MATH 585 – TOPICS IN MATHEMATICAL PHYSICS – FALL 2006
MATHEMATICS OF MEAN FIELD SPIN GLASSES AND THE REPLICA METHOD
LECTURE 10: A CONCENTRATION-OF-MEASURE RESULT

S. STARR
MATHEMATICS DEPARTMENT, UNIVERSITY OF ROCHESTER

CONTENTS

Let us declare a function $f(x)$ to be of "slow-growth" compared to Gaussian decay if $\lim_{\|x\|\to\infty} f(x)e^{-\|x\|^2/2\sigma^2} = 0$ *for every* $\sigma < \infty$*.*

1. ONE OF TALAGRAND'S CONCENTRATION-OF-MEASURE RESULTS

Talagrand is a major contributor (or perhaps *the* major contributor) to the field of "concentration-of-measure inequalities". The strength of these inequalities is that they often bound fluctuations of a random vector in a sufficiently strong way to allow nontrivial results, even when the dimension of the random vector approaches ∞. A central work in the field is his IHES review article [3]. But in this lecture we will present his Theorem 2.2.4 of [4]. This is a newer version of an older concentration of measure results, which Talagrand actually credits to Pisier [2]. But Talagrand has given a very simple new proof using quadratic interpolation, and some extra very clever ideas, whereas, Pisier's proof used linear interpolation.

Recall that a function $\psi : \mathbb{R}^n \to \mathbb{R}$ is called globally Lipschitz, with constant L, if it is true that

$$\frac{|\psi(x) - \psi(y)|}{\|x - y\|} \leq L ,$$

for all $x \neq y$. We will say that ψ is Lipschitz with the constant L, when we mean it is "globally Lipschitz". When we make this statement, we do not necessarily mean that L is the smallest possible constant such that ψ is Lipschitz with that constant.

Let us mention a technical fact. If ψ is Lipschitz with constant L, then there is an approximating family of smooth functions, ψ_ϵ, for $\epsilon > 0$, which are also Lipschitz with constant L, and such that $\psi_\epsilon \to \psi$, uniformly. Moreover, the family can be chosen such that for each $k \in \mathbb{Z}_{>0}$, there are constants $a_k(\epsilon), b_k(\epsilon) < \infty$ such that

$$\|\nabla^k \psi(x)\| \leq a_k(\epsilon)\|x\| + b_k(\epsilon) ,$$

Date: November 26, 2006.

where $\nabla^k \psi$ is the tensor of k-th order partial derivatives, and $\|\cdot\|$ is the ℓ^2-norm of the components. Since ψ_ϵ is Lipschitz, one can take $a_1 = 0$ and $b_1 = L$ for all ϵ, because by the very definition of "Lipschitz" it is apparent that $\sup_x \|\nabla \psi_\epsilon(x)\| \leq L$. For higher derivatives there will be ϵ-dependence which may diverge as $\epsilon \to 0$. Of course, ψ itself (without any derivatives), satisfies a linear bound like these because it is Lipschitz. (E.g., $|\psi(x)| \leq |\psi(0)| + L\|x\|$.) So does each ψ_ϵ. In particular this means that $e^{\lambda \psi_\epsilon(x)}$ is of slow-growth compared to a Gaussian density, for each $\lambda \in \mathbb{R}$, because it grows only exponentially. The derivatives of $e^{\lambda\psi_\epsilon}$ are also of slow-growth for similar reasons.

Theorem 1.1 (Concentration-of-measure for Lipschitz functions of Gaussian r.v.'s) *Suppose $\psi : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz with constant L. Also suppose that $\mathsf{X} \in \mathbb{R}^n$ is a $N(0, I_n)$ random vector (where I_n is the identity matrix on $\mathbb{R}^n$). Then for each $t > 0$,*

$$\mathbb{P}\big\{\big|\psi(\mathsf{X}) - \mathbb{E}[\psi(\mathsf{X})]\big| \geq t\big\} \leq 2e^{-t^2/4L^2}. \tag{1.1}$$

We will break Talagrand's proof into two pieces. The first piece is a lemma.

Lemma 1.2 *Suppose $\psi : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz with constant L. Suppose that X and Y are independent $N(0, I_n)$ random vectors. Then for any $\lambda \in \mathbb{R}$,*

$$\mathbb{E}\left[e^{\lambda[\psi(\mathsf{X})-\psi(\mathsf{Y})]}\right] \leq e^{\lambda^2 L^2}. \tag{1.2}$$

Theorem 1.1 follows easily from this lemma.

