ECE586/AI586 Applied Matrix Analysis - Homework 2

Winter 2024
Instructor: Xiao Fu

School of Electrical Engineering and Computer Science
Oregon State University

January 23, 2025

Woonki Kim
kimwoon@oregonstate.edu

Q1. Show that a Vandermonde matrix with a size of $m \times n$ has full column rank if $n \leq m$, given that all the variables in the Vandermonde matrix are distinct. (5%)

A1.
Let $\mathbf{A}$ be an $m \times n$ Vandermonde matrix defined as:

$$\mathbf{A} = \begin{bmatrix} 1 & z_1 & z_1^2 & \dots & z_1^{n-1} \\ 1 & z_2 & z_2^2 & \dots & z_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_m & z_m^2 & \dots & z_m^{n-1} \end{bmatrix}$$

where $z_1, z_2, \dots, z_m$ are distinct values and $n \leq m$.
We need to prove $\mathbf{A}$ has full column rank.

Proof by Construction:
Let $\mathbf{B}$:

$$\mathbf{B} = \begin{bmatrix} 1 & z_1 & z_1^2 & \dots & z_1^{n-1} \\ 1 & z_2 & z_2^2 & \dots & z_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_n & z_n^2 & \dots & z_n^{n-1} \end{bmatrix}$$

where $\mathbf{B} \in \mathbb{C}^{n \times n}$

From our lecture note we know that:

$$\mathbf{B}\mathbf{x} = \mathbf{0} \text{ for some } \mathbf{x} \neq \mathbf{0} \text{ if and only if } \det(\mathbf{B}) = 0$$

$$\Rightarrow \text{if } \det(\mathbf{B}) \neq 0 \text{ then } \mathbf{B}\mathbf{x} = \mathbf{0} \text{ only when } \mathbf{x} = \mathbf{0}$$

Which means, if $\det(\mathbf{B}) \neq 0$, all rows and columns are linearly independent, implying that matrix is full rank.

The determinant of the Vandermonde matrix $\mathbf{B}$ is given by the well-known formula:

$$\det(\mathbf{B}) = \prod_{1\le i<j\le n} (z_j - z_i).$$

This determinant is nonzero if and only if all z_i are distinct, i.e., $z_i \neq z_j$ for $i \neq j$. This means that if the z_i values are distinct, then:

$$\det(\mathbf{B}) \neq 0.$$

Thus:

$$\text{rank}(\mathbf{B}) = n$$

Since $n \le m$, we can construct matrix $\mathbf{A}$ using $\mathbf{B}$:

$$\mathbf{A^T} = [\mathbf{B}\ \times]$$

here, "$\times$" means parts that do not matter.

Since $\mathbf{B}$ is a sub-matrix of $\mathbf{A}^T$, it is obvious that:

$$\text{rank}(\mathbf{A}^T) \ge \text{rank}(\mathbf{B}) = n$$

While by property of rank:

$$\text{rank}(\mathbf{A}^T) \le \min(n, m) \Rightarrow \text{rank}(\mathbf{A}^T) \le n (\because n \le m)$$

Thus to satisfy both inequalities:

$$\text{rank}(\mathbf{A}^T) = n$$

By the property of rank:

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T) = n$$

Proving that $\mathbf{A}$ has full column rank.

Q2. Let $\{\phi_1, \ldots, \phi_n\}$ be the DFT basis, and observe that

$$\mathbf{A}\phi_i = \frac{1}{\sqrt{n}} \begin{bmatrix} h_0 & h_{n-1} & \cdots & h_1 \\ h_1 & h_0 & h_{n-1} & \cdots \\ h_2 & h_1 & h_0 & \cdots \\ \vdots & \vdots & & \ddots \\ h_{n-1} & \cdots & & h_0 \end{bmatrix} \begin{bmatrix} 1 \\ e^{j2\pi(i-1)/n} \\ e^{j4\pi(i-1)/n} \\ \vdots \\ e^{j2\pi(n-1)(i-1)/n} \end{bmatrix}$$

$$= \frac{1}{\sqrt{n}} \underbrace{\sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n}}_{=d_i} \begin{bmatrix} 1 \\ e^{j2\pi(i-1)/n} \\ e^{j4\pi(i-1)/n} \\ \vdots \\ e^{j2\pi(n-1)(i-1)/n} \end{bmatrix} = d_i\phi_i.$$

Show that the above equality holds. (5%)

A2.

DFT basis:

$$(\phi_i)_m = \frac{1}{\sqrt{n}} e^{j2m\pi(i-1)/n}, \quad \text{for } m = 0, 1, \ldots, n-1.$$

- Now expanding the first row(discard $\frac{1}{\sqrt{n}}$ when expanding equation for convenience):

  $$(\mathbf{A}\phi_i)_0 = h_0 \cdot 1 + h_{n-1} \cdot e^{j2\pi(i-1)/n} + h_{n-2} \cdot e^{j4\pi(i-1)/n} + \cdots + h_1 \cdot e^{j2\pi(n-1)(i-1)/n}$$

  Using the fact introduced in the lecture note that:

  $$e^{j2\pi k(i-1)/n} = e^{-j2\pi(n-k)(i-1)/n}, \quad \forall k \in \{0, 1, \ldots, n-1\}$$

  Substituting:

  $$(\mathbf{A}\phi_i)_0 = h_0 \cdot 1 + h_{n-1} \cdot e^{-j2\pi(n-1)(i-1)/n} + h_{n-2} \cdot e^{-j2\pi(n-2)(i-1)/n} + \cdots + h_1 \cdot e^{-j2\pi(n-(n-1))(i-1)/n}$$

  We can just express $h_0 \cdot 1$ as:

  $$h_0 \cdot 1 = h_0 \cdot e^{-j2\pi(0)(i-1)/n}$$

  Now:

  $$(\mathbf{A}\phi_i)_0 = h_0 \cdot e^{-j2\pi(0)(i-1)/n} + h_{n-1} \cdot e^{-j2\pi(n-1)(i-1)/n} + h_{n-2} \cdot e^{-j2\pi(n-2)(i-1)/n} + \cdots + h_1 \cdot e^{-j2\pi(1))(i-1)/n}$$

  Re-expressing with summation:

  $$(\mathbf{A}\phi_i)_0 = \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n}$$

- Similarly, for the second row:

  $$(\mathbf{A}\phi_i)_1 = h_1 \cdot 1 + h_0 \cdot e^{j2\pi(i-1)/n} + h_{n-1} \cdot e^{j4\pi(i-1)/n} + \cdots + h_2 \cdot e^{j2\pi(n-1)(i-1)/n}$$

  Using the fact introduced in the lecture note that:

  $$e^{j2\pi k(i-1)/n} = e^{-j2\pi(n-k)(i-1)/n}, \quad \forall k \in \{0, 1, \ldots, n-1\}$$

