The Parisi ultrametricity conjecture

By Dmitry Panchenko

Abstract

In this paper we prove that the support of a random measure on the unit ball of a separable Hilbert space that satisfies the Ghirlanda-Guerra identities must be ultrametric with probability one. This implies the Parisi ultrametricity conjecture in mean-field spin glass models, such as the Sherrington-Kirkpatrick and mixed p-spin models, for which Gibbs measures are known to satisfy the Ghirlanda-Guerra identities in the thermodynamic limit.

1. Introduction and main result

Let us consider a random probability measure G on the unit ball of a separable Hilbert space H. We will denote by $(\sigma^l)_{l \geq 1}$ an independent and identically distributed sample from this measure, by $\langle \cdot \rangle$ the average with respect to $G^{\otimes \infty}$ and by $\mathbb{E}$ the expectation with respect to the randomness of G. Let $R_{l,l'} = \sigma^l \cdot \sigma^{l'}$ be the scalar product, or overlap, of σ^l and $\sigma^{l'}$. The random measure G is said to satisfy the Ghirlanda-Guerra identities if for any $n \geq 2$, any bounded measurable function f of the overlaps $(R_{l,l'})_{l,l' \leq n}$ and any bounded measurable function ψ of one overlap,

$$(1.1) \qquad \mathbb{E}\big\langle f\psi(R_{1,n+1})\big\rangle = \frac{1}{n}\mathbb{E}\big\langle f\big\rangle\mathbb{E}\big\langle \psi(R_{1,2})\big\rangle + \frac{1}{n}\sum_{l=2}^{n}\mathbb{E}\big\langle f\psi(R_{1,l})\big\rangle.$$

Another way to express the Ghirlanda-Guerra identities is to say that, under the measure $\mathbb{E}G^{\otimes \infty}$, conditionally on $R^n = (R_{l,l'})_{l,l' \leq n}$ the distribution of $R_{1,n+1}$ is given by the mixture

$$(1.2) \qquad \frac{1}{n}\mu + \frac{1}{n}\sum_{l=2}^{n}\delta_{R_{1,l}},$$

where μ is the distribution of one overlap $R_{1,2}$ under $\mathbb{E}G^{\otimes 2}$. We will prove the following.

THEOREM 1. *Under* (1.1), *the distribution of* $(R_{l,l'})_{l,l' \geq 1}$ *is ultrametric*; *i.e.,*

$$\text{(1.3)} \qquad\qquad \mathbb{E}\big\langle I\big(R_{1,2} \geq \min(R_{1,3}, R_{2,3})\big)\big\rangle = 1.$$

It is known (Theorem 2 in [9]) that if G satisfies the Ghirlanda-Guerra identities and if q^* is the supremum of the support of μ, then with probability one the support of G belongs to the sphere of radius $\sqrt{q^*}$ in H. Therefore, (1.3) means that with probability one over the choice of the random measure G, the distances in the Hilbert space H between three independent replicas σ^1, σ^2 and σ^3 sampled from G must satisfy the ultrametric inequality

$$\text{(1.4)} \qquad\qquad \|\sigma^1 - \sigma^2\| \leq \max\big(\|\sigma^1 - \sigma^3\|, \|\sigma^2 - \sigma^3\|\big).$$

Examples of random measures satisfying the Ghirlanda-Guerra identities arise in several mean-field spin glass models, such as the Sherrington-Kirkpatrick model [16] and mixed p-spin models, for which measures G are defined as the asymptotic analogues of the Gibbs measures in the thermodynamic limit by way of the Dovbysh-Sudakov representation [3]. Originally, the Ghirlanda-Guerra identities (1.1) were proved in [4] on average over the inverse temperature parameters and later, in a closely related formulation, by introducing a small perturbation term to the Hamiltonian of the model ([17], [20], [21]), but in some cases it can be proved in a strong sense without perturbation ([10]).

The ultrametric structure of the overlap array $(R_{l,l'})$ appeared implicitly in the original work of G. Parisi in [14], [15], in which the famous Parisi formula for the free energy in the Sherrington-Kirkpatrick model was discovered. The fact that the particular form of the array $(R_{l,l'})$ suggested in [14], [15] encoded some definite physical properties of the Gibbs measure, including ultrametricity, was found during the subsequent interpretation of the Parisi solution in the work of M. Mézard, G. Parisi, N. Sourlas, G. Toulouse and M.A. Virasoro in [6], [7]; see [8] for more details. The Parisi formula for the free energy was proved rigorously in a celebrated work of M. Talagrand in [18] following the break-through invention of the replica symmetry breaking interpolation scheme by F. Guerra in [5], which gave a very strong indirect support to the entire Parisi ansatz including the ultrametricity conjecture. More recently, several results providing some direct mathematical support to the ultrametricity conjecture were proved under an additional technical assumption that the overlaps take only finitely many values, i.e., $R_{1,2} \in \{q_1, \dots, q_k\}$ with probability one for some nonrandom values $(q_l)_{l \leq k}$. The first such result was proved by L.-P. Arguin and M. Aizenman in [2] as a consequence of the Aizenman-Contucci stochastic stability property [1] of the Gibbs measures in the mixed p-spin models. Inspired by [2], the author proved a similar result based on the Ghirlanda-Guerra identities in [9] (see [11] for an elementary proof) and M. Talagrand gave a different

