MATH 585 – TOPICS IN MATHEMATICAL PHYSICS – FALL 2006
MATHEMATICS OF MEAN FIELD SPIN GLASSES AND THE REPLICA METHOD
LECTURE 8: QUADRATIC REPLICA COUPLING I, SLEPIAN'S LEMMA

S. STARR
MATHEMATICS DEPARTMENT, UNIVERSITY OF ROCHESTER

CONTENTS

All Gaussians are always assumed to be centered.

1. QUADRATIC INTERPOLATION

In 2000 to 2001 Guerra and Toninelli invented an important new idea in the context of spin glasses. This was quadratic replica coupling. Prior to their discovery, linear interpolation was a well-known tool. Namely, if one starts with independent Gaussians X and Y, then defining $\mathsf{Z}_t = t\mathsf{X} + (1-t)\mathsf{Y}$, for $0 < t < 1$, one obtains a 1-parameter family of Gaussian random variables, interpolating between X and Y. (We should say that we will never make use of the explicit coupling. I.e., what is important is to have probability distributions interpolating between the distributions of X and Y, not actual interpolating sample paths.) Guerra and Toninelli's observation was that, for spin glasses, many things work better (or actually only work at all) if you take an interpolation instead $\mathsf{W}_t = \sqrt{t}\,\mathsf{X} + \sqrt{1-t}\,\mathsf{Y}$. In particular, this means that the variances are linearly interpolated

$$\mathrm{Var}[\mathsf{W}_t] \;=\; t\mathrm{Var}[\mathsf{X}] + (1-t)\mathrm{Var}[\mathsf{Y}]\,.$$

Since the variance is quadratic in X and Y, this may be called quadratic interpolation.

In this first lecture on the topic of quadratic coupling, we will present a generalization of Slepian's lemma. This lemma is somewhat distinct from spin glasses. Slepian's lemma is supposedly important in the subject of "probability on Banach spaces". (Talagrand is an expert on this topic, for example.) But the generalized Slepian's lemma, that we will present, was derived before Guerra and Toninelli's breakthrough, and it seems that nobody realized that it applies to spin glasses. On the other hand, after Guerra and Toninelli's results, it was quickly rederived, and now it seems that it allows a pedagogical introduction to the technique. Therefore, we present this first, but the reader should in no way let this detract

Date: November 7, 2006.

from the importance, or originality, of Guerra and Toninelli's results which we will discuss in subsequent lectures.

2. A GAUSSIAN DIFFERENTIATION LEMMA

For $\lambda \in \mathbb{R}^n$, the notation $\frac{\partial}{\partial\lambda}$ denotes the divergence and $\frac{\partial^2}{\partial\lambda^2}$ denotes the Hessian. The Laplacian is denoted Δ. We denote $\mathscr{C}_\kappa^k(\mathbb{R}^n)$ to be functions on $\mathbb{R}^n$ which are k-times continuously differentiable with compact support. By $\mathscr{C}_0^k(\mathbb{R}^n)$ we mean the functions which are k-times continuously differentiable, and such that the function and all its derivatives up to kth order vanish at ∞. We denote the inner-product on $\mathbb{R}^n$ by $\langle\cdot,\cdot\rangle$.

Lemma 2.1 (Infinitesimal Generator) *For $0 < t < 1$, let there be Gaussian random vectors $\mathsf{X}_t \in \mathbb{R}^n$, with covariances $C_t \in M_n(\mathbb{R})$, such that $\dot{C}_t = \frac{d}{dt}C_t$ is continuous on $(0,1)$. Suppose $\psi \in \mathscr{C}_0^2(\mathbb{R}^n)$. Then*

$$\frac{d}{dt}\mathbb{E}[\psi(\mathsf{X}_t)] \,=\, \mathbb{E}\left[(\mathfrak{G}_t\psi)(\mathsf{X}_t)\right] \quad \textit{where} \quad \mathfrak{G}_t = \frac{1}{2}\big\langle\nabla, \dot{C}_t\nabla\big\rangle\,.$$

