Homework 0

A1. (2 points)

(a)

Let X be the event that I test positive for the disease and Y be the event that I have the disease.
We are given:

- $\mathbb{P}\{X|Y\} = \frac{99}{100}$
- $\mathbb{P}\{X^C|Y^C\} = \frac{99}{100}$
- $\mathbb{P}\{Y\} = \frac{1}{10{,}000}$

Our goal is to find the probability that I have the disease given that I tested positive (i.e. $\mathbb{P}\{Y|X\}$).
We can use Bayes' Theorem:

$$\begin{aligned}
\mathbb{P}\{Y|X\} &= \frac{\mathbb{P}\{X|Y\} \cdot \mathbb{P}\{Y\}}{\mathbb{P}\{X\}} \\
&= \frac{\mathbb{P}\{X|Y\} \cdot \mathbb{P}\{Y\}}{\mathbb{P}\{X|Y\} \cdot \mathbb{P}\{Y\} + \mathbb{P}\{X|Y^C\} \cdot \mathbb{P}\{Y^C\}} && \text{(Law of Total Probability on } \mathbb{P}\{X\}\text{)} \\
&= \frac{\mathbb{P}\{X|Y\} \cdot \mathbb{P}\{Y\}}{\mathbb{P}\{X|Y\} \cdot \mathbb{P}\{Y\} + (1 - \mathbb{P}\{X^C|Y^C\}) \cdot (1 - \mathbb{P}\{Y\})} && \text{(Complement Rule)} \\
&= \frac{\frac{99}{100} \cdot \frac{1}{10{,}000}}{\frac{99}{100} \cdot \frac{1}{10{,}000} + (1 - \frac{99}{100}) \cdot (1 - \frac{1}{10{,}000})} && \text{(plugging in given values)} \\
&= \frac{99}{10098} \\
&\approx 0.0098
\end{aligned}$$

Therefore, the probability that I actually have the disease is $\boxed{\frac{99}{10098} \approx 0.0098}$.

A2. (2 points)

(a)

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] && \text{(definition of covariance)}\\
&= \mathbb{E}[XY - X\,\mathbb{E}[Y] - Y\,\mathbb{E}[X] + \mathbb{E}[X]\,\mathbb{E}[Y]]\\
&= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[Y]\,\mathbb{E}[X] + \mathbb{E}[X]\,\mathbb{E}[Y] && \text{(linearity of expectation)}\\
&= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]\\
&= \Big(\sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} xy\cdot\mathbb{P}\{X=x, Y=y\}\Big) - \mathbb{E}[X]\,\mathbb{E}[Y] && \text{(definition of expectation)}\\
&= \Big(\sum_{x\in\Omega_X} x\cdot\Big(\sum_{y\in\Omega_Y} y\cdot\mathbb{P}\{Y|X=x\}\Big)\cdot\mathbb{P}\{X=x\}\Big) - \mathbb{E}[X]\,\mathbb{E}[Y]\\
&= \Big(\sum_{x\in\Omega_X} x\cdot\mathbb{E}[Y|X=x]\cdot\mathbb{P}\{X=x\}\Big) - \mathbb{E}[X]\,\mathbb{E}[Y] && \text{(definition of conditional expectation)}\\
&= \Big(\sum_{x\in\Omega_X} x^2\cdot\mathbb{P}\{X=x\}\Big) - \mathbb{E}[X]\,\mathbb{E}[Y] && \text{(Since } \mathbb{E}[Y|X=x]=x\text{)}\\
&= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]\,\mathbb{E}[Y] && \text{(definition of expectation)}\\
&= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]\cdot\sum_{x\in\Omega_X}\mathbb{E}[Y|X=x]\cdot\mathbb{P}\{X=x\} && \text{(Law of Total Expectation)}\\
&= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]\cdot\sum_{x\in\Omega_X} x\cdot\mathbb{P}\{X=x\} && \text{(Since } \mathbb{E}[Y|X=x]=x\text{)}\\
&= \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 && \text{(definition of expectation)}\\
&= \mathbb{E}\big[X^2\big] - 2\,\mathbb{E}[X]^2 + \mathbb{E}[X]^2\\
&= \mathbb{E}\Big[X^2 - 2X\,\mathbb{E}[X] + \mathbb{E}[X]^2\Big] && \text{(linearaity of expectation)}\\
&= \mathbb{E}\big[(X-\mathbb{E}[X])^2\big]
\end{aligned}
$$

