Lecture 4: Basic Random Matrix Theory/ Spectral method for Planted Clique

Lecturer: Yihong Wu *Scribe: Jiyi Liu, Sep 19, 2018*

Let $Z = (Z_{ij})_{n\times n}$ be a real symmetric matrix. Three main goals for this lecture:

1. When $Z_{ij} \sim_{i.i.d} N(0,1)$ for $i < j$, show rigorously that $\|Z\|_{\text{op}} \le C\sqrt{n}$ w.h.p. with ε-net argument. We have seen intuitions for this from last lecture.
2. Extend the result to sub-Gaussian r.v.s.
3. Apply to the hidden clique problem.

4.1 Gaussian Random Matrix

For simplicity, let's consider Z having independent $N(0,1)$ off-diagonals and $N(0,2)$ diagonals. It will become transparent that the variance of the diagonal is immaterial, provided it is small, say, a constant. This model is referred to as Gaussian Orthogonal Ensemble (GOE).

$$\|Z\|_{\text{op}} = \sigma_{\max} = \max_{\|v\|_2=1} |\langle Z, vv^T\rangle|.$$

For $\forall$ fixed $v \in \mathbb{S}^{n-1}$,

$$\langle Z, vv^T\rangle = \sum_i Z_{ii}v_i^2 + 2\sum_{i<j} Z_{ij}v_iv_j \sim N(0, 2\sum_i v_i^4 + 4\sum_{i<j} v_i^2 v_j^2) = N(0,2).$$

$$\Rightarrow \mathbb{P}(|\langle Z, vv^T\rangle| > t) \le 2e^{-\frac{t^2}{4}}, \forall t > 0.$$

Remark 4.1. The distributions of the diagonals are not important for the operator norm. To see this, note

$$\|Z\|_{\text{op}} \le \|Z_o\|_{\text{op}} + \|\text{diag}(Z)\|_{\text{op}}$$

where Z_o is the same as Z except the diagonals are set to zero, and $\text{diag}(Z) = \text{diag}(Z_{ii})$. By union bound, $\|Z_o\|_{\text{op}} = \max_{1\le i\le n} |Z_i| = O_p(\sqrt{\log n}) \ll O_p(\sqrt{n})$ thus negligible.

To bound $\mathbb{P}(\max_{v\in\mathbb{S}^{n-1}} |\langle Z, vv^T\rangle| > t)$, we would like to apply the union bound. However, the $\mathbb{S}^{n-1}$ here is continuous and $|\mathbb{S}^{n-1}| = \infty$. In order to handle this, we use the discretization technique — the ε-net argument.

Definition 4.1. $V \subset \mathbb{S}^{n-1}$ is called an ε-net (covering), if $\forall u \in \mathbb{S}^{n-1}$, $\exists v \in V$ s.t. $\|u - v\|_2 \le \varepsilon$.

Lemma 4.1. *For V an ε-net,*

$$\max_{v\in V} |\langle Z, vv^T\rangle| \le \|Z\|_{op} \le \frac{1}{1-2\varepsilon} \max_{v\in V} |\langle Z, vv^T\rangle|.$$

Proof. Choose $u \in \mathbb{S}^{n-1}$ such that $|\langle Z, uu^T\rangle| = \|Z\|_{\text{op}}$. $\exists v \in V$, $\|u-v\|_2 \le \varepsilon$.

$$\begin{aligned}\|Z\|_{\text{op}} = |\langle Z, uu^T\rangle| &\le |\langle Z, vv^T\rangle| + |\langle Z, uu^T - vv^T\rangle| \\ &= |\langle Z, vv^T\rangle| + |\langle Z, uu^T - uv^T + uv^T - vv^T\rangle| \\ &\le |\langle Z, vv^T\rangle| + |\langle Z, u(u-v)^T\rangle| + |\langle Z, (u-v)v^T\rangle| \\ &\le |\langle Z, vv^T\rangle| + 2\|Z(u-v)\|_2 \\ &\le \max_{v\in V} |\langle Z, vv^T\rangle| + 2\varepsilon\|Z\|_{\text{op}}.\end{aligned}$$

□

Definition 4.2. For $A \subset \mathbb{R}^d$, $V = \{v_1, \dots, v_m\} \subset A$ is called an ε-packing, if $\forall i \ne j$, $\|v_i - v_j\|_2 \ge \varepsilon$.

