CSE 521 Problem Set 3

Tyler Chen

Problem 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 $U_1, \ldots, U_k$. So, $UU^T = \sum_{i=1}^{k} U_i U_i^T$ is a projection matrix. Show that,

$$\left\|A - UU^T A\right\|_F^2 = \|A\|_F^2 - \left\|U^T A\right\|_F^2$$

(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 nonzero eigenvalues, i.e., if $ABv = \lambda v$, then there exists a vector y such that $BAy = \lambda y$.

Solution

(a) Recall that for any matrices X, Y, $\|X\|_F^2 = \operatorname{tr}(X^T X)$ and that $\operatorname{tr}(X+Y) = \operatorname{tr}(X) + \operatorname{tr}(Y)$. Note further than $U^T U = I_k$, the $k \times k$ identity. Therefore,

$$\begin{aligned}
\left\|A - UU^T A\right\|_F^2 &= \operatorname{tr}\left((A - UU^T A)^T (A - UU^T A)\right) \\
&= \operatorname{tr}\left((A^T - A^T UU^T)(A - UU^T A)\right) \\
&= \operatorname{tr}\left(A^T A - A^T UU^T A - A^T UU^T A + A^T UU^T UU^T A\right) \\
&= \operatorname{tr}\left(A^T A - A^T UU^T A\right) \\
&= \operatorname{tr}\left(A^T A\right) - \operatorname{tr}\left(A^T UU^T A\right) \\
&= \|A\|_F^2 - \left\|U^T A\right\|_F^2
\end{aligned}$$

(b) Suppose $ABv = \lambda v$ for some $\lambda \neq 0$ and v. Then,

$$BA(Bv) = \lambda(Bv)$$

This proves that λ is an eigenvalue of BA (with eigenvector Bv). The reverse direction is proved by relabling A, B, m, n.

Note that if $\lambda = 0$ then we cannot guarantee that $Bv \neq 0$ so $\lambda = 0$ may not be an eigenvalue of BA.

Problem 2

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

(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.

(c) **Extra Credit**: Let $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$ 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.,

$$G_\sigma(i,j) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\|p_i - p_j\|^2/2\sigma}$$

Show that $G_\sigma \succeq 0$.

Solution

(a) We prove a more general result where $(u \otimes v)_{(i-1)n+j} = u_i v_j$.

By definition of inner product, and using the bijection $k \leftrightarrow (i-1)n + j$ between $\{1, \ldots, n^2\}$ and $\{1, \ldots, n\}^2$,

$$\langle u \otimes v, x \otimes y \rangle = \sum_{k=1}^{n^2} (u \otimes v)_k (x \otimes y)_k = \sum_{i=1}^{n} \sum_{i=1}^{n} (u \otimes v)_{(i-1)n+j} (x \otimes y)_{(i-1)n+j}$$

Now, applying the definition of $u \otimes v$ and $x \otimes y$,

$$\langle u \otimes v, x \otimes y \rangle = \sum_{i=1}^{n} \sum_{i=1}^{n} u_i v_j x_i y_j = \left(\sum_{i=1}^{n} u_i x_i \right) \left(\sum_{j=1}^{n} v_j y_j \right) = \langle u, x \rangle \langle v, y \rangle$$

(b) We prove the more general result that for two PSD matrices C and D that the matrix B defined by $B_{i,j} = C_{i,j} D_{i,j}$ is also PSD.

Since C and D are PSD we can write,

$$C = UU^T, \qquad D = VV^T$$

Then, denoting the i-th row of U by U_i and the i-th row of V by V_i,

$$C_{i,j} = \langle U_i, U_j \rangle, \qquad D_{i,j} = \langle D_i, D_j \rangle$$

Therefore,

$$D_{i,j} = C_{i,j} D_{i,j} = \langle U_i, U_j \rangle \langle V_i, V_j \rangle = \langle U_i \otimes V_i, U_j \otimes V_j \rangle$$

But this means $B = XX^T$ where the $(i-1)n + j$-th row of X is $U_i \otimes V_j$. Therefore B is PSD.