proof in [19]. Unfortunately, in the Sherrington-Kirkpatrick and mixed p-spin models one expects the distribution of the overlap to have a continuous component ([8]), so the results in [2], [9] and [19] were not directly applicable to these models.

In this paper we deduce ultrametricity (1.3) without any assumptions on the distribution of the overlap and, as a result, one can now give a more direct approach to the Parisi formula for the free energy in the mixed p-spin models (see [13]). The proof of Theorem 1 utilizes a new representation of the Ghirlanda-Guerra identities that appears in Theorem 2 below, which can be viewed as a new invariance principle for random measures that satisfy (1.1). The idea behind this representation was originally motivated by the stability property proved in [12], which unified the Aizenman-Contucci stochastic stability and the Ghirlanda-Guerra identities; however, the proof we give here is based only on the Ghirlanda-Guerra identities.

2. **Invariance principles**

In this section, we will first prove a new invariance property for random measures that satisfy the Ghirlanda-Guerra identities in Theorem 2 and then deduce from it a modified version of the invariance principle in Theorem 3, which will be used in the proof of Theorem 1 in Section 3. Given $n \geq 1$, consider n bounded measurable functions $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$ and define

$$(2.1) \qquad F(\sigma, \sigma^1, \ldots, \sigma^n) = f_1(\sigma \cdot \sigma^1) + \cdots + f_n(\sigma \cdot \sigma^n).$$

For $1 \leq l \leq n$, we define

$$(2.2) \qquad F_l(\sigma, \sigma^1, \ldots, \sigma^n) = F(\sigma, \sigma^1, \ldots, \sigma^n) - f_l(\sigma \cdot \sigma^l) + \mathbb{E}\langle f_l(R_{1,2}) \rangle,$$

and for $l \geq n + 1$, we define

$$(2.3) \qquad F_l(\sigma, \sigma^1, \ldots, \sigma^n) = F(\sigma, \sigma^1, \ldots, \sigma^n).$$

The definition (2.3) for $l \geq n+1$ will not be used in the statement but will appear in the proof of the next result. Let us recall the notation $R^n = (R_{l,l'})_{l,l' \leq n}$.

THEOREM 2. *Suppose* (1.1) *holds, and let* Φ *be a bounded measurable function of* R^n. *Then*

$$(2.4) \qquad \mathbb{E}\langle \Phi \rangle = \mathbb{E}\Big\langle \frac{\Phi \exp \sum_{l=1}^n F_l(\sigma^l, \sigma^1, \ldots, \sigma^n)}{\langle \exp F(\sigma, \sigma^1, \ldots, \sigma^n) \rangle_-^n} \Big\rangle,$$

where the average $\langle \cdot \rangle_-$ *in the denominator is in* σ *only for fixed* $\sigma^1, \ldots, \sigma^n$ *and the outside average of the ratio is in* $\sigma^1, \ldots, \sigma^n$.

When $n = 1$, it is understood that Φ is a constant. Notice that one can easily recover the original Ghirlanda-Guerra identities from (2.4) by taking $f_1 = t\psi$ and $f_2 = \cdots = f_n = 0$ and computing the derivative at $t = 0$.

Proof. Without loss of generality, let us assume that Φ takes values in $[0, 1]$ and suppose that $|f_l| \leq L$ for $1 \leq l \leq n$ for some large enough L. For $t \geq 0$, let

$$(2.5) \qquad \varphi(t) = \mathbb{E}\Big\langle \frac{\Phi \exp \sum_{l=1}^{n} tF_l(\sigma^l, \sigma^1, \dots, \sigma^n)}{\langle \exp tF(\sigma, \sigma^1, \dots, \sigma^n)\rangle_-^n} \Big\rangle.$$

We will show that the Ghirlanda-Guerra identities (1.1) imply that this function is constant, thus, proving the statement of the theorem, $\varphi(0) = \varphi(1)$. If for $k \geq 1$ we denote

$$D_{n+k} = \sum_{l=1}^{n+k-1} F_l(\sigma^l, \sigma^1, \dots, \sigma^n) - (n+k-1)F_{n+k}(\sigma^{n+k}, \sigma^1, \dots, \sigma^n),$$

then one can easily compute by induction that (recall (2.3) and that we average in σ only in the denominator of (2.4))

$$\varphi^{(k)}(t) = \mathbb{E}\Big\langle \frac{\Phi D_{n+1} \dots D_{n+k} \exp \sum_{l=1}^{n+k} tF_l(\sigma^l, \sigma^1, \dots, \sigma^n)}{\langle \exp tF(\sigma, \sigma^1, \dots, \sigma^n)\rangle_-^{n+k}} \Big\rangle.$$

