Problem 1: Let $\boldsymbol{x} \in \mathbb{R}^M$, $\boldsymbol{y} \in \mathbb{R}^N$ and $\boldsymbol{Z} \in \mathbb{R}^{P \times Q}$. The function $f : \mathbb{R}^M \times \mathbb{R}^N \times \mathbb{R}^{P \times Q} \to \mathbb{R}$ is defined as

$$f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}) = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{y} + \boldsymbol{B} \boldsymbol{x} - \boldsymbol{y}^T \boldsymbol{C} \boldsymbol{Z} \boldsymbol{D} - \boldsymbol{y}^T \boldsymbol{E}^T \boldsymbol{y} + \boldsymbol{F}.$$

What should be the dimensions (shapes) of the matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}, \boldsymbol{F}$ for the expression above to be a valid mathematical expression?

$x^T A y \quad + \quad B x \quad - \quad y^T C Z D \quad - \quad y^T E^T y \quad + F$

x^T: $1 \times M$, y: $N \times 1$, A: $M \times N$, $x^T A y$: 1×1

B: $1 \times M$, x: $M \times 1$, Bx: 1×1

y^T: $1 \times N$, Z: $P \times Q$, C: $N \times P$, D: $Q \times 1$

y^T: $1 \times N$, y: $N \times 1$, E: $N \times N$

F: 1×1

Problem 2: Let $\boldsymbol{x} \in \mathbb{R}^N, \boldsymbol{M} \in \mathbb{R}^{N \times N}$. Express the function $f(\boldsymbol{x}) = \sum_{i=1}^N \sum_{j=1}^N x_i x_j M_{ij}$ using **only** matrix-vector multiplications.

$x^T M x$

$$\sum_{i=1}^N \sum_{j=1}^N x_i x_j M_{ij} = \sum_{i=1}^N x_i \sum_{j=1}^N x_j M_{ij}$$

$$= \sum_{i=1}^N x_i \left(x M_i \right) = x^T x M$$

Problem 3: Let $\boldsymbol{A} \in \mathbb{R}^{M \times N}, \boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{b} \in \mathbb{R}^M$. We are interested in solving the following system of linear equations for $\boldsymbol{x}$

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{1}$$

a) Under what conditions does the system of linear equations have a **unique** solution $\boldsymbol{x}$ for any choice of $\boldsymbol{b}$?

b) Assume that $M = N = 5$ and that $\boldsymbol{A}$ has the following eigenvalues: $\{-5, 0, 1, 1, 3\}$. Does Equation 1 have a unique solution $\boldsymbol{x}$ for any choice of $\boldsymbol{b}$? Justify your answer.

a) $\operatorname{rank}(Ax) = \operatorname{rank}(b) \Rightarrow$ unique

b) The rank of A is equal to the number of its non-zero eigenvalues. which here is $\operatorname{rank}(A) = 4 \Rightarrow \operatorname{rank}(A) = 4 \neq 5 \overset{a)}{\Rightarrow}$ no unique solution

Problem 4: Let $\boldsymbol{A} \in \mathbb{R}^{N \times N}$. Assume that there exists a matrix $\boldsymbol{B} \in \mathbb{R}^{N \times N}$ such that $\boldsymbol{BA} = \boldsymbol{AB} = \boldsymbol{I}$. What can you say about the eigenvalues of $\boldsymbol{A}$? Justify your answer.

$AB = BA = I \Rightarrow B = A^{-1} \Rightarrow$ A invertible

Since A is invertible and it has full rank, its eigenvalues cannot be 0.

Problem 7: Consider the following function $f : \mathbb{R} \to \mathbb{R}$

$$f(x) = \frac{1}{2}ax^2 + bx + c$$

We are interested in solving the following optimization problem

$$\min_{x \in \mathbb{R}} f(x)$$

a) Under what conditions does this optimization problem have (i) a unique solution, (ii) infinitely many solutions or (iii) no solution? Justify your answer.

a) Assume that the optimization problem has a unique solution. Write down the closed-form expression for $x^\star$ that minimizes the objective function, i.e. find $x^\star = \arg\min_{x \in \mathbb{R}} f(x)$.

a) i) $a = 1$ $\quad \cup \quad$ (unique solution is $-b$ (see b))

ii) $a = b = 0$ $\quad - \quad$ $f'(x) = 0 \Rightarrow f(x) = c \;\; \forall x \in \mathbb{R}$

iii) $a = -1$ $\quad \cap \quad$ $f'(x) = -x + b$

$f''(x) = -1 < 0$

$\Rightarrow$ concave $\Rightarrow$ no minimum

b) $f'(x) = ax + b$

$\overset{!}{=} 0$

$\Rightarrow ax + b = 0$

$\Leftrightarrow ax = -b$

$\Leftrightarrow x = -\frac{b}{a}$

Problem 8: Consider the following function $g : \mathbb{R}^N \to \mathbb{R}$

$$g(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$$

where $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is a PSD matrix, $\boldsymbol{b} \in \mathbb{R}^N$ and $c \in \mathbb{R}$.

