Non-Asymptotic Theory of Random Matrices

Lecture 9: Applications of Dudley's Inequality: Sharper bounds for random matrices

Lecturer: Roman Vershynin Scribe: Brendan Farrell

Thursday, February 1, 2006

Recall Dudley's Inequality from lecture 8: Let (T, d) be a metric space and $(X_t)_{t \in T}$ a random process on T with subgaussian increments, that is

$$\mathbb{P}\left(|X_t - X_s| > u \cdot d(t, s)\right) \leq \exp(-u^2)$$

for all $t, s \in T$ and all $u > 0$. Then

$$\mathbb{E} \sup_{t \in T} X_t \leq C \int_0^\infty \sqrt{\log N(t, \epsilon)} d\epsilon,$$

where $N(t, \epsilon)$ is the ϵ-covering number of T. Brownian motion is an example of random process, where T is time and X_t is a random variable that gives the position of the particle at time t. See [2, 3, 1].

Theorem 1. *Let A be an $n \times m$, $m > n$, random matrix with i.i.d, mean zero, variance 1, subgaussian entries. Then, with probability* $1 - C\exp(-m)$,

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

Equivalently, for every vector $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$,

$$\sqrt{m} - C\sqrt{n} \leq \|Ax\|_2 \leq \sqrt{m} + C\sqrt{n}. \tag{1}$$

This shows that $\frac{1}{\sqrt{m}}A$ is an "almost isometry" if $m >> n$. Our approach to prove this theorem is to establish a bound on

$$X_x := \left| \frac{1}{m}\|Ax\|_2^2 - 1 \right|$$

for all $x \in S^{n-1}$. We work with the random process $(X_x)_{x \in T}$ on $T = (S^{n-1}, \|\cdot\|_2)$, and establish a bound on $\mathbb{E}\sup_{x \in T} X_x$ using Dudley's inequality. First we address whether the increments $|X_x - X_y|$ are subgaussian. We rewrite $\|Ax\|_2^2$ as $\|Ax\|_2^2 = \sum_{k=1}^m \langle \overline{x}_k, x \rangle$, where $\overline{x}_k$ denotes the k^{th} row of A.

$$\begin{aligned} |X_x - X_y| &\leq | \frac{1}{m}\sum_{k=1}^m \langle \overline{x}_k, x\rangle^2 - \langle \overline{x}_k, y\rangle^2 | \\ &= |\frac{1}{m}\sum_{k=1}^m z_k|, \end{aligned}$$

where $z_k = \langle \overline{x}_k, x-y\rangle\langle \overline{x}_k, x+y\rangle$. The z_k are not subgaussian, yet $\langle \overline{x}_k, x-y\rangle$ and $\langle \overline{x}_k, x+y\rangle$ are individually subgaussian (by a previous lemma). We have two bounds

$$\mathbb{P}(|\langle \overline{x}_k, x-y\rangle| > \sqrt{u}\|x-y\|_2) \leq Ce^{-cu}$$
$$\mathbb{P}(|\langle \overline{x}_k, x+y\rangle| > \sqrt{u}\|x+y\|_2) \leq Ce^{-cu},$$

and, therefore, considering the product $|z_k| = |\langle \overline{x}_k, x+y\rangle\langle \overline{x}_k, x-y\rangle|$, we have

$$\mathbb{P}(|z_k| > \sqrt{u}\|x-y\|_2\|x+y\|_2) \leq 2Ce^{-cu}.$$

Since $\|x+y\| \leq 2$,

$$\mathbb{P}(|z_k| > \sqrt{u}\|x-y\|_2) \leq 2Ce^{-c'u};$$

that is, z_k are subexponential. Recall from lecture 7 the Large Deviation Inequality; it says that the sum of subexponential random variables is a mixture of subexponential and subgaussian. That is, in its most concentrated region, the sum is distributed like a gaussian, yet its tails are subexponential. We will use the following corallary.

Corollary 2 (Bernstein's Inequality for Subexponential Random Variables)**.** *If Y_k are i.i.d. subexponential random variables, then*

$$\mathbb{P}(|\frac{1}{m}\sum_{k=1}^{m} Y_k| \geq u) \leq C\exp(-Cm \cdot \min(u, u^2)).$$

We apply this corollary to $Y_k = \frac{z_k}{\|x-y\|_2}$ as follows:

$$\begin{aligned} Y_k &= \frac{z_k}{\|x-y\|_2} \\ &= \mathbb{P}(|\frac{1}{m}\sum_{k=1}^{m} z_k| > u\|x-y\|_2) \\ &= \mathbb{P}(|X_x - X_y| > u\|x-y\|_2) \\ &\leq C\exp(-Cm \cdot \min(u, u^2)) \end{aligned}$$

Dudley's Inequality may equivalently be stated for subgaussian/subexpontial tails as

$$\mathbb{E}\sup_{t\in T} X_t \leq C\int_0^{\infty}\left(\sqrt{\frac{\log(N(t,\epsilon))}{m}} + \frac{\log(N(T,\epsilon))}{m}\right)d\epsilon.$$

Exercise 3. *Determine the appropriate weights in the proof of Dudley's inequality to adopt the proof to the subgaussian/subexponential case.*

In our case, $T = (S^{n-1}, \|\cdot\|_2)$. Therefore, $N(T,\epsilon) \leq (\frac{3}{\epsilon})^n$ (see lecture 6).

$$\begin{aligned}
\mathbb{E}\sup_{t\in T} X_t &\leq C\int_0^\infty \left(\sqrt{\frac{n}{m}\log(N(T,\epsilon))} + \frac{\log(N(T,\epsilon))}{m}\right) d\epsilon \\
&\leq C\int_0^\infty \left(\sqrt{\frac{n}{m}\log(\frac{3}{\epsilon})} + \frac{\log(\frac{3}{\epsilon})}{m}\right) d\epsilon \\
&= C\int_0^{1=\text{diam } T} \left(\sqrt{\frac{n}{m}\log(\frac{3}{\epsilon})} + \frac{\log(N(T,\epsilon))}{m}\right) d\epsilon \\
&= C(\sqrt{\frac{m}{n}} + \frac{n}{m}) \\
&= C'\sqrt{\frac{m}{n}}.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{E}X_x &= \mathbb{E}\left|\frac{1}{m}\|Ax\|_2^2 - 1\right| & (2)\\
&\leq C'\sqrt{\frac{m}{n}} & (3)
\end{aligned}$$

Exercise 4. *Show that equation (1), and hence Theorem 1., follows from (3).*

References

[1] M. Ledoux and M. 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.

[2] M. Talagrand. Majorizing measures: the generic chaining. *The Annals of Probability*, 24(3):1049–1103, 1996.

[3] M. Talagrand. *The generic chaining.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.