ECE 269: Linear Algebra and Applications

Solutions to Homework 4

Fall 2020

1 Problem 1

We have the following relations that define $\mathbf{A}^+$ the pseudoinverse of $\mathbf{A}$, which we summarize here.

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}. \tag{1}$$

$$\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+. \tag{2}$$

$$\mathbf{A}^T(\mathbf{A}^+)^T = \mathbf{A}^+\mathbf{A}. \tag{3}$$

$$(\mathbf{A}^+)^T\mathbf{A}^T = \mathbf{A}\mathbf{A}^+. \tag{4}$$

a) Let $\mathbf{B}$ and $\mathbf{C}$ be pseudoinverses of $\mathbf{A}$. Then, we have:

$$\mathbf{AB} = \mathbf{ACAB} \overset{(4)}{=} \mathbf{C}^T\mathbf{A}^T\mathbf{B}^T\mathbf{A}^T = \mathbf{C}^T(\mathbf{ABA})^T = \mathbf{C}^T\mathbf{A}^T \overset{(4)}{=} \mathbf{AC}.$$

Similarly, we can show that $\mathbf{BA} = \mathbf{CA}$. We therefore have

$$\mathbf{B} = \mathbf{BAB} = \mathbf{CAB} = \mathbf{CAC} = \mathbf{C}.$$

b) Since $\mathbf{A}$ is full-rank and tall, $\mathbf{A}^T\mathbf{A}$ is symmetric and non-singular. Defining $\mathbf{B} := (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$, we see that $\mathbf{B} \in \mathbb{R}^{n\times m}$, and

$$\begin{aligned}
\mathbf{ABA} &= \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A} = \mathbf{A}\\
\mathbf{BAB} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{B}\\
\mathbf{A}^T\mathbf{B}^T &= \mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1} = \mathbf{I}_n = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A} = \mathbf{BA}\\
\mathbf{B}^T\mathbf{A}^T &= \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{A}((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T) = \mathbf{AB}.
\end{aligned}$$

Thus, $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the pseudoinverse of $\mathbf{A}$.

c) Since $\mathbf{A}$ is full-rank and fat, $\mathbf{A}\mathbf{A}^T$ is symmetric and non-singular. Defining $\mathbf{B} := \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$, we see that $\mathbf{B} \in \mathbb{R}^{n\times m}$, and

$$\begin{aligned}
\mathbf{ABA} &= \mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A} = \mathbf{A}\\
\mathbf{BAB} &= \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{B}\\
\mathbf{A}^T\mathbf{B}^T &= \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A} = \ (\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1})\mathbf{A} = \mathbf{BA}\\
\mathbf{B}^T\mathbf{A}^T &= (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{A}^T = \mathbf{I}_m = (\mathbf{A}\mathbf{A}^T)(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{AB}.
\end{aligned}$$

Thus, $\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$ is the pseudoinverse of $\mathbf{A}$.

d) If $\mathbf{A}$ is full-rank and square, then so is $\mathbf{A}^T$, and using part (c) and the uniqueness of the pseudoinverse (proved in part (a)), we can compute the pseudoinverse of $\mathbf{A}$ as

$$\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{A}^T(\mathbf{A}^T)^{-1}\mathbf{A}^{-1} = \mathbf{A}^{-1}.$$

e) Since $\mathbf{A}$ is an (orthogonal) projection matrix, it is symmetric and satisfies $\mathbf{A}^2 = \mathbf{A}$. We then have

$$\mathbf{A}\cdot\mathbf{A}\cdot\mathbf{A} = \mathbf{A}^2\cdot\mathbf{A} = \mathbf{A}\cdot\mathbf{A} = \mathbf{A}^2 = \mathbf{A}.$$

It also holds that $\mathbf{A}^T\mathbf{A}^T = \mathbf{A}\mathbf{A}$ since $\mathbf{A}$ is symmetric, which show that $\mathbf{A}$ is its own pseudoinverse.