(c) Note that $\|p_i - p_j\|^2 = \|p_i\|^2 + \|p_j\|^2 - \langle p_i, p_j \rangle = 2 - 2\langle p_i, p_j \rangle$. Therefore,

$$G_\sigma(i,j) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\|p_i - p_j\|^2/2\sigma} = \frac{e^{-1/\sigma}}{\sqrt{2\pi\sigma}} e^{\langle p_i, p_j \rangle/\sigma}$$

We then have,

$$G_\sigma(i,j) = \frac{e^{-1/\sigma}}{\sqrt{2\pi\sigma}} \sum_{k=1}^{\infty} \frac{\langle p_i, p_j \rangle^k}{k!}$$

Let $C_{i,j} = \langle p_i, p_j \rangle$ and define $P = [p_1, \ldots, p_t]$ to be the matrix with rows p_i. Then,

$$C_{i,j} = \langle p_i, p_j \rangle = (P^T P)_{i,j}$$

We have show that the matrix with entries $C_{i,j}^k$ is also PSD. Moreover, the scalar multiple of a PSD matrix is PSD, as is the sum of PSD matrices. Therefore each partial sum is PSD, and since the total sum converges, it is also PSD.

This proves G_σ is PSD.

Problem 3

Let $A \in R^{n\times n}$. 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 $\mathcal{O}(n\log(1/\delta)/2)$ 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 $i \neq j$, $A_{i,j} \leq 1$.

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

(c) Now, we solve the general case: In this case, we sample $A_{i,j}$ with probability $p_{i,j} = (A_{i,i}A_{j,j})/(\sum_{k,l} A_{k,k}A_{l,l})$ 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 $\mathcal{O}(n\log(1/\delta)/\epsilon^2)$ coordinates of A) gives a multiplicative $1\pm\epsilon$ approximation of $\|A\|_F^2$ with probability at least $1-\delta$.

Solution

Recall the theorem from lecture 5: Let $X_1,\ldots,X_k$ be iid with mean μ and relative variance t. Then the average $(X_1+\cdots+X_k)/k$ is a $1\pm\epsilon$ approximation to μ with probability $9/10$ for some $k = \mathcal{O}(t/\epsilon^2)$.

Recall that by returning the median of $\log(1/\delta)$ of such trials the probability of success is boosted to $1-\delta$. That is, a $1\pm\epsilon$ multiplicative approximation to μ can be obtained with probability $1-\delta$ using $\mathcal{O}(t\log(1/\delta)/\epsilon^2)$ samples.

(a) Assume A is symmetric positive-semidefinite. Then, $x^TAx \geq 0$ for all x. In particular,

$$0 \leq (e_i - e_j)^T A(e_i - e_j) = A_{i,i} - A_{i,j} - A_{j,i} + A_{j,j} = 2 - 2A_{i,j}$$

Therefore $A_{i,j} \leq A_{i,i} = 1$.

Note also that,

$$0 \leq (e_i + e_j)^T A(e_i + e_j) = A_{i,i} + A_{i,j} + A_{j,i} + A_{j,j}$$

Therefore $A_{i,j} \geq -A_{i,i}$.

(b) Let $X = n^2(A_{i,j})^2$ with probability $1/n^2$. Then,

$$\mathbb{E}[X] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{P}[\text{choose } i,j]\, n^2(A_{i,j})^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}(A_{i,j})^2 = \|A\|_F^2$$

Similarly,

$$\mathbb{E}[X^2] = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{P}[\text{choose } i,j]\, (n^2A_{i,j}^2)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} n^2(A_{i,j})^4$$

When the diagonal entries of $A = A^T$ are all ones we have,

$$\|A\|_F^2 = \sum_i \sum_j (A_{i,j})^2 = \sum_i (A_{i,i})^2 + \sum_i \sum_{j\neq i} (A_{i,j})^2 = n + \sum_i \sum_{j\neq i} (A_{i,j})^2$$

Now observe that since $|A_{i,j}| \leq 1$ and all terms of the sum are positive,

$$\|A\|_F^4 = \left(\sum_{i,j} (A_{i,j})^2 \right) \left(n + \sum_{i\neq j} (A_{i,j})^2 \right) \geq n \sum_{i,j} (A_{i,j})^2 \geq n \sum_{i,j} (A_{i,j})^4$$

