Shengjie Li
RUID: 188008047

Homework 1

Problem 1: Consider the matrix
$$A = \begin{bmatrix} 1 & 0.5 \\ 0 & 1+\epsilon \end{bmatrix}.$$

a) Find the eigenvalues/eigenvectors of A assuming $\epsilon \neq 0$. Force your eigenvectors to have unit norm.

$$A - \lambda I = \begin{bmatrix} 1-\lambda & 0.5 \\ 0 & 1+\epsilon-\lambda \end{bmatrix}$$
$$\begin{aligned} \det(A-\lambda I) &= (1-\lambda)*(1+\epsilon-\lambda) - 0.5*0 \\ &= (1-\lambda)*(1+\epsilon-\lambda) \end{aligned}$$

Let $\det(A - \lambda I) = 0$ we can get eigenvalues $\lambda_1 = 1, \lambda_2 = 1+\epsilon$.
For $\lambda = 1$, solve $A\vec{\boldsymbol{x}} = \lambda\vec{\boldsymbol{x}}$:

$$\begin{bmatrix} 1 & 0.5 \\ 0 & 1+\epsilon \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
$$\downarrow$$
$$\begin{aligned} x_1 + 0.5x_2 &= x_1 \\ (1+\epsilon)x_2 &= x_2 \end{aligned}$$
$$\downarrow$$
$$\begin{aligned} x_1 &= x_1 (x_1 \neq 0) \\ x_2 &= 0 \end{aligned}$$

Therefore, $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is a eigenvector of A associated with the eigenvalue $\lambda = 1$.
For $\lambda = 1+\epsilon$, solve $A\vec{\boldsymbol{x}} = \lambda\vec{\boldsymbol{x}}$:

$$\begin{bmatrix} 1 & 0.5 \\ 0 & 1+\epsilon \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (1+\epsilon) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
$$\downarrow$$
$$\begin{aligned} x_1 + 0.5x_2 &= (1+\epsilon)x_1 \\ (1+\epsilon)x_2 &= (1+\epsilon)x_2 \end{aligned}$$
$$\downarrow$$
$$\begin{aligned} x_1 &= \frac{x_2}{2\epsilon} \\ x_2 &= x_2 \end{aligned}$$

Therefore, $\begin{bmatrix} \frac{1}{\sqrt{1+4\epsilon^2}} \\ \frac{2\epsilon}{\sqrt{1+4\epsilon^2}} \end{bmatrix}$ is a eigenvector of A associated with the eigenvalue $\lambda = 1+\epsilon$.

b) Diagonalize A using the eigenvalues/eigenvectors you computed.

Let T be the matrix with eigenvectors as its columns.

$$
\begin{aligned}
T &= \begin{bmatrix} 1 & \frac{1}{\sqrt{1+4\epsilon^2}} \\ 0 & \frac{2\epsilon}{\sqrt{1+4\epsilon^2}} \end{bmatrix} \\
T^{-1} &= \frac{1}{\frac{2\epsilon}{\sqrt{1+4\epsilon^2}}} \begin{bmatrix} \frac{2\epsilon}{\sqrt{1+4\epsilon^2}} & -\frac{1}{\sqrt{1+4\epsilon^2}} \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & -\frac{1}{2\epsilon} \\ 0 & \frac{\sqrt{1+4\epsilon^2}}{2\epsilon} \end{bmatrix} \\
\Lambda = T^{-1}AT &= \begin{bmatrix} 1 & -\frac{1}{2\epsilon} \\ 0 & \frac{\sqrt{1+4\epsilon^2}}{2\epsilon} \end{bmatrix} \begin{bmatrix} 1 & 0.5 \\ 0 & 1+\epsilon \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{\sqrt{1+4\epsilon^2}} \\ 0 & \frac{2\epsilon}{\sqrt{1+4\epsilon^2}} \end{bmatrix} \\
&= \begin{bmatrix} 1 & -\frac{1}{2\epsilon} \\ 0 & \frac{(1+\epsilon)\sqrt{1+4\epsilon^2}}{2\epsilon} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{\sqrt{1+4\epsilon^2}} \\ 0 & \frac{2\epsilon}{\sqrt{1+4\epsilon^2}} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1+\epsilon \end{bmatrix}
\end{aligned}
$$

c) Start now sending $\epsilon \to 0$. What do you observe is happening to the matrices you use for diagonalization as ϵ becomes smaller and smaller? So what do you conclude when $\epsilon = 0$?