First, we note that $\varphi^{(k)}(0) = 0$. Indeed, if we denote $\Phi' = \Phi D_{n+1} \cdots D_{n+k-1}$, then Φ' is a function of the overlaps $(R_{l,l'})_{l,l' \leq n+k-1}$ and

(2.6)

$$\varphi^{(k)}(0) = \mathbb{E}\Big\langle \Phi'\Big(\sum_{l=1}^{n+k-1} F_l(\sigma^l, \sigma^1, \dots, \sigma^n) - (n+k-1)F_{n+k}(\sigma^{n+k}, \sigma^1, \dots, \sigma^n)\Big)\Big\rangle$$

$$= \sum_{j=1}^{n} \mathbb{E}\Big\langle \Phi'\Big(\sum_{l\neq j, l=1}^{n+k-1} f_j(R_{j,l}) + \mathbb{E}\langle f_j(R_{1,2})\rangle - (n+k-1)f_j(R_{j,n+k})\Big)\Big\rangle = 0$$

by the Ghirlanda-Guerra identities (1.1) applied to each term j. Now, since $|F_l| \leq Ln$ and $|D_{n+k}| \leq 2L(n+k-1)n$, we get

$$|\varphi^{(k)}(t)| \leq \Big(\prod_{l=1}^{k} 2L(n+l-1)n \Big) \mathbb{E}\Big\langle \frac{\Phi \exp \sum_{l=1}^{n+k} tF_l(\sigma^l, \sigma^1, \dots, \sigma^n)}{\langle \exp tF(\sigma, \sigma^1, \dots, \sigma^n)\rangle_-^{n+k}} \Big\rangle$$

$$= \Big(\prod_{l=1}^{k} 2L(n+l-1)n \Big) \mathbb{E}\Big\langle \frac{\Phi \exp \sum_{l=1}^{n} tF_l(\sigma^l, \sigma^1, \dots, \sigma^n)}{\langle \exp tF(\sigma, \sigma^1, \dots, \sigma^n)\rangle_-^{n}} \Big\rangle$$

$$= \prod_{l=1}^{k}(n+l-1)\,(2Ln)^k\,\varphi(t).$$

Consider arbitrary $T > 0$. Again, using that $|F_l| \leq Ln$ it is obvious that $\varphi(t) \leq e^{2LTn^2}$ for $0 \leq t \leq T$ and, therefore,

$$|\varphi^{(k)}(t)| \leq e^{2LTn^2} \frac{(n+k-1)!}{(n-1)!}(2Ln)^k.$$

By (2.6) and Taylor's expansion,

$$|\varphi(t) - \varphi(0)| \leq \max_{0 \leq s \leq t} \frac{|\varphi^{(k)}(s)|}{k!} t^k \leq e^{2LTn^2} \frac{(n+k-1)!}{k!\,(n-1)!}(2Lnt)^k.$$

Letting $k \to \infty$, we get that $\varphi(t) = \varphi(0)$ for $t < (2Ln)^{-1}$. Therefore, for any $t_0 < (2Ln)^{-1}$, we again have $\varphi^{(k)}(t_0) = 0$ for all $k \geq 1$, and by Taylor's expansion for $t_0 \leq t \leq T$,

$$|\varphi(t) - \varphi(t_0)| \leq \max_{t_0 \leq s \leq t} \frac{|\varphi^{(k)}(s)|}{k!}(t-t_0)^k \leq e^{2LTn^2} \frac{(n+k-1)!}{k!\,(n-1)!}(2Ln(t-t_0))^k.$$

Letting $k \to \infty$ proves that $\varphi(t) = \varphi(0)$ for $0 \leq t < 2(2Ln)^{-1}$. We can continue in the same fashion to prove this equality for all $0 \leq t < T$; note that T was arbitrary. $\qquad\square$

Let us write down a corollary of Theorem 2 on which the proof of Theorem 1 will be based. Consider a finite index set $\mathcal{A}$. Given $n \geq 1$ and configurations $\sigma^1, \ldots, \sigma^n$, let $(B_\alpha)_{\alpha \in \mathcal{A}}$ be a partition of the Hilbert space H such that for each $\alpha \in \mathcal{A}$, the indicator $I_{B_\alpha} = I(\sigma \in B_\alpha)$ is a measurable function of R^n and $(\sigma \cdot \sigma^l)_{l \leq n}$, and let

$$(2.7) \qquad\qquad W_\alpha = W_\alpha(\sigma^1, \ldots, \sigma^n) = G(B_\alpha).$$

Let us define a map T by

$$(2.8) \qquad W = (W_\alpha)_{\alpha \in \mathcal{A}} \to T(W) = \Big(\frac{\langle I_{B_\alpha} \exp F(\sigma, \sigma^1, \ldots, \sigma^n)\rangle_-}{\langle \exp F(\sigma, \sigma^1, \ldots, \sigma^n)\rangle_-} \Big)_{\alpha \in \mathcal{A}}.$$

The following holds.

