**Department of Electronics and Telecommunication Engineering**

**University of Moratuwa**

**EN4573 - Pattern Recognition and Machine Intelligence**

# Homework Assignment - 1

**Instructor - Dr. Prathapasinghe Dharmawansa**

| **Name** | **Index** |
|---|---|
| H. D. M. Premathilaka | 180497C |

This report is submitted in partial fulfilment of the requirements for the module EN4573 - Pattern Recognition and Machine Intelligence.

Date of Submission: June 24, 2023

## Question 1 - Matrix Derivatives

**Definition 1.1.** For $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^T$ with $(\cdot)^T$ denoting the transpose operator, we define the vector,

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}} \triangleq \begin{pmatrix} \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \cdots & \frac{\partial}{\partial x_n} \end{pmatrix}^T$$

**Definition 1.2.** For a scalar $x$ and matrix $\mathbf{S} \in \mathbb{R}^{m \times n}$, we define the matrix,

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}x} \triangleq \begin{pmatrix} \frac{\mathrm{d}s_{11}}{\mathrm{d}x} & \frac{\mathrm{d}s_{12}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}s_{1n}}{\mathrm{d}x} \\ \frac{\mathrm{d}s_{21}}{\mathrm{d}x} & \frac{\mathrm{d}s_{22}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}s_{2n}}{\mathrm{d}x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}s_{m1}}{\mathrm{d}x} & \frac{\mathrm{d}s_{m2}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}s_{mn}}{\mathrm{d}x} \end{pmatrix} = \left(\frac{\mathrm{d}s_{ij}}{\mathrm{d}x}\right)_{\substack{1 \le i \le m \\ 1 \le j \le n}}$$

### 01). a).

Consider $\mathbf{P} \in \mathbb{R}^{m \times n}$ and $\mathbf{Q} \in \mathbb{R}^{n \times q}$. Let $x$ be a scalar. According to Definition 1.2, we can write,

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}x} = \begin{pmatrix} \frac{\mathrm{d}p_{11}}{\mathrm{d}x} & \frac{\mathrm{d}p_{12}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}p_{1n}}{\mathrm{d}x} \\ \frac{\mathrm{d}p_{21}}{\mathrm{d}x} & \frac{\mathrm{d}p_{22}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}p_{2n}}{\mathrm{d}x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}p_{m1}}{\mathrm{d}x} & \frac{\mathrm{d}p_{m2}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}p_{mn}}{\mathrm{d}x} \end{pmatrix} \tag{1.1}$$

$$\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}x} = \begin{pmatrix} \frac{\mathrm{d}q_{11}}{\mathrm{d}x} & \frac{\mathrm{d}q_{12}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}q_{1q}}{\mathrm{d}x} \\ \frac{\mathrm{d}q_{21}}{\mathrm{d}x} & \frac{\mathrm{d}q_{22}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}q_{2q}}{\mathrm{d}x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}q_{n1}}{\mathrm{d}x} & \frac{\mathrm{d}q_{n2}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}q_{nq}}{\mathrm{d}x} \end{pmatrix} \tag{1.2}$$

Now, let's consider $\mathbf{PQ} \in \mathbb{R}^{m \times q}$ and then apply Definition 1.2,

$$\mathbf{PQ} = \begin{pmatrix} \sum_{i=1}^{n} p_{1i}q_{i1} & \sum_{i=1}^{n} p_{1i}q_{i2} & \cdots & \sum_{i=1}^{n} p_{1i}q_{iq} \\ \sum_{i=1}^{n} p_{2i}q_{i1} & \sum_{i=1}^{n} p_{2i}q_{i2} & \cdots & \sum_{i=1}^{n} p_{2i}q_{iq} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} p_{mi}q_{i1} & \sum_{i=1}^{n} p_{mi}q_{i2} & \cdots & \sum_{i=1}^{n} p_{mi}q_{iq} \end{pmatrix}$$

$$
\begin{aligned}
\frac{\mathrm{d}(\mathbf{PQ})}{\mathrm{d}x} &= \begin{pmatrix}
\dfrac{\mathrm{d}\left(\sum\limits_{i=1}^{n} p_{1i}q_{i1}\right)}{\mathrm{d}x} & \dfrac{\mathrm{d}\left(\sum\limits_{i=1}^{n} p_{1i}q_{i2}\right)}{\mathrm{d}x} & \cdots & \dfrac{\mathrm{d}\left(\sum\limits_{i=1}^{n} p_{1i}q_{iq}\right)}{\mathrm{d}x} \\
\dfrac{\mathrm{d}\left(\sum\limits_{i=1}^{n} p_{2i}q_{i1}\right)}{\mathrm{d}x} & \dfrac{\mathrm{d}\left(\sum\limits_{i=1}^{n} p_{2i}q_{i2}\right)}{\mathrm{d}x} & \cdots & \dfrac{\mathrm{d}\left(\sum\limits_{i=1}^{n} p_{2i}q_{iq}\right)}{\mathrm{d}x} \\
\vdots & \vdots & \ddots & \vdots \\
\dfrac{\mathrm{d}\left(\sum\limits_{i=1}^{n} p_{mi}q_{i1}\right)}{\mathrm{d}x} & \dfrac{\mathrm{d}\left(\sum\limits_{i=1}^{n} p_{mi}q_{i2}\right)}{\mathrm{d}x} & \cdots & \dfrac{\mathrm{d}\left(\sum\limits_{i=1}^{n} p_{mi}q_{iq}\right)}{\mathrm{d}x}
\end{pmatrix} \\
&= \begin{pmatrix}
\sum\limits_{i=1}^{n}\left(\dfrac{\mathrm{d}\,p_{1i}q_{i1}}{\mathrm{d}x}\right) & \sum\limits_{i=1}^{n}\left(\dfrac{\mathrm{d}\,p_{1i}q_{i2}}{\mathrm{d}x}\right) & \cdots & \sum\limits_{i=1}^{n}\left(\dfrac{\mathrm{d}\,p_{1i}q_{iq}}{\mathrm{d}x}\right) \\
\sum\limits_{i=1}^{n}\left(\dfrac{\mathrm{d}\,p_{2i}q_{i1}}{\mathrm{d}x}\right) & \sum\limits_{i=1}^{n}\left(\dfrac{\mathrm{d}\,p_{2i}q_{i2}}{\mathrm{d}x}\right) & \cdots & \sum\limits_{i=1}^{n}\left(\dfrac{\mathrm{d}\,p_{2i}q_{iq}}{\mathrm{d}x}\right) \\
\vdots & \vdots & \ddots & \vdots \\
\sum\limits_{i=1}^{n}\left(\dfrac{\mathrm{d}\,p_{mi}q_{i1}}{\mathrm{d}x}\right) & \sum\limits_{i=1}^{n}\left(\dfrac{\mathrm{d}\,p_{mi}q_{i2}}{\mathrm{d}x}\right) & \cdots & \sum\limits_{i=1}^{n}\left(\dfrac{\mathrm{d}\,p_{mi}q_{iq}}{\mathrm{d}x}\right)
\end{pmatrix} \\
&= \begin{pmatrix}
\sum\limits_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{1i}}{\mathrm{d}x}q_{i1} + p_{1i}\frac{\mathrm{d}\,q_{i1}}{\mathrm{d}x}\right) & \sum\limits_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{1i}}{\mathrm{d}x}q_{i2} + p_{1i}\frac{\mathrm{d}\,q_{i2}}{\mathrm{d}x}\right) & \cdots & \sum\limits_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{1i}}{\mathrm{d}x}q_{iq} + p_{1i}\frac{\mathrm{d}\,q_{iq}}{\mathrm{d}x}\right) \\
\sum\limits_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{2i}}{\mathrm{d}x}q_{i1} + p_{2i}\frac{\mathrm{d}\,q_{i1}}{\mathrm{d}x}\right) & \sum\limits_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{2i}}{\mathrm{d}x}q_{i2} + p_{2i}\frac{\mathrm{d}\,q_{i2}}{\mathrm{d}x}\right) & \cdots & \sum\limits_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{2i}}{\mathrm{d}x}q_{iq} + p_{2i}\frac{\mathrm{d}\,q_{iq}}{\mathrm{d}x}\right) \\
\vdots & \vdots & \ddots & \vdots \\
\sum\limits_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{mi}}{\mathrm{d}x}q_{i1} + p_{mi}\frac{\mathrm{d}\,q_{i1}}{\mathrm{d}x}\right) & \sum\limits_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{mi}}{\mathrm{d}x}q_{i2} + p_{mi}\frac{\mathrm{d}\,q_{i2}}{\mathrm{d}x}\right) & \cdots & \sum\limits_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{mi}}{\mathrm{d}x}q_{iq} + p_{mi}\frac{\mathrm{d}\,q_{iq}}{\mathrm{d}x}\right)
\end{pmatrix}
\end{aligned}
$$

$$\frac{\mathrm{d}(\mathbf{PQ})}{\mathrm{d}x} = \begin{pmatrix} \sum_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{1i}}{\mathrm{d}x}q_{i1}\right) & \sum_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{1i}}{\mathrm{d}x}q_{i2}\right) & \cdots & \sum_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{1i}}{\mathrm{d}x}q_{iq}\right) \\ \sum_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{2i}}{\mathrm{d}x}q_{i1}\right) & \sum_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{2i}}{\mathrm{d}x}q_{i2}\right) & \cdots & \sum_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{2i}}{\mathrm{d}x}q_{iq}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{mi}}{\mathrm{d}x}q_{i1}\right) & \sum_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{mi}}{\mathrm{d}x}q_{i2}\right) & \cdots & \sum_{i=1}^{n}\left(\frac{\mathrm{d}\,p_{mi}}{\mathrm{d}x}q_{iq}\right) \end{pmatrix}$$

$$+ \begin{pmatrix} \sum_{i=1}^{n}\left(p_{1i}\frac{\mathrm{d}\,q_{i1}}{\mathrm{d}x}\right) & \sum_{i=1}^{n}\left(p_{1i}\frac{\mathrm{d}\,q_{i2}}{\mathrm{d}x}\right) & \cdots & \sum_{i=1}^{n}\left(p_{1i}\frac{\mathrm{d}\,q_{iq}}{\mathrm{d}x}\right) \\ \sum_{i=1}^{n}\left(p_{2i}\frac{\mathrm{d}\,q_{i1}}{\mathrm{d}x}\right) & \sum_{i=1}^{n}\left(p_{2i}\frac{\mathrm{d}\,q_{i2}}{\mathrm{d}x}\right) & \cdots & \sum_{i=1}^{n}\left(p_{2i}\frac{\mathrm{d}\,q_{iq}}{\mathrm{d}x}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n}\left(p_{mi}\frac{\mathrm{d}\,q_{i1}}{\mathrm{d}x}\right) & \sum_{i=1}^{n}\left(p_{mi}\frac{\mathrm{d}\,q_{i2}}{\mathrm{d}x}\right) & \cdots & \sum_{i=1}^{n}\left(p_{mi}\frac{\mathrm{d}\,q_{iq}}{\mathrm{d}x}\right) \end{pmatrix}$$

$$\frac{\mathrm{d}(\mathbf{PQ})}{\mathrm{d}x} = \begin{pmatrix} \frac{\mathrm{d}p_{11}}{\mathrm{d}x} & \frac{\mathrm{d}p_{12}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}p_{1n}}{\mathrm{d}x} \\ \frac{\mathrm{d}p_{21}}{\mathrm{d}x} & \frac{\mathrm{d}p_{22}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}p_{2n}}{\mathrm{d}x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}p_{m1}}{\mathrm{d}x} & \frac{\mathrm{d}p_{m2}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}p_{mn}}{\mathrm{d}x} \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1q} \\ q_{21} & q_{22} & \cdots & q_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nq} \end{pmatrix}$$

$$+ \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{pmatrix} \begin{pmatrix} \frac{\mathrm{d}q_{11}}{\mathrm{d}x} & \frac{\mathrm{d}q_{12}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}q_{1q}}{\mathrm{d}x} \\ \frac{\mathrm{d}q_{21}}{\mathrm{d}x} & \frac{\mathrm{d}q_{22}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}q_{2q}}{\mathrm{d}x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}q_{n1}}{\mathrm{d}x} & \frac{\mathrm{d}q_{n2}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}q_{nq}}{\mathrm{d}x} \end{pmatrix} \tag{1.3}$$

Substituting Eq. 1.1 and Eq. 1.2 in Eq. 1.3

$$\frac{\mathrm{d}(\mathbf{PQ})}{\mathrm{d}x} = \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}x}\mathbf{Q} + \mathbf{P}\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}x}$$

## 01). b).

Let's consider an orthogonal matrix $\mathbf{\Phi} \in \mathbb{R}^{n\times n}$. Then, $\mathbf{\Phi}\mathbf{\Phi}^T = \mathbf{\Phi}^T\mathbf{\Phi} = \mathbf{I}_n$
From Q1). a).,

$$\frac{\mathrm{d}(\mathbf{PQ})}{\mathrm{d}x} = \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}x}\mathbf{Q} + \mathbf{P}\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}x}$$

Accordingly, we can write,

$$\frac{\mathrm{d}(\mathbf{\Phi}\mathbf{\Phi}^T)}{\mathrm{d}x} = \frac{\mathrm{d}\mathbf{\Phi}}{\mathrm{d}x}\mathbf{\Phi}^T + \mathbf{\Phi}\frac{\mathrm{d}\mathbf{\Phi}^T}{\mathrm{d}x} \tag{1.4}$$

Moreover,

$$\frac{\mathrm{d}(\mathbf{\Phi}\mathbf{\Phi}^T)}{\mathrm{d}x} = \frac{\mathrm{d}\mathbf{I}_n}{\mathrm{d}x} = \mathbf{0} \tag{1.5}$$

From Eq. 1.4 and Eq. 1.5,

$$\frac{\mathrm{d}\mathbf{\Phi}}{\mathrm{d}x}\mathbf{\Phi}^T + \mathbf{\Phi}\frac{\mathrm{d}\mathbf{\Phi}^T}{\mathrm{d}x} = \mathbf{0}$$

## 01). c).

Take a symmetric positive definite matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$.

$$\frac{\mathrm{d}\ln(\det(\mathbf{A}))}{\mathrm{d}x} = \frac{1}{\det(\mathbf{A})}\frac{\mathrm{d}\det(\mathbf{A})}{\mathrm{d}x} \tag{1.6}$$

An expression for the determinant could be written using Laplace expansion.

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij}C_{ij} = \sum_{j=1}^{n}(-1)^{i+j}a_{ij}M_{ij}$$

where $C_{ij}$ and $M_{ij}$ represent the cofactor and minor of $a_{ij}$ ($i^{th}$ row $j^{th}$ column element of A), respectively.

$$\frac{\mathrm{d}\det(\mathbf{A})}{\mathrm{d}x} = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial\det(\mathbf{A})}{\partial a_{ij}}\frac{\partial a_{ij}}{\partial x} \tag{1.7}$$

$$\begin{aligned}
\frac{\partial\det(\mathbf{A})}{\partial a_{ij}} &= \frac{\partial\left(\sum\limits_{k=1}^{n} a_{ik}C_{ik}\right)}{\partial a_{ij}} \\
&= \sum_{k=1}^{n}\frac{\partial\left(a_{ik}C_{ik}\right)}{\partial a_{ij}} \\
&= \sum_{k=1}^{n}\left(\frac{\partial a_{ik}}{\partial a_{ij}}C_{ik} + a_{ik}\frac{\partial C_{ik}}{\partial a_{ij}}\right) \\
&= \sum_{k=1}^{n}\left((-1)^{i+k}M_{ik}\frac{\partial a_{ik}}{\partial a_{ij}} + a_{ik}\frac{\partial\left((-1)^{i+k}M_{ik}\right)}{\partial a_{ij}}\right)
\end{aligned} \tag{1.8}$$

$\forall k = 1, 2, \cdots, n, M_{ik}$ doesn't depend on any of the elements in the $i^{th}$ row of $\mathbf{A}$. Thus, for a given $j = 1, 2, \cdots, n$, $\dfrac{\partial\left((-1)^{i+k}M_{ik}\right)}{\partial a_{ij}} = 0$. Accordingly, Eq. 1.8 will be reduced to,

$$\begin{aligned}\frac{\partial \det(\mathbf{A})}{\partial a_{ij}} &= \sum_{k=1}^{n}\left((-1)^{i+k}M_{ik}\frac{\partial a_{ik}}{\partial a_{ij}}\right)\\ \frac{\partial \det(\mathbf{A})}{\partial a_{ij}} &= (-1)^{i+j}M_{ij}\end{aligned}$$

The adjoint of a matrix is the transpose of its cofactor matrix. That is, $(\mathrm{Adj}(\mathbf{A}))_{ji} = (-1)^{i+j}M_{ij}$. Then we get,

$$\frac{\partial \det(\mathbf{A})}{\partial a_{ij}} = (\mathrm{Adj}(\mathbf{A}))_{ji} \tag{1.9}$$

Substituting Eq. 1.9 in Eq. 1.7,

$$\frac{\mathrm{d}\det(\mathbf{A})}{\mathrm{d}x} = \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathrm{Adj}(\mathbf{A}))_{ji}\frac{\partial a_{ij}}{\partial x} \tag{1.10}$$

Substituting Eq. 1.10 in Eq. 1.6,

$$\begin{aligned}\frac{\mathrm{d}\ln(\det(\mathbf{A}))}{\mathrm{d}x} &= \frac{1}{\det(\mathbf{A})}\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathrm{Adj}(\mathbf{A}))_{ji}\frac{\partial a_{ij}}{\partial x}\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{A}^{-1})_{ji}\frac{\partial a_{ij}}{\partial x}\\ \frac{\mathrm{d}\ln(\det(\mathbf{A}))}{\mathrm{d}x} &= \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{A}^{-1})_{ji}\frac{\mathrm{d}\,a_{ij}}{\mathrm{d}x} \qquad (1.11)\end{aligned}$$

Let's consider $\mathbf{B} = \mathbf{A}^{-1}\dfrac{\mathrm{d}\mathbf{A}}{\mathrm{d}x}$.

$$\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n}\\ b_{21} & b_{22} & \cdots & b_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ b_{n1} & b_{n2} & \cdots & b_{nn}\end{pmatrix} = \mathbf{A}^{-1}\begin{pmatrix} \frac{\mathrm{d}a_{11}}{\mathrm{d}x} & \frac{\mathrm{d}a_{12}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}a_{1n}}{\mathrm{d}x}\\ \frac{\mathrm{d}a_{21}}{\mathrm{d}x} & \frac{\mathrm{d}a_{22}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}a_{2n}}{\mathrm{d}x}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\mathrm{d}a_{n1}}{\mathrm{d}x} & \frac{\mathrm{d}a_{n2}}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}a_{nn}}{\mathrm{d}x}\end{pmatrix}$$

$$b_{ii} = \sum_{j=1}^{n}(\mathbf{A}^{-1})_{ij}\frac{\mathrm{d}\,a_{ji}}{\mathrm{d}x} \qquad (\forall i = 1, 2, \cdots, n)$$

Since $\mathbf{A}$ is a symmetric matrix,

$$\begin{aligned}b_{ii} &= \sum_{j=1}^{n}(\mathbf{A}^{-1})_{ij}\frac{\mathrm{d}\,a_{ji}}{\mathrm{d}x}\\ &= \sum_{j=1}^{n}(\mathbf{A}^{-1})_{ji}\frac{\mathrm{d}\,a_{ij}}{\mathrm{d}x}\end{aligned}$$

$$\operatorname{tr}\left(\mathbf{A}^{-1}\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}x}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{A}^{-1})_{ji}\frac{\mathrm{d}\,a_{ij}}{\mathrm{d}x} \tag{1.12}$$

From Eq. 1.11 and Eq. 1.12,

$$\frac{\mathrm{d}\ln(\det(\mathbf{A}))}{\mathrm{d}x} = \operatorname{tr}\left(\mathbf{A}^{-1}\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}x}\right)$$

### 01). d).

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n\times 1}$. That is, $\mathbf{u}^T = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix}$ and $\mathbf{v}^T = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}$. Using Definition 1.1,

$$\frac{\mathrm{d}\,\mathbf{u}^T}{\mathrm{d}\mathbf{x}} = \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_2}{\partial x_1} & \cdots & \frac{\partial u_n}{\partial x_1} \\ \frac{\partial u_1}{\partial x_2} & \frac{\partial u_2}{\partial x_2} & \cdots & \frac{\partial u_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_1}{\partial x_n} & \frac{\partial u_2}{\partial x_n} & \cdots & \frac{\partial u_n}{\partial x_n} \end{pmatrix} \tag{1.13}$$

$$\frac{\mathrm{d}\,\mathbf{v}^T}{\mathrm{d}\mathbf{x}} = \begin{pmatrix} \frac{\partial v_1}{\partial x_1} & \frac{\partial v_2}{\partial x_1} & \cdots & \frac{\partial v_n}{\partial x_1} \\ \frac{\partial v_1}{\partial x_2} & \frac{\partial v_2}{\partial x_2} & \cdots & \frac{\partial v_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial v_1}{\partial x_n} & \frac{\partial v_2}{\partial x_n} & \cdots & \frac{\partial v_n}{\partial x_n} \end{pmatrix} \tag{1.14}$$

Moreover, we can write,

$$\frac{\mathrm{d}\,\mathbf{u}^T\mathbf{v}}{\mathrm{d}\mathbf{x}} = \begin{pmatrix} \frac{\partial \sum\limits_{i=1}^{n} u_i v_i}{\partial x_1} \\ \frac{\partial \sum\limits_{i=1}^{n} u_i v_i}{\partial x_2} \\ \vdots \\ \frac{\partial \sum\limits_{i=1}^{n} u_i v_i}{\partial x_n} \end{pmatrix}$$

$$
\begin{aligned}
\frac{\mathrm{d}\,\mathbf{u}^T\mathbf{v}}{\mathrm{d}\mathbf{x}} &= \begin{pmatrix} \sum_{i=1}^{n}\left(\frac{\partial u_i}{\partial x_1}v_i + \frac{\partial v_i}{\partial x_1}u_i\right) \\ \sum_{i=1}^{n}\left(\frac{\partial u_i}{\partial x_2}v_i + \frac{\partial v_i}{\partial x_2}u_i\right) \\ \vdots \\ \sum_{i=1}^{n}\left(\frac{\partial u_i}{\partial x_n}v_i + \frac{\partial v_i}{\partial x_n}u_i\right) \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{n}\left(\frac{\partial u_i}{\partial x_1}v_i\right) \\ \sum_{i=1}^{n}\left(\frac{\partial u_i}{\partial x_2}v_i\right) \\ \vdots \\ \sum_{i=1}^{n}\left(\frac{\partial u_i}{\partial x_n}v_i\right) \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^{n}\left(\frac{\partial v_i}{\partial x_1}u_i\right) \\ \sum_{i=1}^{n}\left(\frac{\partial v_i}{\partial x_2}u_i\right) \\ \vdots \\ \sum_{i=1}^{n}\left(\frac{\partial v_i}{\partial x_n}u_i\right) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_2}{\partial x_1} & \cdots & \frac{\partial u_n}{\partial x_1} \\ \frac{\partial u_1}{\partial x_2} & \frac{\partial u_2}{\partial x_2} & \cdots & \frac{\partial u_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_1}{\partial x_n} & \frac{\partial u_2}{\partial x_n} & \cdots & \frac{\partial u_n}{\partial x_n} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} \frac{\partial v_1}{\partial x_1} & \frac{\partial v_2}{\partial x_1} & \cdots & \frac{\partial v_n}{\partial x_1} \\ \frac{\partial v_1}{\partial x_2} & \frac{\partial v_2}{\partial x_2} & \cdots & \frac{\partial v_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial v_1}{\partial x_n} & \frac{\partial v_2}{\partial x_n} & \cdots & \frac{\partial v_n}{\partial x_n} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}
\end{aligned} \tag{1.15}
$$