Therefore,

$$\frac{\mathbb{E}[X^2] - \mathbb{E}[X]^2}{\mathbb{E}[X]^2} = \frac{n^2 \sum_{i,j} (A_{i,j})^4}{\left(\|A\|_F^2 \right)^2} - 1 \leq n - 1$$

We can then obtain a $1 \pm \epsilon$ multiplicative approximation to $\|A\|_F^2$ with probability $9/10$ by sampling $\mathcal{O}((n-1)/\epsilon^2) = \mathcal{O}(n/\epsilon^2)$ entries of A.

(c) Let $X = (A_{i,j})^2 / p_{i,j}$ with probability $p_{i,j}$. Then,

$$\mathbb{E}[X] = \sum_{i=1}^n \sum_{j=1}^n \mathbb{P}[\text{choose } i,j] \; (A_{i,j})^2/p_{i,j} = \sum_{i=1}^n \sum_{j=1}^n (A_{i,j})^2 = \|A\|_F^2$$

Similarly,

$$\mathbb{E}[X^2] = \sum_{i=1}^n \sum_{j=1}^n \mathbb{P}[\text{choose } i,j] \; \left(\frac{(A_{i,j})^2}{p_{i,j}} \right)^2 = \sum_{i=1}^n \sum_{j=1}^n \frac{(A_{i,j})^4}{p_{i,j}}$$

With the choice of $p_{i,j}$ we have,

$$\mathbb{E}[X^2] = \left(\sum_{k=1}^n \sum_{\ell=1}^n A_{k,k} A_{\ell,\ell} \right) \sum_{i=1}^n \sum_{j=1}^n \frac{(A_{i,j})^4}{A_{i,i} A_{j,j}}$$

We again want to show the relative variance of X is $\mathcal{O}(n)$. The result will then directly follow by the Theorem from lecture 5.

Note that since $|A_{i,j}| \leq \min(A_{i,i}, A_{j,j})$,

$$\sum_{i=1}^n \sum_{j=1}^n \frac{(A_{i,j})^4}{A_{i,i} A_{j,j}} \leq \sum_{i=1}^n \sum_{j=1}^n (A_{i,j})^2 = \|A\|_F^2$$

Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A so that,

$$\|A\|_F^2 = \text{tr}(A^T A) = \text{tr}(A^2) = \sum_{i=1}^n \lambda_i^2$$

Now note that,

$$\sum_{k=1}^{n}\sum_{\ell=1}^{n} A_{k,k}A_{\ell,\ell} = \left(\sum_{k=1}^{n} A_{k,k}\right)\left(\sum_{\ell=1}^{n} A_{\ell,\ell}\right) = \operatorname{tr}(A)^2$$

Therefore,

$$\operatorname{tr}(A)^2 = \left(\sum_{i=1}^{n} \lambda_i\right)^2 \leq n\sum_{i=1}^{n} \lambda_i^2 = n\operatorname{tr}(A^2) = n\operatorname{tr}(A^TA) = n\left\|A\right\|_F^2$$

Where we have used the fact that for $a_i \geq 0$,

$$\left(\sum_{i=1}^{n} a_i\right)^2 \leq n\sum_{i=1}^{n} a_i^2$$

which follows from the fact that $2ab \leq a^2 + b^2$ whenever $a, b \geq 0$ (expand $(a-b)^2$). Putting these together we have that,

$$\mathbb{E}[X^2] = \left(\sum_{k=1}^{n}\sum_{\ell=1}^{n} A_{k,k}A_{\ell,\ell}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(A_{i,j})^4}{A_{i,i}A_{j,j}} \leq n\left\|A\right\|_F^2\left\|A\right\|_F^2$$

Therefore,

$$\frac{\mathbb{E}[X^2] - \mathbb{E}[X]^2}{\mathbb{E}[X]^2} = n - 1$$

The algorithm to estimate $\|A\|_F$ within a $1 \pm \epsilon$ multiplicative error with probability $1-\delta$ is as follows: 1. Take the average of $k = \mathcal{O}(n/\epsilon^2)$ independent observations of X, 2. repeat this $\log(1/\delta)$ time and return the median of these repetitions.

