**CSE 521: Design and Analysis of Algorithms** **Fall 2018**

# Problem Set 3

*Deadline: Nov 12th in Canvas*

1) Prove the following Matrix equations:

   a) Let $A \in \mathbb{R}^{n\times n}$ and let $U \in \mathbb{R}^{n\times k}$ be a matrix with orthonormal columns $U_1, \dots, U_k$. So, $UU^T = \sum_{i=1}^k U_i U_i^T$ is a projection matrix. Show that
   $$\|A - UU^T A\|_F^2 = \|A\|_F^2 - \|U^T A\|_F^2.$$

   b) Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$. Show that $AB$ and $BA$ have the same set of nonzero eigenvalues, i.e., if $ABx = \lambda x$, then there exists a vector $y$ such that $BAy = \lambda y$.

2) For a vector $u \in \mathbb{R}^n$, we write $u \otimes u$ to denote the vector in $\mathbb{R}^{n^2}$ where for any $1 \le i, j \le n$, $(u \otimes u)_{in+j} = u_i \cdot u_j$.

   a) Show that for any pair of vectors $u, v \in \mathbb{R}^n$,
   $$\langle u \otimes u, v \otimes v\rangle = \langle u, v\rangle^2.$$

   b) Let $A \in \mathbb{R}^{n\times n}$ be a PSD matrix, and let $B \in \mathbb{R}^{n\times n}$ be the matrix where $B_{i,j} = A_{i,j}^2$. Prove that $B$ is PSD.
   **Hint:** Use part (a) and that any matrix $A$ is PSD iff it can be written as $A = CC^\intercal$ for some matrix $C \in \mathbb{R}^{n\times k}$, for some integer $k$.

   c) **Extra Credit:** Let $P = \{p_1, \dots, p_n\} \subseteq \mathbb{R}^d$ be a set of points of norm 1. For $\sigma > 0$, let $G_\sigma \in \mathbb{R}^{n\times n}$ be the Gaussian kernel on $P$, i.e.,
   $$G_\sigma(i,j) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\|p_i - p_j\|^2/2\sigma}.$$
   Prove that $G_\sigma \succeq 0$.

3) 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 $i \neq j$, $A_{i,j} \le 1$.
   **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 $1 \le i, j \le n$, 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 $1-\delta$.

4) 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 $\mathrm{Var}(X) = \mathbb{E}\left[\|X\|^2\right] - \|\mathbb{E}X\|^2$.

   b) Next, we use a similar idea to approximate $A$. For $1 \le i \le s$ let $X_i = \frac{A_j}{\sqrt{s p_j}}$ with probability $p_j$ where $1 \le j \le n$. 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 $\mathbb{E}\|XX^T - AA^T\|_F^2 \le \frac{1}{s}\|A\|_F^4$.

   c) **Extra Credit:** Let $X = \sum_{i=1}^s \sigma_i u_i v_i^T$ be the SVD decomposition of $X$ where $\sigma_1 \ge \cdots \ge \sigma_s$. Let $U_k$ be the matrix with columns $u_1, \ldots, u_k$. So, $U_k U_k^T = \sum_{i=1}^k u_i u_i^T$ is a projection matrix. We want to show that for any such matrix $X$ and $U_k$,

   $$\|A - U_k U_k^T A\|_F^2 \le \|A - A_k\|_F^2 + 2\sqrt{k}\|AA^T - XX^T\|_F, \tag{3.1}$$

   where $A_k$ is the best rank $k$ approximation of $A$. Note that if this is true we can simply let $s = O(k/\epsilon^2)$ and then a random $X$ chosen from part (b) would give

   $$\|A - U_k U_k^T A\|_F^2 \le \|A - A_k\|_F^2 + \epsilon\|A\|_F^2.$$

   Also, note that the algorithm runs in time $\mathrm{nnz}(A) + O(mk^2/\epsilon^4)$ as we need to compute the SVD of $X$.

   It remains to prove (3.1). First, by part (a) of Problem 1, we have

   $$\|A - U_k U_k^T A\|_F^2 \le \|A\|_F^2 - \|A^T U_k\|_F^2.$$

   Show that

   $$\left| \|A^T U_k\|_F^2 - \sum_{i=1}^k \sigma_i^2 \right| \le \sqrt{k}\|AA^T - XX^T\|_F.$$

   You can use without proof that

   $$\left| \sum_{i=1}^k \sigma_i^2 - \sum_{i=1}^k \sigma_i(A)^2 \right| \le \sqrt{k}\|AA^T - XX^T\|_F,$$

   where $\sigma_i(A)$ is the $i$-th largest singular value of $A$. Use the above two equations to conclude (3.1).

   d) Use the above algorithm to approxiamte the Einstein image we used in class. Specify how large $s$ should be to obtain a "good" approximation. Note that you do not need to calculate $s$ based on the bound on part c; instead just choose enough samples until the approximate image is close to the actual image. This is supposed to show that for $s$ much smaller than what the theory suggests you will get a good approximation. Upload the approximate image together with your code.