(b)

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] && \text{(definition of covariance)}\\
&= \mathbb{E}[XY - X\,\mathbb{E}[Y] - Y\,\mathbb{E}[X] + \mathbb{E}[X]\,\mathbb{E}[Y]]\\
&= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[Y]\,\mathbb{E}[X] + \mathbb{E}[X]\,\mathbb{E}[Y] && \text{(linearity of expectation)}\\
&= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]\\
&= \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[X]\,\mathbb{E}[Y] && \text{(since } X \text{ and } Y \text{ are independent)}\\
&= 0
\end{aligned}
$$

A3. (2 points)

(a)

$$
\begin{aligned}
h(z)\,\mathrm{d}z &= \mathbb{P}\{z \le Z \le z + \mathrm{d}z\} && \text{(Probability of } Z = z \text{ (continuous))}\\
\Rightarrow h(z)\,\mathrm{d}z &= \mathbb{P}\{z \le X + Y \le z + \mathrm{d}z\} && \text{(Since } Z = X + Y\text{)}\\
\Rightarrow h(z)\,\mathrm{d}z &= \int_{-\infty}^{\infty} f(x)\cdot \mathbb{P}\{z - x \le Y \le z - x + \mathrm{d}z | X = x\}\,\mathrm{d}x \\
& && \text{(Law of Total Probability (continuous))}\\
\Rightarrow h(z)\,\mathrm{d}z &= \int_{-\infty}^{\infty} f(x)\cdot \mathbb{P}\{z - x \le Y \le z - x + \mathrm{d}z\}\,\mathrm{d}x && \text{(Since } X \text{ and } Y \text{ are independent)}\\
\Rightarrow h(z)\,\mathrm{d}z &= \int_{-\infty}^{\infty} f(x)\cdot g(z-x)\,\mathrm{d}x\,\mathrm{d}z && \text{(Def. PDF)}\\
\Rightarrow h(z) &= \int_{-\infty}^{\infty} f(x)\cdot g(z-x)\,\mathrm{d}x && \text{(}\mathrm{d}z \text{ cancels out)}
\end{aligned}
$$

(b)

When $0 \le z \le 1$, since $z = x + y$, we know $0 \le x \le z$.
Since X and Y are uniform on $[0, 1]$, we have:

$$h(z) = \int_{-\infty}^{\infty} f(x)\cdot g(z-x)\,\mathrm{d}x = \int_{0}^{z} 1\cdot 1\,\mathrm{d}x = z$$

When $1 < z \le 2$, since $z = x + y$, we know $z - 1 \le x \le 1$.
Since X and Y are uniform on $[0, 1]$, we have:

$$h(z) = \int_{-\infty}^{\infty} f(x)\cdot g(z-x)\,\mathrm{d}x = \int_{z-1}^{1} 1\cdot 1\,\mathrm{d}x = 2 - z$$

Since X and Y are uniform on $[0, 1]$, we know when $z > 2$ or $z < 0$, $h(z) = 0$.
Therefore, the PDF of Z is:

$$\boxed{h(z) = \begin{cases} z & 0 \le z \le 1 \\ 2 - z & 1 < z \le 2 \\ 0 & \text{otherwise} \end{cases}}$$

A4. (4 points)

(a)

Since $X_1 \sim \mathcal{N}(\mu, \sigma^2)$, we know $aX_1 + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
Since $X_1 \sim \mathcal{N}(0, 1)$, we have:

$$\begin{cases} a\mu + b = 0 \\ a^2\sigma^2 = 1 \end{cases} \underset{\text{solving } a, b \text{ in terms of } \sigma, \mu}{\Longrightarrow} \boxed{\begin{cases} a = \frac{1}{\sigma} \\ b = -\frac{\mu}{\sigma} \end{cases}}$$

(b)