Problem 4

In this problem we discuss a fast 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 $\mathbb{V}[X] = \mathbb{E}[\|X\|^2] - \|\mathbb{E}[X]\|^2$.

(b) Next, we use a similar idea to approximate A. For $1 \le i \le s$, let $X_i = A_j / \sqrt{sp_j}$ with probability p_j where $1 \le i \le n$. Let $X \in \mathbb{R}^{m\times s}$ and let X_i be the i-th column 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,

$$\mathbb{E}\left[\left\|XX^T - AA^T\right\|_F^2\right] \le \frac{1}{s}\|A\|_F^4$$

(c) **Extra Credit**: Let $X = \sum_{i=1}^{s} \sigma_i u_i v_i^T$ be the SVD of X where $\sigma_1 \le \cdots \le \sigma_s$. 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,

$$\left\|A - U_k U_k^T A\right\|_F^2 \le \|A - A_k\|_F^2 + 2\sqrt{k}\left\|AA^T - XX^T\right\|_F \qquad (*)$$

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

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

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

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

$$\left\|A - U_k U_k^T A\right\|_F^2 \le \|A\|_F^2 - \left\|A^T U_k\right\|_F^2$$

Show that,

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

You can use without proof,

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

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

(d) Use the above algorithm to approximate the Einstein image we used in class. Specify how large k should be to obtain a "good" approximation. Upload the approximate image together with your code.

Solution

(a) By definition,

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{P}\left[X = A_i v_i / p_i\right] \frac{A_i v_i}{p_i} = \sum_{i=1}^{n} A_i v_i = Av$$

Similarly,

$$\mathbb{E}[\|X\|^2] = \sum_{i=1}^{n} \mathbb{P}\left[X = A_i v_i / p_i\right] \left\| \frac{A_i v_i}{p_i} \right\|^2 = \sum_{i=1}^{n} \frac{\|A_i\|^2 v_i^2}{p_i}$$

Now using the definition of p_i we have,

$$\mathbb{E}[\|X\|^2] = \sum_{i=1}^{n} \|A\|_F^2 v_i^2 = \|A\|_F^2 \sum_{i=1}^{n} v_i^2 = \|A\|_F^2 \|v\|^2$$

Therefore,

$$\mathbb{V}[X_j] = \mathbb{E}[\|X\|^2] - \|\mathbb{E}[X]\|^2 = \|A\|_F^2 \|v\|^2 - \|Av\|^2$$

(b) We have,

$$\mathbb{E}\left[XX^T\right] = \mathbb{E}\left[\sum_{i=1}^{s} X_i X_i^T\right] = \sum_{i=1}^{s} \sum_{j=1}^{n} \mathbb{P}\left[X_i = \frac{A_j}{\sqrt{sp_j}}\right] \left(\frac{A_j}{\sqrt{sp_j}}\right) \left(\frac{A_j}{\sqrt{sp_j}}\right)^T$$

Therefore, since $\mathbb{P}[X_i = A_j/\sqrt{sp_j}] = p_j$,

$$\mathbb{E}\left[XX^T\right] = \sum_{i=1}^{s} \sum_{j=1}^{n} \frac{1}{s} A_j A_j^T = AA^T$$

We now approach the second part. We start with,

$$\left\|XX^T - AA^T\right\|_F^2 = \left\|XX^T\right\|_F^2 + \left\|AA^T\right\|_F^2 - 2\left\langle XX^T, AA^T \right\rangle_F$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\left\|XX^T - AA^T\right\|_F^2\right] &= \mathbb{E}\left[\left\|XX^T\right\|_F^2\right] + \mathbb{E}\left[\left\|AA^T\right\|_F^2\right] - \mathbb{E}\left[2\left\langle XX^T, AA^T \right\rangle_F\right] \\
&= \mathbb{E}\left[\left\|XX^T\right\|_F^2\right] + \left\|AA^T\right\|_F^2 - 2\left\langle \mathbb{E}[XX^T], AA^T \right\rangle \\
&= \mathbb{E}\left[\left\|XX^T\right\|_F^2\right] + \left\|AA^T\right\|_F^2 - 2\left\langle AA^T, AA^T \right\rangle \\
&= \mathbb{E}\left[\left\|XX^T\right\|_F^2\right] - \left\|AA^T\right\|_F^2
\end{aligned}$$