  Substituting:

  $$(\mathbf{A}\phi_i)_1 = h_1 \cdot e^{-j2\pi(0)(i-1)/n} + \underline{h_0 \cdot e^{-j2\pi(n-1)(i-1)/n}} + h_{n-1} \cdot e^{-j2\pi(n-2)(i-1)/n} + \cdots + h_2 \cdot e^{-j2\pi(1)(i-1)/n}$$

Since the underlined part seems to break the pattern of:

$$h_k e^{-j2\pi(k-1)(i-1)/n}$$

Break down:

$$h_0 \cdot e^{-j2\pi(n-1)(i-1)/n} = h_0 \cdot e^{j2\pi(1)(i-1)/n} = h_0 \cdot e^{-j2\pi(-1)(i-1)/n}$$

Substituting in:

$$(\mathbf{A}\phi_i)_1 = h_1 \cdot e^{-j2\pi(0)(i-1)/n} + h_0 \cdot e^{-j2\pi(-1)(i-1)/n} + h_{n-1} \cdot e^{-j2\pi(n-2)(i-1)/n} + \cdots + h_2 \cdot e^{-j2\pi(1)(i-1)/n}$$

Re-expressing with summation:

$$(\mathbf{A}\phi_i)_1 = \sum_{k=0}^{n-1} h_k e^{-j2\pi(k-1)(i-1)/n} = \sum_{k=0}^{n-1} h_k e^{(-j2\pi k + j2\pi)(i-1)/n} = \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n} \cdot e^{j2\pi(i-1)/n}$$

Thus, final form:

$$(\mathbf{A}\phi_i)_1 = \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n} \cdot e^{j2\pi(i-1)/n}$$

Acknowledging a pattern here, but to make sure let's try two more, one for $(\mathbf{A}\phi_i)_2$ and $(\mathbf{A}\phi_i)_{n-1}$

- Similarly, for the third row:

$$(\mathbf{A}\phi_i)_2 = h_2 \cdot 1 + h_1 \cdot e^{j2\pi(i-1)/n} + h_0 \cdot e^{j4\pi(i-1)/n} + \cdots + h_3 \cdot e^{j2\pi(n-1)(i-1)/n}$$

Using the fact introduced in the lecture note that:

$$e^{j2\pi k(i-1)/n} = e^{-j2\pi(n-k)(i-1)/n}, \quad \forall k \in \{0, 1, \ldots, n-1\}$$

Substituting:

$$(\mathbf{A}\phi_i)_2 = h_2 \cdot e^{-j2\pi(0)(i-1)/n} + \underline{h_1 \cdot e^{-j2\pi(n-1)(i-1)/n}} + \underline{h_0 \cdot e^{-j2\pi(n-2)(i-1)/n}} + \cdots + h_3 \cdot e^{-j2\pi(1)(i-1)/n}$$

Since the underlined part seems to break the pattern of:

$$h_k e^{-j2\pi(k-2)(i-1)/n}$$

Break down:

$$h_1 \cdot e^{-j2\pi(n-1)(i-1)/n} = h_1 \cdot e^{j2\pi(1)(i-1)/n} = h_1 \cdot e^{-j2\pi(-1)(i-1)/n}$$

$$h_0 \cdot e^{-j2\pi(n-2)(i-1)/n} = h_0 \cdot e^{j2\pi(2)(i-1)/n} = h_0 \cdot e^{-j2\pi(-2)(i-1)/n}$$

Substituting in:

$$(\mathbf{A}\phi_i)_2 = h_2 \cdot e^{-j2\pi(0)(i-1)/n} + h_1 \cdot e^{-j2\pi(-1)(i-1)/n} + h_0 \cdot e^{-j2\pi(n-2)(i-1)/n} + \cdots + h_3 \cdot e^{-j2\pi(1)(i-1)/n}$$

Re-expressing with summation:

$$(\mathbf{A}\phi_i)_2 = \sum_{k=0}^{n-1} h_k e^{-j2\pi(k-2)(i-1)/n} = \sum_{k=0}^{n-1} h_k e^{(-j2\pi k+2j2\pi)(i-1)/n} = \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n} \cdot e^{2j2\pi(i-1)/n}$$

Thus, final form:

$$(\mathbf{A}\phi_i)_2 = \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n} \cdot e^{2j2\pi(i-1)/n}$$

- Similarly, for the n'th row:

$$(\mathbf{A}\phi_i)_{n-1} = h_{n-1} \cdot 1 + h_{n-2} \cdot e^{j2\pi(i-1)/n} + h_{n-3} \cdot e^{j4\pi(i-1)/n} + \cdots + h_0 \cdot e^{j2\pi(n-1)(i-1)/n}$$

Using the fact introduced in the lecture note that:

$$e^{j2\pi k(i-1)/n} = e^{-j2\pi(n-k)(i-1)/n}, \quad \forall k \in \{0, 1, \ldots, n-1\}$$

Substituting:

$$(\mathbf{A}\phi_i)_{n-1} = h_{n-1} \cdot e^{-j2\pi(0)(i-1)/n} + \underline{h_{n-2} \cdot e^{-j2\pi(n-1)(i-1)/n}} + \underline{h_{n-3} \cdot e^{-j2\pi(n-2)(i-1)/n}} + \cdots + \underline{h_0 \cdot e^{-j2\pi(1)(i-1)/n}}$$

Since the underlined part seems to break the pattern of:

$$h_k e^{-j2\pi(k-(n-1))(i-1)/n}$$

Break down:

$$h_{n-2} \cdot e^{-j2\pi(n-1)(i-1)/n} = h_{n-2} \cdot e^{j2\pi(1)(i-1)/n} = h_{n-2} \cdot e^{-j2\pi(-1)(i-1)/n}$$

$$h_{n-3} \cdot e^{-j2\pi(n-2)(i-1)/n} = h_{n-3} \cdot e^{j2\pi(2)(i-1)/n} = h_{n-3} \cdot e^{-j2\pi(-2)(i-1)/n}$$

$$\cdots$$

$$h_0 \cdot e^{-j2\pi(1)(i-1)/n} = h_0 \cdot e^{j2\pi(n-1)(i-1)/n} = h_0 \cdot e^{-j2\pi(-(n-1))(i-1)/n}$$

Substituting in:

$$(\mathbf{A}\phi_i)_{n-1} = h_{n-1} \cdot e^{-j2\pi(0)(i-1)/n} + h_{n-2} \cdot e^{-j2\pi(-1)(i-1)/n} + h_{n-3} \cdot e^{-j2\pi(-2)(i-1)/n} + \cdots + h_0 \cdot e^{-j2\pi(-(n-1))(i-1)/n}$$

