Non-Asymptotic Theory of Random Matrices
Lecture 10: Slepian's Inequality
Sharpness Bounds for Gaussian Matrices

Date: February 6, 2007

Lecturer: Roman Vershynin Scribe: Matthew Herman

1 Comparison Inequalities in Probability Theory

We will estimate the size of a random process $(X_t)_{t\in T}$ by the size of another (simpler) random process $(Y_t)_{t\in T}$. Define a *Gaussian centered process* to be when X_t is a mean-zero random variable. For a random variable X

$$\|X\|_2 = \Big(\mathbb{E}\,|X|^2\Big)^{1/2} = \Big(\int_\Omega |X(w)|^2 d\mathbb{P}\Big)^{1/2}$$

where $(\Omega, \mathbb{P}, \Sigma)$ is the underlying probability space in $L^2(\Omega, \mathbb{P}, \Sigma)$.

Theorem 1 (Slepian's Inequality). *Assume $(X_t)_{t\in T}, (Y_t)_{t\in T}$ are Gaussian centered processes. If for all $s, t \in T$*

$$\|X_t - X_s\|_2 \;\le\; \|Y_t - Y_s\|_2,$$

then

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

Proof. See Ledoux-Talagrand §3.1 [2]. □

Observation 2. *If additionally, the moments $\|X_t\|_2 = \|Y_t\|_2$ for all $t \in T$, then*

$$\mathbb{P}(\sup_{t\in T} X_t > u) \;\le\; \mathbb{P}(\sup_{t\in T} Y_t > u)$$

for all $u \ge 0$.

Definition 3 (Canonical Gaussian Process). *Let $T \subset \mathbb{R}^n$ be a subset. Then for all $t \in T$*

$$X_t = \langle g, t\rangle = \sum_{k=1}^{n} g_k t_k$$

where g is the canonical Gaussian random vector in $\mathbb{R}^n$ (i.e., $g_k \sim N(0, 1)$ i.i.d.).

Example 4. *If $T =$ the canonical vector basis of $\mathbb{R}^n$, then*

$$(X_t)_{t \in T} = \{g_1, g_2, \ldots, g_n\}$$

and the process is trivially equal to n i.i.d. $N(0,1)$ random variables.

Definition 5 ("Mean width" of T)**.**

$$\mathbb{E} \sup_{t \in T} X_t = \ell(T).$$

This gives a sense of the "size" of T. Contrast this with the measure of volume. For instance, consider a set T which has a non-negligible width and an infinitesimal cross section. This might lead one to dismiss T as being "small" since it would have a negligible volume. However, examining $\ell(T)$ would reveal that T has a non-negligible size, at least in some direction. Note that there are other definitions of the mean width as well.

Fact 6. *The "size" of an increment equals the distance between indices. That is,*

$$\|X_t - X_s\|_2 \; = \; \left(\mathbb{E}\,|\langle g, t-s\rangle|^2\right)^{1/2} \; = \; \|t-s\|_2.$$

2 Application for Gaussian Random Processes

Recall that A is an $m \times n$ Gaussian matrix (i.e., $A_{ij} = N(0,1)$ i.i.d. random variables). We want to find an upper bound on the largest singular value of A

$$s_1(A) \; = \; \|A\| \; \leq \; ?.$$

Our goal will be to put this problem in the form of a canonical Gaussian process.

Side Remark

$$\|A\| \; = \sup_{\substack{u \in S^{n-1} \\ v \in S^{m-1}}} \langle Au, v\rangle$$

For this inner product we want the matrix A to play the gole of the Gaussian vector g in Definition 3. First we need the following definitions and lemma.

Definition 7 (Inner Product of Two Matrices)**.** *For two matrices* $A, B \in \mathbb{R}^{m\times n}$*, their inner product is defined as*

$$\langle A, B\rangle_{tr} \;=\; \sum_{i,j} A_{ij}B_{ij} \;=\; tr(AB^T).$$

It follows that this induces the Hilbert-Schmidt norm (also known as the Froebenius norm):

$$\|A\|_{HS} \;=\; \Big(\sum_{i,j} A_{ij}^2\Big)^{1/2} \;=\; tr(AB^T).$$

Definition 8 (Tensor Product)**.** *Let* H, G *be two Hilbert spaces. For any* $u \in H, v \in G$ *we can make the rank-one linear operator* $u \otimes v : H \to G$ *defined as*

$$(u \otimes v)\, x \;=\; \langle u, x\rangle\, v$$

for all $x \in H$*.*

Note: In the case of finite dimensional space with $u \in \mathbb{R}^n, v \in \mathbb{R}^m$ we have that $(u \otimes v)$ is the $m \times n$ matrix $(v_i u_j)_{i=1,j=1}^{m\;\;\;n}$.

