36-710, Fall 2018
Homework 5

Due Mon Nov 19, by 5:00pm in JaeHyeok's mailbox.

1. **Matrix Algebra Problems.**
   (a) Exercise 8.3 (You may assume the result of Problem 8.1 as given).
   (b) Exercise 8.4.
   (c) Recall the spiked covariance model: $\Sigma = \theta v v^\top + I_d$, where $\theta > 0$ and $v \in \mathbb{S}^{d-1}$. Let $\hat{v}$ be another unit vector in $\mathbb{S}^{d-1}$. Show that
   $$v^\top \Sigma v - \hat{v}^\top \Sigma \hat{v} = \theta \sin^2(\angle(v, \hat{v}))$$
   where $\angle(v, \hat{v}) = \cos^{-1}(|v^\top \hat{v}|)$
   (d) Show that
   $$\left\| \hat{v}\hat{v}^\top - v v^\top \right\|_F^2 = 2 \sin^2(\angle(v, \hat{v})),$$
   where, for a matrix $A = (A_{i,j})$, $\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$ and v and $\hat{v}$ are unit vectors.

2. In this exercise you will fill in some of the details from the proof of the upper bound for sparse PCA under a spike covariance model.
   (a) For $p \geq 1$ the Schatten p-norm of a $n \times m$ matrix A is the ℓ_p norm of its singular values:
   $$\|A\|_p = \left(\sum_{i=1}^r \sigma_i^p \right)^{1/p},$$
   where $\sigma \geq \sigma_2 \geq \ldots \sigma_r \geq 0$ are the singular values of A and $r = \min\{m, n\}$. Prove the non-commutative Hölder inequality for conformal matrices A and B:
   $$|\mathrm{tr}(A^\top B)| \leq \|A\|_p \|B\|_q,$$
   for all $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$.
   (b) Let u and v be two unit norm vectors in $\mathbb{R}^d$. Show that
   $$\sqrt{2}\|uu^\top - vv^\top\|_{\mathrm{op}} = \|uu^\top - vv^\top\|_F.$$

3. Suppose we observe an i.i.d. sample $X_1, \ldots, X_n$ from the mixture distribution
   $$\frac{1}{2}P_1 + \frac{1}{2}P_2,$$
   where $P_1 = N_d(\mu, I_d)$ and $P_2 = N_d(-\mu, I_d)$, with $\mu \in \mathbb{R}^d$ a non-zero vector. Ours task is to cluster the sample points into two groups, where points in the same group originated from the same component of the mixture.

   We will use spectral clustering: compute the leading eigenvector $\hat{\nu}$ of the empirical covariance matrix and cluster the points depending on the signs of $X_i^\top \hat{\nu}$, $i = 1, \ldots, n$.

   Use the Davis-Kahan theorem to derive an upper bound on the proportion of misclustered nodes.

4. **Reading Exercise.** Read the paper "Sparse principal component analysis via random projections," by Milana Gataric, Tengyao Wang, Richard J. Samworth.