When ϵ is becoming closer to 0, the determinant of matrix T is becoming closer to 0.
Thus, when $\epsilon = 0$, matrix T will become non-invertible. Matrix A will become non-diagonalizable.

Problem 2: Let A, B be two matrices of the same dimensions k × m.

a) With direct computation show that $trace(AB^T) = trace(B^TA) = trace(BA^T) = trace(A^TB)$.

The $i-th$ element in the diagonal of matrix AB^T is $\sum\limits_{j=1}^{m} a_{ij}b_{ij}$. Thus, $trace(AB^T) = \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{m} a_{ij}b_{ij}$.

The $j-th$ element in the diagonal of matrix B^TA is $\sum\limits_{i=1}^{k} b_{ij}a_{ij}$. Thus, $trace(B^TA) = \sum\limits_{j=1}^{m}\sum\limits_{i=1}^{k} b_{ij}a_{ij} = \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{m} a_{ij}b_{ij}$.

The $i-th$ element in the diagonal of matrix BA^T is $\sum\limits_{j=1}^{m} b_{ij}a_{ij}$. Thus, $trace(BA^T) = \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{m} b_{ij}a_{ij} = \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{m} a_{ij}b_{ij}$.

The $j-th$ element in the diagonal of matrix A^TB is $\sum\limits_{i=1}^{k} a_{ij}b_{ij}$. Thus, $trace(A^TB) = \sum\limits_{j=1}^{m}\sum\limits_{i=1}^{k} a_{ij}b_{ij} = \sum\limits_{i=1}^{k}\sum\limits_{j=1}^{m} a_{ij}b_{ij}$.

Therefore, $trace(AB^T) = trace(B^TA) = trace(BA^T) = trace(A^TB)$.

b) Use question a) to compute $\mathsf{E}[\mathbf{x}^T A\mathbf{x}]$ where $\mathsf{E}[\cdot]$ denotes expectation, A is a constant matrix and $\mathbf{x}$ is a random vector for which we know that $\mathsf{E}[\mathbf{x}\mathbf{x}^T] = Q$. *Hint: The trace of a scalar is the scalar itself.*

Because $\mathbf{x}^T A\mathbf{x}$ is a scalar, Therefore:

$$\begin{aligned}
\mathbf{x}^T A\mathbf{x} &= trace(\mathbf{x}^T A\mathbf{x}) \\
\mathsf{E}[\mathbf{x}^T A\mathbf{x}] &= \mathsf{E}[trace(\mathbf{x}^T A\mathbf{x})] \\
&= \mathsf{E}[trace(A\mathbf{x}\mathbf{x}^T)] \\
&= trace(\mathsf{E}[A\mathbf{x}\mathbf{x}^T])
\end{aligned}$$

Because A is constant, Therefore:

$$\begin{aligned}
\mathsf{E}[\mathbf{x}^T A\mathbf{x}] &= trace(A \cdot \mathsf{E}[\mathbf{x}\mathbf{x}^T]) \\
&= trace(A \cdot Q)
\end{aligned}$$

c) Using the previous properties show that for any matrix A of dimensions $k \times k$ we have $trace(A) = trace(UAU^{-1})$ for any nonsingular matrix U of dimensions $k \times k$. In other words that the trace does not change if we apply a similarity transformation.

$$trace(UAU^{-1}) = trace(AU^{-1}U) = trace(AI) = trace(A)$$

d) Use question c) to prove that if matrix A of dimensions $k \times k$ is diagonalizable then its trace is equal to the sum of its eigenvalues (actually this is true even if the matrix is not diagonalizable).

