



Parisi Formula for Balanced Potts Spin Glass

Erik Bates¹ , Youngtak Sohn²

¹ Department of Mathematics, North Carolina State University, SAS Hall, 2311 Stinson Drive, NC, Raleigh 27695-8205, USA. E-mail: ebates@ncsu.edu

² Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-4307, USA. E-mail: youngtak@mit.edu

Received: 2 February 2024 / Accepted: 17 July 2024

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract: The Potts spin glass is a generalization of the Sherrington–Kirkpatrick (SK) model that allows for spins to take more than two values. Based on a novel synchronization mechanism, Panchenko (Ann Probab 46(2):829–864, 2018) showed that the limiting free energy is given by a Parisi-type variational formula. The functional order parameter in this formula is a probability measure on a monotone path in the space of positive-semidefinite matrices. By comparison, the order parameter for the SK model is much simpler: a probability measure on the unit interval. Nevertheless, a longstanding prediction by Elderfield and Sherrington (J Phys C Solid State Phys 16(15):L497–L503, 1983) is that the order parameter for the Potts spin glass can be reduced to that of the SK model. We prove this prediction for the balanced Potts spin glass, where the model is constrained so that the fraction of spins taking each value is asymptotically the same. It is generally believed that the limiting free energy of the balanced model is the same as that of the unconstrained model, in which case our results reduce the functional order parameter of Panchenko’s variational formula to probability measures on the unit interval. The intuitive reason—for both this belief and the Elderfield–Sherrington prediction—is that no spin value is a priori preferred over another, and the order parameter should reflect this inherent symmetry. This paper rigorously demonstrates how symmetry, when combined with synchronization, acts as the desired reduction mechanism. Our proof requires that we introduce a generalized Potts spin glass model with mixed higher-order interactions, which is interesting in its own right. We prove that the Parisi formula for this model is differentiable with respect to inverse temperatures. This is a key ingredient for guaranteeing the Ghirlanda–Guerra identities without perturbation, which then allow us to exploit symmetry and synchronization simultaneously.

1. Introduction

1.1. The model. The Potts spin glass was introduced by Elderfield and Sherrington [31] and has been extensively studied in statistical mechanics [12, 14, 28, 32, 35, 48]. In dimension $\kappa \geq 2$, the model is defined as follows. At volume N , the configuration space is the product set Σ^N , where $\Sigma = \{\mathbf{e}_1, \dots, \mathbf{e}_\kappa\}$ is the standard basis in \mathbb{R}^κ . Each configuration $\sigma = (\sigma_1, \dots, \sigma_N) \in \Sigma^N$ is thought of as a $\kappa \times N$ matrix, and its energy is given by the Hamiltonian

$$H_N(\sigma) := \frac{\beta}{\sqrt{N}} \sum_{i,j=1}^N g_{i,j} \mathbb{1}\{\sigma_i = \sigma_j\} = \frac{\beta}{\sqrt{N}} \sum_{i,j=1}^N g_{i,j} \langle \sigma_i, \sigma_j \rangle, \quad (1.1)$$

where $(g_{i,j})_{i,j=1}^N$ are independent standard normal random variables, $\langle \lambda, \sigma \rangle = \lambda^\top \sigma$ is the inner product of $\lambda, \sigma \in \mathbb{R}^\kappa$, and $\beta > 0$ is an inverse temperature parameter. The associated free energy is

$$\mathbf{F}_N := \frac{1}{N} \mathbb{E} \log Z_N, \quad \text{where} \quad Z_N := \sum_{\sigma \in \Sigma^N} \exp H_N(\sigma). \quad (1.2)$$

For $\kappa = 2$, the Hamiltonian (1.1) is equivalent to the classical Sherrington–Kirkpatrick (SK) model [68] by the mapping $\sigma \mapsto \tau = (\tau_1, \dots, \tau_N)$, where

$$\tau_i = \begin{cases} +1 & \text{if } \sigma_i = \mathbf{e}_1 \\ -1 & \text{if } \sigma_i = \mathbf{e}_2. \end{cases} \quad (1.3)$$

By this transformation, the free energy \mathbf{F}_N equals the free energy of the SK model up to rescaling β . In this case, the limiting free energy $\lim_{N \rightarrow \infty} \mathbf{F}_N$ is given by the celebrated Parisi formula, which is a variational expression predicted by Parisi [60–63] but not proved until the seminal work of Guerra [36] and Talagrand [70].

The fundamental insight behind the Parisi formula is that the SK free energy can be understood by keeping track of a single random variable, namely the *replica overlap*. In our context, this is the quantity $N^{-1} \sum_{i=1}^N \langle \sigma_i^1, \sigma_i^2 \rangle$, where σ^1 and σ^2 are independent samples from the Gibbs measure $G_N(\sigma) = \exp H_N(\sigma) / Z_N$. This quantity is exactly the fraction of coordinates at which σ^1 and σ^2 agree. The law of this random variable is a probability measure on $[0, 1]$, and Parisi’s formula is a minimization problem over such measures.

For $\kappa \geq 3$, this perspective runs into difficulty because of the additional degrees of freedom. Namely, the transformation (1.3) does not have a natural generalization, and so there is no obvious way to relate the free energy of the Potts spin glass to a scalar statistic. In principle, one needs to keep track of the entire $\kappa \times \kappa$ matrix $N^{-1} \sigma^1 (\sigma^2)^\top$, whereas the replica overlap is just the trace. Nevertheless, using an ingenious synchronization mechanism, Panchenko [58] showed that in the large- N limit, this overlap matrix is some deterministic map of its trace (Theorem 2.3). This led to a generalized Parisi formula that optimizes over probability measures on $[0, 1]$ *together with* so-called synchronization maps (Theorem 1.1).

The program initiated by this paper is to go even further: there is only one possible choice for the synchronization map. Theorem 1.3 accomplishes this for the balanced Potts spin glass and is presented in the next section.

1.2. Limiting free energy. A crucial fact of synchronization is that it requires the self-overlap matrix $N^{-1}\sigma\sigma^\top \in \mathbb{R}^{\kappa \times \kappa}$ to take a fixed value (see (2.11) in Theorem 2.3). But of course the function $\sigma \mapsto N^{-1}\sigma\sigma^\top$ is not constant over the configuration space Σ^N ; it can be any diagonal matrix whose entries belong to $\{0, \frac{1}{N}, \dots, \frac{N-1}{N}, 1\}$ and sum to 1. Therefore, the strategy of Panchenko [57] is to consider subsets of Σ^N on which this function is approximately constant, and then derive a Parisi formula for the model constrained to these subsets. The number of subsets needed grows only polynomially in N , whereas (1.2) concerns an exponential growth rate. Therefore, classical Gaussian concentration allows one to determine that the limiting free energy of the unconstrained model is simply the largest limiting free energy among the constrained models. We now proceed to make things precise.

Denote the κ -dimensional unit simplex by

$$\mathcal{D} := \left\{ d = (d_1, \dots, d_\kappa) \in [0, 1]^\kappa : \sum_{k=1}^{\kappa} d_k = 1 \right\}.$$

An element $d \in \mathcal{D}$ is called a *magnetization* of the Potts spin glass. The configuration space with magnetization d and approximation parameter $\varepsilon \geq 0$ is

$$\Sigma^N(d, \varepsilon) := \left\{ \sigma \in \Sigma^N : \left| \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{\sigma_i = \mathbf{e}_k\} - d_k \right| \leq \varepsilon \text{ for each } k \in \{1, \dots, \kappa\} \right\}. \quad (1.4)$$

The associated *constrained* free energy is

$$F_N(d, \varepsilon) := \frac{1}{N} \mathbb{E} \log Z_N(d, \varepsilon), \quad \text{where} \quad Z_N(d, \varepsilon) := \sum_{\sigma \in \Sigma^N(d, \varepsilon)} \exp H_N(\sigma). \quad (1.5)$$

When $\varepsilon = 0$, we write

$$\Sigma^N(d) := \Sigma^N(d, 0), \quad Z_N(d) := Z_N(d, 0), \quad F_N(d) := F_N(d, 0).$$

This last free energy only makes sense if $\Sigma^N(d)$ is nonempty, which occurs precisely when d belongs to the set

$$\mathcal{D}_N := \mathcal{D} \cap (\mathbb{Z}/N)^\kappa = \{d \in \mathcal{D} : \Sigma^N(d) \neq \emptyset\}. \quad (1.6)$$

Next we introduce the order parameter for Panchenko's variational formula. Define

$$\Gamma_\kappa := \{\gamma \in \mathbb{R}^{\kappa \times \kappa} : \gamma \text{ is symmetric and positive-semidefinite}\}.$$

Given $d \in \mathcal{D}$, let $\Gamma_\kappa(d)$ be the subset of Γ_κ consisting of matrices with nonnegative entries whose row sums are given by d :

$$\Gamma_\kappa(d) := \left\{ \gamma \in \Gamma_\kappa \cap [0, 1]^{\kappa \times \kappa} : \sum_{k'=1}^{\kappa} \gamma_{k,k'} = d_k \text{ for each } k \in \{1, \dots, \kappa\} \right\}, \quad d \in \mathcal{D}. \quad (1.7)$$

We write \preceq for the positive-semidefinite order on symmetric matrices. Then consider the following collection of paths on $\Gamma_\kappa(d)$:

$$\Pi_d := \{\pi : (0, 1] \rightarrow \Gamma_\kappa(d) : \pi \text{ is left-continuous, } \pi(s) \preceq \pi(t) \text{ if } s \leq t\}. \quad (1.8)$$

To place this definition into context with the previous section: each $\pi \in \Pi_d$ is the combination of a probability measure μ on $[0, 1]$ with a synchronization map $\Phi : [0, 1] \rightarrow \Gamma_\kappa(d)$. More specifically, π is the composition $\Phi \circ Q_\mu$, where Q_μ is the quantile function of μ . See Remark 1.4 for an important special case, and Sect. 2.1 for the origin of the map Φ .

The following result summarizes the outcome of Panchenko's synchronization scheme for the Potts spin glass. The statement below combines the results of two papers. The Parisi functional \mathcal{P} is defined in Sect. 1.4, specifically (1.35) with ξ given in (1.15).

Theorem 1.1. [57, 58] *There is an explicit functional $\mathcal{P} : \cup_{d \in \mathcal{D}} \Pi_d \rightarrow \mathbb{R}$ such that for every $d \in \mathcal{D}$, the constrained free energy has the following limit:*

$$\lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} F_N(d, \varepsilon) = \lim_{\varepsilon \searrow 0} \liminf_{N \rightarrow \infty} F_N(d, \varepsilon) = \inf_{\pi \in \Pi_d} \mathcal{P}(\pi). \quad (1.9)$$

Furthermore, the limiting unconstrained free energy is given by

$$\lim_{N \rightarrow \infty} F_N = \sup_{d \in \mathcal{D}} \inf_{\pi \in \Pi_d} \mathcal{P}(\pi). \quad (1.10)$$

Remark 1.2. In [58], the definition of Π_d included additional stipulations that $\pi(0) = 0$ and $\pi(1) = \text{diag}(d)$. The first condition is unimportant because the value of $\pi(0)$ plays no role in determining the value of $\mathcal{P}(\pi)$. Regarding the second condition $\pi(1) = \text{diag}(d)$, our definition (1.8) only implies $\pi(1) \preceq \text{diag}(d)$, since $\text{diag}(d)$ is the maximal element of $\Gamma_\kappa(d)$ (Lemma A.3). The infimum in (1.9) is not sensitive to this difference because every $\pi \in \Pi_d$ can be approximated by $\tilde{\pi} \in \Pi_d$ such that $\tilde{\pi}(1) = \text{diag}(d)$, and the Parisi functional is continuous (Proposition 2.13(b)). On the other hand, our definition guarantees the existence of a minimizer.

It was further predicted by Elderfield and Sherrington [31] when they introduced the Potts spin glass model that the variational expression $\sup_{d \in \mathcal{D}} \inf_{\pi \in \Pi_d} \mathcal{P}(\pi)$ in (1.10) is achieved on a restricted set. In our notation, Elderfield and Sherrington [31, Eq. (15)] predicts the following:

- (a) The supremum over $d \in \mathcal{D}$ is achieved at the balanced case $d = d_{\text{bal}}$, where

$$d_{\text{bal}} := \kappa^{-1} \mathbf{1},$$

and $\mathbf{1} \in \mathbb{R}^\kappa$ is the vector of all ones. This prediction was echoed in [58, Rmk. 3].

- (b) The infimum over $\pi \in \Pi_{d_{\text{bal}}}$ is achieved by some π such that for every $t \in [0, 1]$, the matrix $\pi(t)$ is constant on its diagonal and also constant off its diagonal.

In fact, these predictions were central to analysis of [31] regarding the phase transitions of the Potts spin glass. See [12, Sec. VI.H.1] for further discussion.

Our main result confirms prediction (b). To this end, let $\mathbf{I}_\kappa \in \mathbb{R}^{\kappa \times \kappa}$ denote the identity matrix, and consider the following subset of $\Gamma_\kappa(d_{\text{bal}})$:

$$\Gamma^\star := \left\{ \gamma \in \Gamma_\kappa : \gamma = \frac{q}{\kappa} \mathbf{I}_\kappa + \frac{1-q}{\kappa^2} \mathbf{1} \mathbf{1}^\top \text{ for some } q \in [0, 1] \right\}. \quad (1.11)$$

Then define the corresponding subset of the path space $\Pi_{d_{\text{bal}}}$:

$$\Pi^* := \{\pi : (0, 1] \rightarrow \Gamma^* : \pi \text{ is left-continuous, } \pi(s) \leq \pi(t) \text{ if } s \leq t\}. \quad (1.12)$$

Theorem 1.3. *For $d_{\text{bal}} = \kappa^{-1}\mathbf{1}$, we have*

$$\lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} F_N(d_{\text{bal}}, \varepsilon) = \lim_{\varepsilon \searrow 0} \liminf_{N \rightarrow \infty} F_N(d_{\text{bal}}, \varepsilon) = \inf_{\pi \in \Pi^*} \mathcal{P}(\pi).$$

Moreover, for any $d^N \in \mathcal{D}_N$ converging to d_{bal} as $N \rightarrow \infty$, we have

$$\lim_{N \rightarrow \infty} F_N(d^N) = \inf_{\pi \in \Pi^*} \mathcal{P}(\pi).$$

We note that Theorem 1.3 is only interesting when $\kappa \geq 3$ since for $\kappa = 2$, we have $\Pi^* = \Pi_{d_{\text{bal}}}$. For $\kappa \geq 3$, however, Γ^* remains a 1-dimensional space, whereas $\Gamma_\kappa(d_{\text{bal}})$ has dimension $\frac{\kappa(\kappa-1)}{2}$. Theorem 1.3 thus shows that when $d = d_{\text{bal}}$, the infimum in (1.10) can be taken over a much smaller space. In this way Theorem 1.3 can be viewed as a deterministic statement on top of Theorem 1.1, although our proof is probabilistic. In fact, we will deduce Theorem 1.3 from a more general result, Theorem 1.8, presented in Sect. 1.3.

One may naively hope to prove Theorem 1.3 by showing that $\Pi_{d_{\text{bal}}} \ni \pi \mapsto \mathcal{P}(\pi)$ is convex, since then the infimum in (1.10) must be achieved at a “symmetric” π , by which we mean $\pi \in \Pi^*$. But unfortunately this convexity fails when $\kappa \geq 3$, because of complications in comparing synchronization maps with different images. In fact, a recurring challenge in settings relying on synchronization—such as multi-species models or vector spin models—is to extract information about the minimizer of the Parisi functional without relying on convexity, e.g. [8, 9, 29]. Theorem 1.3 is a new example of how to accomplish this, and we anticipate it can actually enable convexity arguments as in the SK model [4, 13, 40, 50]. This is because the synchronization map is now fixed, as explained below in Remark 1.4. Indeed, following the initial release of this manuscript, Chen [16] proved a version of Theorem 1.3 for the model with self-overlap correction, and the resulting Parisi functional is strictly convex [18].

Remark 1.4. There is a one-to-one correspondence between Π^* and $\text{Prob}([0, 1])$, the set of Borel probability measures on $[0, 1]$. Given $\mu \in \text{Prob}([0, 1])$, denote its quantile function by $Q_\mu(t) = \inf\{q \geq 0 : \mu([0, q]) \geq t\}$. Then the correspondence is achieved through the relation $\pi = \Phi^* \circ Q_\mu$, where $\Phi^*(q) = q\kappa^{-1}\mathbf{I}_\kappa + (1 - q)\kappa^{-2}\mathbf{1}\mathbf{1}^\top$. In this way, the functional order parameter in the balanced Potts spin glass is understood as an element of $\text{Prob}([0, 1])$. Moreover, if prediction (a) is also true, then Theorems 1.1 and 1.3 together express the unconstrained free energy $\lim_{N \rightarrow \infty} F_N$ as an infimum over probability measures on $[0, 1]$, just as for the classical SK model.

Regardless of the status of prediction (a), the balanced Potts spin glass has an important application in combinatorial optimization. Namely, denote by $\text{MaxCut}(G, \kappa)$ the size of the maximum cut of graph G into κ parts, and consider either the sparse Erdős–Rényi graph $G_N \sim G(N, d/N)$ or the sparse random regular graph $G_N \sim G_{\text{reg}}(N, d)$ with degree d . It was shown by Sen [67, Sec. 2.2] that for d sufficiently large, the following asymptotic holds with high probability as $N \rightarrow \infty$:

$$\frac{\text{MaxCut}(G_N, \kappa)}{N} = \frac{d}{2} \left(1 - \frac{1}{\kappa}\right) + P_\star(\kappa) \frac{\sqrt{d}}{2} + o(\sqrt{d}),$$

where $P_\star(\kappa)$ denotes the ground-state energy of the balanced Potts spin glass model:

$$P_\star(\kappa) := \lim_{\beta \rightarrow \infty} \frac{1}{\beta} \inf_{\pi \in \Pi_{d_{\text{bal}}}} \mathcal{P}(\pi; \beta).$$

Here we are making explicit the dependence of the Parisi functional on β . Now Theorem 1.3 simplifies this to a lower-dimensional minimization problem:

$$P_\star(\kappa) = \lim_{\beta \rightarrow \infty} \frac{1}{\beta} \inf_{\pi \in \Pi^\star} \mathcal{P}(\pi; \beta). \quad (1.13)$$

For $\kappa = 2$ (in which case $\Pi_{d_{\text{bal}}} = \Pi^\star$), numerical studies [27, 66] give $P_\star(2) \approx 0.763166$. Since Π^\star has no dimensional dependence on κ (unlike $\Pi_{d_{\text{bal}}}$), (1.13) opens the possibility of numerically computing $P_\star(\kappa)$ for $\kappa \geq 3$.

1.3. Potts spin glass with general covariance function. We will obtain Theorem 1.3 from the analogous result for a more general Hamiltonian. But this is not simply for the sake of generality: our proof of Theorem 1.3 actually requires that we first replace (1.1) with a Hamiltonian that includes higher-order interactions. Namely, we assume henceforth that $(H_{N,\xi}(\sigma))_{\sigma \in \Sigma^N}$ is a centered Gaussian process with covariance

$$\mathbb{E}[H_{N,\xi}(\sigma)H_{N,\xi}(\sigma')] = N\xi\left(\frac{\sigma\sigma'^T}{N}\right), \quad \sigma, \sigma' \in \Sigma^N, \quad (1.14)$$

where $\xi: \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}$ is any function satisfying assumptions (A1) and (A2) below. The ordinary Potts spin glass (1.1) is the special case

$$\xi(R) = \beta^2 \sum_{k,k'=1}^{\kappa} R_{k,k'}^2 = \beta^2 \text{tr}(R^T R). \quad (1.15)$$

To state our more general assumptions, we require some definitions.

Let $\mathbb{N} = \{1, 2, \dots\}$. Then define the following set of parameters:

$$\Theta := \left\{ (p, m, n_1, \dots, n_m, w_1, \dots, w_m) : p, m, n_1, \dots, n_m \in \mathbb{N}, w_1, \dots, w_m \in [-1, 1]^{\kappa} \right\}. \quad (1.16)$$

For $\theta \in \Theta$, define the function $\xi_\theta: \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}$ by

$$\xi_\theta(R) := \prod_{j=1}^m \langle R^{\circ p} w_j, w_j \rangle^{n_j}, \quad (1.17)$$

where $R^{\circ p}$ denotes the p th Hadamard power of the matrix R , i.e. $(R^{\circ p})_{k,k'} = (R_{k,k'})^p$. The function ξ_θ is a homogeneous polynomial of degree

$$\deg(\theta) := p(n_1 + \dots + n_m). \quad (1.18)$$

Then we assume the following about the covariance function $\xi: \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}$ in (1.14):

(A1) There exists a countable subset $\Theta_0 \subset \Theta$ and coefficients $(\alpha_\theta)_{\theta \in \Theta_0}$ such that

$$\xi(R) = \sum_{\theta \in \Theta_0} \alpha_\theta^2 \xi_\theta(R) \quad \text{and} \quad \sum_{\theta \in \Theta_0} \alpha_\theta^2 (1 + \varepsilon)^{\deg(\theta)} < \infty \quad \text{for some } \varepsilon > 0.$$

(A2) ξ is convex when restricted to the following set:

$$\mathcal{R}_\kappa := \{R \in [0, 1]^{\kappa \times \kappa} : \|R\|_1 \leq 1\}. \quad (1.19)$$

Remark 1.5. We always have

$$\|\xi_\theta(R)\| \leq \|R\|_1^{\deg(\theta)}. \quad (1.20)$$

Therefore, the decay condition on α_θ in (A1) ensures that $\sum_{\theta \in \Theta_0} \alpha_\theta^2 \xi_\theta(R)$ converges uniformly on the set $\{R \in \mathbb{R}^{\kappa \times \kappa} : \|R\|_1 \leq 1\}$. In particular, ξ is smooth on this set. Outside this set, the sum may diverge, and this is okay because we will only ever need to apply ξ on the set \mathcal{R}_κ defined in (A2). This is because for any $\sigma, \sigma' \in \Sigma^N$, we have $N^{-1} \sigma \sigma'^T \in \mathcal{R}_\kappa$. Nevertheless, we will continue to write $\mathbb{R}^{\kappa \times \kappa}$ as the domain of ξ simply to remind the reader of the ambient matrix space.

Remark 1.6. Comparing (1.15) and (1.17), one sees that the ordinary Potts spin glass (1.1) corresponds to the case when Θ_0 has a single element: $(p = 1, m = 1, n = 2, w = 1)$.

Extending the notation of (1.5), we write

$$F_{N,\xi}(d, \varepsilon) = \frac{1}{N} \mathbb{E} \log Z_{N,\xi}(d, \varepsilon), \quad \text{where} \quad Z_{N,\xi}(d, \varepsilon) = \sum_{\sigma \in \Sigma^N(d, \varepsilon)} \exp H_{N,\xi}(\sigma), \quad (1.21)$$

and $F_{N,\xi}(d) = F_{N,\xi}(d, 0)$ and $Z_{N,\xi}(d) = Z_{N,\xi}(d, 0)$ as before. We first generalize Theorem 1.1 to this setting.

Theorem 1.7. Assume $\xi : \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}$ satisfies (A1) and (A2). Then for any $d \in \mathcal{D}$,

$$\lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} F_{N,\xi}(d, \varepsilon) = \lim_{\varepsilon \searrow 0} \liminf_{N \rightarrow \infty} F_{N,\xi}(d, \varepsilon) = \inf_{\pi \in \Pi_d} \mathcal{P}_\xi(\pi), \quad (1.22a)$$

where the Parisi functional $\mathcal{P}_\xi(\pi)$ is defined in (1.35). Moreover, for any $d^N \in \mathcal{D}_N$ such that $d^N \rightarrow d$ as $N \rightarrow \infty$, we have

$$\lim_{N \rightarrow \infty} F_{N,\xi}(d^N) = \inf_{\pi \in \Pi_d} \mathcal{P}_\xi(\pi). \quad (1.22b)$$

More importantly, we show that when $d = d_{\text{bal}}$ and ξ has a certain symmetry—which is the case for the usual Potts spin glass (1.15)—the Parisi formula (1.22) can be reduced to have order parameter $\pi \in \Pi^*$. First we define the symmetry condition. Let S_κ denote the symmetric group on $\{1, \dots, \kappa\}$. Given a permutation $\omega \in S_\kappa$ and a matrix $R \in \mathbb{R}^{\kappa \times \kappa}$, let $\omega \bullet R$ be the matrix obtained by permuting the rows and columns according to ω . That is,

$$(\omega \bullet R)_{k,k'} := R_{\omega^{-1}(k), \omega^{-1}(k')}, \quad k, k' \in \{1, \dots, \kappa\}. \quad (1.23)$$

We then make the following assumption:

$$\xi(\omega \bullet R) = \xi(R) \quad \text{for every } \omega \in S_\kappa \text{ and } R \in \mathbb{R}^{\kappa \times \kappa}. \quad (A3)$$

The following is our main result: it generalizes Theorem 1.3 to the setting (1.14).

Theorem 1.8. Assume $\xi: \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}$ satisfies (A1), (A2), (A3). Then for any $d^N \in \mathcal{D}_N$ such that $d^N \rightarrow d_{\text{bal}}$ as $N \rightarrow \infty$, we have

$$\lim_{N \rightarrow \infty} F_{N,\xi}(d^N) = \lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} F_{N,\xi}(d_{\text{bal}}, \varepsilon) = \lim_{\varepsilon \searrow 0} \liminf_{N \rightarrow \infty} F_{N,\xi}(d_{\text{bal}}, \varepsilon) = \inf_{\pi \in \Pi^*} \mathcal{P}_\xi(\pi), \quad (1.24)$$

where Π^* is defined in (1.12) and $\mathcal{P}_\xi(\pi)$ is defined in (1.35) (see also (1.36)).

We close this section by explaining why Theorem 1.8 only applies to the balanced case $d = d_{\text{bal}}$. Given a permutation $\omega \in S_\kappa$ and a vector $u = (u_1, \dots, u_\kappa) \in \mathbb{R}^\kappa$, let $\omega \bullet u$ be the vector obtained by permuting the coordinates of u :

$$(\omega \bullet u)_k = u_{\omega^{-1}(k)}, \quad k \in \{1, \dots, \kappa\}. \quad (1.25a)$$

In particular, for any standard basis vector \mathbf{e}_a , we have $\omega \bullet \mathbf{e}_a = \mathbf{e}_{\omega(a)}$. Now extend this action to $\sigma = (\sigma_1, \dots, \sigma_N) \in \Sigma^N$ by

$$\omega \bullet \sigma = (\omega \bullet \sigma_1, \dots, \omega \bullet \sigma_N), \quad \sigma \in \Sigma^N. \quad (1.25b)$$

The original Potts Hamiltonian (1.1) is clearly unchanged in distribution under any such action. This fact is generalized by assumption (A3), which implies (see Lemma 6.7(a))

$$(H_{N,\xi}(\sigma))_{\sigma \in \Sigma^N} \stackrel{\text{law}}{=} (H_{N,\xi}(\omega \bullet \sigma))_{\sigma \in \Sigma^N}. \quad (1.26)$$

Unfortunately this symmetry is not particularly helpful for the constrained model (1.21), since the constrained configuration space $\Sigma^N(d)$ from (1.4) is not invariant under these permutations. But there is one exception: when $d = d_{\text{bal}}$. In this case, the map $\sigma \mapsto \omega \bullet \sigma$ is a bijection $\Sigma^N(d_{\text{bal}}) \rightarrow \Sigma^N(d_{\text{bal}})$. The central objective of the paper is to capitalize on this basic observation. Roughly speaking, Theorem 1.8 is the statement that when $d = d_{\text{bal}}$, the order parameter $\pi: (0, 1] \rightarrow \Gamma_\kappa(d_{\text{bal}})$ in (1.34) must reflect the symmetry offered by (1.26). That is, π should actually map to the space Γ^* defined in (1.11), which is precisely the subset of $\Gamma_\kappa(d_{\text{bal}})$ consisting of those matrices γ which satisfy $\omega \bullet \gamma = \gamma$ for every $\omega \in S_\kappa$.

1.4. Parisi functional. In this section we define the Parisi functional \mathcal{P}_ξ for any ξ satisfying (A1). We denote the derivative of ξ by $\nabla \xi: \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}^{\kappa \times \kappa}$, which is a matrix-valued function with entries given by

$$[\nabla \xi(R)]_{k,k'} := \frac{\partial \xi}{\partial R_{k,k'}}(R), \quad k, k' \in \{1, \dots, \kappa\}.$$

In addition, define the function $\vartheta_\xi: \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}$ by

$$\vartheta_\xi(R) := \langle R, \nabla \xi(R) \rangle - \xi(R), \quad (1.27)$$

where $\langle A, B \rangle = \text{tr}(A^\top B)$ denotes the inner product of two matrices A, B . Under assumption (A1), the functions ξ , $\nabla \xi$, and ϑ_ξ are nondecreasing when restricted to the set Γ_κ of positive-semidefinite matrices; see Proposition 2.7(b).

Recall the set of paths Π_d from (1.8). We first define $\mathcal{P}_\xi(\pi)$ for any discrete path π , and then extend continuously to general π . By a discrete path we mean an element of the set

$$\Pi_d^{\text{disc}} := \{\pi \in \Pi_d : \pi \text{ takes only finitely many values, } \pi(1) = \text{diag}(d)\}. \quad (1.28)$$

Thus every $\pi \in \Pi_d^{\text{disc}}$ has the form

$$\pi(t) = \gamma_r \quad \text{for } t \in (m_{r-1}, m_r], r \in \{1, \dots, s\}, \quad (1.29a)$$

for some sequence of weights

$$0 = m_0 < m_1 < \dots < m_{s-1} < m_s = 1, \quad (1.29b)$$

and some sequence of matrices

$$0 = \gamma_0 < \gamma_1 < \dots < \gamma_{s-1} < \gamma_s = \text{diag}(d). \quad (1.29c)$$

Given these sequences, consider independent centered Gaussian random vectors z_0, \dots, z_{s-1} in \mathbb{R}^κ , with covariance structure

$$\mathbb{E}(z_r z_r^\top) = \nabla \xi(\gamma_{r+1}) - \nabla \xi(\gamma_r) \mathbf{1}\{r > 0\}. \quad (1.30)$$

Given a parameter $\lambda = (\lambda_1, \dots, \lambda_\kappa) \in \mathbb{R}^\kappa$ which serves as a Lagrange multiplier, define

$$X_s = \log \sum_{k=1}^{\kappa} \exp \left\langle \sum_{r=0}^{s-1} z_r + \lambda, \mathbf{e}_k \right\rangle.$$

Using \mathbb{E}_r to denote expectation with respect to z_r , we then define inductively

$$X_r = \frac{1}{m_r} \log \mathbb{E}_r \exp(m_r X_{r+1}) \quad \text{for } r \in \{1, \dots, s-1\}, \quad \text{and } X_0 = \mathbb{E}_0(X_1).$$

We then record the non-random quantity X_0 as a function of π and λ :

$$\mathcal{P}_\xi^{(1)}(\pi, \lambda) := X_0. \quad (1.31)$$

We show in Lemma 5.1 that $\Pi_d^{\text{disc}} \ni \pi \mapsto \mathcal{P}_\xi^{(1)}(\pi, \lambda)$ is Lipschitz continuous with respect to the following L^1 norm on paths:

$$\pi \mapsto \int_0^1 \|\pi(t)\|_1 dt. \quad (1.32)$$

Since $\Pi_d^{\text{disc}} \subset \Pi_d$ is dense with respect to this norm, there is a unique continuous extension of $\mathcal{P}_\xi^{(1)}$ to all of Π_d . Next define

$$\mathcal{P}_\xi^{(2)}(\pi) := \frac{1}{2} \int_0^1 \vartheta_\xi(\pi(t)) dt - \frac{1}{2} \vartheta_\xi(\text{diag}(d)), \quad \pi \in \Pi_d, \quad (1.33)$$

and set

$$\mathcal{P}_\xi(\pi, \lambda) := \mathcal{P}_\xi^{(1)}(\pi, \lambda) + \mathcal{P}_\xi^{(2)}(\pi). \quad (1.34)$$

Finally, the Parisi functional is defined as

$$\mathcal{P}_\xi(\pi) := \inf_{\lambda \in \mathbb{R}^\kappa} [\mathcal{P}_\xi(\pi, \lambda) - \langle \lambda, d \rangle], \quad \pi \in \Pi_d. \quad (1.35)$$

Remark 1.9. The reason for the Lagrange multiplier λ and the infimum in (1.35) is later explained in Remark 3.8. It is straightforward to check that $\mathcal{P}_\xi(\pi, \lambda) = \mathcal{P}_\xi(\pi, \lambda + c\mathbf{1})$ holds for any $c \in \mathbb{R}$. Therefore, one could fix the last coordinate $\lambda_\kappa = 0$ so that the infimum in (1.35) is over $\mathbb{R}^{\kappa-1}$; this was done in [58]. More importantly, we show in Lemma 5.4 that for $\pi \in \Pi^*$ and ξ satisfying the symmetry condition (A3), the infimum is achieved at $\lambda = 0$:

$$\mathcal{P}_\xi(\pi) = \mathcal{P}_\xi(\pi, 0) \quad \text{for any } \pi \in \Pi^*. \quad (1.36)$$

Consequently, Theorem 1.8 shows that the Lagrange parameter λ is not needed for symmetric models, although it remains necessary for the proof.

1.5. Related literature. The Potts spin glass model is closely related to the max κ -cut problem on random graphs. As mentioned after Remark 1.4, the balanced case addressed in this paper corresponds to sparse Erdős–Rényi graphs or random regular graphs [67], both of which have distributional symmetry analogous to (A3). An unbalanced version (with $\kappa = 2$) was studied in [39], corresponding to the generalized SK model introduced in [49]. Motivated by the max κ -cut problem on inhomogeneous random graphs, Jaganath et al. [38] introduced a vector version of multi-species SK model and combined approaches from [56, 58] to obtain a Parisi formula for the limiting free energy. As in this paper, a key technical component was the synchronization mechanism discussed in Sect. 1.1.

The synchronization technique was first introduced by Panchenko [56] to study the multi-species SK model. More specifically, synchronization enabled a free energy lower bound that matched the upper bound obtained in [7] via Guerra’s replica symmetry breaking interpolation [36]. These methods were generalized in [58] to obtain the variational formula (1.10), which is a special case of the mixed vector spin model considered in [57]. Synchronization has since been used in a variety of generalized spin glasses, including a multi-scale SK model [26], the quantum SK model [1], spherical spin glasses with constrained overlaps [42], and multi-species spherical models [10]. These generalized models are related to a number of problems in statistical inference and combinatorial optimization, such as spiked random tensors [22, 23], principal submatrix recovery [33], and the ℓ^p –Gaussian–Grothendieck problem [25, 30].