We can write,

$$\left\|AA^T\right\|_F^2 = \operatorname{tr}\left(\left(\sum_{i=1}^n A_iA_i^T\right)\left(\sum_{j=1}^n A_iA_j^T\right)\right) = \operatorname{tr}\left(\sum_{i=1}^n\sum_{j=1}^n A_iA_i^TA_iA_j^T\right)$$

Similarly,

$$\left\|XX^T\right\|_F^2 = \operatorname{tr}\left(\left(\sum_{p=1}^s X_pX_p^T\right)\left(\sum_{q=1}^s X_qX_q^T\right)\right) = \operatorname{tr}\left(\sum_{p=1}^s\sum_{q=1}^s X_pX_p^TX_qX_q^T\right)$$

Observe that for $p \neq q$, since the columns of X are chosen independently,

$$\mathbb{E}\left[X_pX_p^TX_qX_q^T\right] = \sum_{i=1}^n\sum_{j=1}^n p_ip_j\frac{A_iA_i^T}{sp_i}\frac{A_jA_j^T}{sp_j} = \frac{1}{s^2}\sum_{i=1}^n\sum_{j=1}^n A_iA_i^TA_jA_j^T$$

Therefore, by the linearity of trace,

$$\operatorname{tr}\left(\mathbb{E}\left[X_pX_p^TX_qX_q^T\right]\right) = \frac{1}{s^2}\sum_{i=1}^n\sum_{j=1}^n\operatorname{tr}\left(A_iA_i^TA_jA_j^T\right) = \frac{1}{s^2}\left\|AA^T\right\|_F^2$$

For $p = q$,

$$\mathbb{E}\left[X_pX_p^TX_pX_p^T\right] = \sum_{i=1}^n p_i\frac{A_iA_i^T}{sp_i}\frac{A_iA_i^T}{sp_i} = \frac{1}{s^2}\sum_{i=1}^n\frac{1}{p_i}A_iA_i^TA_iA_i^T$$

Therefore, since $\operatorname{tr}(A_iA_i^TA_iA_i^T) = A_i^TA_i\operatorname{tr}(A_iA_i^T) = A_i^TA_i\operatorname{tr}(A_i^TA_i) = \|A_i\|^4$, $p_i = \|A_i\|^2 / \|A\|_F^2$, and $\sum_i \|A_i\|^2 = \|A\|_F^2$,

$$\operatorname{tr}\left(\mathbb{E}\left[X_pX_p^TX_pX_p^T\right]\right) = \frac{1}{s^2}\sum_{i=1}^n\frac{\|A\|_F^2}{\|A_i\|^2}\operatorname{tr}\left(A_iA_i^TA_iA_i^T\right) = \frac{1}{s^2}\sum_{i=1}^n\|A\|_F^2\|A_i\|^2 = \frac{1}{s^2}\|A\|_F^4$$

We now put everything together. First separate the sum for XX^T into terms with $p = q$ and $p \neq q$. That is,

$$\begin{aligned}\mathbb{E}\left[\operatorname{tr}\left(XX^T\right)\right] &= \sum_p\operatorname{tr}\left(\mathbb{E}\left[X_pX_p^TX_pX_p^T\right]\right) + \sum_p\sum_{q\neq p}\operatorname{tr}\left(\mathbb{E}\left[X_pX_p^TX_qX_q^T\right]\right)\\ &= \sum_p\frac{1}{s^2}\|A\|_F^4 + \sum_p\sum_{q\neq p}\frac{1}{s^2}\left\|AA^T\right\|_F^2\\ &= \frac{1}{s}\|A\|_F^4 + \frac{s(s-1)}{s^2}\left\|AA^T\right\|_F^2\end{aligned}$$