THEOREM 3. *Under* (1.1), *for any bounded measurable function* $\varphi : \mathbb{R}^{n^2} \times \mathbb{R}^{|\mathcal{A}|} \to \mathbb{R}$,

$$(2.9) \qquad \mathbb{E}\big\langle \varphi(R^n, W)\big\rangle = \mathbb{E}\Big\langle \frac{\varphi(R^n, T(W)) \exp \sum_{l=1}^n F_l(\sigma^l, \sigma^1, \ldots, \sigma^n)}{\langle \exp F(\sigma, \sigma^1, \ldots, \sigma^n)\rangle_-^n}\Big\rangle.$$

Proof. For each $\alpha \in \mathcal{A}$, let us take integer $n_\alpha \geq 0$ and let $m = n + \sum_{\alpha \in \mathcal{A}} n_\alpha$. Let $(S_\alpha)_{\alpha \in \mathcal{A}}$ be any partition of $\{n+1, \ldots, m\}$ such that $|S_\alpha| = n_\alpha$. Consider a continuous function $\Phi : \mathbb{R}^{n^2} \to \mathbb{R}$, and let $\Phi' = \Phi(R^n) \prod_{\alpha \in \mathcal{A}} \varphi_\alpha$, where

$$(2.10) \qquad\qquad \varphi_\alpha = I\big(\sigma^l \in B_\alpha, \forall l \in S_\alpha\big).$$

Let f_l for $l \leq n$ be as in (2.1) and $f_{n+1} = \cdots = f_m = 0$. Let us now apply Theorem 2 with these choices of functions Φ' and f_l. First of all, integrating out the coordinates $(\sigma^l)_{l>n}$, the left-hand side of (2.4) can be written as

$$(2.11) \qquad \mathbb{E}\langle \Phi' \rangle = \mathbb{E}\Big\langle \Phi(R^n) \prod_{\alpha \in \mathcal{A}} \varphi_\alpha \Big\rangle = \mathbb{E}\Big\langle \Phi(R^n) \prod_{\alpha \in \mathcal{A}} W_\alpha^{n_\alpha}(\sigma^1, \ldots, \sigma^n) \Big\rangle,$$

where W_α's were defined in (2.7). Let us now compute the right-hand side of (2.4). Since $f_{n+1} = \cdots = f_m = 0$, the denominator will be $\langle \exp F(\sigma, \sigma^1, \ldots, \sigma^n) \rangle_-^m$ and

$$(2.12) \quad \sum_{l=1}^m F_l(\sigma^l, \sigma^1, \ldots, \sigma^m) = \sum_{l=1}^n F_l(\sigma^l, \sigma^1, \ldots, \sigma^n) + \sum_{l=n+1}^m F(\sigma^l, \sigma^1, \ldots, \sigma^n).$$

Since the denominator does not depend on $(\sigma^l)_{l>n}$, integrating in the coordinate σ^l for $l \in S_\alpha$ will produce a factor

$$\langle I_{B_\alpha} \exp F(\sigma, \sigma^1, \ldots, \sigma^n) \rangle_-.$$

For each $\alpha \in \mathcal{A}$, we have $|S_\alpha| = n_\alpha$ such coordinates. Therefore, the right-hand side of (2.4) is equal to

$$(2.13)$$
$$\mathbb{E}\Big\langle \frac{\Phi(R^n) \exp \sum_{l=1}^n F_l(\sigma^l, \sigma^1, \ldots, \sigma^n)}{\langle \exp F(\sigma, \sigma^1, \ldots, \sigma^n) \rangle_-^n} \prod_{\alpha \in \mathcal{A}} \Big(\frac{\langle I_{B_\alpha} \exp F(\sigma, \sigma^1, \ldots, \sigma^n) \rangle_-}{\langle \exp F(\sigma, \sigma^1, \ldots, \sigma^n) \rangle_-} \Big)^{n_\alpha} \Big\rangle.$$

Comparing with (2.11), recalling (2.8) and approximating a continuous function ϕ on $[0,1]^{|\mathcal{A}|}$ by polynomials we get (2.9) first for products $\Phi(R^n)\phi(W)$, then for continuous functions $\varphi(R^n, W)$ and then for arbitrary bounded measurable functions. $\qquad \square$

