COMPSCI 514: Problem Set 3

Due: 11/14 by 11:59pm in Gradescope.

Instructions:

- You are allowed to, and highly encouraged to, work on this problem set in a group of up to three members.
- 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 may talk to members of other groups at a high level about the problems but **not work through the solutions in detail together**.
- You must show your work/derive any answers as part of the solutions to receive full credit.

1. Formulations of Low-Rank Approximation (5 points)

Prove that for any matrix $A \in \mathbb{R}^{n \times d}$ the quantities o_1, o_2, o_3, o_4 defined below are all equal.

1. $o_1 = \min_{B \in \mathbb{R}^{n \times d} \; s.t. \; \text{rank}(B) \leq k} \|A - B\|_F^2$.
2. $o_2 = \min_{M \in \mathbb{R}^{n \times k}, N \in \mathbb{R}^{d \times k}} \|A - MN^T\|_F^2$.
3. $o_3 = \min_{U \in \mathbb{R}^{n \times k} \; s.t. \; U^T U = I} \|A - UU^T A\|_F^2$.
4. $o_4 = \min_{V \in \mathbb{R}^{d \times k} \; s.t. \; V^T V = I} \|A - AVV^T\|_F^2$.

Hint 1: To formally prove that o_i is equal to o_j it may be helpful to argue that $o_i \not< o_j$ and also $o_j \not< o_i$, which implies that $o_i = o_j$.
Hint 2: You do not need to use anything about the SVD or eigendecomposition to prove that these quantities are equivalent.

2. Inner Products for Matrices (6 points)

In this question we will show that for two matrices $A \in \mathbb{R}^{n \times d}$ and $B \in \mathbb{R}^{d \times n}$ the quantity $\text{tr}(AB)$ behaves much like the standard inner product over vectors.

1. (2 points) Prove that $\text{tr}(AB) = \sum_{i=1}^{n} \sum_{j=1}^{d} A_{ij} \cdot B_{ji}$. **Hint:** Use the definition of matrix multiplication.
2. (1 point) Use part (1) to prove that $\text{tr}(AB) = \text{tr}(BA)$.

3. (1 point) Use part (1) to prove that for any $A \in \mathbb{R}^{n \times d}$, $\operatorname{tr}(AA^T) = \|A\|_F^2$.
4. (2 points) Prove that $|\operatorname{tr}(AB)| \leq \|A\|_F \cdot \|B\|_F$. **Hint:** Apply the Cauchy-Schwartz inequality to vectors in R^{nd} that correspond to A and B.

3. Random Projection for Faster Matrix Multiplication (10 points)

Let $\pi \in \mathbb{R}^n$ be a random vector with each entry set independently to 1 with probability 1/2 and -1 with probability 1/2. Let $A \in \mathbb{R}^{n \times d}$ be any matrix.

1. (2 points) Show that $\mathbb{E}[A^T \pi \pi^T A] = A^T A$.

   Hint: Fix $i, j \in [d]$ and show that $\mathbb{E}[(A^T \pi \pi^T A)_{ij}] = (A^T A)_{ij}$.

2. (2 points) Show that $\mathbb{E}[\|A^T \pi \pi^T A - A^T A\|_F^2] \leq 2\|A\|_F^4$.

   Hint: Fix $i, j \in [d]$ and show that $\mathbb{E}[(A^T \pi \pi^T A - A^T A)_{ij}^2] = \operatorname{Var}((A^T \pi \pi^T A)_{ij}) \leq 2\|a_i\|_2^2 \cdot \|a_j\|_2^2$ where $a_i, a_j \in \mathbb{R}^n$ are the i^{th} and j^{th} columns of A. Then sum over all $i, j \in [d]$

3. (2 points) Let $\Pi \in \mathbb{R}^{n \times m}$ be a random matrix with each entry set independently to $1/\sqrt{m}$ with probability 1/2 and $-1/\sqrt{m}$ with probability 1/2. Show that $\mathbb{E}[\|A^T \Pi \Pi^T A - A^T A\|_F^2] \leq \frac{2\|A\|_F^4}{m}$.

   Hint: Show that $A^T \Pi \Pi^T A = \frac{1}{m} \sum_{t=1}^m A^T \pi_t \pi_t^T A$, where $\pi_1, \ldots, \pi_t \in \mathbb{R}^n$ are independent random vectors distributed as in parts (1) and (2). Then leverage your work from part (2).

4. (2 points) Show that if $m = \frac{20}{\epsilon^2}$, then with probability at least 9/10, $\|A^T \Pi \Pi^T A - A^T A\|_F \leq \epsilon \|A\|_F^2$. **Note:** Here we are looking at the Frobenius norm of $A^T \Pi \Pi^T A - A^T A$, not the squared Frobenius norm.

5. (2 points) In terms of n, d, m, what is the runtime required to compute the approximate matrix product $A^T \Pi \Pi^T A$ as compared to the exact product $A^T A$.

4. Random Projection for Faster Low-Rank Approximation (10 points)

1. (2 points) In class we showed that for any $B \in \mathbb{R}^{n \times d}$, a span for the optimal rank-k subspace to approximate B in the Frobenius norm is given by:

$$Z = \underset{Z \in \mathbb{R}^{d \times k}, s.t.\ Z^T Z = I}{\arg\min} \|B - BZZ^T\|_F^2 = \operatorname{argmax}_{Z \in \mathbb{R}^{d \times k}, s.t.\ Z^T Z = I} \|BZ\|_F^2.$$

   Show that equivalently, $Z = \operatorname{argmax}_{Z \in \mathbb{R}^{d \times k}, s.t.\ Z^T Z = I} \operatorname{tr}(B^T B Z Z^T)$.

   Hint: Use Problems 2.3 and 2.2.