Since X_1 and X_2 are i.i.d., we have:

$$\begin{aligned} \mathbb{E}[X_1 + 2X_2] &= \mathbb{E}[X_1] + 2\,\mathbb{E}[X_2] && \text{(linearity of expectation)} \\ &= \mu + 2\mu && \text{(since } X_1, X_2 \sim \mathcal{N}(\mu, \sigma^2)\text{)} \\ &= \boxed{3\mu} \end{aligned}$$

and,

$$\begin{aligned} \mathrm{Var}[X_1 + 2X_2] &= \mathrm{Var}[X_1] + 2^2\,\mathrm{Var}[X_2] && \text{(since } X_1 \text{ and } X_2 \text{ are independent)} \\ &= \sigma^2 + 4\sigma^2 && \text{(since } X_1, X_2 \sim \mathcal{N}(\mu, \sigma^2)\text{)} \\ &= \boxed{5\sigma^2} \end{aligned}$$

(c)

$$
\begin{aligned}
\mathbb{E}\big[\sqrt{n}(\widehat{\mu}_n - \mu)\big] &= \mathbb{E}\left[\sqrt{n}((\frac{1}{n}\sum_{i=1}^{n} X_i) - \mu)\right] && \text{(definition of } \widehat{\mu}_n\text{)} \\
&= (\frac{\sqrt{n}}{n}\sum_{i=1}^{n} \mathbb{E}[X_i]) - (\sqrt{n}\,\mathbb{E}[\mu]) && \text{(linearity of expectation)} \\
&= (\frac{\sqrt{n}}{n}\sum_{i=1}^{n} \mu) - (\sqrt{n}\,\mathbb{E}[\mu]) && \text{(since } X_i \sim \mathcal{N}(\mu, \sigma^2)\text{)} \\
&= \frac{\sqrt{n}}{n} \cdot n \cdot \mu - \sqrt{n} \cdot \mu \\
&= \boxed{0}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}\big[\sqrt{n}(\widehat{\mu}_n - \mu)\big] &= \mathrm{Var}\left[\sqrt{n}((\frac{1}{n}\sum_{i=1}^{n} X_i) - \mu)\right] && \text{(definition of } \widehat{\mu}_n\text{)} \\
&= \sqrt{n}^2\,\mathrm{Var}\left[(\frac{1}{n}\sum_{i=1}^{n} X_i) - \mu\right] && \text{(since } \sqrt{n} \text{ is a constant)} \\
&= n\,\mathrm{Var}\left[(\frac{1}{n}\sum_{i=1}^{n} X_i) - \mu\right] \\
&= n\,\mathrm{Var}\left[(\frac{1}{n}\sum_{i=1}^{n} X_i)\right] && \text{(since } \mu \text{ is a constant)} \\
&= n \cdot \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[X_i] && \text{(since } X_i \text{ are independent)} \\
&= \frac{1}{n}\sum_{i=1}^{n} \sigma^2 && \text{(since } X_i \sim \mathcal{N}(\mu, \sigma^2)\text{)} \\
&= \frac{1}{n} \cdot n \cdot \sigma^2 \\
&= \boxed{\sigma^2}
\end{aligned}
$$

A5. (4 points)

(a)

We know the rank of a matrix is the number of linearly independent columns in the matrix.
We can use Gaussian Elimination to find the reduced row echelon form of the matrices:

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow[R_3=R_3-R_1]{R_2=R_2-R_1} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{R_2=-\frac{1}{2}\cdot R_2} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix} \xrightarrow{R_3=R_2+R_3} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \Longrightarrow \boxed{\text{rank: } 2}$$

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \xrightarrow[R_3=R_1-R_3]{R_2=\frac{1}{2}\cdot(R_1-R_2)} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow{R_3=R_2-R_3} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \Longrightarrow \boxed{\text{rank: } 2}$$

(b)

From (a.), we know the reduced row echelon form of A is: $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$.

Thus, the basis for A's column span is $\boxed{\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}}$.

From (a.), we know the reduced row echelon form of B is: $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

Thus, the basis for B's column span also is $\boxed{\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \right\}}$.