3. **Proof of Theorem 1**

We mentioned in the introduction that G is concentrated on the sphere of radius $\sqrt{q^*}$ so all σ below will be of length $\|\sigma\| = \sqrt{q^*}$. Consider a symmetric nonnegative definite matrix $A = (a_{l,l'})_{l,l' \leq n}$ such that $a_{l,l} = q^*$ for $l \leq n$. Given $\varepsilon > 0$, we will write $x \approx a$ to denote that $x \in (a - \varepsilon, a + \varepsilon)$ and $R^n \approx A$ to denote that $R_{l,l'} \approx a_{l,l'}$ for all $l \neq l' \leq n$. For simplicity of notation, we will keep the dependence of $\approx$ on ε implicit. Below, the matrix A will be used to describe a set of constraints such that the overlaps in R^n can take values close to A,

$$(3.1) \qquad\qquad\qquad \mathbb{E}\big\langle I(R^n \approx A) \big\rangle > 0,$$

for a given $\varepsilon > 0$. Let us introduce the notation

$$(3.2) \qquad\qquad\qquad a_n^* = \max(a_{1,n}, \ldots, a_{n-1,n}).$$

The main step in the proof of Theorem 1 is the following result, which will be based on the invariance principle of Theorem 3.

THEOREM 4. *Under* (1.1), *given* $\varepsilon > 0$, *if the matrix A satisfies* (3.1) *and* $a_n^* + \varepsilon < q^*$, *then*

(3.3)
$$\mathbb{E}\Big\langle I\Big(R^n \approx A, R_{1,n+1} \approx a_{1,n}, \ldots, R_{n-1,n+1} \approx a_{n-1,n}, R_{n,n+1} < a_n^* + \varepsilon\Big)\Big\rangle > 0.$$

Theorem 4 will be used in the following way. Suppose that the matrix A is such that $a_n^* < q^*$ and A is in the support of the distribution of R^n under $\mathbb{E}G^{\otimes\infty}$, which means that (3.1) holds for all $\varepsilon > 0$. Since $a_n^* + \varepsilon < q^*$ for small $\varepsilon > 0$, (3.3) holds for all $\varepsilon > 0$. Therefore, the support of the distribution of R^{n+1} under $\mathbb{E}G^{\otimes\infty}$ intersects the event in (3.3) for every $\varepsilon > 0$, and since the support is compact, it contains a point in the set

(3.4) $\Big\{ A' : a'_{l,l'} = a_{l,l'} \text{ for } l, l' \leq n, a'_{l,n+1} = a_{l,n} \text{ for } l \leq n-1, a'_{n,n+1} \leq a_n^* \Big\}.$

Proof of Theorem 4. We will prove (3.3) by contradiction, so suppose that the left-hand side is equal to zero. We will apply Theorem 3 with $\mathcal{A} = \{1, 2\}$ and the partition

$$B_1 = \Big\{ \sigma : \sigma \cdot \sigma^n \geq a_n^* + \varepsilon \Big\}, \; B_2 = B_1^c.$$

Since we assume that $a_n^* + \varepsilon < q^*$, the set B_1 contains a small neighborhood of σ^n, and on the event $\{R^n \approx A\}$ its complement $B_2 = B_1^c$ contains small neighborhoods of $\sigma^1, \ldots, \sigma^{n-1}$ since $R_{l,n} < a_{l,n} + \varepsilon \leq a_n^* + \varepsilon$. Thus, on this event and for $\sigma^1, \ldots, \sigma^n$ in the support of G, the weights $W_1 = G(B_1), W_2 = G(B_2) = 1 - W_1$ are strictly positive. Then, (3.1) implies that we can find $0 < p < p' < 1$ and small $\delta > 0$ such that

(3.5) $$\mathbb{E}\Big\langle I\Big(R^n \approx A, W_1 \in (p, p')\Big)\Big\rangle \geq \delta.$$

Let us apply Theorem 3 and (2.9) with the above partition, the choice of

(3.6) $$\varphi(R^n, W) = I\Big(R^n \approx A, W_1 \in (p, p')\Big),$$

and the choices of functions $f_1 = \cdots = f_{n-1} = 0$ and $f_n(x) = tI(x \geq a_n^* + \varepsilon)$ for $t \in \mathbb{R}$. The terms that appear on the right-hand side of (2.9) will become

$$\sum_{l=1}^{n} F_l(\sigma^l, \sigma^1, \ldots, \sigma^n) = \sum_{l=1}^{n-1} tI(R_{l,n} \geq a_n^* + \varepsilon) + t\mathbb{E}\Big\langle I(R_{1,2} \geq a_n^* + \varepsilon)\Big\rangle$$