Because matrix A is diagonalizable, then there exists a matrix T such that:

$$\Lambda = T^{-1}AT$$

$$trace(\Lambda) = trace(T^{-1}AT) = trace(A)$$

Notice that Λ is a diagonal matrix with all eigenvalues of A as its diagonal entries. Thus,

$$trace(\Lambda) = \text{sum of eigenvalues of } A.$$

e) Regarding question d) how do you explain this equality given that when A is real the trace is also real whereas the eigenvalues can be complex?

Because the complex eigenvalues of A will always come in complex conjugate pairs. Therefore, when we sum them up, we will always get real numbers.

f) Using again question d) what can you say about the coefficient c_{k-1} of the characteristic polynomial $\lambda^k + c_{k-1}\lambda^{k-1} + \cdots + c_0$ of A. We recall that we already know that $c_0 = (-1)^k \lambda_1 \cdots \lambda_k = (-1)^k \det(A)$.

$$\begin{aligned}
\det(\lambda I - A) &= \lambda^k + c_{k-1}\lambda^{k-1} + \cdots + c_0 \\
&= (\lambda - \lambda_1)(\lambda - \lambda_2) \ldots (\lambda - \lambda_k) \\
&= \lambda^k - \lambda^{k-1}(\lambda_1 + \lambda_2 + \cdots + \lambda_k) + \cdots \\
&= \lambda^k - trace(A)\lambda^{k-1} + \cdots
\end{aligned}$$

Therefore,

$$c_{k-1} = -trace(A)$$

Problem 3: A matrix A is called *nilpotent* if $A^r = 0$ for some integer $r > 1$.

a) Show that all eigenvalues of A must be equal to 0.

Take $\vec{x}$ to be an eigenvector of A associated with eigenvalue λ, then

$$\begin{aligned} A\vec{x} &= \lambda\vec{x} \\ A^2\vec{x} &= A(A\vec{x}) = A(\lambda\vec{x}) = \lambda^2\vec{x} \\ &\vdots \\ A^r\vec{x} &= \lambda^r\vec{x} \end{aligned}$$

From $A^r = 0$ we can get $\lambda^r = 0$. Thus, $\lambda = 0$. Note that all eigenvalues of A apply to the above equation. Therefore, all eigenvalues of A must be equal to 0.

b) Is such a matrix diagonalizable?

Proof. Assume matrix A is diagonalizable, then there exits an invertible matrix T such that

$$A = T^{-1}\Lambda T.$$

The diagonal entries of matrix Λ are the eigenvalues of A. Since all eigenvalues of A are equal to 0, $\Lambda = 0$.
Thus, the right side of the equation equals to 0, while the left side does not.
Therefore, no matrix T can satisfy this equation. This contradicts the assumption that there exits an invertible matrix T such that $A = T^{-1}\Lambda T$.
By contradiction, matrix A is non-diagonalizable. □

c) Give an example of a 2×2 matrix which is *nilpotent*.

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Therefore,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \text{is } \textit{nilpotent}.$$

d) If we apply a similarity transformation to a *nilpotent*, is the resulting matrix *nilpotent*? Apply the previous observation to your example in c) to obtain a second example of a *nilpotent* matrix.

$$\begin{aligned} B &= T^{-1}AT \\ B^2 &= T^{-1}ATT^{-1}AT = T^{-1}A^2T \\ &\vdots \\ B^r &= T^{-1}A^rT \\ B^r &= 0 \end{aligned}$$

Thus, if we apply a similarity transformation to a *nilpotent* A, the resulting matrix B is *nilpotent*.

$$\begin{aligned}
A &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \\
T &= \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, T^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \\
B &= T^{-1}AT \\
&= \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} \\
B^2 &= \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}
\end{aligned}$$

Thus, after rotating matrix $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ 90 degrees counterclockwisely, matrix $B = \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}$ is *nilpotent* for r = 2.