Definition 4.3. An ε-packing V is a maximal packing, if $\forall u \in A \backslash V$, $V \cup \{u\}$ is not an ε-packing.

We make two key observation for these concepts:

- Any maximal ε-packing is an ε-net.
- $\forall$ ε-packing V of A, $|V| \le Vol(A + \frac{\varepsilon}{2}B)/Vol(\frac{\varepsilon}{2}B)$. Here the sum of two sets $A + B := \{x + y : x \in A, y \in B\}$ is the Minkowski sum.

The first observation is just by definition. We can construct a maximal ε-packing through greedy search. The second one is because we can put $|V|$ balls of radius $\frac{\varepsilon}{2}$ into $A + \frac{\varepsilon}{2}B$ and keep them disjoint. So the total volume of balls should not exceed that of the $A + \frac{\varepsilon}{2}B$. Among many measures of objects, we choose volume because it's location invariant. We can summarize the observations as

covering # ≤ packing # ≤ volume ratio.

Now set $A = \mathbb{S}^{n-1}$. Then $A + \frac{\varepsilon}{2}B \subset B + \frac{\varepsilon}{2}B = (1+\frac{\varepsilon}{2})B$.[1] The volume ratio

$$\frac{Vol(A + \frac{\varepsilon}{2}B)}{Vol(\frac{\varepsilon}{2}B)} \le \frac{Vol((1+\frac{\varepsilon}{2})B)}{Vol(\frac{\varepsilon}{2}B)} = \frac{(1+\frac{\varepsilon}{2})^n Vol(B)}{(\frac{\varepsilon}{2})^n Vol(B)} = (1 + \frac{2}{\varepsilon})^n.$$

What we discussed above concludes the following lemma.

Lemma 4.2 (Size of ε-net)**.** *There exists an ε-net V for S^{n-1}, of size $|V| \le (1+\frac{2}{\varepsilon})^n$.*

Theorem 4.1. $\|Z\|_{op} \le C\sqrt{n}$ *w.h.p..*

Proof. Set $\varepsilon = \frac{1}{4}$ and choose V as in Lemma 4.2. By Lemma 4.1, $\|Z\|_{\text{op}} \le 2\max_{v\in V} |\langle Z, vv^T\rangle|$. For $\forall t > 0$,

$$\begin{aligned}\mathbb{P}(\max_{v\in V} |\langle Z, vv^T\rangle| > t) &\le \sum_{v\in V} \mathbb{P}(|\langle Z, vv^T\rangle| > t) \\ &\le |V| \cdot 2e^{-\frac{t^2}{4}} = 2e^{n\log 9 - \frac{t^2}{4}}.\end{aligned}$$

Choose $t = \frac{C}{2}\sqrt{n}$ with $C > 4\sqrt{\log 9}$ a universal constant, then we know $\|Z\|_{\text{op}} \le C\sqrt{n}$ with probability at lesat $1 - 2e^{-C'n}$, where $C' = C^2/16 - \log 9 > 0$.

□

[1] The first inclusion does not lose much volume, because the volume of a ball in high dimension is concentrated near the shell anyway.

4.2 sub-Gaussian Random Matrix

Reviewing the whole proof of Theorem 4.1, we can see there is only one part that the Gaussian assumption is used: the tail bound $\mathbb{P}(|\langle Z, vv^T\rangle| > t) \leq 2e^{-\frac{t^2}{4}}$. Thus the result of Theorem 4.1 can be naturally extended to other r.v.s with such tail bound.

Definition 4.4. A r.v. X is sub-Gaussian (SG) with parameter σ^2 if $\forall\lambda$, $\mathbb{E}e^{\lambda(X-\mathbb{E}X)} \leq e^{\sigma^2\lambda^2/2}$.