f) Let $\mathbf{B} = (\mathbf{A}^+)^T$. Then $\mathbf{B} \in \mathbb{R}^{m\times n}$, and we have

$$\begin{aligned} \mathbf{A}^T\mathbf{B}\mathbf{A}^T &= \mathbf{A}^T(\mathbf{A}^+)^T\mathbf{A}^T = (\mathbf{A}\mathbf{A}^+\mathbf{A})^T = \mathbf{A}^T \\ \mathbf{B}\mathbf{A}^T\mathbf{B} &= (\mathbf{A}^+)^T\mathbf{A}^T(\mathbf{A}^+)^T = (\mathbf{A}^+\mathbf{A}\mathbf{A}^+)^T = (\mathbf{A}^+)^T = \mathbf{B} \\ (\mathbf{A}^T)^T\mathbf{B}^T &= \mathbf{A}\mathbf{A}^+ = \ (\mathbf{A}^+)^T\mathbf{A}^T = \mathbf{B}\mathbf{A}^T \\ \mathbf{B}^T(\mathbf{A}^T)^T &= \mathbf{A}^+\mathbf{A} = \mathbf{A}^T(\mathbf{A}^+)^T = \mathbf{A}^T\mathbf{B}. \end{aligned}$$

This shows that $(\mathbf{A}^T)^+ = (\mathbf{A}^+)^T$.

g) Let $\mathbf{B} := (\mathbf{A}^+)^T\mathbf{A}^+$. Then we have

$$(\mathbf{A}\mathbf{A}^T)\mathbf{B}(\mathbf{A}\mathbf{A}^T\ ) \ = \mathbf{A}\left(\mathbf{A}^T(\mathbf{A}^+)^T\right)\mathbf{A}^+\mathbf{A}\mathbf{A}^T = \ \mathbf{A}\left(\mathbf{A}^+\mathbf{A}\right)\mathbf{A}^+\mathbf{A}\mathbf{A}^T = \mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T,$$

and

$$\mathbf{B}(\mathbf{A}\mathbf{A}^T)\mathbf{B} = (\mathbf{A}^+)^T\mathbf{A}^+\mathbf{A}\left(\mathbf{A}^T(\mathbf{A}^+)^T\right)\mathbf{A}^+ = (\mathbf{A}^+)^T\mathbf{A}^+\mathbf{A}\left(\mathbf{A}^+\mathbf{A}\right)\mathbf{A}^+ = (\mathbf{A}^+)^T\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = (\mathbf{A}^+)^T\mathbf{A}^+ = \mathbf{B}.$$

Now, since $\mathbf{B}$ and $\mathbf{A}\mathbf{A}^T$ are both symmetric, their products will be symmetric if and only if they commute. We have

$$\mathbf{B}\mathbf{A}\mathbf{A}^T = (\mathbf{A}^+)^T\mathbf{A}^+\mathbf{A}\mathbf{A}^T = (\mathbf{A}^+)^T\mathbf{A}^T(\mathbf{A}^+)^T\mathbf{A}^T = (\mathbf{A}^+\mathbf{A}\mathbf{A}^+)^T\mathbf{A}^T = (\mathbf{A}^+)^T\mathbf{A}^T = \mathbf{A}\mathbf{A}^+,$$

and

$$\mathbf{A}\mathbf{A}^T\mathbf{B} = \mathbf{A}\mathbf{A}^T(\mathbf{A}^+)^T\mathbf{A}^+ = \mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}\mathbf{A}^+.$$

Thus, $(\mathbf{A}\mathbf{A}^T)^+ = (\mathbf{A}^+)^T\mathbf{A}^+$. Call this relationship (g1). Now, using this relation and part (f), we have

$$(\mathbf{A}^T\mathbf{A})^+ = [\mathbf{A}^T(\mathbf{A}^T)^T]^+ \overset{(g1)}{=} ((\mathbf{A}^T)^+)^T(\mathbf{A}^T)^+ \overset{(f)}{=} ((\mathbf{A}^+)^T)^T(\mathbf{A}^+)^T = \mathbf{A}^+(\mathbf{A}^+)^T\ .$$

h) We will show that $\mathcal{R}(\mathbf{A}^+) \subseteq \mathcal{R}(\mathbf{A}^T\ )$ and $\mathcal{R}(\mathbf{A}^+) \supseteq \mathcal{R}(\mathbf{A}^T\ )$. For showing the first part, let $\mathbf{y} \in \mathcal{R}(\mathbf{A}^+)$, then $\mathbf{y} = \mathbf{A}^+\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^m$. Then, we have

$$\mathbf{y} = \mathbf{A}^+\mathbf{x} = \mathbf{A}^+\mathbf{A}\mathbf{A}^+\mathbf{x} = \mathbf{A}^+\mathbf{A}\mathbf{y} = \mathbf{A}^T(\mathbf{A}^+)^T\mathbf{y}$$

so defining $\tilde{\mathbf{x}} := (\mathbf{A}^+)^T\mathbf{y}$, we see that $\mathbf{y}$ can be written as $\mathbf{A}^T\tilde{\mathbf{x}}$, which shows that $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T\ )$.

