CS 839 Probability and Learning in High Dimension	Lecture 9 - 02/23/2022

Lecture 9: Convex Relaxation for Community Detection II

Lecturer: Yudong Chen *Scribe: Shuyao Li*

In this lecture,[1] we will introduce explicit clustering error bounds based on the ℓ_1-bound that we derived in the last elcture. We'll also present a simple proof of the Grothendieck's inequality for $K \approx 1.783$ based on the angular kernel trick.

1 Explicit Clustering Error Bounds

Recall the stochastic block model with n nodes, k truth clusters represented by Y^*, and in/cross-cluster edge probability p and q. In the class, we considered an estimator $\hat{Y}$ given by convex SDP relaxation, and establish the following for recovering Y^*.

Theorem 1. *If $p \geq \frac{1}{n}$, then with probability $\geq 1 - 2\left(\frac{2}{e}\right)^n$, we have*

$$\frac{1}{n^2}\left\|\hat{Y} - Y^*\right\|_1 \lesssim \sqrt{\frac{p}{(p-q)^2 n}}.$$

Below we focus on the special case with $k = 2$ clusters. To extract from $\hat{Y}$ an explicit estimate of the clusters, we use the following spectral algorithm.

1. Compute the top unit norm singular vector $\hat{u}$ of $\hat{Y} - \frac{1}{2}$
2. For each i, assign node i to $\begin{cases} \text{the first cluster} & \text{if } \hat{u}_i^* \geq 0 \\ \text{the second cluster} & \text{if } \hat{u}_i^* < 0 \end{cases}$

We now derive the guarantee for this spectral recovering algorithm.

Let u^* be the top unit norm singular vector of $Y^* - \frac{1}{2}$, then we have

$$u_i^* = \begin{cases} \frac{1}{\sqrt{n}} & \text{if } i \text{ is in cluster 1} \\ -\frac{1}{\sqrt{n}} & \text{otherwise} \end{cases}$$

Lemma 1. *If $\frac{1}{n^2}\|\hat{Y} - Y^*\|_1 \leq \frac{1}{20}$, then*

$$\frac{1}{n}\min_{\alpha\in\{\pm 1\}} |\{i : \operatorname{sgn}(\hat{u}_i) \neq \alpha \operatorname{sgn}(u_i^*)\}| \lesssim \frac{1}{n^2}\|\hat{Y} - Y^*\|_1$$

Note that we use α to tolerate a global sign flip of singular vector. The left hand side is the fraction of nodes that are incorrectly classified. Lemma 1 together with Theorem 1 provides an explicit upper bound on the clustering error.

Proof WLOG assume $\alpha = 1$ and sort the nodes so that $u^* = (\frac{1}{\sqrt{n}} \cdots \frac{1}{\sqrt{n}}, -\frac{1}{\sqrt{n}} \cdots -\frac{1}{\sqrt{n}})^\top$.

[1] *Reading:*

- Analysis of SDP relaxation:[Guédon and Vershynin, 2016]
- Proof of Grothendieck's inequality Sec 3.5–3.7 of [Vershynin, 2018]
- High-rank Grothendieck's inequality and applications: [Montanari and Sen, 2016]
- Additional background: [Khot and Naor, 2012]

Our strategy is to apply Wedin's SinΘ Theorem, for which we need to compute the singular value gap and the perturbation magnitude $\|\hat{Y} - Y^*\|_{op}$.

Since $Y^* - \frac{1}{2}$ is rank-1, we have $\sigma_1(Y^* - \frac{1}{2}) = \|Y^* - \frac{1}{2}\|_F = \frac{n}{2\sqrt{2}}$ and $\sigma_2(Y^* - \frac{1}{2}) = 0$, so gap $= \frac{n}{2\sqrt{2}}$. Moreover, we have

$$\|\hat{Y} - Y^*\|_{op}^2 \le \|\hat{Y} - Y^*\|_F^2 = \sum_{i,j} |\hat{Y}_{ij} - Y^*_{i,j}|^2 \le \sum_{i,j} 2|\hat{Y}_{ij} - Y^*_{ij}| = 2\|\hat{Y} - Y^*\|_1 \le \frac{1}{10}n^2$$