Proof. By Jensen's inequality, and (1.2),

$$\mathbb{E}\left[e^{\lambda(\psi(\mathsf{X})-\mathbb{E}[\psi(\mathsf{Y})])}\right] \leq \mathbb{E}\left[e^{\lambda[\psi(\mathsf{X})-\psi(\mathsf{Y})]}\right] \leq e^{\lambda^2 L^2}.$$

Note that this holds by convexity of $e^{\lambda x}$ regardless of the sign of λ. By Chebyshev's inequality, for $t \geq 0$,

$$\mathbb{P}\{\psi(\mathsf{X}) - \mathbb{E}[\psi(\mathsf{Y})] \geq t\} \leq e^{\lambda^2 L^2 - \lambda t},$$

for every $\lambda \geq 0$. Optimizing in λ gives

$$\mathbb{P}\{\psi(\mathsf{X}) - \mathbb{E}[\psi(\mathsf{Y})] \geq t\} \leq e^{-t^2/(4L^2)}.$$

Taking $\lambda < 0$ and repeating the Chebyshev argument also leads to

$$\mathbb{P}\{\psi(\mathsf{X}) - \mathbb{E}[\psi(\mathsf{Y})] \leq -t\} \leq e^{-t^2/(4L^2)},$$

for $t > 0$. Putting the two together gives (1.1). □

Now it is time to prove the lemma.

Proof. To begin with, assume that ψ is smooth and all derivatives of ψ satisfy linear bounds, in addition to ψ being Lipschitz for constant L. Given $\lambda \in \mathbb{R}$, let $\Psi_\lambda : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be the function

$$\Psi_\lambda(x, y) = e^{\lambda[\psi(x)-\psi(y)]}.$$

Because of the linear bounds on ψ (which hold without further assumptions just because ψ is Lipschitz) the function Ψ_λ is of slow growth compared to a Gaussian decay. (It grows at most exponentially with exponent equal to λL.) Because ψ is smooth with linear bounds on its derivatives, all derivatives of Ψ_λ are also of slow growth compared to the Gaussian decay. Let us define a vector $u \in \mathbb{R}^{2n}$ by $u = (x, y)$. So $u_i = x_i$ for $i \leq n$, while $u_{n+i} = y_i$.

Let $\mathsf{X}, \mathsf{Y}, \mathsf{Z}$ be independent $N(0, I_n)$ random variables. Define $\mathsf{U} = (\mathsf{X}, \mathsf{Y}) \in \mathbb{R}^{2n}$ which is $N(0, I_{2n})$. Also define $\mathsf{V} = (\mathsf{Z}, \mathsf{Z})$. This of course is not $N(0, I_{2n})$, rather

$$\mathbb{E}[\mathsf{V}_i \mathsf{V}_j] \;=\; \delta_{ij} + \delta_{i,n+j} + \delta_{j,n+1}\,.$$

Therefore,

$$\mathbb{E}[\mathsf{U}_i \mathsf{U}_j] - \mathbb{E}[\mathsf{V}_i \mathsf{V}_j] \;=\; -\delta_{i,n+j} - \delta_{j,n+i}\,.$$

By a calculation from Lecture 8, suitably generalized to functions of slow growth (whose derivatives also have slow growth) we know that

$$\mathbb{E}[\Psi_\lambda(\mathsf{U})] - \mathbb{E}[\Psi_\lambda(\mathsf{V})] \;=\; \frac{1}{2}\sum_{i,j=1}^{2n} \left(\mathbb{E}[\mathsf{U}_i \mathsf{U}_j] - \mathbb{E}[\mathsf{V}_i \mathsf{V}_j]\right) \int_0^1 \mathbb{E}\left[\frac{\partial^2 \Psi_\lambda}{\partial u_i \partial u_j}(\sqrt{t}\,\mathsf{U} + \sqrt{1-t}\,\mathsf{V})\right] dt\,. \tag{1.3}$$

Note: *I accidentally missed the factor $\frac{1}{2}$ in the statement of Lemma 3.1 of Lecture 8. Compare to Lemma 2.1 of that lecture. It did not matter before now, because we only worried about the sign of the right-hand-side. But now it becomes important.*

On the other hand, the sum is restricted to i, j such that $i = n + j$ or $j = n + i$. Note that

$$\frac{\partial^2 \Psi_\lambda}{\partial u_i \partial u_{n+i}}(u) \;=\; \frac{\partial^2 \Psi_\lambda}{\partial x_i \partial y_i}(x, y) \;=\; -\lambda^2 \frac{\partial \psi}{\partial x_i}(x)\, \frac{\partial \psi}{\partial x_i}(y)\, \Psi_\lambda(x, y)\,.$$