Synchronization also been used in conjunction with a strategy initiated by Mourrat [43, 45] that identifies the Parisi formula as a solution to a Hamilton–Jacobi equation. This program has also led to results on spin glasses enriched by a certain magnetic field [47], non-convex models such as the bipartite SK model [44], and vector spin models [20, 46]. Building on these developments, Chen [19] recently showed that by introducing a self-overlap correction term in the free energy, one can remove the supremum over magnetizations in (1.10). Moreover, the self-overlap concentrates in that setting, thereby softly enforcing a balanced constraint [17].

For spherical vector spin glasses, Ko [41] obtained the Crisanti–Sommers variational formula for the limiting free energy. Various properties of the minimizer to this formula were obtained by Auffinger and Zhou [6], who also extended the formula to zero temperature. In the SK case, the limiting free energy was computed earlier by Panchenko and Talagrand [59, Thm. 2]. More recently, Husson and Ko [37] gave an alternative proof using spherical integrals, even allowing κ to grow sublinearly with N .

1.6. Organization of the paper. The paper’s format is meant to prioritize new and central ideas, by postponing certain technical aspects that could be otherwise distracting. Section 2 contains the proof of our main result, Theorem 1.8. The proof invokes various inputs that are introduced in Sect. 2 but proved later in the paper. Section 3 establishes some notation involving the Parisi functional and Ruelle probability cascades that will be needed in all subsequent parts. Section 4 concerns the differentiability of the Parisi functional with respect to certain inverse temperatures, while Sect. 5 contains various continuity properties of the Parisi functional. Section 6 unites these ingredients with symmetry to complete the lower bound portion of Theorem 1.8.

Several intermediate results are either standard or very similar to previous works. For some of these results, the proofs can be read essentially verbatim elsewhere and are thus omitted. For others, we defer the proofs to one of several appendices. These include the existence of certain Gaussian processes (“Appendix A”), proof of the Aizenman–Sims–Starr scheme (“Appendix B”), checking various lemmas involving Ruelle probability cascades (“Appendix C”), generalizing a variational argument from [58] related to the infimum in (1.35) (“Appendix D”), and finally verification of the upper bound via Guerra interpolation (“Appendix E”).

2. Proof Overview

We introduce a series of intermediate results in Sects. 2.1–2.3, and then use them to prove Theorem 1.8 (and Theorem 1.3) in Sect. 2.4. The results labeled as propositions will be proved in the remainder of the paper, while lemmas are argued immediately. Meanwhile, Theorem 1.7 is proved using the same strategy as in [58], and so we will not focus on it here. Instead a review of the proof is provided in “Appendix E”.

Toward our goal of Theorem 1.8, we first observe that the upper bound

$$\lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} F_{N,\xi}(d_{\text{bal}}, \varepsilon) \leq \inf_{\pi \in \Pi^*} \mathcal{P}_\xi(\pi)$$

is immediate from Theorem 1.7, since Π^* is a subset of $\Pi_{d_{\text{bal}}}$. What requires novel justification is the lower bound, which amounts to showing

$$\liminf_{N \rightarrow \infty} F_{\kappa N, \xi}(d_{\text{bal}}) \geq \mathcal{P}_\xi(\pi) \quad \text{for some } \pi \in \Pi^*. \quad (2.1)$$

The factor κ in κN is to guarantee that $\Sigma^{\kappa N}(d_{\text{bal}})$ is nonempty.

The standard approach for proving a Parisi formula lower bound such as (2.1) is the Aizenman–Sims–Starr (A.S.S.) scheme [2, 3]. Indeed, this method is what yields the lower bound portion of Theorem 1.7. The shortcoming is that this implementation alone yields a minimizer π belonging to $\Pi_{d_{\text{bal}}}$, whereas we want more specific information, namely that π can be found in Π^* . So we will actually use the A.S.S. scheme a *second* time, but with some interventions that preserve the symmetry needed to infer Theorem 1.8 from Theorem 1.7. The next three sections explain how this is accomplished.

2.1. Synchronized overlap arrays. The main characters in the A.S.S. scheme are overlap arrays. Given any $\sigma^1, \sigma^2, \dots \in \Sigma^N$, we can define an array $\mathcal{R} = (\mathcal{R}_{\ell, \ell'})_{\ell, \ell' \geq 1}$, where $\mathcal{R}_{\ell, \ell'}$ is the $\kappa \times \kappa$ overlap matrix associated to σ^ℓ and $\sigma^{\ell'}$:

$$\mathcal{R}_{\ell, \ell'} = R(\sigma^\ell, \sigma^{\ell'}) := \frac{\sigma^\ell(\sigma^{\ell'})^\top}{N} = \frac{1}{N} \sum_{i=1}^N \sigma_i^\ell (\sigma_i^{\ell'})^\top \in \mathbb{R}^{\kappa \times \kappa}. \quad (2.2)$$

In other words, the (k, k') entry of $R(\sigma, \sigma')$ is the fraction of columns i that satisfy $\sigma_i = \mathbf{e}_k$ and $\sigma'_i = \mathbf{e}_{k'}$:

$$R(\sigma, \sigma')_{k,k'} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{\sigma_i = \mathbf{e}_k\} \mathbb{1}\{\sigma'_i = \mathbf{e}_{k'}\}. \quad (2.3)$$

The case of interest is when $(\sigma^\ell)_{\ell \geq 1}$ are independent samples from the Gibbs measure $G_{\kappa N, \xi}$ associated to the Hamiltonian $H_{\kappa N, \xi}$ from (1.14):

$$G_{\kappa N, \xi}(\sigma) \propto \exp H_{\kappa N, \xi}(\sigma), \quad \sigma \in \Sigma^{\kappa N}(d_{\text{bal}}). \quad (2.4)$$

This measure is random depending on the Gaussian process $H_{\kappa N, \xi}$. Therefore, the array \mathcal{R} is generated by first fixing the realization of $G_{\kappa N, \xi}$, and then drawing i.i.d. samples $(\sigma^\ell)_{\ell \geq 1}$ from $G_{\kappa N, \xi}$. We denote the resulting distribution by $\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N, \xi}^{\otimes \infty}))$. This is a probability measure on the space $(\mathbb{R}^{\kappa \times \kappa})^{\mathbb{N} \times \mathbb{N}}$ equipped with the usual product σ -algebra. Its properties generalize those of classical Gram–de Finetti arrays, as captured by the following definition. Recall that $R^{\circ p}$ denotes the p th Hadamard power of matrix R , and the permutation action $R \mapsto \omega \bullet R$ from (1.23).

Definition 2.1. A κ -dimensional Gram–de Finetti array is a random array of $\kappa \times \kappa$ matrices $\mathcal{R} = (\mathcal{R}_{\ell, \ell'})_{\ell, \ell' \geq 1}$ such that for every positive integer p and $w \in \mathbb{R}^\kappa$, the array $\mathcal{Q} = (\langle \mathcal{R}_{\ell, \ell'}^{\circ p} w, w \rangle)_{\ell, \ell' \geq 1}$ is a (1-dimensional) Gram–de Finetti array. That is, \mathcal{Q} is almost surely symmetric and positive-semidefinite, and is exchangeable under any finite permutation applied to its rows and columns simultaneously. In this case, we say $\text{Law}(\mathcal{R})$ is a *Gram–de Finetti law*.

Roughly speaking, the A.S.S. scheme works by identifying a functional Ψ_ξ on Gram–de Finetti laws (see Sect. 3.1) such that

$$\mathbf{F}_{\kappa N, \xi} + o(1) \geq \Psi_\xi(\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N, \xi}^{\otimes \infty}))). \quad (2.5)$$

As a first step to go from (2.5) to (2.1), one assumes (by passing to a subsequence) that

$$\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N, \xi}^{\otimes \infty})) \text{ converges weakly as } N \rightarrow \infty \text{ to some } \mathcal{L}. \quad (2.6)$$

By continuity of Ψ_ξ (see Corollary 3.3), this results in

$$\liminf_{N \rightarrow \infty} \mathbf{F}_{\kappa N, \xi}(d_{\text{bal}}) \geq \Psi_\xi(\mathcal{L}). \quad (2.7)$$

But this limiting law \mathcal{L} is now divorced from the representation (2.2). There is not necessarily any way of realizing \mathcal{L} by sampling from a Gibbs measure, unless \mathcal{L} satisfies a certain family of identities introduced in [58] and recalled in the following definition.

Definition 2.2. A Gram–de Finetti law $\mathcal{L} = \text{Law}(\mathcal{R})$ is said to satisfy the (κ -dimensional) Ghirlanda–Guerra (G.G.) identities if the following holds for every $p \geq 1$, $m \geq 1$, $w_1, \dots, w_m \in [-1, 1]^\kappa$, and bounded measurable function $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$. Define the scalar array

$$\mathcal{A}_{\ell, \ell'} := \varphi(\langle \mathcal{R}_{\ell, \ell'}^{\circ p} w_1, w_1 \rangle, \dots, \langle \mathcal{R}_{\ell, \ell'}^{\circ p} w_m, w_m \rangle). \quad (2.8)$$

Then for any $n \geq 1$ and any bounded measurable function f of the finite subarray $\mathcal{R}^{(n)} = (\mathcal{R}_{\ell, \ell'})_{1 \leq \ell, \ell' \leq n}$, we have

$$\mathbb{E}[f(\mathcal{R}^{(n)}) \mathcal{A}_{1, n+1}] = \frac{1}{n} \mathbb{E}[f(\mathcal{R}^{(n)})] \mathbb{E}[\mathcal{A}_{1, 2}] + \frac{1}{n} \sum_{\ell=2}^n \mathbb{E}[f(\mathcal{R}^{(n)}) \mathcal{A}_{1, \ell}]. \quad (2.9)$$

The theorem quoted below clarifies the role of these identities: to guarantee *synchronization*. In fact, they also guarantee that each matrix in the array is positive-semidefinite. The condition (2.11) is why the Gibbs measure in (2.4) is restricted to $\Sigma^{\kappa N}(d_{\text{bal}})$. This restriction ensures that the array in (2.2) satisfies

$$\mathcal{R}_{\ell,\ell} = \text{diag}(d_{\text{bal}}) = \kappa^{-1} \mathbf{I}_{\kappa} \quad \text{for every } \ell. \quad (2.10)$$

Theorem 2.3. [57, Thm. 4] *Assume $\text{Law}(\mathcal{R})$ is a Gram–de Finetti law in the sense of Definition 2.1 and satisfies the Ghirlanda–Guerra identities in Definition 2.2. Suppose further that there is some deterministic $D \in \Gamma_{\kappa}$ with $\text{tr}(D) = 1$ such that*

$$\mathcal{R}_{\ell,\ell} = D \quad \text{for all } \ell \geq 1, \text{ with probability one.} \quad (2.11)$$

Then there exists a map $\Phi: [0, 1] \rightarrow \Gamma_{\kappa}$ such that with probability one,

$$\mathcal{R}_{\ell,\ell'} = \Phi(\text{tr}(\mathcal{R}_{\ell,\ell'})) \quad \text{for every } \ell, \ell' \geq 1.$$

Furthermore, this map can be taken to be nondecreasing,

$$\Phi(q) \leq \Phi(q') \quad \text{for } q \leq q',$$

and Lipschitz continuous,

$$\|\Phi(q) - \Phi(q')\|_1 \leq C_{\kappa} |q - q'|,$$

for some constant C_{κ} depending only on κ .

So when the κ -dimensional G.G. identities are satisfied, the matrix array $(\mathcal{R}_{\ell,\ell'})_{\ell,\ell' \geq 1}$ is simply a function of the scalar array $(\text{tr}(\mathcal{R}_{\ell,\ell'}))_{\ell,\ell' \geq 1}$. Moreover, this scalar array satisfies the 1-dimensional (or canonical) G.G. identities (see Lemma 6.6). In turn, the 1-dimensional G.G. identities [34] are known to yield the following existence and uniqueness result via ultrametricity [53]. We use \mathcal{Q} and $\tilde{\mathcal{L}}$ (instead of \mathcal{R} and \mathcal{L}) as notational cues that we are speaking of 1-dimensional Gram–de Finetti arrays.

Theorem 2.4. [54, Thm. 2.13, 2.16, and 2.17] *Let μ be a probability measure on $[0, 1]$. There is exactly one Gram–de Finetti law $\tilde{\mathcal{L}}_{\mu} = \text{Law}(\mathcal{Q})$ that has all three of the following properties: (i) $\tilde{\mathcal{L}}_{\mu}$ satisfies the 1-dimensional G.G. identities; (ii) $\text{Law}(\mathcal{Q}_{1,2}) = \mu$; and (iii) $\mathcal{Q}_{\ell,\ell} = 1$ with probability one, for every $\ell \geq 1$. Furthermore, the map $\mu \mapsto \tilde{\mathcal{L}}_{\mu}$ is continuous with respect to weak convergence.*

In summary, every Gram–de Finetti law $\mathcal{L} = \text{Law}(\mathcal{R})$ satisfying the κ -dimensional G.G. identities can be described by a pair (Φ, μ) , where Φ is the synchronization map from Theorem 2.3, and $\mu = \text{Law}(\text{tr}(\mathcal{R}_{1,2}))$. This pair induces a path $\pi = \Phi \circ \mathcal{Q}_{\mu}$, where \mathcal{Q}_{μ} is the quantile function of μ . There is thus a correspondence $\mathcal{L} \leftrightarrow \pi$, and under this correspondence one checks that $\Psi_{\xi}(\mathcal{L}) = \mathcal{P}_{\xi}(\pi)$ (see Lemma 3.9). In our case, the \mathcal{L} of interest satisfies (2.10), which means Φ maps into $\Gamma_{\kappa}(d_{\text{bal}})$, and thus $\pi \in \Pi_{d_{\text{bal}}}$. Hence (2.7) leads to

$$\liminf_{N \rightarrow \infty} F_{\kappa N, \xi}(d_{\text{bal}}) \geq \mathcal{P}_{\xi}(\pi) \quad \text{for some } \pi \in \Pi_{d_{\text{bal}}}. \quad (2.12)$$

To go further and say π belongs to the smaller set Π^* from (1.12), we make a simple but important observation stated in Lemma 2.6 below. We preface the result with a definition.

Definition 2.5. For an array of $\kappa \times \kappa$ matrices $\mathcal{R} = (\mathcal{R}_{\ell, \ell'})_{\ell, \ell' \geq 1}$ and a permutation $\omega \in S_\kappa$, we write $\omega \bullet \mathcal{R}$ to denote the array $(\omega \bullet \mathcal{R}_{\ell, \ell'})_{\ell, \ell' \geq 1}$, where $\omega \bullet \mathcal{R}_{\ell, \ell'}$ is defined in (1.23). When \mathcal{R} is a random array, we say that $\text{Law}(\mathcal{R})$ is *symmetric* if $\text{Law}(\mathcal{R}) = \text{Law}(\omega \bullet \mathcal{R})$ for every $\omega \in S_\kappa$.

The following result shows that once this symmetry condition is added to the hypotheses of Theorem 2.3, each matrix in the array has all of its diagonal entries equal, and all of its off-diagonal entries equal. In the statement below, $\mathcal{R}_{\ell, \ell'}^{k, k'}$ denotes the (k, k') entry of $\mathcal{R}_{\ell, \ell'}$.

Lemma 2.6. Assume $\text{Law}(\mathcal{R})$ satisfies the hypotheses of Theorem 2.3 and is symmetric in the sense of Definition 2.5. Then for all $k_1 \neq k'_1$ and $k_2 \neq k'_2$ and all $\ell, \ell' \geq 1$,

$$\mathcal{R}_{\ell, \ell'}^{k_1, k'_1} = \mathcal{R}_{\ell, \ell'}^{k_2, k'_2} \quad \text{and} \quad \mathcal{R}_{\ell, \ell'}^{k_1, k'_1} = \mathcal{R}_{\ell, \ell'}^{k_2, k'_2} \quad \text{a.s.} \quad (2.13)$$

In particular, if $\mathcal{R}_{\ell, \ell'} \in \Gamma_\kappa(d_{\text{bal}})$ almost surely, then $\mathcal{R}_{\ell, \ell'} \in \Gamma^*$ almost surely.

Proof. Consider any $k, k' \in \{1, \dots, \kappa\}$, $\ell, \ell' \geq 1$, and $\omega \in S_\kappa$. Since $\mathcal{R}_{\ell, \ell'} \stackrel{\text{law}}{=} \omega \bullet \mathcal{R}_{\ell, \ell'}$ by assumption, it follows that

$$(\mathcal{R}_{\ell, \ell'}^{k, k'}, \text{tr}(\mathcal{R}_{\ell, \ell'})) \stackrel{\text{law}}{=} (\mathcal{R}_{\ell, \ell'}^{\omega^{-1}(k), \omega^{-1}(k')}, \text{tr}(\mathcal{R}_{\ell, \ell'})),$$

where we used the fact that $\text{tr}(\mathcal{R}_{\ell, \ell'}) = \text{tr}(\omega \bullet \mathcal{R}_{\ell, \ell'})$ holds for any $\omega \in S_\kappa$. In particular,

$$\mathbb{E}[\mathcal{R}_{\ell, \ell'}^{k, k'} \mid \text{tr}(\mathcal{R}_{\ell, \ell'})] = \mathbb{E}[\mathcal{R}_{\ell, \ell'}^{\omega^{-1}(k), \omega^{-1}(k')} \mid \text{tr}(\mathcal{R}_{\ell, \ell'})] \quad \text{a.s.}$$

On the other hand, by Theorem 2.3, $\mathcal{R}_{\ell, \ell'}$ is measurable with respect to $\text{tr}(\mathcal{R}_{\ell, \ell'})$. Hence

$$\mathcal{R}_{\ell, \ell'}^{k, k'} = \mathbb{E}[\mathcal{R}_{\ell, \ell'}^{k, k'} \mid \text{tr}(\mathcal{R}_{\ell, \ell'})] = \mathbb{E}[\mathcal{R}_{\ell, \ell'}^{\omega^{-1}(k), \omega^{-1}(k')} \mid \text{tr}(\mathcal{R}_{\ell, \ell'})] = \mathcal{R}_{\ell, \ell'}^{\omega^{-1}(k), \omega^{-1}(k')} \quad \text{a.s.}$$

As this holds for any permutation $\omega \in S_\kappa$, (2.13) follows.

To justify the final sentence in the proposition, recall from definition (1.11) that Γ^* is precisely the subset of $\Gamma_\kappa(d_{\text{bal}})$ consisting of matrices that are constant both on the diagonal and off the diagonal. \square

Under the condition of symmetry, Lemma 2.6 specifies exactly what the synchronization map in Theorem 2.3 must be:

$$\Phi^*(q) = \frac{q}{\kappa} \mathbf{I}_\kappa + \frac{1-q}{\kappa^2} \mathbf{1}\mathbf{1}^\top, \quad q \in [0, 1].$$

This suggests that under the assumption (A3), the previous lower bound (2.12) can be improved to (2.1). Indeed, (1.26) implies that $\text{Law}(\mathcal{R}; \mathbb{E}(G_{N, \xi}^{\otimes \infty}))$ is symmetric in the sense of Definition 2.5 (see Lemma 6.7(b)), and this symmetry trivially passes to the limit law \mathcal{L} in (2.6). The crucial detail is that \mathcal{L} must also satisfy the G.G. identities for us to obtain (2.12) in the first place, but there is a major obstacle: the G.G. identities are not known to hold!

There are two natural options for moving forward, both of which run into difficulty:

- (I) The standard approach for obtaining the G.G. identities is to perturb the Hamiltonian by lower-order terms with random coefficients, and then average over these coefficients [34, 72]. Indeed, this strategy was used in [57, 58] to obtain Theorem 1.1. Because the perturbations do not change the limiting free energy, one can obtain a law \mathcal{L} satisfying both (2.7) and the G.G. identities, but it would not necessarily be equal to the limit in (2.6). Instead, the Gibbs measure $G_{\kappa N, \xi}$ would be replaced by one corresponding to the perturbed Hamiltonian, whose covariance function only satisfies the symmetry condition (A3) *asymptotically*. Since Gibbs measures in spin glass theory are known to exhibit temperature chaos [11, 24], this is not enough to ensure \mathcal{L} is still symmetric in the sense of Definition 2.5. Without this property, Lemma 2.6 is not applicable.
- (II) One could instead try to obtain the G.G. identities directly, without any perturbation. This would allow symmetry to be preserved, but the task of proving the G.G. identities (or some version thereof) remains a major open problem even for SK model [71]. Therefore, this approach does not presently seem viable.

It would thus appear that the G.G. identities needed for the existence of a synchronization are at odds with the symmetry needed to say more about the map itself. We thus employ a different strategy: we introduce a *generic* version of the Potts spin glass for which we can obtain the G.G. identities via differentiability, without averaging over perturbations.

Inspired by [52] and [54, Sec. 3.7], the generic model works by augmenting the covariance function ξ with countably many polynomials of the form ξ_θ from (1.17). If these polynomials are selected so that they span a dense subset of all continuous functions, and if the array $(\xi_\theta(R_{\ell, \ell'}))_{\ell, \ell' \geq 1}$ satisfies the 1-dimensional G.G. identities for each θ , then $(R_{\ell, \ell'})_{\ell, \ell' \geq 1}$ will satisfy the κ -dimensional G.G. identities. The same principle underlies approach (I), but here the new terms added to the Hamiltonian will be of the same order as the Hamiltonian itself. Therefore, the limiting free energy of the augmented model will be different from the original, but still fall under the purview of Theorem 1.7. Each new term in the Hamiltonian will be modulated by an inverse temperature parameter β_θ which is fixed and *does not require averaging*. We will ultimately choose these parameters to be sufficiently small, but all nonzero, and preserve symmetry of ξ in (A3). The next section defines this generic model.

2.2. The generic model and differentiability. Recall the set Θ in (1.16) and the function $\xi_\theta: \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}$ in (1.17). Specializing the notation from (1.27), we denote

$$\vartheta_\theta(R) := \vartheta_{\xi_\theta}(R) = \langle R, \nabla \xi_\theta(R) \rangle - \xi_\theta(R). \quad (2.14)$$

The following result provides several important properties of the functions $\xi_\theta, \nabla \xi_\theta, \vartheta_\theta$. Under assumption (A1), parts (b), (c), and (d) obviously remain true when ξ_θ is replaced by ξ .

Proposition 2.7. *For any $\theta \in \Theta$ and $N \geq 1$, the following hold.*

- (a) $\vartheta_\theta = (\deg(\theta) - 1)\xi_\theta$.
- (b) For any $Q, R \in \Gamma_\kappa$ such that $Q \preceq R$, we have

$$0 \leq \xi_\theta(Q) \leq \xi_\theta(R), \quad 0 \leq \nabla \xi_\theta(Q) \leq \nabla \xi_\theta(R), \quad 0 \leq \vartheta_\theta(Q) \leq \vartheta_\theta(R). \quad (2.15)$$

(c) *There exists a centered Gaussian process $H_{N,\theta} : \Sigma^N \rightarrow \mathbb{R}$ with covariance*

$$\mathbb{E}[H_{N,\theta}(\sigma)H_{N,\theta}(\sigma')] = N\xi_\theta\left(\frac{\sigma\sigma'^T}{N}\right). \quad (2.16)$$

(d) *There exist centered Gaussian processes $Z_{N,\theta} : \Sigma^N \rightarrow \mathbb{R}^\kappa$ and $Y_{N,\theta} : \Sigma^N \rightarrow \mathbb{R}$ with covariances*

$$\mathbb{E}\left[Z_{N,\theta}(\sigma)Z_{N,\theta}(\sigma')^T\right] = \nabla\xi_\theta\left(\frac{\sigma\sigma'^T}{N}\right), \quad \mathbb{E}[Y_{N,\theta}(\sigma)Y_{N,\theta}(\sigma')] = \vartheta_\theta\left(\frac{\sigma\sigma'^T}{N}\right).$$

The proof of Proposition 2.7 is given in “Appendix A”. Note that part (b) was used in (1.30) to define the Parisi functional \mathcal{P}_ξ . Part (c) guarantees the existence of the Gaussian process $(H_{N,\xi}(\sigma))_{\sigma \in \Sigma^N}$ in (1.14), under assumption (A1). Finally, part (d) will be needed for the A.S.S. scheme in Sect. 6.1.

While (A1) allows the covariance function $\xi = \sum_{\theta \in \Theta_0} \alpha_\theta^2 \xi_\theta$ to use any countable subset Θ_0 of the parameter space Θ , here we consider a particular countable subset:

$$\Theta_{\mathbb{Q}} = \left\{ (p, m, n_1, \dots, n_m, w_1, \dots, w_m) : p, m, n_1, \dots, n_m \geq 1, w_1, \dots, w_m \in (\mathbb{Q} \cap [-1, 1])^\kappa \right\}.$$

Given ξ and a family of parameters $\beta = (\beta_\theta)_{\theta \in \Theta_{\mathbb{Q}}}$, we define the modified covariance function $\xi_\beta : \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}$ by

$$\xi_\beta(R) := \xi(R) + \sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_\theta^2 \xi_\theta(R). \quad (2.17)$$

As in Remark 1.5, the following decay condition on $(\beta_\theta)_{\theta \in \Theta_{\mathbb{Q}}}$ guarantees that ξ_β is smooth on the set $\{R \in \mathbb{R}^{\kappa \times \kappa} : \|R\|_\infty \leq 1\}$

$$\sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_\theta^2 (1 + \varepsilon)^{\deg(\theta)} < \infty \quad \text{for some } \varepsilon > 0. \quad (\text{B1})$$

Nevertheless, the modified function ξ_β may fail to be convex even if ξ is convex, in which case Theorem 1.7 would not apply to ξ_β . To avoid this scenario, we impose an additional decay condition on $(\beta_\theta)_{\theta \in \Theta_{\mathbb{Q}}}$ as follows.

For twice continuously differentiable $f : \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}$, denote the Hessian of f at R by

$$\nabla^2 f(R) := \left(\frac{\partial^2 f}{\partial R_{k_1, k'_1} \partial R_{k_2, k'_2}}(R) \right)_{(k_1, k'_1), (k_2, k'_2) \in \{1, \dots, \kappa\}^2}.$$

Thinking of $\nabla^2 f(R)$ as a (self-adjoint) linear operator $\mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}^{\kappa \times \kappa}$, we can speak of its minimum eigenvalue

$$\lambda_{\min}(\nabla^2 f(R)) := \min_{\|Q\|_2=1} \langle \nabla^2 f(R)Q, Q \rangle,$$

and its spectral radius

$$\|\nabla^2 f(R)\|_{\text{op}} := \max_{\|Q\|_2=1} |\langle \nabla^2 f(R)Q, Q \rangle|.$$

Now define the following constant for each $\theta \in \Theta$:

$$C_\theta := \sup_{\|R\|_1 \leq 1} \{ \|\nabla \xi_\theta(R)\|_\infty \vee \|\nabla^2 \xi_\theta(R)\|_{\text{op}} \}, \quad (2.18)$$

which is finite by Remark 1.5. Convexity assumption (A2) means $\lambda_{\min}(\nabla^2 \xi(R)) \geq 0$ for all $R \in \mathcal{R}_K$. The following lemma shows how this condition can be maintained after modification (2.17) provided we assume a strict inequality for the original covariance function ξ .

Lemma 2.8. *Assume ξ satisfies (A1) and*

$$\inf_{R \in \mathcal{R}_K} \lambda_{\min}(\nabla^2 \xi(R)) > 0. \quad (\text{A2}')$$

Assume β satisfies (B1) and

$$\sum_{\theta \in \Theta_Q} \beta_\theta^2 C_\theta < \inf_{R \in \mathcal{R}_K} \lambda_{\min}(\nabla^2 \xi(R)). \quad (\text{B2})$$

Then ξ_β defined in (2.17) satisfies (A1) and (A2).

Proof. That ξ_β satisfies (A1) is immediate from (B1). For (A2), we simply observe that for any $R \in \mathcal{R}_K$,

$$\lambda_{\min}(\nabla^2 \xi_\beta(R)) \geq \lambda_{\min}(\nabla^2 \xi(R)) - \sum_{\theta \in \Theta_Q} \beta_\theta^2 \|\nabla^2 \xi_\theta(R)\|_{\text{op}} \stackrel{(\text{B2})}{\geq} 0.$$

Hence ξ_β is convex on the set \mathcal{R}_K . □

The following result is the reason for introducing the modified model (2.17). It is proved in Sect. 4.

Proposition 2.9. *Assume ξ satisfies (A1), (A2'), and β satisfies (B1), (B2). Then for any $d \in \mathcal{D}$ and $\theta \in \Theta_Q$, when only the value of β_θ is varied, the function $\beta_\theta \mapsto \inf_{\pi \in \Pi_d} \mathcal{P}_{\xi_\beta}(\pi)$ is differentiable on an open interval.*

To see the utility of Proposition 2.9, let us consider a Hamiltonian corresponding to the modified covariance function ξ_β from (2.17). Namely,

$$H_{N, \xi_\beta}(\sigma) = H_{N, \xi}(\sigma) + \sum_{\theta \in \Theta_Q} \beta_\theta H_{N, \theta}(\sigma), \quad (2.19)$$

where $H_{N, \xi}$ is as in (1.14), and each $H_{N, \theta}$ is an independent Gaussian process satisfying (2.16). It is well-known [5, 15, 51] that the differentiability of the Parisi formula guarantees the concentration of $H_{N, \theta}(\sigma)$ under the Gibbs measure. This concentration in turn leads to the G.G. identities in Definition 2.2, provided that $\beta_\theta \neq 0$ for each $\theta \in \Theta_Q$. In this way we will be able to obtain synchronization from Theorem 2.3 without any perturbation. The tradeoff is that we need the model to have a generic covariance function ξ_β , instead of the original function ξ whose symmetry (A3) was meant to enable the crucial Lemma 2.6. But as the next section explains, this symmetry can be preserved by selecting β in a special way.

2.3. Parisi formula for symmetric generic models. Section 2.1 outlined how to obtain the desired lower bound (2.1) under two assumptions:

- (i) in the large N -large limit, the array of overlap matrices satisfies the G.G. identities;
- (ii) each matrix in that array is invariant (in distribution) under permutation of its rows and columns.

Section 2.2 described how to secure assumption (i) for the generic model (2.17) with parameters $\beta = (\beta_\theta)_{\theta \in \Theta_{\mathbb{Q}}}$ satisfying $\beta_\theta \neq 0$ for every θ . Now we address assumption (ii) for the generic model.

Given a permutation $\omega \in S_\kappa$ and a vector $w = (w(k))_{k=1}^\kappa \in \mathbb{R}^\kappa$, define the permuted vector $\omega \bullet w := (w(\omega^{-1}(k)))_{k=1}^\kappa$. We now extend this action to the parameter space Θ :

$$\begin{aligned} & \text{for } \theta = (p, m, n_1, \dots, n_m, w_1, \dots, w_m) \in \Theta, \\ & \text{define } \omega \bullet \theta = (p, m, n_1, \dots, n_m, \omega \bullet w_1, \dots, \omega \bullet w_m) \in \Theta. \end{aligned} \quad (2.20)$$

We are then interested in β satisfying the following symmetry condition:

$$\beta_\theta = \beta_{\omega \bullet \theta} \quad \text{for every } \theta \in \Theta \text{ and } \omega \in S_\kappa. \quad (\text{B3})$$

Lemma 2.10. *If ξ satisfies (A3) and β satisfies (B3), then ξ_β satisfies (A3).*

Proof. For any given $\theta = (p, m, n_1, \dots, n_m, w_1, \dots, w_m)$, we have

$$\begin{aligned} \xi_\theta(\omega \bullet R) & \stackrel{(1.17), (1.23)}{=} \prod_{j=1}^m \left(\sum_{k, k'=1}^\kappa R_{\omega^{-1}(k), \omega^{-1}(k')}^p w_j(k) w_j(k') \right)^{n_j} \\ & = \prod_{j=1}^m \left(\sum_{k, k'=1}^\kappa R_{k, k'}^p w_j(\omega(k)) w_j(\omega(k')) \right)^{n_j} \stackrel{(1.17), (2.20)}{=} \xi_{\omega^{-1} \bullet \theta}(R). \end{aligned}$$

The assumption (B3) means $\beta_\theta = \beta_{\omega^{-1} \bullet \theta}$ no matter the choice of ω , hence

$$\sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_\theta^2 \xi_\theta(\omega \bullet R) = \sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_{\omega^{-1} \bullet \theta}^2 \xi_{\omega^{-1} \bullet \theta}(R) = \sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_\theta^2 \xi_\theta(R).$$

Since $\xi(\omega \bullet R) = \xi(R)$ by assumption, we have now shown that the same is true for the sum $\xi_\beta = \xi + \sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_\theta^2 \xi_\theta$. \square

By implementing the strategy discussed after Proposition 2.9, we will prove the following result in Sect. 6. Note that $\mathcal{P}_{\xi_\beta}(\pi, 0)$ is the Parisi functional (1.34) without the Lagrange multiplier λ (see Remark 1.9 for further discussion).

Proposition 2.11. *Assume ξ satisfies (A1), (A2'), (A3), and β satisfies (B1), (B2), (B3). Provided $\beta_\theta \neq 0$ for every $\theta \in \Theta_{\mathbb{Q}}$, there exists $\pi \in \Pi^*$ such that*

$$\liminf_{N \rightarrow \infty} F_{\kappa N, \xi_\beta}(d_{\text{bal}}) \geq \mathcal{P}_{\xi_\beta}(\pi, 0). \quad (2.21)$$

From here the end goal (2.1) is not too much further. There are just two missing items:

1. Construct β that satisfies the hypotheses of Proposition 2.11.

2. Justify sending $\beta \rightarrow 0$.

Item 1. is taken care of by the following lemma. Recall the constant C_θ defined in (2.18).

Lemma 2.12. *For any $\varepsilon > 0$, there exists $\beta = (\beta_\theta)_{\theta \in \Theta_\mathbb{Q}}$ satisfying (B1), (B3), $\beta_\theta \neq 0$ for every $\theta \in \Theta_\mathbb{Q}$, and*

$$\sum_{\theta \in \Theta_\mathbb{Q}} \beta_\theta^2 (C_\theta \vee 1) \leq \varepsilon. \quad (2.22)$$

Proof. The operation $(\omega, \theta) \mapsto \omega \bullet \theta$ defined in (2.20) is a group action by the symmetric group S_κ on $\Theta_\mathbb{Q}$. Since $\Theta_\mathbb{Q}$ is countable, we can enumerate the orbits of $\Theta_\mathbb{Q}$ under this action: $\mathcal{O}_1, \mathcal{O}_2$, and so on. Both the constant C_θ in (2.18) and the degree in (1.18) are invariant under the action, and so we can write $C_i = C_\theta$ and $\deg(\mathcal{O}_i) = \deg(\theta)$ for any $\theta \in \mathcal{O}_i$. Given $\varepsilon > 0$, define

$$\beta_i := \sqrt{\frac{1}{2^i (1 + \varepsilon)^{\deg(\mathcal{O}_i)}}} \wedge \sqrt{\frac{\varepsilon}{2^i k! (C_i \vee 1)}} > 0, \quad (2.23)$$

and then set $\beta_\theta = \beta_i$ for all $\theta \in \mathcal{O}_i$. The collection $\beta = (\beta_\theta)_{\theta \in \Theta_\mathbb{Q}}$ satisfies (B3) by construction. Inequality (B1) holds because of the first term on the right-hand side of (2.23), while (2.22) holds because of the second term and the fact $|\mathcal{O}_i| \leq \kappa!$. \square

Meanwhile, item 2. will be possible thanks to the following proposition, which shows that both sides of (2.21) are continuous with respect to the covariance function ξ .