It follows that

$$\begin{aligned}
\|\hat{u} - u^*\|_2 &= 2\sin(\angle(\hat{u}, u^*)/2) \\
&\le 2\sin(\angle(\hat{u}, u^*)) \\
&\le 2\frac{\|\hat{Y} - Y^*\|_{op}}{\text{gap} - \|\hat{Y} - Y^*\|_{op}} \\
&\lesssim \frac{\sqrt{\|\hat{Y} - Y^*\|_1}}{\frac{n}{2\sqrt{2}} - \frac{n}{\sqrt{10}}} \\
&\lesssim \sqrt{\frac{\|\hat{Y} - Y^*\|_1}{n^2}},
\end{aligned} \tag{1}$$

where the first equality is high-school geometry and the second inequality is Wedin's SinΘ Theorem.

To proceed, we make use of the following inequality.

Exercise 2.

$$\|\hat{u} - u^*\|_2^2 \gtrsim \frac{1}{n}|\{i : \text{sgn}(\hat{u}_i) \neq \text{sgn}(u^*_i)| \tag{2}$$

Proof is left as exercise. Hint: $\text{sgn}(\hat{u}_i) \neq \text{sgn}(u^*_i)$ implies that $|\hat{u}_i - u^*_i| \ge \frac{1}{\sqrt{n}}$.

Combining Equation (1) and (2) gives

$$\frac{1}{n}\{i : \text{sgn}(\hat{u}_i) \neq \text{sgn}(u^*_i)\}| \lesssim \frac{1}{n^2}\|\hat{Y} - Y^*\|_1$$

as claimed. □

2 Comparison with Spectral Method

In the proof of Theorem 1, we showed the intermediate inequality

$$\frac{p-q}{2}\left\|\hat{Y} - Y^*\right\|_1 \le 2\max_{\substack{Y \succeq 0 \\ Y_{ii}=1}} |\langle Y, A - \mathbb{E}A\rangle| \tag{3}$$

and them control RHS by Grothendieck's inequality, followed by a high probability bound via Bernstein's inequality and union bound.

Alternatively, we may use the duality between nuclear and spectral norms:

$$\max_{\substack{Y \succeq 0 \\ Y_{ii}=1}} |\langle Y, A - \mathbb{E}A\rangle| \le \max_{\substack{Y \succeq 0 \\ Y_{ii}=1}} \|Y\|_* \|A - \mathbb{E}\,A\|_{op} = n\|A - \mathbb{E}\,A\|_{op} \lesssim n(\sqrt{pn\log n} + \log n), \tag{4}$$

where the first inequality is generalized Hölder's inequality for nuclear and spectral norms, and the last inequality is due to matrix Bernstein and holds with high probability. The matrix Bernstein turns out to be not tight for controlling $\|A - \mathbb{E}\,A\|_{op}$. One may use more sophisticated method to remove a $\sqrt{\log n}$ factor:

$$\|A - \mathbb{E}\,A\|_{op} \lesssim \sqrt{pn} + \sqrt{\log n}. \tag{5}$$

However, the remaining $\sqrt{\log n}$ factor cannot be avoided, since

$$\begin{aligned}
\|A - \mathbb{E}\, A\|_{op} &\geq \max_i \|(A - \mathbb{E}\, A)_{i-}\|_2 && \text{(largest } \ell_2 \text{ norm of a row)} \\
&\gtrsim \sqrt{\log n} && \text{with high probability when } p \gtrsim \tfrac{1}{n}.
\end{aligned}$$

The last inequality is left as an exercise. Note that in the sparse graph regime $p \asymp \frac{1}{n}$, even though the *expected* degree is bounded, $\mathbb{E} \sum_j A_{ij} = pn \asymp 1$, the *maximum* degree is unbounded, $\max_i \sum_j A_{ij} \asymp \log n$ with high probability.

Combining Equation (3), (4), and (5) gives

$$\frac{1}{n^2}\|\hat{Y} - Y^*\|_1 \lesssim \frac{1}{n^2}\frac{n(\sqrt{pn} + \sqrt{\log n})}{p-q} \lesssim \sqrt{\frac{p}{n(p-q)^2}} + \sqrt{\frac{\log n}{n^2(p-q)^2}} \tag{6}$$

Note that the first term is the same as in Theorem 1 but the additional second term makes the bound useful only when $p \gtrsim \frac{\sqrt{\log n}}{n}$, which is not sparse regime since the expected degree $\asymp \sqrt{\log n}$.

