Elect. Comm. in Probab. **12** (2007), 401–410

ELECTRONIC COMMUNICATIONS in PROBABILITY

A NOTE ON TALAGRAND'S POSITIVITY PRINCIPLE

DMITRY PANCHENKO[1]
Department of Mathematics: Texas A&M University, College Station, TX and Massachusetts Institute of Technology, Cambridge, MA.
email: panchenk@math.tamu.edu

Submitted July 15, 2007, *accepted in final form* October 17, 2003

AMS 2000 Subject classification: 60K35, 82B44
Keywords: Talagrand's positivity principle, Ghirlanda-Guerra identities.

Abstract

Talagrand's positivity principle, Section 6.6 in [9], states that one can slightly perturb a Hamiltonian in the Sherrington-Kirkpatrick model in such a way that the overlap of two configurations under the perturbed Gibbs' measure will become typically nonnegative. In this note we observe that abstracting from the setting of the SK model only improves the result and does not require any modifications in Talagrand's argument. In this version, for example, positivity principle immediately applies to the setting in [10]. Also, abstracting from the SK model improves the conditions in the Ghirlanda-Guerra identities and as a consequence results in a perturbation of smaller order necessary to ensure positivity of the overlap.

1 Introduction.

Let us consider a unit sphere $S = \{z \in \mathbb{R}^N : |z| = 1\}$ on euclidean space $\mathbb{R}^N$ and let ν be a probability measure on S. Given a measurable function $g : S \to \mathbb{R}$ let us define a probability measure ν_g on S by a change of density

$$\frac{d\nu_g}{d\nu} = \frac{e^{g(z)}}{\int e^{g(z)} d\nu(z)}.$$

We assume that the denominator on the right hand side is finite. Let us make a specific choice of $g(z)$ given by

$$g(z) = v \sum_{p\geq 1} 2^{-p} x_p \, g_p(z) \quad \text{for} \quad g_p(z) = \sum_{1\leq i_1,\ldots,i_p \leq N} g_{i_1,\ldots,i_p} z_{i_1} \cdots z_{i_p}, \tag{1}$$

where $v \geq 0$, x_p are i.i.d. random variables uniform on $[0, 1]$ and $g_{i_1,\ldots,i_p}$ are i.i.d. standard Gaussian for all $i_1, \ldots, i_p$ and all $p \geq 1$.

[1]RESEARCH PARTIALLY SUPPORTED BY NSF

Let us denote by $\mathbb{E}_g$ the expectation with respect to Gaussian random variables and by $\mathbb{E}_x$ the expectation with respect to uniform random variables in the definition of $g(z)$ in (1); the expectation in all randomness will be denoted by $\mathbb{E}$. The following is the main result of this note.

Theorem 1. (Talagrand's positivity principle) *For any $\varepsilon > 0$ there exists large enough v in (1) such that*

$$\mathbb{E}\nu_g^{\otimes 2}\big\{z^1 \cdot z^2 \le -\varepsilon\big\} \le \varepsilon. \tag{2}$$

The choice of v does not depend on N and ν.

This means that one can define a random perturbation ν_g of an arbitrary measure ν such that the scalar product $z^1 \cdot z^2$ of two vectors drawn independently from distribution ν_g will be typically nonnegative. This result was proved in Section 6.6 [9] in the setting of the Sherrington-Kirkpatrick model where ν was a random Gibbs' measure and in which case the expectation on the left hand side of (2) was also in the randomness of ν. The main ingredient of the proof was the Ghirlanda-Guerra identities that are typically given in the setting of the SK model as well.

The main contribution of this note is an observation that abstracting ourselves from the setting of the SK model results in some qualitative improvements of the positivity principle of Theorem 1. First of all, we notice that Talagrand's proof in [9] requires no modifications in order to prove a positivity principle that holds uniformly over ν rather than on average over a random Gibbs' measure in the SK model. This observation, for example, implies the result in [10] without any additional work as will be shown in Example below. Another important qualitative improvement is the fact that the choice of v in (2) is independent of N. In [9], one needed $v \gg N^{1/4}$ - a condition that appears in the proof of Ghirlanda-Guerra identities due to the fact that one controls random Gibbs' measure from the very beginning. We will show below that the supremum of $g(z)$ on the sphere is of order $v\sqrt{N}$ which means that one can perturb any measure on S by a change of density of order $\exp v\sqrt{N}$ and force the scalar product $z^1 \cdot z^2$ to be essentially nonnegative.

