

# ST 790 Assignment 3

David Elsheimer and Jimmy Hickey

10/15/2020

## Instruction

This assignment consists of 3 problems. The assignment is due on **Friday, October 16** at 11:59pm EDT. Please submit your assignment electronically through the **Moodle** webpage. The assignment can be done as a group with at most 3 members per group (please include the name of the group members on the front page of the assignment).

## Problem 1

In class, we only proved matrix concentration inequalities for symmetric matrices. Specifically, Theorem 9 in the lecture notes states that

**Theorem** (Tropp) Let  $X_1, X_2, \dots, X_n$  be independent mean 0  $d \times d$  **symmetric** random matrices and suppose that  $\|X_i\| \leq M$  almost surely for all  $i$ . Define

$$\sigma^2 = \left\| \sum_{i=1}^n \mathbb{E}[X_i^2] \right\|$$

Then for every  $t > 0$

$$\mathbb{P}\left(\left\| \sum_{i=1}^n X_i \right\| > t\right) \leq 2d \exp\left(\frac{-t^2}{\sigma^2 + Mt/3}\right)$$

Extend this result to general matrices with real-valued entries. That is, show the following corollary

**Corollary** Let  $X_1, X_2, \dots, X_n$  be independent mean 0  $d_1 \times d_2$  random matrices and suppose that  $\|X_i\| \leq M$  almost surely for all  $i$ . Define

$$\sigma^2 = \max\left\{\left\| \sum_{i=1}^n \mathbb{E}[X_i X_i^\top] \right\|, \left\| \sum_{i=1}^n \mathbb{E}[X_i^\top X_i] \right\|\right\}.$$

Then for every  $t > 0$

$$\mathbb{P}\left(\left\| \sum_{i=1}^n X_i \right\| > t\right) \leq (d_1 + d_2) \exp\left(\frac{-t^2}{\sigma^2 + Mt/3}\right).$$

**Hint** Relate the singular values of a rectangular matrix  $M$  with the eigenvalues of its **symmetric dilation**

$$\begin{bmatrix} 0 & M \\ M^\top & 0 \end{bmatrix}.$$

**Answer:**

Note that the Hermitean dilation  $H = \begin{bmatrix} 0 & M \\ M^\top & 0 \end{bmatrix}$  is symmetric. Then note that  $H^2 = H \cdot H = \begin{bmatrix} MM^\top & 0 \\ 0 & M^\top M \end{bmatrix}$ . This is a symmetric, block diagonal matrix. The eigenvalues of  $H^2$  are the eigenvalues of  $MM^\top, M^\top M$ . In turn, these are the same as the square of the singular values of  $M, M^\top$ . As such, the spectral radius of  $M$  is the same as the largest eigenvalues of  $H$ .

Where  $H$  is as previously defined, we can then consider  $H = \sum_{i=1}^n H_i$ , such that  $H_i = \begin{bmatrix} 0 & X_i \\ X_i^\top & 0 \end{bmatrix}$ .

Using this, the proof follows similarly to the proof for Tropp.

Note first that  $\left\| \sum_{i=1}^n X_i \right\| = \left\| \sum_{i=1}^n H_i \right\| = \lambda_{\max}(Z_n)$ , where  $Z_n = \sum_{i=1}^n H_i$ .

Note further that  $\sigma^2 = \max \left\{ \left\| \sum_{i=1}^n \mathbb{E}[X_i X_i^\top] \right\|, \left\| \sum_{i=1}^n \mathbb{E}[X_i^\top X_i] \right\| \right\} = \left\| \sum_{i=1}^n H_i^2 \right\|$ .

Thus, the remainder of the proof is as follows

$$\begin{aligned}
P\left(\left\| \sum_{i=1}^n H_i \right\| > t\right) &\leq e^{-st} E(\exp(s\lambda_{\max}(Z_n))) \\
&\leq e^{-st} E(\text{tr}(\exp(\sum_{i=1}^n sH_i))) \\
&= e^{-st} E(E(\text{tr} \exp\left(\sum_{i=1}^{n-1} sH_i + sH_n\right) | \{H_i\}_{i=1}^{n-1})) \\
&\leq e^{-st} E(\text{tr} \exp\left(\sum_{i=1}^{n-1} sH_i\right) + \log E(e^{sH_n})) \\
&\leq \dots \\
&\leq e^{-st} \text{tr} \exp\left(\sum_{i=1}^1 \log E(e^{sH_i})\right), \text{ [Via corollary 13.]} \\
&\leq e^{-st} \text{tr} \exp\left(\sum_{i=1}^n \frac{s^2/2}{1-|s|M/3} E(H_i^2)\right), \text{ [Via Proposition 14.]} \\
&\leq (d_1 + d_2) e^{-st} \lambda_{\max}\left(\exp\left(\frac{s^2/2}{1-|s|M/3} \sum_{i=1}^n E(H_i^2)\right)\right), \text{ [Via Golden-Thompson.]} \\
&\leq (d_1 + d_2) e^{-st} \exp\left(\frac{s^2/2}{1-|s|M/3} \sum_{i=1}^n \lambda_{\max}(E(H_i^2))\right) \\
&= (d_1 + d_2) e^{-st} \exp\left(\frac{s^2/2}{1-|s|M/3} \left\| \sum_{i=1}^n H_i^2 \right\| \right) \\
&= (d_1 + d_2) e^{-st} \exp\left(\frac{\sigma^2 s^2/2}{1-|s|M/3}\right)
\end{aligned}$$

