Non-Asymptotic Theory of Random Matrices
Lecture 11: Gordon's Inequality

Lecturer: Roman Vershynin Scribe: Yuting Yang

Thursday, Feburuary 08, 2007

1 Review

Slepian's Inequality: Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two Gaussian processes. If

$$\|X_t - X_s\|_2 \leq \|Y_t - Y_s\|_2 \;\forall\, t, s \in T,$$

then

$$\mathbb{E} \sup_{t\in T} X_t \leq \mathbb{E} \sup_{t\in T} Y_t.$$

We used Slepian's inequality to estimate the largest singular value of an $m \times n$ Gaussian random matrix A (i.e. the operator norm of A, $\|A\|$):

$$s_1(A) = \sup_{u\in S^{n-1}} \|Au\|_2 = \sup_{\substack{u\in S^{n-1}\\ v\in S^{m-1}}} \langle\, Au, v \,\rangle = \sup_{\substack{u\in S^{n-1}\\ v\in S^{m-1}}} X_{u,v},$$

where $X_{u,v} = \langle\, Au, v \,\rangle$ is a Gausian random variable so that the supremum can be considered as a Gaussian process. Then we compared it to a simpler process $Y_{u,v}$, applied the Slepian's inequality and concluded that $s_1(A) \leq \sqrt{m} + \sqrt{n}$.

Now, we want to estimate the smallest singular value of A (i.e. $1/\|\, A^{-1} \,\|$):

$$s_n(A) = \inf_{u\in S^{n-1}} \|\, Au \,\|_2 = \inf_{u\in S^{n-1}} \sup_{v\in S^{m-1}} \langle\, Au, v \,\rangle \;\geq ?$$

Note that this is a minimization problem, so Slepian's inequality does not apply.

2 Gordon's Inequality; Estimate for Smallest Singular Value

Gordon's Inequality: Let $(X_{u,v})_{\substack{u\in U\\ v\in V}}$ and $(Y_{u,v})_{\substack{u\in U\\ v\in V}}$ be centered Gaussian processes (Here, "centered" means that the expectations of all the random variables are zero).
Assume that:

(1) $\| X_{u,v} - X_{u',v'} \|_2 \leq \| Y_{u,v} - Y_{u',v'} \|_2$ if $u \neq u'$;

(2) $\| X_{u,v} - X_{u,v'} \|_2 = \| Y_{u,v} - Y_{u,v'} \|_2$.

Then, $\mathbb{E} \sup_{u \in U} \inf_{v \in V} X_{u,v} \leq \mathbb{E} \sup_{u \in U} \inf_{v \in V} Y_{u,v}$.

Remarks :

1. Gordon's inequality contains Slepian's inequality by taking the index set V to be a singleton set (i.e. $|V| = 1$).
2. If we apply Gordon's inequality for $-X_{u,v}$ and $-Y_{u,v}$, we get $\mathbb{E} \inf_{u \in U} \sup_{v \in V} X_{u,v} \leq \mathbb{E} \inf_{u \in U} \sup_{v \in V} Y_{u,v}$.
3. For the proof of Gordon's inequality, see [7],Chapter 3.
4. Gordon's inequality also holds for V replaced by V_U, that is, the index space V can depend on U.

Now, we will use Gordon's inequality to get an estimate for $s_n(A)$.

Recall that we recognize the inner product $\langle Au, v \rangle$ as the trace inner product of the random vector A and the tensor product $u \otimes v$:

$$\langle Au, v \rangle = \langle A, u \otimes v \rangle_{tr}.$$

We let $X_{u,v} = \langle A, u \otimes v \rangle_{tr}$, and compare the process $(X_{u,v})$ with the process $(Y_{u,v})$, where $Y_{u,v} = \langle (g, h), (u, v) \rangle = \langle g, u \rangle + \langle h, v \rangle$, and (g,h) is considered as a Gaussian vector in $\mathbb{R}^{m+n}$.

