CSCI 6971/4971: Homework 4

Assigned Monday April 15 2019. Due at beginning of class Monday April 22 2019.

Submission Instructions: Submit an appropriately typeset solution at the start of class. Do not email me your solution.

1. [100 points] (Analysis of matrix sparsification)

   Consider the problem of finding a sparse approximant $\mathbf{Z}$ to an arbitrary given matrix $\mathbf{A} \in \mathrm{R}^{m \times n}$ that has low spectral norm.

   We will show that if we take $\mathbf{Z} = \frac{1}{c}\sum_{k=1}^{c} \mathbf{X}_k$ where the $\mathbf{X}_k$ are i.i.d. random matrices with exactly one non-zero entry, selected from a probability distribution to be described, and $c = O(\max\{m,n\}\,\mathrm{srank}(A)\ln((m+n)/\delta)\varepsilon^{-2})$, then $\mathbf{Z}$ has at most c non-zero entries and satisfies

   $$\|\mathbf{A} - \mathbf{Z}\|_2 \leq \varepsilon\|\mathbf{A}\|_2$$

   with probability at least $1 - \delta$.

   Specifically, we will take the summands $\mathbf{X}_k$ to be distributed as

   $$\mathbf{X} = \frac{1}{p_{ij}} a_{ij}\mathbf{E}_{ij} \quad \text{with probability } p_{ij},$$

   where $\mathbf{E}_{ij}$ denotes the standard basis matrix that has a one in the (i,j) entry and zeros elsewhere, and

   $$p_{ij} = \frac{1}{2}\left[\frac{a_{ij}^2}{\|\mathbf{A}\|_F^2} + \frac{|a_{ij}|}{\|\mathbf{A}\|_1}\right],$$

   where $\|\mathbf{A}\|_1 = \sum_{i,j}|a_{ij}|$.

   The matrix Bernstein inequality presented in class is not directly applicable (because it requires the summands to be symmetric matrices). However, the matrix Bernstein inequality given below applies to sums of general random matrices, and can be used to establish the desired sparsification guarantees[1].

   > Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent 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
   >
   > $$\max\left\{\left\|\sum_{i=1}^{n}\mathbb{E}\left(\mathbf{X}_i\mathbf{X}_i^T\right)\right\|_2, \left\|\sum_{i=1}^{n}\mathbb{E}\left(\mathbf{X}_i^T\mathbf{X}_i\right)\right\|_2\right\} \leq \sigma^2.$$
   >
   > Then
   >
   > $$\mathbb{P}\left[\left\|\sum_i \mathbf{X}_i\right\|_2 \geq t\right] \leq (n+d)\exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right).$$

[1] This inequality follows from the inequality presented in class, using some linear algebraic manipulations.

As a corollary,

$$\mathbb{P}\left[\left\|\sum_i \mathbf{X}_i\right\|_2 \geq t\right] \leq \begin{cases} (n+d)\exp\left(\frac{-3t^2}{8\sigma^2}\right) & t \leq \frac{\sigma^2}{R} \\ (n+d)\exp\left(\frac{-3t}{8L}\right) & t \geq \frac{\sigma^2}{R} \end{cases}$$

To use this result, we must establish that the conditions of the theorem are satisfied, and identify appropriate quantities R and σ^2.

1. Verify that p_{ij} is a probability distribution over the set of indices into $\mathbf{A}$.
2. Identify summands $\mathbf{Y}_k$ such that $\mathbf{A} - \mathbf{Z} = \sum_{k=1}^{c} \mathbf{Y}_k$, and prove that $\mathbb{E}\mathbf{Y}_k = 0$.
3. Note that the choice of p_{ij} ensures that $p_{ij}^{-1} \leq 2\|\mathbf{A}\|_1/|a_{ij}|$. It is also the case that $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_1$ for any matrix $\mathbf{A}$. Use these facts and the triangle inequality to argue that
$$\|\mathbf{Y}_k\|_2 \leq R = \frac{3}{c}\|\mathbf{A}\|_1$$
always.
4. Argue that $\sigma^2 = c\max\{\|\mathbb{E}\mathbf{Y}\mathbf{Y}^T\|_2, \|\mathbb{E}\mathbf{Y}^T\mathbf{Y}\|_2\}$ is a valid choice.
5. Compute an exact expression for $\mathbb{E}[\mathbf{Y}\mathbf{Y}^T]$. Use this expression and the fact that $p_{ij}^{-1} \leq 2\|\mathbf{A}\|_F^2/|a_{ij}|^2$ to argue that
$$\mathbb{E}[\mathbf{Y}\mathbf{Y}^T] \preceq \frac{2n}{c^2}\|\mathbf{A}\|_F^2 \cdot \mathbf{I}_m.$$
6. Similarly, argue that $\mathbb{E}[\mathbf{Y}^T\mathbf{Y}] \preceq \frac{2m}{c^2}\|\mathbf{A}\|_F^2 \cdot \mathbf{I}_n$.
7. Argue that $\sigma^2 \leq \frac{2}{c}\max\{m, n\}\|\mathbf{A}\|_F^2$.
8. Apply the small deviations version of the matrix Bernstein bound from Problem 1 to conclude that the claimed c is sufficient to ensure the claimed approximation error with the claimed success probability.