Note that similar analysis applies to the naive spectral method, that is, clustering based on the eigenvectors of the graph adjacency matrix or graph Laplacian. Note that the top eigenvector of the shifted adjacency matrix $A - \frac{p+q}{2}$ coincides with the solution of the following convex program:

$$\begin{aligned}
&\max \langle Y, A - \frac{p+q}{2} \rangle \\
&s.t.\ Y \succeq 0 \\
&\quad tr(Y) = n.
\end{aligned}$$

In comparsion, Theorem 1 corresponds to the following formulation:

$$\begin{aligned}
&\max \langle Y, A - \frac{p+q}{2} \rangle \\
&s.t.\ Y \succeq 0 \\
&\quad Y_{ii} = 1, \forall i \\
&\quad Y_{ij} \in [0,1], \forall i, j.
\end{aligned}$$

Note that the diagonal constraint $Y_{ii} = 1, \forall i$ is significantly stronger than the trace constraint $tr(Y) = n$.

3 Proof of Grothendieck's inequality

As we have seen in the last lecture, the proof of Theorem 1 relies on the Grothendieck's inequality.

Theorem 3 (Grothendieck's Inequality)**.** *For any $A \in \mathbb{R}^{n\times n}$, it holds that*

$$\max_{u_i, v_j : \|u_i\|_2 = \|v_j\|_2 = 1} \left| \sum_{\forall i,j=1,\dots,n,} A_{ij} \langle u_i, v_j \rangle \right| \leq K \max_{x_i, y_j \in \{\pm 1\}} \left| \sum_{i,j} A_{ij} x_i y_j \right|$$

where $K = \frac{\pi}{2\log(1+\sqrt{2})} \approx 1.783$.

Remark The smallest possible value for K is called the Grothendieck's constant and is denoted by K_G. It has been proved that K_G is strictly smaller than $\frac{\pi}{2\log(1+\sqrt{2})}$, but the latter remains the best known upper bound on K_G. See [Khot and Naor, 2012] for a survey of the use of Grothendieck's inequality in combinatorial optimization.

We give a proof of Theorem 3 for $K_G \leq K = \frac{\pi}{2\log(1+\sqrt{2})}$. The key step of the proof is the following lemma, where $S(H)$ denote the unit sphere of a Hilbert space H.

Lemma 2. *There exists a Hilbert space H and functions $\Phi, \Psi : S(\mathbb{R}^n) \to S(H)$ such that*

$$\langle u, v\rangle = \frac{2K}{\pi} \arcsin\langle \Phi(u), \Psi(v)\rangle, \quad \forall u, v \in S(\mathbb{R}^n),$$

where $K = \frac{\pi}{2\log(1+\sqrt{2})}$.

Proof We introduce some tensor notations.

For a vector $u \in \mathbb{R}^n$, the order-3 rank-one tensor $T := u^{\otimes 3} \in \mathbb{R}^{n\times n\times n}$ is defined via $T_{ijl} = u_i u_j u_l$. The inner product of two tensors is defined as $\langle T, S\rangle = \sum_{i,j,l} T_{ijl} S_{ijl}$. These definitions can be generalized to higher order tensors of the form $u^{\otimes k}$ for $k = 2, 3, \ldots$ Note that $u^{\otimes 2} = uu^\top$.

Taking Taylor's expalnsion of $\sin(\cdot)$ at 0, we have

$$\begin{aligned}
\sin(\langle u, v\rangle) &= \langle u, v\rangle - \frac{1}{3!}\langle u, v\rangle^3 + \frac{1}{5!}\langle u, v\rangle^5 - \cdots \\
&= \langle u, v\rangle - \frac{1}{3!}\langle u^{\otimes 3}, v^{\otimes 3}\rangle + \frac{1}{5!}\langle u^{\otimes 5}, v^{\otimes 5}\rangle - \cdots \\
&= \left\langle \begin{bmatrix} u \\ -\sqrt{\frac{1}{3!}}u^{\otimes 3} \\ \sqrt{\frac{1}{5!}}u^{\otimes 5} \\ \vdots \end{bmatrix}, \begin{bmatrix} v \\ \sqrt{\frac{1}{3!}}v^{\otimes 3} \\ \sqrt{\frac{1}{5!}}v^{\otimes 5} \\ \vdots \end{bmatrix} \right\rangle.
\end{aligned}$$