Re-expressing with summation:

$$(\mathbf{A}\phi_i)_{n-1} = \sum_{k=0}^{n-1} h_k e^{-j2\pi(k-(n-1))(i-1)/n} = \sum_{k=0}^{n-1} h_k e^{(-j2\pi k+2j\pi(n-1))(i-1)/n} = \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n} \cdot e^{2j\pi(n-1)(i-}$$

Thus, final form:

$$(\mathbf{A}\phi_i)_{n-1} = \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n} \cdot e^{j2\pi(n-1)(i-1)/n}$$

- Thus we have shown the pattern of:

$$(\mathbf{A}\phi_i)_0 = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n} \cdot 1$$

$$(\mathbf{A}\phi_i)_1 = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n} \cdot e^{j2\pi(i-1)/n}$$

$$(\mathbf{A}\phi_i)_2 = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n} \cdot e^{j4\pi(i-1)/n}$$

$$\cdots$$

$$(\mathbf{A}\phi_i)_{n-1} = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n} \cdot e^{j2\pi(n-1)(i-1)/n}$$

Showing constant pattern, since the matrix $\mathbf{A}$ is the circulant matrix, unproven parts could be assumed by the 4 cases above.

Thus we can conclude :

$$\mathbf{A}\phi_i = \underbrace{\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} h_k e^{-j2\pi k(i-1)/n}}_{=d_i} \begin{bmatrix} 1 \\ e^{j2\pi(i-1)/n} \\ e^{j4\pi(i-1)/n} \\ \vdots \\ e^{j2\pi(n-1)(i-1)/n} \end{bmatrix} = d_i \phi_i.$$

Q3. Consider a weighted linear model fitting problem

$$\min_{\mathbf{x}} \sum_{i=1}^{n} w_i (y_i - \mathbf{a}_i^\top \mathbf{x})^2,$$

where $w_1, \ldots, w_n$ are nonnegative weights corresponding to the data pairs $(y_i, \mathbf{a}_i)$ for $i = 1, \ldots, n$. Derive the solution of this fitting problem using (i) projection theorem and (ii) convex optimization's optimality condition (verify that the function to be minimized is convex in $\mathbf{x}$ first). (10%)

A3.

First, let's write the problem in standard LS problem form.

Since $w_1, \ldots, w_n$ are nonnegative weights corresponding to the data pairs $(y_i, \mathbf{a_i})$, we can represent $w_1, \ldots, w_n$ as:

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 & 0 \\ 0 & w_2 & \cdots & 0 & 0 \\ 0 & 0 & w_3 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & \cdots & 0 & 0 & w_n \end{bmatrix}$$

where $\mathbf{W} \in \mathbb{R}^{n \times n}$.

Let:

$$\mathbf{y} = [y_1, y_2, \cdots, y_n]^T$$

Now considering $\mathbf{a_i}^T$. Let $\mathbf{a_i}, \mathbf{x} \in \mathbb{R}^d$, then we can reshape vector $\mathbf{a}$ into matrix:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a_1}^T \\ \mathbf{a_2}^T \\ \vdots \\ \mathbf{a_n}^T \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{bmatrix}$$

where, $\mathbf{A} \in \mathbb{R}^{n \times d}$.
Now reformulating objective function in matrix multiplication and using norm:

$$\mathbf{W}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$$

Thus, rewriting the optimization problem:

$$\min_{\mathbf{x}} \mathbf{W}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$$

(i) Projection Theorem
Let $\mathbf{z} = \mathbf{A}\mathbf{x}$, then:

$$\mathbf{z} \in \mathcal{R}(\mathbf{A})$$

Let $\mathbf{x}_{LS}$ be an LS solution, then:

$$\Pi_{\mathcal{R}(\mathbf{A})}(\mathbf{y}) = \min_{\mathbf{z} \in \mathcal{R}(\mathbf{A})} \mathbf{W}\|\mathbf{y} - \mathbf{z}\|_2^2 = \mathbf{W}\mathbf{A}\mathbf{x}_{LS}$$

The projection theorem states:

$$\mathbf{y}_s = \Pi_{\mathcal{S}}(\mathbf{y}) \iff \mathbf{y} \in \mathcal{S}, \mathbf{z}^T(\mathbf{y}_s - \mathbf{y}) = 0 \text{ for all } \mathbf{z} \in \mathcal{S}$$

Using the theorem:

$$\Pi_{\mathcal{R}(A)}(\mathbf{y}) = \mathbf{W}\mathbf{A}\mathbf{x}_{LS} \iff \mathbf{z}^T\mathbf{W}(\mathbf{y} - \mathbf{A}\mathbf{x}_{LS}) = 0, \forall \mathbf{z} \in \mathcal{R}(\mathbf{A})$$

Eventually:

$$\mathbf{W}\mathbf{A}\mathbf{x}_{LS} \iff \mathbf{x}^T\mathbf{A}^T\mathbf{W}(\mathbf{y} - \mathbf{A}\mathbf{x}_{LS}) = 0, \forall \mathbf{x} \in \mathbb{R}^d$$

To satisfy the equation for $\forall \mathbf{x} \in \mathbb{R}^d$, following must hold:

$$\mathbf{A}^T\mathbf{W}(\mathbf{y} - \mathbf{A}\mathbf{x}_{LS}) = 0$$

It is because, there is no such vector orthogonal to all vectors in a subspace.
Thus we have:

$$\mathbf{A}^T\mathbf{W}(\mathbf{y} - \mathbf{A}\mathbf{x}_{LS}) = 0$$

Expanding:

$$\mathbf{A}^T\mathbf{W}\mathbf{y} - \mathbf{A}^T\mathbf{W}\mathbf{A}\mathbf{x}_{LS} = 0$$

Now verify if $(\mathbf{A}^T\mathbf{W}\mathbf{A})$ can be invertible

We know that for $\mathbf{A}^T\mathbf{A}$ if the matrix $\mathbf{A}$ is full column rank, it is non-singular, thus invertible. And even if the $\mathbf{A}^T A$ is non-singular, for $(\mathbf{A}^T\mathbf{W}\mathbf{A})$ if diagonal weight matrix $\mathbf{W}$ has elements that has value of 0, it will ruin linearly independency.

Thus, for $\mathbf{A}^T\mathbf{W}\mathbf{A}$ it is non-singular only if $\mathbf{A}$ is full column rank and diagonal weight matrix $\mathbf{W}$ has strictly positive elements.

And since we do not have these conditions, $\mathbf{A}^T\mathbf{W}\mathbf{A}$ is not invertible.
Furthermore, it cannot be written as pseudo inverse since there is no guarantee that $\mathbf{A}^T\mathbf{W}\mathbf{A}$ is full-column rank and not a tall matrix.

Thus this problem has no single LS solution

If $\mathbf{A}$ is full column rank and diagonal weight matrix $\mathbf{W}$ has strictly positive elements then:

$$\mathbf{x}^* = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y}$$

(ii)Convex optimization's optimality condition

- Verifying that the function is Convex.

