Fall 2018

Homework 2

S&DS 684: Statistical Inference on Graphs

Due: December 19, 2018

`Prof. Yihong Wu`

Rules:

- It is mandatory to type your solutions in LaTeX. Email your solution in pdf by midnight of the due date to yihong.wu@yale.edu with subject line `Homework XX: your name`.
- Justify your work rigorously. As long as you are able to prove the result or a stronger version, there is no need to follow the hints.

1. (Spiked Wigner model) Consider the following rank-one perturbation to a Gaussian random matrix:

$$W = \sqrt{\frac{\mu}{n}}\sigma\sigma^\top + Z$$

where $Z = (Z_{ij})$ is a symmetric matrix with $\{Z_{ij} : 1 \le i \le j \le n\}$ being iid $N(0, 1)$, and the membership vector σ is uniformly drawn from the set of all bisections, i.e., $\{\sigma \in \{\pm\}^n : \sum_i \sigma_i = 0\}$.

(a) (Detection) Consider the hypothesis testing problem of testing $H_0 : W = Z$ (i.e. $\mu = 0$) versus $H_1 : W = \sqrt{\frac{\mu}{n}}\sigma\sigma^\top + Z$. Assume that μ is a constant. Show that reliable detection (i.e. both Type-I and Type-II error probabilities vanish as $n \to \infty$) is impossible if $\mu < 1$.[1] (Hint: compute the χ^2-divergence using the second moment method).

(b) (Correlated recovery) We say an estimator $\hat{\sigma} = \hat{\sigma}(W)$ achieves correlated recovery, if it has an nontrivial overlap with the true partition, i.e., $\mathbb{E}|\langle\sigma, \hat{\sigma}\rangle| = \Omega(n)$ as $n \to \infty$. Instead of using conditional second-moment argument as we did in class, we show that correlated recovery is impossible if $\mu < 1$ by a reduction argument:[2]

Suppose correlated recovery is possible. Let's construct a test statistic. Write $\sigma = \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix}$, where $\sigma_1 \in \{\pm\}^{(1-\epsilon)n}$ and $\sigma_2 \in \{\pm\}^{\epsilon n}$ with appropriately chosen ϵ. Write W accordingly in a block form $W = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}$. Apply correlated recovery estimator on W_{11} to obtain $\hat{\sigma}_1$, and compute $y = W_{21}\hat{\sigma}_1$. Under the null, we expect the variance of each coordinate of y is roughly 1; under the alternative, thanks to the correlation between σ_1 and $\hat{\sigma}_1$, we expect the variance of each coordinate is strictly bigger than 1. Make this argument rigorous by analyzing the test statistic $\frac{1}{n}\|y\|_2^2$.

(c) (Almost exact recovery) We say an estimator $\hat{\sigma} = \hat{\sigma}(W)$ achieves almost exact recovery, if the fraction of misclassification is vanishing, i.e., $\mathbb{E}|\langle\sigma, \hat{\sigma}\rangle| = n - o(n)$ as $n \to \infty$. Show that almost exact recovery is possible if and only if $\mu \to \infty$.

(Hint: for achievability, consider spectral method and perturbation bound).

(d) (Exact recovery: impossibility) We say an estimator $\hat{\sigma} = \hat{\sigma}(W)$ achieves exact recovery, if $\mathbb{P}[\sigma = \pm\hat{\sigma}] \to 1$ as $n \to \infty$. Show that exact recovery is impossible if $\mu = \sqrt{(2-\epsilon)\log n}$ for any fixed $\epsilon > 0$.

(Hint: show that even the maximum likelihood estimator fails in this case).

[1] In fact, $\mu = 1$ is also impossible, but we have to resort more advanced techniques than second moment method.

[2] This idea was suggested by Prof. Zhou Fan.

(e) (Exact recovery: SDP) Consider the following SDP relaxation:

$$\hat{X} = \arg\max\{\langle W, X\rangle : X \succeq 0, X_{ii} = 1, \langle X, \mathbf{J}\rangle = 0\}$$

where $\mathbf{J}$ is the all-one matrix. Show that exact recovery is achieved, i.e., $\hat{X} = \sigma\sigma^\top$ with probability tending to one, if $\mu = \sqrt{(2+\epsilon)\log n}$ for any fixed $\epsilon > 0$.

