The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381V: Large Scale Learning — Spring 2013

Assignment 3

Caramanis/Sanghavi — Due: April 8, 2013 (Monday, **10am**)

---

1. Consider a collection of (arbitrary) points: $\mathbf{x}_1, \dots, \mathbf{x}_n$ in p-dimensional Euclidean space, $\mathbb{R}^p$. Assume that these points are centered, i.e., $\sum_i \mathbf{x}_i = 0$. For V a k-dimensional subspace, let P_V denote the orthogonal projection onto V. Then the approximation error of V with respect to our points is given as:
$$\sum_{i=1}^{n} \|\mathbf{x}_i - P_V(\mathbf{x}_i)\|_2^2.$$
   (a) Write this optimization problem explicitly, and show that it is non-convex.
   (b) Show that the top k singular vectors of the empirical covariance matrix of the points $\{\mathbf{x}_i\}$ give the optimal solution.

2. Suppose a $m \times n$ matrix A has a low-rank approximate factorization:
$$\|A - CB\| \leq \varepsilon,$$
   where C is $m \times k$ and B is $k \times n$.
   (a) Use this factorization to come up with an approximate QR-factorization. How many flops does it take?
   (b) Use this factorization to come up with an approximate SVD-factorization. How many flops does it take?

3. The power method for finding the leading eigenvector of a square matrix A proceeds as follows:
$$\begin{aligned} \tilde{v}_{t+1} &= Av_t \\ v_{t+1} &= \tilde{v}_{t+1}/\|\tilde{v}_{t+1}\|. \end{aligned}$$
   Let v^* be the true leading eigenvector and v_0 the initial iterate.
   a) What happens when v_0 is orthogonal to v^* ?
   b) Give an upper bound on $||v_t - v^*||$ in terms of $||v_0 - v^*||$. Your bound should depend on t and the two leading singular values σ_1 and σ_2.

4. Show that for any $m \times n$ matrix M, $||M||_F^2 = \sum_{i=1}^{\min(m,n)} \sigma_i(M)^2$, where $\sigma_i(M)$ is the i^{th} singular value of M.

5. Suppose we have a $n \times d$ matrix A where each element A_{ij} is drawn i.i.d. to be $\frac{1}{\sqrt{n}}$ w.p. 1/2, and $-\frac{1}{\sqrt{n}}$ otherwise. Find a lower bound on n so that the matrix A satisfies (ϵ, s) RIP w.p. at least $1-\delta$. Your bound should depend on s, d, ϵ, δ.

   Follow the steps in the proof of Theorem 4 in the paper "Compressed Sensing: Basic results and self contained proofs" by Shai Shalev-Shwartz. (this paper is available on the class webpage). You can use any lemmas therein, provided you refer to them appropriately.

6. The adjacency matrix A of a graph $G=(V,E)$ is a $0-1$ matrix where $a_{ij}=1$ if and only if $(i,j) \in E$, and 0 otherwise. By convention $a_{ii}=0$. The Laplacian L of a graph is $L = D - A$ where D is a diagonal matrix with d_{ii} equal to the degree of node i, $i \in V$.

   a) Show that L is positive semi-definite.

   b) Show that the smallest eigenvalue $\lambda_n(L) = 0$, for any graph with n nodes.

   c) Show that if the graph has k components, then $\lambda_n(L) = \lambda_{n-1}(L) = \ldots = \lambda_{n-k+1}(L) = 0$. [i.e. L has k eigenvalues equal to 0]

7. Consider a $m \times n$ matrix A. Consider the sparsification technique where we sample $E \subseteq [m] \times [n]$ entries of the matrix A. In particular, let us consider the sampling-with-replacement model, where we form matrix Y such that

$$Y = \begin{cases} A_{ij}/p, & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases}$$

   We set $p = |E|/mn$ so that the expected number of entries sampled equals $|E|$.

   Let Q_k denote the top left singular vectors of Y. In this exercise you will show that there exists a universal constant c such that if $m \le n$, and $|E| \ge cn \log n$, then with probability at least $1 - n^{-3}$,

$$\|A - Q_k Q_k^\top Y\| \le \|A - A_k\| + cA_{\max}\sqrt{mn}\sqrt{\frac{n}{|E|}}.$$

   To simplify the expression, suppose $m = n$, and $A_{\max} = 1$.

   As discussed in class, there is a common approach to many problems like this. The tools we have seen in class control the deviations of sums of zero-mean independent random matrices. Therefore, we need to make the error look like a sum of zero-mean independent matrices. This exercise proceeds in a very similar way as our proof for the column-sub-selection case we did in class.

   (a) Show that for any two matrices, A and B (of compatible dimensions)

$$\|A - B_k\|_2 \le \|A - A_k\|_2 + 2\|A - B\|_2.$$

   (Recall that we are using M_k to denote the best k-rank approximation of a matrix). In our setting, this lemma says that to control $\|A - Y_k\|$, we have to control $\|A - Y\|$.

   (b) Let $X = Y - A$, and write it as a sum of mn independent matrices.

   (c) Now bound the key parameters R and σ^2 in Bernstein's Theorem.

   (d) Use this, to show that

$$\mathbb{P}(\|X\|_2 \ge t\sqrt{n/p}) \le 2n \cdot \exp\left(-\frac{t^2}{12}\right).$$

(e) Choose an appropriate t, and conclude the proof of the theorem.

8. Show that if a matrix W satisfies $(\epsilon, 2s)$-RIP (as defined in class), for some $\epsilon < 1$, then there cannot exist two s-sparse solutions to the system of equations $y = Wx$.