Exercises

Exercises are for fun and edification, please do not submit. But please do try to solve them, we may need ideas from there later in the course (even in this very homework)!

1. Show that $\|a + b\|^2 = \|a\|^2 + \|b\|^2 - 2\langle a, b \rangle$, where we’re looking at the Euclidean norm.

2. Complete the proof (given on Piazza and sketched in lecture for the case $k = 1, 2$): if the columns of $V$ are the top $k$ eigenvalues of $A^TA$, then $V$ solves the PCA problem for dimension $k$. Namely, it is a maximum for $\max_V \|AV\|_F^2$, where the maximum is taken over all matrices with orthonormal columns.

3. If square matrix $A$ has orthonormal rows, show that it has orthonormal columns.

4. Check that the SVD $A = UDV^\top$ means that $A = \sum_i \sigma_i u_i v_i^\top$. Note that each $u_i v_i^\top$ is multiplying the column vector $u_i$ and the row vector $v_i^\top$ to get a matrix! Each such matrix has rank one (why?). So the rank-$r$ matrix $A$ is being written as the sum of $r$ rank-1 matrices.

Recall that $\text{Trace}(PQ) = \text{Trace}(QP)$, as long as matrices $P$ and $Q^\top$ have the same dimensions. So show that the trace of a square rank-1 matrix $\sigma uv^\top$ equals $\sigma \langle u, v \rangle$.

5. For a symmetric matrix $A$ with distinct singular values, show that the left and right singular vectors are the same up to signs. Namely, $A = \hat{V}DV^\top$ for some orthonormal $V$ and $\hat{V}$ and diagonal $D$, where each column $V^i$ of $V$ is either the same as the column $\hat{V}^i$, or it is $-\hat{V}^i$. (You may use that the SVD of a matrix with distinct singular values is unique, up to the flipping the signs of columns of $U$ and $V$.)

6. For a symmetric matrix $A$, show that all its eigenvalues are real. Show that $A$ is PSD if and only if all its eigenvalues are non-negative. (See Problem 3 for a definition of PSD.)

Problems

Please solve both of #3 and #4, and any one of #1 or #2.

1. **How Many Almost-Orthonormal Vectors.**

Recall that there are exactly $d$ orthonormal vectors in $\mathbb{R}^d$. Call two unit-length vectors “near-orthonormal” if their inner product $\langle x, y \rangle$ has small value; in this problem we show there can be exponentially more near-orthonormal vectors. Recall the inner product $\langle x, y \rangle := \sum_{i=1}^d x_i y_i$, and let $\|\vec{x}\| = \sqrt{\sum_{i=1}^d x_i^2} = \sqrt{\langle x, x \rangle}$ be the Euclidean norm.

---

1This is often called the *outer product* of two vectors.
2. Faster Approximate MatMult by “Preconditioning”. We now use the randomized dimension reduction procedure (the Johnson Lindenstrauss Flattening Lemma) to speed up matrix multiplications, at least approximately. Say that the matrix $S \in \mathbb{R}^{k \times D}$ $\varepsilon$-preserves vector $x \in \mathbb{R}^D$ if

$$1 - \varepsilon \leq \frac{\|Sx\|^2}{\|x\|^2} \leq 1 + \varepsilon.$$ 

(i) For unit vectors $x, y$, suppose $S$ $\varepsilon$-preserves all four of $x$, $y$, $x + y$, $x - y$. Show that

$$\langle x, y \rangle - \varepsilon \|x\|\|y\| \leq \langle Sx, Sy \rangle \leq \langle x, y \rangle + \varepsilon \|x\|\|y\|.$$ 

Hint: recall the parallelogram law that

$$2\|x\|^2 + 2\|y\|^2 = \|x + y\|^2 + \|x - y\|^2.$$ 

(ii) (Nothing to do here.) Given $A \in \mathbb{R}^{m \times D}$ and $B \in \mathbb{R}^{D \times n}$, let $V$ be the set of all $(m + n)$ unit vectors obtained by normalizing rows of $A$ or columns of $B$. If we set $k = O(\log(mn/\eta)/\varepsilon^2)$ and choose $S$ to be a random $k \times D$ Gaussian matrix scaled down by $\frac{1}{\sqrt{k}}$, the JL theorem gives us

$$\Pr \left[ \forall x, y \in V, \text{ matrix } S \varepsilon\text{-preserves all four of } x, y, x + y, x - y \right] \geq 1 - \eta. \quad (1)$$
(iii) Assuming the event in (1) is true, show that for all \( i, j \)

\[
| (AB)_{ij} - (AS^T SB)_{ij} | \leq \varepsilon \| A_{i*} \| \| B_{*j} \|
\]

(Be careful: the row \( A_{i*} \) and column \( B_{*j} \) are not unit vectors.)