$$= t\mathbb{E}\Big\langle I(R_{1,2} \geq a_n^* + \varepsilon)\Big\rangle =: t\gamma$$

since, again, on the event $\{R^n \approx A\}$ the overlaps $R_{l,n} < a_{l,n} + \varepsilon \leq a_n^* + \varepsilon$ for $l \leq n-1$ and

$$\Big\langle \exp F(\sigma, \sigma^1, \ldots, \sigma^n)\Big\rangle_- = \Big\langle \exp tI(\sigma \cdot \sigma^n \geq a_n^* + \varepsilon)\Big\rangle_- = \Delta_t(W),$$

where

(3.7) $$\Delta_t(W) = G(B_1)e^t + G(B_2) = W_1 e^t + 1 - W_1.$$

If $W = (W_1, W_2)$, the map $T_t(W)$ corresponding to (2.8) can now be written as

$$(3.8) \qquad\qquad T_t(W) = \Big(\frac{W_1 e^t}{\Delta_t(W)}, \frac{1 - W_1}{\Delta_t(W)} \Big).$$

Since $\Delta_t(W) \geq 1$ for $t \geq 0$, in this case equation (2.9) together with (3.5) implies

$$(3.9) \qquad\qquad \delta \leq \mathbb{E}\Big\langle \frac{I(R^n \approx A, (T_t(W))_1 \in (p, p')) \, e^{t\gamma}}{\Delta_t(W)^n} \Big\rangle$$

$$\leq \mathbb{E}\Big\langle I\big(R^n \approx A, (T_t(W))_1 \in (p, p')\big) \, e^{t\gamma} \Big\rangle.$$

In the average $\langle \cdot \rangle$ on the right-hand side let us fix $\sigma^1, \dots, \sigma^{n-1}$ and consider the average with respect to σ^n first. Clearly, on the event $\{R^n \approx A\}$ such average will be taken over the set

$$(3.10) \qquad \Omega(\sigma^1, \dots, \sigma^{n-1}) = \big\{ \sigma : \sigma \cdot \sigma^l \approx a_{l,n} \text{ for } l \leq n - 1 \big\}.$$

Let us look at the diameter of this set on the support of G. Suppose that with positive probability over the choice of the measure G and replicas $\sigma^1, \dots, \sigma^{n-1}$ from G satisfying the constraints in A (i.e., $R_{l,l'} \approx a_{l,l'}$ for $l, l' \leq n - 1$) we can find two points σ', σ'' in the support of G that belong to the set $\Omega(\sigma^1, \dots, \sigma^{n-1})$ and such that $\sigma' \cdot \sigma'' < a_n^* + \varepsilon$. This would then imply (3.3) since for (σ^n, σ^{n+1}) in a small neighborhood of (σ', σ''), the vector $(\sigma^1, \dots, \sigma^n, \sigma^{n+1})$ would belong to the event

$$\big\{ R^n \approx A, R_{1,n+1} \approx a_{1,n}, \dots, R_{n-1,n+1} \approx a_{n-1,n}, R_{n,n+1} < a_n^* + \varepsilon \big\}$$

on the left-hand side of (3.3). Since we assume that the left-hand side of (3.3) is equal to zero, we must have that for almost all choices of the measure G and replicas $\sigma^1, \dots, \sigma^{n-1}$ satisfying the constraints in A any two points σ', σ'' in the support of G that belong to the set $\Omega(\sigma^1, \dots, \sigma^{n-1})$ satisfy $\sigma' \cdot \sigma'' \geq a_n^* + \varepsilon$. Now, let us also recall that in (3.9) we are averaging over σ^n that satisfy the condition $(T_t(W))_1 \in (p, p')$. If we fix any such σ' in the support of G that satisfies this condition and belongs to the set (3.10), then the Gibbs average in σ^n will be taken over its neighborhood $B_1 = B_1(\sigma') = \{\sigma'' : \sigma' \cdot \sigma'' \geq a_n^* + \varepsilon\}$ of measure $W_1 = W_1(\sigma') = G(B_1(\sigma'))$ that satisfies $(T_t(W))_1 \in (p, p')$. One can easily check that the map in (3.8) satisfies $T_t^{-1} = T_{-t}$, and using this for $(T_t(W))_1 \in (p, p')$ implies that

$$W_1(\sigma') \in \Big\{ \frac{q e^{-t}}{q e^{-t} + 1 - q} : q \in (p, p') \Big\}$$

and, thus, $W_1(\sigma') \leq (1 - p')^{-1} e^{-t}$. This means that the average on the right-hand side of (3.9) over σ^n for fixed $\sigma^1, \dots, \sigma^{n-1}$ is bounded by $(1 - p')^{-1} e^{-t} e^{t\gamma}$

and, therefore, for $t \geq 0$,

(3.11) $\quad 0 < \delta \leq \mathbb{E}\langle I\big(R^n \approx A, (T_t(W))_1 \in (p, p')\big)e^{t\gamma}\rangle \leq (1 - p')^{-1}e^{-t(1-\gamma)}.$