Substituting Eq. 1.13 and Eq. 1.14 in Eq. 1.15,

$$
\frac{\mathrm{d}\,\mathbf{u}^T\mathbf{v}}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{d}\,\mathbf{u}^T}{\mathrm{d}\mathbf{x}}\mathbf{v} + \frac{\mathrm{d}\,\mathbf{v}^T}{\mathrm{d}\mathbf{x}}\mathbf{u}
$$

## 01). e).

Let $\mathbf{x} \in \mathbb{R}^{n\times 1}$. That is, $\mathbf{x}^T = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$. Using Definition 1.1,

$$\frac{\mathrm{d}\,\mathbf{x}^T}{\mathrm{d}\mathbf{x}} = \begin{pmatrix} \frac{\partial x_1}{\partial x_1} & \frac{\partial x_2}{\partial x_1} & \cdots & \frac{\partial x_n}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2} & \frac{\partial x_2}{\partial x_2} & \cdots & \frac{\partial x_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial x_n} & \frac{\partial x_2}{\partial x_n} & \cdots & \frac{\partial x_n}{\partial x_n} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

$$\frac{\mathrm{d}\,\mathbf{x}^T}{\mathrm{d}\mathbf{x}} = \mathbf{I}_n$$

## 01). f).

Let $\mathbf{x}, \mathbf{u} \in \mathbb{R}^{n\times 1}$. That is, $\mathbf{x}^T = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$ and $\mathbf{u}^T = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix}$. Further, suppose that $\mathbf{u}$ is not a function of $\mathbf{x}$.
That is,

$$\frac{\mathrm{d}\,\mathbf{u}^T}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{d}\,\mathbf{u}}{\mathrm{d}\mathbf{x}^T} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{n\times n} = \mathbf{0}$$

From 01). d).,

$$\frac{\mathrm{d}\,\mathbf{u}^T\mathbf{v}}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{d}\,\mathbf{u}^T}{\mathrm{d}\mathbf{x}}\mathbf{v} + \frac{\mathrm{d}\,\mathbf{v}^T}{\mathrm{d}\mathbf{x}}\mathbf{u}$$

From 01). e).,

$$\frac{\mathrm{d}\,\mathbf{x}^T}{\mathrm{d}\mathbf{x}} = \mathbf{I}_n$$

Accordingly, we can write,

$$\frac{\mathrm{d}\,\mathbf{u}^T\mathbf{x}}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{d}\,\mathbf{u}^T}{\mathrm{d}\mathbf{x}}\mathbf{x} + \frac{\mathrm{d}\,\mathbf{x}^T}{\mathrm{d}\mathbf{x}}\mathbf{u}$$

Substituting previous results, we get,

$$
\begin{aligned}
\frac{\mathrm{d}\,\mathbf{u}^T\mathbf{x}}{\mathrm{d}\mathbf{x}} &= \mathbf{0}\mathbf{x} + \mathbf{I}_n\mathbf{u} \\
\frac{\mathrm{d}\,\mathbf{u}^T\mathbf{x}}{\mathrm{d}\mathbf{x}} &= \mathbf{u} \qquad (1.16)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\frac{\mathrm{d}\,\mathbf{x}^T\mathbf{u}}{\mathrm{d}\mathbf{x}} &= \frac{\mathrm{d}\,\mathbf{x}^T}{\mathrm{d}\mathbf{x}}\mathbf{u} + \frac{\mathrm{d}\,\mathbf{u}^T}{\mathrm{d}\mathbf{x}}\mathbf{x} \\
&= \mathbf{I}_n\mathbf{u} + \mathbf{0}\mathbf{x} \\
\frac{\mathrm{d}\,\mathbf{x}^T\mathbf{u}}{\mathrm{d}\mathbf{x}} &= \mathbf{u} \qquad (1.17)
\end{aligned}
$$

From Eq. 1.16 and Eq. 1.17,

$$
\frac{\mathrm{d}\,\mathbf{u}^T\mathbf{x}}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{d}\,\mathbf{x}^T\mathbf{u}}{\mathrm{d}\mathbf{x}} = \mathbf{u}
$$

### 01). g).

Consider the matrix $\mathbf{R} \in \mathbb{R}^{n\times m}$.

$$
\begin{aligned}
\mathbf{x}^T\mathbf{R} &= \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{pmatrix} \\
\mathbf{x}^T\mathbf{R} &= \begin{pmatrix} \sum\limits_{i=1}^{n} x_i r_{i1} & \sum\limits_{i=1}^{n} x_i r_{i2} & \cdots & \sum\limits_{i=1}^{n} x_i r_{im} \end{pmatrix}
\end{aligned}
$$

Applying Definition 1.1,

$$
\frac{\mathrm{d}\,\mathbf{x}^T\mathbf{R}}{\mathrm{d}\mathbf{x}} = \begin{pmatrix}
\frac{\partial \sum\limits_{i=1}^{n} x_i r_{i1}}{\partial x_1} & \frac{\partial \sum\limits_{i=1}^{n} x_i r_{i2}}{\partial x_1} & \cdots & \frac{\partial \sum\limits_{i=1}^{n} x_i r_{im}}{\partial x_1} \\
\frac{\partial \sum\limits_{i=1}^{n} x_i r_{i1}}{\partial x_2} & \frac{\partial \sum\limits_{i=1}^{n} x_i r_{i2}}{\partial x_2} & \cdots & \frac{\partial \sum\limits_{i=1}^{n} x_i r_{im}}{\partial x_2} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial \sum\limits_{i=1}^{n} x_i r_{i1}}{\partial x_n} & \frac{\partial \sum\limits_{i=1}^{n} x_i r_{i2}}{\partial x_n} & \cdots & \frac{\partial \sum\limits_{i=1}^{n} x_i r_{im}}{\partial x_n}
\end{pmatrix}
$$

$$
\frac{\mathrm{d}\,\mathbf{x}^T\mathbf{R}}{\mathrm{d}\mathbf{x}} = \begin{pmatrix}
\sum_{i=1}^{n}\frac{\partial x_i}{\partial x_1}r_{i1} & \sum_{i=1}^{n}\frac{\partial x_i}{\partial x_1}r_{i2} & \cdots & \sum_{i=1}^{n}\frac{\partial x_i}{\partial x_1}r_{im} \\
\sum_{i=1}^{n}\frac{\partial x_i}{\partial x_2}r_{i1} & \sum_{i=1}^{n}\frac{\partial x_i}{\partial x_2}r_{i2} & \cdots & \sum_{i=1}^{n}\frac{\partial x_i}{\partial x_2}r_{im} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{n}\frac{\partial x_i}{\partial x_n}r_{i1} & \sum_{i=1}^{n}\frac{\partial x_i}{\partial x_n}r_{i2} & \cdots & \sum_{i=1}^{n}\frac{\partial x_i}{\partial x_n}r_{im}
\end{pmatrix}
$$

$$
\frac{\partial x_i}{\partial x_j} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{otherwise} \end{cases}
$$

$$
\therefore \quad \frac{\mathrm{d}\,\mathbf{x}^T\mathbf{R}}{\mathrm{d}\mathbf{x}} = \begin{pmatrix}
r_{11} & r_{12} & \cdots & r_{1m} \\
r_{21} & r_{22} & \cdots & r_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
r_{n1} & r_{n2} & \cdots & r_{nm}
\end{pmatrix}
$$

$$
\frac{\mathrm{d}\,\mathbf{x}^T\mathbf{R}}{\mathrm{d}\mathbf{x}} = \mathbf{R}
$$

### 01). h).

Take a symmetric matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$. Then $\forall i, j \in \{1, 2, \cdots, n\}, b_{ij} = b_{ji}$.

$$
\mathbf{x}^T\mathbf{B}\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix}
b_{11} & b_{12} & \cdots & b_{1n} \\
b_{21} & b_{22} & \cdots & b_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
b_{n1} & b_{n2} & \cdots & b_{nn}
\end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}
$$

$$
= \begin{pmatrix} \sum_{j=1}^{n} x_j b_{j1} & \sum_{j=1}^{n} x_j b_{j2} & \cdots & \sum_{j=1}^{n} x_j b_{jn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}
$$

$$
\mathbf{x}^T\mathbf{B}\mathbf{x} = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} x_j b_{ji} \right) x_i
$$

Using Definition 1.1,

$$\frac{\mathrm{d}\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)}{\mathrm{d}\mathbf{x}} = \begin{pmatrix} \dfrac{\partial \sum\limits_{i=1}^{n}\left(\sum\limits_{j=1}^{n} x_j b_{ji}\right) x_i}{\partial x_1} \\ \dfrac{\partial \sum\limits_{i=1}^{n}\left(\sum\limits_{j=1}^{n} x_j b_{ji}\right) x_i}{\partial x_2} \\ \vdots \\ \dfrac{\partial \sum\limits_{i=1}^{n}\left(\sum\limits_{j=1}^{n} x_j b_{ji}\right) x_i}{\partial x_n} \end{pmatrix}$$

$$\frac{\mathrm{d}\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)}{\mathrm{d}\mathbf{x}} = \begin{pmatrix} \sum\limits_{i=1}^{n} \frac{\partial \left(\sum\limits_{j=1}^{n} x_j b_{ji}\right) x_i}{\partial x_1} \\ \sum\limits_{i=1}^{n} \frac{\partial \left(\sum\limits_{j=1}^{n} x_j b_{ji}\right) x_i}{\partial x_2} \\ \vdots \\ \sum\limits_{i=1}^{n} \frac{\partial \left(\sum\limits_{j=1}^{n} x_j b_{ji}\right) x_i}{\partial x_n} \end{pmatrix} \tag{1.18}$$

Let's take any $k = 1, 2, \cdots, n$.

$$\begin{aligned} \sum_{i=1}^{n} \frac{\partial \left(\sum\limits_{j=1}^{n} x_j b_{ji} x_i\right)}{\partial x_k} &= \sum_{i=1}^{n} \left( \frac{\partial x_i}{\partial x_k} \sum_{j=1}^{n} x_j b_{ji} + x_i \frac{\partial \left(\sum\limits_{j=1}^{n} x_j b_{ji}\right)}{\partial x_k} \right) \\ &= \sum_{i=1}^{n} \left( \frac{\partial x_i}{\partial x_k} \sum_{j=1}^{n} x_j b_{ji} + x_i \sum_{j=1}^{n} \frac{\partial x_j}{\partial x_k} b_{ji} \right) \\ &= \sum_{i=1}^{n} \left( \frac{\partial x_i}{\partial x_k} \sum_{j=1}^{n} x_j b_{ji} \right) + \sum_{i=1}^{n} \left( x_i \sum_{j=1}^{n} \frac{\partial x_j}{\partial x_k} b_{ji} \right) \\ &= \sum_{j=1}^{n} x_j b_{jk} + \sum_{i=1}^{n} x_i b_{ki} \end{aligned}$$

Since B is symmetric, $b_{kj} = b_{jk}$.

$$\therefore \quad \sum_{i=1}^{n} \frac{\partial \left(\sum\limits_{j=1}^{n} x_j b_{ji} x_i\right)}{\partial x_k} = \sum_{j=1}^{n} x_j b_{kj} + \sum_{i=1}^{n} x_i b_{ki} \tag{1.19}$$

Note that $i$ and $j$ are dummy variables used as indices of summation. Accordingly, we can rewrite Eq. 1.19 as,

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\partial \left( \sum\limits_{j=1}^{n} x_j b_{ji} x_i \right)}{\partial x_k} &= \sum_{j=1}^{n} x_j b_{kj} + \sum_{j=1}^{n} x_j b_{kj} \\
\sum_{i=1}^{n} \frac{\partial \left( \sum\limits_{j=1}^{n} x_j b_{ji} x_i \right)}{\partial x_k} &= 2 \sum_{j=1}^{n} x_j b_{kj} \qquad (1.20)
\end{aligned}$$

Using Eq. 1.20 in Eq. 1.18,

$$\begin{aligned}
\frac{\mathrm{d}\left(\mathbf{x}^T \mathbf{B} \mathbf{x}\right)}{\mathrm{d}\mathbf{x}} &= \begin{pmatrix} 2 \sum\limits_{j=1}^{n} x_j b_{1j} \\ 2 \sum\limits_{j=1}^{n} x_j b_{2j} \\ \vdots \\ 2 \sum\limits_{j=1}^{n} x_j b_{nj} \end{pmatrix} \\
&= 2 \begin{pmatrix} \sum\limits_{j=1}^{n} b_{1j} x_j \\ \sum\limits_{j=1}^{n} b_{2j} x_j \\ \vdots \\ \sum\limits_{j=1}^{n} b_{nj} x_j \end{pmatrix} \\
\frac{\mathrm{d}\left(\mathbf{x}^T \mathbf{B} \mathbf{x}\right)}{\mathrm{d}\mathbf{x}} &= 2 \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\
\frac{\mathrm{d}\left(\mathbf{x}^T \mathbf{B} \mathbf{x}\right)}{\mathrm{d}\mathbf{x}} &= 2\mathbf{B}\mathbf{x} \qquad (1.21)
\end{aligned}$$

**01). i).**

Take a symmetric matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$. Then $\forall i,j \in \{1,2,\cdots,n\}, b_{ij} = b_{ji}$.
From 01). h).,

$$\frac{\mathrm{d}\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)}{\mathrm{d}\mathbf{x}} = 2\mathbf{B}\mathbf{x}$$

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)\right)^T &= \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(2\mathbf{B}\mathbf{x}\right)^T \\
&= 2\frac{\mathrm{d}\left(\mathbf{B}\mathbf{x}\right)^T}{\mathrm{d}\mathbf{x}} \\
&= 2\frac{\mathrm{d}\left(\mathbf{x}^T\mathbf{B}^T\right)}{\mathrm{d}\mathbf{x}} \\
\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)\right)^T &= 2\frac{\mathrm{d}\left(\mathbf{x}^T\mathbf{B}\right)}{\mathrm{d}\mathbf{x}} \qquad (1.22)
\end{aligned}$$

From 01). g)., we know that $\dfrac{\mathrm{d}\,\mathbf{x}^T\mathbf{R}}{\mathrm{d}\mathbf{x}} = \mathbf{R}$ for any $\mathbf{R} \in \mathbb{R}^{n\times m}$. Accordingly, $\dfrac{\mathrm{d}\,\mathbf{x}^T\mathbf{B}}{\mathrm{d}\mathbf{x}} = \mathbf{B}$ and thus, we can further simplify Eq. 1.22,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)\right)^T &= 2\mathbf{B} \qquad (1.23) \\
\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\right)^T\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right) &= \frac{\mathrm{d}^2}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^T}\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}^T}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right)\right) \\
&= \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}^T}\left(2\mathbf{B}\mathbf{x}\right) \qquad \text{(From 01).h).)} \\
\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\right)^T\left(\mathbf{x}^T\mathbf{B}\mathbf{x}\right) &= 2\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}^T}\left(\mathbf{B}\mathbf{x}\right) \qquad (1.24)
\end{aligned}$$

Let's simplify $\dfrac{\mathrm{d}}{\mathrm{d}\mathbf{x}^T}\left(\mathbf{B}\mathbf{x}\right)$.

$$\begin{aligned}
\mathbf{B}\mathbf{x} &= \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{n} b_{1i}x_i \\ \sum_{i=1}^{n} b_{2i}x_i \\ \vdots \\ \sum_{i=1}^{n} b_{ni}x_i \end{pmatrix}
\end{aligned}$$

Applying Definition 1.1,

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}^T}(\mathbf{Bx}) = \begin{pmatrix} \frac{\partial \sum\limits_{i=1}^{n} b_{1i}x_i}{\partial x_1} & \frac{\partial \sum\limits_{i=1}^{n} b_{1i}x_i}{\partial x_2} & \cdots & \frac{\partial \sum\limits_{i=1}^{n} b_{1i}x_i}{\partial x_n} \\ \frac{\partial \sum\limits_{i=1}^{n} b_{2i}x_i}{\partial x_1} & \frac{\partial \sum\limits_{i=1}^{n} b_{2i}x_i}{\partial x_2} & \cdots & \frac{\partial \sum\limits_{i=1}^{n} b_{2i}x_i}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sum\limits_{i=1}^{n} b_{ni}x_i}{\partial x_1} & \frac{\partial \sum\limits_{i=1}^{n} b_{ni}x_i}{\partial x_2} & \cdots & \frac{\partial \sum\limits_{i=1}^{n} b_{ni}x_i}{\partial x_n} \end{pmatrix}$$

$$= \begin{pmatrix} \sum\limits_{i=1}^{n} \frac{\partial x_i}{\partial x_1} b_{1i} & \sum\limits_{i=1}^{n} \frac{\partial x_i}{\partial x_2} b_{1i} & \cdots & \sum\limits_{i=1}^{n} \frac{\partial x_i}{\partial x_n} b_{1i} \\ \sum\limits_{i=1}^{n} \frac{\partial x_i}{\partial x_1} b_{2i} & \sum\limits_{i=1}^{n} \frac{\partial x_i}{\partial x_2} b_{2i} & \cdots & \sum\limits_{i=1}^{n} \frac{\partial x_i}{\partial x_n} b_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ \sum\limits_{i=1}^{n} \frac{\partial x_i}{\partial x_1} b_{ni} & \sum\limits_{i=1}^{n} \frac{\partial x_i}{\partial x_2} b_{ni} & \cdots & \sum\limits_{i=1}^{n} \frac{\partial x_i}{\partial x_n} b_{ni} \end{pmatrix}$$

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

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}^T}(\mathbf{Bx}) = \mathbf{B} \tag{1.25}$$

Substituting Eq. 1.25 in Eq. 1.24,

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\right)^T (\mathbf{x}^T\mathbf{Bx}) = 2\mathbf{B} \tag{1.26}$$

From Eq. 1.23 and Eq. 1.26,

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}(\mathbf{x}^T\mathbf{Bx})\right)^T = \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\right)^T (\mathbf{x}^T\mathbf{Bx}) = 2\mathbf{B}$$

**01). j).**

Let's consider $\mathbf{x} \in \mathbb{R}^{n\times 1}$ where $\mathbf{x}^T = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$. Moreover, $\|\mathbf{x}\|_2 = \|\mathbf{x}\| = \left(\mathbf{x}^T\mathbf{x}\right)^{\frac{1}{1}} = \left(\sum\limits_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$.

$$\frac{\mathrm{d}\,\|\mathbf{x}\|}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{d}\left(\sum\limits_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}}{\mathrm{d}\mathbf{x}}$$

Using Definition 1.1,

$$\frac{\mathrm{d}\,\|\mathbf{x}\|}{\mathrm{d}\mathbf{x}} = \begin{pmatrix} \dfrac{\partial\left(\sum\limits_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}}{\partial x_1} \\ \dfrac{\partial\left(\sum\limits_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}}{\partial x_2} \\ \vdots \\ \dfrac{\partial\left(\sum\limits_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}}{\partial x_n} \end{pmatrix}$$

$$= \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2\right)^{-\frac{1}{2}} \begin{pmatrix} \dfrac{\partial \sum\limits_{i=1}^{n} x_i^2}{\partial x_1} \\ \dfrac{\partial \sum\limits_{i=1}^{n} x_i^2}{\partial x_2} \\ \vdots \\ \dfrac{\partial \sum\limits_{i=1}^{n} x_i^2}{\partial x_n} \end{pmatrix}$$

$$= \frac{1}{2\|\mathbf{x}\|} \begin{pmatrix} \sum\limits_{i=1}^{n} \dfrac{\partial x_i^2}{\partial x_1} \\ \sum\limits_{i=1}^{n} \dfrac{\partial x_i^2}{\partial x_2} \\ \vdots \\ \sum\limits_{i=1}^{n} \dfrac{\partial x_i^2}{\partial x_n} \end{pmatrix}$$

$$
\begin{aligned}
\frac{\mathrm{d}\,\|\mathbf{x}\|}{\mathrm{d}\mathbf{x}} &= \frac{1}{2\|\mathbf{x}\|}\begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{pmatrix} \\
&= \frac{1}{\|\mathbf{x}\|}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\
\frac{\mathrm{d}\,\|\mathbf{x}\|}{\mathrm{d}\mathbf{x}} &= \frac{\mathbf{x}}{\|\mathbf{x}\|} \qquad (1.27)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}^2(\|\mathbf{x}\|)}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^T} &= \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}^T}\left(\frac{\mathrm{d}\|\mathbf{x}\|}{\mathrm{d}\mathbf{x}}\right) \\
\frac{\mathrm{d}^2(\|\mathbf{x}\|)}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^T} &= \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}^T}\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \qquad \text{(From Eq. 1.27)}
\end{aligned}
$$

Using Definition 1.1,

$$
\begin{aligned}
\frac{\mathrm{d}^2(\|\mathbf{x}\|)}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^T} &= \begin{pmatrix}
\frac{\partial\left(\frac{x_1}{\|\mathbf{x}\|}\right)}{\partial x_1} & \frac{\partial\left(\frac{x_1}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & \frac{\partial\left(\frac{x_1}{\|\mathbf{x}\|}\right)}{\partial x_n} \\
\frac{\partial\left(\frac{x_2}{\|\mathbf{x}\|}\right)}{\partial x_1} & \frac{\partial\left(\frac{x_2}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & \frac{\partial\left(\frac{x_2}{\|\mathbf{x}\|}\right)}{\partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial\left(\frac{x_n}{\|\mathbf{x}\|}\right)}{\partial x_1} & \frac{\partial\left(\frac{x_n}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & \frac{\partial\left(\frac{x_n}{\|\mathbf{x}\|}\right)}{\partial x_n}
\end{pmatrix} \\
&= \begin{pmatrix}
\frac{1}{\|\mathbf{x}\|}\frac{\partial x_1}{\partial x_1} + x_1\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_1} & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_1}{\partial x_2} + x_1\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_1}{\partial x_n} + x_1\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_n} \\
\frac{1}{\|\mathbf{x}\|}\frac{\partial x_2}{\partial x_1} + x_2\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_1} & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_2}{\partial x_2} + x_2\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_2}{\partial x_n} + x_2\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{\|\mathbf{x}\|}\frac{\partial x_n}{\partial x_1} + x_n\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_1} & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_n}{\partial x_2} + x_n\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_n}{\partial x_n} + x_n\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_n}
\end{pmatrix}
\end{aligned}
$$

$$
\frac{\mathrm{d}^2(\|\mathbf{x}\|)}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^T} = \begin{pmatrix} \frac{1}{\|\mathbf{x}\|}\frac{\partial x_1}{\partial x_1} & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_1}{\partial x_2} & \cdots & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_1}{\partial x_n} \\ \frac{1}{\|\mathbf{x}\|}\frac{\partial x_2}{\partial x_1} & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_2}{\partial x_2} & \cdots & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\|\mathbf{x}\|}\frac{\partial x_n}{\partial x_1} & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_n}{\partial x_2} & \cdots & \frac{1}{\|\mathbf{x}\|}\frac{\partial x_n}{\partial x_n} \end{pmatrix} + \begin{pmatrix} x_1\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_1} & x_1\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & x_1\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_n} \\ x_2\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_1} & x_2\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & x_2\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ x_n\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_1} & x_n\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & x_n\frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_n} \end{pmatrix}
$$

$$
= \frac{1}{\|\mathbf{x}\|}\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} + \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} \frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_1} & \frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & \frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_n} \end{pmatrix}
$$

$$
\frac{\mathrm{d}^2(\|\mathbf{x}\|)}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^T} = \frac{\mathbf{I}_n}{\|\mathbf{x}\|} + \mathbf{x}\frac{\mathrm{d}\left(\frac{1}{\|\mathbf{x}\|}\right)}{\mathrm{d}\mathbf{x}^T} \tag{1.28}
$$

