CSC4008 Assignment 1

CHEN Ang (118010009)

1.2

We have that $\boldsymbol{X} = [X_1, X_2, X_3]^* \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Denote

$$\mathbf{\Sigma} \equiv \begin{bmatrix} \mathbf{\Sigma}_{12} & \mathbf{\Sigma}_{12,3} \\ \mathbf{\Sigma}_{3,12} & \Sigma_3 \end{bmatrix}$$

where $\mathbf{\Sigma}_{12} \in \mathbb{R}^{2\times 2}$, $\mathbf{\Sigma}_{12,3} = \mathbf{\Sigma}^*_{3,12} \in \mathbb{R}^2$, and $\Sigma_3 \in \mathbb{R}$.

By block matrix inversion,

$$\mathbf{\Theta} = \begin{bmatrix} \mathbf{\Theta}_{12} & \mathbf{\Theta}_{12,3} \\ \mathbf{\Theta}_{3,12} & \Theta_3 \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{\Theta}_{12} &= (\mathbf{\Sigma}_{12} - \mathbf{\Sigma}_{12,3}\Sigma_3^{-1}\mathbf{\Sigma}^*_{12,3})^{-1} \\ \Theta_3 &= (\Sigma_3 - \mathbf{\Sigma}^*_{12,3}\mathbf{\Sigma}_{12}^{-1}\mathbf{\Sigma}_{12,3})^{-1} = \Sigma_3^{-1} + \Sigma_3^{-1}\mathbf{\Sigma}^*_{12,3}\mathbf{\Theta}_{12}\mathbf{\Sigma}_{12,3}\Sigma_3^{-1} \\ \mathbf{\Theta}_{12,3} &= \mathbf{\Theta}^*_{3,12} = -\mathbf{\Theta}_{12}\mathbf{\Sigma}_{12,3}\Sigma_3^{-1} \end{aligned} \tag{1}$$

Lemma 0. The marginal distribution of X_3 is $\mathcal{N}(0, \Sigma_3)$.

Proof. Consider the joint PDF

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{3}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{x}^*\mathbf{\Theta}\boldsymbol{x}} = \frac{1}{(2\pi)^{\frac{3}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}Q(\boldsymbol{x})}$$

where we define

$$\begin{aligned} Q(\boldsymbol{x}) &\equiv \boldsymbol{x}^*\mathbf{\Theta}\boldsymbol{x} \\ &= \boldsymbol{x}^*_{12}\mathbf{\Theta}_{12}\boldsymbol{x}_{12} + 2\boldsymbol{x}^*_{12}\mathbf{\Theta}_{12,3}x_3 + \Theta_3 x_3^2 \\ &= \boldsymbol{x}^*_{12}\mathbf{\Theta}_{12}\boldsymbol{x}_{12} \\ &\quad - 2\boldsymbol{x}^*_{12}\mathbf{\Theta}_{12}\mathbf{\Sigma}_{12,3}\Sigma_3^{-1}x_3 \\ &\quad + \left(\Sigma_3^{-1} + \Sigma_3^{-1}\mathbf{\Sigma}^*_{12,3}\mathbf{\Theta}_{12}\mathbf{\Sigma}_{12,3}\Sigma_3^{-1}\right)x_3^2 \\ &= \Sigma_3^{-1}x_3^2 \\ &\quad + \left(\boldsymbol{x}^*_{12}\mathbf{\Theta}_{12}\boldsymbol{x}_{12} - 2\boldsymbol{x}^*_{12}\mathbf{\Theta}_{12}\mathbf{\Sigma}_{12,3}\Sigma_3^{-1}x_3 + \Sigma_3^{-1}\mathbf{\Sigma}^*_{12,3}\mathbf{\Theta}_{12}\mathbf{\Sigma}_{12,3}\Sigma_3^{-1}x_3^2\right) \\ &= \Sigma_3^{-1}x_3^2 \\ &\quad + \left(\boldsymbol{x}_{12} - \mathbf{\Sigma}_{12,3}\Sigma_3^{-1}x_3\right)^*\mathbf{\Theta}_{12}\left(\boldsymbol{x}_{12} - \mathbf{\Sigma}_{12,3}\Sigma_3^{-1}x_3\right) \end{aligned}$$