To utilize Gordon's inequality, we need first to compare the increments:

$$\| X_{u,v} - X_{u',v'} \|_2 = \| u \otimes v - u' \otimes v' \|_{HS}$$

$$\| Y_{u,v} - Y_{u',v'} \|_2 = \| (u, v) - (u', v') \|_2 = \sqrt{\|u - u'\|_2^2 + \|v - v'\|_2^2}$$

(Recall that for a canonical Gaussian process $X_t = \langle g, t \rangle$, we have that $\|X_t - X_s\|_2 = \|t - s\|_2$)

We proved that $\| u \otimes v - u' \otimes v' \|_{HS}^2 \leq \|u - u'\|_2^2 + \|v - v'\|_2^2$ for all unit vectors u, u', v, v'. This gives condition (1) in Gordon's inequality, where we take $U = S^{n-1}$ and $V = S^{m-1}$.

If $u = u'$, then $\|u \otimes v - u' \otimes v'\|_{HS} = \|v - v'\|_2$ since $\|u\|_2 = 1$ (This is left as an exercise to check).
So condition (2) is also satisfied. Therefore, we may apply Gordon's inequality to get

$$\mathbb{E}\ s_n(A) = \mathbb{E}\ \inf_u \sup_v X_{u,v} \geq \mathbb{E}\ \inf_u \sup_v Y_{u,v} = \mathbb{E}\ \inf_u \sup_v \{\langle g, u\rangle + \langle h, v\rangle\} = \mathbb{E}\ \inf_u \langle g, u\rangle + \mathbb{E}\ \sup_v \langle h, v\rangle = -\mathbb{E}\ \|g\|_2 + \mathbb{E}\ \|h\|_2 \geq \sqrt{m} - \sqrt{n}$$

by explicit computation (the gap between $\mathbb{E}\|g\|_2$ and $\sqrt{n}$ is larger than the gap between $\mathbb{E}\|h\|_2$ and $\sqrt{m}$).

3 Deviation Inequalities: How far are $s_1(A)$ and $s_n(A)$ from their averages?

Concentration of Measure:
Let A be a canonical Gaussian vector in $\mathbb{R}^{mn}$. Then, we can view $s_1(A)$ and $s_n(A)$ as functions of A from $\mathbb{R}^{mn}$ to $\mathbb{R}$.
If $s_1(A)$ and $s_n(A)$ are $1-$*Lipschitz* with respect to A, then by concentration of measure, they will be well concentrated around their means. The following proposition shows that the functions $A \mapsto s_1(A)$ and $A \mapsto s_n(A)$ are actually $1-$ *Lipschitz*.

Proposition 1 (1-Lipschitz Conditions). *$|s_1(A) - s_1(B)| \leq \|A - B\|$ and $|s_n(A) - s_n(B)| \leq \|A - B\|$.*

Proof. $|s_1(A) - s_1(B)| = \Big| \|A\| - \|B\| \Big| \leq \|A - B\|$,

$$s_n(A) = \inf_u \|Au\|_2 \leq \inf_u \{\|Bu\|_2 + \|(A - B)u\|_2\} \leq s_n(B) + \|A - B\|$$

□

Remark :

1. Note that $\|A-B\| \leq \|A-B\|_{HS}$ so that $|s_1(A)-s_1(B)| \leq \|A-B\|_{HS}$ and $|s_n(A) - s_n(B)| \leq \|A - B\|_{HS}$.

2. Perturbation theory of eigenvalues:(general results for our proposition)
   - H.Weyl (1912): $|s_k(A) - s_k(B)| \leq \|A - B\| \ \forall\ k$. See [8].
   - Mirsky (1960): $\sum_k |s_k(A) - s_k(B)|^2 \leq \|A - B\|_{HS}^2$. See [4].

