COMPSCI 514: Problem Set 4

Due: 11/25 by 11:59pm in Gradescope.

Instructions:

- You are allowed to work on this problem set in a group of up to three members.
- You should choose your group from within your own class (either online or in-person).
- You may talk to members of other groups at a high level about the problems but **not work through the solutions in detail together**.
- Each group should **submit a single solution set**: one member should upload a pdf to Gradescope, marking the other members as part of their group in Gradescope.
- You must show your work/derive any answers as part of the solutions to receive full credit.

Core Competency Problems

1. Rank One Matrices and Matrix Completion (10 points)

1. (2 points) Suppose $A \in \mathbb{R}^{n \times d}$ is a rank 1 matrix, i.e., all rows are multiples some row vector y^T. Without loss of generality you may assume that y is a unit vector. Prove, without appealing to the existence of the SVD, that A can be written as $\alpha x y^T$ for some unit vector $x \in \mathbb{R}^n$. Specifically, express α and each entry of x in terms of entries of A and y.

2. (2 points) If $A = \alpha x y^T$, find the eigenvector of $A^T A$ with the largest eigenvalue and derive an expression for this largest eigenvalue. To get full marks you need to fully explain your working.

3. (2 points) Let $B \in \mathbb{R}^{n \times d}$ be a partially observed matrix, i.e., we know the values of some of the entries but not others. Assume that $d \leq n$. We say rows $i, j \in [n]$ of B are *directly comparable* if there exists k just that $B_{i,k}$ and $B_{j,k}$ are both known. Prove that the unobserved entries of B can be deduced from the observed entries of B given the following four assumptions:

    - B has rank 1
    - All observed entries are non-zero.
    - There is at least one observed entry in each column.
    - For all $i \in \{1, 2, \ldots, n-1\}$, the ith and $(i+1)$th row are directly comparable.

4. (2 points) Suppose each entry of B is observed independently with probability p. Prove that if $p \geq \frac{\log(cd)}{n}$ for some sufficiently large constant c then with probability at least $9/10$, there exists an observed entry in every column.

5. (2 points) Suppose each entry of B is observed independently with probability p. Prove that if $p \geq \frac{\log(cn)}{\sqrt{d}}$ for some sufficiently large constant c then with probability at least 9/10, for all $i \in \{1, 2, \ldots, n-1\}$, the ith and $(i+1)$th row are directly comparable. **Note:** It is also possible to prove that $p \geq \sqrt{\frac{\log(cn)}{d}}$ suffices. This is only stronger than the bound we are asking for so will receive full credit.

2. SVD Practice (10 points)

Consider any matrix $A \in \mathbb{R}^{n \times d}$.

1. (2 points) Let v be a unit norm eigenvector of $A^T A$ with eigenvalue λ. Prove that $\lambda \geq 0$ and that $\|Av\|_2 = \sqrt{\lambda}$.

2. (2 points) Let v_1, v_2 be two eigenvectors of $A^T A$ that are orthogonal to each other and correspond to non-zero eigenvalues. Prove that Av_1 and Av_2 are also orthogonal to each other.

3. (2 points) Let $V \in \mathbb{R}^{d \times d}$ contain the d eigenvectors of $A^T A$ as its columns and let $\Lambda \in \mathbb{R}^{d \times d}$ contain their corresponding eigenvalues. Assume that $A^T A$ is full rank and so all of its eigenvalues (i.e., the diagonal entries of Λ) are positive. Prove that $U = AV\Lambda^{-1/2}$ has orthonormal columns. **Hint** Apply parts (1) and (2).

4. (2 points) Prove that if $V \in \mathbb{R}^{d \times d}$ has orthonormal columns then $VV^T = I$. Conclude using part (3) that we can write $A = U\Lambda^{1/2}V^T$, where U and V have orthonormal columns and Λ is diagonal. **Hint:** Use the interpretation of VV^T as a projection matrix in your proof.

5. (2 points) Let v be an eigenvector of $A^T A$ with eigenvalue λ. Prove that Av is is an eigenvector of AA^T with eigenvalue λ. Conclude that the columns of U from part (4) are eigenvectors of AA^T.

3. Eigendecomposition and Optimal Low-Rank Approximation (10 points)

Consider any matrix $A \in \mathbb{R}^{n \times d}$.

1. (2 points) Prove that

$$\underset{Z \in \mathbb{R}^{d \times k}: Z^T Z = I}{\arg\min} \|A - AZZ^T\|_F^2 = \text{argmax}_{Z \in \mathbb{R}^{d \times k}: Z^T Z = I} \operatorname{tr}(Z^T A^T A Z).$$

2. (2 points) Let $V \in \mathbb{R}^{d \times d}$ have orthonormal columns. Prove that for any $Z \in \mathbb{R}^{d \times k}$ with orthonormal columns, $V^T Z \in \mathbb{R}^{d \times k}$ also has orthonormal columns. Further prove that any $Z \in \mathbb{R}^{d \times k}$ with orthonormal columns can be written as $Z = V^T U$ for some $U \in \mathbb{R}^{d \times k}$ with orthonormal columns. **Hint:** Use Problem 2.4.