Define

$$\begin{aligned} \boldsymbol{m} &\equiv \boldsymbol{m}(x_3) = \mathbf{\Sigma}_{12,3}\Sigma_3^{-1}x_3 \\ Q_{12}(\boldsymbol{x}) &\equiv (\boldsymbol{x}_{12} - \boldsymbol{m})^*\mathbf{\Theta}_{12}(\boldsymbol{x}_{12} - \boldsymbol{m}) \\ Q_3(x_3) &\equiv \Sigma_3^{-1}x_3^2 \end{aligned}$$

we have

$$\begin{aligned} Q(\boldsymbol{x}) &= (\boldsymbol{x}_{12} - \boldsymbol{m})^*\mathbf{\Theta}_{12}(\boldsymbol{x}_{12} - \boldsymbol{m}) + \Sigma_3^{-1}x_3^2 \\ &= Q_{12}(\boldsymbol{x}) + Q_3(x_3) \end{aligned}$$

Thus the joint PDF can rewritten as

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{3}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}Q_{12}(\boldsymbol{x})} e^{-\frac{1}{2}Q_3(x_3)}$$

Computing the determinant block-wise, one has

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{12,3}\Sigma_3^{-1}\boldsymbol{\Sigma}_{12,3}^*||\Sigma_3| = |\boldsymbol{\Theta}_{12}^{-1}|\Sigma_3$$

Hence

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) &= \frac{1}{(2\pi)^{\frac{2}{2}}|\boldsymbol{\Theta}_{12}^{-1}|^{\frac{1}{2}}} e^{-\frac{1}{2}Q_{12}(\boldsymbol{x})} \cdot \frac{1}{(2\pi)^{\frac{1}{2}}\Sigma_3^{\frac{1}{2}}} e^{-\frac{1}{2}Q_3(x_3)} \\ &= \mathcal{N}\left(\boldsymbol{x}_{12};\ \boldsymbol{m}, \boldsymbol{\Theta}_{12}^{-1}\right) \cdot \mathcal{N}\left(\boldsymbol{x}_3;\ 0, \Sigma_3\right) \end{aligned} \tag{2}$$

Therefore marginal distribution of X_3 has density

$$\begin{aligned} f_{X_3}(x_3) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\boldsymbol{X}}(\boldsymbol{x})\, dx_1 dx_2 \\ &= \mathcal{N}\left(\boldsymbol{x}_3;\ 0, \Sigma_3\right) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathcal{N}\left(\boldsymbol{x}_{12};\ \boldsymbol{m}, \boldsymbol{\Theta}_{12}^{-1}\right)\, dx_1 dx_2 \\ &= \mathcal{N}\left(\boldsymbol{x}_3;\ 0, \Sigma_3\right) \end{aligned}$$

which completes the proof.

Lemma 1. The conditional distribution of $[X_1, X_2]^*$ given $X_3 = x_3$ is normal with mean vector

$$\boldsymbol{\mu}_{12|3} = \boldsymbol{\Sigma}_{12,3}\Sigma_3^{-1}x_3$$

and covariance matrix

$$\begin{aligned} \boldsymbol{\Sigma}_{12|3} &= \boldsymbol{\Sigma}_{12} - \Sigma_3^{-1}\boldsymbol{\Sigma}_{12,3}\boldsymbol{\Sigma}_{12,3}^* \\ &= \begin{bmatrix} \Sigma_1 - \Sigma_{1,3}^2\Sigma_3^{-1} & \Sigma_{1,2} - \Sigma_{1,3}\Sigma_{2,3}\Sigma_3^{-1} \\ \Sigma_{2,1} - \Sigma_{2,3}\Sigma_{1,3}\Sigma_3^{-1} & \Sigma_2 - \Sigma_{2,3}^2\Sigma_3^{-1} \end{bmatrix} \end{aligned}$$

Proof. Using Lemma 0 and equation (2),

$$\begin{aligned} f_{\boldsymbol{X}_{12}|X_3=x_3}(\boldsymbol{x}_{12}) &= \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{X_3}(x_3)} \\ &= \frac{\mathcal{N}\left(\boldsymbol{x}_{12};\ \boldsymbol{m}, \boldsymbol{\Theta}_{12}^{-1}\right) \cdot \mathcal{N}\left(\boldsymbol{x}_3;\ 0, \Sigma_3\right)}{\mathcal{N}\left(\boldsymbol{x}_3;\ 0, \Sigma_3\right)} \\ &= \mathcal{N}\left(\boldsymbol{x}_{12};\ \boldsymbol{m}, \boldsymbol{\Theta}_{12}^{-1}\right) \end{aligned}$$

where $\boldsymbol{m} = \boldsymbol{\Sigma}_{12,3}\Sigma_3^{-1}x_3 = \boldsymbol{\mu}_{12|3}$ and $\boldsymbol{\Theta}_{12}^{-1} = \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{12,3}\Sigma_3^{-1}\boldsymbol{\Sigma}_{12,3}^* = \boldsymbol{\Sigma}_{12|3}$, the desired result.