Let's further simplify $\frac{\mathrm{d}\left(\frac{1}{\|\mathbf{x}\|}\right)}{\mathrm{d}\mathbf{x}^T}$.

$$
\begin{aligned}
\frac{\mathrm{d}\left(\frac{1}{\|\mathbf{x}\|}\right)}{\mathrm{d}\mathbf{x}^T} &= \begin{pmatrix} \frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_1} & \frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_2} & \cdots & \frac{\partial\left(\frac{1}{\|\mathbf{x}\|}\right)}{\partial x_n} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial\left(\sum\limits_{i=1}^{n} x_i^2\right)^{-\frac{1}{2}}}{\partial x_1} & \frac{\partial\left(\sum\limits_{i=1}^{n} x_i^2\right)^{-\frac{1}{2}}}{\partial x_2} & \cdots & \frac{\partial\left(\sum\limits_{i=1}^{n} x_i^2\right)^{-\frac{1}{2}}}{\partial x_n} \end{pmatrix} \\
&= -\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2\right)^{-\frac{3}{2}} \begin{pmatrix} \frac{\partial\left(\sum\limits_{i=1}^{n} x_i^2\right)}{\partial x_1} & \frac{\partial\left(\sum\limits_{i=1}^{n} x_i^2\right)}{\partial x_2} & \cdots & \frac{\partial\left(\sum\limits_{i=1}^{n} x_i^2\right)}{\partial x_n} \end{pmatrix} \\
&= -\frac{1}{2\left(\|\mathbf{x}\|\right)^3} \begin{pmatrix} \sum\limits_{i=1}^{n}\left(\frac{\partial x_i^2}{\partial x_1}\right) & \sum\limits_{i=1}^{n}\left(\frac{\partial x_i^2}{\partial x_2}\right) & \cdots & \sum\limits_{i=1}^{n}\left(\frac{\partial x_i^2}{\partial x_n}\right) \end{pmatrix} \\
&= -\frac{1}{2\left(\|\mathbf{x}\|\right)^3} \begin{pmatrix} 2x_1 & 2x_2 & \cdots & 2x_n \end{pmatrix} \\
&= -\frac{1}{\left(\|\mathbf{x}\|\right)^3} \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}
\end{aligned}
$$

$$
\frac{\mathrm{d}\left(\frac{1}{\|\mathbf{x}\|}\right)}{\mathrm{d}\mathbf{x}^T} = -\frac{\mathbf{x}^T}{\left(\|\mathbf{x}\|\right)^3} \tag{1.29}
$$

Using Eq. 1.29 in Eq. 1.28,

$$
\frac{\mathrm{d}^2(\|\mathbf{x}\|)}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^T} = \frac{\mathbf{I}_n}{\|\mathbf{x}\|} + \mathbf{x}\left(-\frac{\mathbf{x}^T}{\left(\|\mathbf{x}\|\right)^3}\right)
$$

$$
\frac{\mathrm{d}^2(\|\mathbf{x}\|)}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^T} = \frac{\mathbf{I}_n}{\|\mathbf{x}\|} - \frac{\mathbf{x}\mathbf{x}^T}{\|\mathbf{x}\|^3} \tag{1.30}
$$

From Eq. 1.27 and Eq. 1.30, it is evident that for a given $\mathbf{x} \in \mathbb{R}^{n\times 1}$,

$$\frac{\mathrm{d}\,\|\mathbf{x}\|}{\mathrm{d}\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$$
$$\frac{\mathrm{d}^2(\|\mathbf{x}\|)}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^T} = \frac{\mathbf{I}_n}{\|\mathbf{x}\|} - \frac{\mathbf{x}\mathbf{x}^T}{\|\mathbf{x}\|^3}$$

### Question 2 - Concentration of Measure/ High Dimensionality

Pick a point $\mathbf{x}$ uniformly at random from the following set in $p$-dimensional space.

$$\mathcal{Q} = \left\{\mathbf{x} : x_1^4 + x_2^4 + \cdots + x_p^4 \leq 1\right\}$$

Suppose the volume enclosed by the set $\mathcal{Q}$ is $V_{Q_p}$. Then, the probability density function of $\mathbf{x}$ could be expressed as,

$$\mathrm{f}(\mathbf{x}) = \begin{cases} \frac{1}{V_{Q_p}} & ; \quad x_1^4 + x_2^4 + \cdots + x_p^4 \leq 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

### 02). a).

Take $t \in [0,1]$. Then we can write,

$$\begin{aligned} \Pr\left\{x_1^4 + x_2^4 + \cdots + x_p^4 \leq 1-t\right\} &= \underset{\sum\limits_{i=1}^{p} x_i^4 \leq 1-t}{\int \cdots \int} \frac{1}{V_{Q_p}}\, dx_1 dx_2 \dots dx_p \\ &= \frac{1}{V_{Q_p}} \underset{\sum\limits_{i=1}^{p} x_i^4 \leq 1-t}{\int \cdots \int} dx_1 dx_2 \dots dx_p \end{aligned}$$

Let's apply the substitution $x_i = (1-t)^{\frac{1}{4}} y_i$ where $i = 1, 2, \cdots, p$. Note that, $dx_i = (1-t)^{\frac{1}{4}} dy_i$.

$$\begin{aligned} \Pr\left\{x_1^4 + x_2^4 + \cdots + x_p^4 \leq 1-t\right\} &= \frac{1}{V_{Q_p}} \underset{\sum\limits_{i=1}^{p} y_i^4 \leq 1}{\int \cdots \int} (1-t)^{\frac{p}{4}}\; dy_1 dy_2 \dots dy_p \\ &= \frac{(1-t)^{\frac{p}{4}}}{V_{Q_p}} \underbrace{\underset{\sum\limits_{i=1}^{p} y_i^4 \leq 1}{\int \cdots \int} dy_1 dy_2 \dots dy_p}_{V_{Q_p}} \\ &= \frac{(1-t)^{\frac{p}{4}}}{V_{Q_p}} V_{Q_p} \end{aligned}$$

$$\Pr\left\{x_1^4 + x_2^4 + \cdots + x_p^4 \leq 1-t\right\} = (1-t)^{\frac{p}{4}} \tag{2.1}$$

Accordingly, when $t = \frac{1}{2}$, we can reduce Eq. 2.1 to,

$$\Pr\left\{x_1^4 + x_2^4 + \cdots + x_p^4 \le 1 - \frac{1}{2}\right\} = \left(1 - \frac{1}{2}\right)^{\frac{p}{4}}$$
$$\Pr\left\{x_1^4 + x_2^4 + \cdots + x_p^4 \le \frac{1}{2}\right\} = \frac{1}{2^{\frac{p}{4}}}$$

## 02). b).

Take $t \in [0, 1]$.

$$\begin{aligned}\Pr\left\{x_1^4 + x_2^4 + \cdots + x_p^4 \ge 1 - t\right\} &= 1 - \Pr\left\{x_1^4 + x_2^4 + \cdots + x_p^4 \le 1 - t\right\} \\ &= 1 - (1 - t)^{\frac{p}{4}} \qquad \text{(From Eq. 2.1)}\end{aligned}$$

Note that $\forall t, (1 - t) \le e^{-t}$. Then for any $p > 0$, $(1 - t)^{\frac{p}{4}} \le e^{-\frac{tp}{4}} \implies 1 - (1 - t)^{\frac{p}{4}} \ge 1 - e^{-\frac{tp}{4}}$. Thus we can write,

$$\Pr\left\{x_1^4 + x_2^4 + \cdots + x_p^4 \ge 1 - t\right\} \ge 1 - e^{-\frac{tp}{4}}$$

Suppose $t$ is of the form $\mathcal{O}\left(p^{-1}\right)$. Then, $t = \frac{\alpha}{p}$ for some $\alpha > 0$.

$$\begin{aligned}\Pr\left\{x_1^4 + x_2^4 + \cdots + x_p^4 \ge 1 - \mathcal{O}\left(p^{-1}\right)\right\} &\ge 1 - e^{-\frac{\alpha p}{4p}} \\ &\ge 1 - \underbrace{e^{-\frac{\alpha}{4}}}_{\delta} \\ \Pr\left\{x_1^4 + x_2^4 + \cdots + x_p^4 \ge 1 - \mathcal{O}\left(p^{-1}\right)\right\} &\ge 1 - \delta\end{aligned}$$

It is clear that $\delta$ diminishes as $\alpha$ increases.
$\therefore$ we can state with high probability, $x_1^4 + x_2^4 + \cdots + x_p^4 \ge 1 - \mathcal{O}\left(p^{-1}\right)$.

## 02). c).

$$\begin{aligned}\Pr\left\{|x_1| \le t\right\} &= \int_{-t}^{t} \left[ \int \cdots \int_{\sum_{j=2}^{p} x_j^4 \le 1 - x_1^4} \frac{1}{V_{Q_p}} \, dx_2 \ldots dx_p \right] dx_1 \\ &= \frac{1}{V_{Q_p}} \int_{-t}^{t} \left[ \int \cdots \int_{\sum_{j=2}^{p} x_j^4 \le 1 - x_1^4} dx_2 \ldots dx_p \right] dx_1\end{aligned}$$

Let's apply the substitution $x_i = (1 - x_1^4)^{\frac{1}{4}} y_{i-1}$ where $i = 2, \cdots, p$. Note that, $dx_i = (1 - x_1^4)^{\frac{1}{4}} dy_{i-1}$.

$$\Pr\{|x_1| \le t\} = \frac{1}{V_{Q_p}} \int\limits_{-t}^{t} \left[ \int_{\sum_{j=1}^{p-1} y_j^4 \le 1} \cdots \int \left(1 - x_1^4\right)^{\frac{p-1}{4}} dy_1 dy_2 \dots dy_{p-1} \right] dx_1$$

$$= \frac{1}{V_{Q_p}} \int\limits_{-t}^{t} \left(1 - x_1^4\right)^{\frac{p-1}{4}} dx_1 \underbrace{\int_{\sum_{j=1}^{p-1} y_j^4 \le 1} \cdots \int dy_1 dy_2 \dots dy_{p-1}}_{V_{Q_{p-1}}}$$

$$\Pr\{|x_1| \le t\} = \frac{V_{Q_{p-1}}}{V_{Q_p}} \int\limits_{-t}^{t} \left(1 - x_1^4\right)^{\frac{p-1}{4}} dx_1$$

$$= \frac{2V_{Q_{p-1}}}{V_{Q_p}} \int\limits_{0}^{t} \left(1 - x_1^4\right)^{\frac{p-1}{4}} dx_1$$

$$= \frac{2V_{Q_{p-1}}}{V_{Q_p}} \left[ \int\limits_{0}^{1} \left(1 - x_1^4\right)^{\frac{p-1}{4}} dx_1 - \int\limits_{t}^{1} \left(1 - x_1^4\right)^{\frac{p-1}{4}} dx_1 \right] \tag{2.2}$$

Now, let's try to evaluate $\frac{2V_{Q_{p-1}}}{V_{Q_p}}$.

$$V_{Q_p} = \int_{\sum_{i=1}^{p} x_i^4 \le 1} \cdots \int dx_1 dx_2 \dots dx_p$$

$$= \int\limits_{-1}^{1} \left[ \int_{\sum_{i=2}^{p} x_i^4 \le 1 - x_1^4} \cdots \int dx_2 \dots dx_p \right] dx_1$$

Let's apply the substitution $x_i = (1 - x_1^4)^{\frac{1}{4}} y_{i-1}$ where $i = 2, \cdots, p$. Note that, $dx_i = (1 - x_1^4)^{\frac{1}{4}} dy_{i-1}$.

$$V_{Q_p} = \int\limits_{-1}^{1} \left[ \int_{\sum_{i=1}^{p-1} y_i^4 \le 1} \cdots \int \left(1 - x_1^4\right)^{\frac{p-1}{4}} dy_1 dy_2 \dots dy_{p-1} \right] dx_1$$

$$= \int\limits_{-1}^{1} \left(1 - x_1^4\right)^{\frac{p-1}{4}} dx_1 \underbrace{\int_{\sum_{i=1}^{p-1} y_i^4 \le 1} \cdots \int dy_1 dy_2 \dots dy_{p-1}}_{V_{Q_{p-1}}}$$

$$
\begin{aligned}
V_{Q_p} &= V_{Q_{p-1}} \int\limits_{-1}^{1} \left(1 - x_1^4\right)^{\frac{p-1}{4}} \, dx_1 \\
&= 2V_{Q_{p-1}} \int\limits_{0}^{1} \left(1 - x_1^4\right)^{\frac{p-1}{4}} \, dx_1 \\
\frac{V_{Q_p}}{2V_{Q_{p-1}}} &= \int\limits_{0}^{1} \left(1 - x_1^4\right)^{\frac{p-1}{4}} \, dx_1
\end{aligned} \tag{2.3}
$$

Consider the substitution $x_1^4 = u$. Then, $dx_1 = \frac{1}{4}u^{-\frac{3}{4}}du$.

$$
\begin{aligned}
\frac{V_{Q_p}}{2V_{Q_{p-1}}} &= \int\limits_{0}^{1} (1-u)^{\frac{p-1}{4}} \frac{1}{4} u^{-\frac{3}{4}} \, du \\
&= \frac{1}{4} \int\limits_{0}^{1} u^{-\frac{3}{4}} (1-u)^{\frac{p-1}{4}} \, du \\
&= \frac{1}{4} \int\limits_{0}^{1} u^{\frac{1}{4}-1} (1-u)^{\frac{p+3}{4}-1} \, du
\end{aligned}
$$

The Beta function can be written in the integral form as $\mathcal{B}(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1} \, \mathrm{d}t$. Moreover, the Beta function can be related to the Gamma function using the equation, $\mathcal{B}(p,q) = \dfrac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$.

Therefore, we can express the above integral in terms of the Beta and Gamma functions.

$$
\frac{V_{Q_p}}{2V_{Q_{p-1}}} = \frac{1}{4}\mathcal{B}\left(\frac{1}{4}, \frac{p+3}{4}\right) = \frac{1}{4} \frac{\Gamma\left(\frac{1}{4}\right)\Gamma\left(\frac{p+3}{4}\right)}{\Gamma\left(\frac{p}{4}+1\right)} \tag{2.4}
$$

Using Eq. 2.3 and 2.4, we can reduce Eq. 2.2 to the following expression.

$$
\begin{aligned}
\Pr\left\{|x_1| \le t\right\} &= 1 - \frac{\int\limits_{t}^{1} \left(1 - x_1^4\right)^{\frac{p-1}{4}} \, dx_1}{\left[\dfrac{\Gamma\left(\frac{1}{4}\right)\Gamma\left(\frac{p+3}{4}\right)}{4\Gamma\left(\frac{p}{4}+1\right)}\right]} \\
&= 1 - \underbrace{\left[\frac{4\Gamma\left(\frac{p}{4}+1\right)}{\Gamma\left(\frac{1}{4}\right)\Gamma\left(\frac{p+3}{4}\right)}\right]}_{\text{say} K (>0)} \int\limits_{t}^{1} \left(1 - x_1^4\right)^{\frac{p-1}{4}} \, dx_1 \\
\Pr\left\{|x_1| \le t\right\} &= 1 - K \int\limits_{t}^{1} \left(1 - x_1^4\right)^{\frac{p-1}{4}} \, dx_1
\end{aligned} \tag{2.5}
$$

Note that $\forall x_1, 1 - x_1^4 \leq e^{-x_1^4}$. Then for any $p > 0$, $(1 - x_1^4)^{\frac{p-1}{4}} \leq e^{-x_1^4\left(\frac{p-1}{4}\right)}$. Thus we can write,

$$\int_t^1 \left(1 - x_1^4\right)^{\frac{p-1}{4}} \, dx_1 \leq \int_t^1 e^{-x_1^4\left(\frac{p-1}{4}\right)} \, dx_1 < \int_t^\infty e^{-x_1^4\left(\frac{p-1}{4}\right)} \, dx_1 \tag{2.6}$$

Let's try to obtain an upper bound for the integral $\int_t^\infty e^{-x_1^4\left(\frac{p-1}{4}\right)} \, dx_1$.

$$\begin{aligned}
\int_t^\infty e^{-x_1^4\left(\frac{p-1}{4}\right)} \, dx_1 &= \int_t^\infty \frac{1}{-x_1^3(p-1)} \frac{\mathrm{d}\, e^{-x_1^4\left(\frac{p-1}{4}\right)}}{\mathrm{d}x_1} \, dx_1 \\
&= \left. \frac{e^{-x_1^4\left(\frac{p-1}{4}\right)}}{-x_1^3(p-1)} \right|_t^\infty - 3 \int_t^\infty \frac{e^{-x_1^4\left(\frac{p-1}{4}\right)}}{x_1^4(p-1)} \, dx_1 \\
&= \frac{e^{-t^4\left(\frac{p-1}{4}\right)}}{t^3(p-1)} - 3 \int_t^\infty \frac{e^{-x_1^4\left(\frac{p-1}{4}\right)}}{x_1^4(p-1)} \, dx_1
\end{aligned}$$

$$\int_t^\infty e^{-x_1^4\left(\frac{p-1}{4}\right)} \, dx_1 < \frac{e^{-t^4\left(\frac{p-1}{4}\right)}}{t^3(p-1)} \tag{2.7}$$

By combining the inequalities 2.6 and 2.7,

$$\int_t^1 \left(1 - x_1^4\right)^{\frac{p-1}{4}} \, dx_1 < \frac{e^{-t^4\left(\frac{p-1}{4}\right)}}{t^3(p-1)} \tag{2.8}$$

Using Eq. 2.5 and inequality 2.8, we can obtain,

$$\begin{aligned}
\Pr\{|x_1| \leq t\} &= 1 - K \int_t^1 \left(1 - x_1^4\right)^{\frac{p-1}{4}} \, dx_1 > 1 - K \frac{e^{-t^4\left(\frac{p-1}{4}\right)}}{t^3(p-1)} \\
\Pr\{|x_1| \leq t\} &> 1 - \left[\frac{4\Gamma\left(\frac{p}{4}+1\right)}{\Gamma\left(\frac{1}{4}\right)\Gamma\left(\frac{p+3}{4}\right)}\right] \frac{e^{-t^4\left(\frac{p-1}{4}\right)}}{t^3(p-1)}
\end{aligned}$$

Suppose $t$ is of the form $\mathcal{O}\left(p^{-\frac{1}{4}}\right)$. Then, $t = \frac{\alpha}{p^{\frac{1}{4}}}$ for some $\alpha > 0$.

$$\begin{aligned}
\Pr\left\{|x_1| \leq \mathcal{O}\left(p^{-\frac{1}{4}}\right)\right\} &> 1 - \left[\frac{4\Gamma\left(\frac{p}{4}+1\right)}{\Gamma\left(\frac{1}{4}\right)\Gamma\left(\frac{p+3}{4}\right)}\right] \frac{e^{-\alpha^4\left(\frac{p-1}{4p}\right)}}{\frac{\alpha^3}{p^{\frac{3}{4}}}(p-1)} \\
&> 1 - \left[\frac{4\Gamma\left(\frac{p}{4}+1\right)}{\Gamma\left(\frac{1}{4}\right)\Gamma\left(\frac{p+3}{4}\right)}\right] \frac{e^{-\frac{\alpha^4}{4}\left(\frac{p-1}{p}\right)}}{p^{\frac{1}{4}}\alpha^3\left(\frac{p-1}{p}\right)}
\end{aligned} \tag{2.9}$$

For large enough $p$, we can say that $\frac{p-1}{p} \approx 1$. Moreover, $\frac{\Gamma\left(\frac{p}{4}+1\right)}{\Gamma\left(\frac{p}{4}+\frac{3}{4}\right)} \approx \left(\frac{p}{4}\right)^{1-\frac{3}{4}}$. That is, $\frac{\Gamma\left(\frac{p}{4}+1\right)}{\Gamma\left(\frac{p}{4}+\frac{3}{4}\right)} \approx \left(\frac{p}{4}\right)^{\frac{1}{4}}$. Accordingly, inequality 2.9 can be simplified to,

$$
\begin{aligned}
\Pr\left\{|x_1| \leq \mathcal{O}\left(p^{-\frac{1}{4}}\right)\right\} &> 1 - \frac{4\left(\frac{p}{4}\right)^{\frac{1}{4}}}{\Gamma\left(\frac{1}{4}\right)} \frac{e^{-\frac{\alpha^4}{4}}}{p^{\frac{1}{4}}\alpha^3} \\
&> 1 - \underbrace{\frac{4^{\frac{3}{4}}}{\Gamma\left(\frac{1}{4}\right)} \frac{e^{-\frac{\alpha^4}{4}}}{\alpha^3}}_{\delta} \\
\Pr\left\{|x_1| \leq \mathcal{O}\left(p^{-\frac{1}{4}}\right)\right\} &> 1 - \delta
\end{aligned}
$$

It is clear that $\delta$ diminishes as $\alpha$ increases.
$\therefore$ we can state with high probability, $|x_1| \leq \mathcal{O}\left(p^{-\frac{1}{4}}\right)$.

