S&DS 684: Statistical Inference on Graphs	Fall 2018
Lecture 12: Grothendieck Inequality and its Application in SBM	
Lecturer: Yihong Wu	*Scribe: Jiyi Liu, Nov 28, 2018*

Previously in Lecture **??**, we discussed the exact recovery of SBM. In this lecture, we turn to the almost exact recovery. Let σ be the community labels of nodes. Recall that the loss to evaluate a community estimate $\hat{\sigma}$ is $l(\sigma, \hat{\sigma}) = \frac{1}{n}\min_{\pm} d_H(\sigma, \hat{\sigma})$, where $d_H(x, y)$ is the Hamming distance $\sum_i I\{x_i \neq y_i\}$. We call $\hat{\sigma}$ an almost exact recovery if $l(\sigma, \hat{\sigma}) = o(1)$, and an Exact recovery if $l(\sigma, \hat{\sigma}) = 0$ w.h.p. We have seen that the requirement is $H^2(P, Q) \geq \frac{(2+)\log n}{n}$ for $\forall > 0$ for exact recovery. Here we are going to show the necessary and sufficient condition for almost exact recovery is $H^2(P, Q) \gg \frac{1}{n}$, which can be achieved by SDP relaxation.

We first introduce the key technical tool: Grothendieck Inequality (Theorem **??**). Then we discuss its application to SBM following Guédon-Vershynin [**?**].

12.1 $\|\cdot\|_{\infty\to 1}$ norm

Consider $A \in \mathbb{R}^{n\times m}$. We look at the following optimization

$$\max_{x_i, y_j = \pm} \sum_{1\leq i\leq n, 1\leq j\leq m} a_{ij} x_i y_j = \max_{x\in\{\pm\}^n, y\in\{\pm\}^m} \langle A, xy^\top\rangle. \tag{12.1}$$

Remark 12.1. The objective above (12.1) is a norm of A, denoted as $\|A\|_{\infty\to 1} = \max_{\|x\|_\infty\leq 1} \|Ax\|_1$. This is easily seen by writing $\|\cdot\|_1$ in the dual form.

Remark 12.2. $\|A\|_{\infty\to 1}$ is closely related to the cut norm. The cut norm $\|A\|_{\text{cut}}$ is defined as (cf. the min cut in Lecture **??**)

$$\|A\|_{\text{cut}} = \max_{I\subset[n], J\subset[m]} \left|\sum_{i\in I, j\in J} a_{ij}\right|.$$

The relation of the two norms is

$$\|A\|_{\text{cut}} \leq \|A\|_{\infty\to 1} \leq 4\|A\|_{\text{cut}}.$$

The left side inequality can be seen by

$$\|A\|_{\text{cut}} = \max_{I\subset[n], J\subset[m]} \left|\sum_{i\in I, j\in J} a_{ij}\right| \leq \max_{I\subset[n], J\subset[m]} \sum_{i\in I}\left|\sum_{j\in J} a_{ij}\right| \leq \max_J \|Ax_J\|_1 \leq \|A\|_{\infty\to 1}.$$

x_J is the indicator vector of J. The right side inequality can be shown by writing $x = I\{I\} - I\{I^c\}, y = I\{J\} - I\{J^c\}$ in (12.1).

We have the SDP relaxation of $\|A\|_{\infty\to 1}$: for $r \geq n + m$ (otherwise it is nonconvex),

$$\text{SDP}(A) = \max_{u_i, v_j\in\mathbb{R}^r, \|u_i\|=\|v_j\|=1} \sum_{i=1}^n \sum_{j=1}^m a_{ij}\langle u_i, v_j\rangle. \tag{12.2}$$

Remark 12.3. When $r = 1$, SDP(A) corresponds to $\|A\|_{\infty\to 1}$. Thus it is indeed a "relaxation" of the norm: $\|A\|_{\infty\to 1} \leq \text{SDP}(A)$.