Proposition 2.13. *Assume ξ and $\tilde{\xi}$ satisfy (A1). Then the following statements hold.*

(a) *For any $d \in \mathcal{D}_N$, we have*

$$|\mathbf{F}_{N,\xi}(d) - \mathbf{F}_{N,\tilde{\xi}}(d)| \leq \sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)|.$$

(b) *For any $d \in \mathcal{D}$ and any $\pi, \tilde{\pi} \in \Pi_d$, we have*

$$\begin{aligned} |\mathcal{P}_\xi(\pi) - \mathcal{P}_{\tilde{\xi}}(\tilde{\pi})| &\leq 2 \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \tilde{\xi}(R)\|_\infty \\ &\quad + \sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty \int_0^1 \|\pi(t) - \tilde{\pi}(t)\|_1 dt. \end{aligned} \quad (2.24)$$

Proposition 2.13 is proved in Sect. 5.1.

2.4. Proofs of main results. Finally, we prove our main theorems.

Proof of Theorem 1.8. Consider any $d^N \in \mathcal{D}_N$ such that $d^N \rightarrow d_{\text{bal}}$ as $N \rightarrow \infty$. We wish to establish (1.24). Since Π^\star is a subset of $\Pi_{d_{\text{bal}}}$, Theorem 1.7 provides an upper bound:

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbf{F}_{N,\xi}(d^N) &= \lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} \mathbf{F}_{N,\xi}(d_{\text{bal}}, \varepsilon) = \lim_{\varepsilon \searrow 0} \liminf_{N \rightarrow \infty} \mathbf{F}_{N,\xi}(d_{\text{bal}}, \varepsilon) \\ &= \inf_{\pi \in \Pi_{d_{\text{bal}}}} \mathcal{P}_\xi(\pi) \\ &\leq \inf_{\pi \in \Pi^\star} \mathcal{P}_\xi(\pi). \end{aligned} \quad (2.25)$$

Since the first line of (2.25) is entirely equalities, the matching lower bound only needs to be established along a subsequence. More specifically, we will prove (2.1).

The covariance function ξ is assumed to satisfy weak convexity (A2). To upgrade to strong convexity, we add a small multiple of the usual Potts Hamiltonian: for $\varepsilon > 0$, define

$$\xi^\varepsilon(R) := \xi(R) + \varepsilon \cdot \text{tr}(R^\top R)/2. \quad (2.26)$$

Let us note three properties of ξ^ε . First, by Remark 1.6, ξ^ε satisfies (A1) provided ξ satisfies (A1). Second, by assumption (A2) for ξ , the new function ξ^ε satisfies (A2') with

$$\lambda_{\min}(\nabla^2 \xi^\varepsilon(R)) \geq \varepsilon \quad \text{for every } R \in \mathcal{R}_K. \quad (2.27)$$

Third and finally, since ξ is assumed to satisfy the symmetry condition (A3) and the map $R \mapsto \text{tr}(R^\top R)$ does as well (the expression in (1.15) is clearly invariant under any permutation of the row and column indices), the function ξ^ε satisfies (A3). Therefore, Lemma 2.12 guarantees the existence of $\beta = (\beta_\theta)_{\theta \in \Theta_{\mathbb{Q}}}$ such that (B1) and (B3) hold, $\beta_\theta \neq 0$ for every $\theta \in \Theta_{\mathbb{Q}}$, and

$$\sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_\theta^2 (C_\theta \vee 1) \leq \varepsilon. \quad (2.28)$$

Applying the modification (2.17) to ξ^ε in place of ξ , we obtain $\xi_\beta^\varepsilon = \xi^\varepsilon + \sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_\theta^2 \xi_\theta$. The combination of (2.27) and (2.28) shows that (B2) is satisfied with ξ^ε in place of ξ . We can thus apply Proposition 2.11 to determine that

$$\liminf_{N \rightarrow \infty} F_{\kappa N, \xi_\beta^\varepsilon}(d_{\text{bal}}) \geq \mathcal{P}_{\xi_\beta^\varepsilon}(\pi, 0) \stackrel{(1.35)}{\geq} \inf_{\pi \in \Pi^*} \mathcal{P}_{\xi_\beta^\varepsilon}(\pi). \quad (2.29)$$

The remainder of the proof is to justify sending ε and β to zero.

We first address the left-hand side of (2.29). By Proposition 2.13(a), we have

$$|F_{\kappa N, \xi}(d_{\text{bal}}) - F_{\kappa N, \xi_\beta^\varepsilon}(d_{\text{bal}})| \leq \sup_{\|R\|_1 \leq 1} |\xi(R) - \xi_\beta^\varepsilon(R)|.$$

Whenever $\|R\|_1 \leq 1$, we have

$$\begin{aligned} |\xi(R) - \xi_\beta^\varepsilon(R)| &\leq |\xi(R) - \xi^\varepsilon(R)| + |\xi^\varepsilon(R) - \xi_\beta^\varepsilon(R)| \\ &\stackrel{(2.26), (2.17), (1.20)}{\leq} \frac{\varepsilon}{2} + \sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_\theta^2 \stackrel{(2.28)}{\leq} \frac{3\varepsilon}{2}. \end{aligned}$$

The two previous displays together show

$$|F_{\kappa N, \xi}(d_{\text{bal}}) - F_{\kappa N, \xi_\beta^\varepsilon}(d_{\text{bal}})| \leq 3\varepsilon/2. \quad (2.30)$$

Meanwhile, regarding the right-hand side of (2.29), Proposition 2.13(b) gives

$$\left| \inf_{\pi \in \Pi^*} \mathcal{P}_\xi(\pi) - \inf_{\pi \in \Pi^*} \mathcal{P}_{\xi_\beta^\varepsilon}(\pi) \right| \leq 2 \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \xi_\beta^\varepsilon(R)\|_\infty.$$

Whenever $\|R\|_1 \leq 1$, another application of the triangle inequality leads to

$$\begin{aligned} \|\nabla \xi(R) - \nabla \xi_{\beta}^{\varepsilon}(R)\|_{\infty} &\leq \|\nabla \xi(R) - \nabla \xi^{\varepsilon}(R)\|_{\infty} + \|\nabla \xi^{\varepsilon}(R) - \nabla \xi_{\beta}^{\varepsilon}(R)\|_{\infty} \\ &\stackrel{(2.26),(2.18)}{\leq} \varepsilon + \sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_{\theta}^2 C_{\theta} \stackrel{(2.28)}{\leq} 2\varepsilon. \end{aligned}$$

Putting the two previous displays together, we have

$$\left| \inf_{\pi \in \Pi^*} \mathcal{P}_{\xi}(\pi) - \inf_{\pi \in \Pi^*} \mathcal{P}_{\xi_{\beta}^{\varepsilon}}(\pi) \right| \leq 4\varepsilon. \quad (2.31)$$

Finally, the combination of (2.29), (2.30), and (2.31) yields

$$\liminf_{N \rightarrow \infty} \mathbf{F}_{\kappa N, \xi}(d_{\text{bal}}) \geq \inf_{\pi \in \Pi^*} \mathcal{P}_{\xi}(\pi) - 6\varepsilon.$$

Letting $\varepsilon \searrow 0$ completes the proof. \square

Proof of Theorem 1.3. This is immediate from Theorem 1.8 since the associated covariance function $\xi(R) = \beta^2 \text{tr}(R^{\top} R)$ from (1.15) satisfies (A1), (A2), (A3). \square

3. Parisi Functional Preliminaries

3.1. Prelimit of the Parisi functional. In this section we define the functional Ψ_{ξ} that appears in (2.5). More precisely, there is a family of functionals $(\Psi_{\xi, N})_{N \geq 1}$ that arise naturally in the A.S.S. scheme (Proposition 6.2). These can be thought of as prelimiting versions of the Parisi functional \mathcal{P}_{ξ} . Understanding their behavior as $N \rightarrow \infty$ is an important step in proving that the Parisi formula is a lower bound for the limiting free energy.

Our definition requires three steps:

1. Give an abstract notion of an overlap map R which generalizes (2.2).
2. Using this map and a random measure from which to sample, generate a random array of overlap matrices.
3. Define a functional $\mathcal{L} \mapsto \Psi_{\xi, N}(\mathcal{L})$, where \mathcal{L} is the law of the random array.

3.1.1. Allowable overlap maps Recall from (1.19) that \mathcal{R}_{κ} denotes the set of $\kappa \times \kappa$ matrices whose entries are nonnegative and have sum at most 1. Fix $d \in \mathcal{D}$ and a positive integer N .

Assumption 3.1. $(\mathcal{X}, \mathcal{F})$ is a measurable space, and $R: \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{R}_{\kappa}$ is a map with the following properties.

- (i) R is measurable.
- (ii) For every $x \in \mathcal{X}$, $R(x, x) = \text{diag}(d)$.
- (iii) There exist centered Gaussian processes $Z: \mathcal{X} \rightarrow \mathbb{R}^{\kappa}$ and $Y: \mathcal{X} \rightarrow \mathbb{R}$ with covariances

$$\begin{aligned} \mathbb{E}[Z(x)Z(x')^{\top}] &= \nabla \xi(R(x, x')) \\ \mathbb{E}[Y(x)Y(x')] &= \vartheta_{\xi}(R(x, x')) \end{aligned} \quad \text{for } x, x' \in \mathcal{X}. \quad (3.1)$$

Furthermore, these processes are almost surely measurable functions on \mathcal{X} .

3.1.2. The overlap array Given a random probability measure \mathcal{G} on $(\mathcal{X}, \mathcal{F})$ which is independent of the processes from (3.1), let $(x^\ell)_{\ell \geq 1}$ be i.i.d. samples from \mathcal{G} . Apply R to each pair of samples, and set

$$\mathcal{R}_{\ell, \ell'} = R(x^\ell, x^{\ell'}) + \mathbb{1}_{\{\ell = \ell'\}}(\text{diag}(d) - R(x^\ell, x^{\ell'})).$$

This defines a random array $\mathcal{R} = (\mathcal{R}_{\ell, \ell'})_{\ell, \ell' \geq 1}$. Denote the law of \mathcal{R} under $\mathbb{E}(\mathcal{G}^{\otimes \infty})$ by $\text{Law}(\mathcal{R}; \mathbb{E}(\mathcal{G}^{\otimes \infty}))$, where the dependence on the map R is implicit.

3.1.3. The functional Given a nonempty subset $S \subseteq \Sigma^N$, we now define a functional $\mathcal{L} \mapsto \Psi_{N, \xi}(\mathcal{L}; S)$ that accepts as input any law $\mathcal{L} = \text{Law}(\mathcal{R}; \mathbb{E}(\mathcal{G}^{\otimes \infty}))$ realized as above.

Let Z_1, \dots, Z_N be independent copies of the process Z from (3.1). Let Y be as in (3.1). Let $\langle \cdot \rangle_{\mathcal{G}}$ denote expectation with respect to \mathcal{G} . Finally, let $\mathbb{E}(\cdot)$ denote expectation over both realizations of \mathcal{G} and the Gaussian processes from Assumption 3.1. Now define

$$\Psi_{N, \xi}^{(1)}(\mathcal{L}; d, S) := \frac{1}{N} \mathbb{E} \log \sum_{\sigma \in S} \left\langle \exp \left(\sum_{i=1}^N \langle Z_i(x), \sigma_i \rangle \right) \right\rangle_{\mathcal{G}}, \quad (3.2a)$$

$$\Psi_{N, \xi}^{(2)}(\mathcal{L}; d) := \frac{1}{N} \mathbb{E} \log \left\langle \exp \left(\sqrt{N} Y(x) \right) \right\rangle_{\mathcal{G}}, \quad (3.2b)$$

and then the functional of interest is given by

$$\Psi_{N, \xi}(\mathcal{L}; S) := \Psi_{N, \xi}^{(1)}(\mathcal{L}; S) - \Psi_{N, \xi}^{(2)}(\mathcal{L}). \quad (3.3)$$

Especially important is the fact that $\mathcal{L} \mapsto \Psi_{N, \xi}(\mathcal{L}; S)$ is continuous. To state this precisely, we need to introduce some notation. Given any law $\mathcal{L} = \text{Law}(\mathcal{R}; \mathcal{G})$ realized as above, let $\mathcal{L}^{(n)}$ denote the law of the finite subarray $\mathcal{R}^{(n)} = (\mathcal{R}_{\ell, \ell'})_{1 \leq \ell, \ell' \leq n}$; this is a Borel probability measure on $n \times n$ arrays whose elements belong to \mathcal{R}_κ . Let $\text{Prob}((\mathcal{R}_\kappa)^{n \times n})$ denote the set of all Borel probability measures on this space. Because \mathcal{R}_κ is compact, one can metrize the topology of weak convergence on $\text{Prob}((\mathcal{R}_\kappa)^{n \times n})$ by, say, a Wasserstein distance with respect to the norm $\|\mathcal{R}^{(n)}\| := \sum_{\ell, \ell'=1}^n \|\mathcal{R}_{\ell, \ell'}\|_1$. We can thus speak of continuity with respect to weak convergence.

Proposition 3.2. *Assume ξ satisfies (A1). For any $\varepsilon > 0$, there is n large enough and some continuous function $\phi_N : \text{Prob}((\mathcal{R}_\kappa)^{n \times n}) \rightarrow \mathbb{R}$ such that*

$$|\Psi_{N, \xi}(\mathcal{L}; S) - \phi_N(\mathcal{L}^{(n)})| \leq \varepsilon \quad \text{whenever } \Psi_{N, \xi}(\mathcal{L}; S) \text{ is defined.}$$

It follows from this proposition that $\mathcal{L} \mapsto \Psi_{N, \xi}(\mathcal{L}; S)$ is continuous. In the statement below, weak convergence is understood in the product sense: a sequence $(\mathcal{L}_m)_{m \geq 1}$ converges weakly to \mathcal{L} if for every n , $(\mathcal{L}_m^{(n)})_{m \geq 1}$ converges weakly to $\mathcal{L}^{(n)}$.

Corollary 3.3. *Assume ξ satisfies (A1). If $(\mathcal{L}_j)_{j \geq 1}$ is a weakly convergent sequence of laws such that $\Psi_{N, \xi}(\mathcal{L}_j; S)$ is defined for every j , then $\lim_{j \rightarrow \infty} \Psi_{N, \xi}(\mathcal{L}_j; S)$ exists and depends only on the limit of $(\mathcal{L}_j)_{j \geq 1}$.*

A proof of Corollary 3.3 can be found in [10, Cor. 2.7]. Proposition 3.2 follows from a standard truncation argument that relies only on the compactness of \mathcal{R}_κ and the continuity of $\nabla \xi$ and ϑ . Examples of this argument can be found in [55, Lem. 3], [54, Thm. 1.3], and [10, Prop. 2.6]. Therefore, we omit the proof.

3.2. Ruelle probability cascades. In this section we restrict the functional $\Psi_{N,\xi}$ from (3.3) to overlap arrays generated by a Ruelle probability cascade (RPC). The RPCs are a certain class of random probability measures with ultrametric support. Each RPC is associated to a partition of the unit interval:

$$0 = m_0 < m_1 < \cdots < m_{s-1} < m_s = 1. \quad (3.4)$$

For concreteness, we take the support of the RPC to be the countable set \mathbb{N}^{s-1} , which can be regarded as the leaf set of a tree of depth $s - 1$:

- The root vertex is the empty sequence \emptyset , and by convention $\mathbb{N}^0 = \{\emptyset\}$.
- Each $\alpha = (\alpha_1, \dots, \alpha_r) \in \mathbb{N}^r$ is said to have depth $|\alpha| = r$ and has children $\{(\alpha_1, \dots, \alpha_r, a) : a \in \mathbb{N}\}$ at depth $r + 1$.
- The path to a given leaf $\alpha = (\alpha_1, \dots, \alpha_{s-1}) \in \mathbb{N}^{s-1}$ is the set

$$p(\alpha) = \{\emptyset, \alpha_1, (\alpha_1, \alpha_2), \dots, (\alpha_1, \dots, \alpha_{s-1})\}.$$

Given two leaves $\alpha = (\alpha_1, \dots, \alpha_{s-1})$ and $\alpha' = (\alpha'_1, \dots, \alpha'_{s-1})$ in \mathbb{N}^{s-1} , let $r(\alpha, \alpha')$ denote the smallest depth at which their ancestral paths disagree:

$$r(\alpha, \alpha') := \begin{cases} \inf\{r : \alpha_r \neq \alpha'_r\} & \text{if } \alpha \neq \alpha' \\ s & \text{if } \alpha = \alpha'. \end{cases}$$

The RPCs are characterized by the following fact.

Theorem 3.4. [69, Thm. 15.2.1] *For any sequence of the form (3.4), there is a random measure ν on \mathbb{N}^{s-1} with the following property. Given any sequence $0 \leq q_1 \leq \cdots \leq q_s = 1$, define the measure $\mu = \sum_{r=1}^s (m_r - m_{r-1}) \delta_{q_r}$. If $(\alpha^1, \alpha^2, \dots) \sim \mathbb{E}(\nu^{\otimes \infty})$, then the law of the array $(q_{r(\alpha^\ell, \alpha^{\ell'})})_{\ell, \ell' \geq 1}$ is equal to $\tilde{\mathcal{L}}_\mu$ from Theorem 2.4.*

The measure ν is called the RPC associated to (3.4), and we denote its probability mass function by $(\nu_\alpha)_{\alpha \in \mathbb{N}^{s-1}}$. We will now use RPCs to give a different formulation of the Parisi functional from Sect. 1.4.

Let ξ satisfy (A1). Consider any $d \in \mathcal{D}$ and any discrete path $\pi \in \Pi_d^{\text{disc}}$, meaning there is a partition of the form (3.4) and a sequence in $\Gamma_\kappa(d)$,

$$0 \leq \gamma_1 \leq \cdots \leq \gamma_{s-1} \leq \gamma_s = \text{diag}(d), \quad (3.5a)$$

such that

$$\pi(t) = \gamma_r \quad \text{for } t \in (m_{r-1}, m_r], r \in \{1, \dots, s\}. \quad (3.5b)$$

Let $Z_1, \dots, Z_N : \mathbb{N}^{s-1} \rightarrow \mathbb{R}^\kappa$ and $Y : \mathbb{N}^{s-1} \rightarrow \mathbb{R}$ be independent centered Gaussian processes with covariances

$$\begin{aligned} \mathbb{E}[Z_i(\alpha)Z_i(\alpha')^\top] &= \nabla \xi(\gamma_{r(\alpha, \alpha')}) & \text{for } \alpha, \alpha' \in \mathbb{N}^{s-1}. \\ \mathbb{E}[Y(\alpha)Y(\alpha')] &= \vartheta_\xi(\gamma_{r(\alpha, \alpha')}) \end{aligned} \quad (3.6)$$

We assume these processes are also independent of the weights $(\nu_\alpha)_{\alpha \in \mathbb{N}^{s-1}}$; their existence is addressed in Remark 3.9. Now define, for $\lambda \in \mathbb{R}^\kappa$ and nonempty $S \subseteq \Sigma^N$, the quantity

$$\mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; S) := \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in S} \nu_\alpha \exp \left(\sum_{i=1}^N \langle Z_i(\alpha) + \lambda, \sigma_i \rangle \right). \quad (3.7)$$

Also recall the following quantity from (1.33):

$$\mathcal{P}_\xi^{(2)}(\pi) = \frac{1}{2} \int_0^1 \vartheta_\xi(\pi(t)) \, dt - \frac{1}{2} \vartheta_\xi(\text{diag}(d)), \quad \pi \in \Pi_d. \quad (3.8)$$

Apart from the Lagrange multiplier $\lambda \in \mathbb{R}^\kappa$ in (3.7), the functionals $\mathcal{P}_{N,\xi}^{(1)}$ and $\mathcal{P}_{N,\xi}^{(2)}$ are simply $\Psi_{N,\xi}^{(1)}$ and $\Psi_{N,\xi}^{(2)}$ from (3.2) applied to the RPC setting. This is captured by the following lemma, which is proved in “Appendix C”.

Lemma 3.5. *Assume ξ satisfies (A1). Let ν be the RPC associated to (3.4). With $\pi \in \Pi_d^{\text{disc}}$ given by (3.5), let \mathcal{L} denote the law of the array $(\gamma_{r(\alpha^\ell, \alpha^{\ell'})})_{\ell, \ell' \geq 1}$ under $\mathbb{E}(\nu^{\otimes \infty})$. Then*

$$\mathcal{P}_{N,\xi}^{(1)}(\pi, 0; S) = \Psi_{N,\xi}^{(1)}(\mathcal{L}; S), \quad (3.9a)$$

$$\mathcal{P}_\xi^{(2)}(\pi) = -\Psi_{N,\xi}^{(2)}(\mathcal{L}). \quad (3.9b)$$

In particular, $\mathcal{P}_{N,\xi}^{(1)}(\pi, 0; S) + \mathcal{P}_\xi^{(2)}(\pi) = \Psi_{N,\xi}(\mathcal{L}; S)$.

Remark 3.6. Because (3.5a) allows for equalities, the representation (3.5b) of the path π is not unique. For instance, one could insert duplicate copies of some γ_r and then refine the partition (3.4). Consequently, it is not immediately clear that $\mathcal{P}_{N,\xi}^{(1)}$ is well-defined, but this actually follows from Lemma 3.5. Indeed, no matter the representation of π , the following measure remains the same:

$$\mu = \sum_{r=1}^s (m_r - m_{r-1}) \delta_{q_r}, \quad \text{where } q_r = \sup\{t \in (0, 1] : \pi(t) = \gamma_r\}.$$

Under this definition of q_r , we always have $\pi(q_r) = \gamma_r$ by left-continuity of π , and so the array $(\gamma_{r(\alpha^\ell, \alpha^{\ell'})})_{\ell, \ell' \geq 1}$ in Lemma 3.5 is equal to $(\pi(q_{r(\alpha^\ell, \alpha^{\ell'})}))_{\ell, \ell' \geq 1}$. Since the law of $(q_{r(\alpha^\ell, \alpha^{\ell'})})_{\ell, \ell' \geq 1}$ is equal to $\bar{\mathcal{L}}_\mu$ by Theorem 3.4, it follows that the law \mathcal{L} in Lemma 3.5 does not depend on the representation of π .

Notice that the left-hand side of (3.9b) does not depend on N . Our next lemma shows the same is true of (3.9a) when S is the entire product set Σ^N . More importantly, the functional $\mathcal{P}_\xi^{(1)}$ from (1.31) is recovered. Once again, the proof is postponed to “Appendix C”.

Lemma 3.7. *Assume ξ satisfies (A1). For any $N \geq 1$, $\pi \in \Pi_d^{\text{disc}}$, and $\lambda \in \mathbb{R}^\kappa$, we have*

$$\mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N) = \mathcal{P}_\xi^{(1)}(\pi, \lambda). \quad (3.10)$$

Remark 3.8. It would appear from Lemmas 3.5 and 3.7 that the functional $\Psi_{N,\xi}$ from (3.3) is easily related to \mathcal{P}_ξ from (1.34). The complication is that the A.S.S. scheme presented in Sect. 6.1 will require we evaluate $\mathcal{L} \mapsto \Psi_{N,\xi}(\mathcal{L}; S)$ using the constrained set $S = \Sigma^N(d^N)$ for some $d^N \in \mathcal{D}_N$. In contrast, the identity (3.10) only holds in unconstrained case $S = \Sigma^N$. This gap can only be bridged once $N \rightarrow \infty$, for then the constraint d^N can be replaced by an optimization over the dual variable λ ; see Lemma 5.2. The tradeoff is that an additional optimization must be performed over λ , as in (1.35).

Remark 3.9. In order for the quantities in (3.9) to be defined, we need the existence of processes Z_1, \dots, Z_N, Y satisfying (3.6). Fortunately, it is easy to construct them. Indeed, let $(\eta_{i,\alpha} : i \geq 1, \alpha \in \mathbb{N}^0 \cup \dots \cup \mathbb{N}^{s-1})$ be i.i.d. standard normal random vectors in \mathbb{R}^κ , and then define

$$Z_i(\alpha) = \sum_{r=0}^{s-1} \sqrt{\nabla \xi(\gamma_{r+1}) - \nabla \xi(\gamma_r) \mathbb{1}\{r > 0\}} \eta_{i,(\alpha_1, \dots, \alpha_r)}. \quad (3.11)$$

Here we are using (2.15) to ensure that the matrix inside the square root is positive-semidefinite and thus admits a square root. Since $(\alpha_1, \dots, \alpha_r) = (\alpha'_1, \dots, \alpha'_r)$ if and only if $r < r(\alpha, \alpha')$, we have

$$\mathbb{E}[Z_i(\alpha)Z_i(\alpha')^\top] = \sum_{r=0}^{r(\alpha, \alpha')-1} [\nabla \xi(\gamma_{r+1}) - \nabla \xi(\gamma_r) \mathbb{1}\{r > 0\}] = \nabla \xi(\gamma_{r(\alpha, \alpha')}).$$

Similarly, let $(\eta_\alpha : \alpha \in \mathbb{N}^0 \cup \dots \cup \mathbb{N}^{s-1})$ be i.i.d. standard normal random variables, and then

$$Y(\alpha) = \sum_{r=0}^{s-1} \sqrt{\vartheta_\xi(\gamma_{r+1}) - \vartheta_\xi(\gamma_r)} \eta_{(\alpha_1, \dots, \alpha_r)}. \quad (3.12)$$

The same kind of telescoping calculation as above yields $\mathbb{E}[Y(\alpha)Y(\alpha')] = \vartheta_\xi(\gamma_{r(\alpha, \alpha')})$.

4. Differentiability of the Parisi Formula

In this section, we prove Proposition 2.9. The following lemma provides the key estimate. Recall the functional \mathcal{P}_ξ from (1.34).

Lemma 4.1. *Assume ξ satisfies (A1) and $\beta = (\beta_\theta)_{\theta \in \Theta_\mathbb{Q}}$ satisfies (B1). For any $\theta \in \Theta_\mathbb{Q}$, there exists a constant C depending only on θ and κ , such that*

$$\left| \left(\frac{\partial}{\partial \beta_\theta} \right)^2 \mathcal{P}_{\xi_\beta}(\pi, \lambda) \right| \leq C(1 + \beta_\theta^2) \text{ for any } \pi \in \Pi_d^{\text{disc}}, d \in \mathcal{D}, \lambda \in \mathbb{R}^\kappa.$$

Proof. During the proof, the value of C may change from line to line, but it will only depend on θ and κ . Assume π is given by (1.29). Since $\mathcal{P}_{\xi_\beta}(\pi, \lambda) = \mathcal{P}_{\xi_\beta}^{(1)}(\pi, \lambda) + \mathcal{P}_{\xi_\beta}^{(2)}(\pi)$, it suffices to prove the desired differential inequality for each term on the right-hand side. We begin with the second term.

Recalling the definition of $\mathcal{P}_{\xi_\beta}^{(2)}$ from (1.33), we have

$$\mathcal{P}_{\xi_\beta}^{(2)}(\pi) = \frac{1}{2} \sum_{r=1}^s (m_r - m_{r-1}) \vartheta_{\xi_\beta}(\gamma_r) - \frac{1}{2} \vartheta_{\xi_\beta}(\text{diag}(d)).$$

The right-hand side can be rewritten using summation by parts. Since $m_0 = 0, m_s = 1$, and $\gamma_s = \text{diag}(d)$, this results in

$$\mathcal{P}_{\xi_\beta}^{(2)}(\pi) = -\frac{1}{2} \sum_{r=2}^s m_{r-1} [\vartheta_{\xi_\beta}(\gamma_r) - \vartheta_{\xi_\beta}(\gamma_{r-1})].$$

Now recall from (1.27) that $\vartheta_{\xi_\beta}(R) = \langle R, \nabla \xi_\beta(R) \rangle - \xi_\beta(R)$, where $\xi_\beta = \xi + \sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_\theta^2 \xi_\theta$ as in (2.17). We thus have

$$\left(\frac{\partial}{\partial \beta_\theta} \right)^2 \vartheta_{\xi_\beta}(R) = 2 \langle R, \nabla \xi_\theta(R) \rangle - 2 \xi_\theta(R) \stackrel{(2.14)}{=} 2 \vartheta_\theta(R) \stackrel{\text{Prop. 2.7(a)}}{=} 2(\deg(\theta) - 1) \xi_\theta(R).$$

Recall from Proposition 2.7(b) that $\xi_\theta(\gamma_r) \geq \xi_\theta(\gamma_{r-1}) \geq 0$ for each r . Therefore, the two previous displays together yield

$$\begin{aligned} \left| \left(\frac{\partial}{\partial \beta_\theta} \right)^2 \mathcal{P}_{\xi_\beta}^{(2)}(\pi) \right| &\leq (\deg(\theta) - 1) \sum_{r=2}^s [\xi_\theta(\gamma_r) - \xi_\theta(\gamma_{r-1})] \\ &\leq (\deg(\theta) - 1) \xi_\theta(\gamma_s) \leq (\deg(\theta) - 1) \sup_{\|R\|_1 \leq 1} |\xi_\theta(R)| \leq C. \end{aligned} \quad (4.1)$$

Next we consider the quantity $\mathcal{P}_{\xi_\beta}^{(1)}(\pi, \lambda)$ from (1.31). By Lemma 3.7 with $N = 1$,

$$\mathcal{P}_{\xi_\beta}^{(1)}(\pi, \lambda) = \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in \Sigma} \nu_\alpha \exp \langle Z_{\xi_\beta}(\alpha) + \lambda, \sigma \rangle, \quad (4.2)$$

where $(\nu_\alpha)_{\alpha \in \mathbb{N}^{s-1}}$ is the probability mass function of the Ruelle probability cascade associated to (3.4), and $Z_{\xi_\beta} : \mathbb{N}^{s-1} \rightarrow \mathbb{R}^K$ is an independent centered Gaussian process with covariance

$$\mathbb{E}[Z_{\xi_\beta}(\alpha) Z_{\xi_\beta}(\alpha')^\top] = \nabla \xi_\beta(\gamma_{r(\alpha, \alpha')}).$$

We can realize this process as $Z_{\xi_\beta}(\alpha) = Z_\xi(\alpha) + \sum_{\theta \in \Theta_{\mathbb{Q}}} \beta_\theta Z_\theta(\alpha)$, where all the processes on the right-hand side are independent and centered, with covariances

$$\mathbb{E}[Z_\xi(\alpha) Z_\xi(\alpha')^\top] = \nabla \xi(\gamma_{r(\alpha, \alpha')}), \quad \mathbb{E}[Z_\theta(\alpha) Z_\theta(\alpha')^\top] = \nabla \xi_\theta(\gamma_{r(\alpha, \alpha')}). \quad (4.3)$$

These processes exist by Remark 3.9. Then consider the following probability measure on $\mathbb{N}^{s-1} \times \Sigma$:

$$G(\alpha, \sigma) \propto \nu_\alpha \exp \langle Z_{\xi_\beta}(\alpha) + \lambda, \sigma \rangle.$$

In what follows, we will write $((\alpha^\ell, \sigma^\ell))_{\ell \geq 1}$ to denote i.i.d. samples from G , and $\langle \cdot \rangle_G$ to denote expectation with respect to $G^{\otimes \infty}$. Differentiating (4.2) with respect to β_θ , we have

$$\frac{\partial}{\partial \beta_\theta} \mathcal{P}_{\xi_\beta}^{(1)}(\pi, \lambda) = \mathbb{E} \langle g_1 \rangle_G, \quad \text{where } g_\ell = \langle Z_\theta(\alpha^\ell), \sigma^\ell \rangle.$$

By Gaussian integration by parts [54, Lem. 1.1], this can be rewritten as

$$\frac{\partial}{\partial \beta_\theta} \mathcal{P}_{\xi_\beta}^{(1)}(\pi, \lambda) = \beta_\theta \mathbb{E} \langle f_{1,1} - f_{1,2} \rangle_G, \quad (4.4)$$

where $f_{\ell, \ell'} = C((\alpha^\ell, \sigma^\ell), (\alpha^{\ell'}, \sigma^{\ell'}))$, and $C : (\mathbb{N}^{s-1} \times \Sigma)^2 \rightarrow \mathbb{R}$ is given by

$$C((\alpha, \sigma), (\alpha', \sigma')) = \mathbb{E}[\langle Z_\theta(\alpha), \sigma \rangle \cdot \langle Z_\theta(\alpha'), \sigma' \rangle] \stackrel{(4.3)}{=} \langle \nabla \xi_\theta(\gamma_{r(\alpha, \alpha')}) \sigma, \sigma' \rangle.$$

Because $\sigma, \sigma' \in \Sigma$ are both standard basis vectors, we have

$$|\langle \nabla \xi_\theta(\gamma_{r(\alpha, \alpha')})\sigma, \sigma' \rangle| \leq \|\nabla \xi_\theta(\gamma_{r(\alpha, \alpha')})\|_\infty \leq \sup_{\|R\|_1 \leq 1} \|\nabla \xi_\theta(R)\|_\infty \leq C.$$

Therefore, for any $\ell, \ell' \geq 1$ we have

$$|f_{\ell, \ell'}| \leq C.$$

By differentiating (4.4) using the product rule, we obtain

$$\left(\frac{\partial}{\partial \beta_\theta}\right)^2 \mathcal{P}_{\xi_\beta}^{(1)}(\pi, \lambda) = \mathbb{E}\langle f_{1,1} - f_{1,2} \rangle_G + \beta_\theta \frac{\partial}{\partial \beta_\theta} \mathbb{E}\langle f_{1,1} - f_{1,2} \rangle_G.$$

Observe that $f_{\ell, \ell'}$ has no dependence on β_θ , and so direct calculation yields

$$\begin{aligned} \frac{\partial}{\partial \beta_\theta} \mathbb{E}\langle f_{1,1} \rangle_G &= \mathbb{E}\langle f_{1,1} \cdot (g_1 - g_2) \rangle_G, \\ \frac{\partial}{\partial \beta_\theta} \mathbb{E}\langle f_{1,2} \rangle_G &= \mathbb{E}\langle f_{1,2} \cdot (g_1 + g_2 - 2g_3) \rangle_G = 2\mathbb{E}\langle f_{1,2} \cdot (g_1 - g_3) \rangle_G. \end{aligned}$$

Applying Gaussian integration by parts once more (this time using [54, Lem. 1.2]), we have

$$\begin{aligned} \mathbb{E}\langle f_{1,1} \cdot g_1 \rangle_G &= \beta_\theta \mathbb{E}\langle f_{1,1} \cdot (f_{1,1} - f_{1,2}) \rangle_G, \\ \mathbb{E}\langle f_{1,1} \cdot g_2 \rangle_G &= \beta_\theta \mathbb{E}\langle f_{1,1} \cdot (f_{2,1} + f_{2,2} - 2f_{2,3}) \rangle_G, \\ \mathbb{E}\langle f_{1,2} \cdot g_1 \rangle_G &= \beta_\theta \mathbb{E}\langle f_{1,2} \cdot (f_{1,1} + f_{1,2} - 2f_{1,3}) \rangle_G, \\ \mathbb{E}\langle f_{1,2} \cdot g_3 \rangle_G &= \beta_\theta \mathbb{E}\langle f_{1,2} \cdot (f_{3,1} + f_{3,2} + f_{3,3} - 3f_{3,4}) \rangle_G. \end{aligned}$$

The four previous displays together show

$$\left| \left(\frac{\partial}{\partial \beta_\theta}\right)^2 \mathcal{P}_{\xi_\beta}^{(1)}(\pi, \lambda) \right| \leq C(1 + \beta_\theta^2). \quad (4.5)$$

The combination of (4.1) and (4.5) yields the desired inequality. \square

Lemma 4.2. Assume ξ satisfies (A1), (A2'), and β satisfies (B1), (B2). Then for any $d \in \mathcal{D}$, the following statements hold.