(iv) Complete the argument by showing that by choosing \( S \) as above,

\[
\Pr \left[ \| AB - AS^T SB \|_F \leq \varepsilon \| A \|_F \| B \|_F \right] \geq 1 - \eta.
\]

(v) Assuming we use naive matrix multiplication, what is the runtime of computing \( AB \) versus \( AS^T SB \)? Show that when \( k \ll D, n, m \) the improvement is considerable.

3. (Are you Sure? Definitely.) A square symmetric matrix \( A \in \mathbb{R}^{n \times n} \) is positive semi-definite (PSD) if for all \( x \in \mathbb{R}^n \), \( x^T Ax \geq 0 \).

(a) If \( B \) is a real-valued matrix, then show that \( A = B^T B \) is PSD.

(b) Conversely, show that if \( A \) is PSD, then there exists a real-valued matrix \( B \) such that \( A = B^T B \).

(c) Prove that \( A \) is PSD if and only if there exist a set of vectors \( v_1, v_2, \ldots, v_n \) from \( \mathbb{R}^k \) such that \( A_{i,j} = \langle v_i, v_j \rangle \). Find what is the minimal possible value of \( k \) such that such set of vectors exists.

(d) Let \( B \in \mathbb{R}^{m \times n} \) be the vertex-edge incidence matrix of an undirected graph \( G \) with \( n \) vertices and \( m \) edges: namely the rows of \( B \) are indexed by edges and columns by vertices. The row corresponding to edge \( e = \{ u, v \} \) has zeros in all columns other than \( u, v \), it has an entry \( +1 \) in one of those columns (say \( u \)) and an entry \( -1 \) in the other (say \( v \)).

Define the Laplacian matrix \( L_G \in \mathbb{R}^{n \times n} \) of graph \( G \) to be \( L_G := B^T B \). Show that \( L_G = D - A \), where \( D \) is a diagonal matrix of vertex degrees, and \( A \) is the adjacency matrix of the graph \( G \).

(e) Given graph \( G \), consider \( m \) graphs, one per edge: graph \( G_e \) has \( n \) nodes but just a single edge \( e \). Let \( L_e \) be the Laplacian of this graph. Write down the expression for \( x^T L_e x \), and show it is non-negative.

(f) Show that if \( A \) and \( B \) are PSD, then \( aA + bB \) is also PSD when \( a, b \geq 0 \). Here \( aA \) multiplies every entry of \( A \) by the scalar \( a \), etc. Show that \( L_G = \sum_e L_e \), and hence get another proof of the fact that \( L_G \) is PSD.

(g) Finally, use your answer from (3e) to write down the expression for \( x^T L_G x \).

4. (Some Singular Facts.) Recall that we defined the Frobenius norm of a matrix \( A \) to be \( \| A \|_F := \sqrt{\sum_{i,j} A_{ij}^2} \). The spectral norm \( \| A \|_2 \) is defined as \( \sigma_1 \), the largest singular value of \( A \).

(a) For any matrix \( A \), suppose its singular values are \( \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0 \). Show that \( \| A \|_2^2 = \text{Trace}(A^T A) = \sum_i \sigma_i^2 \), and hence for any \( k \),

\[
\sigma_k \leq \frac{\| A \|_F}{\sqrt{k}}.
\]
(b) Given a matrix $A$ as in part (a) above, show that for any integer $k$, there exists another matrix $B$ of rank at most $k$ such that

$$
\|A - B\|_2 \leq \frac{\|A\|_F}{\sqrt{k}}.
$$

(c) Let $A$ be the identity matrix $I$. Then give a value of $k$ where part (b) does not hold for the Frobenius norm: that is, there is no $B$ of rank at most $k$ such that

$$
\|A - B\|_F \leq \frac{\|A\|_F}{\sqrt{k}}.
$$

(d) Suppose $A$ is a square matrix with SVD $A = UDV^\top$, and hence $A = \sum_{i=1}^{r} \sigma_i u_i v_i^\top$ where $r$ is the rank of $A$. Define $B := \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^\top$ and $B := \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^\top$. Show that for all vectors $x$ in the column span of $V$, we have $B A x = x$. Hence $B$ is called the pseudo-inverse of $A$. 
