CS663 Assignment-4

Saksham Rathi, Kavya Gupta, Shravan Srinivasa Raghavan
Department of Computer Science,
Indian Institute of Technology Bombay

Question 3

Solution

Let A be a real $m \times n$ matrix. A can always be expressed as $A = U\Sigma V^T$ where $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ and $\Sigma \in \mathbb{R}^{m\times n}$, U, V being orthogonal matrices and Σ being a diagonal matrix with non negative values (called singular values) on the diagonal. Let $\Sigma = diag(\sigma_1, \sigma_2, \dots, \sigma_{\min(m,n)})$. We will show that the squares of the non-zero singular values of A are the eigenvalues of either AA^T or A^TA. This is equivalent to showing that the the non-zero singular values of A are the squareroots of either AA^T or A^TA because, the singular values are non-negative by definition.

Part a

We will show that the non zero singular values of A are equal to the square roots of the eigen values of either AA^T or A^TA. We define $\Sigma_m^2 = \Sigma\Sigma^T$ and $\Sigma_n^2 = \Sigma^T\Sigma$.

$$\begin{aligned}
AA^T &= (U\Sigma V^T) \cdot (U\Sigma V^T)^T \\
&= U\Sigma V^T \cdot (V\Sigma^T U^T) \\
&= U\Sigma V^T V\Sigma^T U^T \\
&= U\Sigma(V^T V)\Sigma^T U^T \\
&= U\Sigma I_n \Sigma^T U^T \\
&= U\Sigma\Sigma^T U^T \\
&= U\Sigma_m^2 U^T
\end{aligned}$$

Similarly, $A^TA = V\Sigma^T\Sigma V^T = V\Sigma_n^2 V^T$.

Clearly, $\Sigma_m^2 \in R^{m\times m}$ and equals $diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2) = D$. Now $m = \min(m,n)$ or $n = \min(m,n)$.

Let $m = \min(m,n)$. In this case, Σ_m^2 is a diagonal matrix (D) whose entries are squares of the singular values of A. Since AA^T is a real symmetric matrix, by **spectral theorem** it has an orthogonal decomposition given by $AA^T = \mathcal{U}\mathcal{D}\mathcal{U}^T$ where $\mathcal{D}$ is a diagonal matrix whose entries are the eigenvalues of AA^T and $\mathcal{U}$ is an orthogonal matrix. Therefore the non zero entries of D are the eigenvalues of AA^T. Since $D = \Sigma_m^2$, the non zero entries of D are the squares of the non zero singular values of A and the squares of the any non-zero singular value of A is an eigenvalue of AA^T, we are done.

Let $n = \min(m,n)$. In this case, we deal with the diagonal matrix $D = \Sigma_n^2$ whose entries are the squares of all the singular values of A (this is because $n = \min(m,n)$). The proof is very similar to the case above. Since A^TA is a real symmetric matrix, it has an orthogonal decomposition $\mathcal{V}\mathcal{D}\mathcal{V}^T$ and therefore we have the non-zero entries of D to be the eigenvalues of A^TA and following a similar argument we arrive at the conclusion that the square of any non-zero singular value of A is an eigen value of A^TA and that any eigenvalue of A^TA is the square of some non-zero singular value of A. This completes the proof.

Part b

The Frobenius norm of a matrix $A \in \mathbb{R}^{m\times n}$ is defined as

$$||A||_F = \sqrt{\left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2\right)}$$

For any matrix $A \in \mathbb{R}^{m\times n}$ we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2 = \text{Tr}(AA^T) = \text{Tr}(A^TA)$$

where $\text{Tr}(M) = \sum_{i=1}^{n} M_{ii}$ for any $n \times n$ square matrix M. The trace of a matrix also has the following property,

$$\text{Tr}(AB) = \text{Tr}(BA)$$

whenever AB and BA are both square matrices for two matrices A and B.

WLOG we take $m = \min(n,m)$ (in the other case we simply deal with A^TA and Σ_n^2)

$$\begin{aligned}
||A||_F^2 &= \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2 \\
&= \text{Tr}(AA^T) \\
&= \text{Tr}((U\Sigma V^T V\Sigma^T U^T)) \\
&= \text{Tr}(U\Sigma_m^2 U^T) \\
&= \text{Tr}((U\Sigma_m^2)(U^T)) \\
&= \text{Tr}((U^T)(U\Sigma_m^2)) \\
&= \text{Tr}(U^TU\Sigma_m^2) \\
&= \text{Tr}(I_m\Sigma_m^2) \\
&= \text{Tr}(\Sigma_m^2) \\
&= \sum_{i=1}^{k} \sigma_i^2 \text{ (where k is the number of non-zero singular values of A)} \\
\Rightarrow ||A||_F^2 &= \sum_{i=1}^{k} \sigma_i^2
\end{aligned}$$

Thus we have the square of the Frobenius norm of a matrix to equal the sum of squares of the singular values.

Part c

Part d

Given:

$$A \in \mathbb{R}^{m \times n}, m \leqslant n$$
$$P = A^T A$$
$$Q = AA^T$$

A few results and definitions we will use:

Definition 1. *A matrix $M \in \mathbb{R}^{n \times n}$ is said to be positive semi-definite if $\forall \vec{x} \in R^n$,*

$$\vec{v}^T M \vec{v} \geqslant 0$$

Theorem 1 (Spectral Theorem)**.** *Any real symmetric matrix $A \in \mathbb{R}^{n \times n}$ is orthogonally diagonalizable. That is there exists real numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$, $U \in \mathbb{R}^{n \times n}$ such that $A = UDU^T$ where $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$, $U = [\vec{u}_1 \vec{u}_2 \cdots \vec{u}_n]$ with $A\vec{u}_i = \lambda_i \vec{u}_i \, \forall i \in [1, n]$ and $UU^T = U^T U = I_n$*

Lemma 1. *P and Q are positive semi-definite (defintion 1) and their eigenvalues are non-negative.*

Proof. We will first show that P is positive semi-definite and then we will show that the eigen values of P are all non-negative. The proof for Q is similar.

Let $\vec{x} \in \mathbb{R}^n$.

$$\begin{aligned}
\vec{x}^T P \vec{x} &= \vec{x}^T (A^T A) \vec{x} \\
&= \vec{x}^T A^T A \vec{x} \\
&= (\vec{x}^T A^T)(A\vec{x}) \\
&= (A\vec{x})^T (A\vec{x}) \\
&= \vec{y}^T \vec{y} \text{ where } \vec{y} \in \mathbb{R}^n \; \vec{y} = A\vec{x} \\
&= ||\vec{y}||^2 \geqslant 0 \\
\Rightarrow \vec{x}^T P \vec{x} &\geqslant 0
\end{aligned} \tag{1}$$

Let $\vec{u} \in R^n / \{\vec{0}\}$ be an eigenvector of P with corresponding eigenvalue $\lambda \in \mathbb{C}$. Consider the expression $\vec{u}^T P \vec{u}$. From equation 1 we know that $\vec{u}^T P \vec{u} \geqslant 0$. Therefore we have,

$$\begin{aligned}
\vec{u}^T P \vec{u} = \vec{u}^T (P\vec{u}) \geqslant 0 \\
= \vec{u}^T (\lambda \vec{u}) \geqslant 0 \\
= \lambda \vec{u}^T \vec{u} \geqslant 0 \\
= \lambda ||\vec{u}||^2 \geqslant 0 \\
\Rightarrow \lambda ||\vec{u}||^2 \geqslant 0 \\
\Rightarrow \lambda \geqslant 0
\end{aligned} \tag{2}$$

Therefore we have eigenvalues of P to be non-negative. The proofs hold for Q as well, just replace $\vec{x}$ in equation 1 by $\vec{y} \in \mathbb{R}^m$ and $\vec{u}$ in equation 2 by $\vec{v} \in \mathbb{R}^m$. $\square$

Lemma 2. *The following are true regarding the eigenvectors of P and Q:*

1. *If $\vec{u} \in \mathbb{R}^n$ is an eigenvector of P with eigenvalue λ then $A\vec{u}$ is an eigenvector of Q with eigenvalue λ.*

2. *If $\vec{v} \in \mathbb{R}^m$ is an eigenvector of Q with eigenvalue μ, then $A^T\vec{v}$ is an eigenvector of P with eigenvalue μ.*

Proof. Since $\vec{u} \in \mathbb{R}^n$ is an eigenvector of P with eigenvalue λ we have,

$$\begin{aligned}
P\vec{u} &= \lambda\vec{u} \\
\Rightarrow P\vec{u} &= (A^TA)\vec{u} = \lambda\vec{u} \\
\Rightarrow (A^T)(A\vec{u}) &= \lambda\vec{u} \\
\Rightarrow A\left(A^T(A\vec{u})\right) &= A\left(\lambda\vec{u}\right) \text{ (pre-multiplying both LHS and RHS by a non-null matrix } A\text{)} \\
\Rightarrow (AA^T)(A\vec{u}) &= \lambda(A\vec{u}) \\
\Rightarrow (Q)(A\vec{u}) &= \lambda(A\vec{u}) \\
\Rightarrow Q(A\vec{u}) &= \lambda(A\vec{u}) \qquad (3)
\end{aligned}$$

From equation 3 we have that for every $\vec{u}$ that is eigenvector of P, the vector $A\vec{u}$ is an eigenvector of Q with the same eigenvalue. Now we prove the second part of the lemma whose proof is very similar to what we saw above.

Since $\vec{v} \in \mathbb{R}^m$ is an eigenvector of Q with eigenvalue μ we have,

$$\begin{aligned}
Q\vec{v} &= \mu\vec{v} \\
\Rightarrow Q\vec{v} &= (AA^T)\vec{v} = \mu\vec{v} \\
\Rightarrow (A)(A^T\vec{v}) &= \mu\vec{v} \\
\Rightarrow A^T\left(A(A^T\vec{v})\right) &= A^T\left(\mu\vec{v}\right) \text{ (pre-multiplying both LHS and RHS by a non-null matrix } A^T\text{)} \\
\Rightarrow (A^TA)(A^T\vec{v}) &= \mu(A^T\vec{v}) \\
\Rightarrow (P)(A^T\vec{v}) &= \mu(A^T\vec{v}) \\
\Rightarrow P(A^T\vec{v}) &= \mu(A^T\vec{v}) \qquad (4)
\end{aligned}$$

Equations 3 and 4 complete the proof. □

Lemma 3. *Let $\vec{v}_i \in \mathbb{R}^m$ be an eigenvector of Q. Define $u_i \triangleq \frac{A^T\vec{v}_i}{||A^T\vec{v}_i||_2}$. There exists $\gamma_i \geqslant 0$ such that $A\vec{u}_i = \gamma_i\vec{v}_i$*

Proof. Let $\vec{v}_i \in \mathbb{R}^m$ be an eigenvector of Q with an eigenvalue μ. From lemma 1 we know that eigenvalues of P and Q are non-negative, therefore we have $\mu_i \geqslant 0$. Consider the expression $A\vec{u}_i$,

$$\begin{aligned}
A\vec{u}_i &= A\left(\frac{A^T\vec{v}_i}{||A^T\vec{v}_i||_2}\right) \\
&= \frac{A(A^T\vec{v}_i)}{||A^T\vec{v}_i||_2} \\
&= \frac{(AA^T)\vec{v}_i}{||A^T\vec{v}_i||_2} \\
&= \frac{Q\vec{v}_i}{||A^T\vec{v}_i||_2} \\
&= \frac{\mu_i\vec{v}_i}{||A^T\vec{v}_i||_2} \\
&= \left(\frac{\mu_i}{||A^T\vec{v}_i||_2}\right)\vec{v}_i \\
&= \gamma_i\vec{v}_i
\end{aligned} \tag{5}$$

That is, there exists $\gamma_i = \left(\frac{\mu_i}{||A^T\vec{v}_i||_2}\right) \geqslant 0$ such that $A\vec{u}_i = \gamma_i\vec{v}_i$ □

Theorem 2 (Singular Value Decomposition). *Let $\vec{v}_i$,$\vec{u}_i$ and γ_i be defined as in lemma 3. Let $U = [\vec{v}_1|\vec{v}_2|\cdots|\vec{v}_m]$, $V = [\vec{u}_1|\vec{u}_2|\cdots|\vec{u}_m]$ and $\Gamma = diag(\gamma_1, \gamma_2, \ldots, \gamma_m)$ where U and Γ are both $m \times m$ matrices and V is an $n \times m$ matrix. Then*

$$A = U\Gamma V^T$$

Proof. Two matrices A and B are equal only if they the same dimensions and the values at the corresponding indices for every pair of (i, j) are equal ie, $\forall i \in [1, m], \forall j \in [1, n]$, if $A_{ij} = B_{ij}$ then $A = B$ and vice versa.

As P and Q are real symmetric matrices we can obtain an orthonormal basis for $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively that are the eigenvectors of P and Q respectively (theorem 1).

Since the columns of U and V are the orthonormal eigenvectors of Q and P respectively, we have $UU^T = U^TU = I_m$ and $V^TV = I_m$.

Consider the expression AV.

$$\begin{aligned}
AV &= A\left[\vec{u}_1|\vec{u}_2|\cdots|\vec{u}_m\right] \\
\Rightarrow AV &= \left[A\vec{u}_1|A\vec{u}_2|\cdots|A\vec{u}_m\right] \\
&= \left[\gamma_1\vec{v}_1|\gamma_2\vec{v}_2|\cdots|\gamma_m\vec{v}_m\right] \text{ from lemma 3} \\
&= \left[\vec{v}_1|\vec{v}_2|\cdots|\vec{v}_m\right]\Gamma \\
&= U\Gamma \\
\Rightarrow AV &= U\Gamma \\
\Rightarrow AVV^T &= \text{ if this is somehow equal to } I_n \text{ then we are done.}
\end{aligned} \tag{6}$$

□