Example (Positivity in Guerra's replica symmetry breaking bound, [10]). The main result in [10] states that Guerra's replica symmetry breaking bound [4] applies to odd p-spin interactions as well. The proof utilizes the Aizenman-Sims-Starr version of Guerra's interpolation [1] and a positivity principle that requires a concentration inequality for the free energy along the interpolation. We observe that this positivity principle follows directly from Theorem 1. Let $\mathcal{A}$ be a countable set and let (w_α) be a probability function on $\mathcal{A}$ such that $w_\alpha \ge 0$, $\sum_{\alpha\in\mathcal{A}} w_\alpha = 1$. Let $H(z,\alpha)$ be a function on $\Omega = \Sigma \times \mathcal{A}$ for some finite subset Σ of S. Let us consider a probability measure on Ω given by

$$\mu\{(z,\alpha)\} \sim w_\alpha \exp\big(H(z,\alpha) + g(z)\big)$$

where $g(z)$ is defined in (1). Then its marginal on Σ is equal to ν_g if we define ν by $\nu\{z\} \sim \sum w_\alpha \exp H(z,\alpha)$. Therefore,

$$\mu^{\otimes 2}\big\{z^1 \cdot z^2 \le -\varepsilon\big\} = \nu_g^{\otimes 2}\big\{z^1 \cdot z^2 \le -\varepsilon\big\}.$$

By Theorem 1, for large enough $v > 0$,

$$\mathbb{E}\mu^{\otimes 2}\big\{z^1 \cdot z^2 \le -\varepsilon\big\} \le \varepsilon$$

and this inequality holds uniformly over all choices of H and w. Therefore, we can average over arbitrary random distribution of H and w. In particular, in [10], (w_α) was a random Derrida-Ruelle process on $\mathcal{A} = \mathbb{N}^k$ and H was Guerra's interpolating Hamiltonian in the form of Aizenman-Sims-Starr.

□

General remarks. (i) As we will see below, Talagrand's proof of positivity principle uses very deep information about a joint distribution of $z^1 \cdot z^l$ for $2 \leq l \leq n$ under measure $\mathbb{E}_g \nu_g^{\otimes n}$ for a typical realization of (x_p). However, it is not clear whether this deep information is really necessary to prove positivity and if some simpler argument would not suffice. For example, if we consider only a first order Gaussian term in (1) and define $g'(z) = v\sum_{i\leq N} g_i z_i$ then measure $\nu_{g'}$ under a change of density proportional to $e^{g'(z)}$ would favor a random direction $g = (g_1, \ldots, g_N)$ and it is conceivable that for large enough v independent of ν and N two independent vectors z^1 and z^2 from this measure would typically "point in the same direction", i.e.

$$\mathbb{E}\nu_{g'}^{\otimes 2}\{z^1 \cdot z^2 \leq -\varepsilon\} = \mathbb{E}\,\frac{\int I(z^1 \cdot z^2 \leq -\varepsilon) e^{v(g\cdot z^1 + g\cdot z^2)} d\nu(z^1) d\nu(z^2)}{\left(\int e^{vg\cdot z} d\nu(z)\right)^2} \leq \varepsilon.$$

In fact, even a weaker result with $v = o(\sqrt{N})$ would be sufficient for applications as in [10].

(ii) Theorem 1 implies the following non-random version of positivity principle.

Corollary 1. *For any $\varepsilon > 0$ there exists $v > 0$ large enough such that the following holds. For any distribution Q on the set of measures on the sphere S there exists a (non-random) function $g(z)$ such that for some absolute constant L,*

$$\sup_{z\in S} |g(z)| \leq Lv\sqrt{N} \quad \text{and} \quad \int \nu_g^{\otimes 2}\{z^1 \cdot z^2 \leq -\varepsilon\} dQ(\nu) \leq \varepsilon.$$

It would be of interest to prove this result directly and not as a corollary of Theorem 1. Then, by Hahn-Banach theorem, one can find a distribution P on the set of functions $\{\sup_{z\in S} |g(z)| \leq Lv\sqrt{N}\}$ such that for all probability measures ν on S,

$$\int \nu_g^{\otimes 2}\{z^1 \cdot z^2 \leq -\varepsilon\} dP(g) \leq \varepsilon.$$

This would give another proof of Theorem 1 with non constructive description of P.