We are interested in solving the following optimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} g(\boldsymbol{x})$$

a) Compute the Hessian $\nabla^2 g(\boldsymbol{x})$ of the objective function. Under what conditions does this optimization problem have a unique solution?

b) Why is it necessary for a matrix $\boldsymbol{A}$ to be PSD for the optimization problem to be well-defined? What happens if $\boldsymbol{A}$ has a negative eigenvalue?

c) Assume that the matrix $\boldsymbol{A}$ is positive definite (PD). Write down the closed-form expression for $\boldsymbol{x}^\star$ that minimizes the objective function, i.e. find $\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} g(\boldsymbol{x})$.

a) $\nabla g(x) = Ax + b^T$

$\nabla^2 g(x) = A$

A must be PSD $\Rightarrow$ extreme points are all $b \geq 0$

(every local minimum is a global minimum)

b) - If A is PSD, then $\nabla^2 g(x)$ must also be PSD and hence f is convex.

- If A has an eigenvalue < 0, then it is not PSD anymore and so not neccessarily convex anymore.

c) $\nabla g(x) \stackrel{!}{=} 0$ $\quad$ arg min $g(x)$ = arg max $-g(x)$

$\Leftrightarrow Ax + b^T = 0$

$\Leftrightarrow Ax = -b^T$

If A PD then $x^* = -b^T A^{-1}$

Problem 9: Prove or disprove the following statement

$$p(A|B,C) = p(A|C) \Rightarrow p(A|B) = p(A)$$

$P(A|B,C) = P(A|C)$ Gegenbeispiel, ist falsch

$P(x=1) = P(1)$

Problem 10: Prove or disprove the following statement

$$p(A|B) = p(A) \Rightarrow p(A|B,C) = p(A|C)$$

$P(A,C) = P(A \cap C)$

$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$

$$P(A) = P(A|B) \overset{\text{Bayes}}{=} \frac{P(A,B)}{P(B)} = \frac{P(A)\,P(B)}{P(B)} = P(A)$$

$$P(A|B,C) = \frac{P(A,B|C)}{P(B|C)} = \frac{P(A|C)\,P(B|C)}{P(B|C)} = P(A|C) \qquad \square$$

Problem 11: You are given the joint PDF $p(a,b,c)$ of three continuous random variables. Show how the following expressions can be obtained using the rules of probability

1. $p(a)$
2. $p(c|a,b)$
3. $p(b|c)$

$$p(a) = \int_b \int_c p(a,b,c) \, dc \, db$$

$$p(c|a,b) = \frac{p(a,b,c)}{p(a,b)} = \frac{\int_a \int_b p(a,b,c) \, db \, da}{\int_c p(a,b,c) \, dc}$$

$$p(b|c) = \frac{p(b,c)}{p(c)} = \frac{\int_a \int_c p(a,b,c) \, dc \, da}{\int_a \int_b p(a,b,c) \, db \, da}$$

Problem 12: Researchers have developed a test which determines whether a person has a rare disease. The test is fairly reliable: if a person is sick, the test will be positive with 95% probability, if a person is healthy, the test will be negative with 95% probability. It is known that $\frac{1}{1000}$ of the population have this rare disease. A person (chosen uniformly at random from the population) takes the test and obtains a positive result. What is the probability that the person has the disease?

$P(T)$ = test positive $\quad P(S)$ = sick

$P(\neg T)$ = test negative $\quad P(\neg S)$ = healthy

$P(T|S) = 0.95$

$P(\neg T|\neg S) = 0.95$

$P(T|\neg S) = 0.05$

$P(\neg T|S) = 0.05$

$P(S) = \frac{1}{1000} = 0.001$

$P(S|T)$?

$$P(S|T) = \frac{P(T|S)\,P(S)}{P(T|S)\,P(S) + P(T|\neg S)\,P(\neg S)}$$

$$= \frac{0.95 \cdot 0.001}{0.95 \cdot 0.001 + 0.05 \cdot 0.999}$$

$$\approx 0.019$$

$$= 1.9\%$$

Problem 13: Let $X \sim \mathcal{N}(\mu, \sigma^2)$, and $f(x) = ax + bx^2 + c$. What is $\mathbb{E}[f(x)]$?