  Objective function defined above:

  $$f(\mathbf{x}) = \mathbf{W}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$$

  Expanding:

  $$f(\mathbf{x}) = \mathbf{y}^T\mathbf{W}\mathbf{y} - 2\mathbf{y}^T\mathbf{W}\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{A}^T\mathbf{W}\mathbf{A}\mathbf{x}.$$

  Rearranging:

  $$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}^T\mathbf{W}\mathbf{A}\mathbf{x} - 2\mathbf{y}^T\mathbf{W}\mathbf{A}\mathbf{x} + \mathbf{y}^T\mathbf{W}\mathbf{y}$$

  By our lecture note for quadratic function:

  $$f(\mathbf{x}) = \mathbf{x}^T\mathbf{R}\mathbf{x} + \mathbf{q}^T\mathbf{x} + c$$

  where R is symmetric, function f is convex if R is PSD.

  In our formula, analyze if $\mathbf{A}^T\mathbf{W}\mathbf{A}$ is symmetric:

  $$(\mathbf{A}^T\mathbf{W}\mathbf{A})^T = \mathbf{A}^T\mathbf{W}^T\mathbf{A}$$

  Since, $\mathbf{W}$ is diagonal matrix:

  $$\mathbf{A}^T\mathbf{W}^T\mathbf{A} = \mathbf{A}^T\mathbf{W}\mathbf{A}$$

Thus:

$$(\mathbf{A}^T\mathbf{W}\mathbf{A})^T = \mathbf{A}^T\mathbf{W}\mathbf{A}$$

proving symmetricity.

Now we need to prove that it is PSD.

A matrix is positive semi definite when it is symmetric and $\mathbf{x}^T\mathbf{M}\mathbf{x} \geq 0, \forall \mathbf{x} \in \mathbb{R}^n$ holds.

Since we have proven symmetricity, we need to show that:

$$\mathbf{x}^T\mathbf{A}^T\mathbf{W}\mathbf{A}\mathbf{x} \geq 0, \forall x$$

Reformulating:

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

Let $\mathbf{z} = \mathbf{A}\mathbf{x}$, then:

$$\mathbf{z}^T\mathbf{W}\mathbf{z} = \sum_{i=1}^{n} w_i z_i^2$$

Since $w_i, \forall i$ is non-negative weight and z_i^2 is square:

$$\mathbf{z}^T\mathbf{W}\mathbf{z} = \mathbf{x}^T\mathbf{A}^T\mathbf{W}\mathbf{A}\mathbf{x} \geq 0, \forall x$$

Thus, $\mathbf{x}^T\mathbf{A}^T\mathbf{W}\mathbf{A}\mathbf{x}$ is positive semi definite proving our objective function is convex.

- Convex optimization's optimality condition

  By our lecture note, if function f is convex, a point x is an optimal solution if and only if $\nabla f(x) = 0$

$$\nabla f(\mathbf{x}) = -2\mathbf{y}^T\mathbf{W}\mathbf{A} + 2\mathbf{A}^T\mathbf{W}\mathbf{A}$$

Rearranging,

$$\mathbf{A}^T\mathbf{W}\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{W}\mathbf{y}.$$

As mentioned when solving using projection theorem, this problem has no single LS solution

If $\mathbf{A}$ is full column rank and diagonal weight matrix $\mathbf{W}$ has strictly positive elements then:

$$\mathbf{x}^* = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y}$$

Q4. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a full-column rank matrix. Let $\mathbf{P}_\mathbf{A}$ and $\mathbf{P}_\mathbf{A}^\perp$ be the orthogonal and orthogonal complement projectors of $\mathbf{A}$, respectively. Verify that:

(a) $\mathbf{A}^\top\mathbf{A}$ is nonsingular; (5%)

(b) $\mathbf{P}_\mathbf{A}^2 = \mathbf{P}_\mathbf{A}$; (5%)

(c) $(\mathbf{P}_\mathbf{A}^\perp)^2 = \mathbf{P}_\mathbf{A}^\perp$; (5%)

(d) $\mathcal{R}(\mathbf{P}_\mathbf{A}) = \mathcal{R}(\mathbf{A})$. (5%)

A4.

(a) $\mathbf{A}^\top\mathbf{A}$ is nonsingular

By our lecture note, a square matrix $\mathbf{M}$ is said to be non-singular if the columns of $\mathbf{M}$ are linearly independent. Meaning that if $\mathbf{M}$ is full column rank and symmetric, it is non-singular.

The matrix $\mathbf{A}^\top\mathbf{A}$ is an $n \times n$ Gram matrix, which is always symmetric:

$$\mathbf{A}^\top\mathbf{A} = \sum_{i=1}^{m} \mathbf{a}_i\mathbf{a}_i^\top.$$

Now considering linearly independency of columns.

By our lecture note,

$$\text{rank}(\mathbf{AB}) \leq \min(\text{rank}(\mathbf{A})\,\text{rank}(\mathbf{B}))$$

Equality holds if columns of $\mathbf{A}$ are linearly independent and the rows of $\mathbf{B}$ are linearly independent.

Since $\mathbf{A}$ has full column rank, $\mathbf{A}^T$ has full row rank.

Thus:

$$\text{rank}(\mathbf{A}^T\mathbf{A}) = \min(\text{rank}(\mathbf{A})\,\text{rank}(\mathbf{A}^T))$$
$$\text{rank}(\mathbf{A}^T\mathbf{A}) = \min(n, n) = n$$

This implies that $\mathbf{A}^\top\mathbf{A}$ has full rank n. Thus it is symmetric and full column rank. Proving non-sigularity.

(b) $\mathbf{P}_\mathbf{A}^2 = \mathbf{P}_\mathbf{A}$;

Given a full-column rank matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the orthogonal projection matrix onto the range space of $\mathbf{A}$ is:

$$\mathbf{P}_\mathbf{A} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top.$$

Squaring $\mathbf{P}_\mathbf{A}$:

$$\mathbf{P}_\mathbf{A}^2 = \left(\mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\right)\left(\mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\right).$$

Rearranging:

$$\mathbf{P}_{\mathbf{A}}^2 = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1}(\mathbf{A}^\top \mathbf{A})(\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{A}^\top.$$

Since:

$$\mathbf{A}^\top \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} = I_n.$$

Thus:

$$\mathbf{P}_{\mathbf{A}}^2 = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{A}^\top = \mathbf{P}_{\mathbf{A}}.$$

Finally:

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

(c) $(\mathbf{P}_{\mathbf{A}}^\perp)^2 = \mathbf{P}_{\mathbf{A}}^\perp$;