Lemma 9 (Canonical Form)**.** $\langle Au, v\rangle = \langle A, u \otimes v\rangle_{tr}$

Proof. Examining the RHS we have from Definitions 7 and 8

$$\begin{aligned} tr(A\,(u \otimes v)^T) &= tr(A\,(v \otimes u)) \\ &= \sum_i \langle A\,(v \otimes u)\, e_i,\, e_i\rangle \\ &= \sum_i \langle v_i A u,\, e_i\rangle \\ &= \sum_i v_i (Au)_i \\ &= \langle v, Au\rangle \end{aligned}$$

which, for real-valued vectors, equals the LHS. □

With Lemma 9 we can now write the norm of matrix A as

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

where $\left(X_{(u,v)}\right)_{(u,v)\in T}$ is a canonical Gaussian process on $T = S^{n-1}\times S^{m-1}$ because A is a canonical Gaussian "vector" as viewed in $\mathbb{R}^{mn}$.

Now we compare $\left(X_{(u,v)}\right)_{(u,v)\in T}$ to $\left(Y_{(u,v)}\right)_{(u,v)\in T}$. Let g and h be canonical Gaussian vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Viewing the ordered pair (g,h) as a concatenation of vectors, we have that $(g,h)\in\mathbb{R}^{m+n}$. Similarly, $(u,v)\in\mathbb{R}^{m+n}$. Then by Definition 3

$$Y_{(u,v)} = \langle (g,h),(u,v)\rangle = \langle g,u\rangle + \langle h,v\rangle.$$

Applying Slepian's Increment Inequality of Theorem 1 and Fact 6 we have that

$$\|X_{(u,v)} - X_{(u',v')}\|_2 \le \|Y_{(u,v)} - Y_{(u',v')}\|_2$$

is equivalent to

$$\left\|u\otimes v - u'\otimes v'\right\|_{HS} \le \|(u,v)-(u',v')\|_2$$

for all $u,u'\in S^{n-1}$ and $v,v'\in S^{m-1}$.

Example 10. *Applying Minkowski's inequality we obtain*

$$\sum_{i,j}(u_jv_i - u'_jv'_i)^2 \le \sum_j (u_j-u'_j)^2 + \sum_i (v_i - v'_i)^2.$$

Before continuing, note that $\mathbb{E}\,\|g\|_2 < (\mathbb{E}\,\|g\|_2^2)^{1/2} = \sum_{i=1}^n \mathbb{E}\,g_i^2 = \sqrt{n}$, and similarly $\mathbb{E}\,\|h\|_2 < \sqrt{m}$. We can claim strict inequality here since the first and second moments of a Gaussian are not equal. By Slepian's Lemma we have

$$
\begin{aligned}
\mathbb{E}\,\|A\| &= \mathbb{E}\sup_{\substack{u\in S^{n-1}\\ v\in S^{m-1}}} X_{(u,v)} \\
&\le \mathbb{E}\sup_{\substack{u\in S^{n-1}\\ v\in S^{m-1}}} Y_{(u,v)} \\
&\le \mathbb{E}\sup_{u\in S^{n-1}} \langle g,u\rangle \;+\; \mathbb{E}\sup_{v\in S^{m-1}} \langle h,v\rangle \\
&\le \mathbb{E}\sup_{u\in S^{n-1}} \langle g,u\rangle \;+\; \mathbb{E}\sup_{v\in S^{m-1}} \langle h,v\rangle \\
&= \mathbb{E}\,\|g\|_2 \;+\; \mathbb{E}\,\|h\|_2 \\
&\le \sqrt{n} \;+\; \sqrt{m}.
\end{aligned}
$$

This yields the following theorem.

Theorem 11. *Let A be a real $m \times n$ random Gaussian matrix. Then $\mathbb{E}\,\|A\| < \sqrt{m} + \sqrt{n}$.*

Asymptotic theory gives us: $\frac{1}{\sqrt{m}}\,\|A\| \to 1+\sqrt{\frac{n}{m}}$ almost surely as $n \to \infty$, where $\frac{n}{m} \to$ constant.

Note that the argument in this lecture is due to Gordon and can be found in [1].

In the next lecture we will see that we can find a lower bound for the expectation of the smallest singular value

$$\mathbb{E}\, s_1(A) > \sqrt{m} - \sqrt{n}.$$

References

[1] K. Davidson and S. Szarek. Local operator theory, random matrices and banach spaces. *Handbook of the Geometry of Banach Spaces*, 1:317, 2001.

[2] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991.