For a σ^2-SG r.v. X and $t > 0$, a direct result of Chernoff bound is $\mathbb{P}(X - \mathbb{E}X > t) \leq e^{\sigma^2\lambda^2/2-\lambda t}, \forall\lambda$. Choose $\lambda = t/\sigma^2$, and we have $\mathbb{P}(X - \mathbb{E}X > t) \leq e^{-\frac{t^2}{2\sigma^2}}$. The similar result for the other side tail combines to show $\mathbb{P}(|X - \mathbb{E}X| > t) \leq 2e^{-\frac{t^2}{2\sigma^2}}$. Thus, a σ^2-SG r.v. do have the same tail bound as $N(0, \sigma^2)$. We can also view the tail bound as the definition of σ^2-SG. σ^2-SG r.v.s have variance at most σ^2, which can be shown easily through Taylor expansion.

Lemma 4.3 (Hoeffding)**.** *Bounded r.v.s are SG. If $X \in [-a, a]$ a.s. for some $a > 0$, then it's $4a^2$-SG.*

Proof. First, we prove for X a Rademacher r.v..

$$\mathbb{E}e^{\lambda X} = \frac{1}{2}(e^{\lambda} + e^{-\lambda}) = \sum_{k\geq 0, 2|k} \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} \leq \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{2^k k!} = e^{\lambda^2/2}.$$

Second, when $|X| \leq a$ a.s., we apply so-called symmetrization technique. Let $X' \sim X$ and ε be a Rademacher r.v., and they three are independent. Then $X - X'$ has symmetric distribution and $\varepsilon(X - X') =_D X - X'$.

$$\begin{aligned}
\mathbb{E}e^{\lambda(X-\mathbb{E}X)} &= \mathbb{E}e^{\lambda(X-\mathbb{E}X')} = \mathbb{E}e^{\lambda X}e^{-\lambda\mathbb{E}X'} \\
&\leq \mathbb{E}e^{\lambda X}\mathbb{E}e^{-\lambda X'} = \mathbb{E}e^{\lambda(X-X')} \\
&= \mathbb{E}e^{\lambda\varepsilon(X-X')} \\
&= \mathbb{E}\left(\mathbb{E}(e^{\lambda\varepsilon(X-X')}|X, X')\right) \\
&\leq \mathbb{E}\left(e^{\lambda^2(X-X')^2/2}\right) \leq e^{2\lambda^2 a^2}.
\end{aligned}$$

The last inequality is because $|X - X'| \leq 2a$ a.s.. $\square$

Remark 4.2. This bound is good (tight enough) if $Var(X) \asymp a^2$ and is loose when $Var(X) = o(a^2)$. For example, let's consider $X \sim \text{Bern}(p)$. By Lemma 4.3, it's 1-SG regardless of the value of p. But when $p = o(1)$, $Var(X) \approx p = o(1)$, so X will be strongly concentrated around 0 and the tail bound cannot be tight. For this specific example, we can improve the result as $\sigma^2(p) = \Omega(p \log \frac{1}{p})$, where $\sigma(\cdot)$ is sub-Gaussian norm.

Lemma 4.4.

1. *If X is σ^2-SG, then αX is $\alpha^2\sigma^2$-SG.*
2. *If $X_1, \ldots, X_n$ are independent σ_i^2-SG, then $\sum_{i=1}^n X_i$ is $(\sum_{i=1}^n \sigma_i^2)$-SG.*
3. *If X is σ^2-SG, $\tau^2 > \sigma^2$, then X is τ^2-SG.*

Theorem 4.2. *Suppose $Z = (Z_{ij})_{n\times n}$ a real symmetric matrix with $\mathbb{E}Z = 0$. For $i \le j$, $Z_{ij} \sim \sigma^2$-SG and are independent. Then $\|Z\|_{op} \le C\sqrt{n\sigma^2}$ with probability at least $1 - 2e^{-C'n}$ for some universal constant C, C'.*

Proof. For $\forall v \in \mathbb{S}^{n-1}$, $\langle Z, vv^T\rangle = \sum_i Z_{ii}v_i^2 + 2\sum_{i<j} Z_{ij}v_iv_j$ is SG with parameter

$$\sigma^2\left(\sum_i v_i^4 + 4\sum_{i<j} v_i^2v_j^2\right) \le 2\sigma^2\left(\sum_i v_i^2\right)^2 = 2\sigma^2.$$

We have $\mathbb{P}(|X - \mathbb{E}X| > t) \le 2e^{-\frac{t^2}{4\sigma^2}}$. The rest is identical to the proof of Theorem 4.1. □