3. (2 points) Writing $A^T A = V\Lambda V^T$ in its eigendecomposition, use part (2) to prove that

$$\max_{Z \in \mathbb{R}^{d \times k}: Z^T Z = I} \operatorname{tr}(Z^T A^T A Z) = \max_{Z \in \mathbb{R}^{d \times k}: Z^T Z = I} \operatorname{tr}(Z^T \Lambda Z).$$

4. (2 points) Prove that for $Z \in \mathbb{R}^{d \times k}$ with orthonormal columns, $\operatorname{tr}(ZZ^T) = \sum_{i=1}^d (ZZ^T)_{i,i} = k$ and $0 \leq (ZZ^T)_{i,i} \leq 1$ for all $i \in [d]$. **Hint:** There are several ways to prove the second claim. One is by using Problem 2.4 again.

5. (2 points) Use parts (3) and (4) to show that $\max_{Z\in\mathbb{R}^{d\times k}:Z^TZ=I}\operatorname{tr}(Z^TA^TAZ)=\sum_{i=1}^{k}\lambda_i(A^TA)$. Conclude that Z optimizing the equation in part (1) has as its columns the k eigenvectors of A^TA corresponding to the top eigenvalues $\lambda_1(A^TA),\ldots,\lambda_k(A^TA)$.

Challenge Problems (Complete 1 of 2)

C1. Second Smallest Eigenvalue of a Random Graph (10 points)

1. (2 points) Prove that for any disconnected graph with Laplacian L, $\lambda_{n-1}(L)=0$.

2. (2 points) Prove that for any connected graph with Laplacian L, $\lambda_{n-1}(L)>0$.

3. (2 points) Consider a cut indicator vector $v\in\{-1,1\}^n$ with exactly k entries equal to 1. Let G be a random undirected and unweighted graph where each edge (excluding self-loops) is added independently with probability p. Prove that $\Pr[v^TLv=0]=(1-p)^{k(n-k)}$.

4. (2 points) Consider the setting of part (3). Let $\mathcal{C}$ be the set of all 'non-trivial' cut indicator vectors: i.e., vectors $v\in\{-1,1\}^n$ with k entries equal to 1 for some $k\in\{1,\ldots,n-1\}$. Prove that there exists $v\in\mathcal{C}$ with $v^TLv=0$ with probability at most $\sum_{k=1}^{\lfloor n/2\rfloor}\exp\left(k\ln n-pk(n-k)\right)$. **Hint:** Apply a union bound.

5. (2 points) Argue that if $p=\frac{c\ln n}{n}$ for a large enough constant c, then, with probability at least $99/100$, $v^TLv>0$ for all $v\in\mathcal{C}$. Conclude that with probability at least $99/100$, $\lambda_{n-1}(L)>0$.

C2. More Matrix Completion (10 points)

Consider a $n\times d$ matrix B that is partially observed, i.e., every entry is independently observed with probability p. Assume $d\le n$. We say a subset of rows U is isolated there doesn't exist $i\in U$ and $j\notin U$ such that row i and row j are directly comparable. (See Question 1 for the definition of directly comparable.)

1. (2 points) Prove that if

   - B has rank 1
   - All observed entries are non-zero
   - There is at least one observed entry in each column
   - No subset of rows is isolated.

   then B can be reconstructed from the observed entries.

2. (2 points) Prove that a subset of nodes U is isolated with probability at most

$$((1-p)^{|U|}+(1-p)^{n-|U|})^d\ .$$

3. (2 points) Prove $(1-p)^{|U|}+(1-p)^{n-|U|}\le\exp(-p|U|/2)$ assuming $p\ge c\ln(n)/d$ for a sufficiently large constant c. You may assume $|U|\le n/2$. **Hint:** First show

$$(1-p)^{|U|}+(1-p)^{n-|U|}\le\exp(-p|U|+\exp(-p(n-2|U|)))$$

   and consider the cases $|U|\le n/4$ and $|U|\ge n/4$ separately.

4. (2 points) Prove that the probability there exists a subset of nodes that is isolated is at most $1/n$ if $p \geq c\ln(n)/d$ for a sufficiently large constant c. **Hint:** Use the union bound.

5. (2 points) Describe a $n \times n$ matrix C with rank 1 such that the probability we can reconstruct C is at most $1/2^n$ even if each entry is independently observed with probability $1/2$ and observations are independent. **Hint:** C has zero entries otherwise $p = c(\log n)/n$ would suffice for reconstruction with high probability.