2. (2 points) Show that for any $A \in \mathbb{R}^{n \times d}$ and $C \in \mathbb{R}^{m \times d}$ and any $Z \in \mathbb{R}^{d \times k}$ with orthonormal columns, $|\operatorname{tr}(A^T A Z Z^T) - \operatorname{tr}(C^T C Z Z^T)| \leq \sqrt{k} \cdot \|A^T A - C^T C\|_F$.

   Hint: Use Problem 2.4 applied to the matrices $(A^T A - C^T C)$ and ZZ^T.

3. (2 points) Use parts (1) and (2) to argue that if $\tilde{Z} = \underset{Z \in \mathbb{R}^{d \times k}, s.t.\ Z^T Z = I}{\arg\min} \|C - CZZ^T\|_F^2$ then

$$\|A - A\tilde{Z}\tilde{Z}^T\|_F^2 \leq \left(\min_{Z \in \mathbb{R}^{d \times k}, s.t.\ Z^T Z = I} \|A - AZZ^T\|_F^2 \right) + 2\sqrt{k} \cdot \|A^T A - C^T C\|_F.$$

4. (2 points) Let $\Pi \in \mathbb{R}^{n \times m}$ be a random matrix with each entry set independently to $1/\sqrt{m}$ with probability 1/2 and $-1/\sqrt{m}$ with probability 1/2. Show that if $\tilde{Z} \in \mathbb{R}^{d \times k}$ contains the top k eigenvectors of $A^T \Pi \Pi^T A$ as its columns, then for $m = \frac{80k}{\epsilon^2}$, with probability $\geq 9/10$,

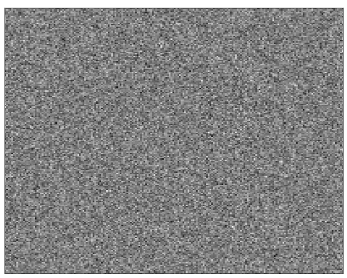
$$\|A - A\tilde{Z}\tilde{Z}^T\|_F^2 \leq \left(\min_{Z \in \mathbb{R}^{d \times k}, s.t.\, Z^T Z = I} \|A - AZZ^T\|_F^2 \right) + \epsilon \|A\|_F^2.$$

Hint: Apply part (3) in conjunction with Problem 3.4.

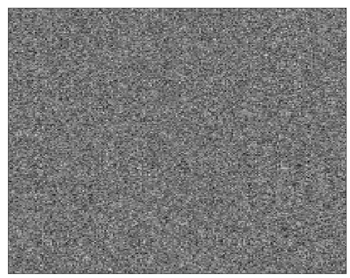
5. (2 points) In terms of n, d, k, and ϵ how does the runtime of computing $\tilde{Z}$ in part (4) compare to that of computing the actual top k eigenvectors of $A^T A$ (which would give an optimal low-rank approximation of A).

5. Distinguishing Random Matrices (10 points)

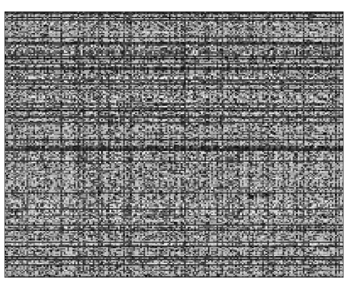
Consider the four 200×200 random matrices shown below. They are represented as 200×200 images, where a pixel is lighter when an entry in the matrix is relatively large, and darker when it is relatively small. The raw matrices can be downloaded in the `four_matrices.mat` file from the assignment page.



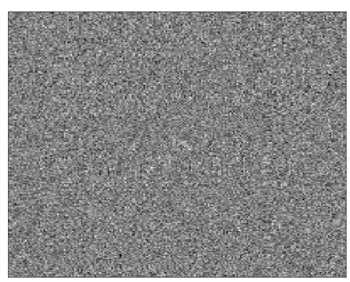
M1



M2



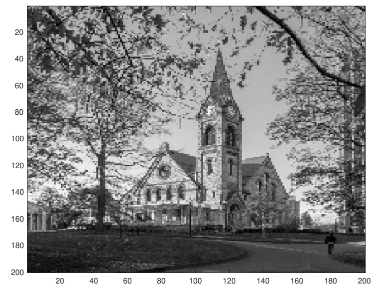
M3



M4

These matrices were generated from the following four distributions:

- **A1:** Each entry of the matrix is i.i.d. $\mathcal{N}(0, 1)$.
- **A2:** The matrix is equal to GV^T where $G \in \mathbb{R}^{200 \times 50}$ has i.i.d. random Gaussian entries and $V \in \mathbb{R}^{200 \times 50}$ is an orthonormal matrix.
- **A3:** The matrix is a mixture of the first two distributions. Specifically, it is equal to $0.1 \cdot B_1 + 0.9 \cdot B_2$ where B_1, B_2 are drawn from A_1 and A_2 respectively.
- **A4:** The matrix is generated by randomly permuting the rows and columns of the following 200×200 pixel image of the UMass Amherst campus:



1. (3 points) Let $M \in \mathbb{R}^{n \times d}$ be an arbitrary matrix and let $P_1 \in \mathbb{R}^{n \times n}$, $P_2 \in \mathbb{R}^{d \times d}$ be permutation matrices. Prove that the singular values of $P_1 M P_2$ are equal to those of M. I.e., if we change the order of the rows and columns of M this does not affect the spectrum of the matrix.

2. (3 points) Write code to compute the singular value spectrums of each of the four matrices. Show a plot of these spectrums and include a print out of your code.

3. (4 points) Use the spectrums computed above to match each matrix $M_1, \ldots M_4$ to the distribution in $A_1, \ldots, A_4$ that it was generated from. Explain why the spectrum is indicative of the distribution described.