Since A satisfies (3.1), $1 - \gamma = \mathbb{E}\langle I(R_{1,2} < a_n^* + \varepsilon)\rangle > 0$. Letting $t \to +\infty$ in (3.11), we arrive at contradiction. $\qquad\square$

Proof of Theorem 1. The proof is again by contradiction. Suppose that ultrametricity is violated, in which case there exist $a < b \leq c < q^*$ such that the matrix

(3.12)
$$\begin{pmatrix} q^* & a & b \\ a & q^* & c \\ b & c & q^* \end{pmatrix}$$

is in the support of the distribution of R^3 under $\mathbb{E}G^{\otimes\infty}$ so it satisfies (3.1) for every $\varepsilon > 0$. In this case Theorem 4 implies the following. Given any $n_1, n_2, n_3 \geq 1$ and $n = n_1 + n_2 + n_3$, we can find a matrix A in the support of the distribution of R^n under $\mathbb{E}G^{\otimes\infty}$ such that for some partition of indices $\{1, \ldots, n\} = I_1 \cup I_2 \cup I_3$ with $|I_j| = n_j$, we have $j \in I_j$ for $j \leq 3$ and

(a) $a_{l,l'} \leq c$ for all $l \neq l' \leq n$;
(b) $a_{l,l'} = a$ if $l \in I_1, l' \in I_2$, $a_{l,l'} = b$ if $l \in I_1, l' \in I_3$ and $a_{l,l'} = c$ if $l \in I_2, l' \in I_3$.

This can be proved by induction on n_1, n_2, n_3. First of all, by the choice of the matrix (3.12), this holds for $n_1 = n_2 = n_3 = 1$. Assuming the claim holds for some n_1, n_2 and n_3 with the matrix A, let us show how one can increase any of the n_j's by one. For example, let us assume for simplicity of notation that $n \in I_3$ and show that the claim holds with $n_3 + 1$. Since $a_n^* \leq c < q^*$, we can use the comment below Theorem 4 to find a matrix A' in the support of the distribution of R^{n+1} under $\mathbb{E}G^{\otimes\infty}$ that belongs to the set (3.4). Hence, $a'_{l,l'} \leq c$ for all $l \neq l' \leq n+1$ and $a'_{l,n+1} = a_{l,n}$ for $l \leq n-1$. Thus, in particular, $a'_{l,n+1} = b$ if $l \in I_1$ and $a'_{l,n+1} = c$ if $l \in I_2$, which means that A' satisfies the conditions (a), (b) with I_3 replaced by $I_3 \cup \{n + 1\}$. In a similar fashion, one can increase the cardinality of I_1 and I_2, which completes the induction. Now, let $n_1 = n_2 = n_3 = m$, find the matrix A as above, and find $\sigma^1, \ldots, \sigma^n$ on the sphere of radius $\sqrt{q^*}$ such that $R_{l,l'} = a_{l,l'}$ for all $l, l' \leq n$. Let $\bar{\sigma}^j$ be the barycenter of the set $\{\sigma^l : l \in I_j\}$. Condition (a) implies that

$$\|\bar{\sigma}^j\|^2 = \frac{1}{m^2}\sum_{l \in I_j}\|\sigma^l\|^2 + \frac{1}{m^2}\sum_{l \neq l' \in I_j}R_{l,l'} \leq \frac{mq^* + m(m-1)c}{m^2},$$

and condition (b) implies that $\bar{\sigma}^1 \cdot \bar{\sigma}^2 = a$, $\bar{\sigma}^1 \cdot \bar{\sigma}^3 = b$ and $\bar{\sigma}^2 \cdot \bar{\sigma}^3 = c$. Therefore,

$$\|\bar{\sigma}^2 - \bar{\sigma}^3\|^2 = \|\bar{\sigma}^2\|^2 + \|\bar{\sigma}^3\|^2 - 2\bar{\sigma}^2 \cdot \bar{\sigma}^3 \leq \frac{2(q^* - c)}{m}$$

and $0 < b - a = \bar{\sigma}^1 \cdot \bar{\sigma}^3 - \bar{\sigma}^1 \cdot \bar{\sigma}^2 \leq K m^{-1/2}$. Letting $m \to \infty$, we arrive at contradiction. $\square$

References

[1] M. AIZENMAN and P. CONTUCCI, On the stability of the quenched state in mean-field spin-glass models, *J. Statist. Phys.* **92** (1998), 765–783. MR 1657840. Zbl 0963.82045. http://dx.doi.org/10.1023/A:1023080223894.