Therefore, returning to our initial expression,

$$\begin{aligned}\mathbb{E}\left[\left\|XX^T - AA^T\right\|_F^2\right] &= \mathbb{E}\left[\left\|XX^T\right\|_F^2\right] - \left\|AA^T\right\|_F^2 \\ &= \frac{1}{s}\left\|A\right\|_F^4 + \frac{s(s-1)}{s^2}\left\|AA^T\right\|_F^2 - \left\|AA^T\right\|_F^2 \\ &= \frac{1}{s}\left\|A\right\|_F^4 - \frac{1}{s^2}\left\|AA^T\right\|_F^2 \\ &\leq \frac{1}{s}\left\|A\right\|_F^4\end{aligned}$$

(c) I couldn't figure out how to prove this result:

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

But using it I was able to prove the main result.

Some attempts at the first part...

Observe,

$$\left\|A^TU_k\right\|_F^2 = \operatorname{tr}\left((A^TU_k)^T(A^TU_k)\right) = \operatorname{tr}\left(U_k^TAA^TU_k\right) = \operatorname{tr}\left(U_kU_k^TAA^T\right)$$

This is a projection of A onto the column space of U_k, so,

$$\operatorname{tr}\left(U_kU_k^TAA^T\right) = \sum_{i=1}^k \sigma_i(U_kU_k^TAA^T) \leq \sum_{i=1}^k \sigma_i(AA^T) = \sum_{i=1}^k \sigma_i(A)^2$$

with equality exactly when the column space of U_k corresponds to the space spanned by the first k singular vectors of A.

However, this is only useful if the absolute value were not there..

$$\left|\left\|A^TU_k\right\|_F^2 - \sum_{i=1}^k \sigma_i^2\right| \leq \sqrt{k}\left\|\right\|$$

Using the definitions of Frobenius norm we have,

$$\begin{aligned}\left|\left\|A^TU_k\right\|_F^2 - \sum_{i=1}^k \sigma_i^2\right| &= \left|\left\|A^TU_k\right\|_F^2 - \left\|X^TU_k\right\|_F^2\right| \\ &= \left|\operatorname{tr}(U_k^TAA^TU_k) - \operatorname{tr}(U_k^TXX^TU_k)\right| \\ &= \left|\operatorname{tr}(U_k^T(AA^T - XX^T)U_k)\right| \\ &= \left|\operatorname{tr}(U_kU_k^T(AA^T - XX^T)\right|\end{aligned}$$

We now prove the main result. By the triangle inequality,

$$\begin{aligned}\left| \left\| A^T U_k \right\|_F^2 - \sum_{i=1}^{k} \sigma_i(A)^2 \right| &\leq \left| \left\| A^T U_k \right\|_F^2 - \sum_{i=1}^{k} \sigma_i^2 \right| + \left| \sum_{i=1}^{k} \sigma_i^2 - \sum_{i=1}^{k} \sigma_i(A)^2 \right| \\ &\leq 2\sqrt{k} \left\| AA^T - XX^T \right\|_F\end{aligned}$$

Note that the best rank k approximation to A (with respect to the Frobenius norm or 2-norm) is the rank k truncated SVD of A. Therefore,

$$\|A\|_F^2 = \sum_{i=1}^{n} \sigma_i^2 = \sum_{i=1}^{k} \sigma_i(A)^2 + \sum_{i=k+1}^{n} \sigma_i(A)^2 = \sum_{i=1}^{k} \sigma_i(A)^2 + \|A - A_k\|_F^2$$

Using the last two results we have,

$$\begin{aligned}\|A\|_F^2 - \left\| A^T U_k \right\|_F^2 &= \|A - A_k\|_F^2 + \sum_{i=1}^{k} \sigma_i(A)^2 - \left\| A^T U_k \right\|_F^2 \\ &\leq \|A - A_k\|_F^2 + 2\sqrt{k} \left\| AA^T - XX^T \right\|_F\end{aligned}$$

Therefore, using the result from 1 we have,

$$\left\| A - U_k U_k^T A \right\|_F^2 \leq \|A - A_k\|_F^2 + 2\sqrt{k} \left\| AA^T - XX^T \right\|_F$$