Remark 12.4 (Dimension-free)**.** SDP(A) is dimension-free in the sense that the value does not depend on r as long as $r \geq n + m$. In particular, if it helps construction, we are free to consider the infinite-dimensional setting, e.g., the decision variables u_i, v_j take values in the Hilbert space of random variables – and we will do so next.

Remark 12.5 (Standard Form)**.** SDP(A) can be written into a standard SDP form

$$\text{SDP}(A) = \max_{X \succeq 0, X_{ii}=1} \langle W, X \rangle$$

where $W = \begin{pmatrix} 0 & A \\ A^\top & 0 \end{pmatrix}$. The correspondence is by writing

$$X = \begin{pmatrix} U^\top \\ V^\top \end{pmatrix} \begin{pmatrix} U & V \end{pmatrix} = \begin{pmatrix} U^T U & U^T V \\ V^T U & V^T V \end{pmatrix}.$$

We can see the role of r in SDP(A) is rank$(X) \leq r$. But X is $n + m$ by $n + m$, so as long as $r \geq n + m$, this constrain disappears.

12.2 Grothendieck Inequality

Theorem 12.1 (Grothendieck Inequality)**.**

$$\|A\|_{\infty\to 1} \leq \text{SDP}(A) \leq k\|A\|_{\infty\to 1}.$$

Here the absolute constant k can be chosen as $k = \frac{1}{\frac{4}{\pi}-1} \approx 3.66$ (with the world record ≈ 1.78).

*Proof: following Rietz [**?**].* The left side is obvious and stated in Remark 12.3. We focus on the right side. The main idea is randomized rounding. Let $u_i, v_j \in \mathbb{S}^{d-1}$ achieve the maximum in SDP(A) (**??**), $d = n + m$. We hope such u_i, v_j can match (not too far away from in objective) the x_i, y_j in (12.1). If we take some random x_i, y_j, then we can have the lower bound

$$\|A\|_{\infty\to 1} \geq \sum a_{ij} x_i y_j = \sum a_{ij} E(x_i y_j).$$

But a_{ij} can be positive or negative, so we cannot go further directly.

Consider $x_i = \text{sign}(\langle g, u_i \rangle), y_j = \text{sgn}(\langle g, v_j \rangle)$, $g \sim N(0, I_d)$.

Fact 12.1. *Note $\frac{g}{\|g\|} \sim$ unif $(\mathbb{S}^{d-1})$, so*

$$x_i y_j = \frac{2}{\pi} \arcsin\langle u_i, v_j \rangle.$$

Denote by $g_{u_i} = \langle g, u_i \rangle, g_{v_j} = \langle g, v_j \rangle$. We consider the generic settign $x_i = f(g_{u_i}), y_j = f(g_{v_j})$ for some $f : \mathbb{R} \to [-1, 1]$. We have the following facts

Fact 12.2. *1.* $g_u g_v = \langle u, v \rangle$.

2. $g_u f(g_v) = \langle u, v \rangle Z f(Z) \triangleq \langle u, v \rangle K$, $Z \sim N(0,1)$.

3. $(g_u - f(g_u))^2 = 1 - 2K + L$, $L \triangleq f^2(Z)$.

The facts hold noticing each $g_u \sim N(0,1)$. Then

$$
\begin{aligned}
\|A\|_{\infty\to 1} \geq & \sum a_{ij} x_i y_j \\
= & \sum a_{ij} f(g_{u_i}) f(g_{v_j}) \\
= & \sum a_{ij} (g_{u_i} - f(g_{u_i}))(g_{v_j} - f(g_{v_j})) - \sum a_{ij} g_{u_i} g_{v_j} \\
& + \sum a_{ij} \left(g_{v_j} f(g_{u_i}) + g_{u_i} f(g_{v_j}) \right) \\
(\text{def of } u_i, v_j) = & \underbrace{\sum a_{ij} (g_{u_i} - f(g_{u_i}))(g_{v_j} - f(g_{v_j}))}_{\star} + (2K - 1)\mathrm{SDP}(A).
\end{aligned}
$$