## Question 3 - Multivariate Density Functions

Consider the simplex

$$
\mathcal{S} = \left\{\mathbf{x} : x_i \geq 0, 0 \leq i \leq p, \sum_{i=1}^{p} x_i \leq 1\right\}
$$

Suppose the volume enclosed by the set $\mathcal{S}$ is $V_{S_p}$. For a random point $\mathbf{x}$ picked with a uniform density on $\mathcal{S}$, we can express the probability density function as,

$$
\mathrm{f}(\mathbf{x}) = \begin{cases} \frac{1}{V_{S_p}} & ; \quad x_1 + x_2 + \cdots + x_p \leq 1 \\ 0 & ; \quad \text{otherwise} \end{cases}
$$

Let's try to come up with a recursive equation for $V_{S_p}$.

$$
\begin{aligned}
V_{S_p} &= \underset{\sum_{i=1}^{p} x_i \leq 1}{\int \cdots \int} dx_1 dx_2 \dots dx_p \\
&= \int_0^1 \left[ \underset{\sum_{i=2}^{p} x_i \leq 1 - x_1}{\int \cdots \int} dx_2 \dots dx_p \right] dx_1
\end{aligned}
$$

Let's apply the substitution $x_i = (1 - x_1) y_{i-1}$ where $i = 2, \cdots, p$. Note that, $dx_i = (1 - x_1) dy_{i-1}$.

$$
V_{S_p} = \int_0^1 \left[ \underset{\sum_{i=1}^{p-1} y_i \leq 1}{\int \cdots \int} (1 - x_1)^{p-1} \, dy_1 dy_2 \dots dy_{p-1} \right] dx_1
$$

$$V_{S_p} = \int_0^1 (1-x_1)^{p-1}\ dx_1 \underbrace{\int \cdots \int_{\sum_{i=1}^{p-1} y_i \le 1} dy_1 dy_2 \dots dy_{p-1}}_{V_{S_{p-1}}}$$

$$= V_{S_{p-1}} \int_0^1 (1-x_1)^{p-1}\ dx_1$$

$$= V_{S_{p-1}} \left[ \left. \frac{(1-x_1)^p}{-p} \right|_0^1 \right]$$

$$V_{S_p} = \frac{V_{S_{p-1}}}{p} \tag{3.1}$$

Take any $x_j$ where $j = 1, 2, \dots p$.

$$\mathbb{E}\{x_j\} = \int \cdots \int_{\sum_{i=1}^{p} x_i \le 1} x_j \frac{1}{V_{S_p}}\ dx_1 dx_2 \dots dx_p$$

$$= \frac{1}{V_{S_p}} \int_0^1 x_j \left[ \int \cdots \int_{\substack{\sum_{i=1}^{p} x_i \le 1 - x_j \\ i=1 \\ i \ne j}} dx_1 dx_2 \dots dx_{j-1} dx_{j+1} \dots dx_p \right] dx_j$$

From the above equation, it is evident that $\mathbb{E}\{x_1\} = \mathbb{E}\{x_2\} = \cdots = \mathbb{E}\{x_p\}$. Accordingly, let's evaluate $\mathbb{E}\{x_1\}$ and consider the substitution $x_i = (1-x_1)y_{i-1}$ where $i = 2, \cdots, p$. Note that, $dx_i = (1-x_1)dy_{i-1}$.

$$\mathbb{E}\{x_1\} = \frac{1}{V_{S_p}} \int_0^1 x_1 \left[ \int \cdots \int_{\sum_{i=1}^{p-1} y_i \le 1} (1-x_1)^{p-1}\ dy_1 dy_2 \dots dy_{p-1} \right] dx_1$$

$$= \frac{1}{V_{S_p}} \int_0^1 x_1 (1-x_1)^{p-1}\ dx_1 \underbrace{\int \cdots \int_{\sum_{i=1}^{p-1} y_i \le 1} dy_1 dy_2 \dots dy_{p-1}}_{V_{S_{p-1}}}$$

$$= \frac{V_{S_{p-1}}}{V_{S_p}} \int_0^1 x_1 (1-x_1)^{p-1}\ dx_1$$

$$= \frac{V_{S_{p-1}}}{V_{S_p}} \mathcal{B}(2, p)$$

$$
\begin{aligned}
\mathbb{E}\{x_1\} &= p\mathcal{B}\left(2,p\right) && \text{(From Eq. 3.1)}\\
&= p\frac{\Gamma(2)\Gamma(p)}{\Gamma(p+2)}\\
&= \frac{p\Gamma(p)}{(p+1)\Gamma(p+1)} && (\because \Gamma(2)=1)\\
&= \frac{\Gamma(p+1)}{(p+1)\Gamma(p+1)} && (\because p\Gamma(p)=\Gamma(p+1))\\
\mathbb{E}\{x_1\} &= \frac{1}{p+1} && (3.2)
\end{aligned}
$$

Now, we will consider $\mathbb{E}\{x_1+x_2+\cdots+x_p\}$.

$$
\begin{aligned}
\mathbb{E}\{x_1+x_2+\cdots+x_p\} &= \mathbb{E}\{x_1\}+\mathbb{E}\{x_2\}+\cdots+\mathbb{E}\{x_p\}\\
&= p\mathbb{E}\{x_1\}\\
\mathbb{E}\{x_1+x_2+\cdots+x_p\} &= p\left(\frac{1}{p+1}\right) && \text{(From Eq. 3.2)}\\
\mathbb{E}\{x_1+x_2+\cdots+x_p\} &= \frac{p}{p+1}
\end{aligned}
$$

## Question 4 - The Gamma and Beta Functions

The Gamma function assumes the following integral representation.

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\ \mathrm{d}t, \quad \mathrm{Re}(z)>0 \tag{4.1}$$

The domain of the definition of $\Gamma(z)$ can easily be extended to the entire complex plane by means of the following simple relationship.

$$\Gamma(z+1) = z\Gamma(z)$$

Therefore, $\Gamma(z)$ is an analytic everywhere except at the simple poles given by $z=-k, k=0,1,\ldots$. Moreover, we have

$$
\begin{aligned}
\Gamma(n+1) &= n!\\
\Gamma(1/2) &= \sqrt{\pi}
\end{aligned}
$$

Another closely related function is the Beta function which can be written in the integral form as,

$$
\begin{aligned}
\mathcal{B}(p,q) &= \int_0^1 t^{p-1}(1-t)^{q-1}\ \mathrm{d}t, \quad \mathrm{Re}(p,q)>0 && (4.2)\\
&= 2\int_0^{\frac{\pi}{2}} \cos^{2p-1}\theta\sin^{2q-1}\theta\mathrm{d}\theta\\
&= \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}
\end{aligned}
$$

**04). a).**

Let $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ with $\sigma \neq 0$. Then, the probability density function of $X$ is given by $f_X(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Consider a function $g(x)$ which is differentiable such that $|g(X)(X-\mu)|$ and $\left|\frac{\mathrm{d}}{\mathrm{d}X} g(X)\right|$ have finite mathematical expectations.

$$
\begin{aligned}
\mathbb{E}\left\{g(X)(X-\mu)\right\} &= \int\limits_{\mathbb{R}} g(x)(x-\mu) f_X(x) \ \mathrm{d}x \\
&= \int\limits_{-\infty}^{\infty} g(x)(x-\mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ \mathrm{d}x \\
&= -\frac{\sigma}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} g(x) \left( \frac{\mathrm{d}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\mathrm{d}x} \right) \ \mathrm{d}x \\
&= -\frac{\sigma}{\sqrt{2\pi}} \left[ \underbrace{g(x) e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Bigg|_{-\infty}^{\infty}}_{=0} - \int\limits_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left( \frac{\mathrm{d}\, g(x)}{\mathrm{d}x} \right) \ \mathrm{d}x \right] \\
&= \frac{\sigma}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left( \frac{\mathrm{d}\, g(x)}{\mathrm{d}x} \right) \ \mathrm{d}x \\
&= \sigma^2 \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left( \frac{\mathrm{d}\, g(x)}{\mathrm{d}x} \right) \ \mathrm{d}x \\
&= \sigma^2 \int\limits_{-\infty}^{\infty} \frac{\mathrm{d}\, g(x)}{\mathrm{d}x} f_X(x) \ \mathrm{d}x \\
\mathbb{E}\left\{g(X)(X-\mu)\right\} &= \sigma^2 \mathbb{E}\left\{ \frac{\mathrm{d}\, g(X)}{\mathrm{d}X} \right\}
\end{aligned}
\tag{4.3}
$$

**04). b).**

Consider $n \in \mathbb{Z}^+$. Note that we can write $\mathbb{E}\left\{(X-\mu)^{2n}\right\} = \mathbb{E}\left\{(X-\mu)^{2n-1}(X-\mu)\right\}$. Suppose, $g(X) = (X-\mu)^{2n-1}$. Then, from Eq. 4.3,

$$
\begin{aligned}
\mathbb{E}\left\{(X-\mu)^{2n-1}(X-\mu)\right\} &= \sigma^2 \mathbb{E}\left\{ \frac{\mathrm{d}\,(X-\mu)^{2n-1}}{\mathrm{d}X} \right\} \\
\mathbb{E}\left\{(X-\mu)^{2n}\right\} &= \sigma^2 \mathbb{E}\left\{(2n-1)(X-\mu)^{2n-2}\right\} \\
&= \sigma^2 (2n-1) \mathbb{E}\left\{(X-\mu)^{2n-2}\right\} \\
&= \sigma^2 (2n-1) \mathbb{E}\left\{(X-\mu)^{2n-3}(X-\mu)\right\} \\
&= \sigma^2 (2n-1) \sigma^2 \mathbb{E}\left\{ \frac{\mathrm{d}\,(X-\mu)^{2n-3}}{\mathrm{d}X} \right\} && \text{(From Eq. 4.3)}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left\{(X-\mu)^{2n}\right\} &= \sigma^4(2n-1)\mathbb{E}\left\{(2n-3)\,(X-\mu)^{2n-4}\right\} \\
&= \sigma^4(2n-1)(2n-3)\mathbb{E}\left\{(X-\mu)^{2n-4}\right\}
\end{aligned}
$$

By applying Eq. 4.3 repeatedly, we can obtain,

$$
\mathbb{E}\left\{(X-\mu)^{2n}\right\} = \sigma^{2n}(2n-1)(2n-3)\ldots(2n-(2n-1))\underbrace{\mathbb{E}\left\{(X-\mu)^{2n-2n}\right\}}_{\mathbb{E}\{1\}=1}
$$

$$
\mathbb{E}\left\{(X-\mu)^{2n}\right\} = \sigma^{2n}(2n-1)(2n-3)\ldots 1 \tag{4.4}
$$

## 04). c).

$$
\mathbb{E}\left\{(X-\mu)^{2n}\right\} = \int_{-\infty}^{\infty} (x-\mu)^{2n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ \mathrm{d}x
$$

Let's apply the substitution $z = (x-\mu)/\sigma$. Note that, $dx = \sigma dz$.

$$
\begin{aligned}
\mathbb{E}\left\{(X-\mu)^{2n}\right\} &= \int_{-\infty}^{\infty} (\sigma z)^{2n}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\ \mathrm{d}z \\
&= \frac{\sigma^{2n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^{2n}e^{-\frac{z^2}{2}}\ \mathrm{d}z \\
&= \frac{2\sigma^{2n}}{\sqrt{2\pi}}\int_{0}^{\infty} z^{2n}e^{-\frac{z^2}{2}}\ \mathrm{d}z
\end{aligned}
$$

Consider the substitution $t = \frac{z^2}{2}$. Note that, $dz = \frac{1}{\sqrt{2t}}dt$.

$$
\begin{aligned}
\mathbb{E}\left\{(X-\mu)^{2n}\right\} &= \frac{2\sigma^{2n}}{\sqrt{2\pi}}\int_{0}^{\infty}(2t)^n e^{-t}\frac{1}{\sqrt{2t}}\ \mathrm{d}t \\
&= \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\int_{0}^{\infty} t^{\left(n-\frac{1}{2}\right)}e^{-t}\ \mathrm{d}t \\
&= \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\int_{0}^{\infty} t^{\left(n+\frac{1}{2}\right)-1}e^{-t}\ \mathrm{d}t \\
&= \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\Gamma\left(n+\frac{1}{2}\right) && \text{(From Eq. 4.1)} \\
&= \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\left(n-\frac{1}{2}\right)\Gamma\left(n-\frac{1}{2}\right) \\
&= \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\Gamma\left(n-\frac{3}{2}\right) \\
&= \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\ldots\left(n-\frac{2n-1}{2}\right)\Gamma\left(n-\frac{2n-1}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left\{(X-\mu)^{2n}\right\} &= \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\dots\frac{1}{2}\underbrace{\Gamma\left(\frac{1}{2}\right)}_{\sqrt{\pi}} \\
&= \frac{2^n\sigma^{2n}}{\sqrt{\pi}}\left(\frac{2n-1}{2}\right)\left(\frac{2n-3}{2}\right)\dots\frac{1}{2}\sqrt{\pi} \\
&= 2^n\sigma^{2n}\frac{(2n-1)(2n-3)\dots 1}{2^n} \\
\mathbb{E}\left\{(X-\mu)^{2n}\right\} &= \sigma^{2n}(2n-1)(2n-3)\dots 1
\end{aligned}
$$

## 04). d).

Let $X \sim \mathcal{G}(\alpha,\lambda)$ with the p.d.f.

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, \quad x,\lambda,\alpha > 0$$

Let's evaluate $\mathbb{E}\{X\}$.

$$
\begin{aligned}
\mathbb{E}\{X\} &= \int_0^\infty x f_X(x)\ \mathrm{d}x \\
&= \int_0^\infty x\left(\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}\right)\ \mathrm{d}x \\
&= \frac{1}{\Gamma(\alpha)}\int_0^\infty (\lambda x)^\alpha\, e^{-\lambda x}\ \mathrm{d}x
\end{aligned}
$$

Let's apply the substitution $t = \lambda x$. Note that, $dt = \lambda dx$.

$$
\begin{aligned}
\mathbb{E}\{X\} &= \frac{1}{\Gamma(\alpha)}\int_0^\infty t^\alpha e^{-t}\left(\frac{1}{\lambda}\right)\ \mathrm{d}t \\
&= \frac{1}{\lambda\Gamma(\alpha)}\int_0^\infty t^{(\alpha+1)-1}e^{-t}\ \mathrm{d}t \\
&= \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} \\
&= \frac{\alpha\Gamma(\alpha)}{\lambda\Gamma(\alpha)} \\
\mathbb{E}\{X\} &= \frac{\alpha}{\lambda}
\end{aligned}
\tag{4.5}
$$

## 04). e).

Let $\lambda \sim \mathcal{G}(\beta,\mu)$ with the p.d.f.

$$f_\lambda(\lambda) = \frac{\mu^\beta}{\Gamma(\beta)}\lambda^{\beta-1}e^{-\mu\lambda}, \quad \lambda,\mu,\beta > 0$$

$$
\begin{aligned}
f_{\lambda|X}(\lambda|x) &= \frac{f_{\lambda,X}(\lambda,x)}{f_X(x)} \\
&= \frac{f_{X|\lambda}(x|\lambda) f_\lambda(\lambda)}{\int\limits_0^\infty f_{\lambda,X}(\lambda,x)\ \mathrm{d}\lambda}
\end{aligned}
$$

$$
\begin{aligned}
f_{\lambda|X}(\lambda|x) &= \frac{f_{X|\lambda}(x|\lambda) f_\lambda(\lambda)}{\int\limits_0^\infty f_{X|\lambda}(x|\lambda) f_\lambda(\lambda)\ \mathrm{d}\lambda} \\
&= \frac{\left(\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\right)\left(\frac{\mu^\beta}{\Gamma(\beta)} \lambda^{\beta-1} e^{-\mu\lambda}\right)}{\int\limits_0^\infty \left(\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\right)\left(\frac{\mu^\beta}{\Gamma(\beta)} \lambda^{\beta-1} e^{-\mu\lambda}\right)\ \mathrm{d}\lambda} \\
&= \frac{\lambda^{\alpha+\beta-1} e^{-\lambda(x+\mu)}}{\int\limits_0^\infty \lambda^{\alpha+\beta-1} e^{-\lambda(x+\mu)}\ \mathrm{d}\lambda}
\end{aligned}
$$

Let's apply the substitution $t = \lambda\,(x+\mu)$ in the integral. Note that, $dt = (x+\mu)\ d\lambda$.

$$
\begin{aligned}
f_{\lambda|X}(\lambda|x) &= \frac{\lambda^{\alpha+\beta-1} e^{-\lambda(x+\mu)}}{\int\limits_0^\infty \left(\frac{t}{x+\mu}\right)^{\alpha+\beta-1} e^{-t} \frac{1}{x+\mu}\ \mathrm{d}\lambda} \\
&= \frac{(x+\mu)^{\alpha+\beta}\ \lambda^{\alpha+\beta-1} e^{-\lambda(x+\mu)}}{\int\limits_0^\infty t^{(\alpha+\beta)-1} e^{-t}\ \mathrm{d}t}
\end{aligned}
$$

$$
f_{\lambda|X}(\lambda|x) = \frac{(x+\mu)^{\alpha+\beta}\ \lambda^{\alpha+\beta-1} e^{-\lambda(x+\mu)}}{\Gamma\left(\alpha+\beta\right)} \tag{4.6}
$$

Accordingly, from Eq. 4.6, we can realize that $\lambda|X \sim \mathcal{G}(\alpha+\beta, X+\mu)$.

By comparing with the Eq. 4.5, it is clear that $\mathbb{E}\left\{\lambda|X\right\} = \dfrac{\alpha+\beta}{x+\mu}$.

**04). f).**

In the previous part, we showed that, $\lambda|X \sim \mathcal{G}(\alpha+\beta, X+\mu)$. For convenience, we will use $f(\lambda|x)$ to represent $f_{\lambda|X}(\lambda|x)$.

$$\frac{\mathrm{d}\, f(\lambda|x)}{\mathrm{d}\lambda} = \frac{\mathrm{d}\left(\dfrac{(x+\mu)^{\alpha+\beta}\,\lambda^{\alpha+\beta-1}e^{-\lambda(x+\mu)}}{\Gamma(\alpha+\beta)}\right)}{\mathrm{d}\lambda}$$

$$\frac{\mathrm{d}\, f(\lambda|x)}{\mathrm{d}\lambda} = \frac{(x+\mu)^{\alpha+\beta}}{\Gamma(\alpha+\beta)}\left[(\alpha+\beta-1)\lambda^{\alpha+\beta-2}e^{-\lambda(x+\mu)} - \lambda^{\alpha+\beta-1}(x+\mu)e^{-\lambda(x+\mu)}\right]$$

$$\frac{\mathrm{d}\, f(\lambda|x)}{\mathrm{d}\lambda} = \frac{(x+\mu)^{\alpha+\beta}\,\lambda^{\alpha+\beta-2}e^{-\lambda(x+\mu)}}{\Gamma(\alpha+\beta)}\left[(\alpha+\beta-1)-\lambda(x+\mu)\right] \tag{4.7}$$

Let's find the stationary point (i.e. $\lambda$ at which $\frac{\mathrm{d}\, f(\lambda|x)}{\mathrm{d}\lambda}=0$).

$$\underbrace{\frac{(x+\mu)^{\alpha+\beta}\,\lambda^{\alpha+\beta-2}e^{-\lambda(x+\mu)}}{\Gamma(\alpha+\beta)}}_{\neq 0 \quad (\because \quad x,\lambda,\mu,\alpha,\beta>0)}\left[(\alpha+\beta-1)-\lambda(x+\mu)\right] = 0$$

$$(\alpha+\beta-1)-\lambda(x+\mu) = 0$$

$$\lambda = \frac{\alpha+\beta-1}{x+\mu} \tag{4.8}$$

Note that Eq. 4.8 is valid when $\alpha+\beta \geq 1$.

We need to determine the nature of this stationary point. Accordingly, let's use the second derivative test.

$$\begin{aligned}
\frac{\mathrm{d}^2 f(\lambda|x)}{\mathrm{d}\lambda^2} &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\mathrm{d}\, f(\lambda|x)}{\mathrm{d}\lambda}\right)\\
&= \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{(x+\mu)^{\alpha+\beta}\,\lambda^{\alpha+\beta-2}e^{-\lambda(x+\mu)}}{\Gamma(\alpha+\beta)}\left[(\alpha+\beta-1)-\lambda(x+\mu)\right]\right)
\end{aligned}$$

(From Eq. 4.7)

$$\begin{aligned}
\left.\frac{\mathrm{d}^2 f(\lambda|x)}{\mathrm{d}\lambda^2}\right|_{\lambda=\frac{\alpha+\beta-1}{x+\mu}} &= \frac{-(x+\mu)^{\alpha+\beta+1}}{\Gamma(\alpha+\beta)}\left\{\left(\frac{\alpha+\beta-1}{x+\mu}\right)^{\alpha+\beta-2} e^{-(\alpha+\beta-1)}\right\}\\
&= -\frac{(x+\mu)^3(\alpha+\beta-1)^{\alpha+\beta-2}e^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} < 0
\end{aligned}$$

Using Eq. 4.8 and the above inequality, we can say that $\lambda = \dfrac{\alpha+\beta-1}{x+\mu}$ maximizes the p.d.f. of $\lambda|X$.

**04). g).**

Let $X_i \sim \mathcal{G}(\alpha, \lambda)$, $i = 1, 2, \ldots, n$, be i.i.d. random variables. Then,

$$f_{X_i}(x_i) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\lambda x_i}, \quad x_i, \lambda, \alpha > 0$$

$$\begin{aligned}
f_{\lambda|X_1,X_2,\ldots,X_n}(\lambda|x_1, x_2, \ldots, x_n) &= \frac{f_{X_1,X_2,\ldots,X_n|\lambda}(x_1, x_2, \ldots, x_n|\lambda) f_\lambda(\lambda)}{\int\limits_0^\infty f_{X_1,X_2,\ldots,X_n|\lambda}(x_1, x_2, \ldots, x_n|\lambda) f_\lambda(\lambda) \, \mathrm{d}\lambda} \\
&= \frac{\left(\prod\limits_{i=1}^n \frac{\lambda^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\lambda x_i}\right)\left(\frac{\mu^\beta}{\Gamma(\beta)} \lambda^{\beta-1} e^{-\mu\lambda}\right)}{\int\limits_0^\infty \left(\prod\limits_{i=1}^n \frac{\lambda^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\lambda x_i}\right)\left(\frac{\mu^\beta}{\Gamma(\beta)} \lambda^{\beta-1} e^{-\mu\lambda}\right) \, \mathrm{d}\lambda} \\
&= \frac{\lambda^{n\alpha+\beta-1} e^{-\lambda\left(\sum\limits_{i=1}^n x_i + \mu\right)}}{\int\limits_0^\infty \lambda^{n\alpha+\beta-1} e^{-\lambda\left(\sum\limits_{i=1}^n x_i + \mu\right)} \, \mathrm{d}\lambda}
\end{aligned}$$

Let's apply the substitution $t = \lambda \left(\sum\limits_{i=1}^n x_i + \mu\right)$ in the integral. Note that, $dt = \left(\sum\limits_{i=1}^n x_i + \mu\right) d\lambda$.

$$\begin{aligned}
f_{\lambda|X_1,X_2,\ldots,X_n}(\lambda|x_1, x_2, \ldots, x_n) &= \frac{\lambda^{n\alpha+\beta-1} e^{-\lambda\left(\sum\limits_{i=1}^n x_i + \mu\right)}}{\int\limits_0^\infty \frac{t^{n\alpha+\beta-1}}{\left(\sum\limits_{i=1}^n x_i + \mu\right)^{n\alpha+\beta-1}} \frac{e^{-t}}{\sum\limits_{i=1}^n x_i + \mu} \, \mathrm{d}\lambda} \\
&= \frac{\left(\sum\limits_{i=1}^n x_i + \mu\right)^{n\alpha+\beta} \lambda^{n\alpha+\beta-1} e^{-\lambda\left(\sum\limits_{i=1}^n x_i + \mu\right)}}{\int\limits_0^\infty t^{(n\alpha+\beta)-1} e^{-t} \, \mathrm{d}t}
\end{aligned}$$

$$f_{\lambda|X_1,X_2,\ldots,X_n}(\lambda|x_1, x_2, \ldots, x_n) = \frac{\left(\sum\limits_{i=1}^n x_i + \mu\right)^{n\alpha+\beta} \lambda^{n\alpha+\beta-1} e^{-\lambda\left(\sum\limits_{i=1}^n x_i + \mu\right)}}{\Gamma\left(n\alpha + \beta\right)} \tag{4.9}$$

Accordingly, from Eq. 4.9, we can realize that $\lambda|X_1, X_2, \ldots, X_n \sim \mathcal{G}(n\alpha + \beta, \sum\limits_{i=1}^n X_i + \mu)$.

By comparing with the Eq. 4.5, it is clear that $\mathbb{E}\left\{\lambda|X_1, X_2, \ldots, X_n\right\} = \dfrac{n\alpha + \beta}{\sum\limits_{i=1}^n x_i + \mu}$.

Further, we can find the $\lambda$ that maximizes the p.d.f. of $\lambda|X_1, X_2, \ldots, X_n$ by comparing with Eq. 4.8.