Lemma 2.2 (Wick's rule) *Let $\mathsf{X} \in \mathbb{R}^n$ be a Gaussian random vector with covariance matrix $C \in M_n(\mathbb{R})$. Suppose $N \in \mathbb{Z}_{>0}$, and suppose $a_1,\dots,a_{2N} \in \mathbb{R}^n$. Then*

$$\mathbb{E}\left[\prod_{k=1}^{2N}\langle a_k, \mathsf{X}\rangle\right] \,=\, \sum_{\pi\in\mathfrak{P}_{2N}}\prod_{k=1}^{N}\langle a_{\pi(2k-1)}, C a_{\pi(2k)}\rangle\,,$$

where $\mathfrak{P}_{2N} \subset \mathfrak{S}_{2N}$ is the set of all $2^{-N}\binom{2N}{N}$ "pairing" permutations, π, satisfying

$$\pi(1) < \pi(3) < \pi(5) < \dots < \pi(2N-1) \quad \textit{and}$$
$$\pi(1) < \pi(2)\,, \quad \pi(3) < \pi(4)\,, \quad \dots\,, \quad \pi(2N-1) < \pi(2N)\,.$$

Proof. For $\mu \in \mathscr{M}_1(\mathbb{R}^n)$ and $\psi \in \mathscr{C}_0(\mathbb{R}^n)$, define the Fourier transforms

$$\hat{\mu}(\lambda) \,=\, \int_{\mathbb{R}^n} e^{-2\pi i\langle\lambda,x\rangle}\,\mu(dx) \quad \text{and} \quad \hat{\psi}(\lambda) \,=\, \int_{\mathbb{R}^n} e^{-2\pi i\langle\lambda,x\rangle}\,\psi(x)\,dx\,.$$

Then Plancherel's formula implies that

$$\int_{\mathbb{R}^n}\psi(x)\,\mu(dx) \,=\, \int_{\mathbb{R}^n}\hat{\psi}(\lambda)\,\hat{\mu}(\lambda)\,d\lambda\,.$$

For μ_t the distribution of X_t, we have $\hat{\mu}_t(\lambda) \,=\, e^{-2\pi^2\langle\lambda,C_t\lambda\rangle}$. Therefore,

$$\frac{d}{dt}\int_{\mathbb{R}^n}\hat{\psi}(\lambda)\,\hat{\mu}_t(\lambda)\,d\lambda \,=\, \int_{\mathbb{R}^n}(-2\pi^2\langle\lambda,C_t\lambda\rangle)\,\hat{\psi}(\lambda)\,\hat{\mu}_t(\lambda)\,d\lambda\,.$$

But by I-B-P, when $\psi \in \mathscr{C}_0^2(\mathbb{R}^n)$, $(\frac{\partial^2}{\partial\lambda^2}\psi)\hat{}\,(\lambda) = -4\pi^2(\lambda\otimes\lambda)\hat{\psi}(\lambda)$, where $\lambda\otimes\lambda$ denotes the outer-product matrix $(\lambda\otimes\lambda)_{ij} = \lambda_i\lambda_j$. Therefore, by the Plancherel formula again,

$$\int_{\mathbb{R}^n}(-2\pi^2\langle\lambda,C_t\lambda\rangle)\,\hat{\psi}(\lambda)\,\hat{\mu}_t(\lambda)\,d\lambda \,=\, \frac{1}{2}\int_{\mathbb{R}^n}\big(\big\langle\nabla,\dot{C}_t\nabla\big\rangle\psi\big)(x)\,\mu(dx)\,.$$

That proves Lemma 2.1.

For Lemma 2.2, note that

$$\begin{aligned}\mathbb{E}\left[\prod_{k=1}^{2N}\langle a_k, \mathsf{X}\rangle\, e^{-2\pi i\langle\lambda,\mathsf{X}\rangle}\right] &= (-4\pi^2)^{-N}\left[\prod_{k=1}^{2N}\left\langle a_k, \frac{\partial}{\partial\lambda}\right\rangle\right]\mathbb{E}\left[e^{-2\pi i\langle\lambda,\mathsf{X}\rangle}\right] \\ &= (-4\pi^2)^{-N}\left[\prod_{k=1}^{2N}\left\langle a_k, \frac{\partial}{\partial\lambda}\right\rangle\right] e^{-2\pi^2\langle\lambda,C\lambda\rangle}\,.\end{aligned}$$

By induction, one can prove that

$$\left[\prod_{k=1}^{2N}\left\langle a_k, \frac{\partial}{\partial\lambda}\right\rangle\right] e^{-2\pi^2\langle\lambda,C\lambda\rangle}\Bigg|_{\lambda=0} = (-4\pi^2)^{N}\sum_{\pi\in\mathfrak{P}_{2N}}\prod_{k=1}^{N}\langle a_{\pi(2k-1)}, C a_{\pi(2k)}\rangle\,.$$

□

3. A generalization of Slepian's lemma

Lemma 3.1 *Suppose that* $\mathsf{X}\in\mathbb{R}^n$ *and* $\mathsf{Y}\in\mathbb{R}^n$ *are independent Gaussian vectors. Suppose that* $\psi\in\mathscr{C}_0^2(\mathbb{R}^n)$*. Then*

$$\mathbb{E}[\psi(\mathsf{X})]-\mathbb{E}[\psi(\mathsf{Y})] = \sum_{j,k=1}^{n}\left(\mathbb{E}[\mathsf{X}_j\mathsf{X}_k]-\mathbb{E}[\mathsf{Y}_j\mathsf{Y}_k]\right)\int_0^1\mathbb{E}\left[\frac{\partial^2\psi}{\partial x_j\partial x_k}(\sqrt{t}\,\mathsf{X}+\sqrt{1-t}\,\mathsf{Y})\right]dt\,.$$

Proof. For each $t\in[0,1]$, let $\mathsf{Z}_t=\sqrt{t}\,\mathsf{X}+\sqrt{1-t}\,\mathsf{Y}$. Then Z_t has covariance $C_t=tC_{\mathsf{X}}+(1-t)C_{\mathsf{Y}}$, where C_{X} and C_{Y} are the covariance matrices for X and Y, respectively. So $\dot{C}_t=C_{\mathsf{X}}-C_{\mathsf{Y}}$. Therefore, by Lemma 2.1, we know that for $0<t_1<t_2<1$,

$$\mathbb{E}[\psi(\mathsf{Z}_{t_2})]-\mathbb{E}[\psi(\mathsf{Z}_{t_1})] = \int_{t_1}^{t_2}\sum_{j,k=1}^{n}\left(\mathbb{E}[\mathsf{X}_j\mathsf{X}_k]-\mathbb{E}[\mathsf{Y}_j\mathsf{Y}_k]\right)\mathbb{E}\left[\frac{\partial^2\psi}{\partial x_j\partial x_k}(\mathsf{Z}_t)\right]dt\,.$$

But also, it is clear that there are distributional limits

$$\mathscr{D}-\lim_{t\to1}\mathsf{Z}_t = \mathsf{X}\quad\text{and}\quad\mathscr{D}-\lim_{t\to0}\mathsf{Z}_t = \mathsf{Y}\,.$$

□

Next we will state a result of Joag-dev, Perlman and Pitt [1].

Lemma 3.2 (A generalized Slepian's lemma) *Suppose that* $\mathsf{X}\in\mathbb{R}^n$ *and* $\mathsf{Y}\in\mathbb{R}^n$ *are Gaussian vectors. Suppose that* $\psi\in\mathscr{C}^2(\mathbb{R}^n)$*, and that for some* $N\in\mathbb{Z}_+$ *and* $c<\infty$*,*

$$\max\left\{|\psi(x)|,\ \left\|\frac{\partial\psi}{\partial x}(x)\right\|,\ \left\|\frac{\partial^2\psi}{\partial x^2}(x)\right\|\right\}\le c(1+\|x\|^{2N})\,.$$

Furthermore, suppose that there is a subset $A\subset\{1,\dots,n\}\times\{1,\dots,n\}$ *such that: for* $(j,k)\in A$*,*

$$\mathbb{E}[\mathsf{X}_j\mathsf{X}_k] = \mathbb{E}[\mathsf{Y}_j\mathsf{Y}_k]\,;$$

while for $(j,k) \notin A$,

$$\mathbb{E}[\mathsf{X}_j\mathsf{X}_k] \,\le\, \mathbb{E}[\mathsf{Y}_j\mathsf{Y}_k] \quad \textit{and} \quad \frac{\partial^2\psi}{\partial x_j \partial x_k}(x) \,\ge\, 0\,, \textit{ for all } x \in \mathbb{R}^n.$$

Then

$$\mathbb{E}[\psi(\mathsf{X})] \,\le\, \mathbb{E}[\psi(\mathsf{Y})]\,.$$

Proof. If ψ is actually in $\mathscr{C}_0^2(\mathbb{R}^n)$, then this is a straightforward consequence of Lemma 3.1. Therefore, all that is required is to see how to approximate ψ when it is not in $\mathscr{C}_0^2(\mathbb{R}^n)$. Let $\eta \in \mathbb{C}_\kappa^2(\mathbb{R}^n)$ be any function with $\eta(0) = 1$. Let

$$\psi_\epsilon(x) \,=\, \eta(\epsilon x)\psi(x)\,.$$

Then $\psi_\epsilon \in \mathbb{C}_\kappa^2(\mathbb{R}^n)$ for each $\epsilon > 0$, and

$$\left|\frac{\partial^2\psi_\epsilon}{\partial x_j \partial x_k}(x) - \eta(\epsilon x)\,\frac{\partial^2\psi}{\partial x_j \partial x_k}(x)\right| \,\le\, \epsilon K \max\left\{\left\|\frac{\partial\psi}{\partial x}(x)\right\|, \left\|\frac{\partial^2\psi}{\partial x^2}(x)\right\|\right\} \,\le\, \epsilon c K(1+\|x\|^{2N})\,.$$

By Lemma 2.2, the right-hand-side is ϵ times an integrable function, relative to the distribution of $\mathsf{Z}_t \,=\, \sqrt{t}\,\mathsf{X} + \sqrt{1-t}\,\mathsf{Y}$, for all t. This converges pointwise to 0, which means that by DCT the integral converges to 0, as $\epsilon \to 0^+$. On the other hand, the left hand side is also dominated by a constant times an integrable function, and it converges pointwise to something positive. Therefore, by the Dominated Convergence Theorem, we still have

$$\mathbb{E}[\psi(\mathsf{X})] - \mathbb{E}[\psi(\mathsf{Y})] \,=\, \lim_{\epsilon\to 0^+}\left(\mathbb{E}[\psi_\epsilon(\mathsf{X})] - \mathbb{E}[\psi_\epsilon(\mathsf{Y})]\right) \,\le\, 0\,.$$

□

Corollary 3.3 (Slepian's lemma) *Suppose that* $\mathsf{X} \in \mathbb{R}^n$ *and* $\mathsf{Y} \in \mathbb{R}^n$ *are Gaussian vectors. Suppose that*

$$\textit{Var}(\mathsf{X}_k) \,=\, \textit{Var}(\mathsf{Y}_k)$$

for $k = 1, \dots, n$*, while*

$$\mathbb{E}[\mathsf{X}_j\mathsf{X}_k] \,\le\, \mathbb{E}[\mathsf{Y}_j\mathsf{Y}_k]\,,$$

for $1 \le j < k \le n$*. Then*

$$\mathbb{E}\left[\max\{\mathsf{X}_1,\dots,\mathsf{X}_n\}\right] \,\ge\, \mathbb{E}\left[\max\{\mathsf{Y}_1,\dots,\mathsf{Y}_n\}\right]\,.$$

Proof. For each $\beta < \infty$, the function

$$\psi_\beta(x) \,=\, \frac{1}{\beta}\log\left(\sum_{k=1}^n e^{\beta x_k}\right)\,,$$

satisfies

$$\frac{\partial^2\psi_\beta}{\partial x_j \partial x_k}(x) \,=\, -\beta\frac{e^{\beta x_j}e^{\beta x_k}}{\left(\sum_{i=1}^n e^{\beta x_i}\right)^2} \,\le\, 0\,,$$

for $j \neq k$. Therefore, by Lemma 3.2,

$$\mathbb{E}[\psi_\beta(\mathsf{X})] \,\ge\, \mathbb{E}[\psi_\beta(\mathsf{Y})]\,,$$

for each β. But, by an argument similar to that used in the LDP lectures, we know that

$$\lim_{\beta\to\infty}\psi_\beta(x) \,=\, \max_{1\le k\le n} x_k\,,$$

and the convergence is uniform. Therefore, by DCT, we obtain the result. □

REFERENCES

[1] K. Joag-dev, M. D. Perlman and L. D. Pitt. Association of normal random variables and Slepian's inequality. *Ann Probab.* **11** 451–455, 1983.

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