(d) We first implement a rank-k SVD approximation.

```
def compress_image(A,k):
    u,s,vt = np.linalg.svd(A, full_matrices=False)

    return (u[:,:k]*s[:k])@vt[:k]
```

We then implement the fast approximate algorithm.

```
def col_sample_compress_image(A,k,eps):
    m,n = np.shape(A)

    column_norms = np.linalg.norm(A,axis=0)

    column_probs = column_norms**2
    column_probs /= np.sum(column_probs)

    s = int(k/eps**2)

    indices = np.random.choice(np.arange(n), size=s, p=column_probs)

    X = A[:,indices]

    u,s,vt = np.linalg.svd(X, full_matrices=False)

    return u[:,:k]@(u[:,:k].T@A)
```

Note that U from the fast algorithm is simply an approximation of the range (column space) of A obtained though columns sampling. Another way to obtain such a basis would be to orthogonalize $A\omega_i$, $i = 1, \ldots, k$, where ω_i are Guassian random vectors.

```
def gaus_sample_compress_image(A,k):
    m,n = np.shape(A)

    X = A@np.random.randn(n,k)
    u,s,vt = np.linalg.svd(X, full_matrices=False)

    return u[:,:k]@(u[:,:k].T@A)
```

The optimal rank 34 approximation looks very similar to the original image. The is comparable to the rank 206 column sampling matrix and rank 84 guassian sampling matrix.

We compare the time it takes to run the three methods. Note that we only time the matrix operations, and not any I/O. There is a lot of overhead with displaying the image, so it doesn't make sense to include this in the comparison.

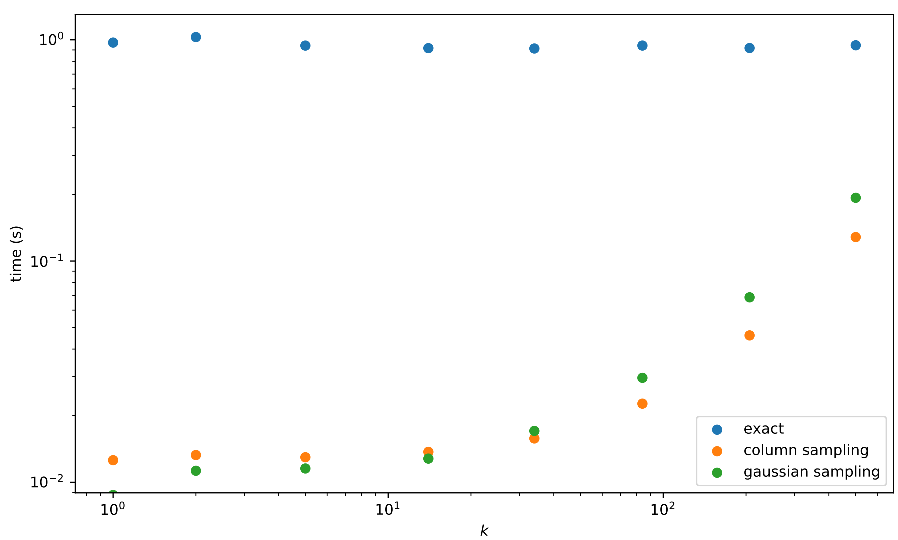
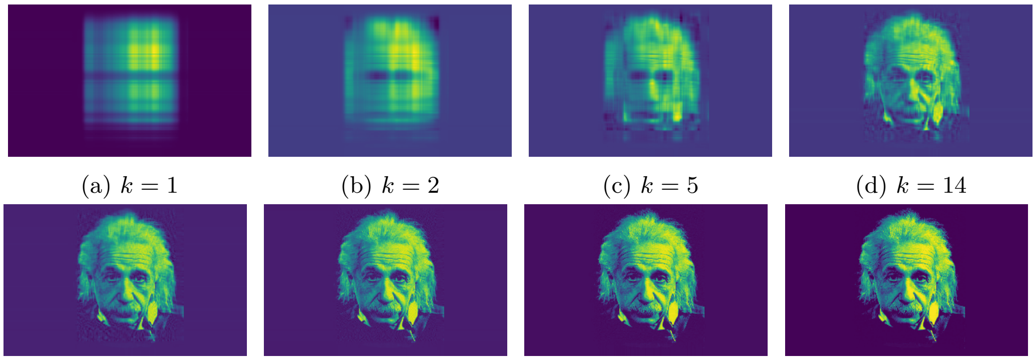