Then, we can see that, $\lambda = \dfrac{n\alpha + \beta - 1}{\sum\limits_{i=1}^{n} x_i + \mu}$ maximizes the p.d.f. of $\lambda | X_1, X_2, \ldots, X_n$.

**04). h).**

Suppose $Y \sim \text{Binomial}(N, p)$ has the p.m.f.,

$$\Pr\{Y = k \mid p\} = \begin{pmatrix} N \\ k \end{pmatrix} p^k (1-p)^{N-k}, \quad p \in (0,1), k = 0, 1, \ldots, N$$

Since the likelihood function follows a binomial distribution, we will assume a Beta distribution for the prior p.d.f. Accordingly,

$$\Pr\{p\} = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{\int\limits_0^1 p^{\alpha-1}(1-p)^{\beta-1} \, \mathrm{d}p} \qquad \alpha, \beta > 0$$

$$\Pr\{p\} = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{\mathcal{B}(\alpha, \beta)}$$

Let's see if the assumed prior acts as a conjugate prior by evaluating the posterior density.

$$\begin{aligned}
\Pr\{p \mid Y = k\} &= \frac{\Pr\{Y = k \mid p\} \Pr\{p\}}{\int\limits_0^1 \Pr\{Y = k \mid p\} \Pr\{p\} \, \mathrm{d}p} \\
&= \frac{\begin{pmatrix} N \\ k \end{pmatrix} p^k (1-p)^{N-k} \left(\dfrac{p^{\alpha-1}(1-p)^{\beta-1}}{\mathcal{B}(\alpha, \beta)}\right)}{\int\limits_0^1 \begin{pmatrix} N \\ k \end{pmatrix} p^k (1-p)^{N-k} \left(\dfrac{p^{\alpha-1}(1-p)^{\beta-1}}{\mathcal{B}(\alpha, \beta)}\right) \mathrm{d}p} \\
&= \frac{p^{k+\alpha-1}(1-p)^{N-k+\beta-1}}{\int\limits_0^1 p^{k+\alpha-1}(1-p)^{N-k+\beta-1} \, \mathrm{d}p}
\end{aligned}$$

$$\Pr\{p \mid Y = k\} = \frac{p^{k+\alpha-1}(1-p)^{N-k+\beta-1}}{\mathcal{B}(k+\alpha, N-k+\beta)} \tag{4.10}$$

From Eq. 4.10, it is clear that the posterior distribution ($\Pr\{p \mid Y = k\}$) is in the same distribution as the assumed prior distribution ($\Pr\{p\}$). Accordingly, we can say that the prior and posterior distributions are conjugate distributions.

Therefore, we can take $\Pr\{p\} = \dfrac{p^{\alpha-1}(1-p)^{\beta-1}}{\mathcal{B}(\alpha, \beta)}$ as the conjugate prior p.d.f.

### Question 5 - Multivariate Normal Density

Let $\mathbf{x} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the p.d.f.

$$f_{\mathbf{X}}(\mathbf{x}) = \sqrt{\frac{\det\left(\boldsymbol{\Sigma}^{-1}\right)}{(2\pi)^n}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

where $\boldsymbol{\mu} \in \mathbb{R}^{n\times 1}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix. Here the quantity $(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ is called the Mahalanobis distance or statistical distance from $\boldsymbol{\mu}$ to $\mathbf{x}$.

Suppose $\boldsymbol{\Sigma}$ assumes the Eigen-decomposition $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T$ where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ with $\lambda_i > 0$ and $\mathbf{P}\mathbf{P}^T = \mathbf{P}^T\mathbf{P} = \mathbf{I}_n$. $\lambda_i$ denotes an eigen value of $\boldsymbol{\Sigma}$ and the $i^{th}$ column of $\mathbf{P}$ indicates the corresponding eigen vector.

#### 05). a).

Consider $\mathbf{y} = \mathbf{x} - \boldsymbol{\mu}$. Clearly, $\mathbf{y}$ would follow a Normal distribution.

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{y}} &= \mathbb{E}\{\mathbf{y}\} \\
&= \mathbb{E}\{\mathbf{x} - \boldsymbol{\mu}\} \\
&= \mathbb{E}\{\mathbf{x}\} - \boldsymbol{\mu} \\
&= \boldsymbol{\mu} - \boldsymbol{\mu} \\
\boldsymbol{\mu}_{\mathbf{y}} &= \mathbf{0} \qquad (5.1) \\
\boldsymbol{\Sigma}_{\mathbf{y}} &= \mathbb{E}\{(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})^T\} \\
&= \mathbb{E}\{\mathbf{y}\mathbf{y}^T\} \\
&= \mathbb{E}\{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T\} \\
\boldsymbol{\Sigma}_{\mathbf{y}} &= \boldsymbol{\Sigma} \qquad (5.2)
\end{aligned}$$

From Eq. 5.1 and Eq. 5.2, $\mathbf{y} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma})$.

#### 05). b).

Since the exponential function is a scalar function, the geometry of the level set of $f_{\mathbf{X}}$ will be reflected by the term $(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ inside the exponent. That is, $f_{\mathbf{X}} = k_1 \iff (\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = k_2$.

Let's consider $(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c^2$ where $c$ is a scalar. Using the Eigen-decomposition of $\boldsymbol{\Sigma}$, we can write, $(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}^T(\mathbf{x}-\boldsymbol{\mu}) = c^2$.

Take $\mathbf{w} = \mathbf{P}^T(\mathbf{x}-\boldsymbol{\mu})$. Note that the coordinate axes of $\mathbf{w}$ are obtained by first translating the coordinate axes of $\mathbf{x}$ to $\mathbf{x} = \boldsymbol{\mu}$ and then by applying a rotation to that translated $\mathbf{x}$ coordinate axes. Accordingly, the origin of $\mathbf{w}$ coordinate axes is at $\mathbf{x} = \boldsymbol{\mu}$.

$$\begin{aligned}
(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) &= c^2 \\
(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}^T(\mathbf{x}-\boldsymbol{\mu}) &= c^2 \\
\mathbf{w}^T\boldsymbol{\Lambda}^{-1}\mathbf{w} &= c^2 \\
\begin{pmatrix} w_1 & w_2 & \cdots & w_n \end{pmatrix}\begin{pmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_n} \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} &= c^2 \\
\frac{w_1^2}{\lambda_1} + \frac{w_2^2}{\lambda_2} + \cdots + \frac{w_n^2}{\lambda_n} &= c^2 \\
\frac{w_1^2}{\left(c\sqrt{\lambda_1}\right)^2} + \frac{w_2^2}{\left(c\sqrt{\lambda_2}\right)^2} + \cdots + \frac{w_n^2}{\left(c\sqrt{\lambda_n}\right)^2} &= 1 \qquad (5.3)
\end{aligned}$$

We can see that Eq. 5.3 represents an n-dimensional ellipsoid centred at $\mathbf{w} = \mathbf{0}$ with axes lengths $k_i = c\sqrt{\lambda_i}, \quad i = 1, 2, \ldots, n$.

In other words, the level set of $f_{\mathbf{X}}$ is an n-dimensional ellipsoid centred at $\mathbf{x} = \boldsymbol{\mu}$.

### 05). c).

Let $\boldsymbol{\Sigma}^{1/2}$ be the symmetric positive definite square root of $\boldsymbol{\Sigma}$ (i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}$).

Consider the transformation given by $\mathbf{y} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})$ for which we define the Jacobian of the transformation as $\mathbf{J} = \dfrac{\mathrm{d}\mathbf{y}^T}{\mathrm{d}\mathbf{x}}$.

$$\begin{aligned}
\mathbf{J} &= \frac{\mathrm{d}\,\mathbf{y}^T}{\mathrm{d}\mathbf{x}} \\
&= \frac{\mathrm{d}\,(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1/2}}{\mathrm{d}\mathbf{x}} \\
&= \begin{pmatrix} \frac{\partial(x_1-\mu_1)}{\partial x_1} & \frac{\partial(x_2-\mu_2)}{\partial x_1} & \cdots & \frac{\partial(x_n-\mu_n)}{\partial x_1} \\ \frac{\partial(x_1-\mu_1)}{\partial x_2} & \frac{\partial(x_2-\mu_2)}{\partial x_2} & \cdots & \frac{\partial(x_n-\mu_n)}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial(x_1-\mu_1)}{\partial x_n} & \frac{\partial(x_2-\mu_2)}{\partial x_n} & \cdots & \frac{\partial(x_n-\mu_n)}{\partial x_n} \end{pmatrix}\boldsymbol{\Sigma}^{-1/2} \\
&= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}\boldsymbol{\Sigma}^{-1/2} \\
&= \mathbf{I_n}\boldsymbol{\Sigma}^{-1/2} \\
\mathbf{J} &= \boldsymbol{\Sigma}^{-1/2}
\end{aligned}$$

$$
\begin{aligned}
\det(\mathbf{J}) &= \det\left(\mathbf{\Sigma}^{-1/2}\right) && (5.4)\\
\det\left(\mathbf{\Sigma}^{-1/2}\right) &= \det\left(\mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}^{T}\right)\\
&= \det\left(\mathbf{P}\right)\det\left(\Lambda^{-1/2}\right)\det\left(\mathbf{P}^{T}\right)\\
&= \det\left(\mathbf{P}\right)\det\left(\mathbf{P}^{T}\right)\det\left(\Lambda^{-1/2}\right)\\
&= \det\left(\mathbf{P}\mathbf{P}^{T}\right)\det\left(\Lambda^{-1/2}\right)\\
&= \det\left(\mathbf{I_n}\right)\det\left(\Lambda^{-1/2}\right)\\
&= \prod_{i=1}^{n}\frac{1}{\sqrt{\lambda_i}}\\
&= \sqrt{\prod_{i=1}^{n}\frac{1}{\lambda_i}}\\
\det\left(\mathbf{\Sigma}^{-1/2}\right) &= \sqrt{\det\left(\mathbf{\Sigma}^{-1}\right)} && (5.5)
\end{aligned}
$$

From Eq. 5.4 and Eq. 5.5,

$$
\det(\mathbf{J}) = \det\left(\mathbf{\Sigma}^{-1/2}\right) = \sqrt{\det\left(\mathbf{\Sigma}^{-1}\right)}
$$

### 05). d).

Note that $\mathbf{y} = \mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}) \implies \mathbf{x} = \mathbf{\Sigma}^{1/2}\mathbf{y}+\boldsymbol{\mu}$. We can write,

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{Y}}\left(\mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\right)\left|\det\left(\frac{\mathrm{d}\mathbf{y}^{T}}{\mathrm{d}\mathbf{x}}\right)\right|\\
&= f_{\mathbf{Y}}\left(\mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\right)|\det\left(\mathbf{J}\right)|\\
f_{\mathbf{Y}}(\mathbf{y}) &= f_{\mathbf{X}}\left(\mathbf{\Sigma}^{1/2}\mathbf{y}+\boldsymbol{\mu}\right)\frac{1}{|\det\left(\mathbf{J}\right)|}\\
&= f_{\mathbf{X}}\left(\mathbf{\Sigma}^{1/2}\mathbf{y}+\boldsymbol{\mu}\right)\frac{1}{\sqrt{\det\left(\mathbf{\Sigma}^{-1}\right)}} && \text{(From Eq. 5.4 and 5.5)}\\
&= \sqrt{\frac{\det\left(\mathbf{\Sigma}^{-1}\right)}{(2\pi)^{n}}}e^{-\frac{1}{2}\left(\mathbf{\Sigma}^{1/2}\mathbf{y}\right)^{T}\mathbf{\Sigma}^{-1}\left(\mathbf{\Sigma}^{1/2}\mathbf{y}\right)}\frac{1}{\sqrt{\det\left(\mathbf{\Sigma}^{-1}\right)}}\\
&= \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\mathbf{y}^{T}\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2}\mathbf{y}}\\
f_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\mathbf{y}^{T}\mathbf{y}} && (\because \mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{-1/2} = \mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2} = \mathbf{I_n})\\
f_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\|\mathbf{y}\|^{2}} && (5.6)
\end{aligned}
$$

## 05). e).

Let $\mathbf{z} = \mathbf{Q}\mathbf{y}$, where $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_n$. Then, $\mathbf{y} = \mathbf{Q}^T\mathbf{z}$. As $\mathbf{Q}$ is an orthogonal matrix, the absolute value of the determinant is equal to 1 (i.e., $|\det(\mathbf{Q})| = 1$).

$$\begin{aligned}
\mathbf{J} &= \frac{\mathrm{d}\,\mathbf{y}^T}{\mathrm{d}\mathbf{z}} \\
&= \frac{\mathrm{d}\,\mathbf{z}^T\mathbf{Q}}{\mathrm{d}\mathbf{z}} \\
&= \frac{\mathrm{d}\,\mathbf{z}^T}{\mathrm{d}\mathbf{z}}\mathbf{Q} \\
&= \mathbf{I}_n\mathbf{Q} \\
\mathbf{J} &= \mathbf{Q} \\
|\det(\mathbf{J})| &= |\det(\mathbf{Q})| \\
|\det(\mathbf{J})| &= 1
\end{aligned} \tag{5.7}$$

Now, let's determine the p.d.f. of $\mathbf{Z}$.

$$\begin{aligned}
f_{\mathbf{Z}}(\mathbf{z}) &= f_{\mathbf{Y}}\left(\mathbf{Q}^T\mathbf{z}\right)\left|\det\left(\frac{\mathrm{d}\,\mathbf{y}^T}{\mathrm{d}\mathbf{z}}\right)\right| \\
&= f_{\mathbf{Y}}\left(\mathbf{Q}^T\mathbf{z}\right)|\det(\mathbf{J})| \\
&= \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\left(\mathbf{Q}^T\mathbf{z}\right)^T\left(\mathbf{Q}^T\mathbf{z}\right)} && \text{(From Eq. 5.6 and Eq. 5.7)} \\
&= \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\mathbf{z}^T\mathbf{Q}\mathbf{Q}^T\mathbf{z}} \\
f_{\mathbf{Z}}(\mathbf{z}) &= \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\mathbf{z}^T\mathbf{z}} \\
f_{\mathbf{Z}}(\mathbf{z}) &= \frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\|\mathbf{z}\|^2}
\end{aligned} \tag{5.8}$$

## 05). f).

Consider the sample variance of $n$ i.i.d. standard Gaussian observations (i.e., $Z_1, Z_2, \ldots, Z_n$) given by,

$$s^2 = \frac{1}{n}\sum_{k=1}^{n}\left(Z_k - \bar{Z}\right)^2 = \frac{1}{n}\mathbf{Z}^T\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}$$

Let's determine the eigenvalues of $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)$.

$$\begin{aligned}
\det\left(\lambda\mathbf{I}_n - \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\right) &= 0 \\
\det\left((\lambda-1)\mathbf{I}_n + \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) &= 0 \\
\det\left((\lambda-1)\left(\mathbf{I}_n + \frac{1}{(\lambda-1)}\frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\right) &= 0 \\
(\lambda-1)^n\det\left(\mathbf{I}_n + \frac{1}{(\lambda-1)n}\mathbf{1}\mathbf{1}^T\right) &= 0
\end{aligned} \tag{5.9}$$

We can use the property; $\det(\mathbf{I}_n + \mathbf{AB}) = \det(\mathbf{I}_m + \mathbf{BA})$ for any two matrices $\mathbf{A} \in \mathbb{R}^{n\times m}, \mathbf{B} \in \mathbb{R}^{m\times n}$. Accordingly, $\det\left(\mathbf{I}_n + \frac{1}{(\lambda-1)n}\mathbf{1}\mathbf{1}^T\right) = \det\left(\mathbf{I}_1 + \frac{1}{(\lambda-1)n}\mathbf{1}^T\mathbf{1}\right) = 1 + \frac{1}{\lambda-1}$. Thus, Eq. 5.9 will be reduced to,

$$
\begin{aligned}
(\lambda-1)^n\left(1+\frac{1}{\lambda-1}\right) &= 0\\
\lambda(\lambda-1)^{n-1} &= 0
\end{aligned}
$$

Therefore, the eigen values of $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)$ are given by $\lambda = 1$ with multiplicity $(n-1)$ and $\lambda = 0$.

### 05). g).

In part 05). f)., the sample variance was defined for i.i.d. random variables $Z_i \sim \mathcal{N}(0,1)$ where $i = 1, 2, \ldots, n$,

$$
\begin{aligned}
s^2 &= \frac{1}{n}\mathbf{Z}^T\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}\\
ns^2 &= \mathbf{Z}^T\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}
\end{aligned}
$$

Further, we derived that $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)$ has the eigen values $\lambda = 1$ with multiplicity $(n-1)$ and $\lambda = 0$. Accordingly, we can write the Eigen-decomposition of $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)$ as $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) = \mathbf{Q\Lambda Q}^T$ where $\mathbf{\Lambda} = \mathrm{diag}\left(\underbrace{1, 1, \ldots, 1}_{(n-1)1's}, 0\right)$ and $\mathbf{QQ}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_n$.

$$
\begin{aligned}
ns^2 &= \mathbf{Z}^T\left(\mathbf{Q\Lambda Q}^T\right)\mathbf{Z}\\
ns^2 &= \left(\mathbf{Z}^T\mathbf{Q}\right)\mathbf{\Lambda}\left(\mathbf{Q}^T\mathbf{Z}\right)
\end{aligned}
$$

Take $\mathbf{Y} = \mathbf{Q}^T\mathbf{Z}$. Using Eq. 5.6 and 5.8, it is clear that $\mathbf{Y}$ also contains n i.i.d. standard Gaussian observations. This is a consequence of the rotational invariance of multivariate Gaussian vectors.

$$
\begin{aligned}
ns^2 &= \mathbf{Y}^T\mathbf{\Lambda}\mathbf{Y}\\
&= \begin{pmatrix} Y_1 & Y_2 & \cdots & Y_n \end{pmatrix}
\begin{pmatrix}
1 & 0 & \cdots & 0 & 0\\
0 & 1 & \cdots & 0 & 0\\
\vdots & \vdots & \vdots & \ddots & \vdots\\
0 & 0 & \cdots & 1 & 0\\
0 & 0 & \cdots & 0 & 0
\end{pmatrix}
\begin{pmatrix} Y_1\\ Y_2\\ \vdots\\ Y_n \end{pmatrix}\\
ns^2 &= Y_1^2 + Y_2^2 + \cdots + Y_{n-1}^2 \qquad (5.10)
\end{aligned}
$$

Eq. 5.10 represents $ns^2$ as a sum of squares of $(n-1)$ i.i.d. standard Gaussian observations. Thus, it is clear that $ns^2$ follows a Chi-square distribution with $(n-1)$ degrees of freedom.

$$
ns^2 \sim \chi^2_{n-1} \qquad (5.11)
$$

**05). h).**

Consider the following ratio between the two quadratic forms.

$$U = \frac{s^2}{\bar{Z}^2} = \frac{\mathcal{P}(\mathbf{Z})}{\mathcal{Q}(\mathbf{Z})} = \frac{\mathbf{Z}^T\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}}{\mathbf{Z}^T\frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{Z}}$$

Note that $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)$ and $\frac{\mathbf{1}\mathbf{1}^T}{n}$ are idempotent matrices. Therefore,

$$\begin{aligned}
\frac{\mathcal{P}(\mathbf{Z})}{\mathcal{Q}(\mathbf{Z})} &= \frac{\mathbf{Z}^T\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}}{\mathbf{Z}^T\frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{Z}} \\
\frac{\mathcal{P}(\mathbf{Z})}{\mathcal{Q}(\mathbf{Z})} &= \frac{\mathbf{Z}^T\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)^T\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}}{\mathbf{Z}^T\left(\frac{\mathbf{1}\mathbf{1}^T}{n}\right)^T\frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{Z}} \qquad (5.12)
\end{aligned}$$

Take $\mathbf{U} = \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}$ and $\mathbf{V} = \frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{Z}$.

$$\begin{aligned}
\mathbb{E}\{\mathbf{U}\} &= \mathbb{E}\left\{\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}\right\} \\
&= \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\underbrace{\mathbb{E}\{\mathbf{Z}\}}_{\mathbf{0}} \\
\mathbb{E}\{\mathbf{U}\} &= \mathbf{0} \\
\mathbb{E}\{\mathbf{V}\} &= \mathbb{E}\left\{\left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}\right\} \\
&= \left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\underbrace{\mathbb{E}\{\mathbf{Z}\}}_{\mathbf{0}} \\
\mathbb{E}\{\mathbf{V}\} &= \mathbf{0} \\
\mathbb{E}\left\{(\mathbf{U} - \mathbb{E}\{\mathbf{U}\})(\mathbf{V} - \mathbb{E}\{\mathbf{V}\})^T\right\} &= \mathbb{E}\left\{\mathbf{U}\mathbf{V}^T\right\} \\
&= \mathbb{E}\left\{\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}\mathbf{Z}^T\left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right)^T\right\} \\
&= \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\underbrace{\mathbb{E}\left\{\mathbf{Z}\mathbf{Z}^T\right\}}_{\mathbf{I}_n}\left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \\
&= \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \\
&= \left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right) - \left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \\
&= \left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right) - \left(\frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \qquad \left(\because \tfrac{1}{n}\mathbf{1}\mathbf{1}^T \text{ is idempotent}\right) \\
\mathbb{E}\left\{(\mathbf{U} - \mathbb{E}\{\mathbf{U}\})(\mathbf{V} - \mathbb{E}\{\mathbf{V}\})^T\right\} &= \mathbf{0}_{n\times n}
\end{aligned}$$

Thus, it is evident that $\mathbf{U}$ and $\mathbf{V}$ are uncorrelated Gaussian random variables. Consequently, $\mathbf{U} = \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z}$ and $\mathbf{V} = \frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{Z}$ are independent Gaussian random variables.

Using Eq. 5.12 and the above statement, we can say that $\mathcal{P}(\mathbf{Z})$ and $\mathcal{Q}(\mathbf{Z})$ are independent.

Let's determine the eigenvalues of $\frac{1}{n}\mathbf{1}\mathbf{1}^T$.

$$
\begin{aligned}
\det\left(\lambda \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) &= 0 \\
\det\left(\lambda\left(\mathbf{I}_n - \frac{1}{\lambda n}\mathbf{1}\mathbf{1}^T\right)\right) &= 0 \\
\lambda^n \det\left(\mathbf{I}_n - \frac{1}{\lambda n}\mathbf{1}\mathbf{1}^T\right) &= 0
\end{aligned} \tag{5.13}
$$

We can use the property; $\det\left(\mathbf{I}_n - \mathbf{AB}\right) = \det\left(\mathbf{I}_m - \mathbf{BA}\right)$ for any two matrices $\mathbf{A} \in \mathbb{R}^{n\times m}, \mathbf{B} \in \mathbb{R}^{m\times n}$. Accordingly, $\det\left(\mathbf{I}_n - \frac{1}{\lambda n}\mathbf{1}\mathbf{1}^T\right) = \det\left(\mathbf{I}_1 - \frac{1}{\lambda n}\mathbf{1}^T\mathbf{1}\right) = 1 - \frac{1}{\lambda}$. Thus, Eq. 5.13 will be reduced to,

$$
\begin{aligned}
\lambda^n\left(1 - \frac{1}{\lambda}\right) &= 0 \\
\lambda^{n-1}\left(\lambda - 1\right) &= 0
\end{aligned}
$$

Therefore, the eigen values of $\frac{1}{n}\mathbf{1}\mathbf{1}^T$ are given by $\lambda = 0$ with multiplicity $(n-1)$ and $\lambda = 1$.