Definition of $\mathbf{P}_{\mathbf{A}}^\perp$:

$$\mathbf{P}_{\mathbf{A}}^\perp = \mathbf{I} - \mathbf{P}_{\mathbf{A}},$$

where $\mathbf{P}_{\mathbf{A}}$ is:

$$\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{A}^\top.$$

Compute $(\mathbf{P}_{\mathbf{A}}^\perp)^2$:

$$(\mathbf{P}_{\mathbf{A}}^\perp)^2 = (\mathbf{I} - \mathbf{P}_{\mathbf{A}})(\mathbf{I} - \mathbf{P}_{\mathbf{A}}).$$

Expanding:

$$(\mathbf{P}_{\mathbf{A}}^\perp)^2 = \mathbf{I} - 2\mathbf{P}_{\mathbf{A}} + \mathbf{P}_{\mathbf{A}}^2.$$

Since we previously verified that $\mathbf{P}_{\mathbf{A}}^2 = \mathbf{P}_{\mathbf{A}}$, substitute:

$$(\mathbf{P}_{\mathbf{A}}^\perp)^2 = \mathbf{I} - 2\mathbf{P}_{\mathbf{A}} + \mathbf{P}_{\mathbf{A}} = \mathbf{I} - \mathbf{P}_{\mathbf{A}}.$$

Thus:

$$(\mathbf{P}_{\mathbf{A}}^\perp)^2 = \mathbf{P}_{\mathbf{A}}^\perp.$$

(d) $\mathcal{R}(\mathbf{P}_{\mathbf{A}}) = \mathcal{R}(\mathbf{A})$.

$\mathcal{R}(\mathbf{P}_{\mathbf{A}}) = \mathcal{R}(\mathbf{A})$ holds if and only if $\mathcal{R}(\mathbf{P}_{\mathbf{A}}) \subseteq \mathcal{R}(\mathbf{A})$ and $\mathcal{R}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{P}_{\mathbf{A}})$ both holds

- Prove $\mathcal{R}(\mathbf{P_A}) \subseteq \mathcal{R}(\mathbf{A})$

  To show that $\mathcal{R}(\mathbf{P_A}) \subseteq \mathcal{R}(\mathbf{A})$, we must show that for any $\mathbf{y} \in \mathcal{R}(\mathbf{P_A})$, there exists some $\mathbf{x}$ such that:

  $$\mathbf{y} = \mathbf{A}\mathbf{x}.$$

  By definition, the range space of $\mathbf{P_A}$ is given by:

  $$\mathcal{R}(\mathbf{P_A}) = \{\mathbf{y} \mid \mathbf{y} = \mathbf{P_A}\mathbf{b} \text{ for some } \mathbf{b} \in \mathbb{R}^m\}.$$

  Since $\mathbf{P_A}$ is:

  $$\mathbf{P_A} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top,$$

  we can write $\mathbf{y}$ as:

  $$\mathbf{y} = \mathbf{P_A}\mathbf{b} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}.$$

  Now, define:

  $$\mathbf{x}' = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}$$

  Now, let's determine the dimensions of each term in:

  $$(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}.$$

  Since $\mathbf{A}$ is $m \times n$, multiplying it by its transpose gives an $n \times n$ matrix. $\mathbf{A}^\top\mathbf{A}$

  $$\mathbf{A}^\top\mathbf{A} \in \mathbb{R}^{n\times n}.$$

  Since $\mathbf{A}$ has full column rank, $\mathbf{A}^\top\mathbf{A}$ is invertible. The inverse of an $n \times n$ matrix remains $n \times n$.

  $$(\mathbf{A}^\top\mathbf{A})^{-1} \in \mathbb{R}^{n\times n}.$$

  $\mathbf{A}^\top$ is $n \times m$, and $\mathbf{b}$ is $m \times 1$. The result of multiplying $n \times m$ by $m \times 1$ gives an $n \times 1$ vector:

  $$\mathbf{A}^\top\mathbf{b} \in \mathbb{R}^n.$$

  Full Expression $(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b}$ Since we are multiplying an $n \times n$ matrix by an $n \times 1$ vector, the result is an $n \times 1$ vector.

  Thus:

  $$\mathbf{x}' = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b} \in \mathbb{R}^n.$$

  Since $\mathbf{x}'$ is an n-dimensional vector, we can rewrite $\mathbf{y}$ as:

$$\mathbf{y} = \mathbf{P_A}\mathbf{b} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{b} = \mathbf{A}\mathbf{x}', \mathbf{x}' \in \mathbb{R}^n$$

which is clearly in the range space of $\mathbf{A}$
Since we have shown that for every $\mathbf{y} \in \mathcal{R}(\mathbf{P_A})$, there exists an $\mathbf{x}'$ such that:

$$\mathbf{y} = \mathbf{A}\mathbf{x}', \mathbf{x}' \in \mathbb{R}^n$$

Thus:

$$\mathcal{R}(\mathbf{P_A}) \subseteq \mathcal{R}(\mathbf{A}).$$

- Prove $\mathcal{R}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{P_A})$
  For a vector $\mathbf{y} \in \mathbb{R}(\mathbf{A})$ let:

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \mathbf{x} \in \mathbb{R}^n$$

Then $\mathbf{P_A}\mathbf{y}$:

$$\begin{aligned} \mathbf{P_A}\mathbf{y} &= \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top(\mathbf{A}\mathbf{x}). \\ &= \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}(\mathbf{A}^\top\mathbf{A})\mathbf{x} \end{aligned}$$

Since $\mathbf{A}$ is full column rank, $\mathbf{A}^\top\mathbf{A}$ is invertible,

$$\mathbf{A}^\top\mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1} = I,$$

Thus:

$$\mathbf{P_A}\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{y}, \mathbf{y} \in \mathcal{R}(\mathbf{A})$$

This shows that $\mathbf{y}$ and $\mathbf{P_A}\mathbf{y}$ is equal, where $\mathbf{y} \in \mathcal{R}(\mathbf{A})$.
Then we can say

$$\mathbf{y} = \mathbf{P_A}\mathbf{y} = \mathbf{P_A}(\mathbf{A}\mathbf{x}), \mathbf{A}\mathbf{x} \in \mathbb{R}^m$$

Thus:

$$\mathcal{R}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{P_A}).$$

Since we have shown both $\mathcal{R}(\mathbf{P_A}) \subseteq \mathcal{R}(\mathbf{A})$ and $\mathcal{R}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{P_A})$:

$$\mathcal{R}(\mathbf{P_A}) = \mathcal{R}(\mathbf{A}).$$

Q5. Consider the following problem:

$$\min_{\mathbf{x}_2 \in \mathbb{R}^k} \left(\min_{\mathbf{x}_1 \in \mathbb{R}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}_1 - \mathbf{B}\mathbf{x}_2\|_2^2 \right),$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{m \times k}$. Suppose that $\mathbf{A}$ has full column rank, and $\mathbf{P}_{\mathbf{A}}^{\perp}\mathbf{B}$ has full column rank. Use LS to show that the solution $\mathbf{x}^*$ to the above problem is

$$\mathbf{x}_2^* = (\mathbf{B}^\top \mathbf{P}_{\mathbf{A}}^{\perp} \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{P}_{\mathbf{A}}^{\perp} \mathbf{y}.$$

(20%)

A5.

