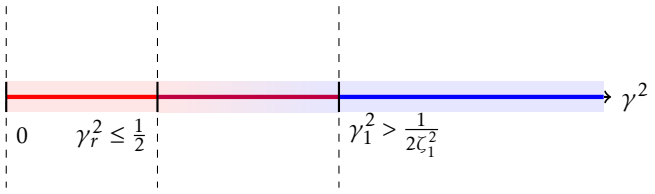

Figure 3 Synthetic representation of our findings.

Figure 3 summarizes our results until Theorem 5, in absence of external magnetic field. All γ's have been collapsed on a 1-D line for the sake of visualization. The red part stands for the annealed region, where $\gamma_r^2 \leq 1/2$. Afterwards there is an intermediate phase where a *partial annealing* may take place, namely the conditional probabilities $\rho_{\mu,\ell}$ have all the mass in 0 up to a certain ℓ. In this shaded region, some of the γ_ℓ^2's can be greater than $1/(2\zeta_\ell^2)$, and some smaller. Following that, we have the strong coupling region highlighted in blue, where partial annealing together with finite RSB is forbidden by Theorem 5. We believe moreover that partial annealing is not possible in this region in any cases. We plan to return to this matter in a subsequent work.

Finally, in Theorem 6 we proved that, under suitable conditions, the CDF associated to a solution of the variational problem (31), must have plateaus in correspondence of the values ζ_ℓ. In said Theorem we give a lower bound on the size of such plateaus, which also holds under a local (i.e. for every ℓ) weak-coupling condition.

The existence of at least one plateau is ensured only if all the mass is not concentrated at zero. This for sure happens when the external magnetic field is present, i.e. $\mathbb{E}h^2 > 0$, as proved in Theorem 5, or when $\gamma_r^2 > 1/2$ as proved in Theorem 4. We also note that for "very strong" couplings full replica symmetry breaking can occur without the Parisi measure manifesting gaps in the support.

Let us now discuss some possible perspectives on this model. It would be interesting to further investigate the shaded region. More precisely, we were able to prove a sufficient condition for an atom to detach from zero at a certain time scale, which is the content of Lemma 12. It is thus natural to wonder if it is also necessary, thus leading to a condition similar to that for the total annealing of Theorem 4. This in turn would give us a fine control on the possible partial annealing phenomenon. Furthermore, it is not clear if the condition $\gamma_{\ell+1}^2 < 1/(2\zeta_\ell)$ of Theorem 6 is fundamental, or it is just a technical artifact. Finally, in presence of external fields instead, it would be interesting to establish a type of de Almeida-Thouless condition for the multiscale SK model, namely, to check if there exists a range of the parameters where each or part of the conditional measures $\rho_{\mu^*,\ell}$ are Dirac deltas centered away from zero. In more physical terms, this would amount to check replica symmetry for each of the thermodynamic scales.

To conclude, we plan to investigate the effect of the multiscale structure combined with other mean-field models, such as multispecies models [43, 54] where different types of synchronization (species and scales) must coexist.

6 Acknowledgements

FC, PC and EM were supported by the EU H2020 ICT48 project Humane AI Net contract number 952026; by the Italian Extended Partnership PE01 - FAIR Future Artificial Intelligence Research - Proposal code PE00000013 under the MUR National Recovery and Resilience Plan; by the project PRIN22CONTUCCI, 2022B5LF52, Boltzmann Machines beyond the independent identically distributed Paradigm: a Mathematical Physics Approach, CUP J53D23003690006.
DT was supported by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU; by the project PRIN22TANTARI, 20229T9EAT, Statistical Mechanics of Learning Machines: from algorithmic and information-theoretical limits to new biologically inspired paradigms, CUP J53D23003640001. The work is partially supported by GNFM (Indam). PC, EM, and DT are affiliated to GNFM-INdAM.

References

[1] Sherrington, D., Kirkpatrick, S.: Solvable model of a spin-glass. Phys. Rev. Lett. **35**, 1792–1796 (1975) https://doi.org/10.1103/PhysRevLett.35.1792

[2] Mezard, M., Parisi, G., Virasoro, M.A.: Spin glass theory and beyond (1987) https://doi.org/10.1142/0271

[3] Talagrand, M.: Mean field models for spin glasses: Volume i: Basic examples, (2010)

[4] Talagrand, M.: Mean field models for spin glasses: Volume ii: Advanced replica-symmetry and low temperature. vol. 55. Springer (2011). https://doi.org/10.1007/978-3-642-22253-5

[5] Panchenko, D.: The Sherrington-Kirkpatrick model. Springer (2013). https://doi.org/10.1007/978-1

[6] Parisi, G.: Infinite number of order parameters for spin-glasses. Phys. Rev. Lett. **43**, 1754–1756 (1979) https://doi.org/10.1103/PhysRevLett.43.1754