Accordingly, we can write the Eigen-decomposition of $\frac{1}{n}\mathbf{1}\mathbf{1}^T$ as $\frac{1}{n}\mathbf{1}\mathbf{1}^T = \mathbf{P}\boldsymbol{\Omega}\mathbf{P}^T$ where $\boldsymbol{\Omega} = \mathrm{diag}\left(1, 0, \ldots, 0, 0\right)$ and $\mathbf{P}\mathbf{P}^T = \mathbf{P}^T\mathbf{P} = \mathbf{I}_n$.

We know that $\bar{Z}\mathbf{1} = \frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{Z}$. Then, $n\bar{Z}^2 = \mathbf{Z}^T\left(\frac{\mathbf{1}\mathbf{1}^T}{n}\right)^T\frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{Z}$ which could be further simplified to $n\bar{Z}^2 = \mathbf{Z}^T\frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{Z}$.

$$
\begin{aligned}
n\bar{Z}^2 &= \mathbf{Z}^T\left(\mathbf{P}\boldsymbol{\Omega}\mathbf{P}^T\right)\mathbf{Z} \\
n\bar{Z}^2 &= \left(\mathbf{Z}^T\mathbf{P}\right)\boldsymbol{\Omega}\left(\mathbf{P}^T\mathbf{Z}\right)
\end{aligned}
$$

Take $\mathbf{W} = \mathbf{P}^T\mathbf{Z}$. Using rotational invariance of multivariate Gaussian vectors, it is clear that $\mathbf{W}$ also contains n i.i.d. standard Gaussian random variables.

$$
\begin{aligned}
n\bar{Z}^2 &= \mathbf{W}^T\boldsymbol{\Omega}\mathbf{W} \\
&= \begin{pmatrix} W_1 & W_2 & \cdots & W_n \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{pmatrix} \\
n\bar{Z}^2 &= W_1^2
\end{aligned} \tag{5.14}
$$

Eq. 5.14 represents $n\bar{Z}^2$ as a square of a standard Gaussian random variable. Thus, it is clear that $n\bar{Z}^2$ follows a Chi-square distribution with 1 degree of freedom.

$$
n\bar{Z}^2 \sim \chi_1^2 \tag{5.15}
$$

Note that $\mathcal{P}(\mathbf{Z}) = \mathbf{Z}^T\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{Z} = ns^2$ and $\mathcal{Q}(\mathbf{Z}) = \mathbf{Z}^T\frac{\mathbf{1}\mathbf{1}^T}{n}\mathbf{Z} = n\bar{Z}^2$.

We showed that $\mathcal{P}(\mathbf{Z})$ and $\mathcal{Q}(\mathbf{Z})$ are independent which implies that $ns^2$ and $n\bar{Z}^2$ are independent.

From Eq. 5.11 and Eq. 5.15, we obtain $ns^2 \sim \chi^2_{n-1}$ and $n\bar{Z}^2 \sim \chi^2_1$.

It was defined that $U = \dfrac{s^2}{\bar{Z}^2}$ which could be written as $U = \dfrac{ns^2}{n\bar{Z}^2}$. Accordingly, $U$ is a ratio between two independent Chi-square random variables. Accordingly, we can write,

$$U \sim \frac{\chi^2_{n-1}}{\chi^2_1}$$

**05). i).**

Let $\mathbf{R} \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Suppose $\mathbf{R}$ assumes the Eigen-decomposition $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ where $\mathbf{\Lambda} = \text{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_n\right)$ with $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_n$.

$$\left(\mathbf{Z}^T\mathbf{R}\mathbf{Z}\right)^2 = \left(\mathbf{Z}^T\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\mathbf{Z}\right)^2$$

We will use the substitution $\mathbf{Y} = \mathbf{Q}^T\mathbf{Z}$.

$$\begin{aligned}
\left(\mathbf{Z}^T\mathbf{R}\mathbf{Z}\right)^2 &= \left(\mathbf{Y}^T\mathbf{\Lambda}\mathbf{Y}\right)^2 \\
&= \left(\sum_{i=1}^{n}\lambda_i Y_i^2\right)^2 \\
\left(\mathbf{Z}^T\mathbf{R}\mathbf{Z}\right)^2 &= \sum_{i=1}^{n}\lambda_i^2 Y_i^4 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_i\lambda_j Y_i^2 Y_j^2 \\
\mathbb{E}\left\{\left(\mathbf{Z}^T\mathbf{R}\mathbf{Z}\right)^2\right\} &= \mathbb{E}\left\{\sum_{i=1}^{n}\lambda_i^2 Y_i^4 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_i\lambda_j Y_i^2 Y_j^2\right\} \\
&= \mathbb{E}\left\{\sum_{i=1}^{n}\lambda_i^2 Y_i^4\right\} + \mathbb{E}\left\{\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_i\lambda_j Y_i^2 Y_j^2\right\} \\
\mathbb{E}\left\{\left(\mathbf{Z}^T\mathbf{R}\mathbf{Z}\right)^2\right\} &= \sum_{i=1}^{n}\lambda_i^2\mathbb{E}\left\{Y_i^4\right\} + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_i\lambda_j\mathbb{E}\left\{Y_i^2 Y_j^2\right\} \qquad (5.16)
\end{aligned}$$

For a given $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ with $\sigma \neq 0$, we proved that $\mathbb{E}\left\{(X-\mu)^{2k}\right\} = \sigma^{2k}(2k-1)(2k-3)\ldots 1$ where k is a positive integer (Given in Eq. 4.4).

Note that for $i = 1, 2, \ldots, n$, $Y_i \sim \mathcal{N}(0,1)$ because $\mathbf{Y}$ contains n i.i.d standard Gaussian random variables. Accordingly,

$$\mathbb{E}\left\{\left(Y_i - \underbrace{\mathbb{E}\{Y_i\}}_{0}\right)^{2k}\right\} = \underbrace{\sigma_{Y_i}^{2k}}_{1}(2k-1)(2k-3)\dots 1$$

$$\mathbb{E}\left\{Y_i^{2k}\right\} = (2k-1)(2k-3)\dots 1 \tag{5.17}$$

From Eq. 5.17, $\mathbb{E}\{Y_i^4\} = 3$ and $\mathbb{E}\{Y_i^2\} = 1$.

As, $Y_i$'s are independent, it is clear that $Y_i^2$s are independent for $i = 1, 2, \dots, n$. Then, $\mathbb{E}\left\{Y_i^2 Y_j^2\right\} = \mathbb{E}\{Y_i^2\}\,\mathbb{E}\left\{Y_j^2\right\} = 1$.

Using the above results in Eq. 5.16, we obtain,

$$\begin{aligned}
\mathbb{E}\left\{\left(\mathbf{Z}^T\mathbf{R}\mathbf{Z}\right)^2\right\} &= 3\sum_{i=1}^{n}\lambda_i^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_i\lambda_j \\
&= 2\sum_{i=1}^{n}\lambda_i^2 + \left(\sum_{i=1}^{n}\lambda_i^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_i\lambda_j\right) \\
\mathbb{E}\left\{\left(\mathbf{Z}^T\mathbf{R}\mathbf{Z}\right)^2\right\} &= 2\sum_{i=1}^{n}\lambda_i^2 + \left(\sum_{i=1}^{n}\lambda_i\right)^2
\end{aligned} \tag{5.18}$$

We know that $\mathrm{tr}(\mathbf{R}) = \sum_{i=1}^{n}\lambda_i$. Moreover, $\mathbf{R}^2 = \mathbf{Q}\mathbf{\Lambda}^2\mathbf{Q}^T$. Therefore, $\mathrm{tr}(\mathbf{R^2}) = \sum_{i=1}^{n}\lambda_i^2$. Thus, we can rewrite Eq. 5.18 as,

$$\mathbb{E}\left\{\left(\mathbf{Z}^T\mathbf{R}\mathbf{Z}\right)^2\right\} = 2\,\mathrm{tr}(\mathbf{R^2}) + \mathrm{tr}^2(\mathbf{R})$$

## Question 6 - Limit Laws, Concentration and Stochastic Convergence

Let $X_1, X_2, \cdots, X_n$, be i.i.d. uniformly distributed random variables each with zero mean and variance 1. Consider the following random variable.

$$Y = \frac{1}{n}\sum_{j=1}^{n}X_j$$

### 06). a).

$$\begin{aligned}
\mathbb{E}\{Y\} &= \mathbb{E}\left\{\frac{1}{n}\sum_{j=1}^{n}X_j\right\} \\
&= \frac{1}{n}\sum_{j=1}^{n}\underbrace{\mathbb{E}\{X_j\}}_{=0} \\
\mathbb{E}\{Y\} &= 0
\end{aligned}$$

$$
\begin{aligned}
\operatorname{Var}\{Y\} &= \operatorname{Var}\left\{\frac{1}{n}\sum_{j=1}^{n} X_j\right\} \\
&= \frac{1}{n^2}\operatorname{Var}\left\{\sum_{j=1}^{n} X_j\right\} \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\underbrace{\operatorname{Var}\{X_j\}}_{=1} && (\because X_js \text{ are i.i.d}) \\
\operatorname{Var}\{Y\} &= \frac{1}{n}
\end{aligned}
$$

**06). b).**

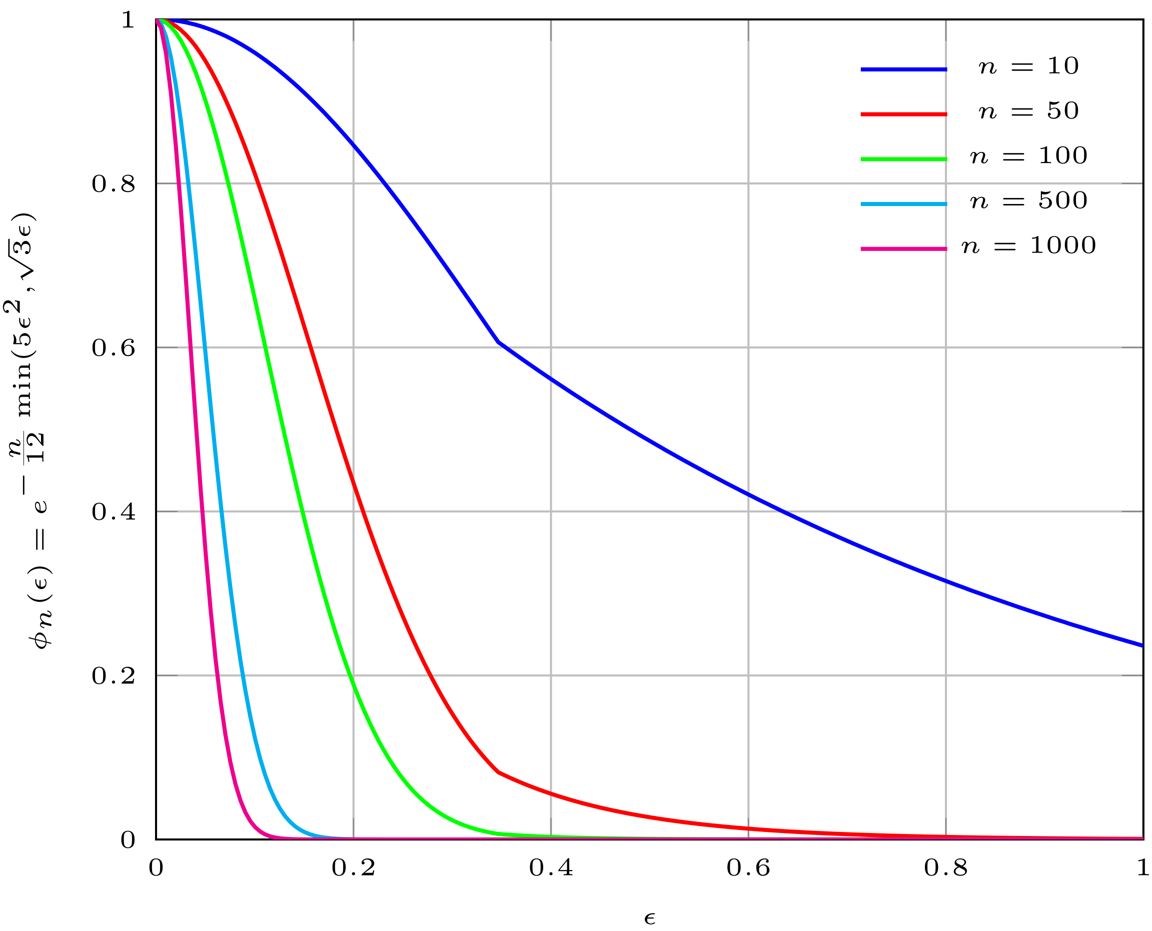


Figure 1: $\phi_n(\epsilon) = e^{-\frac{n}{12}\min(5\epsilon^2,\sqrt{3}\epsilon)}$ for the values of $n = 10, 50, 100, 500$ and $1000$

## 06). c). & d).

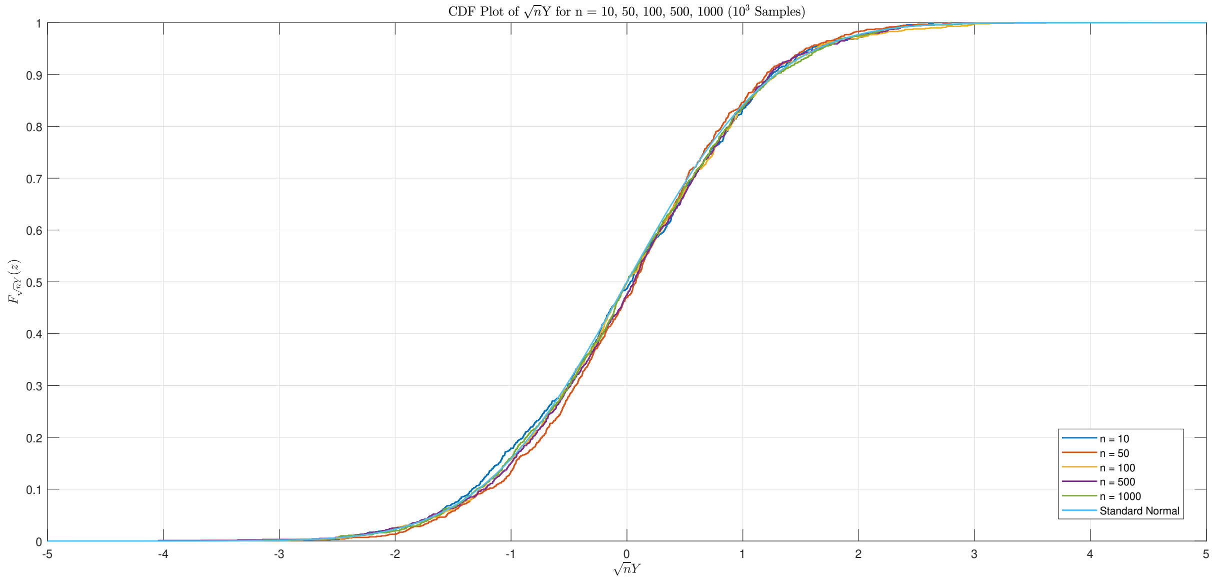


Figure 2: CDFs for $10^3$ Samples of $\sqrt{n}Y$ $(n = 10, 50, 100, 500, 1000)$ and the standard normal distribution

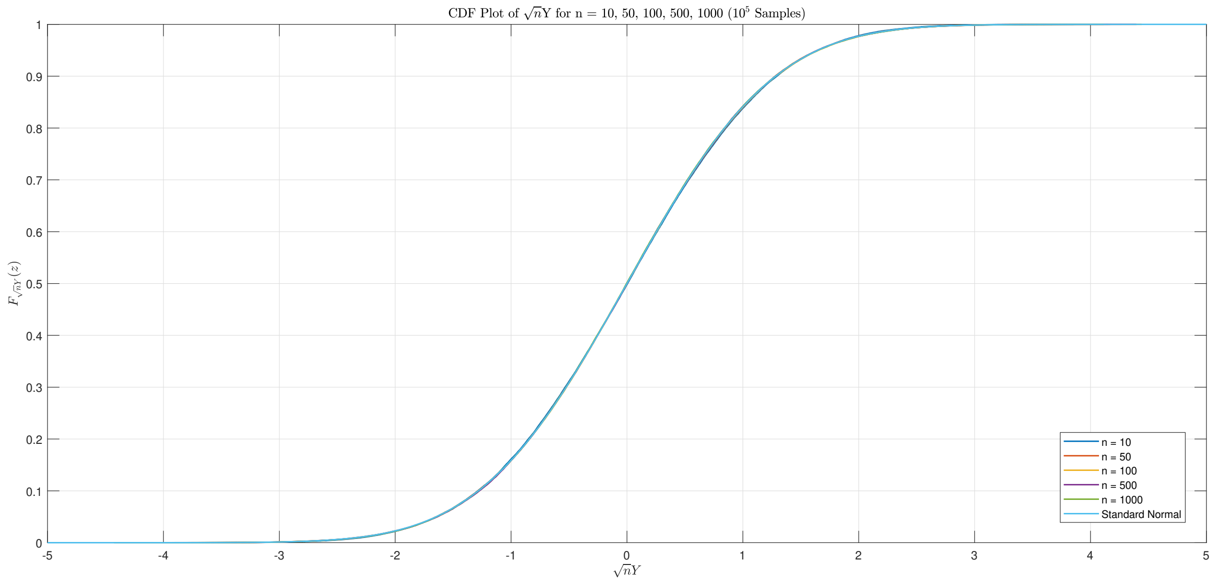


Figure 3: CDFs for $10^5$ Samples of $\sqrt{n}Y$ $(n = 10, 50, 100, 500, 1000)$ and the standard normal distribution

Suppose the CDF of the standard normal distribution is given by $\Phi(z)$. From Fig. 2 and Fig. 3, it is clear that $\lim_{n\to\infty} F_{\sqrt{n}Y}(z) \to \Phi(z)$ which indicates that $\sqrt{n}Y$ converges in distribution to a standard normal random variable as $n \to \infty$.

In addition, we can further state by comparing Fig. 2 and Fig. 3 that $\sqrt{n}Y$ converges more (in distribution) to a standard normal random variable when the number of samples is larger.

## Question 7 - Central Limit Theorem

Let $Z_1, Z_2, \cdots, Z_n$ be i.i.d. random variables each with mean $\mu$ and variance $\sigma^2$. Consider the following random sum,

$$S_n = \frac{\sum\limits_{j=1}^{n} Z_j - n\mu}{\sqrt{n}\sigma}$$

$$S_n = \frac{\sum\limits_{j=1}^{n} \left(\frac{Z_j-\mu}{\sigma}\right)}{\sqrt{n}}$$

Take $V_j = \dfrac{Z_j - \mu}{\sigma}$ for $j = 1, 2, \ldots, n$. Note that, $V_1, V_2, \ldots, V_n$ are i.i.d. random variables each with zero mean and variance 1.

$$S_n = \frac{\sum\limits_{j=1}^{n} V_j}{\sqrt{n}}$$

### 07). a).

Suppose the characteristic function of $S_n$ is given by $\phi_{S_n}(t)$. Moreover, for $k = 1, 2, \ldots, n$, $\phi_{V_k}(t) = \mathbb{E}\left\{e^{jtv_k}\right\}$

$$\begin{aligned}
\phi_{S_n}(t) &= \mathbb{E}\left\{e^{jts_n}\right\} \\
&= \mathbb{E}\left\{ e^{jt\left(\frac{\sum\limits_{k=1}^{n} v_k}{\sqrt{n}}\right)} \right\} \\
&= \mathbb{E}\left\{ e^{j\frac{t}{\sqrt{n}}\left(\sum\limits_{k=1}^{n} v_k\right)} \right\} \\
&= \mathbb{E}\left\{ \prod_{k=1}^{n} e^{j\frac{t}{\sqrt{n}}v_k} \right\} \\
&= \prod_{k=1}^{n} \mathbb{E}\left\{ e^{j\frac{t}{\sqrt{n}}v_k} \right\} && (\because V_1, V_2, \ldots, V_n \text{ are independent}) \\
\phi_{S_n}(t) &= \prod_{k=1}^{n} \phi_{V_k}\left(\frac{t}{\sqrt{n}}\right)
\end{aligned}$$

Since $V_1, V_2, \ldots, V_n$ are i.i.d., $\phi_{V_1}(t) = \phi_{V_2}(t) = \cdots = \phi_{V_n}(t)$. Accordingly, we will use $\phi(t) = \mathbb{E}\left\{e^{jtv}\right\}$ to denote the characteristic funcions of $V_1, V_2, \ldots, V_n$.

$$\therefore \phi_{S_n}(t) = \left\{\phi\left(\frac{t}{\sqrt{n}}\right)\right\}^n \tag{7.1}$$

**07). b).**

$$\phi\left(\frac{t}{\sqrt{n}}\right) = \mathbb{E}\left\{e^{j\frac{t}{\sqrt{n}}v}\right\}$$

Using the series expansion of $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots,$

$$\begin{aligned}
\phi\left(\frac{t}{\sqrt{n}}\right) &= \mathbb{E}\left\{1 + j\frac{t}{\sqrt{n}}v + \frac{\left(j\frac{t}{\sqrt{n}}v\right)^2}{2!} + \frac{\left(j\frac{t}{\sqrt{n}}v\right)^3}{3!} + \frac{\left(j\frac{t}{\sqrt{n}}v\right)^4}{4!} + \ldots\right\} \\
&= \mathbb{E}\left\{1 + j\frac{t}{\sqrt{n}}v - \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 v^2 - j\frac{1}{6}\left(\frac{t}{\sqrt{n}}\right)^3 v^3 + \frac{1}{24}\left(\frac{t}{\sqrt{n}}\right)^4 v^4 + \ldots\right\} \\
&= 1 + j\frac{t}{\sqrt{n}}\underbrace{\mathbb{E}\{v\}}_{=0} - \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 \underbrace{\mathbb{E}\{v^2\}}_{=1} - j\frac{1}{6}\left(\frac{t}{\sqrt{n}}\right)^3 \underbrace{\mathbb{E}\{v^3\}}_{=0} + \frac{1}{24}\left(\frac{t}{\sqrt{n}}\right)^4 \underbrace{\mathbb{E}\{v^4\}}_{=3} + \ldots \\
&= 1 - \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + \frac{1}{8}\left(\frac{t}{\sqrt{n}}\right)^4 + \ldots \\
&= 1 - \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 + o\left(\left(\frac{t}{\sqrt{n}}\right)^2\right) \\
\phi\left(\frac{t}{\sqrt{n}}\right) &= 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right) \qquad (7.2)
\end{aligned}$$

**07). c).**

From Eq. 7.1 and Eq. 7.2, we have $\phi_{S_n}(t) = \left\{\phi\left(\frac{t}{\sqrt{n}}\right)\right\}^n$ and $\phi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)$. Then,

$$\begin{aligned}
\phi_{S_n}(t) &= \left\{1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right\}^n \\
\lim_{n\to\infty} \phi_{S_n}(t) &= \lim_{n\to\infty}\left\{1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right\}^n \\
&= \lim_{n\to\infty}\left\{1 - \frac{t^2}{2n}\right\}^n
\end{aligned}$$

The limit definition of $e^x$ is $e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n$. Accordingly,

$$\begin{aligned}
\lim_{n\to\infty} \phi_{S_n}(t) &= \lim_{n\to\infty}\left\{1 + \frac{\left(-\frac{t^2}{2}\right)}{n}\right\}^n \\
\therefore \lim_{n\to\infty} \phi_{S_n}(t) &= e^{-\frac{t^2}{2}} \qquad (7.3)
\end{aligned}$$

Note that $e^{-\frac{t^2}{2}}$ is the characteristic function ($\phi(t)$) of a standard normal random variable. Thus, using Eq. 7.3, we can say that $\lim_{n\to\infty} S_n \longrightarrow Z$ in distribution, where $Z \sim \mathcal{N}(0,1)$.