□

Sketch of the proof of positivity principle. The main ingredient in Talagrand's proof is the extended Ghirlanda-Guerra identities (Section 6.4 in [9]) that state that if we sample $(z^1, \ldots, z^{n+1})$ from a measure $\mathbb{E}_g \nu_g^{\otimes(n+1)}$ (which is a mixture of product measures of ν_g over the randomness of Gaussian r.v.s) then for a typical realization of (x_p) the scalar product $z^1 \cdot z^{n+1}$ with probability $1/n$ is independent of $z^1 \cdot z^l, 2 \leq l \leq n$, and with probabilities $1/n$ it is equal to one of them. Given a function f on S^n, $\langle f \rangle$ will denote its average with respect to measure $\nu_g^{\otimes n}$ and we will use this notation for any n. The following holds.

Theorem 2. (Ghirlanda-Guerra identities) *For any measurable function f on S^n such that $|f| \leq 1$ and for any continuous function ψ on $[-1, 1]$ we have*

$$\mathbb{E}_x \Big| \mathbb{E}_g \langle f\psi(z^1 \cdot z^{n+1}) \rangle - \frac{1}{n}\mathbb{E}_g \langle f \rangle \mathbb{E}_g \langle \psi(z^1 \cdot z^2) \rangle - \frac{1}{n} \sum_{2\leq l\leq n} \mathbb{E}_g \langle f\psi(z^1 \cdot z^l) \rangle \Big| \leq \delta(\psi, n, v) \qquad (3)$$

where $\delta(\psi, n, v) \to 0$ as $v \to \infty$ and δ does not depend on N and ν.

The main idea of Talagrand's proof can be shortly described as follows. Suppose that we sample $z^1, \ldots, z^n$ independently from any measure on S. Then the event that all $z^1 \cdot z^l \leq -\varepsilon$ simultaneously is very unlikely and its probability is or order $1/(n\varepsilon)$. The bound on this probability is uniform over all measures and therefore can be averaged over some distribution on measures and holds, for example, for $\mathbb{E}_g \nu_g^{\otimes n}$. On the other hand, by Ghirlanda-Guerra identities under measure $\mathbb{E}_g \nu_g^{\otimes n}$ the events $\{z^1 \cdot z^l \leq -\varepsilon\}$ are strongly correlated due to the fact that with some prescribed probabilities $z^1 \cdot z^l$ can be equal. As a consequence, the simultaneous occurrence of these events will have probability of order a/n^{1-a} where a is the probability of one of them. But since this quantity is bounded by $1/(n\varepsilon)$, taking n large enough shows that a should be small enough.

□

2 Proofs.

The proof of Theorem 2 follows exactly the same argument as the proof of Ghirlanda-Guerra identities in [3] or in [9]. Since we consider a fixed measure ν, we do not need to control the fluctuations of a random Hamiltonian and, as a result, we get a better condition on v. The main part of the proof is the following lemma. For other generalizations see, for example, [2].

Lemma 1. *For any $p \geq 1$ there exists a constant L_p that depends on p only such that*

$$\mathbb{E}\big\langle \big| g_p(z) - \mathbb{E}_g \big\langle g_p(z) \big\rangle \big| \big\rangle \leq L_p \sqrt{v}. \tag{4}$$

Proof. Let us fix $(x_p)_{p\geq 1}$ in the definition of $g(z)$ and until the end of the proof let $\mathbb{E}$ denote the expectation in Gaussian random variables only. Define

$$\theta = \log \int e^{g(z)} d\nu(z), \quad \psi = \mathbb{E}\theta.$$

Given $p \geq 1$, let us think of θ and ψ as functions of $x = x_p$ only and define $v_p = v2^{-p}$. Then $\theta'(x) = v_p \big\langle g_p(z) \big\rangle$ and

$$\psi'(x) = v_p \mathbb{E}\big\langle g_p(z) \big\rangle = v_p^2 x \big(1 - \mathbb{E}\langle (z^1 \cdot z^2)^p \rangle\big) \leq v_p^2 x. \tag{5}$$