4.3 Spectral methods for Planted Clique Model

We now apply the random matrix theory above to the planted clique model. Let $G \sim G(\frac{1}{2}, n, k)$. In other words, $K \subset [n]$ is a hidden k-clique, and G has adjacency matrix

$$A_{ij} = \begin{cases} 1 & i, j \in K \\ \text{Bern}(\frac{1}{2}) & o/w \end{cases}.$$

Let W be a real symmetric matrix that

$$W_{ij} = \begin{cases} 2A_{ij} - 1 & i \ne j \\ 0 & i = j \end{cases}.$$

The following spectral method to find the clique is by [AKS98]:

1. Find the top eigenvector u of W.
2. Let $\tilde{K}$ be the index vector of k largest $|u_i|$.
3. (Clean up) Define $\hat{K}$ = the set of vertices in G having $\ge \frac{3k}{4}$ neighbors in $\tilde{K}$. In other words, $\hat{K} = \{v \in G : d_{\tilde{K}}(v) \ge \frac{3k}{4}\}$.

Theorem 4.3 ([AKS98])**.** *$k \ge C\sqrt{n}$ for C large enough, $\mathbb{P}(\hat{K} = K) \to 1$.*

Proof. First, we show $\tilde{K}$ is approximately correct: $|\tilde{K} \cap K| \ge (1 - \varepsilon)k$ w.h.p for some $\varepsilon = \varepsilon(C)$. Fix some small $\varepsilon > 0$ that we will choose later. Let $W^* = \xi\xi^T$, $\xi = I\{K\} = (I\{i \in K\})_{1\le i\le n}$. Now W^* is rank one, so $v = \frac{1}{\sqrt{k}}\xi$ is its top eigenvector. By $\sin\Theta$-theorem (Davis-Kahan): Providing $\lambda_1(W^*) - \lambda_2(W) > 0$,

$$\min_{\pm} \|u \pm v\|_2 \le \frac{\|W - W^*\|_{\text{op}}}{\lambda_1(W^*) - \lambda_2(W)}. \tag{4.1}$$

WLOG, assume the LHS is $\|u - v\|_2$. $\|W - W^*\|_{\text{op}} \le \|\mathbb{E}W - W\|_{\text{op}} + \|\mathbb{E}W - W^*\|_{\text{op}} \le \|\mathbb{E}W - W\|_{\text{op}} + 1 \le C_0\sqrt{n} + 1$ w.h.p for some universal $C_0 > 1$ by Theorem 4.2. By Weyl's inquality,

$|\lambda_2(W)| = |\lambda_2(W^*) - \lambda_2(W)| \le \|W - W^*\|_{\text{op}} \le C_0\sqrt{n} + 1$ under the event above. And we know $\lambda_1(W^*) = \|\xi\|_2^2 = k$. Plug back into (4.1),

$$\|u - v\|_2 \le \frac{C_0\sqrt{n} + 1}{C\sqrt{n} - C_0\sqrt{n} - 1} \le \varepsilon \tag{4.2}$$

w.h.p for C big enough.

Second, $\|u - v\|_2 \le \varepsilon$ actually implies

$$|\tilde{K} \cap K| \ge (1 - \varepsilon')k. \tag{4.3}$$

To see this, note that $|K| = |\tilde{K}| = k$, thus $|K \backslash \tilde{K}| = |\tilde{K} \backslash K|$.

$$\varepsilon^2 \ge \|u - v\|_2^2 = \sum_{i \in K} (u_i - \frac{1}{\sqrt{k}})^2 + \sum_{i \notin K} u_i^2.$$

If $|u_i| \le \frac{1}{2\sqrt{k}}$, $\forall i \notin \tilde{K}$, then

$$\varepsilon^2 \ge \sum_{i \in K \backslash \tilde{K}} (\frac{1}{\sqrt{k}} - u_i)^2 \ge \frac{1}{4k}|K \backslash \tilde{K}|.$$

If $\exists j \notin \tilde{K}$, $|u_j| > \frac{1}{2\sqrt{k}}$, by the definition of $\tilde{K}$, $\forall i \in \tilde{K}$, $|u_i| > \frac{1}{2\sqrt{k}}$.

$$\varepsilon^2 \ge \sum_{i \in \tilde{K} \backslash K} u_i^2 \ge \frac{1}{4k}|\tilde{K} \backslash K|.$$

In all, in either case, (4.3) holds with $\varepsilon' = 8\varepsilon^2$.