A6. (3 points)

(a)

$$Ac = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top = \begin{bmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0+2+4 \\ 2+4+2 \\ 3+3+1 \end{bmatrix} = \boxed{\begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}}$$

(b)

$$Ax = b \Rightarrow A^{-1}Ax = A^{-1}b \Rightarrow x = A^{-1}b$$

$$\begin{bmatrix} A & | & I \end{bmatrix} = \begin{bmatrix} 0 & 2 & 4 & | & 1 & 0 & 0 \\ 2 & 4 & 2 & | & 0 & 1 & 0 \\ 3 & 3 & 1 & | & 0 & 0 & 1 \end{bmatrix} \xrightarrow[R_2 = \frac{1}{2}R_1]{R_1 = \frac{1}{2}R_2} \begin{bmatrix} 1 & 2 & 1 & | & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 2 & | & \frac{1}{2} & 0 & 0 \\ 3 & 3 & 1 & | & 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_3 = R_3 - 3R_1}$$

$$\begin{bmatrix} 1 & 2 & 1 & | & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 2 & | & \frac{1}{2} & 0 & 0 \\ 0 & -3 & -2 & | & 0 & -\frac{3}{2} & 1 \end{bmatrix} \xrightarrow[R_3 = R_3 + 3R_2]{R_1 = R_1 - 2R_2} \begin{bmatrix} 1 & 0 & -3 & | & -1 & \frac{1}{2} & 0 \\ 0 & 1 & 2 & | & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 4 & | & \frac{3}{2} & -\frac{3}{2} & 1 \end{bmatrix} \xrightarrow{R_3 = \frac{1}{4}R_3}$$

$$\begin{bmatrix} 1 & 0 & -3 & | & -1 & \frac{1}{2} & 0 \\ 0 & 1 & 2 & | & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & | & \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{bmatrix} \xrightarrow[R_2 = R_2 - 2R_3]{R_1 = R_1 + 3R_3} \begin{bmatrix} 1 & 0 & 0 & | & \frac{1}{8} & -\frac{5}{8} & \frac{3}{4} \\ 0 & 1 & 0 & | & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} \\ 0 & 0 & 1 & | & \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} I & | & A^{-1} \end{bmatrix}$$

$$x = A^{-1}b = \begin{bmatrix} \frac{1}{8} & -\frac{5}{8} & \frac{3}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} \\ \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{bmatrix} \cdot \begin{bmatrix} -2 & -2 & -4 \end{bmatrix}^\top = \begin{bmatrix} \frac{1}{8} & -\frac{5}{8} & \frac{3}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} \\ \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{bmatrix} \cdot \begin{bmatrix} -2 \\ -2 \\ -4 \end{bmatrix} =$$

$$\begin{bmatrix} \frac{1}{8} \cdot -2 + -\frac{5}{8} \cdot -2 + \frac{3}{4} \cdot -4 \\ -\frac{1}{4} \cdot -2 + \frac{3}{4} \cdot -2 + -\frac{1}{2} \cdot -4 \\ \frac{3}{8} \cdot -2 + -\frac{3}{8} \cdot -2 + \frac{1}{4} \cdot -4 \end{bmatrix} = \boxed{\begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}}$$

A7. (6 points)

(a)

Assume $x, y \in \mathbb{R}^n$, we know:

$$x^\top \mathbf{A} x = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \mathbf{A}_{i,j} x_j \text{ and } y^\top \mathbf{B} x = \sum_{i=1}^{n} \sum_{j=1}^{n} y_i \mathbf{B}_{i,j} x_j$$

Therefore, the function is:

$$f(x, y) = x^\top \mathbf{A} x + y^\top \mathbf{B} x + c = \boxed{\sum_{i=1}^{n} \sum_{j=1}^{n} x_i \mathbf{A}_{i,j} x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} y_i \mathbf{B}_{i,j} x_j + c}$$