Since

$$\psi''(x) = v_p^2 \mathbb{E}\big(\big\langle g_p(z)^2 \big\rangle - \big\langle g_p(z) \big\rangle^2\big) = v_p^2 \mathbb{E}\big\langle \big(g_p(z) - \langle g_p(z) \rangle\big)^2 \big\rangle.$$

we have

$$v_p^2 \int_0^1 \mathbb{E}\big\langle \big(g_p(z) - \langle g_p(z) \rangle\big)^2 \big\rangle dx = \psi'(1) - \psi'(0) \leq v_p^2$$

and by Cauchy inequality

$$\int_0^1 \mathbb{E}\big\langle \big| g_p(z) - \langle g_p(z) \rangle \big| \big\rangle dx \leq 1. \tag{6}$$

To prove (4) is remains to approximate $\langle g_p(z) \rangle$ by $\mathbb{E}\langle g_p(z) \rangle$ and to achieve that we will use a simple consequence of convexity of θ and ψ given in the inequality (8) below. Since

$$\mathbb{E} g(z)^2 = v^2 \sum_{p\geq 1} 2^{-2p} x_p^2 \leq 2v^2$$

if all $|x_p| \le 2$, we can apply a well-known Gaussian concentration inequality, for completeness given in Lemma 3 in Appendix A, to get

$$\mathbb{E}|\theta(x) - \psi(x)| \le 4v \quad \text{and} \quad \mathbb{E}|\theta(x \pm y) - \psi(x \pm y)| \le 4v.$$

Below we will choose $|y| \le 1$ so that $|x \pm y| \le 2$. (*Remark.* At the same step in the setting of the SK model one also needs to control a random Hamiltonian which produces another term of order $\sqrt{N}$ and this results in unnecessary condition on v in the positivity principle.) Inequality (8) then implies that

$$\mathbb{E}|\theta'(x) - \psi'(x)| \le \psi'(x+y) - \psi'(x-y) + \frac{12v}{y}.$$

Since

$$\int_0^1 (\psi'(x+y) - \psi'(x-y))dx = \psi(1+y) - \psi(1-y) - \psi(y) + \psi(-y) \le 6v_p^2 y$$

using (5) and $|y| \le 1$ we get

$$\int_0^1 \mathbb{E}|\theta'(x) - \psi'(x)|dx \le 6v_p^2 y + \frac{12v}{y} \le L_p v^{3/2}$$

if we take $y = v^{-1/2} \le 1$. Using explicit expressions for θ' and ψ' we finally get

$$\int_0^1 \mathbb{E}|\langle g_p(z)\rangle - \mathbb{E}\langle g_p(z)\rangle|dx \le L_p\sqrt{v}. \tag{7}$$

Together with (6) this gives

$$\int_0^1 \mathbb{E}\big\langle |g_p(z) - \mathbb{E}\langle g_p(z)\rangle|\big\rangle dx \le L_p\sqrt{v}.$$

We now recall that $\mathbb{E}$ was expectation $\mathbb{E}_g$ with respect to Gaussian random variables only. Also, integral over $x = x_p \in [0,1]$ is nothing but expectation with respect to x_p. Therefore, averaging over all remaining x_p finishes the proof.

□

The following inequality was used in the previous proof and it quantifies the fact that if two convex functions are close to each other then their derivatives are also close.

Lemma 2. *If $\theta(x)$ and $\psi(x)$ are convex differentiable functions then*

$$\begin{aligned} |\theta'(x) - \psi'(x)| &\le \psi'(x+y) - \psi'(x-y) \\ &+ \frac{1}{y}\big(|\psi(x+y) - \theta(x+y)| + |\psi(x-y) - \theta(x-y)| + |\psi(x) - \theta(x)|\big). \end{aligned} \tag{8}$$

Proof. By convexity, for any $y > 0$

$$\frac{\theta(x) - \theta(x-y)}{y} \le \theta'(x) \le \frac{\theta(x+y) - \theta(x)}{y}$$

and

$$\frac{\psi(x) - \psi(x-y)}{y} \le \psi'(x) \le \frac{\psi(x+y) - \psi(x)}{y}.$$

