ORIE 7790 High Dimensional Probability and Statistics Lecture 7-8 - 02/16/2020

Lecture 8: Random Matrices II: ε-net

Lecturer: Yudong Chen *Scribe: Liwei Jiang*

References:

- R. Vershynin, High dimensional Probability, Sections 4.2, 4.4, 4.6.
- M. J. Wainwright, High-dimensional statistics: A non-asymptotic viewpoint, Sections 6.1, 6.3.

1 ε-net Argument

This argument contains three steps:

- Discretize $\mathbb{S}^{n-1}$.
- Bound $\|Xu\|_2$ for each fixed $u \in \mathbb{S}^{n-1}$.
- Union Bound.

Definition 1 (ε-net). *$T_\varepsilon \subset T$ is called an ε-net of T (w.r.t. ℓ_2 norm) if*

$$\forall u \in T \;\; \exists u_0 \in T_\varepsilon : \|u - u_0\| \leq \varepsilon.$$

Definition 2 (Covering Number). *The smallest cardinality of an ε-net of T is called the covering number of T and denoted by $N(\varepsilon, T)$. The quantity $\log N(\varepsilon, T)$ is called the metric entropy of T.*

Now we have a lemma about covering number of unit sphere.

Lemma 1 (Covering the l_2 ball and sphere). *Recall that*

$$B^n = \{u \in \mathbb{R}^n : \|u\| \leq 1\}, \;\; \mathbb{S}^{n-1} = \{u \in \mathbb{R}^n : \|u\| \leq 1\}.$$

We have

$$N(\varepsilon, \mathbb{S}^{n-1}) \leq N(\frac{\varepsilon}{2}, B^n) \leq (\frac{4}{\varepsilon} + 1)^n$$

Proof Exercise. □

Remark We also have $N(\frac{\varepsilon}{2}, B^n) \geq (\frac{2}{\varepsilon})^n$, so the upper bound is quite tight.

Lemma 2. *For any $\varepsilon \in [0, 1)$ and ε-net $\mathbb{S}_\varepsilon$ of $\mathbb{S}^{n-1}$, we have*

$$\|X\|_{\text{op}} \leq \frac{1}{1-\varepsilon} \sup_{u \in \mathbb{S}_\varepsilon} \|Xu\|_2$$

Proof By compactness of $\mathbb{S}^{n-1}$, we can choose $u \in \mathbb{S}^{n-1}$ such that $\|X\|_{\text{op}} = \|Xu\|_2$. By definition of ε-net, we can find $u_0 \in \mathbb{S}_\varepsilon$ such that $\|u - u_0\|_2 \leq \varepsilon$. Then we have

$$\begin{aligned}
\|X\|_{\text{op}} &= \|Xu\|_2 \\
&\leq \|Xu_0\|_2 + \|Xu - Xu_0\|_2 \\
&\leq \sup_{u \in \mathbb{S}_\varepsilon} \|Xu\|_2 + \varepsilon \|X\|_{\text{op}}
\end{aligned}$$

Rearrange it, we get

$$\|X\|_{\text{op}} \leq \frac{1}{1-\varepsilon} \sup_{u\in\mathbb{S}_\varepsilon} \|Xu\|_2$$

□

With a slight modification of the proof of Lemma 2, we can get

Lemma 3. *Suppose $X \in \mathbb{R}^{n\times n}$ is symmetric. For any $\varepsilon \in [0, \frac{1}{2})$ and ε-net $\mathbb{S}_\varepsilon$ of $\mathbb{S}^{n-1}$, we have*

$$\|X\|_{\text{op}} \leq \frac{1}{1-2\varepsilon} \sup_{u\in\mathbb{S}_\varepsilon} |u^T X u|.$$