[7] Contucci, P., Mingione, E.: A multi-scale spin-glass mean-field model. Communications in Mathematical Physics **368**, 1323–1344 (2019)

[8] Derrida, B.: A generalization of the random energy model which includes correlations between energies. J. Physique Lett. **46**(9), 401–407 (1985) https://doi.org/10.1051/jphyslet:01985004609040100

[9] Ruelle, D.: A mathematical reformulation of Derrida's REM and GREM. Communications in Mathematical Physics **108**(2), 225–239 (1987) https://doi.org/10.1007/BF01210613

[10] Bolthausen, E., Sznitman, A.-S.: On Ruelle's probability cascades and an abstract cavity method. Communications in Mathematical Physics **197**(2), 247–276 (1998) https://doi.org/10.1007/s002200050450

[11] Bovier, A., Kurkova, I.: Derrida's generalised random energy models 1 : models with finitely many hierarchies. Annales de l'I.H.P. Probabilités et statistiques **40**(4), 439–480 (2004) https://doi.org/10.1016/j.anihpb.2003.09.002

[12] Penney, R.W., Coolen, A.C.C., Sherrington, D.: Coupled dynamics of fast spins and slow interactions in neural networks and spin systems. Journal of Physics A: Mathematical and General **26**(15), 3681–3695 (1993)

[13] Allahverdyan, A.E., Nieuwenhuizen, T.M.: Steady adiabatic state: its thermodynamics, entropy production, energy dissipation, and violation of Onsager relations. Physical Review E **62**(1), 845–850 (2000)

[14] Coolen, A.C.C., Penney, R.W., Sherrington, D.: Coupled dynamics of fast neurons and slow interactions **6**, 447–454 (1994)

[15] Contucci, P., Kurchan, J., Mingione, E.: Equilibrium and dynamics of a multi-bath Sherrington-Kirkpatrick model. Journal of Physics A: Mathematical and Theoretical **52**(32), 324001 (2019)

[16] Contucci, P., Corberi, F., Kurchan, J., Mingione, E.: Stationarization and multithermalization in spin glasses. SciPost Physics **10**(5), 113 (2021)

[17] Alberici, D., Macris, N., Mingione, E.: On the convergence to the non-equilibrium steady state of a langevin dynamics with widely separated time scales and different temperatures. In: Annales Henri Poincaré, pp. 1–62 (2024). Springer

[18] Cugliandolo, L.F., Kurchan, J.: A scenario for the dynamics in the small entropy production limit. Journal of the Physical Society of Japan **69**, 247 (2000)

[19] Bouchaud, J.-P., Cugliandolo, L.F., Kurchan, J., Mézard, M.: Out of equilibrium dynamics in spin-glasses and other glassy systems. Spin glasses and random fields **12**, 161 (1998)

[20] Cugliandolo, L.F.: The effective temperature. Journal of Physics A: Mathematical and Theoretical **44**(48), 483001 (2011)

[21] Franz, S., Mezard, M., Parisi, G., Peliti, L.: The response of glassy systems to random perturbations: A bridge between equilibrium and off-equilibrium. Journal of Statistical Physics **97**, 459–488 (1999) https://doi.org/10.1023/A:1004602906332

[22] Guerra, F.: Broken replica symmetry bounds in the mean field spin glass model. Communications in mathematical physics **233**, 1–12 (2003)

[23] Talagrand, M.: The Parisi formula. Annals of Mathematics **163**(1), 221–263 (2006). Accessed 2024-09-16

[24] Auffinger, A., Chen, W.-K.: The Parisi formula has a unique minimizer. Communications in Mathematical Physics **335** (2014) https://doi.org/10.1007/s00220-014-2254-z

[25] Auffinger, A., Chen, W.-K., Zeng, Q.: The SK model is full-step replica symmetry breaking at zero temperature. arXiv: Probability (2017)

[26] Jagannath, A., Tobasco, I.: Some properties of the phase diagram for mixed p-spin glasses. Probability Theory and Related Fields **167**, 615–672 (2015)

[27] Leuzzi, L., Nieuwenhuizen, T.M.: Thermodynamics of the glassy state (2007)

[28] Barra, A., Biasio, A.D., , Guerra, F.: Replica symmetry breaking in mean-field spin glasses through the Hamilton–Jacobi technique. Journal of Statistical Mechanics: Theory and Experiment, 2010(09) (2010)

[29] Mourrat, J.-C., Panchenko, D.: Extending the Parisi formula along a Hamilton-Jacobi equation. Electronic Journal of Probability **25**, 1–17 (2020) https://doi.org/10.1214/20-EJP432

[30] Hong-Bin Chen, J.-C.M.: On the free energy of vector spin glasses with non-convex interactions. Preprint arXiv:2311.08980 (2023)