For  $s = \frac{t}{\sigma^2 + Mt/3}$ , this bound is  $(d_1 + d_2) \exp(\frac{-t^2}{\sigma^2 + Mt/3}) \exp(t^2/2)$  which is always greater than or equal to  $(d_1 + d_2) \exp(\frac{-t^2}{\sigma^2 + Mt/3})$ .

Thus for independent matrices with dimension  $d_1 \times d_2$ ,  $P(\left\|\sum_{i=1}^n X_i\right\| > t) \leq (d_1 + d_2) \exp(\frac{-t^2}{\sigma^2 + Mt/3})$ .

## Problem 2

To continue with the theme of the previous problem, in our class we only prove matrix concentration inequalities for the case when the summands are bounded almost surely in spectral norm. See for example the statement of the result in the previous problem. For this problem, you are asked to derive a matrix concentration inequality in the setting where the matrix powers of  $X_i$  satisfy a Bernstein-like tail conditions. In particular, prove the following result.

**Theorem** Let  $X_1, X_2, \dots, X_n$  be independent mean 0, symmetric random  $d \times d$  matrices. Furthermore, assume that there exists a constant  $R > 0$  and symmetric matrices  $A_1, A_2, \dots, A_n$  such that for any  $i$

$$\mathbb{E}[X_i^p] \preceq \frac{p!}{2} R^{p-2} A_i^2, \quad \text{for } p = 2, 3, 4, \dots$$

Here  $A \preceq B$  if  $B - A$  is a positive semidefinite matrix. Now define

$$\sigma^2 = \left\| \sum_{i=1}^n A_i^2 \right\|$$

Then for every  $t > 0$

$$\mathbb{P}\left(\left\| \sum_{i=1}^n X_i \right\| > t\right) \leq 2d \exp\left(\frac{-t^2/2}{\sigma^2 + Rt}\right)$$

Next apply the above result to derive a tail bound for the following (weighted) random graph model. Let  $P$  be a  $n \times n$  symmetric matrix with non-negative, finite entries  $p_{ij}$ . Let  $A$  be a  $n \times n$  symmetric matrix whose entries  $a_{ij}$  for  $i \leq j$  are independent and that  $a_{ij} \sim \text{Pois}(p_{ij})$ . Sketch a tail bound for  $\|A - P\|$  similar in flavor to Theorem 12 in the set of lecture notes for matrix concentration inequalities. (You don't have to pay too much attention to constant factors; we are only interested in important quantities similar to the  $\sqrt{\Delta \log n}$  in the statement of Theorem 12 there).

## Problem 3

Sketch a proof for the weak recovery of SBMs for the spectral clustering using the normalized Laplacian. In particular, consider the algorithm on page 1881 of [Rohe et al.](#). Note that you do not need to use the eigenvectors associated with the largest  $k$  eigenvalues (in modulus), but rather the eigenvectors associated with the  $d$  largest eigenvalues where  $d = \text{rk}(\mathbf{B})$  which is assumed known.

In particular, proceed as follows

- Relate the eigenvectors of  $\mathcal{L}(\mathbf{P}) = \mathbf{T}^{-1/2} \mathbf{P} \mathbf{T}^{-1/2}$  to the block structures. Here  $\mathbf{T}$  is a diagonal matrix with  $T_{ii} = \sum_j p_{ij}$  is the **expected degree** of vertex  $i$ .
- Next invoke matrix concentration inequality for  $\|\mathcal{L}(\mathbf{A}) - \mathcal{L}(\mathbf{P})\|$ .
- Then invoke the Davis-Kahan theorem to bound the difference between the top eigenvectors of  $\mathcal{L}(\mathbf{A})$  and the top eigenvectors of  $\mathcal{L}(\mathbf{P})$ .
- Finally, show that the above difference yield that the **proportion** of mis-clustered vertices (assuming clustering is done via  $K$ -means) will converges to 0.