(a) $\inf_{\pi \in \Pi_d} \mathcal{P}_{\xi_\beta}(\pi) \geq 0$.

(b) For any $\theta \in \Theta_{\mathbb{Q}}$, the function $\beta_\theta \mapsto \inf_{\pi \in \Pi_d} \mathcal{P}_{\xi_\beta}(\pi)$ is convex on an open interval.

Proof. By Theorem 1.7 we have

$$f \lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} F_{N, \xi_\beta}(d, \varepsilon) = \inf_{\pi \in \Pi_d} \mathcal{P}_{\xi_\beta}(\pi). \quad (4.6)$$

Given any $\varepsilon > 0$, the set $\Sigma^N(d, \varepsilon)$ from (1.4) is nonempty for all large N . For any $\sigma \in \Sigma^N(d, \varepsilon)$, we have the trivial lower bound

$$F_{N, \xi_\beta}(d, \varepsilon) \geq \frac{1}{N} \mathbb{E} H_{N, \xi_\beta}(\sigma) = 0. \quad (4.7)$$

Part (a) now follows from (4.6) and (4.7).

For part (b), note that (4.6) remains true for all β_θ in some open interval, since (B1) and (B2) remain true for all β_θ in some open interval. From the decomposition (2.19), a standard application of Hölder's inequality shows that whenever $\Sigma^N(d, \varepsilon)$ is nonempty, the map $\beta_\theta \mapsto F_{N, \xi_\beta}(d, \varepsilon)$ is convex for any $\theta \in \Theta_{\mathbb{Q}}$. Therefore, $\beta_\theta \mapsto \inf_{\pi \in \Pi_d} \mathcal{P}_{\xi_\beta}(\pi)$ is a limit of convex functions (via (4.6)) and hence convex. \square

Finally, we show that Lemmas 4.1 and 4.2 imply Proposition 2.9.

Proof of Proposition 2.9. Fix $\theta \in \Theta_{\mathbb{Q}}$ and $d \in \mathcal{D}$.

For clarity, we make the dependence of $\mathcal{P}_{\xi\beta}(\pi, \lambda)$ on β_{θ} explicit by writing $F_{\pi, \lambda}(\beta_{\theta}) := \mathcal{P}_{\xi\beta}(\pi, \lambda)$, as well as

$$F(\beta_{\theta}) := \inf_{\pi \in \Pi_d} \mathcal{P}_{\xi\beta}(\pi) = \inf_{\pi \in \Pi_d, \lambda \in \mathbb{R}^K} F_{\pi, \lambda}(\beta_{\theta}).$$

By Lemma 4.2(a), $F(\beta_{\theta})$ is a finite number. So for any $\varepsilon > 0$, there exist $\pi_{\varepsilon} \in \Pi_d$ and $\lambda_{\varepsilon} \in \mathbb{R}^K$ such that

$$F_{\pi_{\varepsilon}, \lambda_{\varepsilon}}(\beta_{\theta}) \leq F(\beta_{\theta}) + \varepsilon. \quad (4.8)$$

Since $\beta_{\theta} \mapsto F(\beta_{\theta})$ is convex by Lemma 4.2(b), the differentiability we seek is equivalent to the subdifferential $\partial F(\beta_{\theta})$ consisting of a single point. So consider $x \in \partial F(\beta_{\theta})$. For all small enough $h > 0$, we have

$$x \leq \frac{F(\beta_{\theta} + h) - F(\beta_{\theta})}{h}. \quad (4.9)$$

Combining (4.8) and (4.9) with the trivial inequality $F(\beta_{\theta} + h) \leq F_{\pi_{\varepsilon}, \lambda_{\varepsilon}}(\beta_{\theta} + h)$, we see

$$x \leq \frac{F_{\pi_{\varepsilon}, \lambda_{\varepsilon}}(\beta_{\theta} + h) - F_{\pi_{\varepsilon}, \lambda_{\varepsilon}}(\beta_{\theta}) + \varepsilon}{h} \leq \frac{\partial}{\partial \beta_{\theta}} F_{\pi_{\varepsilon}, \lambda_{\varepsilon}}(\beta_{\theta}) + Ch(1 + (|\beta_{\theta}| + h)^2) + \frac{\varepsilon}{h}, \quad (4.10)$$

where the second inequality follows from Taylor's theorem together with Lemma 4.1.

By analogous reasoning,

$$x \geq \frac{\partial}{\partial \beta_{\theta}} F_{\pi_{\varepsilon}, \lambda_{\varepsilon}}(\beta_{\theta}) - Ch(1 + (|\beta_{\theta}| + h)^2) - \frac{\varepsilon}{h}. \quad (4.11)$$

Finally, take $h = \sqrt{\varepsilon}$ and send $\varepsilon \rightarrow 0$. The inequalities (4.10) and (4.11) together show that x can take at most one value. \square

5. Continuity and Duality

The purpose of this section is threefold. First we prove Proposition 2.13, part (b) of which gives Lipschitz continuity of the Parisi functional. Next we use this continuity in the proof of a duality statement (Lemma 5.3) that is needed in Sect. 6 to relate the functional appearing in the Aizenman–Sims–Starr Scheme to the Parisi functional. Finally, we solve this duality in the balanced case (Lemma 5.4), thereby eliminating the infimum appearing in (1.35) and fulfilling the promise of Remark 1.9.

5.1. Continuity of free energy and the Parisi functional. Both parts of Proposition 2.13 follow from interpolation arguments. We begin with part (a).

Proof of Proposition 2.13(a). For $t \in [0, 1]$, define the following interpolating free energy:

$$\phi(t) = \frac{1}{N} \mathbb{E} \log \sum_{\sigma \in \Sigma^N(d)} \exp \mathcal{H}_t(\sigma), \quad \text{where } \mathcal{H}_t(\sigma) = \sqrt{t} H_{N,\xi}(\sigma) + \sqrt{1-t} H_{N,\tilde{\xi}}(\sigma).$$

We assume $H_{N,\xi}$ is independent of $H_{N,\tilde{\xi}}$. Let $\langle \cdot \rangle_t$ denote expectation with respect to the Gibbs measure $G_t(\sigma) \propto \exp \mathcal{H}_t(\sigma)$ on $\Sigma^N(d)$. By differentiating the expression for $\phi(t)$ with respect to t , we have

$$\phi'(t) = \frac{1}{N} \mathbb{E} \left\langle \frac{\partial \mathcal{H}_t(\sigma)}{\partial t} \right\rangle_t.$$

By Gaussian integration by parts [54, Lem. 1.1], this can be rewritten as

$$\phi'(t) = N^{-1} \mathbb{E} \langle \mathcal{C}(\sigma^1, \sigma^1) - \mathcal{C}(\sigma^1, \sigma^2) \rangle_t,$$

where σ^1, σ^2 denote independent samples from G_t , and $\mathcal{C}: \Sigma^N(d) \times \Sigma^N(d) \rightarrow \mathbb{R}$ is given by

$$\mathcal{C}(\sigma, \sigma') = \mathbb{E} \left[\frac{\partial \mathcal{H}_t}{\partial t}(\sigma) \mathcal{H}_t(\sigma') \right] \stackrel{(1.14)}{=} \frac{1}{2} \left[\xi \left(\frac{\sigma \sigma'^T}{N} \right) - \tilde{\xi} \left(\frac{\sigma \sigma'^T}{N} \right) \right].$$

Since $\|\sigma \sigma'^T\|_1 \leq N$ for any $\sigma, \sigma' \in \Sigma^N$, we deduce that

$$\sup_{t \in (0,1)} |\phi'(t)| \leq \sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)|.$$

In summary,

$$|\mathbf{F}_{N,\xi} - \mathbf{F}_{N,\tilde{\xi}}| = |\phi(1) - \phi(0)| \leq \sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)|.$$

□

We now turn our attention to Proposition 2.13(b), which will be an easy consequence of the following lemma. In fact, this lemma is what allows the Parisi functional to be extended to arbitrary paths in the first place.

Lemma 5.1. *Assume ξ and $\tilde{\xi}$ satisfy (A1). Then for any $d \in \mathcal{D}$, discrete paths $\pi, \tilde{\pi} \in \Pi_d^{\text{disc}}$, $\lambda \in \mathbb{R}^\kappa$, and nonempty $S \subseteq \Sigma^N$, we have*

$$\begin{aligned} & |\mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; S) - \mathcal{P}_{N,\tilde{\xi}}^{(1)}(\tilde{\pi}, \lambda; S)| \\ & \leq \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \tilde{\xi}(R)\|_\infty + \frac{1}{2} \sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty \int_0^1 \|\pi(t) - \tilde{\pi}(t)\|_1 dt. \end{aligned} \tag{5.1}$$

Similarly,

$$\begin{aligned} & |\mathcal{P}_\xi^{(2)}(\pi) - \mathcal{P}_\xi^{(2)}(\tilde{\pi})| \\ & \leq \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \tilde{\xi}(R)\|_\infty + \frac{1}{2} \sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty \int_0^1 \|\pi(t) - \tilde{\pi}(t)\|_1 dt. \end{aligned} \quad (5.2)$$

Proof. As in (3.5), we assume π and $\tilde{\pi}$ are of the form

$$\pi(t) = \gamma_r \quad \text{and} \quad \tilde{\pi}(t) = \tilde{\gamma}_r \quad \text{for } t \in (m_{r-1}, m_r], r \in \{1, \dots, s\}.$$

There is no loss of generality in using the same partition $0 = m_0 < m_1 < \dots < m_s = 1$ for both paths, thanks to Remark 3.6.

First we prove (5.1). Let $(\tilde{Z}_i(\alpha))_{1 \leq i \leq N, \alpha \in \mathbb{N}^{s-1}}$ be an independent copy of the process $(Z_i(\alpha))_{1 \leq i \leq N, \alpha \in \mathbb{N}^{s-1}}$ from (3.6). We assume both processes are independent of the RPC weights $(\nu_\alpha)_{\alpha \in \mathbb{N}^{s-1}}$. Then define an interpolating process

$$\mathcal{H}_t(\alpha, \sigma) = \sum_{i=1}^N \left(\sqrt{t} \langle Z_i(\alpha), \sigma_i \rangle + \sqrt{1-t} \langle \tilde{Z}_i(\alpha), \sigma_i \rangle + \langle \lambda, \sigma_i \rangle \right), \quad t \in [0, 1], \quad (5.3)$$

as well as the corresponding energy

$$\phi(t) = \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in S} \nu_\alpha \exp \mathcal{H}_t(\alpha, \sigma). \quad (5.4)$$

We then have $\phi(1) = \mathcal{P}_{N, \xi}^{(1)}(\pi, \lambda; S)$ while $\phi(0) = \mathcal{P}_{N, \xi}^{(1)}(\tilde{\pi}, \lambda; S)$, and so we seek an upper bound on $|\phi(1) - \phi(0)|$. For (5.1) it suffices to show a uniform bound on the derivative of ϕ :

$$\sup_{t \in (0, 1)} |\phi'(t)| \leq \text{R.H.S. of (5.1)}. \quad (5.5)$$

To this end, let $\langle \cdot \rangle_t$ denote expectation with respect to the probability measure $G_t(\alpha, \sigma) \propto \nu_\alpha \exp \mathcal{H}_t(\alpha, \sigma)$ on $\mathbb{N}^{s-1} \times S$. By differentiating (5.4) with respect to t , we have

$$\phi'(t) = \frac{1}{N} \mathbb{E} \left\langle \frac{\partial \mathcal{H}_t(\alpha, \sigma)}{\partial t} \right\rangle_t. \quad (5.6)$$

Define $\mathcal{C}: (\mathbb{N}^{s-1} \times S)^2 \rightarrow \mathbb{R}$ by

$$\begin{aligned} \mathcal{C}((\alpha, \sigma), (\alpha', \sigma')) &= \frac{1}{N} \mathbb{E} \left[\frac{\partial \mathcal{H}_t(\alpha, \sigma)}{\partial t} \mathcal{H}_t(\alpha', \sigma') \right] \\ &\stackrel{(5.3)}{=} \frac{1}{2N} \sum_{i=1}^N \mathbb{E} [\langle Z_i(\alpha), \sigma_i \rangle \langle Z_i(\alpha'), \sigma'_i \rangle - \langle \tilde{Z}_i(\alpha), \sigma_i \rangle \langle \tilde{Z}_i(\alpha'), \sigma'_i \rangle] \\ &\stackrel{(3.6)}{=} \frac{1}{2N} \sum_{i=1}^N [\langle \nabla \xi(\gamma_{r(\alpha, \alpha')}) \sigma_i, \sigma'_i \rangle - \langle \nabla \tilde{\xi}(\tilde{\gamma}_{r(\alpha, \alpha')}) \sigma_i, \sigma'_i \rangle], \end{aligned} \quad (5.7)$$

Then Gaussian integration by parts [54, Lem. 1.1] transforms (5.6) to

$$\phi'(t) = \mathbb{E} \left[\mathcal{C}((\alpha^1, \sigma^1), (\alpha^1, \sigma^1)) - \mathcal{C}((\alpha^1, \sigma^1), (\alpha^2, \sigma^2)) \right]_t, \quad (5.8)$$

where $(\alpha^1, \sigma^1), (\alpha^2, \sigma^2)$ denote independent samples from G_t . Since both σ_i and σ'_i are standard basis vectors, it follows from (5.7) that

$$\begin{aligned} |\mathcal{C}((\alpha, \sigma), (\alpha', \sigma'))| &\leq \frac{1}{2} \|\nabla \xi(\gamma_{r(\alpha, \alpha')}) - \nabla \tilde{\xi}(\tilde{\gamma}_{r(\alpha, \alpha')})\|_\infty \\ &\leq \frac{1}{2} (\|\nabla \xi(\gamma_{r(\alpha, \alpha')}) - \nabla \xi(\tilde{\gamma}_{r(\alpha, \alpha')})\|_\infty + \|\nabla \xi(\tilde{\gamma}_{r(\alpha, \alpha')}) - \nabla \tilde{\xi}(\tilde{\gamma}_{r(\alpha, \alpha')})\|_\infty) \\ &\leq \frac{1}{2} \left(\sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty \cdot \|\gamma_{r(\alpha, \alpha')} - \tilde{\gamma}_{r(\alpha, \alpha')}\|_1 + \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \tilde{\xi}(R)\|_\infty \right). \end{aligned} \quad (5.9)$$

In the special case $\alpha = \alpha'$, we can use the assumption that both π and $\tilde{\pi}$ belong to Π_d^{disc} in order to conclude that

$$\gamma_{r(\alpha, \alpha)} = \gamma_s = \tilde{\gamma}_s = \text{diag}(d).$$

That is, the first term in the final line of (5.9) vanishes when $\alpha = \alpha'$. Therefore, using (5.9) in (5.8) leads to

$$|\phi'(t)| \leq \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \tilde{\xi}(R)\|_\infty + \frac{1}{2} \sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty \mathbb{E} \left(\|\gamma_{r(\alpha^1, \alpha^2)} - \tilde{\gamma}_{r(\alpha^1, \alpha^2)}\|_1 \right)_t.$$

By [54, Thm. 4.4], the marginal of $r(\alpha^1, \alpha^2)$ under $\mathbb{E}(G_t^{\otimes 2})$ is the same with the marginal of $r(\alpha^1, \alpha^2)$ under $\mathbb{E}(\nu^{\otimes 2})$, where ν is the RPC from Theorem 3.4. In particular,

$$\mathbb{E} \left(\|\gamma_{r(\alpha^1, \alpha^2)} - \tilde{\gamma}_{r(\alpha^1, \alpha^2)}\|_1 \right)_t = \sum_{r=1}^s \|\gamma_r - \tilde{\gamma}_r\|_1 (m_r - m_{r-1}) = \int_0^1 \|\pi(t) - \tilde{\pi}(t)\|_1 dt.$$

Using this calculation in the previous inequality yields (5.5).

Now we turn to proving (5.2). Since π and $\tilde{\pi}$ both belong to Π_d , we have

$$|\mathcal{P}_\xi^{(2)}(\pi) - \mathcal{P}_\xi^{(2)}(\tilde{\pi})| \stackrel{(1.33)}{\leq} \frac{1}{2} \int_0^1 |\vartheta_\xi(\pi(t)) - \vartheta_\xi(\tilde{\pi}(t))| dt. \quad (5.10)$$

For any γ and $\tilde{\gamma}$, the triangle inequality gives

$$|\vartheta_\xi(\gamma) - \vartheta_\xi(\tilde{\gamma})| \leq |\vartheta_\xi(\gamma) - \vartheta_\xi(\tilde{\gamma})| + |\vartheta_\xi(\tilde{\gamma}) - \vartheta_\xi(\tilde{\gamma})|. \quad (5.11)$$

When $\|\gamma\|_1, \|\tilde{\gamma}\|_1 \leq 1$, the first term on the right-hand side of (5.11) obeys the upper bound

$$|\vartheta_\xi(\gamma) - \vartheta_\xi(\tilde{\gamma})| \leq \sup_{\|R\|_1 \leq 1} \|\nabla \vartheta_\xi(R)\|_\infty \cdot \|\gamma - \tilde{\gamma}\|_1. \quad (5.12)$$

Now recall from (1.27) that $\vartheta_\xi(R) = \langle R, \nabla \xi(R) \rangle - \xi(R)$. It follows that the (a, b) entry of $\nabla \vartheta_\xi(R)$ is

$$\frac{\partial \vartheta}{\partial R_{a,b}} = R_{a,b} \sum_{k,k'=1}^{\kappa} \frac{\partial^2 \xi}{\partial R_{a,b} \partial R_{k,k'}},$$

and thus $\|\nabla \vartheta_\xi(R)\|_\infty \leq \|R\|_1 \cdot \|\nabla^2 \xi(R)\|_\infty$. Using this in (5.12) results in

$$|\vartheta_\xi(\gamma) - \vartheta_\xi(\tilde{\gamma})| \leq \sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty \cdot \|\gamma - \tilde{\gamma}\|_1. \quad (5.13)$$

Meanwhile, the second term on the right-hand side of (5.11) satisfies

$$\begin{aligned} |\vartheta_\xi(\tilde{\gamma}) - \vartheta_{\tilde{\xi}}(\tilde{\gamma})| &\leq |\langle \tilde{\gamma}, \nabla \xi(\tilde{\gamma}) - \nabla \tilde{\xi}(\tilde{\gamma}) \rangle| + |\xi(\tilde{\gamma}) - \tilde{\xi}(\tilde{\gamma})| \\ &\leq \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \tilde{\xi}(R)\|_\infty + \sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)|. \end{aligned} \quad (5.14)$$

Since $\xi(0) = \tilde{\xi}(0) = 0$ by (A1), the second supremum above can be further controlled as

$$\sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)| \leq \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \tilde{\xi}(R)\|_\infty.$$

Therefore, (5.14) can be rewritten as

$$|\vartheta_\xi(\tilde{\gamma}) - \vartheta_{\tilde{\xi}}(\tilde{\gamma})| \leq 2 \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \tilde{\xi}(R)\|_\infty. \quad (5.15)$$

Applying both (5.13) and (5.15) in (5.11) results in

$$|\vartheta_\xi(\gamma) - \vartheta_{\tilde{\xi}}(\tilde{\gamma})| \leq 2 \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \tilde{\xi}(R)\|_\infty + \sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty \cdot \|\gamma - \tilde{\gamma}\|_1.$$

Applying this uniform bound to the integrand in (5.10) yields (5.2) as desired. \square

A key consequence of Lemma 5.1 is that the map $\Pi_d^{\text{disc}} \ni \pi \mapsto \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; S)$ has a unique continuous extension to all of Π_d . We continue to write $\mathcal{P}_{N,\xi}^{(1)}$ for this extension, including the special case $\mathcal{P}_{\xi}^{(1)}(\pi, \lambda) = \mathcal{P}_{1,\xi}^{(1)}(\pi, \lambda; \Sigma)$ used in (1.31). Of course, these extensions admit the same estimates (5.1) and (5.2) for any $\pi, \tilde{\pi} \in \Pi_d$. We now record how these estimates lead to the continuity of Parisi functional, as claimed in Proposition 2.13(b).

Proof of Proposition 2.13(b). Recall the definition of the Parisi functional from (1.35):

$$\begin{aligned} \mathcal{P}_\xi(\pi) &= \inf_{\lambda \in \mathbb{R}^K} [\mathcal{P}_\xi^{(1)}(\pi, \lambda) + \mathcal{P}_\xi^{(2)}(\pi, \lambda) - \langle \lambda, d \rangle] \\ &\stackrel{(3.10)}{=} \inf_{\lambda \in \mathbb{R}^K} [\mathcal{P}_{1,\xi}^{(1)}(\pi, \lambda; \Sigma) + \mathcal{P}_\xi^{(2)}(\pi, \lambda) - \langle \lambda, d \rangle]. \end{aligned}$$

By combining (5.1) and (5.2), we have

$$\begin{aligned} &|\mathcal{P}_{1,\xi}^{(1)}(\pi, \lambda; \Sigma) + \mathcal{P}_\xi^{(2)}(\pi, \lambda) - \mathcal{P}_{1,\tilde{\xi}}^{(1)}(\tilde{\pi}, \lambda; \Sigma) - \mathcal{P}_{\tilde{\xi}}^{(2)}(\tilde{\pi}, \lambda)| \\ &\leq 2 \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R) - \nabla \tilde{\xi}(R)\|_\infty + \sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty \int_0^1 \|\pi(t) - \tilde{\pi}(t)\|_1 dt. \end{aligned} \quad (5.16)$$

Since the right-hand side of (5.16) does not depend on λ , it follows that

$$|\mathcal{P}_\xi(\pi) - \mathcal{P}_{\tilde{\xi}}(\tilde{\pi})| \leq \text{R.H.S. of (5.16)}.$$

This is exactly what was claimed in (2.24). \square

5.2. Duality of magnetization with Lagrange multiplier. The following lemma makes precise the discussion in Remark 3.8. It is a generalization of [58, Lem. 2] and is proved similarly. The proof is included in “Appendix D” for completeness.

Lemma 5.2. *Assume ξ satisfies (A1). Assume $d^N \in \mathcal{D}_N$ converges to $d \in \mathcal{D}$ as $N \rightarrow \infty$. Then for any $d' \in \mathcal{D}$ and $\pi \in \Pi_{d'}^{\text{disc}}$, we have*

$$\lim_{N \rightarrow \infty} \mathcal{P}_{N,\xi}^{(1)}(\pi, 0; \Sigma^N(d^N)) = \inf_{\lambda \in \mathbb{R}^K} [\mathcal{P}_\xi^{(1)}(\pi, \lambda) - \langle \lambda, d \rangle]. \quad (5.17)$$

On a technical note, we point out that d' does not need to be related to d . We will, however, only need the case $d' = d$. Now we bootstrap to a stronger statement.

Lemma 5.3. *Assume ξ satisfies (A1). Assume $d^N \in \mathcal{D}_N$ converges to $d \in \mathcal{D}$ as $N \rightarrow \infty$. Let $\pi \in \Pi_{d'}$ where $d' \in \mathcal{D}$ is arbitrary, and consider any sequence $(\pi_N)_{N \geq 1}$ in $\Pi_{d'}$ satisfying*

$$\lim_{N \rightarrow \infty} \int_0^1 \|\pi_N(t) - \pi(t)\|_1 dt = 0, \quad (5.18)$$

We then have

$$\lim_{N \rightarrow \infty} \mathcal{P}_{N,\xi}^{(1)}(\pi_N, 0; \Sigma^N(d^N)) = \inf_{\lambda \in \mathbb{R}^K} [\mathcal{P}_\xi^{(1)}(\pi, \lambda) - \langle \lambda, d \rangle]. \quad (5.19)$$

Proof. Our goal is to extend (5.17) to general paths $\pi \in \Pi_{d'}$ and to allow for π on the left-hand side of (5.17) to be replaced with an approximating sequence $(\pi_N)_{N \geq 1}$. To this end, take any $\tilde{\pi} \in \Pi_{d'}^{\text{disc}}$ and observe that

$$\begin{aligned} & \left| \mathcal{P}_{N,\xi}^{(1)}(\pi_N, 0; \Sigma^N(d^N)) - \inf_{\lambda \in \mathbb{R}^K} [\mathcal{P}_\xi^{(1)}(\pi, \lambda) - \langle \lambda, d \rangle] \right| \\ & \leq \left| \mathcal{P}_{N,\xi}^{(1)}(\pi_N, 0; \Sigma^N(d^N)) - \mathcal{P}_{N,\xi}^{(1)}(\tilde{\pi}, 0; \Sigma^N(d^N)) \right| \\ & \quad + \left| \mathcal{P}_{N,\xi}^{(1)}(\tilde{\pi}, 0; \Sigma^N(d^N)) - \inf_{\lambda \in \mathbb{R}^K} [\mathcal{P}_\xi^{(1)}(\tilde{\pi}, \lambda) - \langle \lambda, d \rangle] \right| \\ & \quad + \left| \inf_{\lambda \in \mathbb{R}^K} [\mathcal{P}_\xi^{(1)}(\tilde{\pi}, \lambda) - \langle \lambda, d \rangle] - \inf_{\lambda \in \mathbb{R}^K} [\mathcal{P}_\xi^{(1)}(\pi, \lambda) - \langle \lambda, d \rangle] \right| \\ & \stackrel{(5.1), (5.17)}{\leq} o(1) + \sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty \int_0^1 (\|\pi_N(t) - \tilde{\pi}(t)\|_1 + \|\tilde{\pi}(t) - \pi(t)\|_1) dt \\ & \stackrel{(5.18)}{\leq} o(1) + \sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty \int_0^1 2\|\tilde{\pi}(t) - \pi(t)\|_1 dt. \end{aligned} \quad (5.20)$$

Since $\Pi_{d'}^{\text{disc}}$ is dense in $\Pi_{d'}$ (with respect to the norm (1.32)), we can choose $\tilde{\pi}$ to make the integral on the final line of (5.20) arbitrarily small. The claim (5.19) follows. \square

5.3. Duality in the balanced case. When $d = d_{\text{bal}}$, we can solve the infimum on the right-hand side of (5.19), at least for paths π belonging to the set Π^* from (1.12). Namely, the infimum is achieved at $\lambda = 0$.

Lemma 5.4. *Assume ξ satisfies (A1) and (A3). Then for any $\pi \in \Pi^*$, we have*

$$\inf_{\lambda \in \mathbb{R}^\kappa} [\mathcal{P}_\xi^{(1)}(\pi, \lambda) - \langle \lambda, d_{\text{bal}} \rangle] = \mathcal{P}_\xi^{(1)}(\pi, 0). \quad (5.21)$$

In what follows, it is necessary to recall the permutation action $R \mapsto \omega \bullet R$ from (1.23). Namely, the (k, k') entry of R is equal to the $(\omega(k), \omega(k'))$ entry of $\omega \bullet R$.

Lemma 5.5. *Assume ξ satisfies (A1) and (A3). Then for any permutation $\omega \in S_\kappa$,*

$$\omega \bullet \nabla \xi(R) = \nabla \xi(\omega \bullet R) \quad \text{for all } R \in \mathbb{R}^{\kappa \times \kappa}. \quad (5.22)$$

Proof. First make a general observation not requiring symmetry: if we define the function $\xi^\omega(R) = \xi(\omega \bullet R)$, then

$$(k, k') \text{ entry of } \nabla \xi^\omega(R) = (\omega(k), \omega(k')) \text{ entry of } \nabla \xi(\omega \bullet R). \quad (5.23)$$

The assumption that ξ is symmetric means $\xi^\omega(R) = \xi(R)$. In this case, we can now chase definitions from the left-hand side of (5.22) to the right-hand side of (5.22):

$$\begin{aligned} (k, k') \text{ entry of } \omega \bullet \nabla \xi(R) &\stackrel{(1.23)}{=} (\omega^{-1}(k), \omega^{-1}(k')) \text{ entry of } \nabla \xi(R) \\ &\stackrel{(5.23)}{=} (k, k') \text{ entry of } \nabla \xi(\omega \bullet R). \end{aligned}$$

□

Proof of Lemma 5.4. It suffices to verify (5.21) for discrete paths $\pi \in \Pi_d^{\text{disc}}$, since such paths are dense in Π_d , and both sides of (5.21) are continuous in π thanks to (5.1). So let us assume $\pi \in \Pi_d^{\text{disc}}$ is given by (3.5). It follows from [58, Lem. 6] that $\lambda \mapsto \mathcal{P}_\xi^{(1)}(\pi, \lambda)$ is convex. This implies $\lambda \mapsto \mathcal{P}_\xi^{(1)}(\pi, \lambda) - \langle \lambda, d_{\text{bal}} \rangle$ is convex and thus achieves a global minimum wherever its gradient is zero. Therefore, it suffices to show

$$\left. \frac{\partial \mathcal{P}_\xi^{(1)}(\pi, \lambda)}{\partial \lambda_k} \right|_{\lambda=0} = \frac{\partial}{\partial \lambda_k} \langle \lambda, d_{\text{bal}} \rangle \quad \text{for each } k \in \{1, \dots, \kappa\}.$$

Since $d_{\text{bal}} = \kappa^{-1} \mathbf{1}$, this amounts to showing

$$\left. \frac{\partial \mathcal{P}_\xi^{(1)}(\pi, \lambda)}{\partial \lambda_k} \right|_{\lambda=0} = \frac{1}{\kappa} \quad \text{for each } k \in \{1, \dots, \kappa\}. \quad (5.24)$$

With $(v_\alpha)_{\alpha \in \mathbb{N}^{s-1}}$ denoting the RPC associated to (3.4), we have

$$\mathcal{P}_\xi^{(1)}(\pi, \lambda) \stackrel{(3.10)}{=} \mathcal{P}_{1,\xi}^{(1)}(\pi, \lambda; \Sigma) \stackrel{(3.7)}{=} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{k=1}^{\kappa} v_\alpha \exp \langle Z(\alpha) + \lambda, \mathbf{e}_k \rangle, \quad (5.25)$$

where $(Z(\alpha))_{\alpha \in \mathbb{N}^{s-1}}$ is a centered \mathbb{R}^κ -valued Gaussian process with covariance structure

$$\mathbb{E}[Z(\alpha)Z(\alpha')^T] = \nabla \xi(\gamma_{r(\alpha, \alpha')}), \quad \alpha, \alpha' \in \mathbb{N}^{s-1}. \quad (5.26)$$

Now differentiate (5.25) with respect to λ_k and evaluate at $\lambda = 0$:

$$\frac{\partial \mathcal{P}_\xi^{(1)}(\pi, \lambda)}{\partial \lambda_k} \Big|_{\lambda=0} = \mathbb{E} \left[\frac{\sum_{\alpha \in \mathbb{N}^{s-1}} \nu_\alpha \exp \langle Z(\alpha), \mathbf{e}_k \rangle}{\sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{k'=1}^K \nu_\alpha \exp \langle Z(\alpha), \mathbf{e}_{k'} \rangle} \right]. \quad (5.27)$$

If we sum over k , then the numerators on the right-hand side add up to the denominator:

$$\sum_{k=1}^K \frac{\partial \mathcal{P}_\xi^{(1)}(\pi, \lambda)}{\partial \lambda_k} \Big|_{\lambda=0} = 1. \quad (5.28)$$

On the other hand, we make the following claim.

Claim 5.6. $(\langle Z(\alpha), \mathbf{e}_k \rangle)_{\alpha \in \mathbb{N}^{s-1}, 1 \leq k \leq K}$ is exchangeable in k .

Proof. For any permutation $\omega \in S_K$, we have

$$\begin{aligned} \mathbb{E}[\langle Z(\alpha), \mathbf{e}_{\omega(k)} \rangle \langle Z(\alpha'), \mathbf{e}_{\omega(k')} \rangle] &\stackrel{(5.26)}{=} \langle \nabla \xi(\gamma_{r(\alpha, \alpha')} \mathbf{e}_{\omega(k)}, \mathbf{e}_{\omega(k')}) \rangle \\ &\stackrel{(1.23)}{=} \langle (\omega^{-1} \bullet \nabla \xi(\gamma_{r(\alpha, \alpha')})) \mathbf{e}_k, \mathbf{e}_{k'} \rangle \\ &\stackrel{(5.22)}{=} \langle (\nabla \xi(\omega^{-1} \bullet \gamma_{r(\alpha, \alpha')})) \mathbf{e}_k, \mathbf{e}_{k'} \rangle. \end{aligned}$$

Since $\pi \in \Pi^\star$, the matrix $\gamma_{r(\alpha, \alpha')}$ has identical diagonal entries and identical off-diagonal entries. In particular, it is invariant under the group action (1.23): $\omega^{-1} \bullet \gamma_{r(\alpha, \alpha')} = \gamma_{r(\alpha, \alpha')}$. Therefore, the previous display shows

$$\mathbb{E}[\langle Z(\alpha), \mathbf{e}_{\omega(k)} \rangle \langle Z(\alpha'), \mathbf{e}_{\omega(k')} \rangle] = \mathbb{E}[\langle Z(\alpha), \mathbf{e}_k \rangle \langle Z(\alpha'), \mathbf{e}_{k'} \rangle].$$

Since $(\langle Z(\alpha), \mathbf{e}_k \rangle)_{\alpha \in \mathbb{N}^{s-1}, 1 \leq k \leq K}$ is a centered Gaussian process, this equivalence of covariances is enough to establish exchangeability. \square (Claim)

Recall that $(Z(\alpha))_{\alpha \in \mathbb{N}^{s-1}}$ is independent of the RPC weights $(\nu_\alpha)_{\alpha \in \mathbb{N}^{s-1}}$. Consequently, it follows from Claim 5.6 that the right-hand side of (5.27) does not depend on k . In light of (5.28), this forces (5.24). \square

6. Lower Bound

In this section we prove Proposition 2.11. The A.S.S. scheme is presented in Sect. 6.1, and the G.G. identities are proved in Sect. 6.2. These two inputs are then combined with symmetry to prove the proposition in Sect. 6.3.