The magical next step is observing $(\star)$ is a feasible representation (after normalizing) in (**??**). Thus

$$
|(\star)| \leq (1 - 2K + L)\mathrm{SDP}(A).
$$
$$
\Rightarrow \|A\|_{\infty\to 1} \geq (4K - L - 2)\mathrm{SDP}(A).
$$

Let $f = \mathrm{sgn}$. Then $L = 1, K = |Z| = \sqrt{\frac{2}{\pi}}$. $4K - L - 2 = 4\sqrt{\frac{2}{\pi}} - 3 \approx 0.19 < 0.2$. So we proved that we can choose $k = 5$ in the theorem.

Moreover, there is a natural way to improve the constant. If we replace f by αf in the derivation above, then

$$
\begin{aligned}
\alpha^2 \|A\|_{\infty\to 1} \geq & \sum a_{ij} \alpha f(g_{u_i}) \cdot \alpha f(g_{v_j}) \\
= & \sum a_{ij} (g_{u_i} - \alpha f(g_{u_i}))(g_{v_j} - \alpha f(g_{v_j})) - \sum a_{ij} g_{u_i} g_{v_j} \\
& + \sum a_{ij} \left(\alpha g_{v_j} f(g_{u_i}) + \alpha g_{u_i} f(g_{v_j}) \right) \\
= & (\star)(\alpha) + (2\alpha K - 1)\mathrm{SDP}(A).
\end{aligned}
$$

And

$$
|(\star)(\alpha)| \leq (1 - 2\alpha K + \alpha^2 L)\mathrm{SDP}(A).
$$

Then the statement would be

$$
\|A\|_{\infty\to 1} \geq \left(\frac{4K}{\alpha} - \frac{2}{\alpha^2} - L \right) \mathrm{SDP}(A).
$$

The optimal α is $\frac{1}{K}$, and the bound is

$$
\|A\|_{\infty\to 1} \geq \left(2K^2 - L \right) \mathrm{SDP}(A)
$$

In the case $f = \mathrm{sgn}$, $2K^2 - L = \frac{4}{\pi} - 3$, as suggested in the theorem.

Clearly, the best strategy is to let

$$
f = \arg \max_{|f| \leq 1, Zf(Z) > 0} 2K^2 - L.
$$

The solution is given by the bounded linear function, $f(x) = \max(0, \min(x, 1))$. □

Remark 12.6 (PSD A)**.** When A is psd, the optimal bound is

$$\|A\|_{\infty\to 1} \geq \frac{2}{\pi}\mathrm{SDP}(A).$$

In this case, we can lower bound ($\star$) by 0 instead, since by design $u_i = v_i$ in this case. Then

$$\frac{\|A\|_{\infty\to 1}}{\mathrm{SDP}(A)} \geq \sup_{\alpha} \frac{2\alpha K - 1}{\alpha^2} = K^2 = \frac{2}{\pi}.$$

To show the sharpness, we construct examples that (asymptotically) achieve the bound. Let $l_i \overset{\text{i.i.d}}{\sim} \mathrm{unif}(\mathbb{S}^{d-1})$, $A_{ij} = \frac{1}{n^2}\langle l_i, l_j\rangle$. Choose $u_i = v_i = l_i$ in (**??**), we have

$$\mathrm{SDP}(A) \geq \frac{1}{n^2}\sum_{i,j}\langle l_i, l_j\rangle^2 \overset{\text{LLN}}{\approx} \langle l, l'\rangle^2 = \frac{1}{d}(1 + o(1)).$$

But

$$\|A\|_{\infty\to 1} \overset{\text{w.h.p}}{\leq} \frac{1}{d}(\frac{2}{\pi} + o(1)).$$