(b)

$$
\begin{aligned}
\frac{\partial}{\partial x_k} x^\top \mathbf{A} x &= \frac{\partial}{\partial x_k} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \mathbf{A}_{i,j} x_j && \text{(from (a.))} \\
&= \frac{\partial}{\partial x_k} [\sum_{i \neq k} \sum_{j \neq k} x_i \mathbf{A}_{i,j} x_j + \sum_{i \neq k} x_i \mathbf{A}_{i,k} x_k + \sum_{j \neq k} x_k \mathbf{A}_{k,j} x_j + \mathbf{A}_{k,k} x_k^2] \\
&= \sum_{i \neq k} x_i \mathbf{A}_{i,k} + \sum_{j \neq k} \mathbf{A}_{k,j} x_j + 2\mathbf{A}_{k,k} x_k \\
&= \sum_{i=1}^{n} \mathbf{A}_{i,k} x_i + \sum_{j=1}^{n} \mathbf{A}_{k,j} x_j
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial x_k} y^\top \mathbf{B} x &= \frac{\partial}{\partial x_k} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i \mathbf{B}_{i,j} x_j && \text{(from (a.))} \\
&= \frac{\partial}{\partial x_k} [\sum_{i \neq k} \sum_{j \neq k} y_i \mathbf{B}_{i,j} x_j + \sum_{i \neq k} y_i \mathbf{B}_{i,k} x_k + \sum_{j \neq k} y_k \mathbf{B}_{k,j} x_j + y_k \mathbf{B}_{k,k} x_k] \\
&= \sum_{i \neq k} y_i \mathbf{B}_{i,k} + y_k \mathbf{B}_{k,k} \\
&= \sum_{i=1}^{n} \mathbf{B}_{i,k} y_i \\
\frac{\partial}{\partial x_k} c &= 0
\end{aligned}
$$

Thus, we get:

$$
\frac{\partial f}{\partial x_k}(x, y) = \sum_{i=1}^{n} \mathbf{A}_{i,k} x_i + \sum_{j=1}^{n} \mathbf{A}_{k,j} x_j + \sum_{i=1}^{n} \mathbf{B}_{i,k} y_i
$$

Therefore we have:

$$
\begin{aligned}
\nabla_x f(x, y) &= \left[\tfrac{\partial f}{\partial x_1}(x, y) \quad \tfrac{\partial f}{\partial x_2}(x, y) \quad \cdots \quad \tfrac{\partial f}{\partial x_n}(x, y) \right]^\top \in \mathbb{R}^n \\
&= \boxed{\begin{bmatrix} \sum_{i=1}^{n} \mathbf{A}_{i,1} x_i + \sum_{j=1}^{n} \mathbf{A}_{1,j} x_j + \sum_{i=1}^{n} \mathbf{B}_{i,1} y_i \\ \sum_{i=1}^{n} \mathbf{A}_{i,2} x_i + \sum_{j=1}^{n} \mathbf{A}_{2,j} x_j + \sum_{i=1}^{n} \mathbf{B}_{i,2} y_i \\ \cdots \\ \sum_{i=1}^{n} \mathbf{A}_{i,n} x_i + \sum_{j=1}^{n} \mathbf{A}_{n,j} x_j + \sum_{i=1}^{n} \mathbf{B}_{i,n} y_i \end{bmatrix} \in \mathbb{R}^n} && \text{(summations over indices)} \\
&= \nabla_x (x^\top \mathbf{A} x + y^\top \mathbf{B} x + c) \\
&= \nabla_x (x^\top \mathbf{A} x) + \nabla_x (y^\top \mathbf{B} x) + \nabla_x (c) \\
&= (\mathbf{A} + \mathbf{A}^\top) x + \mathbf{B}^\top y + 0 \\
&= \boxed{(\mathbf{A} + \mathbf{A}^\top) x + \mathbf{B}^\top y} && \text{(vector notation)}
\end{aligned}
$$

(c)

$$\frac{\partial}{\partial y_k} x^\top \mathbf{A} x = \frac{\partial}{\partial y_k} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i \mathbf{A}_{i,j} x_j \qquad \text{(from (a.))}$$
$$= 0$$

$$\frac{\partial}{\partial y_k} y^\top \mathbf{B} x = \frac{\partial}{\partial y_k} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i \mathbf{B}_{i,j} x_j \qquad \text{(from (a.))}$$
$$= \frac{\partial}{\partial y_k} [\sum_{i \neq k} \sum_{j \neq k} y_i \mathbf{B}_{i,j} x_j + \sum_{i \neq k} y_i \mathbf{B}_{i,k} x_k + \sum_{j \neq k} y_k \mathbf{B}_{k,j} x_j + y_k \mathbf{B}_{k,k} x_k]$$
$$= \sum_{j \neq k} \mathbf{B}_{k,j} x_j + \mathbf{B}_{k,k} x_k$$
$$= \sum_{j=1}^{n} \mathbf{B}_{k,j} x_j$$