Lemma 2. The conditional distribution of X_1 given $X_3 = x_3$ is normal with mean

$$\mu_{1|3} = (\boldsymbol{\mu}_{12|3})_1 = x_3\Sigma_{1,3}\Sigma_3{}^{-1}$$

and variance

$$\sigma_{1|3}^2 = (\boldsymbol{\Sigma}_{12|3})_{1,1} = \Sigma_1 - \Sigma_{1,3}^2\Sigma_3^{-1}$$

and similarly for X_2 given $X_3 = x_3$.

Proof. This result directly follows from Lemma 0 and 1.

Proof of the main theorem. For clarity, denote

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \equiv \begin{bmatrix} \Sigma_1 - \Sigma_{1,3}^2 \Sigma_3^{-1} & \Sigma_{1,2} - \Sigma_{1,3}\Sigma_{2,3}\Sigma_3^{-1} \\ \Sigma_{2,1} - \Sigma_{2,3}\Sigma_{1,3}\Sigma_3^{-1} & \Sigma_2 - \Sigma_{2,3}^2 \Sigma_3^{-1} \end{bmatrix} = \mathbf{\Sigma}_{12|3}$$

so that

$$\sigma_{1|3}^2 = a,\ \sigma_{2|3}^2 = c$$

" $\implies$ ":

Suppose X_1, X_2 are conditionally independent given X_3. Then

$$f_{X_1,X_2|X_3=x_3}(x_1,x_2) = f_{X_1|X_3=x_3}(x_1) \cdot f_{X_2|X_3=x_3}(x_2)$$

Using Lemma 1 and 2, we substitute the PDFs to get

$$\frac{1}{(2\pi)^{\frac{2}{2}} |\mathbf{\Sigma}_{12|3}|^{\frac{1}{2}}} e^{-\frac{1}{2}[x_1,x_2]\mathbf{\Sigma}_{12|3}^{-1}[x_1,x_2]^*} = \frac{1}{(2\pi)^{\frac{1+1}{2}} \sigma_{1|3}\sigma_{2|3}} e^{-\frac{1}{2}\left(x_1^2/\sigma_{1|3}^2 + x_2^2/\sigma_{2|3}^2\right)}$$

Equating coefficients of x_1, x_2 in the exponents, the matrix

$$\mathbf{\Sigma}_{12|3}^{-1} = \begin{bmatrix} a^{-1} - b^2(a^2\Delta)^{-1} & b(a\Delta)^{-1} \\ b(a\Delta)^{-1} & -\Delta^{-1} \end{bmatrix}, \quad \Delta = b^2/a - c$$

must be diagonal, and so

$$b = \Sigma_{1,2} - \Sigma_{1,3}\Sigma_{2,3}\Sigma_3^{-1} = 0 \iff \Sigma_{1,2}\Sigma_3 - \Sigma_{1,3}\Sigma_{2,3} = 0$$

Solving the inverse covariance matrix gives

$$\theta_{1,2} = \frac{\Sigma_{1,2}\Sigma_3 - \Sigma_{1,3}\Sigma_{2,3}}{\Sigma_1(\Sigma_{2,3}^2 - \Sigma_2\Sigma_3) + \Sigma_{1,2}^2\Sigma_3 - 2\Sigma_{1,2}\Sigma_{1,3}\Sigma_{2,3} + \Sigma_{1,3}^2\Sigma_2} = 0$$

which is to be shown.