- First solve LS problem over $\mathbf{x}_1$ for a fixed $\mathbf{x}_2$:

  $$\min_{\mathbf{x}_1} \|\mathbf{y} - \mathbf{A}\mathbf{x}_1 - \mathbf{B}\mathbf{x}_2\|_2^2.$$

  Then the optimal $\mathbf{x}_1$ is given by:

  $$\mathbf{x}_1 \mathbf{A}^\top (\mathbf{y} - \mathbf{A}\mathbf{x}_1 - \mathbf{B}\mathbf{x}_2) = 0, \forall \mathbf{x}_1$$

  Since there is no vector orthogonal to all vectors in a subspace:

  $$\mathbf{A}^\top (\mathbf{y} - \mathbf{A}\mathbf{x}_1 - \mathbf{B}\mathbf{x}_2) = 0.$$

  must hold.

  Rearranging,

  $$\mathbf{A}^\top \mathbf{A}\mathbf{x}_1 = \mathbf{A}^\top (\mathbf{y} - \mathbf{B}\mathbf{x}_2).$$

  Since $\mathbf{A}$ has full column rank, $\mathbf{A}^\top \mathbf{A}$ is non-singular

  Thus:

  $$\mathbf{x}_1^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top (\mathbf{y} - \mathbf{B}\mathbf{x}_2).$$

- Substituting $\mathbf{x}_1^*$ into original formula to solve for $\mathbf{x}_2$:

  $$\mathbf{y} - \mathbf{A}\mathbf{x}_1^* - \mathbf{B}\mathbf{x}_2.$$

  Substituting in $\mathbf{x}_1^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top (\mathbf{y} - \mathbf{B}\mathbf{x}_2)$:

  $$\mathbf{y} - \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top (\mathbf{y} - \mathbf{B}\mathbf{x}_2) - \mathbf{B}\mathbf{x}_2.$$

$$= (\mathbf{y} - \mathbf{B}\mathbf{x}_2) - \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top(\mathbf{y} - \mathbf{B}\mathbf{x}_2)$$

Combining:

$$(\mathbf{I} - \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top)(\mathbf{y} - \mathbf{B}\mathbf{x}_2).$$

Since $\mathbf{P}_{\mathbf{A}}^\perp = \mathbf{I} - \mathbf{P}_{\mathbf{A}} = \mathbf{I} - \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$:

$$\mathbf{y} - \mathbf{A}\mathbf{x}_1^* - \mathbf{B}\mathbf{x}_2 = (\mathbf{I} - \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top)(\mathbf{y} - \mathbf{B}\mathbf{x}_2) = \mathbf{P}_{\mathbf{A}}^\perp(\mathbf{y} - \mathbf{B}\mathbf{x}_2).$$

- The optimization problem now reduces to:

$$\min_{\mathbf{x}_2} \|\mathbf{P}_{\mathbf{A}}^\perp(\mathbf{y} - \mathbf{B}\mathbf{x}_2)\|_2^2.$$

Expanding:

$$\min_{\mathbf{x}_2} \|\mathbf{P}_{\mathbf{A}}^\perp\mathbf{y} - \mathbf{P}_{\mathbf{A}}^\perp\mathbf{B}\mathbf{x}_2\|_2^2.$$

This is a LS problem in $\mathbf{x}_2$, thus:

$$(\mathbf{P}_{\mathbf{A}}^\perp\mathbf{B})^\top(\mathbf{P}_{\mathbf{A}}^\perp\mathbf{y} - \mathbf{P}_{\mathbf{A}}^\perp\mathbf{B}\mathbf{x}_2) = 0.$$

Rearrange:

$$(\mathbf{P}_{\mathbf{A}}^\perp\mathbf{B})^\top\mathbf{P}_{\mathbf{A}}^\perp\mathbf{B}\mathbf{x}_2 = (\mathbf{P}_{\mathbf{A}}^\perp\mathbf{B})^\top\mathbf{P}_{\mathbf{A}}^\perp\mathbf{y}.$$

Since $\mathbf{P}_{\mathbf{A}}^\perp\mathbf{B}$ has full column rank, $(\mathbf{P}_{\mathbf{A}}^\perp\mathbf{B})^\top(\mathbf{P}_{\mathbf{A}}^\perp\mathbf{B})$ is invertible, giving the solution:

$$\mathbf{x}_2^* = (\mathbf{B}^\top\mathbf{P}_{\mathbf{A}}^\perp\mathbf{B})^{-1}\mathbf{B}^\top\mathbf{P}_{\mathbf{A}}^\perp\mathbf{y}.$$

Q6. Consider the following problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad \text{subject to } \mathbf{q}_1^\top\mathbf{x} = 1,$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ has full column rank, and $\mathbf{q}_1 \in \mathbb{R}^n$ has $\|\mathbf{q}_1\|_2 = 1$. Use LS to show that the solution to the above problem is

$$\mathbf{x}^* = \mathbf{q}_1 + \mathbf{Q}_2(\mathbf{Q}_2^\top\mathbf{A}^\top\mathbf{A}\mathbf{Q}_2)^{-1}\mathbf{Q}_2^\top\mathbf{A}^\top(\mathbf{y} - \mathbf{A}\mathbf{q}_1),$$

where $\mathbf{Q}_2 \in \mathbb{R}^{n\times(n-1)}$ is any matrix such that $[\mathbf{q}_1\ \mathbf{Q}_2]$ is an orthogonal matrix. (20%)

A6.

Let $\mathbf{q_1}$:

$$\mathbf{q_1} = [a_1, a_2, \ldots, a_n]^T$$

and $\mathbf{Q_2}$:

$$\mathbf{Q_2} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1(n-1)} \\ b_{21} & b_{22} & \cdots & b_{2(n-1)} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{n(n-1)} \end{bmatrix}$$

Then $\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{Q}_2]$:

$$\mathbf{Q} = [\mathbf{q}_1 \ \mathbf{Q}_2] = \begin{bmatrix} a_1 & b_{11} & b_{12} & \cdots & b_{1(n-1)} \\ a_2 & b_{21} & b_{22} & \cdots & b_{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ a_n & b_{n1} & b_{n2} & \cdots & b_{n(n-1)} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Where, $\mathbf{Q}$ is an orthogonal matrix.