Since the A.S.S. scheme is based on the cavity method, we must work with a cavity Hamiltonian $(H_{N,M,\xi}(\sigma))_{\sigma \in \Sigma^N}$. Lemma 6.1 stated below guarantees its existence. Throughout Sect. 6, we will write $G_{N,M,\xi}$ for the associated Gibbs measure. More precisely, given some $d^N \in \mathcal{D}_N$, $G_{N,M,\xi}$ is the probability measure on $\Sigma^N(d^N)$ defined by

$$G_{N,M,\xi}(\sigma) := \frac{\exp H_{N,M,\xi}(\sigma)}{Z_{N,M,\xi}(d^N)} \quad \text{where} \quad Z_{N,M,\xi}(d^N) := \sum_{\sigma \in \Sigma^N(d^N)} \exp H_{N,M,\xi}(\sigma). \quad (6.1)$$

We will write $\text{Law}(\mathcal{R}; \mathbb{E}(G_{N,M,\xi}^{\otimes \infty}))$ to denote the law of the array $\mathcal{R} = (\mathcal{R}_{\ell,\ell'})_{\ell,\ell' \geq 1}$, where

$$\mathcal{R}_{\ell,\ell'} = \frac{\sigma^\ell (\sigma^{\ell'})^\top}{N}, \quad (\sigma^1, \sigma^2, \dots) \sim \mathbb{E}(G_{N,M,\xi}^{\otimes \infty}).$$

Lemma 6.1. *Assume ξ satisfies (A1). Then for any positive integers N and M , there exists a centered Gaussian process $(H_{N,M,\xi}(\sigma))_{\sigma \in \Sigma^N}$ with covariance*

$$\mathbb{E}[H_{N,M,\xi}(\sigma)H_{N,M,\xi}(\sigma')] = (N+M)\xi\left(\frac{\sigma\sigma'^\top}{N+M}\right). \quad (6.2)$$

Proof. By the decomposition in (A1), it suffices to prove the following for each θ in the parameter set Θ from (1.16): there exists a centered Gaussian processes $(H_{N,M,\theta}(\sigma))_{\sigma \in \Sigma^N}$ such that

$$\mathbb{E}[H_{N,M,\theta}(\sigma)H_{N,M,\theta}(\sigma')] = (N+M)\xi_\theta\left(\frac{\sigma\sigma'^\top}{N+M}\right), \quad (6.3)$$

where ξ_θ is defined in (1.17). To this end, let $(H_{N,\theta}(\sigma))_{\sigma \in \Sigma^N}$ be the Gaussian process from Proposition 2.7(c), and then set

$$H_{N,M,\theta}(\sigma) = \left(\frac{N}{N+M}\right)^{(\deg(\theta)-1)/2} H_{N,\theta}(\sigma),$$

where $\deg(\theta)$ is defined in (1.18). This trivially leads to

$$\begin{aligned} \mathbb{E}[H_{N,M,\theta}(\sigma)H_{N,M,\theta}(\sigma')] &= \left(\frac{N}{N+M}\right)^{\deg(\theta)-1} \mathbb{E}[H_{N,\theta}(\sigma)H_{N,\theta}(\sigma')] \\ &\stackrel{(2.16)}{=} \left(\frac{N}{N+M}\right)^{\deg(\theta)-1} N\xi_\theta\left(\frac{\sigma\sigma'^\top}{N}\right). \end{aligned}$$

Then, by the fact that ξ_θ is a homogeneous polynomial, we achieve (6.3):

$$\begin{aligned} \mathbb{E}[H_{N,M,\theta}(\sigma)H_{N,M,\theta}(\sigma')] &= (N+M)\left(\frac{N}{N+M}\right)^{\deg(\theta)} \xi_\theta\left(\frac{\sigma\sigma'^\top}{N}\right) \\ &= (N+M)\xi_\theta\left(\frac{\sigma\sigma'^\top}{N+M}\right). \end{aligned} \quad (6.4)$$

□

6.1. Aizenman–Sims–Starr scheme. The A.S.S. scheme relates ratios of partition functions to the functional $\Psi_{N,\xi}$ from (3.3). In what follows, we use the following notation:

$$f = O_{\nabla^2 \xi}(g) \quad \text{means that} \quad |f| \leq g \cdot C \sup_{\|R\|_1 \leq 1} \|\nabla^2 \xi(R)\|_\infty, \quad (6.5)$$

where C is some universal constant.

Proposition 6.2. (A.S.S. scheme) Assume ξ and $\tilde{\xi}$ satisfy (A1). Let M be a positive integer, and suppose $d^N \in \mathcal{D}_N$, $d^{N+M} \in \mathcal{D}_{N+M}$, $\delta^M \in \mathcal{D}_M$ satisfy

$$Nd^N + M\delta^M = (N + M)d^{N+M}. \quad (6.6)$$

Let $\mathcal{L}_{N,M} = \text{Law}(\mathcal{R}; \mathbb{E}(G_{N,M,\xi}^{\otimes \infty}))$. We then have

$$\begin{aligned} \mathbb{E} \log \frac{Z_{N+M,\tilde{\xi}}(d^{N+M})}{Z_{N,\xi}(d^N)} + O_{\nabla^2 \xi} \left(\frac{M^2}{N+M} \right) + \frac{N+M}{2} \sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)| \\ \geq M \Psi_{M,\xi}(\mathcal{L}_{N,M}; d^N, \Sigma^M(\delta^M)). \end{aligned} \quad (6.7)$$

Proposition 6.2 follows from a standard application of Aizenman–Sims–Starr Scheme [2, 21]. The proof is included in “Appendix B” for completeness. In order to prove Proposition 2.11, we will take N, M to be a multiple of κ and use Proposition 6.2 with $d^N = d^{N+M} = \delta^M = d_{\text{bal}}$.

Remark 6.3. For the right-hand side of (6.7) to make sense, it is essential that the Gibbs measure $G_{N,M,\xi}$ in (6.1) is restricted to the space $\Sigma^N(d^N)$, so that $N^{-1}\sigma\sigma^\top = \text{diag}(d^N)$ for all σ in the support of $G_{N,M,\xi}$. This allows for Assumption 3.1(ii) to hold with

$$(\mathcal{X}, R, d) = (\Sigma^N(d^N), (\sigma, \sigma') \mapsto N^{-1}\sigma(\sigma')^\top, d^N).$$

In this setting, Assumption 3.1(i) is obvious, and 3.1(iii) follows from Proposition 2.7(d).

6.2. Ghirlanda–Guerra identities for generic models. In this section we prove that differentiability of the Parisi formula leads to the G.G. identities for generic models, thus fulfilling the outcomes discussed after Proposition 2.9. But in order to discuss the G.G. identities, we must first check that the relevant arrays are symmetric and positive semi-definite.

Lemma 6.4. Given $\sigma^1, \dots, \sigma^n \in \Sigma^N$, denote $\mathcal{R}_{\ell,\ell'} = N^{-1}\sigma^\ell(\sigma^{\ell'})^\top$. Then for any $w \in \mathbb{R}^\kappa$ and positive integer p , the array $(\mathcal{Q}_{\ell,\ell'})_{1 \leq \ell, \ell' \leq n}$ given by $\mathcal{Q}_{\ell,\ell'} = \langle \mathcal{R}_{\ell,\ell'}^{\circ p} w, w \rangle$ is symmetric and positive-semidefinite.

Proof. Fix $p \geq 1$ and $w \in \mathbb{R}^\kappa$. Consider the parameter $\theta = (p, 1, 1, w) \in \Theta$, for which (1.17) becomes $\xi_\theta(R) = \langle R^{\circ p} w, w \rangle$. By Proposition 2.7(c), there exists a centered Gaussian process $H: \Sigma^N \rightarrow \mathbb{R}$ with covariance

$$\mathbb{E}[H(\sigma)H(\sigma')] = \xi_\theta \left(\frac{\sigma\sigma'^\top}{N} \right) = N \langle R(\sigma, \sigma')^{\circ p} w, w \rangle,$$

where $R(\sigma, \sigma') = N^{-1}\sigma\sigma'^\top$. It follows that $(\sigma, \sigma') \mapsto \langle R(\sigma, \sigma')^{\circ p} w, w \rangle$ is a symmetric and positive-semidefinite map $\Sigma^N \times \Sigma^N \rightarrow \mathbb{R}$. \square

Proposition 6.5. Assume ξ satisfies (A1), (A2'), and $\beta = (\beta_\theta)_{\theta \in \Theta_{\mathbb{Q}}}$ satisfies (B1), (B2). Provided $\beta_\theta \neq 0$ for every $\theta \in \Theta_{\mathbb{Q}}$, the following holds for any positive integer M . Let $\mathcal{L}_{N,M} = \text{Law}(\mathcal{R}; \mathbb{E}(G_{N,M,\xi\beta}^{\otimes \infty}))$. Then any subsequential weak limit of $\mathcal{L}_{N,M}$ as $N \rightarrow \infty$ satisfies the κ -dimensional Ghirlanda–Guerra identities in Definition 2.2.

Proof. To simplify notation, let us assume $\mathcal{L}_{N,M}$ converges weakly as $N \rightarrow \infty$. Even if this convergence occurs only along a subsequence, the argument we give below can be restricted to that subsequence. Under the assumption that $\mathcal{L}_{N,M}$ converges, we must have that d^N converges to some $d \in \mathcal{D}$, since at volume N the self-overlap matrix $\mathcal{R}_{\ell,\ell} = N^{-1} \sigma^\ell (\sigma^\ell)^\top$ is equal to $\text{diag}(d^N)$ with probability one under $G_{N,M,\xi_\beta}^{\otimes \infty}$.

We divide the remainder of the proof into five steps. In the first four steps, we consider a fixed $\theta \in \Theta_{\mathbb{Q}}$.

Step 1. Isolate the role of θ in the Hamiltonian.

Similar to (2.19), we can realize H_{N,M,ξ_β} as a sum:

$$H_{N,M,\xi_\beta}(\sigma) = H_{N,M,\xi}(\sigma) + \sum_{\theta' \in \Theta_{\mathbb{Q}}} \beta_{\theta'} H_{N,M,\theta'}(\sigma), \quad (6.8)$$

where $H_{N,M,\xi}$ is as in (6.2), $H_{N,M,\theta'}$ is as in (6.3), and all Gaussian processes on the right-hand side of (6.8) are independent. We also consider the Hamiltonian without the contribution of $H_{N,M,\theta}$, as follows. Define $\beta^- = (\beta_{\theta'}^-)_{\theta' \in \Theta_{\mathbb{Q}}}$ by

$$\beta_{\theta'}^- := \begin{cases} \beta_{\theta'} & \text{if } \theta' \neq \theta \\ 0 & \text{if } \theta' = \theta. \end{cases}$$

We then have

$$H_{N,M,\xi_{\beta^-}}(\sigma) = H_{N,M,\xi}(\sigma) + \sum_{\substack{\theta' \in \Theta_{\mathbb{Q}} \\ \theta' \neq \theta}} \beta_{\theta'} H_{N,M,\theta'}(\sigma) = H_{N,M,\xi_\beta}(\sigma) - \beta_\theta H_{N,M,\theta}(\sigma). \quad (6.9)$$

From our definitions,

$$\begin{aligned} G_{N,M,\xi_\beta}(\sigma) &\stackrel{(6.1)}{=} \frac{\exp H_{N,M,\xi_\beta}(\sigma)}{Z_{N,M,\xi_\beta}(d^N)} \\ &\stackrel{(6.9)}{=} \frac{\exp(\beta_\theta H_{N,M,\theta}(\sigma))}{Z_{N,M,\xi_\beta}(d^N)/Z_{N,M,\xi_{\beta^-}}(d^N)} \cdot \frac{\exp H_{N,M,\xi_{\beta^-}}(\sigma)}{Z_{N,M,\xi_{\beta^-}}(d^N)} \\ &\stackrel{(6.1)}{=} \frac{\exp(\beta_\theta H_{N,M,\theta}(\sigma))}{Z_{N,M,\xi_\beta}(d^N)/Z_{N,M,\xi_{\beta^-}}(d^N)} G_{N,M,\xi_{\beta^-}}(\sigma). \end{aligned} \quad (6.10)$$

Note that $H_{N,M,\xi_{\beta^-}}(\sigma)$ is independent of $H_{N,M,\theta}$. Therefore, (6.10) expresses the measure of interest G_{N,M,ξ_β} as the Gibbs measure associated to $\beta_\theta H_{N,M,\theta}$ with respect to the independent reference measure $G_{N,M,\xi_{\beta^-}}$. We thus write

$$Z_{N,M,\theta}(\beta_\theta) := \frac{Z_{N,M,\xi_\beta}(d^N)}{Z_{N,M,\xi_{\beta^-}}(d^N)} \quad \text{and} \quad F_{N,M,\theta}(\beta_\theta) := \frac{1}{N} \mathbb{E} \log Z_{N,M,\theta}(\beta_\theta). \quad (6.11)$$

We think of these as functions of just β_θ , keeping $\beta_{\theta'}$ fixed for every $\theta' \neq \theta$.

Step 2. Show that $F_{N,M,\theta}(\beta_\theta)$ converges as $N \rightarrow \infty$.

Upon defining $\tilde{\xi}_\beta: \mathbb{R}^{\kappa \times \kappa} \rightarrow \mathbb{R}$ by

$$\tilde{\xi}_\beta(R) = \frac{N+M}{N} \xi_\beta\left(\frac{N}{N+M} R\right), \quad (6.12)$$

we can rewrite (6.2) for ξ_β as

$$\mathbb{E}[H_{N,M,\xi_\beta}(\sigma)H_{N,M,\xi_\beta}(\sigma')] = N\tilde{\xi}_\beta\left(\frac{\sigma\sigma'^T}{N}\right). \quad (6.13)$$

We can then use the notation of (1.21) to write

$$\frac{1}{N}\mathbb{E}\log Z_{N,M,\xi_\beta}(d^N) = F_{N,\tilde{\xi}_\beta}(d^N).$$

By Proposition 2.13(a), we have

$$|F_{N,\xi_\beta}(d^N) - F_{N,\tilde{\xi}_\beta}(d^N)| \leq \sup_{\|R\|_1 \leq 1} |\xi_\beta(R) - \tilde{\xi}_\beta(R)|.$$

Assuming $\|R\|_1 \leq 1$, we now recall (6.12) and apply the triangle inequality to obtain

$$\begin{aligned} |\xi_\beta(R) - \tilde{\xi}_\beta(R)| &\leq \left| \xi_\beta(R) - \xi_\beta\left(\frac{N}{N+M}R\right) \right| + \frac{M}{N} \left| \xi_\beta(R) - \xi_\beta\left(\frac{N}{N+M}R\right) \right| \\ &\leq \frac{M}{N+M} \sup_{\|Q\|_1 \leq 1} \|\nabla \xi_\beta(Q)\|_\infty + \frac{M}{N} \sup_{\|Q\|_1 \leq 1} |\xi_\beta(Q)|. \end{aligned}$$

Since M is fixed, we conclude from the two previous displays that

$$\lim_{N \rightarrow \infty} |F_{N,\xi_\beta}(d^N) - F_{N,\tilde{\xi}_\beta}(d^N)| = 0.$$

Hence Theorem 1.7 gives

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \log Z_{N,M,\xi_\beta}(d^N) = \inf_{\pi \in \Pi_d} \mathcal{P}_{\xi_\beta}(\pi).$$

By the same argument (just replacing β with β^- , which still satisfies (B1) and (B2)),

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \log Z_{N,M,\xi_{\beta^-}}(d^N) = \inf_{\pi \in \Pi_d} \mathcal{P}_{\xi_{\beta^-}}(\pi).$$

In light of definition (6.11), the two previous displays together yield

$$\lim_{N \rightarrow \infty} F_{N,M,\theta}(\beta_\theta) = \inf_{\pi \in \Pi_d} \mathcal{P}_{\xi_\beta}(\pi) - \inf_{\pi \in \Pi_d} \mathcal{P}_{\xi_{\beta^-}}(\pi). \quad (6.14)$$

Step 3. Show that $Z_{N,M,\theta}(\beta_\theta)$ concentrates.

By (6.13) we have

$$\mathbb{E}[H_{N,M,\xi_\beta}(\sigma)^2] = N\tilde{\xi}_\beta(\text{diag}(d^N)) \quad \text{for all } \sigma \in \Sigma^N(d^N).$$

Hence Gaussian concentration (see [54, Thm. 1.2]) implies that for all $x \geq 0$,

$$\mathbb{P}\left(\left|\frac{\log Z_{N,M,\xi_\beta}(d^N)}{N} - \frac{\mathbb{E} \log Z_{N,M,\xi_\beta}(d^N)}{N}\right| \geq x\right) \leq 2 \exp\left(\frac{-x^2 N}{4\tilde{\xi}_\beta(\text{diag}(d^N))}\right).$$

In particular, we have the following limit:

$$\lim_{N \rightarrow \infty} \mathbb{E}\left|\frac{\log Z_{N,M,\xi_\beta}(d^N)}{N} - \frac{\mathbb{E} \log Z_{N,M,\xi_\beta}(d^N)}{N}\right| = 0.$$

By the same argument (but replacing β with β^-),

$$\lim_{N \rightarrow \infty} \mathbb{E} \left| \frac{\log Z_{N,M,\xi_{\beta^-}}(d^N)}{N} - \frac{\mathbb{E} \log Z_{N,M,\xi_{\beta^-}}(d^N)}{N} \right| = 0.$$

In light of the definition (6.11), the two previous displays together yield

$$\lim_{N \rightarrow \infty} \mathbb{E} \left| \frac{\log Z_{N,M,\theta}(\beta_\theta)}{N} - \frac{\mathbb{E} \log Z_{N,M,\theta}(\beta_\theta)}{N} \right| = 0. \quad (6.15)$$

Step 4. Conclude that $H_{N,M,\theta}(\sigma)$ from (6.8) concentrates around $\mathbb{E}\langle H_{N,M,\theta}(\sigma) \rangle_N$, where $\langle \cdot \rangle_N$ denotes expectation with respect to $G_{N,M,\xi_\beta}^{\otimes \infty}$.

Since (B1) and (B2) remain true if β_θ is varied slightly, the limits (6.14) and (6.15) remain true for all choices of β_θ in some open interval. Furthermore, by Proposition 2.9, the right-hand side of (6.14) is differentiable with respect to β_θ . Therefore, by applying the result of [52] (see the remark after Thm. 1) to the representation (6.10), we have

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \left(\left| H_{N,M,\theta}(\sigma) - \mathbb{E}\langle H_{N,M,\theta}(\sigma) \rangle_N \right| \right)_N = 0. \quad (6.16)$$

Step 5. Conclude that the Ghirlanda–Guerra identities are satisfied in the large- N limit.

Given any positive integer n , consider any bounded measurable function f of the finite subarray $\mathcal{R}^{(n)} = (\mathcal{R}_{\ell,\ell'})_{1 \leq \ell, \ell' \leq n}$. With $\langle \cdot \rangle_N$ denoting expectation with respect to $G_{N,M,\xi_\beta}^{\otimes \infty}$, we have the following for any $\theta \in \Theta_{\mathbb{Q}}$:

$$\left| \mathbb{E} \left\langle f \cdot \left(H_{N,M,\theta}(\sigma^1) - \mathbb{E}\langle H_{N,M,\theta}(\sigma^1) \rangle_N \right) \right\rangle_N \right| \leq \|f\|_\infty \left(\left| H_{N,M,\theta}(\sigma^1) - \mathbb{E}\langle H_{N,M,\theta}(\sigma^1) \rangle_N \right| \right)_N.$$

The right-hand side is $o(N)$ by (6.16), and so the left-hand side is $o(N)$ as well:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \left| \mathbb{E} \left\langle f \cdot \left(H_{N,M,\theta}(\sigma^1) - \mathbb{E}\langle H_{N,M,\theta}(\sigma^1) \rangle_N \right) \right\rangle_N \right| = 0. \quad (6.17)$$

On the other hand, Gaussian integration by parts [54, Lem. 1.2] gives

$$\mathbb{E} \left\langle f \cdot H_{N,M,\theta}(\sigma^1) \right\rangle_N = \beta_\theta \mathbb{E} \left\langle f \cdot \left(\sum_{\ell=1}^n \mathcal{C}(\sigma^1, \sigma^\ell) - n \mathcal{C}(\sigma^1, \sigma^{n+1}) \right) \right\rangle_N, \quad (6.18)$$

where $\mathcal{C}: \Sigma^N(d^N) \times \Sigma^N(d^N) \rightarrow \mathbb{R}$ is given by

$$\mathcal{C}(\sigma, \sigma') = \mathbb{E}[H_{N,M,\theta}(\sigma) H_{N,M,\theta}(\sigma')] \stackrel{(6.4)}{=} (N+M) \left(\frac{N}{N+M} \right)^{\deg(\theta)} \xi_\theta \left(\frac{\sigma \sigma'^T}{N} \right).$$

Hence (6.18) can be rewritten as

$$\frac{1}{N} \mathbb{E} \left\langle f \cdot H_{N,M,\theta}(\sigma^1) \right\rangle_N = \beta_\theta \left(\frac{N}{N+M} \right)^{\deg(\theta)-1} \mathbb{E} \left\langle f \cdot \left(\sum_{\ell=1}^n \xi_\theta(\mathcal{R}_{1,\ell}) - n \xi_\theta(\mathcal{R}_{1,n+1}) \right) \right\rangle_N.$$

Since $\mathcal{R}_{1,1} = \text{diag}(d^N)$ with probability one under G_{N,M,ξ_θ} , we can further rewrite the right-hand side to obtain

$$\begin{aligned} \frac{1}{N} \mathbb{E} \left\langle f \cdot H_{N,M,\theta}(\sigma^1) \right\rangle_N &= \beta_\theta \left(\frac{N}{N+M} \right)^{\deg(\theta)-1} \left[\mathbb{E} \langle f \rangle_N \cdot \xi_\theta(\text{diag}(d^N)) \right. \\ &\quad \left. + \mathbb{E} \left\langle f \cdot \left(\sum_{\ell=2}^n \xi_\theta(\mathcal{R}_{1,\ell}) - n \xi_\theta(\mathcal{R}_{1,n+1}) \right) \right\rangle_N \right]. \end{aligned} \quad (6.19)$$

In the special case of the constant function $f \equiv 1$, we have

$$\begin{aligned} \frac{\mathbb{E} \langle H_{N,M,\theta}(\sigma^1) \rangle_N}{N} &= \beta_\theta \left(\frac{N}{N+M} \right)^{\deg(\theta)-1} \left[\xi_\theta(\text{diag}(d^N)) + \mathbb{E} \left\langle \sum_{\ell=2}^n \xi_\theta(\mathcal{R}_{1,\ell}) - n \xi_\theta(\mathcal{R}_{1,n+1}) \right\rangle_N \right] \\ &= \beta_\theta \left(\frac{N}{N+M} \right)^{\deg(\theta)-1} \left[\xi_\theta(\text{diag}(d^N)) - \mathbb{E} \langle \xi_\theta(\mathcal{R}_{1,2}) \rangle_N \right]. \end{aligned}$$

Upon multiplying this last equation by $\mathbb{E} \langle f \rangle_N$, we obtain

$$\begin{aligned} \frac{1}{N} \mathbb{E} \left\langle f \cdot \mathbb{E} \langle H_{N,M,\theta}(\sigma^1) \rangle_N \right\rangle_N \\ = \beta_\theta \left(\frac{N}{N+M} \right)^{\deg(\theta)-1} \left[\mathbb{E} \langle f \rangle_N \cdot \xi_\theta(\text{diag}(d^N)) - \mathbb{E} \langle f \rangle_N \cdot \mathbb{E} \langle \xi_\theta(\mathcal{R}_{1,2}) \rangle_N \right]. \end{aligned} \quad (6.20)$$

Subtracting (6.20) from (6.19) results in

$$\begin{aligned} \frac{1}{N} \mathbb{E} \left\langle f \cdot \left(H_{N,M,\theta}(\sigma^1) - \mathbb{E} \langle H_{N,M,\theta}(\sigma^1) \rangle_N \right) \right\rangle_N \\ = \beta_\theta \left(\frac{N}{N+M} \right)^{\deg(\theta)-1} \left[\mathbb{E} \left\langle f \cdot \left(\sum_{\ell=2}^n \xi_\theta(\mathcal{R}_{1,\ell}) - n \xi_\theta(\mathcal{R}_{1,n+1}) \right) \right\rangle_N + \mathbb{E} \langle f \rangle_N \cdot \mathbb{E} \langle \xi_\theta(\mathcal{R}_{1,2}) \rangle_N \right]. \end{aligned}$$

By (6.17), the left-hand side of this identity tends to 0 as $N \rightarrow \infty$. By our crucial assumption that $\beta_\theta \neq 0$, it follows that

$$\lim_{N \rightarrow \infty} \left| \mathbb{E} \langle f \cdot \xi_\theta(\mathcal{R}_{1,n+1}) \rangle_N - \frac{1}{n} \mathbb{E} \langle f \rangle_N \cdot \mathbb{E} \langle \xi_\theta(\mathcal{R}_{1,2}) \rangle_N - \frac{1}{n} \sum_{\ell=2}^n \mathbb{E} \langle f \cdot \xi_\theta(\mathcal{R}_{1,\ell}) \rangle_N \right| = 0.$$

Now let \mathcal{L} denote the weak limit of $\mathcal{L}_{N,M}$ as $N \rightarrow \infty$. The previous display means that when the array \mathcal{R} is distributed according to \mathcal{L} , we have

$$\mathbb{E} [f(\mathcal{R}^{(n)}) \cdot \xi_\theta(\mathcal{R}_{1,n+1})] = \frac{1}{n} \mathbb{E} [f(\mathcal{R}^n)] \cdot \mathbb{E} \xi_\theta(\mathcal{R}_{1,2}) + \frac{1}{n} \sum_{\ell=2}^n \mathbb{E} [f(\mathcal{R}^n) \cdot \xi_\theta(\mathcal{R}_{1,\ell})]. \quad (6.21)$$

This is a special case of the G.G. identity (2.9). All that remains is to argue that this special case implies the general case.

Recall that $\xi_\theta(R) = \prod_{j=1}^m \langle R^{\circ p} w_j, w_j \rangle^{n_j}$, where $p, m, n_1, \dots, n_m \geq 1$ and $w_1, \dots, w_m \in [-1, 1]^\kappa$ are the parameters defining θ . Since $(\mathbb{Q} \cap [-1, 1])^\kappa$ is dense in $[-1, 1]^\kappa$, and (6.21) holds for every $\theta \in \Theta_{\mathbb{Q}}$, it follows that (6.21) holds for every $\theta \in \Theta$. That is, (2.9) holds whenever the function φ in (2.8) is a polynomial of the form $\varphi(x_1, \dots, x_m) = \prod_{j=1}^m x_j^{n_j}$. By approximating continuous functions with linear

combinations of such polynomials (together with constant functions, for which (2.9) is trivial), we deduce the same statement for any bounded continuous $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$. Finally, by approximating bounded measurable functions with bounded continuous functions (e.g. using Lusin's theorem [65, Thm. 2.24]), we obtain the identity (2.9) for any bounded measurable φ . Note that these approximations are over compact domains, since every overlap matrix $\mathcal{R}_{\ell, \ell'}$ almost surely belongs to the compact set \mathcal{R}_κ from (1.19). \square

Before we can make use of Proposition 6.5 in the next section, we need one more basic fact about the G.G. identities.

Lemma 6.6. *Assume $\mathcal{L} = \text{Law}(\mathcal{R})$ is a Gram–de Finetti law that satisfies the κ -dimensional Ghirlanda–Guerra identities in Definition 2.2. Then the scalar array $\mathcal{Q} = (\mathcal{Q}_{\ell, \ell'})_{\ell, \ell' \geq 1}$ given by $\mathcal{Q}_{\ell, \ell'} = \text{tr}(\mathcal{R}_{\ell, \ell'})$ satisfies the 1-dimensional Ghirlanda–Guerra identities. That is, for any bounded measurable function f of the finite subarray $\mathcal{Q}^{(n)} = (\mathcal{Q}_{\ell, \ell'})_{1 \leq \ell, \ell' \leq n}$, and any bounded measurable $\psi: \mathbb{R} \rightarrow \mathbb{R}$, we have*

$$\mathbb{E}[f(\mathcal{Q}^{(n)})\psi(\mathcal{Q}_{1, n+1})] = \frac{1}{n}\mathbb{E}[f(\mathcal{Q}^{(n)})]\mathbb{E}[\psi(\mathcal{Q}_{1, 2})] + \frac{1}{n}\sum_{\ell=2}^n \mathbb{E}[f(\mathcal{Q}^{(n)})\psi(\mathcal{Q}_{1, \ell})]. \quad (6.22)$$

Proof. Define $\varphi: \mathbb{R}^\kappa \rightarrow \mathbb{R}$ by $\varphi(x_1, \dots, x_\kappa) = \psi(x_1 + \dots + x_\kappa)$. We then have

$$\psi(\text{tr}(R)) = \varphi(\langle R\mathbf{e}_1, \mathbf{e}_1 \rangle, \dots, \langle R\mathbf{e}_\kappa, \mathbf{e}_\kappa \rangle),$$

and so (6.22) is a special case of (2.9). \square

6.3. Proof of lower bound from symmetry. From here to the end of Sect. 6, we always assume $d^N = d_{\text{bal}}$. Recall that $\Sigma^N(d_{\text{bal}})$ is nonempty if and only if N is a multiple of κ , and so we will frequently replace N with κN so that N continues to be a generic positive integer. For instance, we will write $G_{\kappa N, M, \xi}$ as in (6.1), but now with the understanding that $d^{\kappa N} = d_{\text{bal}}$.

Lemma 6.7. *Assume ξ satisfies (A1) and (A3). Then the following statements hold.*

(a) *For any permutation $\omega \in S_\kappa$, we have*

$$(H_{N, M, \xi}(\sigma))_{\sigma \in \Sigma^N} \stackrel{\text{law}}{=} (H_{N, M, \xi}(\omega \bullet \sigma))_{\sigma \in \Sigma^N}.$$

(b) *If $d^{\kappa N} = d_{\text{bal}}$, then $\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N, M, \xi}^{\otimes \infty}))$ is symmetric in the sense of Definition 2.5.*

Proof. For $k \in \{1, \dots, \kappa\}$ and $\sigma = (\sigma_1, \dots, \sigma_N) \in \Sigma^N$, we isolate the k th coordinates in σ by writing $\sigma(k) = (\sigma_1(k), \dots, \sigma_N(k))^T \in \mathbb{R}^N$. Recall the overlap map $R(\sigma, \sigma') = N^{-1}\sigma\sigma'^T$ from (2.2). That is, the (k, k') entry of $R(\sigma, \sigma')$ is given by an inner product:

$$R(\sigma, \sigma')_{k, k'} = N^{-1}\langle \sigma(k), \sigma'(k') \rangle. \quad (6.23)$$

For any $\omega \in S_\kappa$ and $\sigma, \sigma' \in \Sigma^N$, we claim that

$$\omega \bullet R(\sigma, \sigma') = R(\omega \bullet \sigma, \omega \bullet \sigma'). \quad (6.24)$$

The action by ω on the left-hand side of (6.24) is defined in (1.23), whereas the action on the right-hand side is defined in (1.25). For any $k, k' \in \{1, \dots, \kappa\}$, we have

$$\begin{aligned}
 (k, k') \text{ entry of } \omega \bullet R(\sigma, \sigma') &\stackrel{(1.23)}{=} (\omega^{-1}(k), \omega^{-1}(k')) \text{ entry of } R(\sigma, \sigma') \\
 &\stackrel{(6.23)}{=} \langle \sigma(\omega^{-1}(k)), \sigma'(\omega^{-1}(k')) \rangle \\
 &\stackrel{(1.25)}{=} \langle [\omega \bullet \sigma](k), [\omega \bullet \sigma'](k') \rangle \\
 &\stackrel{(6.23)}{=} (k, k') \text{ entry of } R(\omega \bullet \sigma, \omega \bullet \sigma').
 \end{aligned}$$

This verifies (6.24). We can now write

$$\begin{aligned}
 \mathbb{E}[H_{N,M,\xi}(\omega \bullet \sigma) H_{N,M,\xi}(\omega \bullet \sigma)] &\stackrel{(6.2)}{=} (N+M)\xi \left(\frac{N}{\kappa N + M} R(\omega \bullet \sigma, \omega \bullet \sigma') \right) \\
 &\stackrel{(6.24)}{=} (N+M)\xi \left(\frac{N}{N+M} (\omega \bullet R(\sigma, \sigma')) \right) \\
 &\stackrel{(A3)}{=} (N+M)\xi \left(\frac{N}{N+M} R(\sigma, \sigma') \right) \\
 &\stackrel{(6.2)}{=} \mathbb{E}[H_{N,M,\xi}(\sigma) H_{N,M,\xi}(\sigma)].
 \end{aligned}$$

This proves part (a).