If we define

$$U = \frac{1}{y}\Big(|\psi(x+y) - \theta(x+y)| + |\psi(x-y) - \theta(x-y)| + |\psi(x) - \theta(x)|\Big)$$

then the above inequalities imply

$$\begin{aligned}
\theta'(x) &\leq \frac{\theta(x+y)-\theta(x)}{y} \leq \frac{\psi(x+y)-\psi(x)}{y} + U \\
&= \frac{\psi(x)-\psi(x-y)}{y} + \frac{\psi(x+y)+\psi(x-y)-2\psi(x)}{y} + U \\
&\leq \psi'(x) + \psi'(x+y) - \psi'(x-y) + U.
\end{aligned}$$

Similarly,

$$\theta'(x) \geq \psi'(x) - (\psi'(x+y) - \psi'(x-y)) - U.$$

Combining two inequalities finishes the proof.

□

Proof of Theorem 2. Since $|f| \leq 1$ we can write

$$\Big|\mathbb{E}_g\big\langle f g_p(z)\big\rangle - \mathbb{E}_g\langle f\rangle\mathbb{E}\langle g_p(z)\rangle\Big| \leq \mathbb{E}_g\big\langle\big|g_p(z) - \mathbb{E}_g\langle g_p(z)\rangle\big|\big\rangle.$$

By Gaussian integration by parts, the left hand side is equal to $nv2^{-p}x_p\phi_p$ where

$$\phi_p = \Big|\mathbb{E}_g\big\langle f(z^1\cdot z^{n+1})^p\big\rangle - \frac{1}{n}\mathbb{E}_g\big\langle f\big\rangle\mathbb{E}_g\big\langle(z^1\cdot z^2)^p\big\rangle - \frac{1}{n}\sum_{2\leq l\leq n}\mathbb{E}_g\big\langle f(z^1\cdot z^l)^p\big\rangle\Big|.$$

Lemma 1 then implies $nv2^{-p}\mathbb{E}_x x_p\phi_p \leq L_p\sqrt{v}$ and, thus,

$$\mathbb{E}_x x_p\phi_p \leq Lv^{-1/2}$$

for some constant L that depends only on n and p. Since $\phi_p \leq 2$, for any $x_0 \in (0,1)$,

$$\mathbb{E}_x\phi_p \leq 2x_0 + \frac{1}{x_0}\mathbb{E}_x x_p\phi_p \leq 2x_0 + \frac{Lv^{-1/2}}{x_0}$$

and minimizing over x_0 we get $\mathbb{E}_x\phi_p(x_p) \leq Lv^{-1/4}$. Since any continuous function ψ on $[-1,1]$ can be approximated by polynomials this, obviously, implies the result.

□

Proof of Theorem 1. Step 1. First we use Ghirlanda-Guerra identities to give a lower bound on probability that all $z^1\cdot z^l \leq -\varepsilon$ for $2\leq l\leq n$. In order to use Theorem 2 it will be convenient to denote by δ any quantity that depends on (x_p) and such that $\mathbb{E}_x|\delta|$ does not depend on ν and N and $\mathbb{E}_x|\delta| \to 0$ as $v\to\infty$. Then (3) can be written as

$$\mathbb{E}_g\langle f\psi(z^1\cdot z^{n+1})\rangle - \frac{1}{n}\mathbb{E}_g\langle f\rangle\mathbb{E}_g\langle\psi(z^1\cdot z^2)\rangle - \frac{1}{n}\sum_{2\leq l\leq n}\mathbb{E}_g\langle f\psi(z^1\cdot z^l)\rangle = \delta.$$

Even though a function ψ is assumed to be continuous, let us use this result formally for $\psi(x) = I(x\leq -\varepsilon)$. The argument can be easily modified by using continuous approximations of the indicator function ψ. Let

$$f_n = \prod_{2\leq l\leq n}\psi(z^1\cdot z^l) = I(z^1\cdot z^l \leq -\varepsilon \text{ for } 2\leq l\leq n)$$

and let

$$a = \mathbb{E}_g \langle \psi(z^1 \cdot z^2) \rangle = \mathbb{E}_g \nu_g^{\otimes 2} \{ z^1 \cdot z^2 \le -\varepsilon \}.$$