As learned in class, the orthogonal matrix $\mathbf{Q}$ satisfies $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$:

$$\mathbf{Q}^\top \mathbf{Q} = \begin{bmatrix} \mathbf{q}_1^\top \\ \mathbf{Q}_2^\top \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 & \mathbf{Q}_2 \end{bmatrix}.$$

$$= \begin{bmatrix} \mathbf{q}_1^\top \mathbf{q}_1 & \mathbf{q}_1^\top \mathbf{Q}_2 \\ \mathbf{Q}_2^\top \mathbf{q}_1 & \mathbf{Q}_2^\top \mathbf{Q}_2 \end{bmatrix} = \mathbf{I}_n$$

To satisfy this equality following conditions must hold:

$$\mathbf{q}_1^\top \mathbf{q}_1 = 1, (\text{ which is already given})$$

$$\mathbf{q}_1^\top \mathbf{Q}_2 = \mathbf{0}$$

$$\mathbf{Q}_2^\top \mathbf{q}_1 = \mathbf{0}$$

$$\mathbf{Q}_2^\top \mathbf{Q}_2 = \mathbf{I}_{n-1}$$

Thus we can conclude that:

- $\mathbf{Q_2}$ and $\mathbf{q_1}$ are orthogonal to each other.
- $\mathbf{Q_2}$ is semi-orthogonal matrix that has orthonormal columns.(Since it is tall and $\mathbf{Q_2}^T\mathbf{Q_2} = \mathbf{I_{n-1}}$)

Since the columns of the orthogonal matrix are orthonormal, $\mathbf{Q}$ matrix forms a orthonormal basis for $\mathbb{R}^n$ This implies that any vector $\mathbf{x} \in \mathbb{R}^n$ can be uniquely expressed as a linear combination of the columns of $\mathbf{Q}$:

$$\mathbf{x} = \mathbf{q}_1 c + \mathbf{Q}_2 \mathbf{z}, \quad \text{for some } c \in \mathbb{R}, \ \mathbf{z} \in \mathbb{R}^{n-1}.$$

Substituting into the constraint $\mathbf{q_1}^T\mathbf{x} = \mathbf{1}$:

$$\mathbf{q}_1^\top(\mathbf{q}_1 c + \mathbf{Q}_2\mathbf{z}) = 1.$$
$$\mathbf{q}_1^\top\mathbf{q}_1 c + \mathbf{q}_1^\top\mathbf{Q}_2\mathbf{z} = 1.$$

Since we know that $\mathbf{q}_1^\top\mathbf{Q}_2 = 0$, we get:

$$\mathbf{q}_1^\top\mathbf{q}_1 c = 1.$$

Since $\|\mathbf{q}_1\|_2 = 1$:

$$c = 1.$$

Substituting $c = 1$ back, we get:

$$\mathbf{x} = \mathbf{q}_1 + \mathbf{Q}_2\mathbf{z}.$$

Substituting $\mathbf{x} = \mathbf{q}_1 + \mathbf{Q}_2\mathbf{z}$ into the objective function:

$$\min_{\mathbf{z}\in\mathbb{R}^{n-1}} \|\mathbf{y} - \mathbf{A}(\mathbf{q}_1 + \mathbf{Q}_2\mathbf{z})\|_2^2.$$

Expanding the norm:

$$\|\mathbf{y} - \mathbf{A}\mathbf{q}_1 - \mathbf{A}\mathbf{Q}_2\mathbf{z}\|_2^2.$$

Let $\mathbf{b} = \mathbf{y} - \mathbf{A}\mathbf{q}_1$, then:

$$\min_{\mathbf{z}\in\mathbb{R}^{n-1}} \|\mathbf{b} - \mathbf{A}\mathbf{Q}_2\mathbf{z}\|_2^2.$$

This is a standard least squares problem of the form:

$$\min_{\mathbf{z}} \|\mathbf{b} - \mathbf{A}\mathbf{Q}_2\mathbf{z}\|_2^2.$$

The optimal solution for $\mathbf{z}$ satisfies:

$$(\mathbf{A}\mathbf{Q}_2)^\top(\mathbf{A}\mathbf{Q}_2)\mathbf{z} = (\mathbf{A}\mathbf{Q}_2)^\top\mathbf{b}.$$

Now we need to see if $\mathbf{AQ_2}$ is full column rank:

$$\mathbf{AQ_2} \in R^{n\times(n-1)}$$

where $\mathbf{A}$ is full column rank and $\mathbf{Q_2}$ is semi-orthogonal matrix that has orthonormal columns(which is linearly independent):

$$\text{rank}(\mathbf{A}) = n, \text{rank}(\mathbf{Q_2}) = n-1$$

Thus:

$$\text{rank}(\mathbf{AQ_2}) = \min(n, n-1) = n-1$$

Showing that matrix $\mathbf{AQ_2}$ is full column rank, which implies that $(\mathbf{AQ_2}^T)(\mathbf{AQ_2})$ is non-singular.

Thus, it is invertible:

$$\mathbf{z}^* = (\mathbf{Q}_2^\top \mathbf{A}^\top \mathbf{A}\mathbf{Q}_2)^{-1}\mathbf{Q}_2^\top \mathbf{A}^\top(\mathbf{y} - \mathbf{A}\mathbf{q}_1).$$

Finally, substituting $\mathbf{z}^*$ back into $\mathbf{x} = \mathbf{q}_1 + \mathbf{Q}_2\mathbf{z}$:

$$\mathbf{x}^* = \mathbf{q}_1 + \mathbf{Q}_2(\mathbf{Q}_2^\top \mathbf{A}^\top \mathbf{A}\mathbf{Q}_2)^{-1}\mathbf{Q}_2^\top \mathbf{A}^\top(\mathbf{y} - \mathbf{A}\mathbf{q}_1).$$

Q7. Let $\mathbf{A}$ be an $n \times n$ matrix. Verify the following results:

(a) If $\mathbf{A}$ is non-singular and λ is an eigenvalue of $\mathbf{A}$, then λ^{-1} is an eigenvalue of $\mathbf{A}^{-1}$. (5%)

(b) If $\mathbf{A}$ satisfies $\mathbf{a}_i^\top \mathbf{1} = 1$ for all i, then 1 is an eigenvalue of $\mathbf{A}$. (5%)

(c) If $\mathbf{A}$ is real orthogonal or complex unitary, then any eigenvalue λ of $\mathbf{A}$ has $|\lambda| = 1$. (5%)

(d) If $\mathbf{A}$ is complex Hermitian, the quadratic form $\mathbf{x}^\mathsf{H}\mathbf{A}\mathbf{x}$ is real for any $\mathbf{x} \in \mathbb{C}^n$. (5%)

A7.