For the opposite direction, if $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T\ )$ is written as $\mathbf{y} = \mathbf{A}^T\mathbf{x}$, then we can similarly show that

$$\mathbf{y} = \mathbf{A}^T\mathbf{x} = (\mathbf{A}\mathbf{A}^+\mathbf{A})^T\mathbf{x} = \mathbf{A}^T(\mathbf{A}^+)^T\mathbf{A}^T\mathbf{x} = \mathbf{A}^T(\mathbf{A}^+)^T\mathbf{y} = \mathbf{A}^+\mathbf{A}\mathbf{y}$$

so $\mathbf{y} = \mathbf{A}^+\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} := \mathbf{A}\mathbf{y}$. Thus, $\mathbf{y} \in \mathcal{R}(\mathbf{A}^+)$.

Therefore, we have shown that $\mathcal{R}(\mathbf{A}^+) = \mathcal{R}(\mathbf{A}^T\ )$.

Now, let $\mathbf{x} \in \mathcal{N}(\mathbf{A}^+)$. We then have

$$\mathbf{A}^+\mathbf{x} = 0 \Rightarrow \mathbf{A}\mathbf{A}^+\mathbf{x} = 0 \Rightarrow (\mathbf{A}^+)^T\mathbf{A}^T\mathbf{x} = 0 \Rightarrow \mathbf{A}^T(\mathbf{A}^+)^T\mathbf{A}^T\mathbf{x} = 0 \Rightarrow (\mathbf{A}\mathbf{A}^+\mathbf{A})^T\mathbf{x} = 0 \Rightarrow \mathbf{A}^T\mathbf{x} = 0.$$

Thus, $\mathbf{x} \in \mathcal{N}(\mathbf{A}^T\ )$.

Similarly, if $\mathbf{x} \in \mathcal{N}(\mathbf{A}^T\ )$, we have

$$\mathbf{A}^T\mathbf{x} = 0 \ \Rightarrow\ (\mathbf{A}^+)^T\mathbf{A}^T\mathbf{x} = 0 \ \Rightarrow\ \mathbf{A}\mathbf{A}^+\mathbf{x} = 0 \ \Rightarrow \mathbf{A}^+\mathbf{A}\mathbf{A}^+\mathbf{x} = 0 \ \Rightarrow\ \mathbf{A}^+\mathbf{x} = 0.$$

Thus, $\mathbf{x} \in \mathcal{N}(\mathbf{A}^+)$.

We have therefore shown that $\mathcal{N}(\mathbf{A}^+) = \mathcal{N}(\mathbf{A}^T\ )$.

i) $\mathbf{P}$ and $\mathbf{Q}$ are symmetric by the properties of $\mathbf{A}^+$, and

$$\mathbf{P}^2 = \mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}\mathbf{A}^+ = \mathbf{P}.$$

Similarly,

$$\mathbf{Q}^2 = \mathbf{A}^+\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}^+\mathbf{A} = \mathbf{Q}.$$

Therefore, $\mathbf{P}$ and $\mathbf{Q}$ are projection matrices.

j) Clearly, for every $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} = \mathbf{P}\mathbf{x} = \mathbf{A}\mathbf{A}^+\mathbf{x} \in \mathcal{R}(\mathbf{A})$. Thus, we are done if we can show that $\mathbf{x} - \mathbf{P}\mathbf{x} \in \mathcal{R}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^T\ )$ which is $\mathcal{N}(\mathbf{A}^+)$ by (h).
We have

$$\mathbf{A}^+(\mathbf{x} - \mathbf{P}\mathbf{x}) = \mathbf{A}^+\mathbf{x} - \mathbf{A}^+\mathbf{A}\mathbf{A}^+\mathbf{x} = \mathbf{A}^+\mathbf{x} - \mathbf{A}^+\mathbf{x} = 0,$$

therefore $\mathbf{P}$ is indeed the projection onto $\mathcal{R}(\mathbf{A})$.