This is because

$$\|A\|_{\infty\to 1} = \max_{x\in\{\pm\}^n} \langle A, xx^\top\rangle = \max_{x\in\{\pm\}^n} \frac{1}{n^2}\sum x_i x_j \langle l_i, l_j\rangle = \max_{x\in\{\pm\}^n} \|\frac{1}{n}\sum x_i l_i\|_{l_2}^2 = \|l\|_{2\to 1}.$$

Remark 12.7 (Connection to max-cut)**.** Given a weighted graph with weight matrix W, similar to min-cut in (**??**), define:

$$\mathrm{maxcut}(W) \triangleq \max_{I\subset[n]} \sum_{i\in I, j\in I^c} W_{ij}.$$

Then

$$\begin{aligned}
2\mathrm{maxcut}(W) &= \max_{\sigma\in\{\pm\}^n} \sum W_{ij}(1 - \sigma_i\sigma_j) \\
&= \max_{\sigma\in\{\pm\}^n} \langle W, J - \sigma\sigma^\top\rangle \\
&\leq \max_{X\succeq 0, X_{ii}=1} \langle W, J - X\rangle \triangleq GW(W).
\end{aligned}$$

The same as the proof of Theorem **??**, note here $W_{ij} \geq 0$,

$$\begin{aligned}
2\mathrm{maxcut}(W) &\geq \sum W_{ij}(1 - \frac{2}{\pi}\arccos\langle u_i, v_j\rangle) \\
&\geq 0.878\sum W_{ij}(1 - \langle u_i, v_j\rangle) \\
&= 0.878 GW(W).
\end{aligned}$$

12.3 Application to SBM

Consider $SBM(n, p, q)$, $p = \frac{a}{n}, q = \frac{b}{n}$, and bisection $\langle\sigma, \mathbf{1}\rangle = 0$. Define $d = \frac{a+b}{2}, s = a - b$. Recall in the bisection case (see Remark 10.2), the MLE has the following SDP relaxation

$$\begin{aligned}
\hat{X} = \arg\max\ &\langle A, X\rangle. \\
&X \succeq 0 \\
&X_{ii} = 1 \\
&\langle X, J\rangle = 0
\end{aligned}$$

We claim that the necessary and sufficient condition is

$$\frac{(a-b)^2}{a+b} \to \infty.$$

Here $\frac{(a-b)^2}{a+b}$ can be interpreted as the signal-to-noise ratio (snr). In the more general P/Q model, the condition is $H^2(P,Q) \gg \frac{1}{n}$, which recovers the above when $P = \mathrm{Bern}(p)$ and $Q = \mathrm{Bern}(q)$.

Theorem 12.2 ([**?**]). *Let $\hat{v} =$ the top eigenvector of $\hat{X}$, and $\hat{\sigma} = sgn(\hat{v})$. Then*

$$l(\hat{\sigma}, \sigma) \overset{(also\ w.h.p)}{\lesssim} \frac{1}{\sqrt{snr}}.$$

Note: The above misclassification rate is later sharpened to exponential (optimal) by [**?**]:

$$l(\hat{\sigma}, \sigma) \leq \exp(-\Omega(snr)).$$

Proof. Define the population solution

$$\begin{aligned} X^* = \arg\max \ & \langle A, X \rangle. \\ & X \succeq 0 \\ & X_{ii} = 1 \\ & \langle X, J \rangle = 0 \end{aligned}$$

We can calculate

$$A = \frac{p+q}{2} J + \frac{p-q}{2} \sigma\sigma^\top - pI$$

and justify $\sigma\sigma^\top = X^*$.

$$\begin{aligned} \langle A, \hat{X} \rangle &= \langle A, \hat{X} \rangle - \langle A - A, \hat{X} \rangle \\ &\geq \langle A, X^* \rangle - \langle A - A, \hat{X} \rangle \\ &= \langle A, X^* \rangle + \underbrace{\langle A - A, X^* \rangle - \langle A - A, \hat{X} \rangle}_{\triangleq -\delta}. \end{aligned}$$