Problem 4: Let Q be symmetric and positive definite with dimensions $k \times k$. Denote with Q_i the matrix we obtain from Q by eliminating its ith column and ith row.

a) Show that Q_i is also symmetric and positive definite. Extend this to when we eliminate more than one columns and rows.

Proof. Because Q is symmetric and positive definite.
$Q_{xy} = Q_{yx}$ for $x, y \in [1, k]$.
Thus, $Q_{xy} = Q_{yx}$ is also true for $x, y \in [1, i) \cup (i, k]$.
Therefore, Q_i is symmetric.

For any non-zero $k \times 1$ vector $\vec{\boldsymbol{x}}$, we have $\vec{\boldsymbol{x}}^T Q \vec{\boldsymbol{x}} > 0$.
Let $\hat{\vec{\boldsymbol{x}}}$ be a vector we obtain from $\vec{\boldsymbol{x}}$ by eliminating its ith element.
Let $\hat{\hat{\vec{\boldsymbol{x}}}} = \vec{\boldsymbol{x}}$ except that $\hat{\hat{\vec{\boldsymbol{x}}}}_i = 0$, then we get $\hat{\vec{\boldsymbol{x}}}^T Q_i \hat{\vec{\boldsymbol{x}}} = \hat{\hat{\vec{\boldsymbol{x}}}}^T Q \hat{\hat{\vec{\boldsymbol{x}}}} > 0$.
Therefore, Q_i is symmetric and positive definite. □

b) Show that the largest eigenvalue of Q is larger than the largest eigenvalue of Q_i and, that the smallest eigenvalue of Q is smaller than the smallest eigenvalue of Q_i .

Proof.

$$\lambda_{max}(Q_i) = \max_{\hat{\vec{\boldsymbol{x}}}} \frac{\hat{\vec{\boldsymbol{x}}}^T Q_i \hat{\vec{\boldsymbol{x}}}}{\hat{\vec{\boldsymbol{x}}}^T \hat{\vec{\boldsymbol{x}}}}$$

$$\frac{\hat{\vec{\boldsymbol{x}}}^T Q_i \hat{\vec{\boldsymbol{x}}}}{\hat{\vec{\boldsymbol{x}}}^T \hat{\vec{\boldsymbol{x}}}} = \frac{\hat{\hat{\vec{\boldsymbol{x}}}}^T Q \hat{\hat{\vec{\boldsymbol{x}}}}}{\hat{\hat{\vec{\boldsymbol{x}}}}^T \hat{\hat{\vec{\boldsymbol{x}}}}} \leq \frac{\vec{\boldsymbol{x}}^T Q \vec{\boldsymbol{x}}}{\vec{\boldsymbol{x}}^T \vec{\boldsymbol{x}}} \qquad (\hat{\hat{\vec{\boldsymbol{x}}}} = \vec{\boldsymbol{x}} \text{ except that } \hat{\hat{\vec{\boldsymbol{x}}}}_i = 0)$$

Therefore,

$$\min_{\hat{\vec{\boldsymbol{x}}}} \frac{\hat{\vec{\boldsymbol{x}}}^T Q_i \hat{\vec{\boldsymbol{x}}}}{\hat{\vec{\boldsymbol{x}}}^T \hat{\vec{\boldsymbol{x}}}} \leq \max_{\hat{\vec{\boldsymbol{x}}}} \frac{\hat{\vec{\boldsymbol{x}}}^T Q_i \hat{\vec{\boldsymbol{x}}}}{\hat{\vec{\boldsymbol{x}}}^T \hat{\vec{\boldsymbol{x}}}} \leq \min_{\vec{\boldsymbol{x}}} \frac{\vec{\boldsymbol{x}}^T Q \vec{\boldsymbol{x}}}{\vec{\boldsymbol{x}}^T \vec{\boldsymbol{x}}} \leq \max_{\vec{\boldsymbol{x}}} \frac{\vec{\boldsymbol{x}}^T Q \vec{\boldsymbol{x}}}{\vec{\boldsymbol{x}}^T \vec{\boldsymbol{x}}}$$

$$\lambda_{min}(Q_i) \leq \lambda_{max}(Q_i) \leq \lambda_{min}(Q) \leq \lambda_{max}(Q)$$