Figure 1: Runtime vs. rank of approximation

We also take $s = k$, since there is no point computing the approximate SVD and then discarding extra information. Note further than since we are just looking to approximate the range we could use a QR factorization which would be faster than SVD (by a constant factor).

The time to compute the approximations are shown in Figure 1. Note that the time to compute the optimal rank k approximation is independent of k since the full SVD must be computed.

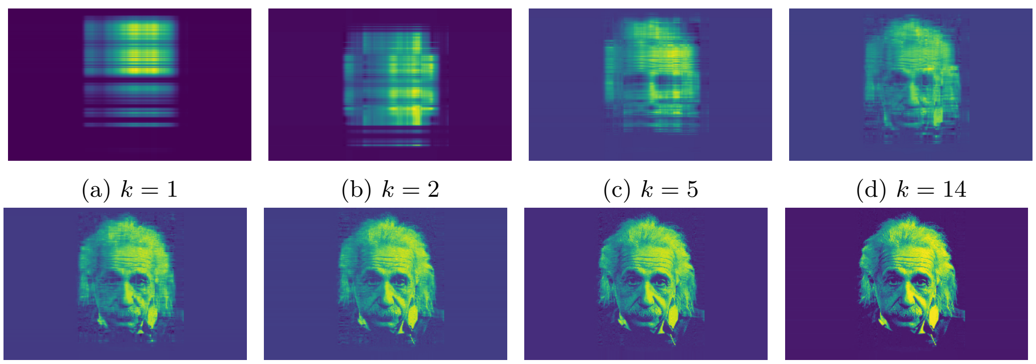
Sample outputs are shown in Figures 2,3,4. Note that the images shown here have been downsampled to reduce the size of this document.



(a) $k = 1$ (b) $k = 2$ (c) $k = 5$ (d) $k = 14$

(e) $k = 34$ (f) $k = 84$ (g) $k = 206$ (h) $k = 501$

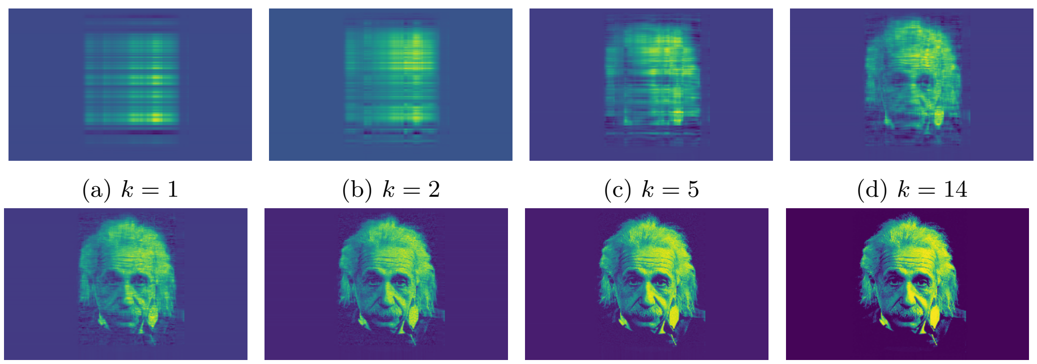
Figure 2: optimal rank-k approximation via SVD truncation



(a) $k = 1$ (b) $k = 2$ (c) $k = 5$ (d) $k = 14$

(e) $k = 34$ (f) $k = 84$ (g) $k = 206$ (h) $k = 501$

Figure 3: rank-k approximation via column sampling



(a) $k = 1$ (b) $k = 2$ (c) $k = 5$ (d) $k = 14$

(e) $k = 34$ (f) $k = 84$ (g) $k = 206$ (h) $k = 501$

Figure 4: rank-k approximation via Gaussian test matrix