Homework 7

Bryan Lee
922649673

#1

Let $A = P\Sigma Q^T$ & b be a column vector where $b \in \mathbb{R}^n$ such that $Ax = b$. Since we know that A is a nonsingular $n \times n$ matrix, then it has a full rank : $\text{rank}(A) = n$. This means that P is an $n \times n$ orthogonal matrix, Σ is a diagonal matrix of singular values $\sigma_1 \geq \ldots \geq \sigma_n > 0$, & Q is an $n \times n$ orthogonal matrix.

To solve for the system of linear equations $Ax = b$, we can use substitution:

$$Ax = b \iff (P\Sigma Q^T)x = b$$

Then, with respect to change of basis:

$$P^T(P\Sigma Q^T)x = P^T b \qquad \leftarrow P^TP = I_n \text{ as } P \text{ is orthogonal matrix}$$

$$\Sigma Q^T x = P^T b$$

$$\Sigma^{-1}\Sigma Q^T x = \Sigma^{-1} P^T b \qquad \leftarrow \Sigma^{-1} \text{ has reciprocal singular values } \therefore \Sigma^{-1}\Sigma = I_n$$

$$Q^T x = \Sigma^{-1} P^T b$$

$$QQ^T x = Q\Sigma^{-1} P^T b \qquad \leftarrow QQ^T = I_n \text{ as } Q \text{ is orthogonal matrix}$$

$$x = (Q\Sigma^{-1}P^T)b$$

Why is A being non-singular important in computing $Ax = b$ via SVD? A needs to be nonsingular as it requires the need for a "n" set of singular values to exist. If $A_{m \times n}$ is singular then $\text{rank}(A) = r$ where $r \leq m$ and so we would only have "r", where $r < n$ set of singular values. This means that Σ has the dimension $r \times r$ & $\Sigma^{-1}\Sigma = I_r$ rather than the necessary I_n for the correct dimension to perform matrix multiplication with the orthogonal matrices.

There could also be a 0 singular value if A is nonsingular $\therefore$ making Σ singular & not invertible.

#2

For a non-singular matrix A with real entries, the singular values of A & the singular values of A^{-1} is reciprocally related, i.e

$\sigma_1, \dots, \sigma_n$ are singular values of A

Then

$\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_n}$ are singular values of A^{-1}.

We can show this by taking the inverse of A such that A^{-1}.

Let $A = P\Sigma Q^T$ where importantly $\Sigma = diag(\sigma_1, \dots, \sigma_n)$

Then:

$$A^{-1} = (P\Sigma Q^T)^{-1} = Q\Sigma^{-1}P^T$$ since P & Q are orthogonal, so $P^{-1} = P^T$ & $Q^{-1} = Q^T$

However, taking the inverse of Σ, its singular values would then be:

$$\Sigma^{-1} = diag\left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_n}\right).$$

Furthermore, as shown in problem #1, this case is only true for non-singular matrix A, as it requires all $\sigma_i > 0$.

#3

a) Let B be a $p \times r$ matrix; Let C be a $r \times p$ matrix.

The trace of a matrix can be defined as the sum of all diagonal elements of a matrix such that $\sum_{i=1}^{n} a_{ii}$.

Now, if we compute the product of BC, we would get a $p \times p$ square matrix and so the trace can be define as:

$$trace(BC) = \sum_{i=1}^{p} (BC)_{ii} = \sum_{i=1}^{p}\sum_{j=1}^{r} B_{ij}C_{ji}$$

& similarly for the product of CB, we would get a $r \times r$ square matrix, and so the trace can be define as:

$$trace(CB) = \sum_{i=1}^{r} (CB)_{ii} = \sum_{j=1}^{r}\sum_{i=1}^{p} C_{ji}B_{ij}$$

which shows that indeed $trace(BC) = \sum_{i=1}^{p}\sum_{j=1}^{r} B_{ij}C_{ji} = trace(CB) = \sum_{j=1}^{r}\sum_{i=1}^{p} C_{ji}B_{ij}$ as per commutative property.

b) Since $||A||$ is define as $\sqrt{\sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^2}$ then its squared is:

$$||A||^2 = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^2$$

Now, since the trace of a matrix is the trace of a matrix, we can define it as $\text{trace}(A) = \sum_{i=1}^{n} a_{ii}$.

We know that AA^T is an $m \times m$ matrix where its trace can be denoted as:

$$\text{trace}(AA^T) = \sum_{i=1}^{m} (AA^T)_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^2$$

& similarly, A^TA is an $n \times n$ matrix where its trace can be denoted as:

$$\text{trace}(A^TA) = \sum_{j=1}^{n} (A^TA)_{jj} = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^2$$

Note the change in variable

$$\therefore\ ||A||^2 = \text{trace}(AA^T) = \text{trace}(A^TA)$$

c) Given that U is an $m \times m$ orthogonal matrix, then $U^TU = UU^T = I_n$

Now, using the definition of $||A||^2 = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^2$, I have also shown that $||A||^2 = \text{trace}(AA^T) = \text{trace}(A^TA)$ in the previous part 3b.