For part (b), consider the following Gibbs measure:

$$G_{\kappa N, M, \xi}^{\omega}(\sigma) := \frac{\exp H_{\kappa N, M, \xi}(\omega^{-1} \bullet \sigma)}{\sum_{\sigma' \in \Sigma^{\kappa N}(d_{\text{bal}})} \exp H_{\kappa N, M, \xi}(\omega^{-1} \bullet \sigma')}, \quad \sigma \in \Sigma^{\kappa N}(d_{\text{bal}}). \quad (6.25)$$

Part (a) implies $\mathbb{E}(G_{\kappa N, M, \xi}^{\otimes \infty}) = \mathbb{E}((G_{\kappa N, M, \xi}^{\omega})^{\otimes \infty})$. In particular,

$$\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N, M, \xi}^{\otimes \infty})) = \text{Law}(\mathcal{R}; \mathbb{E}((G_{\kappa N, M, \xi}^{\omega})^{\otimes \infty})). \quad (6.26)$$

But notice that $\sigma' \mapsto \omega^{-1} \bullet \sigma'$ is a bijection on $\Sigma^{\kappa N}(d_{\text{bal}})$ since $\omega \bullet \text{diag}(d_{\text{bal}}) = \text{diag}(d_{\text{bal}})$. Therefore, the denominator in (6.25) can be rewritten to give

$$G_{\kappa N, M, \xi}^{\omega}(\sigma) = \frac{\exp H_{\kappa N, M, \xi}(\omega^{-1} \bullet \sigma)}{\sum_{\sigma' \in \Sigma^{\kappa N}(d_{\text{bal}})} \exp H_{\kappa N, M, \xi}(\sigma')}.$$

Drawing from $G_{\kappa N, M, \xi}^{\omega}$ is thus equivalent to sampling from $G_{\kappa N, M, \xi}$ and then applying ω to the sample. Consequently,

$$\text{Law}(\mathcal{R}; \mathbb{E}((G_{\kappa N, M, \xi}^{\omega})^{\otimes \infty})) = \text{Law}(\mathcal{R}^{\omega}; \mathbb{E}(G_{\kappa N, M, \xi}^{\otimes \infty})), \quad (6.27)$$

where $\mathcal{R}^{\omega} = (\mathcal{R}_{\ell, \ell'}^{\omega})_{\ell, \ell' \geq 1}$ is the array given by $\mathcal{R}_{\ell, \ell'}^{\omega} = R(\omega \bullet \sigma^{\ell}, \omega \bullet \sigma^{\ell'})$. On the other hand, (6.24) says that $\mathcal{R}^{\omega} = \omega \bullet \mathcal{R}$, hence

$$\text{Law}(\mathcal{R}^{\omega}; \mathbb{E}(G_{\kappa N, M, \xi}^{\otimes \infty})) = \text{Law}(\omega \bullet \mathcal{R}; \mathbb{E}(G_{\kappa N, M, \xi}^{\otimes \infty})). \quad (6.28)$$

Chaining together (6.26)–(6.28) yields $\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N, M, \xi}^{\otimes \infty})) = \text{Law}(\omega \bullet \mathcal{R}; \mathbb{E}(G_{\kappa N, M, \xi}^{\otimes \infty}))$. \square

We are now ready to complete the main objective of this section.

Proof of Proposition 2.11. It is an elementary fact that for any real-valued sequence $(a_N)_{N \geq 1}$ and an integer $M \geq 1$,

$$\liminf_{N \rightarrow \infty} \frac{a_N}{N} \geq \frac{1}{M} \liminf_{N \rightarrow \infty} (a_{N+M} - a_N).$$

Applying this observation to $a_N = \kappa^{-1} \mathbb{E} \log Z_{\kappa N, \xi_\beta}(d_{\text{bal}})$ and replacing M by κM results in

$$\liminf_{N \rightarrow \infty} F_{\kappa N, \xi_\beta}(d_{\text{bal}}) \geq \frac{1}{\kappa M} \liminf_{N \rightarrow \infty} \mathbb{E} \log \frac{Z_{\kappa(N+M), \xi_\beta}(d_{\text{bal}})}{Z_{\kappa N, \xi_\beta}(d_{\text{bal}})}.$$

Now apply Proposition 6.2 with $d^{\kappa N} = d^{\kappa(N+M)} = \delta^{\kappa M} = d_{\text{bal}}$, so that (6.7) yields

$$\liminf_{N \rightarrow \infty} F_{\kappa N, \xi_\beta}(d_{\text{bal}}) \geq \liminf_{N \rightarrow \infty} \Psi_{\kappa M, \xi_\beta} \left(\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N, \kappa M, \xi_\beta}^{\otimes \infty})); \Sigma^{\kappa M}(d_{\text{bal}}) \right). \quad (6.29)$$

Note that because $G_{\kappa N, \kappa M, \xi_\beta}$ is supported on $\Sigma^{\kappa N_k}(d_{\text{bal}})$, we trivially have

$$\mathcal{R}_{\ell, \ell'} = R(\sigma^\ell, \sigma^{\ell'}) = \text{diag}(d_{\text{bal}}) \quad \text{for all } \ell, \ell' \geq 1, \text{ with probability one.} \quad (6.30)$$

By Lemma 6.4, we know

$$\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N, \kappa M, \xi_\beta}^{\otimes \infty})) \text{ is Gram-de Finetti (Definition 2.1).} \quad (6.31)$$

By Lemma 2.10, ξ_β satisfies the symmetry condition (A3). So by Lemma 6.7(b), we know

$$\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N, \kappa M, \xi_\beta}^{\otimes \infty})) \text{ is symmetric (Definition 2.5).} \quad (6.32)$$

Now pass to a subsequence $(N_j)_{j \geq 1}$ that achieves the infimum on the right-hand side of (6.29):

$$\liminf_{N \rightarrow \infty} F_{\kappa N, \xi_\beta}(d_{\text{bal}}) \geq \lim_{j \rightarrow \infty} \Psi_{\kappa M, \xi_\beta} \left(\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N_j, \kappa M, \xi_\beta}^{\otimes \infty})); \Sigma^{\kappa M}(d_{\text{bal}}) \right). \quad (6.33)$$

By passing to a further subsequence, we may also assume there is some Gram-de Finetti law \mathcal{L}_M such that

$$\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N_j, \kappa M, \xi_\beta}^{\otimes \infty})) \text{ converges weakly to } \mathcal{L}_M \text{ as } j \rightarrow \infty. \quad (6.34)$$

Let us check that \mathcal{L}_M satisfies the hypotheses of Lemma 2.6. First, (6.30) obviously remains true under \mathcal{L}_M . Second, \mathcal{L}_M must be a Gram-de Finetti law by (6.31), because $\mathcal{R} \mapsto (\langle \mathcal{R}_{\ell, \ell'}^{\circ p} w, w \rangle)_{\ell, \ell' \geq 1}$ is a continuous operation on arrays for any $p \geq 1$ and $w \in \mathbb{R}^\kappa$. Similarly, \mathcal{L}_M inherits symmetry from (6.32), since $\mathcal{R} \mapsto \omega \bullet \mathcal{R}$ is a continuous operation on arrays for any $\omega \in S_\kappa$. Finally, according to Proposition 6.5, the κ -dimensional G.G. identities are satisfied by \mathcal{L}_M .

We can now invoke Lemma 2.6: if $\mathcal{L}_M = \text{Law}(\mathcal{R})$, then almost surely

$$R_{\ell, \ell'} = \Phi^*(\text{tr}(R_{\ell, \ell'})), \quad \text{where} \quad \Phi^*(q) = \frac{q}{\kappa} \mathbf{I}_\kappa + \frac{1-q}{\kappa^2} \mathbf{1}\mathbf{1}^\top. \quad (6.35)$$

In other words, if we denote the array of traces by $\mathcal{Q} = (\mathcal{Q}_{\ell, \ell'})_{\ell, \ell' \geq 1} = (\text{tr}(\mathcal{R}_{\ell, \ell'}))_{\ell, \ell' \geq 1}$ and define $\tilde{\mathcal{L}}_M = \text{Law}(\mathcal{Q})$, then \mathcal{L}_M is the pushforward¹ of $\tilde{\mathcal{L}}_M$ under Φ^* :

$$\mathcal{L}_M = \tilde{\mathcal{L}}_M \circ (\Phi^*)^{-1}. \quad (6.36)$$

Furthermore, $\tilde{\mathcal{L}}_M$ satisfies the 1-dimensional G.G. identities by Lemma 6.6. Therefore, Theorem 2.4 says that $\tilde{\mathcal{L}}_M = \tilde{\mathcal{L}}_{\mu_M}$, where $\mu_M = \text{Law}(\text{tr}(\mathcal{R}_{1,2}))$. This is a probability measure on $[0, 1]$, and we consider its quantile function

$$Q_{\mu_M}(t) := \inf\{q \geq 0 : \mu([0, q]) \geq t\}, \quad t \in (0, 1].$$

Then define $\pi_M : (0, 1] \rightarrow \Gamma^*$ by $\pi_M(t) = \Phi^*(Q_{\mu_M}(t))$. Since Q_{μ_M} is left-continuous and Φ^* is continuous, this π_M is an element of the path space Π^* defined in (1.12).

Now let $(\mu_{M,j})_{j \geq 1}$ be a sequence of probability measures on $[0, 1]$ such that

- (i) $\mu_{M,j}$ is supported on finitely many points.
- (ii) $\mu_{M,j}(\{1\}) > 0$.
- (iii) $\mu_{M,j}$ converges weakly to μ_M as $j \rightarrow \infty$.

Then define $\pi_{M,j} : (0, 1] \rightarrow \Gamma^*$ by $\pi_{M,j}(t) = \Phi^*(Q_{\mu_{M,j}}(t))$. By properties (i) and (ii), this $\pi_{M,j}$ is an element of the path space $\Pi_{d_{\text{bal}}}^{\text{disc}}$ from (1.28). Meanwhile, property (iii) implies

$$\lim_{j \rightarrow \infty} \int_0^1 |Q_{\mu_M}(t) - Q_{\mu_{M,j}}(t)| dt = 0. \quad (6.37a)$$

As Φ^* is Lipschitz and $\pi_M(t) - \pi_{M,j}(t) = \Phi^*(Q_{\mu_M}(t)) - \Phi^*(Q_{\mu_{M,j}}(t))$, this limit implies

$$\lim_{j \rightarrow \infty} \int_0^1 \|\pi_M(t) - \pi_{M,j}(t)\|_1 dt = 0. \quad (6.37b)$$

From this convergence and Lemma 5.1, it follows that

$$\begin{aligned} \lim_{j \rightarrow \infty} \mathcal{P}_{\kappa M, \xi\beta}^{(1)}(\pi_{M,j}, 0; \Sigma^{\kappa M}(d_{\text{bal}})) &= \mathcal{P}_{\kappa M, \xi\beta}(\pi_M, 0; \Sigma^{\kappa M}(d_{\text{bal}})), \\ \text{and } \lim_{j \rightarrow \infty} \mathcal{P}_{\xi\beta}^{(2)}(\pi_{M,j}) &= \mathcal{P}_{\xi\beta}^{(2)}(\pi_M). \end{aligned} \quad (6.38)$$

On the other hand, we claim the following equality:

$$\mathcal{P}_{\kappa M, \xi\beta}^{(1)}(\pi_{M,j}, 0; \Sigma^{\kappa M}(d_{\text{bal}})) + \mathcal{P}_{\xi\beta}^{(2)}(\pi_{M,j}) = \Psi_{\kappa M, \xi\beta}(\tilde{\mathcal{L}}_{\mu_{M,j}} \circ (\Phi^*)^{-1}; \Sigma^{\kappa M}(d_{\text{bal}})), \quad (6.39)$$

where $\tilde{\mathcal{L}}_{\mu_{M,j}}$ is the 1-dimensional Gram-de Finetti law from Theorem 2.4. Indeed, suppose $\pi_{M,j}$ has the form

$$\pi_{M,j}(t) = \gamma_r \quad \text{for } t \in (m_{r-1}, m_r], r \in \{1, \dots, s\},$$

¹ Here there is a slight abuse of notation: in (6.35), Φ^* is a map $[0, 1] \rightarrow \mathbb{R}^{\kappa \times \kappa}$, whereas in (6.36), Φ^* is thought of as a map $[0, 1]^{\mathbb{N} \times \mathbb{N}} \rightarrow (\mathbb{R}^{\kappa \times \kappa})^{\mathbb{N} \times \mathbb{N}}$ defined by performing (6.35) to every element in an array.

where $0 = m_0 < m_1 < \dots < m_s = 1$ and $0 \leq \gamma_1 < \dots < \gamma_s = \text{diag}(d_{\text{bal}})$. This means $\mu_{M,j}$ has the form

$$\mu_{M,j} = \sum_{r=1}^s (m_r - m_{r-1}) \delta_{q_r}, \quad \text{where } \Phi^*(q_r) = \gamma_r.$$

Now let ν be the RPC from Theorem 3.4, so that when $(\alpha^1, \alpha^2, \dots)$ is sampled from $\mathbb{E}(\nu^{\otimes \infty})$, the induced array $(q_{r(\alpha^\ell, \alpha^{\ell'})})_{\ell, \ell' \geq 1}$ has law $\tilde{\mathcal{L}}_{\mu_{M,j}}$. Applying Φ^* to every entry, we obtain an array of matrices $(\gamma_{r(\alpha^\ell, \alpha^{\ell'})})_{\ell, \ell' \geq 1}$ whose law is $\tilde{\mathcal{L}}_{\mu_{M,j}} \circ (\Phi^*)^{-1}$. Then (6.39) follows from Lemma 3.5.

By Property (iii) and the final statement in Theorem 2.4, $\tilde{\mathcal{L}}_{\mu_{M,j}}$ converges weakly to $\tilde{\mathcal{L}}_{\mu_M}$ as $j \rightarrow \infty$. Since Φ^* is continuous, it follows that

$$\tilde{\mathcal{L}}_{\mu_{M,j}} \circ (\Phi^*)^{-1} \text{ converges weakly to } \tilde{\mathcal{L}}_{\mu_M} \circ (\Phi^*)^{-1} \stackrel{(6.36)}{=} \mathcal{L}_M \text{ as } j \rightarrow \infty. \quad (6.40)$$

Since (6.40) and (6.34) have the same limit, Corollary 3.3 implies

$$\begin{aligned} \lim_{j \rightarrow \infty} \Psi_{\kappa M, \xi_\beta} \left(\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N_j, \kappa M, \xi_\beta}^{\otimes \infty}); \Sigma^{\kappa M}(d_{\text{bal}})) \right) \\ = \lim_{j \rightarrow \infty} \Psi_{\kappa M, \xi_\beta} (\tilde{\mathcal{L}}_{\mu_{M,j}} \circ (\Phi^*)^{-1}; \Sigma^{\kappa M}(d_{\text{bal}})). \end{aligned} \quad (6.41)$$

Now we put our various observations together:

$$\begin{aligned} \liminf_{N \rightarrow \infty} F_{\kappa N, \xi_\beta}(d_{\text{bal}}) &\stackrel{(6.33)}{\geq} \lim_{j \rightarrow \infty} \Psi_{\kappa M, \xi_\beta} \left(\text{Law}(\mathcal{R}; \mathbb{E}(G_{\kappa N_j, \kappa M, \xi_\beta}^{\otimes \infty}); \Sigma^{\kappa M}(d_{\text{bal}})) \right) \\ &\stackrel{(6.41)}{=} \lim_{j \rightarrow \infty} \Psi_{\kappa M, \xi_\beta} (\tilde{\mathcal{L}}_{\mu_{M,j}} \circ (\Phi^*)^{-1}; \Sigma^{\kappa M}(d_{\text{bal}})) \\ &\stackrel{(6.39)}{=} \lim_{j \rightarrow \infty} [\mathcal{P}_{\kappa M, \xi_\beta}^{(1)}(\pi_{M,j}, 0; \Sigma^{\kappa M}(d_{\text{bal}})) + \mathcal{P}_{\xi_\beta}^{(2)}(\pi_{M,j})] \\ &\stackrel{(6.38)}{=} \mathcal{P}_{\kappa M, \xi_\beta}(\pi_M, 0; \Sigma^{\kappa M}(d_{\text{bal}})) + \mathcal{P}_{\xi_\beta}^{(2)}(\pi_M). \end{aligned}$$

Finally, we must send $M \rightarrow \infty$.

By passing to a subsequence, we may assume μ_M converges weakly as $M \rightarrow \infty$ to some probability measure μ on $[0, 1]$. Define $\pi : (0, 1] \rightarrow \Gamma^*$ by $\pi(t) = \Phi^*(Q_\mu(t))$. As before, π belongs to Π^* because of the left-continuity of Q_μ together with the continuity of Φ^* . Using the same logic as in (6.37), we must have $\pi_M \rightarrow \pi$ in the L^1 norm (1.32). Hence Lemmas 5.3 and 5.4 give

$$\lim_{M \rightarrow \infty} \mathcal{P}_{\kappa M, \xi_\beta}^{(1)}(\pi_M, 0; \Sigma^{\kappa M}(d_{\text{bal}})) \stackrel{(5.19)}{=} \inf_{\lambda \in \mathbb{R}^\kappa} [\mathcal{P}_{\xi_\beta}^{(1)}(\pi, \lambda) - \langle \lambda, d_{\text{bal}} \rangle] \stackrel{(5.21)}{=} \mathcal{P}_{\xi_\beta}^{(1)}(\pi, 0).$$

Applying Lemma 5.1 once more, we also have

$$\lim_{M \rightarrow \infty} \mathcal{P}_{\xi_\beta}^{(2)}(\pi_M) = \mathcal{P}_{\xi_\beta}^{(2)}(\pi).$$

Combining the three previous displays, we obtain

$$\liminf_{N \rightarrow \infty} F_{\kappa N, \xi_\beta}(d_{\text{bal}}) \geq \mathcal{P}_{\xi_\beta}^{(1)}(\pi, 0) + \mathcal{P}_{\xi_\beta}^{(2)}(\pi) \stackrel{(1.34)}{=} \mathcal{P}_{\xi_\beta}(\pi, 0).$$

□

Acknowledgement We are grateful to Hong-Bin Chen, Amir Dembo, Justin Ko, and Nike Sun for helpful discussions. We thank the referee for their careful reading and helpful suggestions.

Funding E.B. was supported by National Science Foundation (NSF) grant DMS-2246616/2412473. Y.S. was supported by Simons–NSF Collaboration on Deep Learning grant DMS-2031883 and Vannevar Bush Faculty Fellowship award ONR-N00014-20-1-2826.

Data No data were involved in this research.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Appendix A. Gaussian Processes for Generic Model

Here we prove Proposition 2.7. We fix $\theta = (p, m, n_1, \dots, n_m, w_1, \dots, w_m) \in \Theta$ throughout this appendix and denote the coordinates of $w_j \in [-1, 1]^\kappa$ by $w_j = (w_j(k))_{k=1}^\kappa$. We first establish two lemmas.

Lemma A.1. *For $Q, R \in \Gamma_\kappa$ such that $Q \preceq R$, we have*

$$0 \leq \xi_\theta(Q) \leq \xi_\theta(R), \quad 0 \leq \nabla \xi_\theta(Q) \leq \nabla \xi_\theta(R).$$

Furthermore, $\vartheta_\theta(R) = (\deg(\theta) - 1)\xi_\theta(R)$ for any $R \in \mathbb{R}^{\kappa \times \kappa}$.

Proof. First we make an elementary claim about Hadamard products:

$$0 \leq Q \leq R, \quad 0 \leq Q' \leq R' \quad \implies \quad 0 \leq Q \circ Q' \leq R \circ R'. \quad (\text{A.1})$$

Indeed, we have the decomposition

$$R \circ R' - Q \circ Q' = R \circ (R' - Q') + Q' \circ (R - Q),$$

and the Schur product theorem tells us that both $R \circ (R' - Q')$ and $Q' \circ (R - Q)$ are positive-semidefinite. Hence (A.1) holds.

By repeatedly applying (A.1) with $Q' = Q$ and $R' = R$, we see that

$$0 \leq Q \leq R \quad \implies \quad 0 \leq \langle Q^{\circ p} u, u \rangle \leq \langle R^{\circ p} u, u \rangle \quad \text{for any } p \geq 1 \text{ and } u \in \mathbb{R}^\kappa.$$

In light of the definition of ξ_θ in (1.17), it follows that $0 \leq \xi_\theta(Q) \leq \xi_\theta(R)$.

Next we argue $0 \leq \nabla \xi_\theta(Q) \leq \nabla \xi_\theta(R)$. By differentiating (1.17) with respect to $R_{k,k'}$, we compute the (k, k') entry of $\nabla \xi_\theta(R)$ to be

$$\nabla \xi_\theta(R)_{k,k'} = \sum_{j=1}^m \left[\prod_{\ell \neq j} \langle R^{\circ p} w_\ell, w_\ell \rangle^{n_\ell} \right] n_j \langle R^{\circ p} w_j, w_j \rangle^{n_j-1} \cdot p R_{k,k'}^{p-1} w_j(k) w_j(k'). \quad (\text{A.2})$$

If we write $W_j = \text{diag}(w_j)$, then (A.2) says

$$\nabla \xi_\theta(R) = p \sum_{j=1}^m n_j \xi_{\theta_j}(R) \cdot W_j R^{\circ(p-1)} W_j, \quad (\text{A.3})$$

where $\theta_j \in \Theta$ is obtained from $\theta = (p, m, n_1, \dots, n_m, w_1, \dots, w_m)$ by modifying n_j to $n_j - 1$. For each j we have $0 \leq \xi_{\theta_j}(Q) \leq \xi_{\theta_j}(R)$ by the argument of the previous paragraph. In addition, (A.1) gives $0 \leq Q^{\circ(p-1)} \leq R^{\circ(p-1)}$, and so $0 \leq WQ^{\circ(p-1)}W \leq WR^{\circ(p-1)}W$ for any diagonal matrix W . It is thus apparent from (A.3) that $0 \leq \nabla \xi_\theta(Q) \leq \nabla \xi_\theta(R)$.

To see the final assertion of the lemma, insert (A.2) into the definition of $\vartheta_\theta(R) = \langle R, \nabla \xi_\theta(R) \rangle - \xi_\theta(R)$. This results in

$$\begin{aligned} \vartheta_\theta(R) &= p \sum_{j=1}^m n_j \left[\prod_{\ell \neq j} \langle R^{\circ p} w_\ell, w_\ell \rangle^{n_\ell} \right] \langle R^{\circ p} w_j, w_j \rangle^{n_j} \\ &\quad - \xi_\theta(R) \stackrel{(1.17)}{=} p \sum_{j=1}^m n_j \xi_\theta(R) - \xi_\theta(R), \end{aligned}$$

which is exactly as desired. \square

In the next lemma, we follow the construction in [57, Sec. 5]. Our $H_{N,\theta}$ is equal to $\sqrt{N} h_{N,\theta}$ in the notation of [57].

Lemma A.2. *For each $N \geq 1$, there exist a centered Gaussian process $(H_{N,\theta}(\sigma))_{\sigma \in \mathbb{R}^{\kappa \times N}}$ with covariance*

$$\mathbb{E}[H_{N,\theta}(\sigma)H_{N,\theta}(\sigma')] = N\xi_\theta\left(\frac{\sigma\sigma'^T}{N}\right), \quad (\text{A.4})$$

and a centered κ -dimensional Gaussian process $(Z_{N,\theta}(\sigma))_{\sigma \in \mathbb{R}^{\kappa \times N}}$ with covariance

$$\mathbb{E}\left[Z_{N,\theta}(\sigma)Z_{N,\theta}(\sigma')^T\right] = \nabla \xi_\theta\left(\frac{\sigma\sigma'^T}{N}\right). \quad (\text{A.5})$$

Proof. For any p -tuple of indices $I = (i_1, \dots, i_p) \in \{1, \dots, N\}^p$, let us write

$$\sigma_I = (\sigma_{i_1}, \dots, \sigma_{i_p}) \in (\mathbb{R}^\kappa)^p \quad \text{for } \sigma \in (\mathbb{R}^\kappa)^N.$$

Furthermore, for any n -tuple of p -tuples $\mathcal{I} = (I_1, \dots, I_n) \in (\{1, \dots, N\}^p)^n$, we will write

$$\sigma_{\mathcal{I}} = (\sigma_{I_1}, \dots, \sigma_{I_n}) \in ((\mathbb{R}^\kappa)^p)^n \quad \text{for } \sigma \in (\mathbb{R}^\kappa)^N.$$

Given such \mathcal{I} and some $w = (w(k))_{1 \leq k \leq \kappa} \in \mathbb{R}^\kappa$, define the following polynomial:

$$S_w(\sigma_{\mathcal{I}}) := S_w(\sigma_{I_1}) \dots S_w(\sigma_{I_n}), \quad \text{where} \quad S_w(\sigma_{i_1}, \dots, \sigma_{i_p}) := \sum_{k=1}^{\kappa} w(k) \sigma_{i_1}(k) \dots \sigma_{i_p}(k).$$

Finally, we sum over all choices of $\mathcal{I}_1, \dots, \mathcal{I}_m$ with $\mathcal{I}_j \in (\{1, \dots, N\}^p)^{n_j}$:

$$H_{N,\theta}(\sigma) := \frac{1}{N^{(\deg(\theta)-1)/2}} \sum_{\mathcal{I}_1, \dots, \mathcal{I}_m} g_{\mathcal{I}_1, \dots, \mathcal{I}_m} S_{w_1}(\sigma_{\mathcal{I}_1}) \dots S_{w_m}(\sigma_{\mathcal{I}_m}),$$

where each $g_{\mathcal{I}_1, \dots, \mathcal{I}_m}$ is an independent standard normal random variable. Recall that $\deg(\theta) = p \sum_{j=1}^m n_j$. The covariance of the Gaussian process $(H_{N, \theta}(\sigma))_{\sigma \in \mathbb{R}^{\kappa \times N}}$ is then

$$\begin{aligned} \mathbb{E}[H_{N, \theta}(\sigma) H_{N, \theta}(\sigma')] &= N \prod_{j=1}^m \left[\frac{1}{N^{pn_j}} \sum_{\mathcal{I}_j} S_{w_j}(\sigma_{\mathcal{I}_j}) S_{w_j}(\sigma'_{\mathcal{I}_j}) \right] \\ &= N \prod_{j=1}^m \left(\frac{1}{N^p} \sum_{I \in \{1, \dots, N\}^p} S_{w_j}(\sigma_I) S_{w_j}(\sigma'_I) \right)^{n_j}. \end{aligned}$$

If we write R for the matrix $R(\sigma, \sigma') = \sigma \sigma'^T / N$, then

$$\begin{aligned} \frac{1}{N^p} \sum_{I \in \{1, \dots, N\}^p} S_{w_j}(\sigma_I) S_{w_j}(\sigma'_I) &= \sum_{k, k'=1}^{\kappa} w_j(k) w_j(k') \left(\frac{1}{N} \sum_{i=1}^N \sigma_i(k) \sigma'_i(k') \right)^p \\ &= \langle R^{\circ p} w_j, w_j \rangle. \end{aligned}$$

By combining the two previous displays and recalling the definition of ξ_θ from (1.17), we obtain (A.4).

To prove the existence of the Gaussian process $Z_{N, \theta}$, we first claim that for any $p \geq 1$, there exists a κ -dimensional centered Gaussian process $(z_{N, p}(\sigma))_{\sigma \in \mathbb{R}^{\kappa \times N}}$ with covariance

$$\mathbb{E}[z_{N, p}(\sigma) z_{N, p}(\sigma')^T] = \left(\frac{\sigma \sigma'^T}{N} \right)^{\circ(p-1)}. \quad (\text{A.6})$$

If $p = 1$, then the right-hand side of (A.6) is interpreted as the $\kappa \times \kappa$ identity matrix, and so it suffices to take $z_{N, p}(\sigma)$ equal to a standard normal random vector not depending on σ . If $p \geq 2$, then we make the following construction. For $k \in \{1, \dots, \kappa\}$ define

$$z_{N, p, k}(\sigma) = \frac{1}{N^{p-1}} \sum_{i_1, \dots, i_{p-1}=1}^N g_{i_1, \dots, i_{p-1}} \sigma_{i_1}(k) \dots \sigma_{i_{p-1}}(k),$$

where each $g_{i_1, \dots, i_{p-1}}$ is an independent standard normal random variable. By setting $z_{N, p}(\sigma) = (z_{N, p, k}(\sigma))_{k=1}^{\kappa}$, we obtain a κ -dimensional Gaussian process satisfying (A.6).

Now let $z_{N, p}^1, \dots, z_{N, p}^m$ be independent copies of $z_{N, p}$. As in the proof of Lemma A.1, let $\theta_j \in \Theta$ be obtained from $\theta = (p, m, n_1, \dots, n_m, w_1, \dots, w_m)$ by modifying n_j to $n_j - 1$. From (A.4), consider centered Gaussian processes $H_{N, \theta_1}, \dots, H_{N, \theta_m}$, which we assume are independent of each other and of each $z_{N, p}^1, \dots, z_{N, p}^m$. Finally, recalling the notation $W_j = \text{diag}(w_j)$, we define

$$Z_{N, \theta}(\sigma) := p^{1/2} \sum_{j=1}^m n_j^{1/2} N^{-1/2} H_{N, \theta_j}(\sigma) \cdot W_j z_{N, p}^j(\sigma).$$

It follows from (A.3) and (A.6) that the covariance of $(Z_{N, \theta}(\sigma))_{\sigma \in \mathbb{R}^{\kappa \times N}}$ is given by (A.5). \square

Proof of Proposition 2.7. Parts (a) and (b) follow immediately from Lemma A.1. Part (c) follows from Lemma A.2. Finally, part (d) follows from combining Lemmas A.1 and A.2 since we can set $Y_{N,\theta}(\sigma) = \left(\frac{\deg(\theta)-1}{N}\right)^{1/2} H_{N,\theta}(\sigma)$. \square

The following lemma was used in Remark 1.2.

Lemma A.3. *For any $d \in \mathcal{D}$ and any γ belonging to the set $\Gamma_\kappa(d)$ from (1.7), we have $\gamma \preceq \text{diag}(d)$.*

Proof. For any $u = (u_1, \dots, u_\kappa) \in \mathbb{R}^\kappa$, we have

$$\begin{aligned} \langle \text{diag}(d)u, u \rangle - \langle \gamma u, u \rangle &= \sum_{k=1}^{\kappa} d_k u_k^2 - \sum_{k,k'=1}^{\kappa} \gamma_{k,k'} u_k u_{k'} \\ &\stackrel{(1.7)}{=} \sum_{k,k'=1}^{\kappa} \gamma_{k,k'} (u_k^2 - u_k u_{k'}) \\ &= \sum_{k < k'} \gamma_{k,k'} (u_k^2 + u_{k'}^2 - 2u_k u_{k'}) = \sum_{k < k'} \gamma_{k,k'} (u_k - u_{k'})^2 \geq 0, \end{aligned}$$

where the final inequality uses the fact that every entry $\gamma_{k,k'}$ is nonnegative. \square

Appendix B. Proof of the Aizenman–Sims–Starr Scheme

In this appendix, we prove Proposition 6.2. To begin, we deduce from (6.6) the inclusion

$$\Sigma^{N+M}(d^{N+M}) \supseteq \Sigma^N(d^N) \times \Sigma^M(\delta^M). \quad (\text{B.1})$$

Indeed, for $\sigma = (\sigma_1, \dots, \sigma_N) \in \Sigma^N(d^N)$ and $\tau = (\tau_1, \dots, \tau_M) \in \Sigma^M(\delta^M)$, let us write (σ, τ) for the $\kappa \times (N+M)$ matrix whose first N columns are $\sigma_1, \dots, \sigma_N$ and whose last M columns are τ_1, \dots, τ_M . In this notation, we have

$$(\sigma, \tau)(\sigma, \tau)^\top = \sigma\sigma^\top + \tau\tau^\top = N\text{diag}(d^N) + M\text{diag}(\delta^M) \stackrel{(6.6)}{=} (N+M)\text{diag}(d^{N+M}),$$

which by definition means $(\sigma, \tau) \in \Sigma^{N+M}(d^{N+M})$. From (B.1), it immediately follows that

$$Z_{N+M,\tilde{\xi}}(d^{N+M}) \geq \sum_{\sigma \in \Sigma^N(d^N)} \sum_{\tau \in \Sigma^M(\delta^M)} \exp H_{N+M,\tilde{\xi}}(\sigma, \tau). \quad (\text{B.2})$$

Recall from (1.14) that

$$\begin{aligned} \mathbb{E}[H_{N+M,\tilde{\xi}}(\sigma, \tau) H_{N+M,\tilde{\xi}}(\sigma', \tau')] &= (N+M)\tilde{\xi} \left(\frac{(\sigma, \tau)(\sigma', \tau')^\top}{N+M} \right) \\ &= (N+M)\tilde{\xi} \left(\frac{\sigma\sigma'^\top + \tau\tau'^\top}{N+M} \right). \end{aligned}$$

Replacing $\tilde{\xi}$ with ξ on the right-hand side incurs an error:

$$\begin{aligned} \left| \mathbb{E}[H_{N+M,\tilde{\xi}}(\sigma, \tau) H_{N+M,\tilde{\xi}}(\sigma', \tau')] - (N+M)\xi \left(\frac{\sigma\sigma'^\top + \tau\tau'^\top}{N+M} \right) \right| \\ \leq (N+M) \sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)|. \end{aligned} \quad (\text{B.3})$$

By smoothness (see Remark 1.5), we have the following linearization about $(\sigma\sigma'^T)/(N+M)$:

$$\xi\left(\frac{\sigma\sigma'^T + \tau\tau'^T}{N+M}\right) = \xi\left(\frac{\sigma\sigma'^T}{N+M}\right) + \left\langle \nabla\xi\left(\frac{\sigma\sigma'^T}{N+M}\right), \frac{\tau\tau'^T}{N+M} \right\rangle + O_{\nabla^2\xi}\left(\left(\frac{M}{N+M}\right)^2\right),$$

where we are using the notation from (6.5). Furthermore, we have the approximation

$$\nabla\xi\left(\frac{\sigma\sigma'^T}{N+M}\right) = \nabla\xi\left(\frac{\sigma\sigma'^T}{N}\right) + O_{\nabla^2\xi}\left(\frac{M}{N+M}\right).$$

Combining the two previous displays, we arrive to

$$\begin{aligned} (N+M)\xi\left(\frac{\sigma\sigma'^T + \tau\tau'^T}{N+M}\right) &= (N+M)\xi\left(\frac{\sigma\sigma'^T}{N+M}\right) + \left\langle \nabla\xi\left(\frac{\sigma\sigma'^T}{N}\right), \tau\tau'^T \right\rangle \\ &\quad + O_{\nabla^2\xi}\left(\frac{M^2}{N+M}\right). \end{aligned} \quad (\text{B.4})$$

Regarding the first term on the right-hand side, we have the further linearization

$$\begin{aligned} (N+M)\xi\left(\frac{\sigma\sigma'^T}{N+M}\right) &= (N+M)\left[\xi\left(\frac{\sigma\sigma'^T}{N}\right) - \left\langle \nabla\xi\left(\frac{\sigma\sigma'^T}{N}\right), \frac{M\sigma\sigma'^T}{N(N+M)} \right\rangle + O_{\nabla^2\xi}\left(\left(\frac{M}{N+M}\right)^2\right)\right] \\ &= N\xi\left(\frac{\sigma\sigma'^T}{N}\right) - M\left[\left\langle \nabla\xi\left(\frac{\sigma\sigma'^T}{N}\right), \frac{\sigma\sigma'^T}{N} \right\rangle - \xi\left(\frac{\sigma\sigma'^T}{N}\right)\right] + O_{\nabla^2\xi}\left(\frac{M^2}{N+M}\right). \end{aligned}$$

The term being subtracted on the final line is simply $M\vartheta_\xi\left(\frac{\sigma\sigma'^T}{N}\right)$. Moving this term to the left-hand side, we arrive to

$$(N+M)\xi\left(\frac{\sigma\sigma'^T}{N+M}\right) + M\vartheta_\xi\left(\frac{\sigma\sigma'^T}{N}\right) = N\xi\left(\frac{\sigma\sigma'^T}{N}\right) + O_{\nabla^2\xi}\left(\frac{M^2}{N+M}\right). \quad (\text{B.5})$$

Equipped with the approximations (B.4) and (B.5), we resume our probabilistic argument.