## Question 8 - Least Squares

Consider the overdetermined system given by,

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{x} \in \mathbb{R}^n$, and $\mathbf{b} \in \mathbb{R}^m$ with $m \geq n$. Since usually $\mathbf{b}$ is not in the column space of $\mathbf{A}$, an overdetermined system has no exact solution. To circumvent this problem, we strive to minimize $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$. It turns out that the solution to the latter minimization problem depends on the rank of $\mathbf{A}$.

### 08). a).

Consider the equality, for $\mathbf{z} \in \mathbb{R}^n$,

$$\|\mathbf{A}(\mathbf{x} + \mathbf{z}) - \mathbf{b}\|_2^2 = \|\mathbf{A}\mathbf{x} - \mathbf{b} + \mathbf{A}\mathbf{z}\|_2^2 = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \|\mathbf{A}\mathbf{z}\|_2^2 + 2(\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{z})$$

Suppose both $\mathbf{x}$ and $\mathbf{x} + \mathbf{z}$ are solutions to the above minimization problem. Then,

$$\mathbf{A}(\mathbf{x} + \mathbf{z}) = \mathbf{A}\mathbf{x}$$
$$\therefore \mathbf{A}\mathbf{z} = \mathbf{0}$$

In other words, $\mathbf{z}$ is in the null space of $\mathbf{A}$. (i.e., null space of $\mathbf{A}$ consists of the set of vectors $\mathbf{z}$ for which $\mathbf{A}\mathbf{z} = \mathbf{0}$).

### 08). b).

Take $\mathbf{A} = \begin{pmatrix} \mathbf{a_1} & \mathbf{a_2} & \cdots & \mathbf{a_n} \end{pmatrix}$ and $\mathbf{z} = \begin{pmatrix} z_1 & z_2 & \cdots & z_n \end{pmatrix}^T$. Assume that $\mathbf{A}$ has full column rank which implies that its columns are linearly independent. In such cases,

$$\mathbf{A}\mathbf{z} = \mathbf{0} \implies z_1\mathbf{a_1} + z_2\mathbf{a_2} + \cdots + z_n\mathbf{a_n} = \mathbf{0} \implies \mathbf{z} = \mathbf{0}$$

Accordingly, when $\mathbf{A}$ has full column rank, the only vector $\mathbf{z}$ that is in the null space of $\mathbf{A}$ is $\mathbf{z} = \mathbf{0}$.

Thus, we can say that the dimensionality of $\mathbf{z}$ is 0.

### 08). c).

Under the full column rank assumption on $\mathbf{A}$, let us consider the function,

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \\ &= \frac{1}{2}(\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ &= \frac{1}{2}\left(\mathbf{x}^T\mathbf{A}^T - \mathbf{b}^T\right)(\mathbf{A}\mathbf{x} - \mathbf{b}) \end{aligned}$$

$$
\begin{aligned}
f(\mathbf{x}) &= \frac{1}{2}\left\{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{x}^T\mathbf{A}^T\mathbf{b} + \mathbf{b}^T\mathbf{b}\right\} \\
\frac{\mathrm{d}\, f(\mathbf{x})}{\mathrm{d}\mathbf{x}} &= \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}\left(\frac{1}{2}\left\{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{x}^T\mathbf{A}^T\mathbf{b} + \mathbf{b}^T\mathbf{b}\right\}\right) \\
&= \frac{1}{2}\left\{\frac{\mathrm{d}\,\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}}{\mathrm{d}\mathbf{x}} - 2\frac{\mathrm{d}\,\mathbf{x}^T\mathbf{A}^T\mathbf{b}}{\mathrm{d}\mathbf{x}} + \underbrace{\frac{\mathrm{d}\,\mathbf{b}^T\mathbf{b}}{\mathrm{d}\mathbf{x}}}_{\mathbf{0}}\right\}
\end{aligned}
$$

From Eq. 1.17, $\frac{\mathrm{d}\,\mathbf{x}^T\mathbf{A}^T\mathbf{b}}{\mathrm{d}\mathbf{x}} = \mathbf{A}^T\mathbf{b}$ and from Eq. 1.21, $\frac{\mathrm{d}\,\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}}{\mathrm{d}\mathbf{x}} = 2\mathbf{A}^T\mathbf{A}\mathbf{x}$.

$$
\begin{aligned}
\frac{\mathrm{d}\, f(\mathbf{x})}{\mathrm{d}\mathbf{x}} &= \frac{1}{2}\left\{2\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{A}^T\mathbf{b}\right\} \\
\frac{\mathrm{d}\, f(\mathbf{x})}{\mathrm{d}\mathbf{x}} &= \mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{A}^T\mathbf{b}
\end{aligned}
$$

Suppose $\frac{\mathrm{d}\, f(\mathbf{x})}{\mathrm{d}\mathbf{x}} = \mathbf{0}$ when $\mathbf{x} = \mathbf{x}_{LS}$. Then,

$$
\begin{aligned}
\mathbf{A}^T\mathbf{A}\mathbf{x}_{LS} - \mathbf{A}^T\mathbf{b} &= \mathbf{0} \\
\mathbf{A}^T\mathbf{A}\mathbf{x}_{LS} &= \mathbf{A}^T\mathbf{b}
\end{aligned}
$$

Since $\mathbf{A}$ is a full rank matrix, $\mathbf{A}^T\mathbf{A}$ would also be full rank and it would be invertible. Therefore,

$$
\mathbf{x}_{LS} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}
$$

As $f(\mathbf{x})$ is strictly convex, it would have a global minimum. Accordingly, we can say that $f(\mathbf{x})$ achieves the global minimum at $\mathbf{x}_{\mathrm{LS}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$.

### 08). d).

Consider the singular value decomposition of the full rank matrix $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Suppose $\mathbf{U} = \begin{pmatrix}\mathbf{u_1} & \mathbf{u_2} & \cdots & \mathbf{u_m}\end{pmatrix}, \mathbf{V} = \begin{pmatrix}\mathbf{v_1} & \mathbf{v_2} & \cdots & \mathbf{v_n}\end{pmatrix}$ and $\mathbf{\Sigma}$ contains the singular values of $\mathbf{A}$; with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$.

Note that $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_m$ and $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.

$$
\begin{aligned}
\left\|\mathbf{U}^T\mathbf{A}\mathbf{x} - \mathbf{U}^T\mathbf{b}\right\|_2^2 &= \left(\mathbf{U}^T\mathbf{A}\mathbf{x} - \mathbf{U}^T\mathbf{b}\right)^T\left(\mathbf{U}^T\mathbf{A}\mathbf{x} - \mathbf{U}^T\mathbf{b}\right) \\
&= \left(\mathbf{U}^T\left(\mathbf{A}\mathbf{x} - \mathbf{b}\right)\right)^T\left(\mathbf{U}^T\left(\mathbf{A}\mathbf{x} - \mathbf{b}\right)\right) \\
&= \left(\mathbf{A}\mathbf{x} - \mathbf{b}\right)^T\underbrace{\mathbf{U}\mathbf{U}^T}_{\mathbf{I}_m}\left(\mathbf{A}\mathbf{x} - \mathbf{b}\right) \\
&= \left(\mathbf{A}\mathbf{x} - \mathbf{b}\right)^T\left(\mathbf{A}\mathbf{x} - \mathbf{b}\right) \\
\left\|\mathbf{U}^T\mathbf{A}\mathbf{x} - \mathbf{U}^T\mathbf{b}\right\|_2^2 &= \left\|\mathbf{A}\mathbf{x} - \mathbf{b}\right\|_2^2 \\
\left\|\mathbf{A}\mathbf{x} - \mathbf{b}\right\|_2^2 &= \left\|\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{x} - \mathbf{U}^T\mathbf{b}\right\|_2^2 \\
\left\|\mathbf{A}\mathbf{x} - \mathbf{b}\right\|_2^2 &= \left\|\mathbf{\Sigma}\mathbf{V}^T\mathbf{x} - \mathbf{U}^T\mathbf{b}\right\|_2^2 \\
\therefore \quad \min_{\mathbf{x}\in\mathbb{R}^n}\left\|\mathbf{A}\mathbf{x} - \mathbf{b}\right\|_2^2 &= \min_{\mathbf{x}\in\mathbb{R}^n}\left\|\mathbf{\Sigma}\mathbf{V}^T\mathbf{x} - \mathbf{U}^T\mathbf{b}\right\|_2^2 \qquad (8.1)
\end{aligned}
$$

Let's apply the substitutions, $\mathbf{y} = \mathbf{V}^T\mathbf{x}$ and $\mathbf{a} = \mathbf{U}^T\mathbf{b}$. Then, $\left\|\mathbf{\Sigma}\mathbf{V}^T\mathbf{x} - \mathbf{U}^T\mathbf{b}\right\|_2^2 = \|\mathbf{\Sigma}\mathbf{y} - \mathbf{a}\|_2^2$.

$$\|\mathbf{\Sigma}\mathbf{y} - \mathbf{a}\|_2^2 = (\sigma_1 y_1 - a_1)^2 + (\sigma_2 y_2 - a_2)^2 + \cdots + (\sigma_n y_n - a_n)^2 + a_{n+1}^2 + \cdots + a_m^2$$

In order to minimize $\|\mathbf{\Sigma}\mathbf{y} - \mathbf{a}\|_2^2$, we need to have $\sigma_i y_i = a_i \implies y_i = \dfrac{a_i}{\sigma_i}$ for $i = 1, 2, \ldots, n$.

$$\mathbf{y}_{LS} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \tag{8.2}$$

We used the substitution, $\mathbf{a} = \mathbf{U}^T\mathbf{b}$. Then,

$$\mathbf{a} = \begin{pmatrix} \mathbf{u_1}^T \\ \mathbf{u_2}^T \\ \vdots \\ \mathbf{u_m}^T \end{pmatrix} \mathbf{b} \implies \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \mathbf{u_1}^T\mathbf{b} \\ \mathbf{u_2}^T\mathbf{b} \\ \vdots \\ \mathbf{u_m}^T\mathbf{b} \end{pmatrix} \tag{8.3}$$

From Eq. 8.3, we get $a_i = \mathbf{u_i}^T\mathbf{b}$ for $i = 1, 2, \ldots, m$. Using this result in Eq. 8.2,

$$\mathbf{y}_{LS} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{pmatrix} \begin{pmatrix} \mathbf{u_1}^T\mathbf{b} \\ \mathbf{u_2}^T\mathbf{b} \\ \vdots \\ \mathbf{u_n}^T\mathbf{b} \end{pmatrix} \implies \mathbf{y}_{LS} = \begin{pmatrix} \dfrac{\mathbf{u_1}^T\mathbf{b}}{\sigma_1} \\ \dfrac{\mathbf{u_2}^T\mathbf{b}}{\sigma_2} \\ \vdots \\ \dfrac{\mathbf{u_n}^T\mathbf{b}}{\sigma_n} \end{pmatrix} \implies \mathbf{V}^T\mathbf{x}_{LS} = \begin{pmatrix} \dfrac{\mathbf{u_1}^T\mathbf{b}}{\sigma_1} \\ \dfrac{\mathbf{u_2}^T\mathbf{b}}{\sigma_2} \\ \vdots \\ \dfrac{\mathbf{u_n}^T\mathbf{b}}{\sigma_n} \end{pmatrix}$$

$$\therefore \quad \mathbf{x}_{LS} = \mathbf{V} \begin{pmatrix} \dfrac{\mathbf{u_1}^T\mathbf{b}}{\sigma_1} \\ \dfrac{\mathbf{u_2}^T\mathbf{b}}{\sigma_2} \\ \vdots \\ \dfrac{\mathbf{u_n}^T\mathbf{b}}{\sigma_n} \end{pmatrix} \implies \mathbf{x}_{LS} = \begin{pmatrix} \mathbf{v_1} & \mathbf{v_2} & \cdots & \mathbf{v_n} \end{pmatrix} \begin{pmatrix} \dfrac{\mathbf{u_1}^T\mathbf{b}}{\sigma_1} \\ \dfrac{\mathbf{u_2}^T\mathbf{b}}{\sigma_2} \\ \vdots \\ \dfrac{\mathbf{u_n}^T\mathbf{b}}{\sigma_n} \end{pmatrix}$$

Using the outer product,

$$\mathbf{x}_{LS} = \frac{\mathbf{u_1}^T\mathbf{b}}{\sigma_1}\mathbf{v_1} + \frac{\mathbf{u_2}^T\mathbf{b}}{\sigma_2}\mathbf{v_2} + \cdots + \frac{\mathbf{u_n}^T\mathbf{b}}{\sigma_n}\mathbf{v_n}$$

$$\mathbf{x}_{LS} = \sum_{j=1}^{n} \frac{\mathbf{u_j}^T\mathbf{b}}{\sigma_j}\mathbf{v_j} \tag{8.4}$$

### 08). e).

Suppose rank($\mathbf{A}$) $= r \leq n$. Then the columns of $\mathbf{A}$ are linearly dependent. Therefore, $\exists\, \mathbf{z} \in \mathbb{R}^n, \mathbf{z} \neq \mathbf{0}$ such that $\mathbf{Az} = \mathbf{0}$. Accordingly, the null space of $\mathbf{A}$ contains non-zero vectors other than $\mathbf{0}$.

Consider any $\mathbf{z} \in \mathbb{R}^n, \mathbf{z} \neq \mathbf{0}$ such that $\mathbf{Az} = \mathbf{0}$. Moreover, if $\mathbf{x}^*$ is a solution to the minimization problem $\min\limits_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|_2^2$, then note that for any $k \in \mathbb{R}$, $\mathbf{x}^* + k\mathbf{z}$ would also be a solution to the same minimization problem as, $\mathbf{A}\left(\mathbf{x}^* + k\mathbf{z}\right) = \mathbf{Ax}^* + \mathbf{A}\left(k\mathbf{z}\right) = \mathbf{Ax}^* + k\mathbf{Az} = \mathbf{Ax}^*$.

Accordingly, when rank($\mathbf{A}$) $= r \leq n$, the minimization problem $\min\limits_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|_2^2$ would have infinitely many solutions.

### 08). f).

It is given that rank($\mathbf{A}$) $= r \leq n$. Consider the singular value decomposition of the matrix $\mathbf{A} = \mathbf{U\Sigma V}^T$. Suppose $\mathbf{U} = \begin{pmatrix} \mathbf{u_1} & \mathbf{u_2} & \cdots & \mathbf{u_m} \end{pmatrix}, \mathbf{V} = \begin{pmatrix} \mathbf{v_1} & \mathbf{v_2} & \cdots & \mathbf{v_n} \end{pmatrix}$ and $\mathbf{\Sigma}$ contains the singular values of $\mathbf{A}$; with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Note that $\mathbf{UU}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_m$ and $\mathbf{VV}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.

From Eq. 8.1, $\min\limits_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|_2^2 = \min\limits_{\mathbf{x}\in\mathbb{R}^n} \left\|\mathbf{\Sigma V}^T\mathbf{x} - \mathbf{U}^T\mathbf{b}\right\|_2^2$.

Let's apply the substitutions, $\mathbf{y} = \mathbf{V}^T\mathbf{x}$ and $\mathbf{a} = \mathbf{U}^T\mathbf{b}$. Then, $\left\|\mathbf{\Sigma V}^T\mathbf{x} - \mathbf{U}^T\mathbf{b}\right\|_2^2 = \|\mathbf{\Sigma y} - \mathbf{a}\|_2^2$.

$$\|\mathbf{\Sigma y} - \mathbf{a}\|_2^2 = \left(\sigma_1 y_1 - a_1\right)^2 + \left(\sigma_2 y_2 - a_2\right)^2 + \cdots + \left(\sigma_r y_r - a_r\right)^2 + a_{r+1}^2 + \cdots + a_m^2$$

In order to minimize $\|\mathbf{\Sigma y} - \mathbf{a}\|_2^2$, we need to have $\sigma_i y_i = a_i \implies y_i = \dfrac{a_i}{\sigma_i}$ for $i = 1, 2, \ldots, r$.

$$\mathbf{y}_{LS}^* = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_r} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \tag{8.5}$$

From Eq. 8.3, we get $a_i = \mathbf{u_i}^T\mathbf{b}$ for $i = 1, 2, \ldots, m$. Using this result in Eq. 8.5,

$$\mathbf{y}^*_{LS} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_r} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u_1}^T\mathbf{b} \\ \mathbf{u_2}^T\mathbf{b} \\ \vdots \\ \mathbf{u_n}^T\mathbf{b} \end{pmatrix} \implies \mathbf{y}^*_{LS} = \begin{pmatrix} \dfrac{\mathbf{u_1}^T\mathbf{b}}{\sigma_1} \\ \dfrac{\mathbf{u_2}^T\mathbf{b}}{\sigma_2} \\ \vdots \\ \dfrac{\mathbf{u_r}^T\mathbf{b}}{\sigma_r} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\therefore \quad \mathbf{V}^T\mathbf{x}^*_{LS} = \begin{pmatrix} \dfrac{\mathbf{u_1}^T\mathbf{b}}{\sigma_1} \\ \dfrac{\mathbf{u_2}^T\mathbf{b}}{\sigma_2} \\ \vdots \\ \dfrac{\mathbf{u_r}^T\mathbf{b}}{\sigma_r} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \implies \mathbf{x}^*_{LS} = \mathbf{V} \begin{pmatrix} \dfrac{\mathbf{u_1}^T\mathbf{b}}{\sigma_1} \\ \dfrac{\mathbf{u_2}^T\mathbf{b}}{\sigma_2} \\ \vdots \\ \dfrac{\mathbf{u_r}^T\mathbf{b}}{\sigma_r} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \implies \mathbf{x}^*_{LS} = \begin{pmatrix} \mathbf{v_1} & \mathbf{v_2} & \cdots & \mathbf{v_n} \end{pmatrix} \begin{pmatrix} \dfrac{\mathbf{u_1}^T\mathbf{b}}{\sigma_1} \\ \dfrac{\mathbf{u_2}^T\mathbf{b}}{\sigma_2} \\ \vdots \\ \dfrac{\mathbf{u_r}^T\mathbf{b}}{\sigma_r} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Using the outer product,

$$\begin{aligned} \mathbf{x}^*_{LS} &= \frac{\mathbf{u_1}^T\mathbf{b}}{\sigma_1}\mathbf{v_1} + \frac{\mathbf{u_2}^T\mathbf{b}}{\sigma_2}\mathbf{v_2} + \cdots + \frac{\mathbf{u_r}^T\mathbf{b}}{\sigma_r}\mathbf{v_r} \\ \mathbf{x}^*_{LS} &= \sum_{j=1}^{r} \frac{\mathbf{u_j}^T\mathbf{b}}{\sigma_j}\mathbf{v_j} \end{aligned} \tag{8.6}$$

## Question 9 - MM Algorithm

A random variable $X$ is known to have the following truncated Poisson distribution:

$$\Pr(X = k) = \frac{1}{(e^{\theta} - 1)\, k!}\theta^{k}, \quad k = 1, 2, 3, \dots$$

where $\theta$ is a positive deterministic parameter.

### 09). a).

$$\begin{aligned}
f\left(\theta\right) &= \ln\left(e^{\theta} - 1\right) \\
f'\left(\theta\right) &= \frac{e^{\theta}}{e^{\theta} - 1} \\
f''\left(\theta\right) &= \frac{e^{\theta}(e^{\theta} - 1) - e^{\theta}e^{\theta}}{\left(e^{\theta} - 1\right)^2} = -\frac{e^{\theta}}{\left(e^{\theta} - 1\right)^2}
\end{aligned} \tag{9.1}$$

From Eq. 9.1, we can see that $\forall \theta > 0, f''(\theta)$ exists and $f''(\theta) < 0$.

$\therefore \ln\left(e^{\theta} - 1\right)$ is concave in $\theta \implies (e^{\theta} - 1)$ is log-concave in $\theta$.

### 09). b).

Consider $m$ independent and identically distributed sample values $x_1, x_2, \dots, x_m$. Let's use the maximum likelihood (ML) estimation method to estimate $\theta$ based on the sample values. Then, the likelihood function is given by,

$$\begin{aligned}
\Pr\left[X_1 = x_1, X_2 = x_2, \dots, X_m = x_m\right] &= \prod_{k=1}^{m} \frac{\theta^{x_k}}{(e^{\theta} - 1)\, x_k!} \\
&= \frac{1}{\prod\limits_{k=1}^{m} x_k!} \frac{\theta^{m\bar{x}}}{\left(e^{\theta} - 1\right)^m}
\end{aligned}$$

where $\bar{x} = \frac{1}{m}\left(x_1 + x_2 + \dots + x_m\right)$.

Since $\ln(.)$ is a monotonically increasing function and as the above likelihood function is positive, we can say that,

$$\begin{aligned}
\underset{\theta>0}{\text{maximize}} \quad \frac{1}{\prod\limits_{k=1}^{m} x_k!} \frac{\theta^{m\bar{x}}}{\left(e^{\theta} - 1\right)^m} &\equiv \underset{\theta>0}{\text{maximize}} \quad \ln\left[\frac{1}{\prod\limits_{k=1}^{m} x_k!} \frac{\theta^{m\bar{x}}}{\left(e^{\theta} - 1\right)^m}\right] \\
&\equiv \underset{\theta>0}{\text{maximize}} \quad \left[-\ln\left(\prod_{k=1}^{m} x_k!\right) + \ln\left(\theta^{m\bar{x}}\right) - \ln\left(e^{\theta} - 1\right)^m\right] \\
&\equiv \underset{\theta>0}{\text{maximize}} \quad \left[-\sum_{k=1}^{m} \ln\left(x_k!\right) + m\bar{x}\ln\left(\theta\right) - m\ln\left(e^{\theta} - 1\right)\right]
\end{aligned}$$

Note that $-\sum_{k=1}^{m} \ln(x_k!)$ does not include any $\theta$ term, and hence we can further simplify the above expression,

$$\begin{aligned}
\underset{\theta>0}{\text{maximize}} \quad \left[m\bar{x}\ln(\theta) - m\ln\left(e^{\theta}-1\right)\right] &\equiv \underset{\theta>0}{\text{maximize}} \quad \left[\bar{x}\ln(\theta) - \ln\left(e^{\theta}-1\right)\right] \qquad (\because m > 0) \\
&\equiv \underset{\theta>0}{\text{maximize}} \quad \left\{-\left[\ln\left(e^{\theta}-1\right) - \bar{x}\ln(\theta)\right]\right\} \\
&\equiv \underset{\theta>0}{\text{minimize}} \quad \left[\ln\left(e^{\theta}-1\right) - \bar{x}\ln(\theta)\right]
\end{aligned}$$

Accordingly, the ML estimation problem is equivalent to,

$$\underset{\theta>0}{\text{maximize}} \quad \frac{1}{\prod_{k=1}^{m} x_k!} \frac{\theta^{m\bar{x}}}{\left(e^{\theta}-1\right)^{m}} \equiv \underset{\theta>0}{\text{minimize}} \quad \left[\ln\left(e^{\theta}-1\right) - \bar{x}\ln(\theta)\right]$$

**09). c).**

Take $\mathcal{L}(\theta) = \ln\left(e^{\theta}-1\right) - \bar{x}\ln(\theta)$. Let's assign $f(\theta) = \ln\left(e^{\theta}-1\right)$ and $h(\theta) = -\bar{x}\ln(\theta)$.