Similarly, for every $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{z} = \mathbf{Q}\mathbf{x} = \mathbf{A}^+\mathbf{A}\mathbf{x}\ \in \mathcal{R}(\mathbf{A}^+) = \mathcal{R}(\mathbf{A}^T\ )$, where the last equality follows from part (h). Thus, we are done if we can show that $\mathbf{x} - \mathbf{Q}\mathbf{x} \in \mathcal{N}(\mathbf{A})$. We have

$$\mathbf{A}(\mathbf{x} - \mathbf{Q}\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{A}^+\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{x} = 0,$$

therefore $\mathbf{Q}$ is indeed the projection onto $\mathcal{R}(\mathbf{A}^T\ )$.

k) We have

$$\mathbf{A}^+(\mathbf{A}\mathbf{x}^* - \mathbf{b}) = \mathbf{A}^+\mathbf{A}\mathbf{A}^+\mathbf{b} - \mathbf{A}^+\mathbf{b} = \mathbf{A}^+\mathbf{b} - \mathbf{A}^+\mathbf{b} = 0.$$

Hence $\mathbf{A}\mathbf{x}^* - \mathbf{b}$ is orthogonal to $\mathcal{R}((\mathbf{A}^+)^\mathbf{T}) = \mathcal{R}((\mathbf{A}^\mathbf{T})^+)$ which is $\mathcal{R}(\mathbf{A})$ by (h) and by orthogonality principle, $\mathbf{x}^*$ is indeed a least-squares solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$.

l) Suppose that the linear equation $\mathbf{b} = \mathbf{A}\mathbf{x}$ has a solution $\tilde{\mathbf{x}}$. Then, $\mathbf{A}\mathbf{x}^* = \mathbf{A}\mathbf{A}^+\mathbf{b} = \mathbf{A}\mathbf{A}^+\mathbf{A}\tilde{\mathbf{x}} = \mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}$. Now, let $\mathbf{z}$ be any other solution to $\mathbf{b} = \mathbf{A}\mathbf{x}$, i.e., we have $\mathbf{A}\mathbf{z} = \mathbf{b}$.

Then $(\mathbf{z} - \mathbf{x}^*) \in \mathcal{N}(\mathbf{A})$. Also $\mathbf{x}^* \in \mathcal{R}(\mathbf{A}^+) = \mathcal{R}(\mathbf{A}^T)$.
Since $\mathcal{N}(\mathbf{A}) \perp \mathcal{R}(\mathbf{A}^T)$,

$$(\mathbf{x}^*)^T(\mathbf{z} - \mathbf{x}^*) = 0$$

We then have

$$\|\mathbf{z}\|^2 = \|\mathbf{x}^* + (\mathbf{z} - \mathbf{x}^*)\|^2 = \|\mathbf{x}^*\|^2 + \|\mathbf{z} - \mathbf{x}^*\|^2 \geq\ \|\mathbf{x}^*\|^2 .$$

2 Problem 2

a) Consider the characteristic polynomial of $\mathbf{A}$, namely, $X_\mathbf{A}(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A})$. Clearly, the highest power of λ in $X_\mathbf{A}$, i.e., the n^{th} power, occurs only in the term $\prod_{i=1}^n(\lambda - \mathbf{A}_{ii})$. Therefore, the coefficient of λ^n equals 1. The constant term is given by $X_\mathbf{A}(0) = \det(-\mathbf{A}) = (-1)^n \det(\mathbf{A})$. Therefore, we have

$$\begin{aligned}\lambda_1.\lambda_2 \dots \lambda_n &= \text{product of roots of } \{X_\mathbf{A}(\lambda) = 0\} \\ &= (-1)^n.\frac{\text{constant term}}{\text{coefficient of } \lambda^n} \\ &= \det(\mathbf{A})\end{aligned}$$

Alternate solution:
Using Schur decomposition $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^\mathbf{H}$. Then

$$\det(\mathbf{A}) = \det(\mathbf{U}\mathbf{T}\mathbf{U}^\mathbf{H}) = \det(\mathbf{U})\det(\mathbf{T})\det(\mathbf{U}^H) = \det(\mathbf{U})\det(\mathbf{U}^H)\det(\mathbf{T}) = \det(\mathbf{T})$$

Since $\mathbf{T}$ is an upper triangular matrix $\det(\mathbf{T}) = \prod_{i=1}^{n} t_{ii}$. We also know the elements on the diagonal of $\mathbf{T}$ are equal to the eigenvalues $t_i = \lambda_i$. Hence,