Then by Ghirlanda-Guerra identities

$$\mathbb{E}_g \langle f_{n+1} \rangle = \mathbb{E}_g \langle f_n \psi(z^1 \cdot z^{n+1}) \rangle = \frac{1}{n} \mathbb{E}_g \langle f_n \rangle \mathbb{E}_g \langle \psi(z^1 \cdot z^2) \rangle + \frac{1}{n} \sum_{2 \le l \le n} \mathbb{E}_g \langle f_n \psi(z^1 \cdot z^l) \rangle + \delta$$

and since $f_n \psi(z^1 \cdot z^l) = f_n$ we get

$$\mathbb{E}_g \langle f_{n+1} \rangle = \frac{n-1+a}{n} \mathbb{E}_g \langle f_n \rangle + \delta.$$

By induction,

$$\mathbb{E}_g \langle f_n \rangle = a \prod_{2 \le l \le n-1} \frac{l-1+a}{l} + \delta \ge \frac{a}{L n^{1-a}} + \delta \tag{9}$$

where the last inequality follows from a simple estimate for $l \ge 2$

$$\frac{l-1+a}{l} = 1 - \frac{1-a}{l} \ge \exp\Bigl(-\frac{1-a}{l} - \frac{1}{l^2} \Bigr).$$

Step 2. On the other hand, we will show that $\langle f_n \rangle$ is of order $1/(n\varepsilon)$ and to emphasize the fact that this is true for any measure we now simply write G instead of ν_g. If $z^1, \ldots, z^n$ are i.i.d. from distribution G then

$$\langle f_n \rangle = \Bigl\langle G\bigl\{ z^2 : z^1 \cdot z^2 \le -\varepsilon \bigr\}^{n-1} \Bigr\rangle \le G(U) + \gamma^{n-1} \tag{10}$$

where, given $0 < \gamma < 1$, we defined a set

$$U = \Bigl\{ z^1 \in S : G\{ z^2 : z^1 \cdot z^2 \le -\varepsilon \} \ge \gamma \Bigr\}.$$

We would like to show that if γ is close to 1 then $G(U)$ is small. This follows from the fact that the average of $z^1 \cdot z^2$ is nonnegative with respect to any product measure,

$$\langle z^1 \cdot z^2 \rangle = \sum_{i \le N} \langle z_i^1 z_i^2 \rangle = \sum_{i \le N} \langle z_i \rangle^2 \ge 0.$$

Since $z^1 \cdot z^2 \le 1$,

$$0 \le \langle z^1 \cdot z^2 \rangle \le -\varepsilon G^{\otimes 2}\{ z^1 \cdot z^2 \le -\varepsilon \} + G^{\otimes 2}\{ z^1 \cdot z^2 > -\varepsilon \} = -\varepsilon + (1+\varepsilon) G^{\otimes 2}\{ z^1 \cdot z^2 > -\varepsilon \}$$

and

$$\varepsilon \le (1+\varepsilon) G^{\otimes 2}\{ z^1 \cdot z^2 > -\varepsilon \}. \tag{11}$$

Let us define a conditional distribution $G_U(C) = G(UC)/G(U)$. If $z^1 \in U$ then by definition $G\{ z^2 : z^1 \cdot z^2 > -\varepsilon \} \le 1 - \gamma$ and

$$G_U\{ z^2 : z^1 \cdot z^2 > -\varepsilon \} \le \frac{1-\gamma}{G(U)}.$$

Since G_U is concentrated on U,

$$G_U^{\otimes 2}\{ z^1 \cdot z^2 > -\varepsilon \} \le \frac{1-\gamma}{G(U)}.$$

Using (11) for G_U instead of G we get

$$\varepsilon \le (1+\varepsilon)G_U^{\otimes 2}\{z^1\cdot z^2 > -\varepsilon\} \le \frac{2(1-\gamma)}{G(U)}$$

and, therefore, $G(U)\le 2(1-\gamma)/\varepsilon$. Then, (10) implies that

$$\langle f_n\rangle \le \frac{2(1-\gamma)}{\varepsilon} + \gamma^{n-1} \le \frac{2(1-\gamma)}{\varepsilon} + e^{-(n-1)(1-\gamma)}$$

and we can minimize this bound over $0<\gamma<1$ to get

$$\langle f_n\rangle \le \frac{L}{n\varepsilon}\log\frac{n\varepsilon}{L}.$$

Step 3. Together with (9) this implies

$$\frac{a}{Ln^{1-a}} \le \frac{L}{n\varepsilon}\log\frac{n\varepsilon}{L} + \delta.$$