So (1.3) becomes

$$\mathbb{E}[\Psi_\lambda(\mathsf{X}, \mathsf{Y})] - \mathbb{E}[\Psi_\lambda(\mathsf{Z}, \mathsf{Z})] \;=\; \lambda^2 \sum_{i=1}^{n} \int_0^1 \mathbb{E}\left[\frac{\partial \psi}{\partial x_i}(x)\, \frac{\partial \psi}{\partial x_i}(y)\, \Psi_\lambda(x, y)\Big|_{\substack{x=\sqrt{t}\mathsf{X}+\sqrt{1-t}\mathsf{Z}\\ y=\sqrt{t}\mathsf{Y}+\sqrt{1-t}\mathsf{Z}}}\right] dt\,.$$

But observe, first, that $\Psi(z, z) = 1$ by definition. Then observe that, bringing the sum inside the integral and expectation,

$$\mathbb{E}[\Psi_\lambda(\mathsf{X}, \mathsf{Y})] - 1 \;=\; \lambda^2 \int_0^1 \mathbb{E}\left[\left(\nabla\psi(x) \cdot \nabla\psi(y)\right)\; \Psi_\lambda(x, y)\Big|_{\substack{x=\sqrt{t}\mathsf{X}+\sqrt{1-t}\mathsf{Z}\\ y=\sqrt{t}\mathsf{Y}+\sqrt{1-t}\mathsf{Z}}}\right] dt\,.$$

Using Cauchy-Schwarz and the fact that $\sup_x \|\nabla\psi(x)\| \leq L$ whenever ψ is Lipschitz with constant L (and using the fact that Ψ_λ is positive) one has

$$\mathbb{E}[\Psi_\lambda(\mathsf{X}, \mathsf{Y})] - 1 \;\leq\; \lambda^2 L^2 \int_0^1 \mathbb{E}\left[\Psi_\lambda(\sqrt{t}\mathsf{X} + \sqrt{1-t}\mathsf{Z}, \sqrt{t}\mathsf{Y} + \sqrt{1-t}\mathsf{Z})\right] dt\,.$$

Defining

$$f(t) \;=\; \mathbb{E}\left[\Psi_\lambda(\sqrt{t}\mathsf{X} + \sqrt{1-t}\mathsf{Z}, \sqrt{t}\mathsf{Y} + \sqrt{1-t}\mathsf{Z})\right]\,,$$

for $0 \leq t \leq 1$, one can iterate this inequality, using the fact that for independent $N(0, I_n)$ random vectors Z and Z', one has

$$\sqrt{s(1-t)}\mathsf{Z} + \sqrt{1-s}\mathsf{Z}' \;\stackrel{\mathscr{D}}{=}\; \sqrt{1-st}\mathsf{Z}\,,$$

to determine that

$$f(t) \;\leq\; 1 + \lambda^2 L^2 \int_0^t f(s)\, ds\,.$$

By a version of Gronwall's inequality, this implies that

$$f(t) \le e^{\lambda^2 L^2 t}.$$

Setting $t = 1$ gives the desired result.

When ψ is not smooth, one can approximate ψ by smooth functions ψ_ϵ, which are also all Lipschitz with constant L, and such that the convergence is uniform, as mentioned before. The inequality (1.1) then applies to each ψ_ϵ. Uniform convergence is enough to guarantee convergence of the right hand side to the desired limit, also using the fact that $e^{\psi_\epsilon(x)}$ is of slow-growth compared to a Gaussian. Therefore, one obtains the desired inequality for ψ in the limit. □

2. Application to the pressure of the Sherrington-Kirkpatrick model

Recall that the random pressure for the SK model is

$$\mathsf{p}_N(\beta,x) = \frac{1}{N}\log\left[\sum_{\sigma\in\Omega^N} w_N(\sigma;x)\,\exp\left(-\frac{\beta}{\sqrt{2N}}\sum_{i,j=1}^N \mathsf{J}_{i,j}\sigma_i\sigma_j\right)\right].$$