Set $\Phi(u)$ to be the unit vector proportional to $\begin{bmatrix} u \\ -\sqrt{\frac{1}{3!}}u^{\otimes 3} \\ \sqrt{\frac{1}{5!}}u^{\otimes 5} \\ \vdots \end{bmatrix}$ and $\Psi(v)$ to be the unit vector proportional to $\begin{bmatrix} v \\ \sqrt{\frac{1}{3!}}v^{\otimes 3} \\ \sqrt{\frac{1}{5!}}v^{\otimes 5} \\ \vdots \end{bmatrix}$. One can verify that the normalization constant is $\frac{\pi}{2K}$ (this can be checked by looking at Taylor's expansion of $\sinh(\cdot)$). This completes the proof of Lemma 2. $\square$

Proof of Grothendieck's inequality: For any $u_i, v_j \in S(\mathbb{R}^n)$:

$$\begin{aligned}
&\sum_{i,j} A_{ij}\left\langle u_i, v_j\right\rangle \\
=&K \sum_{i,j} A_{ij} \frac{2}{\pi} \arcsin\left\langle \Phi\left(u_i\right), \Psi\left(v_j\right)\right\rangle && \text{(Lemma 2)} \\
=&K \sum_{i,j} A_{ij} \mathbb{E}_{w\sim\text{Gaussian}}\left[\operatorname{sign}\left\langle w, \Phi\left(u_i\right)\right\rangle \cdot \operatorname{sign}\left\langle w, \Psi\left(v_j\right)\right\rangle\right] && \text{(Recall angular kernel from lecture 5)} \\
\leqslant&K \cdot \max_{x_i, y_j \in\{\pm 1\}}\left|\sum_{i,j} A_{ij} x_i y_j\right|.
\end{aligned}$$

The proof is completed by taking the maximum of both sides over $u_i, v_j \in S(\mathbb{R}^n)$. $\square$

4 Generalization of Grothendieck's inequality

The above proof of the Grothendieck's inequality is called Krivine's method. A related proof uses the called Rietz's method, which can be used to established a high-rank version of the Grothendieck's inequality.

For $A \in \mathbb{R}^{n\times n}$ and $k \in \{1, 2, \cdots, n\}$, define

$$\begin{aligned} \mathrm{SDP}(A) &:= \max\{\langle A, Y\rangle : Y \succeq 0, Y_{ii} = 1\ \forall i\} \\ \mathrm{OPT}_k(A) &:= \max\{\langle A, Y\rangle : Y \succeq 0, Y_{ii} = 1\ \forall i, \mathrm{rank}(Y) \leq k\} \end{aligned}$$

One can check that $\mathrm{OPT}_1(A)$ is the RHS of the Grothendieck's inequality in Theorem 3. We have the following generalization:

Theorem 4 ([Montanari and Sen, 2016], Theorem 4 and Remark 6)**.** *For sufficiently large k, it holds that*

$$\left(1 - \frac{1}{k}\right) \mathrm{SDP}(A) - \frac{1}{k}|\,\mathrm{SDP}(-A)| \leq \mathrm{OPT}_k(A) \leq \mathrm{SDP}(A).$$

In particular, if $\mathrm{SDP}(A) \asymp \mathrm{SDP}(-A)$*, then*

$$\mathrm{SDP}(A) \leq \left(1 + O\Big(\frac{1}{k}\Big)\right) \mathrm{OPT}_k(A).$$

References

[Guédon and Vershynin, 2016] Guédon, O. and Vershynin, R. (2016). Community detection in sparse networks via grothendieck's inequality. *Probability Theory and Related Fields*, 165(3-4):1025–1049.

[Khot and Naor, 2012] Khot, S. and Naor, A. (2012). Grothendieck-type inequalities in combinatorial optimization. *Communications on Pure and Applied Mathematics*, 65(7):992–1035.

[Montanari and Sen, 2016] Montanari, A. and Sen, S. (2016). Semidefinite programs on sparse random graphs. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, volume abs/1504.05910, pages 814–827.

[Vershynin, 2018] Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.