$$\mathbb{E}[f(x)]$$
$$= \mathbb{E}[ax + bx^2 + c]$$
$$= \mathbb{E}[ax] + \mathbb{E}[bx^2] + \mathbb{E}[c]$$
$$= a\mathbb{E}[x] + b\mathbb{E}[x^2] + c$$
$$= a\mu + b \cdot \mathbb{E}[(x - \mu + \mu)^2] + c$$

$$= a\mu + c + b\Big(\mathbb{E}\big[(\overbrace{(x-\mu)}^{\phi}+\mu)(\overbrace{(x-\mu)}^{\phi}+\mu)\big]\Big)$$
$$= a\mu + c + b\left(\mathbb{E}[(\phi + \mu)^2]\right)$$
$$= a\mu + c + b\left(\mathbb{E}[\phi^2 + 2\phi\mu + \mu^2]\right)$$
$$= a\mu + c + b\left(\mathbb{E}[\phi^2] + \mathbb{E}[2\phi\mu] + \mathbb{E}[\mu^2]\right)$$
$$= a\mu + c + b\left(\mathbb{E}[(x-\mu)^2] + 2\mu\mathbb{E}[x-\mu] + \mathbb{E}[\mu^2]\right)$$
$$= a\mu + c + b(\sigma^2 + 0 + \mu^2)$$
$$= a\mu + c + b\sigma^2 + b\mu^2$$

$$\mathbb{E}[x-\mu] = \mathbb{E}[x] - \mathbb{E}[\mu] = \mu - \mu = 0$$
$$\mathrm{Var}[x] = \sigma^2 = \mathbb{E}[(x-\mu)^2]$$
$$= \mathbb{E}[x^2] - (\mathbb{E}[x])^2$$

$$\mathbb{E}[(x-\mu)^2]$$
$$= \mathbb{E}[x^2 - 2x\mu + \mu^2]$$
$$= \mathbb{E}[x^2] - 2\mu^2 + \mu^2$$
$$= \mathbb{E}[x^2] - \mu^2$$

How to get σ^2?

Problem 14: Let $p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $g(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ (where $\boldsymbol{A} \in \mathbb{R}^{N \times N}$). What are the values of the following expressions:

- $\mathbb{E}[g(\boldsymbol{x})]$,
- $\mathbb{E}[g(\boldsymbol{x})g(\boldsymbol{x})^T]$,
- $\mathbb{E}[g(\boldsymbol{x})^T g(\boldsymbol{x})]$,
- the covariance matrix $\mathrm{Cov}[g(\boldsymbol{x})]$.

- $\mathbb{E}[g(x)] = \mathbb{E}[Ax] = A\mathbb{E}[x] = A\mu$
- $\mathbb{E}[g(x)g(x)^T] = \mathbb{E}[Ax(Ax)^T] = A\mathbb{E}[x(Ax)^T] = A\mathbb{E}[x x^T A^T]$

$$= A\mathbb{E}[xx^T]A^T = \underset{N\times N}{A}\,(\underset{N\times N}{\Sigma} + \underset{N\times N}{\mu\mu^T})\,\underset{N\times N}{A^T}$$

$$= AA^T\Sigma + AA^T\mu\mu^T$$

- $\mathbb{E}[g(x)^T g(x)] = \mathbb{E}[(Ax)^T Ax]$

$$= \mathbb{E}[x^T A^T A x] \qquad B := A^T A$$

$$= \mathbb{E}\left[\sum_{i=1}^{N} \sum_{j=1}^{N} B_{i,j} x_i x_j\right]$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} B_{i,j} \mathbb{E}[x_i x_j]$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} B_{i,j} (\sigma_{i,j} + \mu_i \mu_j)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} B_{i,j} \sigma_{i,j} + \sum_{i=1}^{N} \sum_{j=1}^{N} B_{i,j} \mu_i \mu_j$$

$$= \sum_{i=1}^{N} (B\Sigma)_{i,i} + \mu^T B \mu$$

$$= tr(A^T A \Sigma) + \mu^T A^T A \mu$$

- $\mathrm{Cov}[g(x)] = \mathrm{Cov}[Ax]$

$$= \mathbb{E}[(Ax - \mathbb{E}[Ax])(Ax - \mathbb{E}[Ax])^T]$$

$$= \mathbb{E}[(Ax - A\mathbb{E}[x])(Ax - A\mathbb{E}[x])^T]$$

$$= \mathbb{E}[A(x - \mathbb{E}[x])(x - \mathbb{E}[x])^T A^T]$$

$$= A\, \mathbb{E}[(x - \mathbb{E}[x])(x - \mathbb{E}[x])^T] A^T$$

$$= A\, \mathrm{Cov}(x) A^T$$

$$= A \Sigma A^T$$