By definition of δ this implies that

$$\mathbb{E}_x a n^a \le \frac{L}{\varepsilon}\log n\varepsilon + \mathbb{E}_x|\delta|$$

where $\mathbb{E}_x|\delta|\to 0$ as $v\to\infty$. For any $a_0>0$ we can write

$$\mathbb{E}_x a \le a_0 + \mathbb{E}_x a I(a\ge a_0) \le a_0 + \frac{1}{n^{a_0}}\mathbb{E}_x a n^a \le a_0 + \frac{1}{n^{a_0}}\Big(\frac{L}{\varepsilon}\log n\varepsilon + \mathbb{E}_x|\delta|\Big).$$

The right hand side goes to zero if we let $v\to\infty, n\to\infty, a_0\to 0$ which means that $\mathbb{E}_x a\to 0$ as $v\to\infty$ and this finishes the proof.

$\square$

Proof of Corollary 1. Since (2) is uniform in ν, by Chebyshev's inequality

$$\mathbb{P}\Big(\int \nu_g^{\otimes 2}\{z^1\cdot z^2 \le -\varepsilon\}dQ(\nu) \ge 4\varepsilon\Big) \le \frac{1}{4}. \tag{12}$$

Next, we will show that for some absolute constant L,

$$\mathbb{P}\Big(\sup_{z\in S}|g(z)| \ge Lv\sqrt{N}\Big) \le \frac{1}{4}. \tag{13}$$

Indeed, conditionally on (x_p) the process $g(z)$ is Gaussian with covariance

$$\mathbb{E}_g g(z^1)g(z^2) = v^2\xi(z^1\cdot z^2) \ \text{ where } \ \xi(s)=\sum_{p\ge 1}2^{-2p}x_p^2 s^p.$$

Since the first two derivatives of $\xi(s)$ are bounded uniformly over $(x_p)\in[0,1]^\infty$,

$$\begin{aligned}\mathbb{E}_g(g(z^1)-g(z^2))^2 &= v^2\big(\xi(z^1\cdot z^1)+\xi(z^2\cdot z^2)-2\xi(z^1\cdot z^2)\big)\\ &\le Lv^2|z^1-z^2|^2 = Lv^2\mathbb{E}(\eta(z^1)-\eta(z^2))^2\end{aligned}$$

where η is the canonical Gaussian process $\eta(z) = \sum g_i z_i$. By a consequence of Slepian's inequality (see Corollary 3.14 in [5])

$$\mathbb{E}_g \sup_{z\in S} g(z) \le Lv\, \mathbb{E} \sup_{z\in S} \eta(z) \le Lv\sqrt{N}$$

and, by symmetry, $\mathbb{E}_g \sup_{z\in S} |g(z)| \le Lv\sqrt{N}$ which implies (13). (12) and (13) imply that the event

$$\Big\{\int \nu_g^{\otimes 2}\{z^1\cdot z^2 \le -\varepsilon\} dQ(\nu) \le 4\varepsilon\Big\} \bigcap \Big\{\sup_{z\in S} |g(z)| \le Lv\sqrt{N}\Big\}$$

is not empty and this finishes the proof.

□

References

[1] Aizenman, M., Sims, R., Starr, S. (2003) An extended variational principle for the SK spin-glass model. *Phys. Rev. B*, **68**, 214403.

[2] Contucci, P., Giardinà, C. (2007) The Ghirlanda-Guerra identities. *J. Stat. Phys.,* **126**, no. 4-5, 917-931. MR2311890

[3] Ghirlanda, S., Guerra, F. (1998) General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity. *J. Phys. A* **31**, no. 46, 9149-9155. MR1662161

[4] Guerra, F. (2003) Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.* **233**, no. 1, 1-12. MR1957729

[5] Ledoux, M., Talagrand, M. (1991) Probability in Banach spaces. Isoperimetry and Processes. Springer-Verlag. MR1102015

[6] Panchenko, D. (2005). Free energy in the generalized Sherrington-Kirkpatrick mean field model. *Rev. Math. Phys.* **17**, no. 7, 793-857. MR2159369

[7] Parisi, G., Talagrand, M. (2004) On the distribution of the overlaps at given disorder. *C. R. Math. Acad. Sci. Paris,* **339**, no. 4, 303-306. MR2092018

[8] Sherrington, D., Kirkpatrick, S. (1972) Solvable model of a spin glass. *Phys. Rev. Lett.* **35**, 1792-1796.

[9] Talagrand, M. (2003) Spin Glasses: a Challenge for Mathematicians. Springer-Verlag. MR1993891

[10] Talagrand, M. (2003) On Guerra's broken replica-symmetry bound. *C. R. Math. Acad. Sci. Paris* **337**, no. 7, 477-480. MR2023757

[11] Talagrand, M. (2006) Parisi formula. *Ann. of Math. (2)* **163**, no. 1, 221-263. MR2195134

[12] Talagrand, M. (2007) Large deviations, Guerra's and A.S.S. Schemes, and the Parisi hypothesis. *Lecture Notes in Mathematics*, Vol. 1900, Eds: E. Bolthausen, A. Bovier.

A A gaussian concentration inequality.

For completeness, we give a proof of the Gaussian concentration inequality which can be found, for example, in [6].

Lemma 3. *Let ν be a finite measure and $g(z)$ be a Gaussian process on $\mathbb{R}^N$ such that $\mathbb{E}g(z)^2 \le a$ for z in the support of measure ν. If*

$$X = \log \int \exp g(z) d\nu(z)$$

then $\mathbb{E}(X - \mathbb{E}X)^2 \le 8a$.

Proof. Let g^1 and g^2 be two independent copies of g. For $t \in [0,1]$ and $j = 1,2$ we define

$$g_t^j(z) = \sqrt{t} g^j(z) + \sqrt{1-t} g(z) \quad \text{and} \quad g_t(z^1, z^2) = g_t^1(z^1) + g_t^2(z^2).$$

Let

$$F_j = \log \int \exp g_t^j(z) d\nu(z) \tag{14}$$

and let $F = F_1 + F_2$. For $s \ge 0$, let $\varphi(t) = \mathbb{E}\exp s(F_2 - F_1)$. We can write

$$\varphi'(t) = s\mathbb{E}\exp s(F_2 - F_1) \sum_{j\le 2} (-1)^{j+1} \exp(-F_j) \int \frac{\partial g_t^j(z)}{\partial t} \exp g_t^j(z) d\nu(z).$$

If $\zeta(z^1, z^2) := \mathbb{E}g(z^1)g(z^2)$ then

$$2\mathbb{E}\frac{\partial g_t^j(z^1)}{\partial t} g_t^j(z^2) = \zeta(z^1, z^2) - \zeta(z^1, z^2) = 0$$

and

$$2\mathbb{E}\frac{\partial g_t^1(z^1)}{\partial t} g_t^2(z^2) = -\zeta(z^1, z^2).$$

Therefore, Gaussian integration by parts gives

$$\varphi'(t) = -s^2 \mathbb{E}\exp s(F_1 - F_2) \exp(-F) \int \zeta(z^1, z^2) \exp g_t(z^1, z^2) d\nu(z^1) d\nu(z^2)$$

and since by assumption $|\zeta(z^1, z^2)| \le a$ on the support of measure $\nu^{\otimes 2}$ we get

$$\varphi'(t) \le as^2 \mathbb{E}\exp s(F_1 - F_2) \exp(-F) \int \exp g_t(z^1, z^2) d\nu(z^1) d\nu(z^2) = as^2 \varphi(t).$$

By construction $\varphi(0) = 1$ and the above inequality implies that $\varphi(1) \le \exp as^2$. On the other hand, by construction, $\varphi(1) = \mathbb{E}\exp s(X - X')$, where X' is an independent copy of X. Thus,

$$\mathbb{E}\exp s(X - X') \le \exp as^2.$$

By Jensen's inequality, $\mathbb{E}\exp s(X - \mathbb{E}X) \le \exp as^2$ and by Markov's inequality

$$\mathbb{P}\big(X - \mathbb{E}X \ge t\big) \le \inf_{s>0} \exp(as^2 - st) = \exp\Big(-\frac{t^2}{4a}\Big).$$

Obviously, a similar inequality can be written for $\mathbb{E}X - X$ and, therefore,

$$\mathbb{P}\big(|X - \mathbb{E}X| \ge t\big) \le 2\exp\Big(-\frac{t^2}{4a}\Big).$$

The result follows.

$\square$