CSCI 6971/4971: Homework 2

Assigned Monday February 12 2017. Due at beginning of class Thursday February 22 2017. Remember to typeset this submission using LaTeX, and email it to me by the start of class on the due date.

1. [50 points (JLT Lemma for Gaussian Projections)] In class we established the Johnson-Lindenstrauss Lemma for general sub-gaussian random vectors. Specialize the proof to apply to vectors whose entries are independent normal random variables, and give explicit constants. You may use the following concentration inequality for chi-squared random variables with n degrees of freedom: if the variables X_i, $i = 1, 2, \ldots, n$ are independent standard normal random variables, then
$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i^2 - 1\right| \geq t\right\} \leq 2\mathrm{e}^{-nt^2/8}.$$

2. [50 points (Nonsymmetric Bernstein Inequality)] The matrix Bernstein inequality we learned in class allows us to bound the maximum eigenvalue of a sum of symmetric random matrices. Sometimes we need to control the norm of a sum of general random matrices. In this exercise, you will show that such a bound follows from the inequality you already know. Specifically, you will establish the following theorem.

> Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be random n-by-d matrices with zero mean. Choose a constant R that uniformly bounds the norms of the summands almost surely,
> $$\max_{i=1}^{n} \|\mathbf{X}_i\|_2 \leq R \quad \text{almost surely,}$$
> and a "matrix variance" σ^2 that satisfies
> $$\max\left\{\left\|\sum\nolimits_{i=1}^{n} \mathbb{E}\left(\mathbf{X}_i\mathbf{X}_i^T\right)\right\|_2, \left\|\sum\nolimits_{i=1}^{n} \mathbb{E}\left(\mathbf{X}_i^T\mathbf{X}_i\right)\right\|_2\right\} \leq \sigma^2.$$
> Then
> $$\mathbb{P}\left[\left\|\sum\nolimits_i \mathbf{X}_i\right\|_2 \leq t\right] \leq (n+d)\exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right).$$

- The key to proving this theorem lies in two linear algebraic results. The first of these allows replacing the sum of rectangular matrices $\mathbf{X}_i$ with a sum of closely related symmetric matrices. Prove that, for any matrix $\mathbf{A}$,
$$\|\mathbf{A}\|_2 = \lambda_{\max}\left(\begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix}\right).$$
The symmetric matrix on the right-hand side is called the self-adjoint dilation of $\mathbf{A}$. To establish this result, use the full SVD of $\mathbf{A}$ to demonstrate that the singular values of the self-adjoint dilation of $\mathbf{A}$ are the same as those of $\mathbf{A}$: show that if $\mathbf{\Sigma}$ contains the singular values of $\mathbf{A}$, then there is an orthonormal matrix $\mathbf{Z}$ which satisfies
$$\begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} = \mathbf{Z}\begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Sigma} \end{bmatrix}\mathbf{Z}^T$$
Hint: if $\sum_{i=1}^{\min\{n,d\}} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is the full SVD of $\mathbf{A}$, argue that $\left[\mathbf{u}_i^T \mathbf{v}_i^T\right]^T/\sqrt{2}$ and $\left[\mathbf{u}_i^T - \mathbf{v}_i^T\right]^T/\sqrt{2}$ are eigenvectors of the self-adjoint dilation of $\mathbf{A}$.
- Show that this fact implies that
$$\left\|\sum\nolimits_{i=1}^{n} \mathbf{X}_i\right\|_2 = \lambda_{\max}\left(\sum\nolimits_{i=1}^{n} \mathbf{D}_i\right)$$

where $\mathbf{D}_i$ is the self-adjoint dilation of $\mathbf{X}_i$. Thus we can use the matrix Bernstein inequality from class, once we have identified a uniform bound on $\|\mathbf{D}_i\|_2$ that holds with probability one, and an estimate of the matrix variance $\|\sum_{i=1}^n \mathbb{E}\mathbf{D}_i^2\|_2$.

- Argue that the R specified in the theorem statement is indeed a uniform bound on $\|\mathbf{D}_i\|_2$ with probability one. That is, show that

$$\max_{i=1}^n \|\mathbf{D}_i\|_2 \leq \max_{i=1}^n \|\mathbf{X}_i\|_2.$$

- The next key linear algebraic result allows us to find the matrix variance of the sum of the self-adjoint dilations. Prove that if $\mathbf{A}$ and $\mathbf{B}$ are matrices, then there are orthonormal matrices $\mathbf{L}$ and $\mathbf{R}$ such that

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} = \mathbf{L} \begin{bmatrix} \mathbf{\Sigma_A} & 0 \\ 0 & \mathbf{\Sigma_B} \end{bmatrix} \mathbf{R}^T,$$

where $\mathbf{\Sigma_A}$ contains the singular values of $\mathbf{A}$, and similarly $\mathbf{\Sigma_B}$ contains the singular values of $\mathbf{B}$. Argue that it follows that

$$\left\| \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \right\|_2 = \max\{\|\mathbf{A}\|_2, \|\mathbf{B}\|_2\}.$$

- Use the above result to argue that the σ^2 specified in the theorem statement is indeed a suitable matrix variance for the sum of self adjoint dilations. That is, argue that

$$\left\| \sum_{i=1}^n \mathbb{E}\mathbf{D}_i^2 \right\|_2 = \max\left\{ \left\| \sum_{i=1}^n \mathbb{E}\left(\mathbf{X}_i\mathbf{X}_i^T\right) \right\|_2, \left\| \sum_{i=1}^n \mathbb{E}\left(\mathbf{X}_i^T\mathbf{X}_i\right) \right\|_2 \right\}.$$

- Put these arguments together, and apply the matrix Bernstein theorem from class to demonstrate that you obtain the claimed bound.