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$$

b) We have

$$X_{\mathbf{A}^T}(\lambda) = \det(\lambda\mathbf{I} - \mathbf{A}^T) = \det((\lambda\mathbf{I} - \mathbf{A})^T) = \det(\lambda\mathbf{I} - \mathbf{A}) = X_{\mathbf{A}}(\lambda)$$

which shows that $\mathbf{A}^T$ and $\mathbf{A}$ have identical characteristic polynomials and hence, identical eigenvalues.

c) A quick note about upper triangular matrices.
If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ upper triangular matrices, then the elements of $\mathbf{AB}$ are

$$[\mathbf{AB}]_{ij} = \sum_{k=1}^{n} \mathbf{A}_{ik}\mathbf{B}_{kj} = \sum_{i \le k \le j} \mathbf{A}_{ik}\mathbf{B}_{kj}$$

as $\mathbf{A}_{ik} = 0$ for $k < i$ and $\mathbf{B}_{kj} = 0$ for $j < k$. Therefore $[\mathbf{AB}]_{ij} = 0$ if $i > j$ i.e. $\mathbf{AB}$ is upper triangular. Also the elements on the diagonal are $[\mathbf{AB}]_{ii} = \mathbf{A}_{ii}\mathbf{B}_{ii}$. An analogous statement can be made about lower triangular matrices.

Let $\mathbf{A} = \mathbf{UTU^H}$ be the Schur decomposition of $\mathbf{A}$. Note $\mathbf{A}^2 = \mathbf{UTU^H UTU^H} = \mathbf{UT^2U^H}$. Similarly for any positive integer k, we have

$$\mathbf{A}^k = \mathbf{U}\mathbf{T}^k\mathbf{U}^H$$

which is a Schur decomposition as $\mathbf{T}^k$ is also upper triangular. Hence the diagonal of $\mathbf{T}^k$ contains the eigenvalues of $\mathbf{A}^k$ as discussed in part (a). The ith diagonal element of $\mathbf{T}$ is $\mathbf{T}_{ii}$ and thus the ith diagonal element of $\mathbf{T}^k$ is $(\mathbf{T}_{ii})^k$. Since $\mathbf{T}_{ii} = \lambda_i$ where λ_i is an eigenvalue of $\mathbf{A}$, we conclude that $\lambda_1^k, \ldots, \lambda_n^k$ are the eigenvalues of $\mathbf{A}^k$.

d) Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$. We have:

$$\mathbf{A} \text{ is invertible} \iff \det(\mathbf{A}) \neq 0 \iff \prod_{i=1}^{n} \lambda_i \neq 0 \iff \lambda_i \neq 0 \ \forall i$$

e) If $\mathbf{A}$ is invertible, we know that $\lambda_i \neq 0$ for all i. We have

$$\begin{aligned}
X_{\mathbf{A}^{-1}}(\lambda) &= \det(\lambda\mathbf{I} - \mathbf{A}^{-1}) \\
&= \det((\lambda\mathbf{A} - \mathbf{I})\mathbf{A}^{-1}) \\
&= \det(\lambda\mathbf{A} - \mathbf{I}) \det(\mathbf{A}^{-1}) \\
&= \lambda^n \det(\mathbf{A} - \lambda^{-1}\mathbf{I}) \det(\mathbf{A})^{-1} \\
&= (-\lambda)^n \det(\lambda^{-1}\mathbf{I} - \mathbf{A}) \det(\mathbf{A})^{-1} \\
&= (-\lambda)^n X_{\mathbf{A}}(\lambda^{-1}) \det(\mathbf{A})^{-1} \\
&= (-1)^n X_{\mathbf{A}}(\lambda^{-1}) \prod_{i=1}^{n} \frac{\lambda}{\lambda_i} \\
&= (-1)^n \prod_{i=1}^{n} \frac{\lambda}{\lambda_i}(\lambda^{-1} - \lambda_i) \\
&= \prod_{i=1}^{n} (\lambda - \lambda_i^{-1})
\end{aligned}$$

which shows that $\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_n^{-1}$ are the eigenvalues of $\mathbf{A}^{-1}$

f) We have

$$X_{\mathbf{T}^{-1}\mathbf{AT}}(\lambda) = \det(\lambda\mathbf{I} - \mathbf{T}^{-1}\mathbf{AT}) = \det(\mathbf{T}^{-1}(\lambda\mathbf{I} - \mathbf{A})\mathbf{T}) = \det(\mathbf{T})^{-1} \det(\mathbf{T}) \det(\lambda\mathbf{I} - \mathbf{A}) = X_{\mathbf{A}}(\lambda)$$

which shows that $\mathbf{T}^{-1}\mathbf{AT}$ has the same eigenvalues as $\mathbf{A}$.