Using Eq. 9.1, we deduced in 09). a). that $f(\theta) = \ln\left(e^{\theta}-1\right)$ is concave in $\theta$ where $\theta > 0$. We will determine the nature of $h(\theta) = -\bar{x}\ln(\theta)$.

$$h(\theta) = -\bar{x}\ln(\theta) \implies h'(\theta) = -\frac{\bar{x}}{\theta} \implies h''(\theta) = \frac{\bar{x}}{\theta^2} > 0 \quad \forall\, \theta > 0$$

Therefore $h(\theta) = -\bar{x}\ln(\theta)$ is convex in $\theta$ where $\theta > 0$. Accordingly, we can't determine the nature of $\mathcal{L}(\theta)$ by only considering the nature of $f(\theta)$ and $h(\theta)$ separately.

$$\mathcal{L}(\theta) = \ln\left(e^{\theta}-1\right) - \bar{x}\ln(\theta) \implies \mathcal{L}'(\theta) = 1 + \frac{1}{e^{\theta}-1} - \frac{\bar{x}}{\theta} \implies \mathcal{L}''(\theta) = -\frac{e^{\theta}}{\left(e^{\theta}-1\right)^2} + \frac{\bar{x}}{\theta^2}$$

Then, $\mathcal{L}''(\theta) = -\frac{e^{\theta}}{\left(e^{\theta}-1\right)^2} + \frac{\bar{x}}{\theta^2} = \frac{\bar{x}\left(e^{\theta}-1\right)^2 - \theta^2 e^{\theta}}{\theta^2\left(e^{\theta}-1\right)^2}$. Since, $\theta^2\left(e^{\theta}-1\right)^2 > 0 \quad \forall\, \theta > 0$, the sign of $\mathcal{L}''(\theta)$ depends on $\bar{x}\left(e^{\theta}-1\right)^2 - \theta^2 e^{\theta}$.

It is clear that when $\bar{x} < \frac{\theta^2 e^{\theta}}{\left(e^{\theta}-1\right)^2}$, $\quad \mathcal{L}''(\theta) < 0$ and when $\bar{x} > \frac{\theta^2 e^{\theta}}{\left(e^{\theta}-1\right)^2}$, $\quad \mathcal{L}''(\theta) > 0$.

$\therefore \quad L(\theta) = \ln\left(e^{\theta}-1\right) - \bar{x}\ln\theta$, is not convex in $\theta$ in general.

**09). d).**

We proved that $f(\theta) = \ln\left(e^{\theta} - 1\right)$ is concave in $\theta$ where $\theta > 0$.

Let $f_{T_1}(\theta)$ denote the first order Taylor approximation of $f(\theta)$ about $\theta = \theta_0$. Then,

$$
\begin{aligned}
f_{T_1}(\theta) &= f(\theta_0) + f'(\theta_0)(\theta - \theta_0) \\
&= \ln\left(e^{\theta} - 1\right)\Big|_{\theta=\theta_0} + \frac{e^{\theta}}{e^{\theta} - 1}\Big|_{\theta=\theta_0} (\theta - \theta_0) \\
f_{T_1}(\theta) &= \ln\left(e^{\theta_0} - 1\right) + \frac{e^{\theta_0}}{e^{\theta_0} - 1}(\theta - \theta_0)
\end{aligned}
$$

Note that for any $\theta_0 > 0$, the first order Taylor expansion of $f(\theta)$ about $\theta = \theta_0$ represents the equation of the tangential line drawn to $f(\theta)$ at $\theta_0$. As $f(\theta)$ is concave, this tangential line would lie above the curve of $f(\theta)$. In other words,

$$
\begin{aligned}
f(\theta) &\leq f_{T_1}(\theta) \\
\ln\left(e^{\theta} - 1\right) &\leq \ln\left(e^{\theta_0} - 1\right) + \frac{e^{\theta_0}}{e^{\theta_0} - 1}(\theta - \theta_0)
\end{aligned}
\tag{9.2}
$$

**09). e).**

From Eq. 9.2,

$$
\begin{aligned}
\ln\left(e^{\theta} - 1\right) &\leq \ln\left(e^{\theta_0} - 1\right) + \frac{e^{\theta_0}}{e^{\theta_0} - 1}(\theta - \theta_0) \\
\underbrace{\ln\left(e^{\theta} - 1\right) - \bar{x}\ln(\theta)}_{\mathcal{L}(\theta)} &\leq \ln\left(e^{\theta_0} - 1\right) + \frac{e^{\theta_0}}{e^{\theta_0} - 1}(\theta - \theta_0) - \bar{x}\ln(\theta)
\end{aligned}
$$

$$
\mathcal{L}(\theta) \leq \ln\left(e^{\theta_0} - 1\right) + \frac{e^{\theta_0}}{e^{\theta_0} - 1}(\theta - \theta_0) - \bar{x}\ln(\theta) \tag{9.3}
$$

Take $g(\theta|\theta_0) = \ln\left(e^{\theta_0} - 1\right) + \dfrac{e^{\theta_0}}{e^{\theta_0} - 1}(\theta - \theta_0) - \bar{x}\ln(\theta)$.

Note that $g'(\theta|\theta_0) = \dfrac{e^{\theta_0}}{e^{\theta_0} - 1} - \dfrac{\bar{x}}{\theta}$. Moreover, $g''(\theta|\theta_0) = \dfrac{\bar{x}}{\theta^2} > 0 \quad \forall\theta > 0$. Accordingly, we can say that $g(\theta|\theta_0)$ is a convex function.

In addition, we can state that,

$$
g(\theta \mid \theta_0)\,|_{\theta=\theta_0} = \mathcal{L}(\theta_0) = \ln\left(e^{\theta_0} - 1\right) - \bar{x}\ln(\theta_0)
$$

$$
\mathcal{L}(\theta) \leq g(\theta \mid \theta_0), \quad \forall\theta \neq \theta_0, \quad \theta > 0 \qquad \text{(From inequality 9.3)}
$$

Therefore, we can take the surrogate function $g(\theta|\theta_0)$ corresponding to $\mathcal{L}(\theta)$ to be,

$$
g(\theta|\theta_0) = \ln\left(e^{\theta_0} - 1\right) + \frac{e^{\theta_0}}{e^{\theta_0} - 1}(\theta - \theta_0) - \bar{x}\ln(\theta) \tag{9.4}
$$

**09). f).**

From Eq. 9.4, we have $g(\theta|\theta_0) = \ln\left(e^{\theta_0} - 1\right) + \frac{e^{\theta_0}}{e^{\theta_0} - 1}(\theta - \theta_0) - \bar{x}\ln(\theta)$. We know that $g'(\theta|\theta_0) = \frac{e^{\theta_0}}{e^{\theta_0} - 1} - \frac{\bar{x}}{\theta}$.

Suppose $g'(\theta|\theta_0) = 0$ when $\theta = \theta_1$. Then,

$$\begin{aligned} \frac{e^{\theta_0}}{e^{\theta_0} - 1} - \frac{\bar{x}}{\theta_1} &= 0 \\ \frac{\bar{x}}{\theta_1} &= \frac{e^{\theta_0}}{e^{\theta_0} - 1} \\ \theta_1 &= \frac{\left(e^{\theta_0} - 1\right)\bar{x}}{e^{\theta_0}} \end{aligned}$$

We can apply this process repeatedly. Accordingly, we can write in general,

$$\theta_k = \frac{\left(e^{\theta_{k-1}} - 1\right)\bar{x}}{e^{\theta_{k-1}}} \quad k = 1, 2, \ldots \tag{9.5}$$

**09). g).**

For $m = 10, \bar{x} = 2$, starting with $\theta_0 = 1$, the recursive equation (Eq. 9.5) was run for 13 iterations. Obtained $\theta$ and corresponding values of $\mathcal{L}(\theta)$ are tabulated below.

| Iteration | $\theta$ | $\mathcal{L}(\theta)$ |
|---|---|---|
| 0 | 1 | 0.541325 |
| 1 | 1.264241 | 0.463379 |
| 2 | 1.435093 | 0.440703 |
| 3 | 1.523813 | 0.435635 |
| 4 | 1.564241 | 0.434670 |
| 5 | 1.581507 | 0.434501 |
| 6 | 1.588670 | 0.434472 |
| 7 | 1.591606 | 0.434467 |
| 8 | 1.592803 | 0.434466 |
| 9 | 1.593290 | 0.434466 |
| 10 | 1.593489 | 0.434466 |
| 11 | 1.593569 | 0.434466 |
| 12 | 1.593601 | 0.434466 |
| 13 | 1.593615 | 0.434466 |

Table 1: Updated values of $\theta$ and $\mathcal{L}(\theta)$ - 13 Iterations

The shift in the surrogate function over 4 iterations is depicted in the following diagram. (Obtained $\theta$ and corresponding values of $\mathcal{L}(\theta)$ are indicated using blue dots.)

*Note: As $\theta_k$ and $\mathcal{L}(\theta_k)$ values were not changing significantly, only 4 iterations were plotted to maintain clarity in the diagram.*

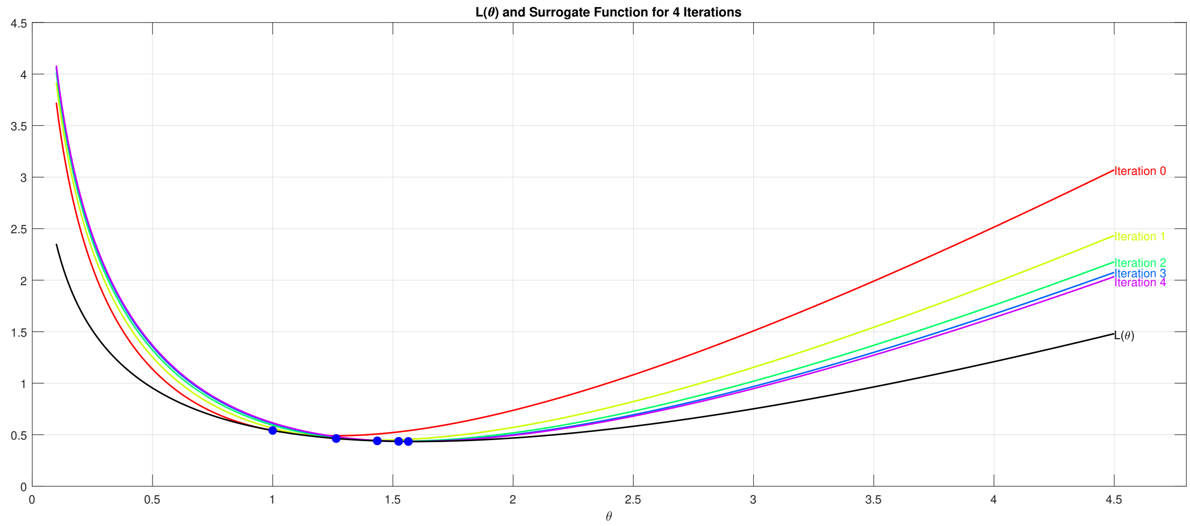


Figure 4: $L(\theta)$ and $g(\theta|\theta_k)$ where $k = 0, 1, \ldots, 4$

## Question 10 - Ridge Regression

Given $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, let us consider the optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_2^2 \qquad (\lambda > 0)$$

### 10). a).

Let $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_2^2$.

$$\begin{aligned}
f(\mathbf{x}) &= (\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \lambda\mathbf{x}^T\mathbf{x} \\
&= \left(\mathbf{x}^T\mathbf{A}^T - \mathbf{b}^T\right)(\mathbf{A}\mathbf{x} - \mathbf{b}) + \lambda\mathbf{x}^T\mathbf{x} \\
&= \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{x}^T\mathbf{A}^T\mathbf{b} + \mathbf{b}^T\mathbf{b} + \lambda\mathbf{x}^T\mathbf{x} \\
\frac{\mathrm{d}\, f(\mathbf{x})}{\mathrm{d}\mathbf{x}} &= \frac{\mathrm{d}\, \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}}{\mathrm{d}\mathbf{x}} - 2\frac{\mathrm{d}\, \mathbf{x}^T\mathbf{A}^T\mathbf{b}}{\mathrm{d}\mathbf{x}} + \underbrace{\frac{\mathrm{d}\, \mathbf{b}^T\mathbf{b}}{\mathrm{d}\mathbf{x}}}_{=\mathbf{0}} + \lambda\frac{\mathrm{d}\, \mathbf{x}^T\mathbf{x}}{\mathrm{d}\mathbf{x}}
\end{aligned}$$

Using Eq. 1.17, Eq. 1.21,

$$\begin{aligned}
\frac{\mathrm{d}\, f(\mathbf{x})}{\mathrm{d}\mathbf{x}} &= 2\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{A}^T\mathbf{b} + 2\lambda\mathbf{x} \\
&= 2\left[\mathbf{A}^T\mathbf{A}\mathbf{x} + \lambda\mathbf{x} - \mathbf{A}^T\mathbf{b}\right] \\
\frac{\mathrm{d}\, f(\mathbf{x})}{\mathrm{d}\mathbf{x}} &= 2\left[\left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}_n\right)\mathbf{x} - \mathbf{A}^T\mathbf{b}\right]
\end{aligned}$$

Suppose $\frac{\mathrm{d}\, f(\mathbf{x})}{\mathrm{d}\mathbf{x}} = \mathbf{0}$ when $\mathbf{x} = \mathbf{x}_R(\lambda)$. Then,

$$\begin{aligned}
2\left[\left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}_n\right)\mathbf{x}_R(\lambda) - \mathbf{A}^T\mathbf{b}\right] &= \mathbf{0} \\
\left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}_n\right)\mathbf{x}_R(\lambda) &= \mathbf{A}^T\mathbf{b} \\
\mathbf{x}_R(\lambda) &= \left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}_n\right)^{-1}\mathbf{A}^T\mathbf{b} \qquad (10.1)
\end{aligned}$$

Accordingly, we can say that $f(\mathbf{x})$ is minimized by $\mathbf{x}_R(\lambda) = \left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}_n\right)^{-1}\mathbf{A}^T\mathbf{b}$.

### 10). b).

Suppose $\text{rank}(\mathbf{A}) = r \leq n$. Consider the singular value decomposition of the matrix $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Take $\mathbf{U} = \begin{pmatrix}\mathbf{u_1} & \mathbf{u_2} & \cdots & \mathbf{u_m}\end{pmatrix}, \mathbf{V} = \begin{pmatrix}\mathbf{v_1} & \mathbf{v_2} & \cdots & \mathbf{v_n}\end{pmatrix}$ and let $\mathbf{\Sigma}$ contain the singular values of $\mathbf{A}$; with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Note that $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_m$ and $\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$.

Using Eq. 10.1, we can write that,

$$\begin{aligned}
\mathbf{x}_R(\lambda) &= \left(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I}_n\right)^{-1}\mathbf{A}^T\mathbf{b} \\
&= \left(\left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\right)^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T + \lambda\mathbf{I}_n\right)^{-1}\left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\right)^T\mathbf{b} \\
&= \left(\mathbf{V}\mathbf{\Sigma}^T\underbrace{\mathbf{U}^T\mathbf{U}}_{\mathbf{I}_m}\mathbf{\Sigma}\mathbf{V}^T + \lambda\mathbf{V}\mathbf{V}^T\right)^{-1}\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{b}
\end{aligned}$$

Take $\mathbf{D} = \mathbf{\Sigma}^T\mathbf{\Sigma} = \begin{pmatrix}\sigma_1^2 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0\end{pmatrix}$.

$$\begin{aligned}
\mathbf{x}_R(\lambda) &= \left(\mathbf{V}\mathbf{D}\mathbf{V}^T + \lambda\mathbf{V}\mathbf{V}^T\right)^{-1}\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{b} \\
&= \left(\mathbf{V}\mathbf{D}\mathbf{V}^T + \mathbf{V}\lambda\mathbf{I}_n\mathbf{V}^T\right)^{-1}\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{b} \\
&= \left(\mathbf{V}\left(\mathbf{D} + \lambda\mathbf{I}_n\right)\mathbf{V}^T\right)^{-1}\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{b} \\
&= \mathbf{V}\left(\mathbf{D} + \lambda\mathbf{I}_n\right)^{-1}\underbrace{\mathbf{V}^T\mathbf{V}}_{\mathbf{I}_n}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{b} \\
\mathbf{x}_R(\lambda) &= \mathbf{V}\left(\mathbf{D} + \lambda\mathbf{I}_n\right)^{-1}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{b} \qquad (10.2)
\end{aligned}$$

Note that $\left(\mathbf{D} + \lambda\mathbf{I}_n\right)^{-1} = \begin{pmatrix}\frac{1}{\sigma_1^2+\lambda} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2+\lambda} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_r^2+\lambda} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \frac{1}{\lambda}\end{pmatrix}$.

$$
(\mathbf{D}+\lambda\mathbf{I}_n)^{-1}\,\mathbf{\Sigma}^T = \begin{pmatrix} \frac{1}{\sigma_1^2+\lambda} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2+\lambda} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_r^2+\lambda} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \frac{1}{\lambda} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_r & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}
$$

$$
= \begin{pmatrix} \frac{\sigma_1}{\sigma_1^2+\lambda} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \frac{\sigma_2}{\sigma_2^2+\lambda} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_r}{\sigma_r^2+\lambda} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}
$$

$$
(\mathbf{D}+\lambda\mathbf{I}_n)^{-1}\,\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{b} = \begin{pmatrix} \frac{\sigma_1}{\sigma_1^2+\lambda} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \frac{\sigma_2}{\sigma_2^2+\lambda} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_r}{\sigma_r^2+\lambda} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u_1}^T \\ \mathbf{u_2}^T \\ \vdots \\ \mathbf{u_m}^T \end{pmatrix} \mathbf{b}
$$

$$
= \begin{pmatrix} \frac{\sigma_1}{\sigma_1^2+\lambda} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \frac{\sigma_2}{\sigma_2^2+\lambda} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_r}{\sigma_r^2+\lambda} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u_1}^T\mathbf{b} \\ \mathbf{u_2}^T\mathbf{b} \\ \vdots \\ \mathbf{u_m}^T\mathbf{b} \end{pmatrix}
$$

$$
(\mathbf{D}+\lambda\mathbf{I}_n)^{-1}\,\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{b} = \begin{pmatrix} \left(\frac{\sigma_1}{\sigma_1^2+\lambda}\right)\mathbf{u_1}^T\mathbf{b} \\ \left(\frac{\sigma_2}{\sigma_2^2+\lambda}\right)\mathbf{u_2}^T\mathbf{b} \\ \vdots \\ \left(\frac{\sigma_r}{\sigma_r^2+\lambda}\right)\mathbf{u_r}^T\mathbf{b} \\ 0 \\ \vdots \\ 0 \end{pmatrix}
$$

$$\underbrace{\mathbf{V}\left(\mathbf{D}+\lambda\mathbf{I}_n\right)^{-1}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{b}}_{=\mathbf{x}_R(\lambda) \quad \text{From Eq. 10.2}} = \mathbf{V}\begin{pmatrix} \left(\frac{\sigma_1}{\sigma_1^2+\lambda}\right)\mathbf{u_1}^T\mathbf{b} \\ \left(\frac{\sigma_2}{\sigma_2^2+\lambda}\right)\mathbf{u_2}^T\mathbf{b} \\ \vdots \\ \left(\frac{\sigma_r}{\sigma_r^2+\lambda}\right)\mathbf{u_r}^T\mathbf{b} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\mathbf{x}_R(\lambda) = \begin{pmatrix} \mathbf{v_1} & \mathbf{v_2} & \cdots & \mathbf{v_n} \end{pmatrix}\begin{pmatrix} \left(\frac{\sigma_1}{\sigma_1^2+\lambda}\right)\mathbf{u_1}^T\mathbf{b} \\ \left(\frac{\sigma_2}{\sigma_2^2+\lambda}\right)\mathbf{u_2}^T\mathbf{b} \\ \vdots \\ \left(\frac{\sigma_r}{\sigma_r^2+\lambda}\right)\mathbf{u_r}^T\mathbf{b} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Using the outer product,

$$\mathbf{x}_R(\lambda) = \left(\frac{\sigma_1\mathbf{u_1}^T\mathbf{b}}{\sigma_1^2+\lambda}\right)\mathbf{v_1} + \left(\frac{\sigma_2\mathbf{u_2}^T\mathbf{b}}{\sigma_2^2+\lambda}\right)\mathbf{v_2} + \cdots + \left(\frac{\sigma_1\mathbf{u_r}^T\mathbf{b}}{\sigma_r^2+\lambda}\right)\mathbf{v_r}$$

$$\mathbf{x}_R(\lambda) = \sum_{j=1}^{r}\left(\frac{\sigma_j\mathbf{u_j}^T\mathbf{b}}{\sigma_j^2+\lambda}\right)\mathbf{v_j} \tag{10.3}$$

## 10). c).

Using the result in Eq. 10.3, we can write,

$$\|\mathbf{x}_R(\lambda)\|_2^2 = \sum_{j=1}^{r}\underbrace{\left(\frac{\sigma_j\mathbf{u_j}^T\mathbf{b}}{\sigma_j^2+\lambda}\right)^2}_{k_j(\lambda)}\|\mathbf{v_j}\|_2^2$$

Note that $\forall j = 1, 2, \ldots, r$, $k_j(\lambda)$ is a decreasing function of $\lambda$. Thus, we will obtain the $\sup_{\lambda}\|\mathbf{x}_R(\lambda)\|_2^2$ when $\lambda = 0$.

$$\sup_{\lambda}\|\mathbf{x}_R(\lambda)\|_2^2 = \sum_{j=1}^{r}\left(\frac{\mathbf{u_j}^T\mathbf{b}}{\sigma_j}\right)^2\|\mathbf{v_j}\|_2^2$$

Accordingly from Eq. 8.6, we can see that $\lambda = 0$ yields,

$$\mathbf{x}_R(\lambda = 0) = \mathbf{x}^*_{LS} = \sum_{j=1}^{r} \frac{\mathbf{u_j}^T\mathbf{b}}{\sigma_j}\mathbf{v_j}$$

Thus, the supremum corresponds to the vector $\mathbf{x}_R(\lambda = 0)$ which is identical to the least squares solution we get when there is no regularization.

### 10). d).

$$\begin{aligned}
\min_{\mathbf{x}\in\mathbb{R}^n} \left\| \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_n \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \right\|_2^2 &= \min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \|\mathbf{I}_n\mathbf{x} - \mathbf{0}\|_2^2 \\
&= \min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \|\mathbf{x}\|_2^2 \\
&= \min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_2^2 && (\lambda = 1) \\
&= \mathbf{x}_R(\lambda = 1) \\
\therefore \quad \min_{\mathbf{x}\in\mathbb{R}^n} \left\| \begin{pmatrix} \mathbf{A} \\ \mathbf{I}_n \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \right\|_2^2 &= \sum_{j=1}^{r} \left( \frac{\sigma_j \mathbf{u_j}^T\mathbf{b}}{\sigma_j^2 + 1} \right) \mathbf{v_j} && \text{(From Eq. 10.3)}
\end{aligned}$$