If we can somehow say $\delta \leq 0$, in other words $\langle A - A, X^* \rangle - \langle A - A, \hat{X} \rangle \geq 0$, then we can conclude $\langle A, \hat{X} \rangle \geq \langle A, X^* \rangle$, and thus $\hat{X} = X^*$. Though this is not possible in general, we can show δ is not too big to get the conclusion. Let $\hat{v} = v_1(\hat{X}), v = v_1(X^*) = \frac{\sigma}{\sqrt{n}}$, then by $\sin\Theta$ law,

$$\min \|\hat{v} \pm v\|_2 \lesssim \frac{\|\hat{X} - X^*\|_{op}}{\lambda_1(X^*) - \lambda_2(\hat{X})} \leq \frac{\|\hat{X} - X^*\|_F}{n - 0} = \frac{\|\hat{X} - X^*\|_F}{n}.$$

Also note that for every $\sigma_i \neq \hat{\sigma}_i$, $\|\hat{v} \pm v\|_2^2$ differs at least $\frac{1}{n}$ at this i. Thus

$$l(\hat{\sigma}, \sigma) \leq \frac{1}{n} \cdot n \min \|\hat{v} \pm v\|_2^2 \lesssim \frac{\|\hat{X} - X^*\|_F^2}{n^2}.$$

Suppose $n\sqrt{d} \overset{\text{w.h.p}}{\gtrsim} \delta \geq \langle A, X^* \rangle - \langle A, \hat{X} \rangle = \frac{p-q}{2}(n^2 - \langle \sigma\sigma^\top, \hat{X} \rangle)$. Then

$$\begin{aligned} \|\hat{X} - X^*\|_F^2 &= \|\hat{X}\|_F^2 + \|X^*\|_F^2 - 2\langle \hat{X}, X^* \rangle \\ &= \|\hat{X}\|_F^2 + n^2 - 2\langle \hat{X}, \sigma\sigma^\top \rangle \\ &\leq \mathrm{Tr}(\hat{X})^2 + n^2 - 2\langle \hat{X}, \sigma\sigma^\top \rangle \\ &= 2(n^2 - \langle \sigma\sigma^\top, \hat{X} \rangle) \lesssim \frac{\delta}{p-q} = \frac{n\delta}{a-b} \leq \frac{n^2}{\sqrt{snr}}. \end{aligned}$$

This completes the proof. So it remains only to show $\delta \overset{\text{w.h.p}}{\lesssim} n\sqrt{d}$. Denote $W = A - A$. We want to show

$$\frac{1}{2}|\delta| \leq \mathrm{SDP}(W) = \max_{X \succeq 0, X_{ii}=1} \langle W, X \rangle \overset{\text{w.h.p}}{\lesssim} n\sqrt{d}.$$

By Grothendieck Inequality,

$$\begin{aligned} \mathrm{SDP}(W) &= \max_{\|u_i\|=1} \sum_{i,j} W_{ij} \langle u_i, u_j \rangle \\ &\leq \max_{\|u_i\|=\|v_j\|=1} \sum_{i,j} W_{ij} \langle u_i, v_j \rangle \\ &\overset{G.I.}{\lesssim} \|W\|_{\infty \to 1} \\ &= \max_{x,y \in \{\pm\}^n} \langle W, xy^\top \rangle. \end{aligned}$$

By Hoeffding's inequality Lemma **??**,

$$\mathbb{P}(|\langle W, xy^\top \rangle| \geq t) \leq \exp(-c \cdot \frac{t^2}{n^2}).$$

To apply union bound on x, y, which in total 4^n, we need to choose $t \sim n^{3/2}$. We apply Bernstein's inequality instead,

$$\mathbb{P}(|\langle W, xy^\top \rangle| \geq t) \leq \exp(-c \cdot \frac{t^2}{t + nd}),$$

then we can choose $t \sim n\sqrt{d}$. □

References