3 Problem 3

a) We have

$$\lambda_1 + \lambda_2 + ... + \lambda_n = -(\text{coefficient of } \lambda^{n-1} \text{ in } X_{\mathbf{A}}(\lambda))$$

Now, in $X_{\mathbf{A}} = \det(\lambda\mathbf{I} - \mathbf{A})$, the only term containing λ^n and λ^{n-1} is $\prod_{i=1}^{n}(\lambda - \mathbf{A}_{ii})$.(This is immediately clear by considering the definition of a determinant in terms of permutations.) Therefore, the coefficient of λ^{n-1} in $X_{\mathbf{A}}$ is the same as the coefficient of λ^{n-1} in $\prod_{i=1}^{n}(\lambda - \mathbf{A}_{ii})$, which is given by $-\sum_{i=1}^{n} \mathbf{A}_{ii} = -\text{tr}(\mathbf{A})$. Therefore,

$$\text{tr}(\mathbf{A}) = \lambda_1 + \lambda_2 + ... + \lambda_n$$

Alternate solution:
The Schur decomposition of the matrix $\mathbf{A}$ is $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$ where $\mathbf{U}$ is a unitary matrix and $\mathbf{T}$ is an upper triangular matrix whose diagonal corresponds to the set of eigenvalues of $\mathbf{A}$. Now by the cyclic property of trace($\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB})$),

$$\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{U}\mathbf{T}\mathbf{U}^H)=\text{tr}(\mathbf{U}^H\mathbf{U}\mathbf{T}) = \text{tr}(\mathbf{T}) = \lambda_1 + \lambda_2 + ... + \lambda_n$$

b) Using problems 3(a) and 2(c), the result is immediate.

$$\text{tr}(\mathbf{A}^k) = \lambda_1^k + \lambda_2^k + ... + \lambda_n^k$$

4 Problem 4

The Frobenius norm of a matrix $\mathbf{A}$ is given by

$$\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}^H\mathbf{A})} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|\mathbf{A}_{ij}|^2}\sqrt{\text{tr}(\mathbf{A}\mathbf{A}^H)} = \left\|\mathbf{A}^H\right\|_F$$

The Schur decomposition of the matrix $\mathbf{A}$ is $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H$ where $\mathbf{U}$ is a unitary matrix and $\mathbf{T}$ is an upper triangular matrix whose diagonal corresponds to the set of eigenvalues of $\mathbf{A}$.
For any matrix $\mathbf{B}$ and a unitary matrix $\mathbf{V}$:

$$\|\mathbf{V}\mathbf{B}\|_F = \sqrt{\text{tr}\left((\mathbf{V}\mathbf{B})^H(\mathbf{V}\mathbf{B})\right)} = \sqrt{\text{tr}\left(\mathbf{B}^H\mathbf{V}^H\mathbf{V}\mathbf{B}\right)} = \sqrt{\text{tr}\left(\mathbf{B}^H\mathbf{B}\right)} = \|\mathbf{B}\|_F$$

Also,

$$\|\mathbf{B}\mathbf{V}\|_F = \left\|\mathbf{V}^H\mathbf{B}^H\right\|_F = \left\|\mathbf{B}^H\right\|_F = \|\mathbf{B}\|_F$$

Hence we have:

$$\|\mathbf{A}\|_F = \left\|\mathbf{U}\mathbf{T}\mathbf{U}^H\right\|_F = \|\mathbf{T}\|_F$$

Now $\mathbf{T}$ is an upper triangular matrix with diagonal elements as $\lambda_1, \lambda_2, .., \lambda_n$. Therefore

$$\|\mathbf{T}\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}|\mathbf{T}_{ij}|^2 \ \geq \sum_{i=1}^{n}|\lambda_i|^2$$

Hence

$$\sum_{i=1}^{n}|\lambda_i|^2 \leq \|\mathbf{T}\|_F^2 = \|\mathbf{A}\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}|\mathbf{A}_{ij}|^2$$

5 Problem 5

a) Let λ be an eigenvalue of A. Then by definition $\det(A-\lambda I)=0$ i.e. $\mathcal{N}(A-\lambda I)$ is non trivial. Therefore there exists a vector $v \neq 0$ such that $(A-\lambda I)v=0$ i.e. $Av=\lambda v$. Premultiplying by A^{k-1} gives us $A^k v=(\lambda)^k v$.But $A^k=0$ and hence $\lambda^k v=0 \implies \lambda^k=0 \implies \lambda=0$.