[31] Dominguez, T., Mourrat, J.-C.: Statistical mechanics of mean-field disordered systems (2024)

[32] Cui, H., Saglietti, L., Zdeborová, L.: Large deviations in the perceptron model and consequences for active learning. Machine Learning: Science and Technology **2**(4), 045001 (2021)

[33] Cui, H., Saglietti, L., Zdeborová, L.: Large deviations of semisupervised learning in the stochastic block model. Physical Review E **105**(3), 034108 (2022)

[34] Barbier, J., Camilli, F., Ko, J., Okajima, K.: On the phase diagram of extensive-rank symmetric matrix denoising beyond rotational invariance (2024) arXiv:2411.01974 [cond-mat.dis-nn]

[35] Nishimori, H.: Statistical physics of spin glasses and information processing: An introduction. Oxford University Press (2001). https://doi.org/10.1093/acprof:oso/9780198509417.001.0001

[36] Alemanno, F., Camanzi, L., Manzan, G., Tantari, D.: Hopfield model with planted patterns: A teacher-student self-supervised learning model. Applied Mathematics and Computation **458**, 128253 (2023)

[37] Thériault, R., Tantari, D.: Dense Hopfield networks in the teacher-student setting. SciPost Phys. 17, 040 (2024)

[38] Thériault, R., Tosello, F., Tantari, D.: Modelling structured data learning with restricted Boltzmann machines in the teacher-student setting. arXiv preprint arXiv:2410.16150 (2024)

[39] Gerace, F., Saglietti, L., Mannelli, S.S., Saxe, A., Zdeborová, L.: Probing transfer learning with a model of synthetic correlated datasets. Machine Learning: Science and Technology **3**(1), 015030 (2022)

[40] Li, C., Huang, Z., Zou, W., Huang, H.: Statistical mechanics of continual learning: Variational principle and mean-field potential. Physical Review E **108**(1), 014309 (2023)

[41] Ingrosso, A., Pacelli, R., Rotondo, P., Gerace, F.: Statistical mechanics of transfer learning in fully-connected networks in the proportional limit. arXiv preprint arXiv:2407.07168 (2024)

[42] Jagannath, A.: Approximate ultrametricity for random measures and applications to spin glasses. Communications on Pure and Applied Mathematics **70**(4), 611–664 (2017) https://doi.org/10.1002/cpa.21685

https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.21685

[43] Panchenko, D.: The free energy in a multi-species Sherrington–Kirkpatrick model. Annals of Probability **43** (2015)

[44] Franz, S., Parisi, G., Virasoro, M.: Ultrametricity in an inhomogeneous simplest spin glass model. Europhysics Letters **17**(1), 5 (1992)

[45] Genovese, G., Tantari, D.: Overlap synchronisation in multipartite random energy models. Journal of Statistical Physics **169**, 1162–1170 (2017)

[46] Talagrand, M.: Large deviations, Guerra's and A.S.S. schemes, and the Parisi hypothesis. Journal of Statistical Physics **126**, 837–894 (2007) https://doi.org/10.1007/s10955-006-9108-9

[47] Barra, A., Guerra, F., Mingione, E.: Interpolating the Sherrington-Kirkpatrick replica trick. Philosophical Magazine - PHILOS MAG **92** (2011) https://doi.org/10.1080/14786435.2011.637979

[48] Gallavotti, G.: Renormalization theory and ultraviolet stability for scalar fields via renormalization group methods. Rev. Mod. Phys., 57 471 (1985)

[49] Contucci, P., Corberi, F., Kurchan, J., Mingione, E.: Stationarization and multithermalization in spin glasses. SciPost Physics **10**(5), 113 (2021)

[50] Jagannath, A., Tobasco, I.: A dynamic programming approach to the Parisi functional. Proceedings of the American Mathematical Society **144**, 1 (2015) https://doi.org/10.1090/proc/12968

[51] Mourrat, J.-C.: Free energy upper bound for mean-field vector spin glasses. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **59**(3), 1143–1182 (2023) https://doi.org/10.1214/22-AIHP1292

[52] Panchenko, D.: On differentiability of the Parisi formula. Electronic Communications in Probability **13** (2007) https://doi.org/10.1214/ECP.v13-1365

[53] Chewi, S., Pooladian, A.-A.: An entropic generalization of Caffarelli's contraction theorem via covariance inequalities. Comptes Rendus. Mathématique **361**, 1471–1482 (2023) https://doi.org/10.5802/crmath.486

[54] Barra, A., Contucci, P., Mingione, E., Tantari, D., et al.: Multi-species mean-field spin-glasses. rigorous results **16**(3), 691–708 (2015) https://doi.org/10.1007/s00023-014-0341-5