(Hint: do not invoke the general result from the lecture; instead, do a direct analysis based on two facts (i) $\|Z\|_{op} = O(\sqrt{n})$ with high probability; (ii) the maximum of n iid standard normals is $\sqrt{(2+o(1))\log n}$ with high probability).

2. ($\|\cdot\|_{2\to 1}$-norm) Denote the rows of $B \in \mathbb{R}^{n\times d}$ by $b_1^\top, \dots, b_n^\top$.

(a) Show that the induced norm $\|B\|_{2\to 1}$ is given by

$$\|B\|_{2\to 1} = \max\left\{\sum_{i=1}^n |\langle b_i, y\rangle| : y \in S^{d-1}\right\}.$$

(b) Suppose $b_1, \dots, b_n$ are iid uniformly drawn from the sphere S^{d-1}. Show that for any fixed d, as $n \to \infty$, $\frac{1}{n}\|B\|_{2\to 1}$ converges in probability to some value c_d as a function of d. Find c_d as explicitly as you can.

(Hint: take an ϵ-net over S^{d-1} and use union bound. For fixed $y \in S^{d-1}$, what is $\mathbb{E}[|\langle b_1, y\rangle|]$?)

(c) Show that $\sqrt{d}c_d \to \sqrt{\frac{2}{\pi}}$ as $d \to \infty$.

3. (Grothendieck inequality for PSD matrices) For $A = (a_{ij}) \in \mathbb{R}^{n\times n}$, consider

$$\|A\|_{\infty\to 1} \triangleq \max\left\{\sum_{i,j\in[n]} a_{ij}x_i y_j : x_i, y_j \in \{\pm\}\right\} = \max\left\{\langle A, xy^\top\rangle : \|x\|_\infty \le 1, \|y\|_\infty \le 1\right\} \quad (1)$$

and its SDP relaxation

$$\mathsf{SDP}(A) \triangleq \max\left\{\sum_{i,j\in[n]} a_{ij}\langle u_i, v_j\rangle : u_i, v_j \in S^{n-1}\right\} = \max\{\langle A, X\rangle : X \succeq 0, X_{ii} = 1\}. \quad (2)$$

(a) Following the argument in class, show that for every positive semidefinite A,

$$\mathsf{SDP}(A) \ge \|A\|_{\infty\to 1} \ge \frac{2}{\pi}\mathsf{SDP}(A). \quad (3)$$

(b) Next we show that the constant $\frac{2}{\pi}$ in (3) is sharp by constructing instances of A so that the ratio $\frac{\|A\|_{\infty\to 1}}{\mathsf{SDP}(A)}$ is arbitrarily close to $\frac{2}{\pi}$.

(i) Show that without loss of optimality, we can restrict to $x_i = y_i$ in (1);
(Hint: $\langle A, xy^\top\rangle^2 = \langle\sqrt{A}x, \sqrt{A}y\rangle^2 \le \langle A, xx^\top\rangle\langle A, yy^\top\rangle$. Why?)

(ii) Show that without loss of optimality, we can restrict to $u_i = v_i$ in (2);
(Hint: $\langle A, U^\top V\rangle^2 = \langle\sqrt{A}U^\top, \sqrt{A}V^\top\rangle^2 \le \langle A, U^\top U\rangle\langle A, V^\top V\rangle$. Why?)

(iii) Show the following deterministic fact: if $A = \frac{1}{n^2} BB^\top$ for $B \in \mathbb{R}^{n\times d}$, then

$$\|A\|_{\infty\to 1} = \frac{1}{n^2}\|B\|_{2\to 1}^2$$

and

$$\mathsf{SDP}(A) \geq \frac{1}{n^2} \sum_{i,j\in[n]} \langle b_i, b_j\rangle^2$$

where $b_i^\top$ is the ith row of B.

(iv) Now take the rows of B to be iid uniform on S^{d-1}, use the previous problem to show that $\|A\|_{\infty\to 1} \to c_d^2$ as $n \to \infty$.

(v) Show that $\mathsf{SDP}(A) \geq r_d$ in probability as $n \to \infty$, where r_d behaves as $\frac{1+o(1)}{d}$ as $d \to \infty$.

(vi) Conclude the sharpness of the constant $\frac{2}{\pi}$ in (3).
(Hint: what is $\mathbb{E}[\langle b_1, b_2\rangle^2]$? Be careful with $\sum_{i,j\in[n]}\langle v_i, v_j\rangle^2$ which is not an iid sum.)