Think of this as a function $\psi((\mathsf{J}_{i,j} : 1 \le i,j \le N))$. Recall that in the context of the Sherrington-Kirkpatrick model, the letter x is supposed to stand for βh. We will define the function ψ to be $\psi(z)$ in order to use the letter x for more than one thing, at one time. So $z \in \mathbb{R}^{N^2}$ and we like to label the components of z by two indices

$$z = (z_{ij} \in \mathbb{R} : 1 \le i,j \le N).$$

Note that $\psi(z)$ is differentiable, and we have

$$\frac{\partial}{\partial z_{i,j}}\psi(z) = \frac{1}{N}\cdot\frac{\sum_{\sigma\in\Omega^N} w_n(\sigma;x)\,\exp\left(-\frac{\beta}{\sqrt{2N}}\sum_{i,j=1}^N z_{i,j}\sigma_i\sigma_j\right)\left(-\frac{\beta}{\sqrt{2N}}\sigma_i\sigma_j\right)}{\sum_{\sigma\in\Omega^N} w_N(\sigma;x)\,\exp\left(-\frac{\beta}{\sqrt{2N}}\sum_{i,j=1}^N z_{i,j}\sigma_i\sigma_j\right)},$$

which we obtain by replacing each $\mathsf{J}_{i,j}$ by $z_{i,j}$. Since $w_N(\sigma,x)$ is positive, and since $|\sigma_i\sigma_j| = 1$, we can bound

$$\left|\frac{\partial\psi}{\partial z_{i,j}}(z)\right| \le \frac{1}{N}\cdot\frac{\sum_{\sigma\in\Omega^N} w_n(\sigma;x)\,\exp\left(-\frac{\beta}{\sqrt{2N}}\sum_{i,j=1}^N z_{i,j}\sigma_i\sigma_j\right)\left|-\frac{\beta}{\sqrt{2N}}\sigma_i\sigma_j\right|}{\sum_{\sigma\in\Omega^N} w_N(\sigma;x)\,\exp\left(-\frac{\beta}{\sqrt{2N}}\sum_{i,j=1}^N z_{i,j}\sigma_i\sigma_j\right)} = \frac{\beta}{2^{1/2}N^{3/2}}.$$

Therefore,

$$\left(\frac{\partial\psi}{\partial z_{i,j}}(z)\right)^2 \le \frac{\beta^2}{2N^3}.$$

Summing over all N^2 components $z_{i,j}$, we obtain

$$\|\nabla\psi(z)\|^2 = \sum_{i,j=1}^N\left(\frac{\partial\psi}{\partial z_{i,j}}(z)\right)^2 \le \frac{\beta^2}{2N}.$$

But, because ψ is differentiable, we have precisely that ψ is Lipschitz with the smallest Lipschitz constant being

$$\mathrm{Lip}(\psi) := \sup_{x \in \mathbb{R}^n} \|\nabla \psi(x)\| .$$

Indeed, if $x, y \in \mathbb{R}^n$ then

$$\begin{aligned} \|\psi(x) - \psi(y)\| &\leq \int_0^1 |\nabla \psi(tx + [1-t]y) \cdot (x-y)|\, dt \\ &\leq \|x-y\| \int_0^1 \|\nabla \psi(tx + [1-t]y)\|\, dt \\ &\leq \|x-y\|\, \mathrm{Lip}(\psi) , \end{aligned}$$

by Cauchy-Schwarz.

Therefore, in the present case ψ is Lipschitz, with constant $L = \frac{\beta}{\sqrt{2N}}$. Therefore, applying Talagrand's theorem, proves

$$\mathbb{P}\{|\mathsf{p}_N(\beta, x) - p_N(\beta, x)| > t\} \leq 2e^{-2Nt^2/\beta^2},$$

for each $t > 0$. This is a strong form of the "self-averaging" property as $N \to \infty$. In particular, this bound can be integrated to give

$$\mathsf{Var}(\mathsf{p}_N(\beta, x)) \leq \frac{\beta^2}{N} .$$

This recovers a result of Pastur and Shcherbina, but in a different, slightly easier way. (See the appendix of [1].) On the other hand, one should not think of the variance bound as the final result because if you are interested in more than just the second moment, the concentration of measure inequality is strictly stronger.

REFERENCES

[1] L. A. Pastur and M. V. Shcherbina. Absence of self-averaging of the order parameter in the Sherrington-Kirkpatrick model. *J. Statist. Phys.* **62** (1991), 1–19.
[2] G. Pisier. Probabilistic methods in the geometry of Banach spaces. in *Probability and analysis (Varenna, 1985)* Lecture Notes in Math., v. 1206, pp. 167–241, Springer, Berlin (1986).
[3] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.* **no. 81**, 73–205 (1995).
[4] M Talagrand. *Spin Glasses: A Challenge for Mathematicians.* Springer, Berlin (2003).

MATHEMATICS DEPARTMENT, UNIVERSITY OF ROCHESTER, ROCHESTER, NY 14627
E-mail address: sstarr@math.rochester.edu