$$\frac{\partial}{\partial y_k} c = 0$$

Thus, we get:

$$\frac{\partial f}{\partial y_k}(x, y) = \sum_{j=1}^{n} \mathbf{B}_{k,j} x_j$$

Therefore, we have:

$$\nabla_x f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial y_1}(x, y) & \frac{\partial f}{\partial y_2}(x, y) & \cdots & \frac{\partial f}{\partial y_n}(x, y) \end{bmatrix}^\top \in \mathbb{R}^n$$
$$= \boxed{\begin{bmatrix} \sum_{j=1}^{n} \mathbf{B}_{1,j} x_j & \sum_{j=1}^{n} \mathbf{B}_{2,j} x_j & \cdots & \sum_{j=1}^{n} \mathbf{B}_{n,j} x_j \end{bmatrix}^\top \in \mathbb{R}^n} \qquad \text{(summations over indices)}$$
$$= \nabla_y (x^\top \mathbf{A} x + y^\top \mathbf{B} x + c)$$
$$= \nabla_y (x^\top \mathbf{A} x) + \nabla_y (y^\top \mathbf{B} x) + \nabla_y (c)$$
$$= 0 + \mathbf{B} x + 0$$
$$= \boxed{\mathbf{B} x} \qquad \text{(vector notation)}$$

A8. (8 points)

(a)

We know $\text{diag}(v) = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_n \end{bmatrix}$

$$\begin{bmatrix}\text{diag}(v) & | & I\end{bmatrix} = \begin{bmatrix} v_1 & 0 & \cdots & 0 & | & 1 & 0 & \cdots & 0 \\ 0 & v_2 & \cdots & 0 & | & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v_n & | & 0 & 0 & \cdots & 1 \end{bmatrix} \xrightarrow{R_i = \frac{1}{v_i} R_i}$$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & | & \frac{1}{v_1} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & | & 0 & \frac{1}{v_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & | & 0 & 0 & \cdots & \frac{1}{v_n} \end{bmatrix} = \begin{bmatrix} I & | & \text{diag}(v)^{-1}\end{bmatrix}$$

Since $\text{diag}(v)^{-1} = \text{diag}(w)$, we know $\text{diag}(w) = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} = \begin{bmatrix} \frac{1}{v_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{v_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{v_n} \end{bmatrix}$

Thus, we know $w_i = \frac{1}{v_i}$.
Since $g(v_i) = w_i$, we have:

$$\boxed{g(x) := \frac{1}{x}}$$

(b)

$$\begin{aligned} ||\mathbf{A}x||_2^2 &= \left(\sqrt{(Ax)^\top (Ax)}\right)^2 && \text{(Def. of } \ell_2 \text{ norm)} \\ &= \left(\sqrt{x^\top A^\top A x}\right)^2 \\ &= \left(\sqrt{x^\top I x}\right)^2 && \text{(Since } A^\top A = I\text{)} \\ &= \left(\sqrt{x^\top x}\right)^2 \\ &= ||x||_2^2 && \text{(Def. of } \ell_2 \text{ norm)} \end{aligned}$$

(c)

$$
\begin{aligned}
& \mathbf{B}^{-1}\mathbf{B} = I && \text{(Since } \mathbf{B} \text{ is invertible)} \\
\Rightarrow \quad & (\mathbf{B}^{-1}\mathbf{B})^\top = I^\top && \\
\Rightarrow \quad & \mathbf{B}^\top(\mathbf{B}^{-1})^\top = I && \text{(Since } (AB)^\top = B^\top A^\top\text{)} \\
\Rightarrow \quad & \mathbf{B}(\mathbf{B}^{-1})^\top = I && \text{(Since } \mathbf{B} = \mathbf{B}^\top\text{)} \\
\Rightarrow \quad & \mathbf{B}^{-1}\mathbf{B}(\mathbf{B}^{-1})^\top = \mathbf{B}^{-1}I && \\
\Rightarrow \quad & I(\mathbf{B}^{-1})^\top = \mathbf{B}^{-1}I && \text{(Since } \mathbf{B} \text{ is invertible)} \\
\Rightarrow \quad & (\mathbf{B}^{-1})^\top = \mathbf{B}^{-1} &&
\end{aligned}
$$

Therefore, $\mathbf{B}^{-1}$ is also symmetric.