Third, we claim $\hat{K} = K$ with high probability. Think under the event of $\|u - v\| \le \varepsilon$. If $v \in K$, $d_{\tilde{K}}(v) \ge d_{\tilde{K} \cap K}(v) \ge |\tilde{K} \cap K| - 1 \ge (1 - \varepsilon')k$. So $v \in \hat{K}$ when $\varepsilon' < \frac{1}{4}$. If $v \notin K$, $d_{\tilde{K}}(v) \le d_K(v) + |\tilde{K} \backslash K|$. From above, we know $|\tilde{K} \backslash K| \le \frac{\varepsilon'}{2}k$. And $d_K(v) \sim \text{Bin}(k, \frac{1}{2})$. By Chernoff bound,

$$\mathbb{P}(d_K(v) \ge (\frac{3}{4} - \frac{\varepsilon'}{2})k) \le \mathbb{P}(d_K(v) \ge \frac{5}{8}k) \le e^{-\frac{k}{32}}.$$

In all, under events $\|u - v\| \le \varepsilon$ and $d_K(v) \le (\frac{3}{4} - \frac{\varepsilon'}{2})k$, we have $\hat{K} = K$. To wrap up the whole proof, we choose $\varepsilon = \frac{1}{8}$. Then $\varepsilon' = 8\varepsilon^2 = \frac{1}{8}$. Choose $C \ge 17C_0$ then the second inequality in (4.2) is guaranteed.

$$\begin{aligned}
\mathbb{P}(\hat{K} \neq K) &\le \mathbb{P}(\|u - v\| > \varepsilon) + \mathbb{P}(d_K(v) > (\frac{3}{4} - \frac{\varepsilon'}{2})k) \\
&\le \mathbb{P}(\|W - \mathbb{E}W\|_{\text{op}} > C_0\sqrt{n}) + e^{-\frac{k}{32}} \\
&\le 2e^{-C_0' n} + e^{-\frac{C}{32}\sqrt{n}} \to 0.
\end{aligned}$$

□

Remark 4.3.

1. An alternative algorithm can take u as the second leading eigenvector of A. The top eigenvector of A is almost deterministic and not informative, since it is almost $\propto \mathbf{1}$.
2. Thresholding technique is widely used in non-parametric estimation. Here, the step 3 (clean up) can be viewed as a version of thresholding.

4.4 Improving the constant

Next we show that the constant C in Theorem 4.3 can be made arbitrarily small, at the pricing of increasing the time complexity (still $\mathsf{poly}(n)$ but with bigger exponent). The part is generic and applies to any algorithm. The idea is as following. Fix a subset of vertices $S \subset V$, $|S| = s$. Define $N_*(S)$ as the common neighbor of S, or $N_*(S) = \{v \in V \backslash S : \exists u \in S, v \sim u\} = \bigcap_{u \in S} N(u) \backslash S$. Let's say $s = 2$. Then $N_*(S) \sim \text{Bin}(n-2, \frac{1}{4}) \approx \frac{n}{4}$. Next, look at the induced subgraph $G' = G[N_*(S)]$, $|V'| \approx \frac{n}{4}$. If $S \subset K$, then $G' = G(N_*(S), \frac{1}{2}, k-2)$. So as we can see, by this subgraph operation, n decays exponentially while k decays linearly. The upgraded algorithm is summarized below:

> Search for $\forall S \subset V$, $|S| = s$. Run the existing algorithm on $G' = G[N_*(S)]$ and output Q. Iterate until $S \cup Q$ is a k-clique. And the final output is $S \cup Q$.

When the search over S finds $S \subset K$, the requirement in Theorem 4.3 asks for $k - s \geq C\sqrt{n \cdot 2^{-s}}$ to guarantee consistency of Q, thus also guarantees the consistency of $S \cup Q$ in the original graph G. Pick $s \approx 2 \log_2 \frac{C}{\delta}$, then the algorithm above is guaranteed to be consistent with requirement $k \geq \delta\sqrt{n}$. The extra search time is at most $\binom{n}{s} \asymp n^s$ that is polynomial in n.

References

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998.