   Therefore if λ is an eigenvalue of A, $\lambda=0$.

b) **(First Solution)** Cayley Hamilton Theorem is indeed valid over any field and also over commutative rings. From a) we know the eigenvalues are 0. Hence the characteristic polynomial is $p_A(x)=(x-0)^n=x^n$. By Cayley Hamiltion for arbitrary field, we know A satisfies its own characteristic equation.

   Hence $A^n=0$. As k is the smallest positive integer for which $A^k=0$, we get $k \leq n$.

   (Alternative Solution) Define the sequence of subspaces

$$S_1=\mathcal{N}(A), S_2=\mathcal{N}(A^2), \ldots S_k=\mathcal{N}(A^k)$$

   Clearly $S_1 \subseteq S_2 \subseteq S_3 \cdots \subseteq S_k$. Let the corresponding dimensions be $d_1 \leq d_2 \leq d_3 \cdots \leq d_k$.

   Suppose there exists an integer i such that $d_i=d_{i+1}$ and let i be the first instance such occurrence. Then $\mathcal{N}(A^i)=\mathcal{N}(A^{i+1})$ as $\mathcal{N}(A^i) \subseteq \mathcal{N}(A^{i+1})$.

   Let t be a non-negative integer. For $j=i+1+t$, let $x \in \mathcal{N}(A^j)$. Then we have $A^j x=0 \implies A^t x \in \mathcal{N}(A^{i+1})=\mathcal{N}(A^i) \implies A^{t+i}x=A^{j-1}x=0 \implies x \in \mathcal{N}(A^{j-1})$. Therefore $\mathcal{N}(A^j)=\mathcal{N}(A^{j-1})$ for all $j \geq i+1$. By induction, this means $\mathcal{N}(A^i)=\mathcal{N}(A^{i+1})=\mathcal{N}(A^{i+2})=\mathcal{N}(A^{i+3})=\ldots$.

   Now we will show such i exists and $i \leq n$. Assume to the contrary $d_1<d_2<d_3\cdots<d_n<d_{n+1}$. Since A is not full-rank, we know $d_1 \geq 1$. Due to chain of inequalities, we would have $d_{n+1} \geq n+1$. However $\mathcal{N}(A^{n+1}) \subset \mathbf{F}^n$ and hence $d_{n+1} \leq n$. We have a contradiction. Therefore $d_n=d_{n+1}$ and $i \leq n$.

   But we know $\mathcal{N}(A^k)=\mathbf{F}^n$ and it is the smallest such k. If $i<k$, then $\mathcal{N}(A^i)=\mathcal{N}(A^k)=\mathbf{F}^n$ which contradicts that k is the smallest such k. Hence $k \leq i$ and as $i \leq n$, we can conclude $k \leq n$.

c) Sufficient to show that $\{x, Ax, A^2x, \ldots, A^{n-1}x\}$ are linearly independent to show that they are a basis for $\mathbf{F}^n$ as the number of vectors is equal to the dimension.

   Let $\sum_{i=0}^{n-1} \alpha_i A^i x=0$. Let j be the smallest index for which $\alpha_j \neq 0$.

$$0=\sum_{i=0}^{n-1} \alpha_i A^i x=\sum_{i=j}^{n-1} \alpha_i A^i x$$

   Premultiplying by A^{n-1-j} and substituting $i=k+j$, we get $0=\sum_{k=0}^{n-1-j} \alpha_{j+k} A^{n-1+k} x=\alpha_j A^{n-1} x$ as $A^{n-1+k}=0$ for $k \geq 1$. We thus have $\alpha_j A^{n-1}x=0$. As $A^{n-1}x \neq 0$, we can conclude that $\alpha_j=0$ which contradicts our assumption that $\alpha_j \neq 0$. Therefore there is no smallest nonnegative index such that $\alpha_i=0$. Hence if the linear combination is 0, all α_i are 0 i.e. they are linearly independent.