Let $H_{N,M,\xi}$ be the cavity Hamiltonian from (6.2). Let $Z_1, \dots, Z_M: \Sigma^N \rightarrow \mathbb{R}^\kappa$ and $Y: \Sigma^N \rightarrow \mathbb{R}$ be centered Gaussian processes with covariances

$$\mathbb{E}\left[Z_i(\sigma)Z_i(\sigma')^T\right] = \nabla\xi\left(\frac{\sigma\sigma'^T}{N}\right), \quad \mathbb{E}[Y(\sigma)Y(\sigma')] = \vartheta_\xi\left(\frac{\sigma\sigma'^T}{N}\right). \quad (\text{B.6})$$

Such processes exist by Proposition 2.7(d) and assumption (A1). We assume all these processes are independent of each other and of $H_{N,M,\xi}$. Now write

$$\frac{Z_{N+M,\tilde{\xi}}(d^{N+M})}{Z_{N,\xi}(d^N)} \stackrel{(\text{B.2})}{\geq} \frac{\sum_{\sigma \in \Sigma^N(d^N)} \sum_{\tau \in \Sigma^M(\delta^M)} \exp H_{N+M,\tilde{\xi}}(\sigma, \tau)}{\sum_{\sigma \in \Sigma^N(d^N)} \exp H_{N,\xi}(\sigma)} = \frac{Q_1 Q_1^{\text{err}}}{Q_2 Q_2^{\text{err}}}, \quad (\text{B.7})$$

where Q_1 , Q_1^{err} , Q_2 , Q_2^{err} are the following quotients:

$$Q_1 := \frac{\sum_{\sigma \in \Sigma^N(d^N)} \sum_{\tau \in \Sigma^M(\delta^M)} \exp(H_{N,M,\xi}(\sigma) + \sum_{j=1}^M \langle Z_j(\sigma), \tau_j \rangle)}{\sum_{\sigma \in \Sigma^N(d^N)} \exp H_{N,M,\xi}(\sigma)},$$

$$Q_1^{\text{err}} := \frac{\sum_{\sigma \in \Sigma^N(d^N)} \sum_{\tau \in \Sigma^M(\delta^M)} \exp H_{N+M,\tilde{\xi}}(\sigma, \tau)}{\sum_{\sigma \in \Sigma^N(d^N)} \sum_{\tau \in \Sigma^M(\delta^M)} \exp(H_{N,M,\xi}(\sigma) + \sum_{j=1}^M \langle Z_j(\sigma), \tau_j \rangle)}, \quad (\text{B.8})$$

$$Q_2 := \frac{\sum_{\sigma \in \Sigma^N(d^N)} \exp(H_{N,M,\xi}(\sigma) + \sqrt{M} Y(\sigma))}{\sum_{\sigma \in \Sigma^N(d^N)} \exp H_{N,M,\xi}(\sigma)},$$

$$Q_2^{\text{err}} := \frac{\sum_{\sigma \in \Sigma^N(d^N)} \exp H_{N,\xi}(\sigma)}{\sum_{\sigma \in \Sigma^N(d^N)} \exp(H_{N,M,\xi}(\sigma) + \sqrt{M} Y(\sigma))}. \quad (\text{B.9})$$

Observe that Q_1 and Q_2 have the desired form for realizing the functional $\Psi_{M,\xi}$ from (3.3). Namely, if $\langle \cdot \rangle_N$ denotes expectation with respect to the measure $G_{N,M,\xi}$ from (6.1), then

$$\mathbb{E} \log Q_1 = \mathbb{E} \log \sum_{\tau \in \Sigma^M(\delta^M)} \left\langle \exp \left(\sum_{j=1}^M \langle Z_j(\sigma), \tau_j \rangle \right) \right\rangle_N \stackrel{(3.2a)}{=} M \Psi_{M,\xi}^{(1)}(\mathcal{L}_{N,M}; d^N, \Sigma^M(\delta^M)),$$

while $\mathbb{E} \log Q_2 = \mathbb{E} \log \langle \exp(\sqrt{M} Y(\sigma)) \rangle_N \stackrel{(3.2b)}{=} M \Psi_{M,\xi}^{(2)}(\mathcal{L}_{N,M}).$

(B.10)

Therefore, the remainder of the proof is to show that

$$|\mathbb{E} \log Q_1^{\text{err}}| \leq O_{\nabla^2 \xi} \left(\frac{M^2}{N+M} \right) + \frac{N+M}{2} \sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)|, \quad (\text{B.11a})$$

$$\text{and } \mathbb{E} \log Q_2^{\text{err}} = O_{\nabla^2 \xi} \left(\frac{M^2}{N+M} \right). \quad (\text{B.11b})$$

Indeed, the claim (6.7) follows from using (B.10) and (B.11) in the initial inequality (B.7).

We now argue each estimate in (B.11) separately, although the two arguments are very similar. Since the one for Q_2^{err} is slightly simpler, we begin there.

Control of Q_2^{err} . Define an interpolating Hamiltonian on $\Sigma^N(d^N)$:

$$\mathcal{H}_t(\sigma) = \sqrt{t} H_{N,\xi}(\sigma) + \sqrt{1-t} (H_{N,M,\xi}(\sigma) + \sqrt{M} Y(\sigma)), \quad t \in [0, 1]. \quad (\text{B.12})$$

Consider the free energy associated to \mathcal{H}_t :

$$\phi(t) = \mathbb{E} \log \sum_{\sigma \in \Sigma^N(d^N)} \exp \mathcal{H}_t(\sigma). \quad (\text{B.13})$$

Recalling the definition of Q_2^{err} from (B.9), we see that

$$\mathbb{E} \log Q_2^{\text{err}} = \phi(1) - \phi(0). \quad (\text{B.14})$$

We control this difference by studying the derivative of (B.13) with respect to t , which is easily computed to be

$$\phi'(t) = \mathbb{E} \left\langle \frac{\partial}{\partial t} \mathcal{H}_t(\sigma, \tau) \right\rangle_t,$$

where $\langle \cdot \rangle_t$ denotes expectation with respect to the Gibbs measure $G_t(\sigma) \propto \exp \mathcal{H}_t(\sigma)$ on $\Sigma^N(d^N)$. Thanks to Gaussian integration by parts [54, Lem. 1.1], we can rewrite this as

$$\phi'(t) = \mathbb{E} \langle \mathcal{C}(\sigma^1, \sigma^1) - \mathcal{C}(\sigma^1, \sigma^2) \rangle_t,$$

where σ^1, σ^2 denote independent samples from G_t , and $\mathcal{C}: \Sigma^N(d^N) \times \Sigma^N(d^N) \rightarrow \mathbb{R}$ is defined by

$$\mathcal{C}(\sigma, \sigma') = \mathbb{E} \left[\frac{\partial \mathcal{H}_t(\sigma)}{\partial t} \mathcal{H}_t(\sigma') \right]. \quad (\text{B.15})$$

Since the Gaussian processes on the right-hand side of (B.12) are independent and centered, the right-hand side of (B.15) reduces to a linear combination of their respective covariances:

$$\begin{aligned} \mathcal{C}(\sigma, \sigma') &\stackrel{(\text{B.12})}{=} \frac{1}{2} \mathbb{E}[H_{N,\xi}(\sigma) H_{N,\xi}(\sigma')] - \frac{1}{2} \mathbb{E}[H_{N,M,\xi}(\sigma) H_{N,M,\xi}(\sigma')] \\ &\quad - \frac{M}{2} \mathbb{E}[Y(\sigma) Y(\sigma')] \\ &\stackrel{(\text{1.14}), (\text{6.2}), (\text{B.6})}{=} \frac{1}{2} \left[N \xi \left(\frac{\sigma \sigma'^\top}{N} \right) - (N+M) \xi \left(\frac{\sigma \sigma'^\top}{N+M} \right) - M \vartheta_\xi \left(\frac{\sigma \sigma'^\top}{N} \right) \right] \\ &\stackrel{(\text{B.5})}{=} O_{\nabla^2 \xi} \left(\frac{M^2}{N+M} \right). \end{aligned}$$

Reading the three previous displays together, we have established the following estimate:

$$\sup_{t \in (0,1)} |\phi'(t)| = O_{\nabla^2 \xi} \left(\frac{M^2}{N+M} \right).$$

Therefore, (B.14) leads to (B.11b).

Control of Q_1^{err} . Define an interpolating Hamiltonian on $\Sigma^N(d^N) \times \Sigma^M(\delta^M)$:

$$\mathcal{H}_t(\sigma, \tau) = \sqrt{t} H_{N+M}(\sigma, \tau) + \sqrt{1-t} \left(H_{N,M,\xi}(\sigma) + \sum_{j=1}^M \langle Z_j(\sigma), \tau_j \rangle \right), \quad t \in [0, 1]. \quad (\text{B.16})$$

From (B.6), one can calculate the covariance of the final term on the right-hand side:

$$\mathbb{E} \left[\sum_{j=1}^M \langle Z_j(\sigma), \tau_j \rangle \langle Z_j(\sigma'), \tau'_j \rangle \right] = \left\langle \nabla \xi \left(\frac{\sigma \sigma'^\top}{N} \right), \tau \tau'^\top \right\rangle. \quad (\text{B.17})$$

Now consider the free energy associated to \mathcal{H}_t :

$$\phi(t) = \mathbb{E} \log \sum_{\sigma \in \Sigma^N(d^N)} \sum_{\tau \in \Sigma^M(\delta^M)} \exp \mathcal{H}_t(\sigma, \tau).$$

Recalling the definition of Q_1^{err} from (B.8), we see that

$$\mathbb{E} \log Q_1^{\text{err}} = \phi(1) - \phi(0). \quad (\text{B.18})$$

We control this difference by studying the derivative of (B.13) with respect to t , which is easily computed to be

$$\phi'(t) = \mathbb{E} \left\langle \frac{\partial}{\partial t} \mathcal{H}_t(\sigma, \tau) \right\rangle_t,$$

where $\langle \cdot \rangle_t$ denotes expectation with respect to the Gibbs measure $G_t(\sigma, \tau) \propto \exp \mathcal{H}_t(\sigma, \tau)$ on $\Sigma^N(d^N) \times \Sigma^M(\delta^M)$. By Gaussian integration by parts [54, Lem. 1.1], we can rewrite this as

$$\phi'(t) = \mathbb{E} \left\langle \mathcal{C}((\sigma^1, \tau^1), (\sigma^1, \tau^1)) - \mathcal{C}((\sigma^1, \tau^1), (\sigma^2, \tau^2)) \right\rangle_t,$$

where $(\sigma^1, \tau^1), (\sigma^2, \tau^2)$ denote independent samples from G_t , and $\mathcal{C}: (\Sigma^N(d^N) \times \Sigma^M(\delta^M))^2 \rightarrow \mathbb{R}$ is defined by

$$\mathcal{C}((\sigma, \tau), (\sigma', \tau')) = \mathbb{E} \left[\frac{\partial \mathcal{H}_t(\sigma, \tau)}{\partial t} \mathcal{H}_t(\sigma', \tau') \right]. \quad (\text{B.19})$$

Since the Gaussian processes on the right-hand side of (B.16) are independent and centered, the right-hand side of (B.19) reduces to a linear combination of their respective covariances:

$$\begin{aligned} & \left| \mathbb{E} \left[\frac{\partial \mathcal{H}_t(\sigma, \tau)}{\partial t} \mathcal{H}_t(\sigma', \tau') \right] \right| \stackrel{(\text{B.16})}{=} \left| \frac{1}{2} \mathbb{E} [H_{N+M, \tilde{\xi}}(\sigma, \tau) H_{N+M, \tilde{\xi}}(\sigma', \tau')] \right. \\ & \quad \left. - \frac{1}{2} \mathbb{E} [H_{N, M, \xi}(\sigma) H_{N, M, \xi}(\sigma')] - \frac{1}{2} \mathbb{E} \left[\sum_{j=1}^M \langle Z_j(\sigma), \tau_j \rangle \langle Z_j(\sigma'), \tau'_j \rangle \right] \right| \\ & \stackrel{(\text{B.3}), (6.2), (\text{B.17})}{\leq} \frac{N+M}{2} \sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)| + \frac{1}{2} \left| (N+M) \xi \left(\frac{\sigma \sigma'^T + \tau \tau'^T}{N+M} \right) \right. \\ & \quad \left. - (N+M) \xi \left(\frac{\sigma \sigma'^T}{N+M} \right) - \left\langle \nabla \xi \left(\frac{\sigma \sigma'^T}{N} \right), \tau \tau'^T \right\rangle \right| \\ & \stackrel{(\text{B.4})}{=} \frac{N+M}{2} \sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)| + O_{\nabla^2 \xi} \left(\frac{M^2}{N+M} \right). \end{aligned}$$

Reading the three previous displays together, we have established the following estimate:

$$\sup_{t \in (0,1)} |\phi'(t)| \leq O_{\nabla^2 \xi} \left(\frac{M^2}{N+M} \right) + \frac{N+M}{2} \sup_{\|R\|_1 \leq 1} |\xi(R) - \tilde{\xi}(R)|.$$

Therefore, (B.18) leads to (B.11a), which concludes the proof of Proposition 6.2.

Appendix C. Properties of Ruelle Probability Cascades

This appendix proves Lemmas 3.5 and 3.7. Both proofs rely on the following fact.

Theorem C.1. [72, Thm. 14.2.1] *Let $\eta^{(0)}, \dots, \eta^{(s-1)}$ be independent random variables taking values in a metric space T . Assume $f: T^s \rightarrow \mathbb{R}$ is a deterministic function such that*

$$\mathbb{E} \exp f(\eta^{(0)}, \eta^{(1)}, \dots, \eta^{(s-1)}) < \infty \quad \text{and} \quad \mathbb{E} |f(\eta^{(0)}, \eta^{(1)}, \dots, \eta^{(s-1)})| < \infty.$$

Let \mathbb{E}_r denote expectation with respect to $\eta^{(r)}$. Given the sequence (3.4), inductively define

$$\begin{aligned} X_s &:= f(\eta^{(0)}, \eta^{(1)}, \dots, \eta^{(s-1)}), \\ X_r &:= \frac{1}{m_r} \log \mathbb{E}_r \exp(m_r X_{r+1}) \quad \text{for } r \in \{1, \dots, s-1\}, \\ X_0 &:= \mathbb{E}_0(X_1). \end{aligned} \tag{C.1}$$

On the other hand, let v be the RPC associated to (3.4). For each $r \in \{0, \dots, s-1\}$, let $(\eta_\beta)_{\beta \in \mathbb{N}^r}$ be i.i.d. copies of $\eta^{(r)}$ that are independent of v . We then have

$$X_0 = \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} v_\alpha \exp f(\eta_\emptyset, \eta_{(\alpha_1)}, \eta_{(\alpha_1, \alpha_2)}, \dots, \eta_{(\alpha_1, \dots, \alpha_{s-1})}). \tag{C.2}$$

Proof of Lemma 3.5. The right-hand sides of (3.9) are defined in (3.2) with

$$\mathcal{X} = \mathbb{N}^{s-1}, \quad R(\alpha, \alpha') = \gamma_{r(\alpha, \alpha')}, \quad \mathcal{G} = v. \tag{C.3}$$

By inspection, (3.2a) is the same as (3.7) with $\lambda = 0$. This proves (3.9a).

Meanwhile, in the setting (C.3), definition (3.2b) becomes

$$\Psi_{N, \xi}^{(2)}(\mathcal{L}; d) = \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} v_\alpha \exp(\sqrt{N} Y(\alpha)). \tag{C.4a}$$

Using the representation (3.12) of Y , we can rewrite this as

$$\Psi_{N, \xi}^{(2)}(\mathcal{L}; d) = \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} v_\alpha \exp\left(\sqrt{N} \sum_{r=0}^{s-1} \sqrt{\vartheta_\xi(\gamma_{r+1}) - \vartheta_\xi(\gamma_r)} \eta_{(\alpha_1, \dots, \alpha_r)}\right).$$

By (C.2), the right-hand side can be transformed to yield

$$\Psi_{N, \xi}^{(2)}(\mathcal{L}; d) = \frac{1}{N} X_0, \tag{C.4b}$$

where X_0, X_1, \dots, X_s are related inductively as in (C.1), and

$$X_s = \sqrt{N} \sum_{r=0}^{s-1} \sqrt{\vartheta_\xi(\gamma_{r+1}) - \vartheta_\xi(\gamma_r)} \eta^{(r)}.$$

Here $\eta^{(0)}, \dots, \eta^{(s-1)}$ are i.i.d. standard normal random variables. Therefore, using the fact that $\mathbb{E} \exp(c\eta^{(r)}) = \frac{1}{2} \exp(c^2)$ for any $c \in \mathbb{R}$, it is straightforward to calculate

$$X_0 = \frac{N}{2} \sum_{r=1}^{s-1} m_r (\vartheta_\xi(\gamma_{r+1}) - \vartheta_\xi(\gamma_r)).$$

Rewriting the right-hand side using summation by parts, we obtain

$$\begin{aligned} X_0 &= \frac{N}{2} \vartheta_\xi(\gamma_s) - \frac{N}{2} \sum_{r=1}^s (m_r - m_{r-1}) \vartheta_\xi(\gamma_r) \\ &= \frac{N}{2} \vartheta_\xi(\text{diag}(d)) - \frac{N}{2} \int_0^1 \vartheta_\xi(\pi(t)) \, dt \stackrel{(3.8)}{=} -N \mathcal{P}_\xi^{(2)}(\pi). \end{aligned} \quad (\text{C.4c})$$

Therefore, (C.4b) is exactly the desired statement (3.9b). \square

Proof of Lemma 3.7. Inserting (3.11) into (3.7), we can express $\mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; S)$ as

$$\frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in S} \nu_\alpha \exp \left(\sum_{i=1}^N \left\langle \sum_{r=0}^{s-1} \sqrt{\nabla \xi(\gamma_{r+1}) - \nabla \xi(\gamma_r) \mathbb{1}\{r > 0\}} \eta_{i,(\alpha_1, \dots, \alpha_r)} + \lambda, \sigma_i \right\rangle \right).$$

By (C.2), we have the alternative representation

$$\mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; S) = \frac{1}{N} X_0(S), \quad (\text{C.5a})$$

where $X_0(S), X_1(S), \dots, X_s(S)$ are related inductively as in (C.1), and

$$X_s(S) = \log \sum_{\sigma \in S} \exp \left(\sum_{i=1}^N \left\langle \sum_{r=0}^{s-1} z_i^{(r)} + \lambda, \sigma_i \right\rangle \right). \quad (\text{C.5b})$$

Here each $z_i^{(r)}$ is an independent centered Gaussian vector in \mathbb{R}^k with covariance matrix

$$\mathbb{E}[z_i^{(r)} (z_i^{(r)})^\top] = \nabla \xi(\gamma_{r+1}) - \nabla \xi(\gamma_r) \mathbb{1}\{r > 0\}. \quad (\text{C.5c})$$

When S is the entire product set Σ^N , (C.5b) can be rewritten as

$$\begin{aligned} X_s(\Sigma^N) &= \log \sum_{\sigma_1, \dots, \sigma_N \in \Sigma} \prod_{i=1}^N \exp \left(\sum_{r=0}^{s-1} \langle z_i^{(r)} + \lambda, \sigma_i \rangle \right) \\ &= \sum_{i=1}^N \log \sum_{\sigma \in \Sigma} \exp \left(\sum_{r=0}^{s-1} \langle z_i^{(r)} + \lambda, \sigma \rangle \right) = \sum_{i=1}^N X_s^{(i)}(\Sigma), \end{aligned}$$

where $X_s^{(1)}(\Sigma), \dots, X_s^{(N)}(\Sigma)$ are i.i.d. random variables. The inductive procedure (C.1) can be applied to each one of these variables separately, resulting in i.i.d. random variables $X_r^{(1)}(\Sigma), \dots, X_r^{(N)}(\Sigma)$ for each $r \in \{0, 1, \dots, s-1\}$. Assuming inductively that

$X_{r+1}(\Sigma^N) = \sum_{i=1}^N X_{r+1}^{(i)}(\Sigma)$, we have

$$\begin{aligned} X_r(\Sigma^N) &\stackrel{\text{(C.1)}}{=} \frac{1}{m_r} \log \mathbb{E}_r \exp \left(m_r \sum_{i=1}^N X_{r+1}^{(i)}(\Sigma) \right) \\ &= \frac{1}{m_r} \log \prod_{i=1}^N \mathbb{E}_r \exp(m_r X_{r+1}^{(i)}(\Sigma)) \stackrel{\text{(C.1)}}{=} \sum_{i=1}^N X_r^{(i)}(\Sigma), \end{aligned}$$

where the middle equality uses independence. Taking $r = 0$, we conclude that

$$\mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N) \stackrel{\text{(C.5a)}}{=} \frac{1}{N} X_0(\Sigma^N) = \frac{1}{N} \sum_{i=1}^N X_0^{(i)}(\Sigma) = X_0^{(1)}(\Sigma) \stackrel{\text{(C.5a)}}{=} \mathcal{P}_{1,\xi}^{(1)}(\pi, \lambda; \Sigma). \quad (\text{C.6})$$

Moreover, a comparison of (C.5a) and (1.31) reveals that

$$\mathcal{P}_{1,\xi}^{(1)}(\pi, \lambda; \Sigma) = \mathcal{P}_{\xi}^{(1)}(\pi, \lambda),$$

and so (C.6) is exactly (3.10). \square

Appendix D. Duality

In this appendix, we prove Lemma 5.2. We closely follow the proof of [58, Lem. 2]. The key ingredient is the following result. Recall the set $\Sigma^N(d, \varepsilon)$ from (1.4): using the notation $R(\sigma, \sigma') = \sigma \sigma'^T / N$, we have

$$\Sigma^N(d, \varepsilon) = \{\sigma \in \Sigma^N : \|R(\sigma, \sigma) - \text{diag}(d)\|_{\infty} \leq \varepsilon\}. \quad (\text{D.1})$$

Lemma D.1. *Assume ξ satisfies (A1). There exists a constant C depending only on ξ and κ , such that for every $N \geq 1$, $d' \in \mathcal{D}$, $\pi \in \Pi_{d'}$, $\lambda \in \mathbb{R}^{\kappa}$, and $\varepsilon > 0$, we have*

$$\sup_{d \in \mathcal{D}_N} \left| \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d, \varepsilon)) - \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d)) \right| \leq C\sqrt{\varepsilon}. \quad (\text{D.2})$$

Proof. We assume without loss of generality that $\varepsilon \in (0, 1/2]$. Note for later that there exists a constant C_0 such that

$$\varepsilon \log(1/\varepsilon) \leq C_0 \sqrt{\varepsilon} \quad \text{for all } \varepsilon \in (0, 1/2]. \quad (\text{D.3})$$

For any $\lambda \in \mathbb{R}^{\kappa}$ and $S \subseteq \Sigma^N$, the map $\Pi_{d'} \ni \pi \mapsto \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; S)$ is by definition a continuous extension of its restriction to the set $\Pi_{d'}^{\text{disc}}$ (see the text following the proof of Lemma 5.1). Therefore, it suffices to verify the lemma for $\pi \in \Pi_{d'}^{\text{disc}}$. We assume π is given by (1.29), with d' replacing d in (1.29c).

Given any $d \in \mathcal{D}_N$, let $P: \Sigma^N(d, \varepsilon) \rightarrow \Sigma^N(d)$ be any projection with respect to Hamming distance. That is, for each $\sigma \in \Sigma^N(d, \varepsilon)$, we choose some $P\sigma \in \Sigma^N(d)$ that minimizes $\sum_{i=1}^N \mathbb{1}\{\sigma_i \neq (P\sigma)_i\}$. Changing a single column σ_i (from one standard

basis vector to another) moves the self-overlap matrix $R(\sigma, \sigma)$ by distance N^{-1} in the ℓ^∞ norm. Therefore, σ and $P\sigma$ differ in at most εN columns:

$$\sum_{i=1}^N \mathbb{1}\{\sigma_i \neq (P\sigma)_i\} \leq \varepsilon N \quad \text{for every } \sigma \in \Sigma^N(d, \varepsilon). \quad (\text{D.4})$$

In particular, if $\varepsilon N < 1$, then P is the identity map and so $\Sigma^N(d, \varepsilon) = \Sigma^N(d)$. The right-hand side of (D.2) is zero in this case, and so we assume henceforth that $\varepsilon N \geq 1$.

Consider the size of the preimage,

$$\mathcal{N}(\sigma) := |\{\tau \in \Sigma^N(d, \varepsilon) : P\tau = \sigma\}|. \quad (\text{D.5})$$

If $P\tau = \sigma$, then (D.4) says there are at most εN many columns in which τ and σ differ. Each such column of τ takes one of κ possible values, and so

$$\mathcal{N}(\sigma) \leq \binom{N}{\lceil \varepsilon N \rceil} \kappa^{\varepsilon N} \leq \left(\frac{\exp(1)N}{\lceil \varepsilon N \rceil} \right)^{\lceil \varepsilon N \rceil} \kappa^{\varepsilon N} \leq \left(\frac{\exp(1)\kappa}{\varepsilon} \right)^{\varepsilon N+1} \leq \left(\frac{\exp(1)\kappa}{\varepsilon} \right)^{2\varepsilon N}.$$

Taking logarithms and dividing by N , we arrive to

$$\frac{1}{N} \log \mathcal{N}(\sigma) \leq 2\varepsilon(1 + \log \kappa + \log(1/\varepsilon)) \stackrel{(\text{D.3})}{\leq} C_\kappa \sqrt{\varepsilon}, \quad (\text{D.6})$$

where $C_\kappa = 2(1 + \log \kappa + C_0)$.

Next let $Z_1, \dots, Z_N, \tilde{Z}_1, \dots, \tilde{Z}_N : \mathbb{N}^{s-1} \rightarrow \mathbb{R}^\kappa$ be i.i.d. centered Gaussian processes with covariance given by (3.6). Define the interpolating process $Z_t : \mathbb{N}^{s-1} \times \Sigma^N(d, \varepsilon) \rightarrow \mathbb{R}$ by

$$Z_t(\alpha, \sigma) = \sum_{i=1}^N \left[\sqrt{t} \langle Z_i(\alpha) + \lambda, \sigma_i \rangle + \sqrt{1-t} \langle \tilde{Z}_i(\alpha) + \lambda, (P\sigma)_i \rangle \right], \quad t \in [0, 1]. \quad (\text{D.7})$$

Upon defining the associated free energy,

$$\phi(t) = \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in \Sigma^N(d, \varepsilon)} v_\alpha \exp Z_t(\alpha, \sigma), \quad (\text{D.8})$$

we have

$$\begin{aligned} \phi(1) &= \mathcal{P}_{N, \xi}^{(1)}(\pi, \lambda; \Sigma^N(d, \varepsilon)), \\ \phi(0) &= \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in \Sigma^N(d, \varepsilon)} \exp \left(\sum_{i=1}^N \langle \tilde{Z}_i(\alpha) + \lambda, (P\sigma)_i \rangle \right). \end{aligned} \quad (\text{D.9})$$

Note that $\mathcal{N}(\sigma) \geq 1$ since $P\sigma = \sigma$, and thus

$$\begin{aligned} \mathcal{P}_{N, \xi}^{(1)}(\pi, \lambda; \Sigma^N(d)) &\stackrel{(3.7)}{=} \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in \Sigma^N(d)} v_\alpha \exp \left(\sum_{i=1}^N \langle Z_i(\alpha) + \lambda, \sigma_i \rangle \right) \\ &\leq \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in \Sigma^N(d)} \mathcal{N}(\sigma) v_\alpha \exp \left(\sum_{i=1}^N \langle Z_i(\alpha) + \lambda, \sigma_i \rangle \right) \stackrel{(\text{D.9})}{=} \phi(0) \\ &\stackrel{(\text{D.6})}{\leq} \mathcal{P}_{N, \xi}^{(1)}(\pi, \lambda; \Sigma^N(d)) + C_\kappa \sqrt{\varepsilon}. \end{aligned} \quad (\text{D.10})$$

Putting these observations together and using a triangle inequality, we have

$$\begin{aligned}
 & \left| \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d, \varepsilon)) - \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d)) \right| \\
 & \stackrel{\text{(D.9)}}{=} \left| \phi(1) - \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d)) \right| \\
 & \leq \left| \phi(1) - \phi(0) \right| + \left| \phi(0) - \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d)) \right| \\
 & \stackrel{\text{(D.10)}}{\leq} \left| \phi(1) - \phi(0) \right| + C_\kappa \sqrt{\varepsilon}.
 \end{aligned} \tag{D.11}$$

To control the first term in the last line, we consider the probability measure $G_t(\alpha, \sigma) \propto \nu_\alpha \exp Z_t(\alpha, \sigma)$ on $\mathbb{N}^{s-1} \times \Sigma^N(d, \varepsilon)$, and let $\langle \cdot \rangle_t$ denote expectation with respect to G_t . Differentiation of (D.8) results in

$$\phi'(t) = \frac{1}{N} \mathbb{E} \left\langle \frac{\partial Z_t(\alpha, \sigma)}{\partial t} \right\rangle_t.$$

Applying Gaussian integration by parts [54, Lem. 1.1], we can rewrite this as

$$\phi'(t) = \frac{1}{N} \mathbb{E} \left\langle \mathcal{C}((\alpha^1, \sigma^1), (\alpha^1, \sigma^1)) - \mathcal{C}((\alpha^1, \sigma^1), (\alpha^1, \sigma^1)) \right\rangle_t,$$

where $(\alpha^1, \sigma^1), (\alpha^2, \sigma^2)$ denote independent samples from G_t , and $\mathcal{C}: (\mathbb{N}^{s-1} \times \Sigma^N(d, \varepsilon))^2 \rightarrow \mathbb{R}$ is given by

$$\begin{aligned}
 & \mathcal{C}((\alpha, \sigma), (\alpha', \sigma')) \\
 & = \mathbb{E} \left[\frac{\partial Z_t(\alpha, \sigma)}{\partial t} Z_t(\alpha', \sigma') \right] \\
 & \stackrel{\text{(D.7)}}{=} \frac{1}{2} \sum_{i=1}^N \mathbb{E} \left[\langle Z_i(\alpha), \sigma_i \rangle \langle Z_i(\alpha'), \sigma'_i \rangle - \langle \tilde{Z}_i(\alpha), (P\sigma)_i \rangle \langle \tilde{Z}_i(\alpha'), (P\sigma')_i \rangle \right] \\
 & \stackrel{\text{(3.6)}}{=} \frac{1}{2} \sum_{i=1}^N \left(\sigma_i^\top \nabla \xi(\gamma_{r(\alpha, \alpha')}) \sigma'_i - (P\sigma)_i^\top \nabla \xi(\gamma_{r(\alpha, \alpha')}) (P\sigma')_i \right).
 \end{aligned}$$

By (D.4), there are at most $2\varepsilon N$ many values of i for which the i th summand on the final line is nonzero. Since $\sigma_i, \sigma'_i, (P\sigma)_i, (P\sigma')_i$ are all standard basis vectors, the nonzero summands trivially satisfy

$$\left| \sigma_i^\top \nabla \xi(\gamma_{r(\alpha, \alpha')}) \sigma'_i - (P\sigma)_i^\top \nabla \xi(\gamma_{r(\alpha, \alpha')}) (P\sigma')_i \right| \leq 2 \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R)\|_\infty.$$

Therefore, the four previous displays together imply

$$\left| \phi(1) - \phi(0) \right| \leq \sup_{t \in (0, 1)} \left| \phi'(t) \right| \leq 4\varepsilon \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R)\|_\infty.$$

Inserting this inequality into (D.11) results in

$$\left| \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d, \varepsilon)) - \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d)) \right| \leq C\sqrt{\varepsilon},$$

where $C = 4 \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R)\|_\infty + C_\kappa$. \square

With Lemma D.1 in hand, the next result follows by the exact same argument as [58, Lem. 4 and Lem. 5], and thus we omit the proof.

Lemma D.2. *Assume ξ satisfies (A1). Fix any $d' \in \mathcal{D}$, $\pi \in \Pi_{d'}$. Suppose $d^N \in \mathcal{D}_N$ converges to $d \in \mathcal{D}$ as $N \rightarrow \infty$. Then the following limit exists, is finite, and depends only on d :*

$$f(d) = \lim_{N \rightarrow \infty} \mathcal{P}_{N,\xi}^{(1)}(\pi, 0; \Sigma^N(d^N)). \quad (\text{D.12})$$

Furthermore, $d \mapsto f(d)$ is a continuous concave function.

We are now ready to complete the objective of this appendix. The following proof is very similar to that of [58, Lem. 6].

Proof of Lemma 5.2. As in the proof of Lemma D.1, let π be given by (1.29), with d' replacing d in (1.29c). We will show that

$$0 \leq \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N) - \max_{d \in \mathcal{D}_N} \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d)) \leq \frac{\kappa \log(N+1)}{m_1 N}. \quad (\text{D.13})$$

Before proving (D.13), we use it to prove the desired identity (5.17). For every $\sigma \in \Sigma^N(d)$, we have $\sum_{i=1}^n \langle \lambda, \sigma_i \rangle = N \langle \lambda, d \rangle$. Therefore, when $S = \Sigma^N(d)$, definition (3.7) results in

$$\mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d)) = \mathcal{P}_{N,\xi}^{(1)}(\pi, 0, \Sigma^N(d)) + \langle \lambda, d \rangle \quad \text{for } d \in \mathcal{D}_N.$$

Using this identity and $\mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N) = \mathcal{P}_{\xi}^{(1)}(\pi, \lambda)$ from (3.10), we infer from (D.13) that

$$\mathcal{P}_{\xi}^{(1)}(\pi, \lambda) = \lim_{N \rightarrow \infty} \max_{d \in \mathcal{D}_N} [\mathcal{P}_{N,\xi}^{(1)}(\pi, 0; \Sigma^N(d)) + \langle \lambda, d \rangle]. \quad (\text{D.14})$$

We then claim that

$$\mathcal{P}_{\xi}^{(1)}(\pi, \lambda) = \sup_{d \in \mathcal{D}} [f(d) + \langle \lambda, d \rangle]. \quad (\text{D.15})$$

Indeed, let $d^N \in \mathcal{D}_N$ be a maximizer in (D.14). By compactness of \mathcal{D} , we may assume (by passing to a subsequence) that d^N converges to some d as $N \rightarrow \infty$. It then follows that

$$\mathcal{P}_{\xi}^{(1)}(\pi, \lambda) \stackrel{(\text{D.14})}{=} \lim_{N \rightarrow \infty} [\mathcal{P}_{N,\xi}^{(1)}(\pi, 0; \Sigma^N(d^N)) + \langle \lambda, d^N \rangle] \stackrel{(\text{D.12})}{=} f(d) + \langle \lambda, d \rangle. \quad (\text{D.16})$$

On the other hand, if we choose any other convergent sequence $d^N \rightarrow d$, not necessarily a sequence of maximizers, then the first equality in (D.16) becomes \geq , and (D.15) follows. Finally, since f is concave and continuous by Lemma D.2, the Fenchel–Moreau theorem (e.g. [64, Thm. 12.2]) implies the following dual version of (D.15):

$$f(d) = \inf_{\lambda \in \mathbb{R}^k} [\mathcal{P}_{\xi}^{(1)}(\pi, \lambda) - \langle \lambda, d \rangle].$$

Once we use (D.12) to rewrite $f(d)$, this identity is exactly (5.17).