Proof Since X is symmetric, we can find $u \in \mathbb{S}^{n-1}$ such that $\|X\|_{\text{op}} = |u^T X u|$. By definition of ε-net, we can find $u_0 \in \mathbb{S}_\varepsilon$ such that $\|u - u_0\|_2 \leq \varepsilon$. Then we have

$$\begin{aligned}
\|X\|_{\text{op}} &= |u^T X u| \\
&= |u_0^T X u_0 + (u-u_0)^T X (u+u_0)| \\
&\leq |u_0^T X u_0| + |(u-u_0)^T X (u+u_0)| \\
&\leq \sup_{u\in\mathbb{S}_\varepsilon} |u^T X u| + 2\varepsilon \|X\|_{\text{op}}
\end{aligned}$$

In the last inequality, we used the fact that $u, u_0 \in \mathbb{S}^{n-1}$ and $\|u - u_0\| \leq \varepsilon$. Rearranging, we get

$$\|X\|_{\text{op}} \leq \frac{1}{1-2\varepsilon} \sup_{u\in\mathbb{S}_\varepsilon} |u^T X u|.$$

□

With all the lemmas above, we have the following theorem about operator norm of random matrix with independent sub-Gaussian columns.

Theorem 1. *Suppose $X \in \mathbb{R}^{m\times n}$, whose columns $X_j \in \mathbb{R}^m$ are independent, zero-mean, isotropic ($\mathbb{E}\left[X_j X_j^T\right] = I_m$) and sub-Gaussian with parameter σ^2. Then*

$$\left\|\frac{1}{n} X X^T - I_m\right\|_{\text{op}} \lesssim \sigma^2 \max\{\sqrt{\frac{m}{n}}, \frac{m}{n}\}$$

with probability at least $1 - 2e^{-m}$. Consequently,

$$\sqrt{n} - c\sigma\sqrt{m} \leq s_m(X) \leq s_1(X) \leq \sqrt{n} + c\sigma\sqrt{m}.$$

Here $s_1(X)$ and $s_m(X)$ are largest and smallest singular values of X, respectively. c is some constant.

Proof Fix $\varepsilon = \frac{1}{4}$, let $\mathbb{S}_\varepsilon$ be the smallest ε-net of $\mathbb{S}^{m-1}$. By lemma 1, we know $|\mathbb{S}_\varepsilon| \leq 17^m$. By lemma 3, we know

$$\begin{aligned}
\left\|\frac{1}{n} X X^T - I_m\right\|_{\text{op}} &\leq 2 \max_{u\in\mathbb{S}_\varepsilon} |u^T (\frac{1}{n} X X^T - I_m) u| \\
&= 2 \max_{u\in\mathbb{S}_\varepsilon} |\frac{1}{n} \left\|X^T u\right\|_2^2 - 1| \qquad (*)
\end{aligned}$$

For fixed $u \in \mathbb{S}_\varepsilon$,

$$\left\|X^T u\right\|_2^2 = \sum_{j=1}^{n} \langle X_j, u\rangle^2.$$

Let $Z_j = \langle X_j, u\rangle$, note that Z_j's are independent of each other, σ^2 sub-Gaussian by definition and $\mathbb{E}\left[Z_j^2\right] = 1$, and since the square of a sub-Gaussian random variable is sub-Exponential, we have

$$\left\|Z_j^2 - 1\right\|_{\psi_1} \lesssim \sigma^2.$$

By Bernstein Inequality, for $t = c\sigma^2 \max\{\sqrt{\frac{m}{n}}, \frac{m}{n}\}$ with sufficient large c, we have

$$\begin{aligned}
\mathbb{P}\left[|\frac{1}{n}\left\|X^T u\right\|_2^2 - 1| \geq \frac{t}{2}\right] &\leq 2\exp(-c'n\min\{\frac{t^2}{\sigma^4}, \frac{t}{\sigma^2}\}) \\
&\leq 2\exp(-c''n\frac{m}{n}) \\
&= 2\exp(-c''m)
\end{aligned}$$

Here c' and c'' are other constants which we don't specify. c'' can be sufficiently large as long as c is sufficiently large. Now, by Union bound and *, we know

$$\begin{aligned}
\mathbb{P}\left[\left\|\frac{1}{n}XX^T - I_m\right\|_{\text{op}} \geq t\right] &\leq \mathbb{P}\left[\max_{u\in\mathbb{S}_\varepsilon} |\frac{1}{n}\left\|X^T u\right\|_2^2 - 1| \geq \frac{t}{2}\right] \\
&\leq 2\cdot 17^m \exp(-c''m) \\
&\leq 2e^{-m}
\end{aligned}$$

The last inequality holds for sufficient large c''.
The proof of "Consequently" part is left as an exercise. So we are done. □

Remark

- In this proof, we only require independent columns(rather that entries) of X.
- We get two-sided bounds on largest/smallest singular values.
- If $m \ll n$, in which case the matrix is very "rectangular", then

$$s_m(X) \geq \sqrt{n} - c\sigma\sqrt{m} > 0 \tag{1}$$

$$\left\|\frac{1}{n}XX^T - I_m\right\|_{\text{op}} \lesssim \sqrt{\frac{m}{n}} < 1 \tag{2}$$

  The first inequality says X is non-singular, and the second one implies XX^T is invertible. Both statements are true with high probability.

- If $X_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$ and $n = m$, then

$$\|X\|_{\text{op}} \leq c\sqrt{n} \quad w.p. \geq 1 - 2e^{-n}$$

  This recovers the Gaussian matrix result from last lecture, but with worse constant. Note that the previous bound is:

$$\|X\|_{\text{op}} \leq (2+\varepsilon)\sqrt{n} \quad w.p. \geq 1 - 2e^{-\frac{\varepsilon^2 n}{2}}$$

- If X_{ij}'s are sub-Exponential, what can you derive using ε-net?

2 Applications to Covariance Estimation

Suppose we observe $Y_1, Y_2, \ldots, Y_n \in \mathbb{R}^m$ sampled i.i.d. from $N(0, \Sigma)$.
Goal: to estimate the covariance matrix Σ.
We use the empirical covariance:

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} Y_i Y_i^T.$$

Note that $\mathbb{E}\left[\hat{\Sigma}\right] = \Sigma$, so it's unbiased. We want to bound $\left\|\hat{\Sigma} - \Sigma\right\|_{\mathrm{op}}$. By change of variables, let

$$Y_i = \Sigma^{\frac{1}{2}} X_i, \text{ where } X_i \sim N(0, I_m)$$

Let $X \triangleq (X_1, X_2, \ldots X_n)$, then $X \in \mathbb{R}^{m \times n}$ has isotropic, independent, sub-Gaussian columns. By Theorem 1, we know

$$\left\|\frac{1}{n}XX^T - I_m\right\|_{\mathrm{op}} \lesssim \sqrt{\frac{m}{n}} + \frac{m}{n}$$

with probability at least $1 - 2e^{-m}$. Consequently, with probability at least $1 - 2e^{-m}$, we have

$$\begin{aligned}
\left\|\hat{\Sigma} - \Sigma\right\|_{\mathrm{op}} &= \left\|\frac{1}{n}YY^T - \Sigma\right\|_{\mathrm{op}} \\
&= \left\|\frac{1}{n}\Sigma^{\frac{1}{2}}XX^T\Sigma^{\frac{1}{2}} - \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}\right\|_{\mathrm{op}} \\
&\leq \left\|\Sigma^{\frac{1}{2}}\right\|_{\mathrm{op}}^2 \left\|\frac{1}{n}XX^T - I_m\right\|_{\mathrm{op}} \\
&\lesssim \|\Sigma\|_{\mathrm{op}} \left(\sqrt{\frac{m}{n}} + \frac{m}{n}\right)
\end{aligned}$$

Remark

- Sample complexity: $n \gtrsim \frac{m}{\varepsilon^2} \Rightarrow \frac{\|\hat{\Sigma}-\Sigma\|_{\mathrm{op}}}{\|\Sigma\|_{\mathrm{op}}} \leq \varepsilon$. So when $n \gg \frac{m}{\varepsilon^2}$, $\frac{\|\hat{\Sigma}-\Sigma\|_{\mathrm{op}}}{\|\Sigma\|_{\mathrm{op}}} \to 0$(consistency.)
- It can be generalized to the case in which Y_i's are sub-Gaussian.