CSE 521: Design and Analysis of Algorithms **Fall 2020**

Problem Set 3

Deadline: Nov 19th in Canvas

1) a) Let $M \in \mathbb{R}^{n\times n}$ be a (symmetric) PSD matrix. Show that for any $P \in \mathbb{R}^{m\times n}$, $PMP^T \succeq 0$.

   b) Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix that has k positive eigenvalues for some integer $k \geq 1$. Use Cauchy's interlacing theorem (below) to show that for any $P \in \mathbb{R}^{m\times n}$, PMP^T has at most k positive eigenvalue.

   Theorem 3.1 (Cauchy's Interlacing Theorem). *Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix with eigenvalues $\lambda_n \leq \cdots \leq \lambda_1$. For a vector $v \in \mathbb{R}^n$, let $\beta_n \leq \cdots \leq \beta_1$ be the eigenvalues of $M + vv^T$. Then,*

$$\lambda_n \leq \beta_n \leq \lambda_{n-1} \leq \beta_{n-1} \leq \cdots \leq \lambda_1 \leq \beta_1.$$

2) Recall that for matrices $A, B \in \mathbb{R}^{n\times n}$ we write $A \succeq B$ iff $A - B \succeq 0$, i.e., $A - B$ is PSD.

   a) Let $A = vv^T$ be a rank 1 matrix. Prove that $A \preceq I$ iff $\text{trace}(A) \leq 1$.

   b) Let $A \in \mathbb{R}^{n\times n}$ be a (symmetric) positive definite matrix, and let $J = 1^{n\times n}$ be the all-ones matrix. Prove that $A \succeq J$ iff $A^{-1} \bullet J \leq 1$. Recall that for matrices $A, B \in \mathbb{R}^{n\times n}$,

$$A \bullet B = \text{trace}(AB^T) = \sum_{i,j} A_{i,j} B_{i,j}.$$

3) In this problem we discuss a fast algorithm for approximately estimating the low rank approximation (up to an additive error) with respect to the Frobenius norm.

   a) Let $A \in \mathbb{R}^{m\times n}$ and suppose we want to estimate Av for a vector $v \in \mathbb{R}^n$. Here is a randomized algorithm for this task. Choose the i-th column of A, A_i, with probability

$$p_i = \frac{\|A_i\|^2}{\|A\|_F^2}$$

   and let $X = A_i v_i / p_i$. Show that $\mathbb{E}[X] = Av$. Calculate $\text{Var}(X) = \mathbb{E}\left[\|X\|^2\right] - \|\mathbb{E}X\|^2$. Note that, with this definition, $\mathbb{E}[X]$ is a vector whereas $\text{Var}(X)$ is a number.

   b) Next, we use a similar idea to approximate A. For $1 \leq i \leq s$ let $X_i = \frac{A_j}{\sqrt{s p_j}}$ with probability p_j where $1 \leq j \leq n$. Let $X \in \mathbb{R}^{m\times s}$ and let X_i be the i-th columns of X. Note that $XX^T = \sum_{i=1}^s X_i X_i^T$. Show that

$$\mathbb{E}XX^T = AA^T.$$

   Show that $\mathbb{E}\|XX^T - AA^T\|_F^2 \leq \frac{1}{s}\|A\|_F^4$.

4) Let $A \in \mathbb{R}^{n\times n}$. Normally, we need to scan all non-zero entries of A to compute $\|A\|_F^2$. In this problem, we see that if A is PSD then we can approximate $\|A\|_F^2$ in time $O(n \log(1/\delta)/\epsilon^2)$ with probability at least $1 - \delta$. Note that this is sublinear in the number of non-zero entries of A. So, indeed our algorithm does not read all non-zero entries of A.

   a) First, assume that all diagonal entries of A are 1, i.e., $A_{i,i} = 1$ for all i. Show that for all $i \neq j$, $A_{i,j} \leq 1$.

   Hint: Use the fact that A is PSD iff $A = BB^\intercal$ for some $B \in \mathbb{R}^{n\times n}$. Note that you are not explicitly given B as part of the input in this problem, but you will use it in the analysis.

b) Still, assume all diagonal entries of A are 1. Show that by uniformly sampling $O(n/\epsilon^2)$ entries of A, we can approximate $\|A\|_F^2$ with a constant probability.

Hint: For $1 \leq i, j \leq n$, let $X = n^2 A_{i,j}^2$ with probability $1/n^2$. Show that X is an unbiased estimator of $\|A\|_F^2$. Compute the relative variance of x and show how to obtain $1 \pm \epsilon$ approximation of $\|A\|_F^2$.

c) Now, we solve the general case: In this case, we sample $A_{i,j}$ with probability $p_{i,j} = \frac{A_{i,i}A_{j,j}}{\sum_{k,l} A_{k,k}A_{l,l}}$ and if i, j is sampled we let $X = A_{i,j}^2 / p_{i,j}$. Show that X gives an unbiased estimator of $\|A\|_F^2$. Design an algorithm that by sampling $O(n \log(1/\delta)/\epsilon^2)$ coordinates of A gives a multiplicative $1 \pm \epsilon$ approximation of $\|A\|_F^2$ with probability at least $1 - \delta$.

Hint: Show that $\frac{\sum_{k,l} A_{k,k}A_{l,l}}{\sum_{i,j} A_{i,j}^2} \leq n$ and use it in your calculations.

d) In this part we want to implement the algorithm that was suggested in part (c). The file "psdnorm" in the course website has entries of a 5000×5000 matrix A. The i-th line of the file has 5000 numbers which are the entries of the row i of the A. First, compute $\|A\|_F^2$ exactly (by reading all entries). Then, estimate $\|A\|_F^2$ by taking empirical mean of 20000 entries of A (according to the above distribution). Do the same estimation this time with 100000 samples; how much does the estimate improve? Compute the median of the 100000 samples of the unbiased estimator. Does that give a better estimate than the empirical mean? Upload your code together with a description of your solutions to the above questions.