It remains to establish the two inequalities in (D.13). The first inequality is immediate from definition (3.7), since $\Sigma^N(d) \subseteq \Sigma^N$. For the second inequality, recall the notation in (C.5). We claim

$$\exp(m_r X_r(\Sigma^N)) \leq \sum_{d \in \mathcal{D}_N} \exp(m_r X_r(\Sigma^N(d))) \quad \text{for each } r \in \{1, \dots, s\}. \quad (\text{D.17})$$

To see this claim, first observe that there is equality when $r = s$, simply by inspecting (C.5b) together with the fact that $m_s = 1$. Now proceed by downward induction. Assuming the claim is true in the $(r + 1)$ th case, and noting that $m_r/m_{r+1} \leq 1$, we have

$$\begin{aligned} \exp(m_r X_r(\Sigma^N)) &\stackrel{(\text{C.1})}{=} \mathbb{E}_r \exp(m_r X_{r+1}(\Sigma^N)) \\ &= \mathbb{E}_r \left[\exp(m_{r+1} X_{r+1}(\Sigma^N))^{m_r/m_{r+1}} \right] \\ &\leq \mathbb{E}_r \left[\left(\sum_{d \in \mathcal{D}_N} \exp(m_{r+1} X_{r+1}(\Sigma^N(d))) \right)^{m_r/m_{r+1}} \right] \\ &\leq \mathbb{E}_r \left[\sum_{d \in \mathcal{D}_N} \exp(m_r X_{r+1}(\Sigma^N(d))) \right] \\ &\stackrel{(\text{C.1})}{=} \sum_{d \in \mathcal{D}_N} \exp(m_r X_r(\Sigma^N(d))). \end{aligned}$$

With (D.17) established, we use the final $r = 1$ case:

$$\mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N) \stackrel{(\text{C.5a})}{=} \frac{1}{N} \mathbb{E}_0 X_1(\Sigma^N) \stackrel{(\text{D.17})}{\leq} \frac{1}{m_1 N} \mathbb{E}_0 \log \sum_{d \in \mathcal{D}_N} \exp(m_1 X_1(\Sigma^N(d))).$$

By definition of \mathcal{D}_N in (1.6), each $d = (d_1, \dots, d_\kappa) \in \mathcal{D}_N$ has $N d_i \in \{0, 1, \dots, N\}$ for each i . Hence $|\mathcal{D}_N| \leq (N + 1)^\kappa$, and so the previous display yields

$$\begin{aligned} \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N) &\leq \frac{\kappa \log(N + 1)}{m_1 N} + \max_{d \in \mathcal{D}_N} \frac{1}{N} \mathbb{E}_0 X_1(\Sigma^N(d)) \\ &\stackrel{(\text{C.5a})}{=} \frac{\kappa \log(N + 1)}{m_1 N} + \max_{d \in \mathcal{D}_N} \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N(d)). \end{aligned}$$

This completes the proof of (D.13), and so we are done. \square

Appendix E. Parisi Formula for General Model

In this appendix, we prove Theorem 1.7 by generalizing the strategy used by Panchenko [57, 58] for Theorem 1.1. The following upper bound uses Guerra-style interpolation [36].

Proposition E.1. *Assume ξ satisfies (A1) and (A2).*

For every $d \in \mathcal{D}$, we have

$$\lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} \mathbf{F}_{N,\xi}(d, \varepsilon) \leq \inf_{\pi \in \Pi_d} \mathcal{P}_\xi(\pi). \quad (\text{E.1})$$

Meanwhile, the lower bound does not require the convexity assumption (A2). The following result follows from the Aizenman–Sims–Starr scheme as in [58, Sec. 7].

Proposition E.2. *Assume ξ satisfies (A1). Given $d = (d_k)_{k=1}^\kappa \in \mathcal{D}$ and any constant L , assume $d^N \in \mathcal{D}_N$ is such that*

$$\|d^N - d\|_\infty \leq L/N \quad \text{and} \quad d_k^N = 0 \text{ whenever } d_k = 0. \quad (\text{E.2})$$

Then

$$\liminf_{N \rightarrow \infty} F_{N,\xi}(d^N) \geq \inf_{\pi \in \Pi_d} \mathcal{P}_\xi(\pi). \quad (\text{E.3})$$

Although [58] considers only the case $\xi(R) = \text{tr}(R^\top R)$, the argument for Proposition E.2 proceeds in exactly the same way, and thus we omit the proof. Proposition E.1, however, is inherently more sensitive to the covariance function ξ , and so we do provide its proof.

Proof of Proposition E.1. Since Π_d^{disc} is a dense subset of Π_d with respect to the L^1 norm (1.32), and $\pi \mapsto \mathcal{P}_\xi(\pi)$ is continuous by Proposition 2.13(b), we have $\inf_{\pi \in \Pi_d} \mathcal{P}_\xi(\pi) = \inf_{\pi \in \Pi_d^{\text{disc}}} \mathcal{P}_\xi(\pi)$. Therefore, to establish (E.1), it suffices to prove that for any path π of the form (1.29), we have

$$\lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} F_{N,\xi}(d, \varepsilon) \leq \mathcal{P}_\xi(\pi). \quad (\text{E.4})$$

To this end, let $Z_1, \dots, Z_N: \mathbb{N}^{s-1} \rightarrow \mathbb{R}^\kappa$ and $Y: \mathbb{N}^{s-1} \rightarrow \mathbb{R}$ be independent centered Gaussian processes with covariances given by (3.6), and define the following process on $\mathbb{N}^{s-1} \times \Sigma^N$:

$$\mathcal{H}_t(\alpha, \sigma) = \sqrt{t} H_{N,\xi}(\sigma) + \sqrt{1-t} \sum_{i=1}^N \langle Z_i(\alpha), \sigma_i \rangle + \sqrt{t} \sqrt{N} Y(\alpha), \quad t \in [0, 1]. \quad (\text{E.5})$$

We assume that $H_{N,\xi}$ is independent of Z_1, \dots, Z_N, Y , and that all of these Gaussian processes are independent of the RPC weights $(\nu_\alpha)_{\alpha \in \mathbb{N}^{s-1}}$ associated with (1.29b). Given $\varepsilon > 0$, consider the associated constrained free energy

$$\phi_N(t) = \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in \Sigma^N(d, \varepsilon)} \nu_\alpha \exp \mathcal{H}_t(\alpha, \sigma). \quad (\text{E.6})$$

Claim E.3. *For any $N \geq 1$ and $t \in (0, 1)$, the derivative of ϕ_N satisfies $\phi'_N(t) \leq C\varepsilon$ for some constant C not depending N, d, ε nor t .*

Proof. Denote by $\langle \cdot \rangle_t$ the average with respect to the following probability measure on $\mathbb{N}^{s-1} \times \Sigma^N(d, \varepsilon)$:

$$G_t(\alpha, \sigma) \propto \nu_\alpha \exp \mathcal{H}_t(\alpha, \sigma).$$

Differentiation of (E.6) results in

$$\phi'_N(t) = \frac{1}{N} \left\langle \frac{\partial \mathcal{H}_t(\alpha, \sigma)}{\partial t} \right\rangle_t.$$

Applying Gaussian integration by parts [54, Lem. 1.1], we rewrite this as

$$\phi'_N(t) = \mathbb{E} \left[\mathcal{C}((\alpha^1, \sigma^1), (\alpha^1, \sigma^1)) - \mathcal{C}((\alpha^1, \sigma^1), (\alpha^2, \sigma^2)) \right]_t, \quad (\text{E.7})$$

where $(\alpha^1, \sigma^1), (\alpha^2, \sigma^2)$ denote independent samples from G_t , and $\mathcal{C}: (\mathbb{N}^{s-1} \times \Sigma^N(d, \varepsilon))^2 \rightarrow \mathbb{R}$ is defined by

$$\mathcal{C}((\alpha, \sigma), (\alpha', \sigma')) = \frac{1}{N} \mathbb{E} \left[\frac{\partial \mathcal{H}_t(\alpha, \sigma)}{\partial t} \mathcal{H}_t(\alpha', \sigma') \right].$$

To compute this expectation, recall that the three Gaussian processes on the right-hand side of (E.5) are independent, centered, and have covariances given by (1.14) and (3.6). Therefore,

$$\begin{aligned} \mathcal{C}((\alpha^1, \sigma^1), (\alpha^2, \sigma^2)) &= \frac{1}{2} \left(\xi(\mathcal{R}_{1,2}) - \langle \nabla \xi(\gamma_{r(\alpha^1, \alpha^2)}), \mathcal{R}_{1,2} \rangle + \vartheta_\xi(\gamma_{r(\alpha^1, \alpha^2)}) \right) \\ &\stackrel{(1.27)}{=} \frac{1}{2} \left(\xi(\mathcal{R}_{1,2}) - \xi(\gamma_{r(\alpha^1, \alpha^2)}) - \langle \nabla \xi(\gamma_{r(\alpha^1, \alpha^2)}), \mathcal{R}_{1,2} \right. \\ &\quad \left. - \gamma_{r(\alpha^1, \alpha^2)} \rangle \right), \end{aligned}$$

where $\mathcal{R}_{1,2} = R(\sigma^1, \sigma^2)$ as in (2.2). By the convexity assumption (A2), we have

$$\mathcal{C}((\alpha^1, \sigma^1), (\alpha^2, \sigma^2)) \geq 0. \quad (\text{E.8})$$

In the special case $(\alpha^1, \sigma^1) = (\alpha^2, \sigma^2)$, we have $\gamma_{r(\alpha^1, \alpha^1)} = \gamma_s = \text{diag}(d)$ by (1.29c), and $\|\mathcal{R}_{1,1} - \text{diag}(d)\|_1 \leq \kappa \varepsilon$ by definition of $\Sigma^N(d, \varepsilon)$ in (D.1). It follows that

$$\mathcal{C}((\alpha^1, \sigma^1), (\alpha^1, \sigma^1)) \leq \frac{1}{2} \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R)\|_\infty \cdot \kappa \varepsilon. \quad (\text{E.9})$$

Combining (E.7), (E.8), and (E.9) concludes the proof. \square (Claim)

Claim E.3 implies

$$\phi_N(1) \leq \phi_N(0) + C \varepsilon. \quad (\text{E.10})$$

When $t = 1$, the α and σ terms in (E.6) fully decouple, resulting in

$$\phi_N(1) = F_{N,\xi}(d, \varepsilon) + \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \nu_\alpha \exp(\sqrt{N} Y_\alpha) \stackrel{(\text{C.4})}{=} F_{N,\xi}(d, \varepsilon) - \mathcal{P}_\xi^{(2)}(\pi). \quad (\text{E.11})$$

On the other hand, evaluating (E.6) at $t = 0$ yields

$$\phi_N(0) = \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in \Sigma^N(d, \varepsilon)} \nu_\alpha \exp \left(\sum_{i=1}^N \langle Z_i(\alpha), \sigma_i \rangle \right). \quad (\text{E.12})$$

For any $\sigma \in \Sigma^N(d, \varepsilon)$ and $\lambda \in \mathbb{R}^\kappa$, we have

$$\frac{1}{N} \sum_{i=1}^N \langle \lambda, \sigma_i \rangle = \langle \lambda, d \rangle + \left\langle \lambda, \frac{1}{N} \sum_{i=1}^N \sigma_i - d \right\rangle \stackrel{(1.4)}{\geq} \langle \lambda, d \rangle - \varepsilon \|\lambda\|_1. \quad (\text{E.13})$$

Now add and subtract $\langle \lambda, \sigma_i \rangle$ within the exponent appearing in (E.12). Using (E.13) and the trivial containment $\Sigma^N(d, \varepsilon) \subseteq \Sigma^N$, we deduce that

$$\begin{aligned} \phi_N(0) &\leq \frac{1}{N} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^{s-1}} \sum_{\sigma \in \Sigma^N} \nu_\alpha \exp \left(\sum_{i=1}^N \langle Z_i(\alpha) + \lambda, \sigma_i \rangle \right) - \langle \lambda, d \rangle + \varepsilon \|\lambda\|_1 \\ &\stackrel{(3.7)}{=} \mathcal{P}_{N,\xi}^{(1)}(\pi, \lambda; \Sigma^N) - \langle \lambda, d \rangle + \varepsilon \|\lambda\|_1 \stackrel{(3.10)}{=} \mathcal{P}_\xi^{(1)}(\pi, \lambda) - \langle \lambda, d \rangle + \varepsilon \|\lambda\|_1. \end{aligned} \quad (\text{E.14})$$

Inserting (E.11) and (E.14) into (E.10), we arrive to

$$\begin{aligned} \mathbf{F}_{N,\xi}(d, \varepsilon) &\leq \mathcal{P}_\xi^{(1)}(\pi, \lambda) + \mathcal{P}_\xi^{(2)}(\pi) - \langle \lambda, d \rangle + (C + \|\lambda\|_1)\varepsilon \\ &\stackrel{(1.34)}{=} \mathcal{P}_\xi(\pi, \lambda) - \langle \lambda, d \rangle + (C + \|\lambda\|_1)\varepsilon. \end{aligned}$$

As this inequality holds for any N , we conclude

$$\lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} \mathbf{F}_{N,\xi}(d, \varepsilon) \leq \mathcal{P}_\xi(\pi, \lambda) - \langle \lambda, d \rangle. \quad (\text{E.15})$$

Finally, recall that $\mathcal{P}_\xi(\pi) = \inf_{\lambda \in \mathbb{R}^\kappa} [\mathcal{P}_\xi(\pi, \lambda) - \langle \lambda, d \rangle]$, and so (E.15) implies (E.4). \square

We are almost ready to prove Theorem 1.7. One technical detail that needs to be resolved is relaxing (E.2) to the weaker condition $d^N \rightarrow d$. The crucial lemma is the following analogue of Lemma D.1 for free energy.

Lemma E.4. *Assume ξ satisfies (A1). There exists a constant C depending only on ξ and κ , such that for every $N \geq 1$ and $\varepsilon > 0$, we have*

$$\sup_{d \in \mathcal{D}_N} |\mathbf{F}_{N,\xi}(d, \varepsilon) - \mathbf{F}_{N,\xi}(d)| \leq C\sqrt{\varepsilon}. \quad (\text{E.16})$$

Proof. As in the proof of Lemma D.1, assume without loss of generality that $\varepsilon \in [1/N, 1/2]$, and consider any map $P: \Sigma^N(d, \varepsilon) \rightarrow \Sigma^N(d)$ that is a projection with respect to Hamming distance.

Let $\tilde{H}_{N,\xi}$ be an independent copy of $H_{N,\xi}$, and then define an interpolating Hamiltonian on $\Sigma^N(d, \varepsilon)$:

$$\mathcal{H}_t(\sigma) = \sqrt{t} H_{N,\xi}(\sigma) + \sqrt{1-t} \tilde{H}_{N,\xi}(P\sigma), \quad t \in [0, 1].$$

Define the associated free energy

$$\phi(t) = \frac{1}{N} \mathbb{E} \log \sum_{\sigma \in \Sigma^N(d, \varepsilon)} \exp \mathcal{H}_t(\sigma), \quad (\text{E.17})$$

so that

$$\phi(1) = \mathbf{F}_{N,\xi}(d, \varepsilon) \quad \text{and} \quad \phi(0) = \frac{1}{N} \mathbb{E} \log \sum_{\sigma \in \Sigma^N(d)} \mathcal{N}(\sigma) \exp \tilde{H}_{N,\xi}(\sigma), \quad (\text{E.18a})$$

where $\mathcal{N}(\sigma)$ is defined in (D.5). Since every $\sigma \in \Sigma^N(d)$ satisfies $P(\sigma) = \sigma$, we have $\mathcal{N}(\sigma) \geq 1$. Also $\mathcal{N}(\sigma) \leq \exp(C_\kappa \sqrt{\varepsilon} N)$ by (D.6), and so

$$\mathbf{F}_{N,\xi}(d) \leq \phi(0) \leq \mathbf{F}_{N,\xi}(d) + C_\kappa \sqrt{\varepsilon}. \quad (\text{E.18b})$$

Meanwhile, differentiation of (E.17) results in

$$\phi'(t) = \frac{1}{N} \left\langle \frac{\partial \mathcal{H}_t(\sigma)}{\partial t} \right\rangle_t,$$

where $\langle \cdot \rangle_t$ denotes expectation with respect to the Gibbs measure $G_t(\sigma) \propto \exp \mathcal{H}_t(\sigma)$ on $\Sigma^N(d, \varepsilon)$. Applying Gaussian integration by parts [54, Lem. 1.1], we rewrite this as

$$\phi'(t) = \mathbb{E} \langle \mathcal{C}(\sigma^1, \sigma^1) - \mathcal{C}(\sigma^1, \sigma^2) \rangle_t,$$

where σ^1, σ^2 denote independent samples from G_t , and $\mathcal{C}: \Sigma^N(d, \varepsilon) \times \Sigma^N(d, \varepsilon) \rightarrow \mathbb{R}$ is defined by

$$\mathcal{C}(\sigma, \sigma') = \frac{1}{N} \left[\frac{\partial \mathcal{H}_t(\sigma)}{\partial t} \mathcal{H}_t(\sigma') \right] \stackrel{(1.14)}{=} \frac{1}{2} \left(\xi(R(\sigma, \sigma')) - \xi(R(P\sigma, P\sigma')) \right).$$

It follows that

$$|\phi'(t)| \leq \sup_{\|R\|_1 \leq 1} \|\nabla \xi(R)\|_\infty \cdot \sup_{\sigma, \sigma' \in \Sigma^N(d, \varepsilon)} \|R(\sigma, \sigma') - R(P\sigma, P\sigma')\|_1.$$

For any $\sigma, \sigma' \in \Sigma^N(d, \varepsilon)$, we have

$$\begin{aligned} & \|R(\sigma, \sigma') - R(P\sigma, P\sigma')\|_1 \\ & \stackrel{(2.3)}{=} \frac{1}{N} \sum_{k, k'=1}^{\kappa} \left| \sum_{i=1}^N \left(\mathbb{1}\{\sigma_i = \mathbf{e}_k\} \mathbb{1}\{\sigma'_i = \mathbf{e}_{k'}\} - \mathbb{1}\{(P\sigma)_i = \mathbf{e}_k\} \mathbb{1}\{(P\sigma')_i = \mathbf{e}_{k'}\} \right) \right| \\ & \leq \frac{1}{N} \sum_{i=1}^N \sum_{k, k'=1}^{\kappa} \left| \mathbb{1}\{\sigma_i = \mathbf{e}_k\} \mathbb{1}\{\sigma'_i = \mathbf{e}_{k'}\} - \mathbb{1}\{(P\sigma)_i = \mathbf{e}_k\} \mathbb{1}\{(P\sigma')_i = \mathbf{e}_{k'}\} \right| \\ & \leq \frac{1}{N} \sum_{i=1}^N \sum_{k, k'=1}^{\kappa} \left(\left| \mathbb{1}\{\sigma_i = \mathbf{e}_k\} - \mathbb{1}\{(P\sigma)_i = \mathbf{e}_k\} \right| + \left| \mathbb{1}\{\sigma'_i = \mathbf{e}_{k'}\} - \mathbb{1}\{(P\sigma')_i = \mathbf{e}_{k'}\} \right| \right) \\ & \stackrel{(D.4)}{\leq} 4\varepsilon. \end{aligned}$$

The two previous displays together show that $\sup_{t \in [0, 1]} |\phi'(t)| \leq C' \varepsilon$ for some constant C' depending only on ξ and κ . Combining this fact with (E.18), we determine that

$$|\mathbf{F}_{N,\xi}(d, \varepsilon) - \mathbf{F}_{N,\xi}(d)| \leq C_\kappa \sqrt{\varepsilon} + C' \varepsilon.$$

We have thus proved (E.16) with $C = C_\kappa + C'$. \square

We now complete the main objective of this appendix.

Proof of Theorem 1.7. For any $d \in \mathcal{D}$ and $N \geq 1$, it follows from the definition of \mathcal{D}_N in (1.6) that there exists $d^N \in \mathcal{D}_N$ satisfying (E.2) with $L = 1$.

So for any fixed $\varepsilon > 0$, we have $\Sigma^N(d, \varepsilon) \supseteq \Sigma^N(d^N)$ once N is large enough that $\varepsilon \geq 1/N$. Hence

$$\liminf_{N \rightarrow \infty} F_{N,\xi}(d, \varepsilon) \geq \liminf_{N \rightarrow \infty} F_{N,\xi}(d^N) \stackrel{(E.3)}{\geq} \inf_{\pi \in \Pi_d} \mathcal{P}_\xi(\pi).$$

Combining this with the upper bound from Proposition E.1, we deduce (1.22a). For (1.22b), we consider any sequence $d^N \in \mathcal{D}_N$ such that $d^N \rightarrow d$ as $N \rightarrow \infty$, not necessarily satisfying (E.2). Once N is large enough that $\|d^N - d\|_\infty \leq \varepsilon/2$, we have

$$\Sigma^N(d, \varepsilon/2) \subseteq \Sigma^N(d^N, \varepsilon) \subseteq \Sigma^N(d, 2\varepsilon).$$

Hence $F_{N,\xi}(d, \varepsilon/2) \leq F_{N,\xi}(d^N, \varepsilon) \leq F_{N,\xi}(d, 2\varepsilon)$ for all large N , and so (1.22a) forces

$$\lim_{\varepsilon \searrow 0} \limsup_{N \rightarrow \infty} F_{N,\xi}(d^N, \varepsilon) = \lim_{\varepsilon \searrow 0} \liminf_{N \rightarrow \infty} F_{N,\xi}(d^N, \varepsilon) = \inf_{\pi \in \Pi_d} \mathcal{P}_\xi(\pi). \quad (E.19)$$

Meanwhile, by Lemma E.4 we have

$$\limsup_{N \rightarrow \infty} |F_{N,\xi}(d^N, \varepsilon) - F_{N,\xi}(d^N)| \leq C\sqrt{\varepsilon}.$$

By sending $\varepsilon \searrow 0$ and appealing to (E.19), we obtain (1.22b). \square

References

- Adhikari, A., Brennecke, C.: Free energy of the quantum Sherrington–Kirkpatrick spin-glass model with transverse field. *J. Math. Phys.* **61**(8), 083302 (2020)
- Aizenman, M., Sims, R., Starr, S.L.: Extended variational principle for the Sherrington–Kirkpatrick spin-glass model. *Phys. Rev. B* **68**, 214403 (2003)
- Aizenman, M., Sims, R., Starr, S.L.: Mean-field spin glass models from the cavity-ROSt perspective. *Contemp. Math.* **437**, 1–30 (2007)
- Auffinger, A., Chen, W.-K.: The Parisi formula has a unique minimizer. *Commun. Math. Phys.* **335**(3), 1429–1444 (2015)
- Auffinger, A., Chen, W.-K.: On concentration properties of disordered Hamiltonians. *Proc. Am. Math. Soc.* **146**(4), 1807–1815 (2018)
- Auffinger, A., Zhou, Y.: On properties of the spherical mixed vector p -spin model. *Stoch. Process. Appl.* **146**, 382–413 (2022)
- Barra, A., Contucci, P., Mingione, E., Tantari, D.: Multi-species mean field spin glasses. Rigorous results. *Ann. Henri Poincaré* **16**(3), 691–708 (2015)
- Bates, E., Sloman, L., Sohn, Y.: Replica symmetry breaking in multi-species Sherrington–Kirkpatrick model. *J. Stat. Phys.* **174**(2), 333–350 (2019)
- Bates, E., Sohn, Y.: Crisanti–Sommers formula and simultaneous symmetry breaking in multi-species spherical spin glasses. *Commun. Math. Phys.* **394**(3), 1101–1152 (2022)
- Bates, E., Sohn, Y.: Free energy in multi-species mixed p -spin spherical models. *Electron. J. Probab.* **27**, 52 (2022)
- Ben Arous, G., Subag, E., Zeitouni, O.: Geometry and temperature chaos in mixed spherical spin glasses at low temperature: the perturbative regime. *Commun. Pure Appl. Math.* **73**(8), 1732–1828 (2020)
- Binder, K., Young, A.P.: Spin glasses: experimental facts, theoretical concepts, and open questions. *Rev. Mod. Phys.* **58**, 801–976 (1986)
- Bovier, A., Klimovsky, A.: The Aizenman–Sims–Starr and Guerra’s schemes for the SK model with multidimensional spins. *Electron. J. Probab.* **14**(8), 161–241 (2009)
- Caltagirone, F., Parisi, G., Rizzo, T.: Dynamical critical exponents for the mean-field Potts glass. *Phys. Rev. E* **85**, 051504 (2012)
- Chatterjee, S.: The Ghirllanda–Guerra identities without averaging. Preprint, available at [arXiv:0911.4520](https://arxiv.org/abs/0911.4520)

16. Chen, H.-B.: On Parisi measures of Potts spin glasses with correction. Preprint, available at [arXiv:2311.11699](https://arxiv.org/abs/2311.11699)
17. Chen, H.-B.: On the self-overlap in vector spin glasses. Preprint, available at [arXiv:2311.09880](https://arxiv.org/abs/2311.09880)
18. Chen, H.-B.: Parisi PDE and convexity for vector spins. Preprint, available at [arXiv:2311.10446](https://arxiv.org/abs/2311.10446)
19. Chen, H.-B.: Self-overlap correction simplifies the Parisi formula for vector spins. *Electron. J. Probab.* **28**, 170 (2023)
20. Chen, H.-B., Mourrat, J.-C.: On the free energy of vector spin glasses with non-convex interactions. Preprint, available at [arXiv:2311.08980](https://arxiv.org/abs/2311.08980)
21. Chen, W.-K.: The Aizenman–Sims–Starr scheme and Parisi formula for mixed p -spin spherical models. *Electron. J. Probab.* **18**(94), 14 (2013)
22. Chen, W.-K.: Phase transition in the spiked random tensor with Rademacher prior. *Ann. Stat.* **47**(5), 2734–2756 (2019)
23. Chen, W.-K., Handschy, M., Lerman, G.: Phase transition in random tensors with multiple independent spikes. *Ann. Appl. Probab.* **31**(4), 1868–1913 (2021)
24. Chen, W.-K., Panchenko, D.: Temperature chaos in some spherical mixed p -spin models. *J. Stat. Phys.* **166**(5), 1151–1162 (2017)
25. Chen, W.-K., Sen, A.: On ℓ_p -Gaussian–Grothendieck problem. *Int. Math. Res. Not. IMRN* **3**, 2344–2428 (2023)
26. Contucci, P., Mingione, E.: A multi-scale spin-glass mean-field model. *Commun. Math. Phys.* **368**(3), 1323–1344 (2019)
27. Crisanti, A., Rizzo, T.: Analysis of the ∞ -replica symmetry breaking solution of the Sherrington–Kirkpatrick model. *Phys. Rev. E* **65**, 046137 (2002)
28. De Santis, E., Parisi, G., Ritort, F.: On the static and dynamical transition in the mean-field Potts glass. *J. Phys. A* **28**(11), 3025–3041 (1995)
29. Dey, P.S., Wu, Q.: Fluctuation results for multi-species Sherrington–Kirkpatrick model in the replica symmetric regime. *J. Stat. Phys.* **185**(3), 22 (2021)
30. Dominguez, T.: The ℓ^p -Gaussian–Grothendieck problem with vector spins. *Electron. J. Probab.* **27**, 70 (2022)
31. Elderfield, D., Sherrington, D.: The curious case of the Potts spin glass. *J. Phys. C Solid State Phys.* **16**(15), L497–L503 (1983)
32. Elderfield, D., Sherrington, D.: Novel non-ergodicity in the Potts spin glass. *J. Phys. C Solid State Phys.* **16**(32), L1169–L1175 (1983)
33. Gamarnik, D., Jagannath, A., Sen, S.: The overlap gap property in principal submatrix recovery. *Probab. Theory Relat. Fields* **181**(4), 757–814 (2021)
34. Ghirlanda, S., Guerra, F.: General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity. *J. Phys. A* **31**(46), 9149–9155 (1998)
35. Gross, D.J., Kanter, I., Sompolinsky, H.: Mean-field theory of the Potts glass. *Phys. Rev. Lett.* **55**, 304–307 (1985)
36. Guerra, F.: Broken replica symmetry bounds in the mean field spin glass model. *Commun. Math. Phys.* **233**(1), 1–12 (2003)
37. Husson, J., Ko, J.: Spherical integrals of sublinear rank. Preprint, available at [arXiv:2208.03642](https://arxiv.org/abs/2208.03642)
38. Jagannath, A., Ko, J., Sen, S.: Max κ -cut and the inhomogeneous Potts spin glass. *Ann. Appl. Probab.* **28**(3), 1536–1572 (2018)
39. Jagannath, A., Sen, S.: On the unbalanced cut problem and the generalized Sherrington–Kirkpatrick model. *Ann. Inst. Henri Poincaré D* **8**(1), 35–88 (2021)
40. Jagannath, A., Tobasco, I.: A dynamic programming approach to the Parisi functional. *Proc. Am. Math. Soc.* **144**(7), 3135–3150 (2016)
41. Ko, J.: The Crisanti–Sommers formula for spherical spin glasses with vector spins. Preprint, available at [arXiv:1911.04355](https://arxiv.org/abs/1911.04355)
42. Ko, J.: Free energy of multiple systems of spherical spin glasses with constrained overlaps. *Electron. J. Probab.* **25**, 28 (2020)
43. Mourrat, J.-C.: Hamilton–Jacobi equations for mean-field disordered systems. *Ann. Henri Lebesgue* **4**, 453–484 (2021)
44. Mourrat, J.-C.: Nonconvex interactions in mean-field spin glasses. *Probab. Math. Phys.* **2**(2), 281–339 (2021)
45. Mourrat, J.-C.: The Parisi formula is a Hamilton–Jacobi equation in Wasserstein space. *Can. J. Math.* **74**(3), 607–629 (2022)
46. Mourrat, J.-C.: Free energy upper bound for mean-field vector spin glasses. *Ann. Inst. Henri Poincaré Probab. Stat.* **59**(3), 1143–1182 (2023)
47. Mourrat, J.-C., Panchenko, D.: Extending the Parisi formula along a Hamilton–Jacobi equation. *Electron. J. Probab.* **25**, 23 (2020)

48. Nishimori, H., Stephen, M.J.: Gauge-invariant frustrated Potts spin-glass. *Phys. Rev. B* **27**, 5644–5652 (1983)
49. Panchenko, D.: Free energy in the generalized Sherrington–Kirkpatrick mean field model. *Rev. Math. Phys.* **17**(7), 793–857 (2005)
50. Panchenko, D.: A question about the Parisi functional. *Electron. Commun. Probab.* **10**, 155–166 (2005)
51. Panchenko, D.: On differentiability of the Parisi formula. *Electron. Commun. Probab.* **13**, 241–247 (2008)
52. Panchenko, D.: The Ghirlanda–Guerra identities for mixed p -spin model. *C. R. Math. Acad. Sci. Paris* **348**(3–4), 189–192 (2010)
53. Panchenko, D.: The Parisi ultrametricity conjecture. *Ann. Math. (2)* **177**(1), 383–393 (2013)
54. Panchenko, D.: The Sherrington–Kirkpatrick Model. Springer Monographs in Mathematics, Springer, New York (2013)
55. Panchenko, D.: The Parisi formula for mixed p -spin models. *Ann. Probab.* **42**(3), 946–958 (2014)
56. Panchenko, D.: The free energy in a multi-species Sherrington–Kirkpatrick model. *Ann. Probab.* **43**(6), 3494–3513 (2015)
57. Panchenko, D.: Free energy in the mixed p -spin models with vector spins. *Ann. Probab.* **46**(2), 865–896 (2018)
58. Panchenko, D.: Free energy in the Potts spin glass. *Ann. Probab.* **46**(2), 829–864 (2018)
59. Panchenko, D., Talagrand, M.: On the overlap in the multiple spherical SK models. *Ann. Probab.* **35**(6), 2321–2355 (2007)
60. Parisi, G.: Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.* **43**, 1754–1756 (1979)
61. Parisi, G.: The order parameter for spin glasses: a function on the interval 0–1. *J. Phys. A Math. Gen.* **13**(3), 1101–1112 (1980)
62. Parisi, G.: A sequence of approximated solutions to the S–K model for spin glasses. *J. Phys. A Math. Gen.* **13**(4), L115–L121 (1980)
63. Parisi, G.: Order parameter for spin-glasses. *Phys. Rev. Lett.* **50**(24), 1946–1948 (1983)
64. Rockafellar, R.T.: *Convex Analysis*. Princeton Mathematical Series No. 28, Princeton University Press, Princeton (1970)
65. Rudin, W.: *Real and Complex Analysis*, 3rd edn. McGraw-Hill Book Co., New York (1987)
66. Schmidt, M.J.: Replica symmetry breaking at low temperatures. PhD thesis, Universität Würzburg (2008)
67. Sen, S.: Optimization on sparse random hypergraphs and spin glasses. *Random Struct. Algorithms* **53**(3), 504–536 (2018)
68. Sherrington, D., Kirkpatrick, S.: Solvable model of a spin-glass. *Phys. Rev. Lett.* **35**(26), 1792–1796 (1975)
69. Talagrand, M.: Free energy of the spherical mean field model. *Probab. Theory Relat. Fields* **134**(3), 339–382 (2006)
70. Talagrand, M.: The Parisi formula. *Ann. Math. (2)* **163**(1), 221–263 (2006)
71. Talagrand, M.: Construction of pure states in mean field models for spin glasses. *Probab. Theory Relat. Fields* **148**(3–4), 601–643 (2010)
72. Talagrand, M.: Mean Field Models for Spin Glasses. Volume II, vol. 55 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Advanced Replica-Symmetry and Low Temperature. Springer, Heidelberg, (2011)

Communicated by J. Ding