[2] L.-P. ARGUIN and M. AIZENMAN, On the structure of quasi-stationary competing particle systems, *Ann. Probab.* **37** (2009), 1080–1113. MR 2537550. Zbl 1177.60050. http://dx.doi.org/10.1214/08-AOP429.

[3] L. N. DOVBYSH and V. N. SUDAKOV, Gram-de Finetti matrices, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **119** (1982), 77–86. MR 0666087. Zbl 0496.60018.

[4] S. GHIRLANDA and F. GUERRA, General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity, *J. Phys. A* **31** (1998), 9149–9155. MR 1662161. Zbl 0953.82037. http://dx.doi.org/10.1088/0305-4470/31/46/006.

[5] F. GUERRA, Broken replica symmetry bounds in the mean field spin glass model, *Comm. Math. Phys.* **233** (2003), 1–12. MR 1957729. Zbl 1013.82023. http://dx.doi.org/10.1007/s00220-002-0773-5.

[6] M. MÉZARD, G. PARISI, N. SOURLAS, G. TOULOUSE, and M. VIRASORO, On the nature of the spin-glass phase, *Phys. Rev. Lett.* **52** (1984), 1156. http://dx.doi.org/10.1103/PhysRevLett.52.1156.

[7] M. MÉZARD, G. PARISI, N. SOURLAS, G. TOULOUSE, and M. VIRASORO, Replica symmetry breaking and the nature of the spin glass phase, *J. Physique* **45** (1984), 843–854. MR 0746889. Zbl 0968.82528. http://dx.doi.org/10.1051/jphys:01984004505084300.

[8] M. MÉZARD, G. PARISI, and M. A. VIRASORO, *Spin Glass Theory and Beyond*, *World Scientific Lecture Notes in Physics* no. 9, World Scientific Publishing Co., Teaneck, NJ, 1987. MR 1026102. Zbl 0992.82500.

[9] D. PANCHENKO, A connection between the Ghirlanda-Guerra identities and ultrametricity, *Ann. Probab.* **38** (2010), 327–347. MR 2599202. Zbl 1196.60167. http://dx.doi.org/10.1214/09-AOP484.

[10] D. PANCHENKO, The Ghirlanda-Guerra identities for mixed p-spin model, *C. R. Math. Acad. Sci. Paris* **348** (2010), 189–192. MR 2600075. Zbl 1204.82036. http://dx.doi.org/10.1016/j.crma.2010.02.004.

[11] D. PANCHENKO, Ghirlanda-Guerra identities and ultrametricity: An elementary proof in the discrete case, *C. R. Math. Acad. Sci. Paris* **349** (2011), 813–816. MR 2825947. Zbl 1226.82030. http://dx.doi.org/10.1016/j.crma.2011.06.021.

[12] D. PANCHENKO, A unified stability property in spin glasses, *Comm. Math. Phys.* **313** (2012), 781–790. MR 2945621. Zbl 06065316. http://dx.doi.org/10.1007/s00220-012-1458-3.

[13] D. PANCHENKO, The Parisi formula for mixed p-spin models, 2011, to appear in *Ann. of Probab.*

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

[15] G. Parisi, A sequence of approximate solutions to the S-K model for spin glasses, *J. Phys. A.* **13** (1980), L–115. http://dx.doi.org/10.1088/0305-4470/13/4/009.

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

[17] M. Talagrand, *Spin Glasses*: *A Challenge for Mathematicians*, *Ergeb. Math. Grenzgeb.* **46**, Springer-Verlag, New York, 2003. MR 1993891. Zbl 1033.82002.

[18] M. Talagrand, The Parisi formula, *Ann. of Math.* **163** (2006), 221–263. MR 2195134. Zbl 1137.82010. http://dx.doi.org/10.4007/annals.2006.163.221.

[19] M. Talagrand, Construction of pure states in mean field models for spin glasses, *Probab. Theory Related Fields* **148** (2010), 601–643. MR 2678900. Zbl 1204.82037. http://dx.doi.org/10.1007/s00440-009-0242-6.

[20] M. Talagrand, *Mean Field Models for Spin Glasses. Volume I*: *Basic Examples*, *Ergeb. Math. Grenzgeb.* 3. *Folge A Series of Modern Surveys in Mathematics* **54**, Springer-Verlag, New York, 2011. MR 2731561. Zbl 1214.82002. http://dx.doi.org/10.1007/978-3-642-15202-3.

[21] M. Talagrand, *Mean Field Models for Spin Glasses. Volume II*: *Advanced Replica-Symmetry and Low Temperature*, *Ergeb. Math. Grenzgeb.* 3. *Folge A Series of Modern Surveys in Mathematics* **55**, 2011. Zbl 1232.82005. http://dx.doi.org/10.1007/978-3-642-22253-5.

Texas A&M University, College Station, TX
E-mail: panchenk@math.tamu.edu