So,

$$||UA||^2 = \text{trace}((UA)^T(UA)) = \text{trace}(A^T \underbrace{U^TU}_{I_m \text{ as } U \text{ is orthogonal}} A^T) = \text{trace}(A^TA)$$

And by the definition of $||A||^2$:

$$||UA||^2 = ||A||^2$$

$$\therefore\ ||UA|| = ||A||$$

d) Let $A = P\Sigma Q^T$ where P & Q are orthogonal matrices & $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$, the diagonal matrix of singular values.

Now, using the definition of $\|A\|^2 = \text{trace}(AA^T) = \text{trace}(A^TA)$ that I have shown in Part 3b; we can substitute $A = P\Sigma Q^T$:

$$AA^T = (P\Sigma Q^T)(P\Sigma Q^T)^T = P\Sigma \underbrace{Q^T Q}_{I_n \text{ as } Q \text{ is orthogonal}} \Sigma^T P^T$$

$$= P\Sigma\Sigma^T P^T = P\Sigma^2 P^T$$

$\Sigma^T = \Sigma$ since Σ is a diagonal matrix

So, $\|A\|^2 = \text{trace}(AA^T) = \text{trace}(P\Sigma^2 P^T)$ & by the cyclic property of trace, we can rewrite $\|A\|^2$ as:

$$\|A\|^2 = \text{trace}(AA^T) = \text{trace}(PP^T\Sigma^2) = \text{trace}(\Sigma^2)$$ since P is an orthogonal matrix $\therefore PP^T = P^TP = I_n$

where:

$$\text{trace}(\Sigma^2) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2$$

$$\therefore \|A\| = \sqrt{\text{trace}(\Sigma^2)} = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2}$$

#4

a) A is an $m \times n$ matrix with real entries, such that $A = P\Sigma Q^T$ be its SVD.

Given that $p_1, p_2, \ldots, p_r$ are the column vectors of P & $q_1, q_2, \ldots, q_r$ are the columns of Q, then we can simply express the product of the matrices multiplication as:

$$A = \sum_{i=1}^{r} \sigma_i p_i q_i^T$$

where σ_i is the i-th singular value (scalar ∴ commuted to front),
p_i is the i-th column vector of P,
q_i^T is the i-th column of Q^T

Note: $\sigma_1, \sigma_2, \ldots, \sigma_r > 0$

∴ we can expand it out as:

$$A = \sigma_1 p_1 q_1^T + \sigma_2 p_2 q_2^T + \ldots + \sigma_r p_r q_r^T$$ where each $\sigma_i p_i q_i^T$ is an $m \times n$ rank 1 matrix.

b) Let $A = P\Sigma Q^T$ & $A_k = P_k \Sigma_k Q_k^T$ (the truncated SVD)

We can then express $A - A_k$ as:

$$A - A_k = \sum_{i=1}^{r} \sigma_i p_i q_i^T - \sum_{i=1}^{k} \sigma_i p_i q_i^T$$ where r represents the rank of A.

This summation expansion is shown in part 4a. This also means that the difference can be expressed as:

$$A - A_k = \sum_{i=k+1}^{r} \sigma_i p_i q_i^T$$

We can now use the definition of $\|A\|^2$ which I've defined in part 3d,

$$\|A\|^2 = \operatorname{trace}(AA^T) = \operatorname{trace}(A^TA)$$

we can then express $\|A - A_k\|^2$ as:

$$\|A - A_k\|^2 = \operatorname{trace}\left((P_{k+1}\Sigma_{k+1}Q_{k+1}^T)(P_{k+1}\Sigma_{k+1}Q_{k+1}^T)^T\right)$$

$$= \operatorname{trace}\left(P_{k+1}\Sigma_{k+1}Q_{k+1}^T Q_{k+1}\Sigma_{k+1}^T P_{k+1}^T\right)$$

using orthogonal matrix property & trace cyclic property, we can reduce down to:

$$\|A - A_k\|^2 = \operatorname{trace}(\Sigma_{k+1}\Sigma_{k+1}^T) = \operatorname{trace}(\Sigma_{k+1}^2) = \sigma_{k+1}^2 + \sigma_{k+2}^2 + \ldots + \sigma_r^2$$

$$\therefore \|A - A_k\| = \sqrt{\sigma_{k+1}^2 + \sigma_{k+2}^2 + \ldots + \sigma_r^2}$$

#5

In MATLAB file. for part a & b

c) By looking at the singular values, I could almost a gook k value. However, it still required a bit of trial & error to reconstruct an approximated image. For example, the SVD did show a drop from 10.06 to 8.3898 but the image was still unrecognizable. The other good singular values to choose from were the ones that started to stagnated between the range 0 to 3, which contains the ideal k value. Choosing the k singular value at the start of the stagnation was the best choice & led me to choosing k = 15.