(d)

Let λ be an eigenvalue of $\mathbf{C}$, and x be the corresponding eigenvector.
We have:

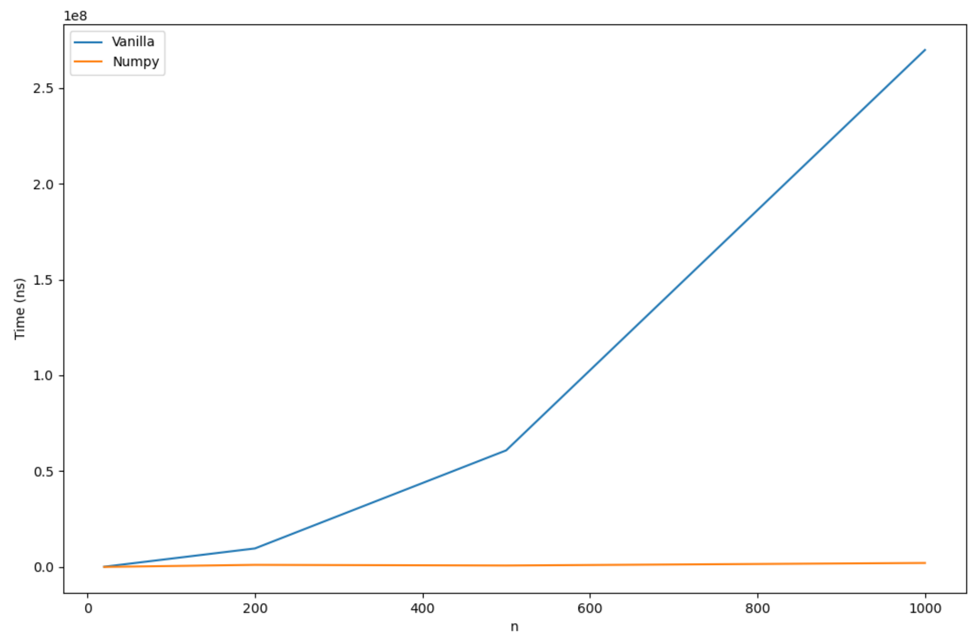
$$
\begin{aligned}
& \mathbf{C}x = \lambda x && \text{(Def. of eigenvalue)} \\
\Rightarrow \quad & x^\top \mathbf{C}x = x^\top \lambda x && \\
\Rightarrow \quad & x^\top \mathbf{C}x = \lambda x^\top x && \text{(Since } \lambda \text{ is a scalar)} \\
\Rightarrow \quad & x^\top \mathbf{C}x = \lambda ||x||_2^2 && \text{(Def. of } \ell_2 \text{ norm)}
\end{aligned}
$$

Since $\mathbf{C}$ is semi-positive definite, we know $x^\top \mathbf{C}x \geq 0$.
Since $x^\top \mathbf{C}x = \lambda||x||_2^2$, we know $\lambda||x||_2^2 \geq 0$.
Since $||x||_2^2 > 0$, we know $\lambda \geq 0$.
Therefore, we have shown the eigenvalues of $\mathbf{C}$ are non-negative.

A9. (3 points)

(c)

Wall-lock time difference plot:



Reason:

For vanilla python, we are manually looping through the list and vectors to do matrix operations on the python level, but NumPy uses optimized low level C precompiles to do the matrix operations. The optimized memory access pattern allowed the running time for NumPy solution to grow much slower than the vanilla python solution.

A10. (4 points)

(a)

n = 40000

(b)

How does the empirical CDF change with k?
As k increases, the empirical CDF becomes more and more similar (converges) to the true CDF.

Plot of $\widehat{F}_n(x) \in [-3, 3]$: