Lecture 32

Lecture date: Nov 9, 2007 *Scribe: Guy Bresler*

1 Matrix Norms

In this lecture we prove central limit theorems for functions of a random matrix with Gaussian entries. We begin by reviewing two matrix norms, and some basic properties and inequalities.

1. Suppose A is a $n \times n$ real matrix. The *operator norm* of A is defined as

$$\|A\| = \sup_{|x|=1} \|Ax\|, \quad x \in \mathbb{R}^n.$$

Alternatively,

$$\|A\| = \sqrt{\lambda_{\max}(A^T A)},$$

where $\lambda_{\max}(M)$ is the maximum eigenvalue of the matrix M.

Basic properties include:

$$\|A + B\| \le \|A\| + \|B\|$$
$$\|\alpha A\| = |\alpha| \|A\|$$
$$\|AB\| \le \|A\| \|B\|.$$

2. The *Hilbert Schmidt* (alternatively called the Schur, Euclidean, Frobenius) *norm* is defined as

$$\|A\|_{\mathrm{HS}} = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\mathrm{Tr}(A^T A)}.$$

Clearly,

$$\|A\|_{\mathrm{HS}} = \sqrt{\text{sum of eigenvalues of } A^T A},$$

which implies that

$$\|A\| \le \|A\|_{\mathrm{HS}} \le \sqrt{n}\|A\|.$$

Of course, $\|A\|_{\mathrm{HS}}$ also satisfies the usual properties of a norm.

Proposition 1 *The following inequality holds:*

$$\|AB\|_{HS} \le \|A\| \|B\|_{HS}.$$

Proof: Let $b_1, \ldots, b_n$ denote the columns of B. Then

$$\|AB\|_{\mathrm{HS}}^2 = \sum_{i=1}^{n} \|Ab_i\|^2 \leq \sum_{i=1}^{n} \|A\|^2 \|b_i\|^2 = \|A\|^2 \|B\|_{\mathrm{HS}}^2.$$

□

3. A simple matrix inequality follows from the Cauchy-Schwarz inequality:

$$|\operatorname{Tr}(AB)| = \sum_{i,j} a_{ij} b_{ji} \leq \|A\|_{\mathrm{HS}} \|B\|_{\mathrm{HS}}.$$

4. Combining the proposition above with observation 3 gives the inequality

$$|\operatorname{Tr}(ACBD)| \leq \|AC\|_{\mathrm{HS}} \|BD\|_{\mathrm{HS}} \leq \|A\| \|B\| \|C\|_{\mathrm{HS}} \|D\|_{\mathrm{HS}}.$$

More generally, it holds that

$$|\operatorname{Tr}(A_1 A_2 \ldots, A_k)| \leq \|A_i\|_{\mathrm{HS}} \|A_j\|_{\mathrm{HS}} \prod_{l \neq i,j} \|A_l\|.$$

Next, recall the theorem from last lecture:

Theorem 2 *Let $X_1, \ldots, X_k$ be i.i.d. $\mathcal{N}(0,1)$ random variables. Let $f \in C^2(\mathbb{R}^n)$ and $W = f(X)$ with $\mathbf{E}\, W = 0$. Then*

$$d_{TV}(\mathcal{L}(W), \mathcal{N}(0, \sigma^2)) \leq \frac{\sqrt{10}}{\sigma^2} (\mathbf{E} \| \operatorname{Hess} f(X)\|^4 \mathbf{E} \|\nabla f(X)\|^4)^{\frac{1}{4}}.$$

We will use this theorem to study the Gaussian random matrix.

2 CLT for $\operatorname{Tr}(A^k)$

Suppose

$$A = \frac{1}{\sqrt{N}} (X_{ij})_{1 \leq i,j \leq N},$$

where $X_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Fix a positive integer k. We would like a CLT for $\operatorname{Tr}(A^k)$.

To begin, note that

$$\operatorname{Tr}(A^k) = \frac{1}{N^{k/2}} \sum_{1 \leq i_1, i_2, \ldots, i_k \leq N} X_{i_1,i_2} X_{i_2,i_3} \ldots X_{i_{k-1},i_k} X_{i_k,i_1}\,. \tag{1}$$

It turns out that the usual dependency graph theorem fails for $k \geq 3$, so a more powerful method must be used.

Exercise 3 *Find a dependency graph theorem that works for all k.*

In order to apply Theorem 2, we identify

$$X = (X_{11}, X_{12}, \ldots, X_{1k}, X_{21}, X_{22}, \ldots, X_{NN})\,,$$

and $f(X) = \mathrm{Tr}(A^k)$. Now

$$\frac{\partial f}{\partial x_{ij}} = \mathrm{Tr}(\frac{\partial}{\partial x_{ij}} A^k) \stackrel{(a)}{=} Tr\left(\sum_{r=0}^{k-1} A^r \frac{\partial A}{\partial x_{ij}} A^{k-1-r}\right) \stackrel{(b)}{=} kTr\left(\frac{\partial A}{\partial x_{ij}} A^{k-1}\right)\,, \tag{2}$$

where (a) follows from the fact that for two matrices A and B, $\frac{\partial}{\partial x} AB = \frac{\partial A}{\partial x} B + A\frac{\partial B}{\partial x}$, and (b) from moving the trace inside the sum and using $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$. But

$$\frac{\partial A}{\partial x_{ij}} = \frac{1}{\sqrt{N}} e_i e_j^T\,,$$

where e_i is the ith standard basis vector, i.e. the vector of all zeros with a 1 in the ith position.

Thus

$$\begin{aligned}\frac{\partial f}{\partial x_{ij}} &= \frac{k}{\sqrt{N}}\,\mathrm{Tr}(e_i e_j^T A^{k-1})\\ &= \frac{k}{\sqrt{N}}\,\mathrm{Tr}(e_j^T A^{k-1} e_i)\\ &= \frac{k}{\sqrt{N}} (A^{k-1})_{ji}\end{aligned}$$

This allows us to calculate

$$\begin{aligned}\|\nabla f(X)\|^2 = \sum \left(\frac{\partial f}{\partial x_{ij}}\right)^2 &= \frac{k^2}{N} \sum_{i,j} (A^{k-1})_{ji}^2\\ &= \frac{k^2}{N} \|A^{k-1}\|_{\mathrm{HS}}^2\\ &\le \frac{k^2}{N} N \|A^{k-1}\|^2\\ &\le k^2 \|A\|^{2(k-1)}\,.\end{aligned} \tag{3}$$

Lemma 4

$$\mathbf{E}\|A\|^p \le C(p) \quad \forall p \in \mathbb{Z}_+\,,$$

where $C(p)$ is a constant independent of N.

Proof: The proof is essentially as follows. For a positive definite random matrix B, $\|B\| = \lambda_{\max}(B)$. Thus

$$\begin{aligned}\mathbf{E}\|B\|^p = \mathbf{E}\lambda_{\max}^p &\le (\mathbf{E}\lambda_{\max}^{pm})^{1/m} \quad \text{for any } m \\ &\le (\mathbf{E}\,\mathrm{Tr}(B^{pm}))^{1/m}\,.\end{aligned}$$

Now let $m \to \infty$ suitably with N.$\square$

This shows that $\|\nabla f(X)\|^2 = O(1)$, and hence the Poincaré inequality implies that $\mathrm{Var}(f(X)) = O(1)$.

Exercise 5 *Show that any two terms in the sum of equation* (1) *have non-negative covariance.*

The exercise implies that

$$\mathrm{Var}(f(X)) \ge \frac{1}{N^k}\sum \mathrm{Var}(X_{i_1,i_2}\ldots X_{i_k,i_1}) \ge C(k) > 0\,.$$

Recalling the result of Theorem 2,

$$d_{TV}(\mathcal{L}(W), \mathcal{N}(0,\sigma^2)) \le \frac{\sqrt{10}}{\sigma^2}(\mathbf{E}\|\,\mathrm{Hess}\, f(X)\|^4 \mathbf{E}\|\nabla f(X)\|^4)^{\frac{1}{4}}\,,$$

we see that $\sigma^2 = \mathrm{Var}(f(X)) \ge C(k)$ and from equation (3) and the fact noted above, $\mathbf{E}\|\nabla f(X)\|^2 \le C(k)$. Therefore it remains only to show that $\mathbf{E}\|\,\mathrm{Hess}\, f(X)\|^4 \to 0$ in order to prove the desired central limit theorem.

We have

$$\frac{\partial A}{\partial x_{ij}} = kTr\left(\frac{\partial A}{\partial x_{ij}} A^{k-1}\right),$$

and

$$\frac{\partial^2 A}{\partial x_{pq} x_{ij}} = kTr\left(\sum_{r=0}^{k-2} \frac{\partial A}{\partial x_{ij}} A^r \frac{\partial A}{\partial x_{pq}} A^{k-r-2}\right)\,.$$

Fact about matrix norms: If A is a symmetric, real matrix then

$$\|A\| = \sup_{\|x\|=\|y\|=1} |x^T A y|\,.$$

Now, $\mathrm{Hess}\, f(X)$ is an $N^2 \times N^2$ symmetric, real matrix:

$$\|\,\mathrm{Hess}\, f(X)\| = \sup\left\{\sum_{ijpq} c_{ij} d_{pq} \frac{\partial^2 f}{\partial x_{ij}\partial x_{pq}} :\ \sum c_{ij}^2 = 1, \sum d_{pq} = 1\right\}\,.$$

Let $C = (c_{ij})$ and $D = (d_{pq})$ be two matrices with $\|C\|_{\mathrm{HS}} = \|D\|_{\mathrm{HS}} = 1$. Fix $0 \le r \le k-2$. Then

$$\sum_{ijpq} c_{ij} d_{pq} Tr\left(\frac{\partial A}{\partial x_{ij}} A^r \frac{\partial A}{\partial x_{pq}} A^{k-2-r}\right) = \frac{1}{N} \sum c_{ij} d_{pq} \operatorname{Tr}(e_i e_j^T A^r e_p e_q^T A^{k-2-r}) = \frac{1}{N} \operatorname{Tr}(C A^r D A^{k-2-r}),$$

where we used the fact that $\sum c_{ij} e_i e_j^T = C$ and similarly for D.

Now

$$|\operatorname{Tr}(C A^r D A^{k-2-r})| \le \|A\|^{k-2} \|C\|_{\mathrm{HS}} \|D\|_{\mathrm{HS}} = \|A\|^{k-2}.$$

Thus

$$\|\operatorname{Hess} f(X)\| \le \frac{k(k-1)\|A\|^{k-2}}{N}.$$

Combining, we get the desired result:

$$d_{TV}(\operatorname{Tr}(A^k), \mathcal{N}(0, \sigma^2)) \le \frac{C(k)}{N}.$$