(a) Let $\mathbf{A}$ be an $n \times n$ non-singular matrix. Suppose λ is an eigenvalue of $\mathbf{A}$, then there exists a nonzero vector $\mathbf{v} \neq 0$ such that:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}.$$

Since $\mathbf{A}$ is invertible, we can multiply both sides of this equation by $\mathbf{A}^{-1}$:

$$\mathbf{A}^{-1}(\mathbf{A}\mathbf{v}) = \mathbf{A}^{-1}(\lambda\mathbf{v}).$$

Using the fact that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$:

$$\mathbf{v} = \lambda\mathbf{A}^{-1}\mathbf{v}.$$

Rearranging:

$$\mathbf{A}^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v}.$$

Thus:

$$\mathbf{A}^{-1}\mathbf{v} = \lambda^{-1}\mathbf{v}.$$

Thus, if λ is an eigenvalue of $\mathbf{A}$, then λ^{-1} is an eigenvalue of $\mathbf{A}^{-1}$

(b) Let $\mathbf{A}$

$$\begin{bmatrix} \mathbf{a_1}^T \\ \mathbf{a_2}^T \\ \vdots \\ \mathbf{a_n}^T \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

where $\mathbf{a}_i^\top \mathbf{1} = 1$

This means :

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Thus:

$$\mathbf{A1} = \mathbf{1}.$$

We can formulate this to:

$$\mathbf{A1} = 1 \cdot \mathbf{1}.$$

Since vector of ones is linearly independent, we can formulate eigenvalue equation:

$$\mathbf{Av} = \lambda \mathbf{v},$$

where the vector $\mathbf{1}$ is the eigenvector corresponding to the eigenvalue $\lambda = 1$.

Thus 1 is an eigenvalue of $\mathbf{A}$

(c)
- Case1: $\mathbf{A}$ is real orthogonal matrix

  We know that by the property of Orthogonal matrix:

  $$\mathbf{A}^\top \mathbf{A} = \mathbf{I}.$$

  Now let λ be an eigenvalue of $\mathbf{A}$, with corresponding eigenvector $\mathbf{x} \neq 0$:

  $$\mathbf{Av} = \lambda \mathbf{v}.$$

  Taking the 2-norm square on both sides:

  $$\|\mathbf{Av}\|_2^2 = \|\lambda \mathbf{v}\|_2^2$$

  We can factor out λ as it is scalar:

  $$\|\mathbf{Av}\|_2 = |\lambda|^2 \|\mathbf{v}\|_2.$$

Since $\mathbf{A}$ is orthogonal matrix, we know that $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, thus:

$$\|\mathbf{Av}\|_2^2 = (\mathbf{v^T A^T})(\mathbf{Av}) = \mathbf{v}^T\mathbf{v} = \|\mathbf{v}\|_2^2$$

Thus, substituting into equation below we derived previously:

$$\|\mathbf{Av}\|_2^2 = |\lambda|^2\|\mathbf{v}\|_2^2$$

We get:

$$\|\mathbf{v}\|_2^2 = |\lambda|^2\|\mathbf{v}\|_2^2$$

Since $\mathbf{v} \neq 0$ and we know $\|\mathbf{v}\|_2^2 > 0$, so we can divide both sides by $\|\mathbf{v}\|_2^2$
Thus we get:

$$|\lambda| = 1.$$

- Case2: $\mathbf{A}$ is complex unitary
  We know that by the property of Unitary matrix:

  $$\mathbf{A}^H\mathbf{A} = \mathbf{I}.$$

  Now let λ be an eigenvalue of $\mathbf{A}$, with corresponding eigenvector $\mathbf{x} \neq 0$:

  $$\mathbf{Av} = \lambda\mathbf{v}.$$

  Taking the 2-norm square on both sides:

  $$\|\mathbf{Av}\|_2^2 = \|\lambda\mathbf{v}\|_2^2$$

  We can factor out λ as it is scalar:

  $$\|\mathbf{Av}\|_2 = |\lambda|^2\|\mathbf{v}\|_2.$$

  Since $\mathbf{A}$ is Unitary matrix, we know that $\mathbf{A}^H\mathbf{A} = \mathbf{I}$, thus:

  $$\|\mathbf{Av}\|_2^2 = (\mathbf{v^H A^H})(\mathbf{Av}) = \mathbf{v}^H\mathbf{v} = \|\mathbf{v}\|_2^2$$

  Thus, substituting into equation below we derived previously:

  $$\|\mathbf{Av}\|_2^2 = |\lambda|^2\|\mathbf{v}\|_2^2$$

  We get:

  $$\|\mathbf{v}\|_2^2 = |\lambda|^2\|\mathbf{v}\|_2^2$$

  Since $\mathbf{v} \neq 0$ and we know $\|\mathbf{v}\|_2^2 > 0$, so we can divide both sides by $\|\mathbf{v}\|_2^2$
  Thus we get:

  $$|\lambda| = 1.$$

(d) A matrix $\mathbf{A}$ is Hermitian if and only if:

$$\mathbf{A}^{\mathsf{H}} = \mathbf{A},$$

Now consider the quadratic form:

$$q = \mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}.$$

let $\mathbf{x}$:

$$\begin{bmatrix} a_1 + ib_1 \\ a_2 + ib_2 \\ \vdots \\ a_n + ib_n \end{bmatrix}$$

then $\mathbf{x}^H$:

$$\mathbf{x}^H = [a_1 - ib_1, a_2 - ib_2, \cdots a_n - ib_n]$$

To check whether this is real, take its complex conjugate:

$$q^* = (\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x})^*$$

here:

$$(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x})^* = [a_1 - ib_1, a_2 - ib_2, \cdots a_n - ib_n]^* \mathbf{A}^* \begin{bmatrix} a_1 + ib_1 \\ a_2 + ib_2 \\ \vdots \\ a_n + ib_n \end{bmatrix}^*$$

$$= [a_1 + ib_1, a_2 + ib_2, \cdots a_n + ib_n] \mathbf{A}^H \begin{bmatrix} a_1 - ib_1 \\ a_2 - ib_2 \\ \vdots \\ a_n - ib_n \end{bmatrix}$$

Since taking conjugate of both $\mathbf{x}^H$ and $\mathbf{x}$ does not affect the total results,

$$(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x})^* = [a_1 - ib_1, a_2 - ib_2, \cdots a_n - ib_n] \mathbf{A}^H \begin{bmatrix} a_1 + ib_1 \\ a_2 + ib_2 \\ \vdots \\ a_n + ib_n \end{bmatrix} = \mathbf{x}^{\mathsf{H}}\mathbf{A}^H\mathbf{x}$$

And since $\mathbf{A} = \mathbf{A}^H$:

$$(\mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x})^* = \mathbf{x}^{\mathsf{H}}\mathbf{A}^H\mathbf{x} = \mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x}$$

Thus:

$$q^* = \mathbf{x}^{\mathsf{H}}\mathbf{A}\mathbf{x} = q.$$

Since a number is real if and only if it is equal to its own conjugate, $q = \mathbf{x}^{\mathbf{H}}\mathbf{A}\mathbf{x}$ is real.