Now, by Theorem 7 (Concentration of Measure in Gauss Space (functional form)) in Lecture 3, we have that

$$\mathbb{P}(|s_k(A) - \mathbb{E}s_k(A)| > t) \leq 2\exp\left(-t^2/2\right) \quad \forall\ k,\ \ \forall\ t > 0$$

Hence,

$$\mathbb{P}(s_1(A) > \sqrt{m} + \sqrt{n} + t) \leq 2\exp\left(-t^2/2\right)$$
$$\mathbb{P}(s_n(A) < \sqrt{m} - \sqrt{n} - t) \leq 2\exp\left(-t^2/2\right).$$

Remark :
The constant 2 can be improved to 1 since we only have one-sided inequalities as our events. That is, we have

$$\mathbb{P}(s_1(A) > \sqrt{m} + \sqrt{n} + t) \leq \exp\left(-t^2/2\right)$$

$$\mathbb{P}(s_n(A) < \sqrt{m} - \sqrt{n} - t) \leq \exp\left(-t^2/2\right).$$

Theorem 2 (Singular Values of Gaussian Matrices)**.** *Let A be an $m \times n$ matrix with i.i.d. standard Gaussian entries. Then,*

$$\sqrt{m} - \sqrt{n} \leq \mathbb{E}s_n(A) \leq \mathbb{E}s_1(A) \leq \sqrt{m} + \sqrt{n}$$

Moreover,

$$\sqrt{m} - \sqrt{n} - t \leq s_n(A) \leq s_1(A) \leq \sqrt{m} + \sqrt{n} + t$$

holds with probability at least $1 - 2\exp\left(-t^2/2\right)$ *for all* $t \geq 0$.

Remark :

1. If we let $A' = \frac{1}{\sqrt{m}}A$, then we have

   $$1 - \sqrt{n/m} \leq \mathbb{E}s_n(A') \leq \mathbb{E}s_1(A') \leq 1 + \sqrt{n/m}.$$

   So if the aspect ratio $n/m < 1 - \epsilon$ (so the matrix is not square), then A is almost a nice isomorphism.

2. Notice that the probability estimate is not sharp. So what are the fluctuations of $s_k(A)$?
   By the concentration inequality $\mathbb{P}(|s_k(A) - \mathbb{E}s_k(A)| > t) \leq 2\exp\left(-t^2/2\right)$, we see that the standard deviation of $s_k(A)$ is of constant order: $\sigma(s_k(A)) = O(1)$.

- Aysmptotic theory implies that for $m = n$ (square matrices), $\sigma(s_1(A)) = O(n^{1/3})$. In particular, it was shown in Johansson [3] and Johnstone [2] that if $n/m =$ constant, then as $n \to \infty$, $\frac{(s_1(A))^2-\mu}{\sigma} \to$ Tracy-Widom Law, where $\mu = (\sqrt{m} + \sqrt{n})^2$, $\sigma = (\sqrt{m} + \sqrt{n})(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}})^{1/3} = O(n^{1/3})$. For results on subgaussian matrices, see [6] and [5].
- Nonasymptotic results: unknown even for Gaussian random matrices. Here is a conjecture that would agree with Johansson-Johnstone:

  $$\mathbb{P}(|s_1(A) - \mathbb{E}s_1(A)| > t) \leq \exp\left(-c(n^{1/6}t)^{3/2}\right).$$

  Aubrun proved this for self-adjoint Gaussian case in 2005. See [1]. For subgaussian matrices, even $\mathbb{P}(|s_1(A) - \mathbb{E}s_1(A)| > t) \leq \exp(-ct^2)$ is unknown.

References

[1] G. Aubrun. A small deviation inequality for random matrices. *Sminaire de probabilits*, 38:1857, 2005.

[2] I.Johnstone. On the distribution of the largest eigenvalue in principal component analysis. *Annals of Statistics*, 29:295, 2001.

[3] K. Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209:437, 2000.

[4] L.Mirsky. Symmetric gauge functions and unitarily invariant norms. *Quart.J.Math.*, 11:50, 1960.

[5] S. Peche. The largest eigenvalue of small rank perturbations of hermitian random matrics. *Preprint : arXiv:math.PR/*, 0411487, 2003.

[6] A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108:1033, 2002.

[7] M. Ledoux; M. Talagrand. *Probability in Banach spaces : isoperimetry and processes.* Springer-Verlag, Berlin ; New York, 1991.

[8] H. Weyl. The laws of asymptotic distribution of the eigenvalues of linear partial differential equation. *Math.Ann.*, 71:441, 1912.