□

c) If A is a square matrix (not necessarily symmetric and positive definite) of dimensions $k \times k$ and $\lambda_1, \ldots, \lambda_k$ are its eigenvalues while $\sigma_1, \ldots, \sigma_k$ are its singular values (obtained by applying SVD) then show

$$\min_i \sigma_i \leq \min_i |\lambda i| \leq \max_i |\lambda i| \leq \max_i \sigma_i.$$

In other words the eigenvalues are located on the complex plane inside an annulus with the two circle radii defined by the largest and smallest singular value. *Use the result we proved in class regarding the bounds of the ratio* $\frac{\mathbf{x}^T C \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$ *for* C *symmetric. Consider that this is true even if* $\mathbf{x}$ *is a vector with complex elements.*

Proof.

We know that:

$$\lambda_{max}(C) \geq \frac{(\vec{\boldsymbol{x}}^*)^T C \vec{\boldsymbol{x}}}{(\vec{\boldsymbol{x}}^*)^T \vec{\boldsymbol{x}}} \geq \lambda_{min}(C)$$

$$\sigma_1^2 \geq \frac{(\vec{\boldsymbol{x}}^*)^T A^T A \vec{\boldsymbol{x}}}{(\vec{\boldsymbol{x}}^*)^T \vec{\boldsymbol{x}}} \geq \sigma_k^2$$

From $A\vec{\boldsymbol{x}}_{\boldsymbol{i}} = \lambda_i \vec{\boldsymbol{x}}_{\boldsymbol{i}}$ and $A(\vec{\boldsymbol{x}}_{\boldsymbol{i}}{}^*) = \lambda_i^* (\vec{\boldsymbol{x}}_{\boldsymbol{i}}{}^*)$ we have:

$$\frac{(\vec{\boldsymbol{x}}_{\boldsymbol{i}}{}^*)^T A^T A \vec{\boldsymbol{x}}_{\boldsymbol{i}}}{(\vec{\boldsymbol{x}}_{\boldsymbol{i}}{}^*)^T \vec{\boldsymbol{x}}_{\boldsymbol{i}}} = \frac{(A\vec{\boldsymbol{x}}_{\boldsymbol{i}}{}^*)^T A \vec{\boldsymbol{x}}_{\boldsymbol{i}}}{(\vec{\boldsymbol{x}}_{\boldsymbol{i}}{}^*)^T \vec{\boldsymbol{x}}_{\boldsymbol{i}}} = \frac{(\lambda_i^* (\vec{\boldsymbol{x}}_{\boldsymbol{i}}{}^*))^T \lambda_i \vec{\boldsymbol{x}}_{\boldsymbol{i}}}{(\vec{\boldsymbol{x}}_{\boldsymbol{i}}{}^*)^T \vec{\boldsymbol{x}}_{\boldsymbol{i}}} = \frac{\lambda_i^* (\vec{\boldsymbol{x}}_{\boldsymbol{i}}{}^*)^T \lambda_i \vec{\boldsymbol{x}}_{\boldsymbol{i}}}{(\vec{\boldsymbol{x}}^*)^T \vec{\boldsymbol{x}}} = \lambda_i^* \lambda_i = |\lambda_i|^2$$

Thus,

$$\sigma_1^2 \geq |\lambda_i|^2 \geq \sigma_k^2.$$

Because σ_k is not negative,

$$\sigma_1 \geq |\lambda_i| \geq \sigma_k.$$

Therefore,

$$\min_i \sigma_i \leq \min_i |\lambda i| \leq \max_i |\lambda i| \leq \max_i \sigma_i.$$

□