" $\impliedby$ ":

The other direction is similar. Suppose now $\theta_{1,2} = 0$. We then have

$$\Sigma_{1,2}\Sigma_3 - \Sigma_{1,3}\Sigma_{2,3} = 0$$

Since $\Sigma_3 \neq 0$, we divide the equation through to get

$$\Sigma_{1,2} - \Sigma_{1,3}\Sigma_{2,3}\Sigma_3^{-1} = b = 0$$

Therefore

$$\mathbf{\Sigma}_{12|3} = \begin{bmatrix} a & 0 \\ 0 & c \end{bmatrix}, \ \mathbf{\Sigma}_{12|3}^{-1} = \begin{bmatrix} a^{-1} & 0 \\ 0 & c^{-1} \end{bmatrix}$$

We evaluate the PDF of $[X_1, X_2]^*$ condition on X_3 using Lemma 1:

$$\begin{aligned} f_{X_1,X_2|X_3=x_3}(x_1,x_2) &= \frac{1}{(2\pi)^{\frac{2}{2}} |\mathbf{\Sigma}_{12|3}|^{\frac{1}{2}}} e^{-\frac{1}{2}[x_1,x_2]\mathbf{\Sigma}_{12|3}^{-1}[x_1,x_2]^*} \\ &= \frac{1}{2\pi(ac)^{\frac{1}{2}}} e^{-\frac{1}{2}(x_1^2/a + x_2^2/c)} \end{aligned}$$

For the other side, we again use Lemma 2 to get

$$
\begin{aligned}
f_{X_1|X_3=x_3}(x_1)\cdot f_{X_2|X_3=x_3}(x_2) &= \frac{1}{(2\pi)^{\frac{1+1}{2}}\sigma_{1|3}\sigma_{2|3}}e^{-\frac{1}{2}\left(x_1^2/\sigma_{1|3}^2+x_2^2/\sigma_{2|3}^2\right)} \\
&= \frac{1}{2\pi a^{\frac{1}{2}}c^{\frac{1}{2}}}e^{-\frac{1}{2}\left(x_1^2/a+x_2^2/c\right)}
\end{aligned}
$$

giving the desired equality, and we are done.

1.3

We have by definition

$$
\mathbf{\Sigma\Theta} = \begin{bmatrix} \mathbf{\Sigma}_o & \mathbf{\Sigma}_{o,h} \\ \mathbf{\Sigma}_{o,h}^* & \mathbf{\Sigma}_h \end{bmatrix} \begin{bmatrix} \mathbf{\Theta}_o & \mathbf{\Theta}_{o,h} \\ \mathbf{\Theta}_{o,h}^* & \mathbf{\Theta}_h \end{bmatrix} = \boldsymbol{I}
$$

Considering the first n_o rows:

$$
\begin{aligned}
\mathbf{\Sigma}_o\mathbf{\Theta}_o + \mathbf{\Sigma}_{o,h}\mathbf{\Theta}_{o,h}^* &= \boldsymbol{I} \\
\mathbf{\Sigma}_o\mathbf{\Theta}_{o,h} + \mathbf{\Sigma}_{o,h}\mathbf{\Theta}_h &= \boldsymbol{O}
\end{aligned}
$$

The second equation gives us

$$
\mathbf{\Sigma}_{o,h} = -\mathbf{\Sigma}_o\mathbf{\Theta}_{o,h}\mathbf{\Theta}_h^{-1}
$$

Plugging back into the first equation,

$$
\mathbf{\Sigma}_o\left(\mathbf{\Theta}_o - \mathbf{\Theta}_{o,h}\mathbf{\Theta}_h^{-1}\right) = \boldsymbol{I} \iff \mathbf{\Sigma}_o^{-1} = \mathbf{\Theta}_o - \mathbf{\Theta}_{o,h}\mathbf{\Theta}_h^{-1}
$$

1.4

We have that

$$
\begin{aligned}
e(\boldsymbol{x}) &\equiv ||\boldsymbol{y} - \boldsymbol{Ax}||_2^2 \\
&= (\boldsymbol{y} - \boldsymbol{Ax})^*(\boldsymbol{y} - \boldsymbol{Ax}) \\
&= (\boldsymbol{y}^* - \boldsymbol{x}^*\boldsymbol{A}^*)(\boldsymbol{y} - \boldsymbol{Ax}) \\
&= \boldsymbol{y}^*\boldsymbol{y} - \boldsymbol{y}^*\boldsymbol{Ax} - \boldsymbol{x}^*\boldsymbol{A}^*\boldsymbol{y} + \boldsymbol{x}^*\boldsymbol{A}^*\boldsymbol{Ax} \\
&= \boldsymbol{y}^*\boldsymbol{y} - 2\boldsymbol{y}^*\boldsymbol{Ax} + \boldsymbol{x}^*\boldsymbol{A}^*\boldsymbol{Ax}
\end{aligned}
$$

First Order Necessary Condition for local minima:

$$
\nabla e(\boldsymbol{x}) = 2(\boldsymbol{A}^*\boldsymbol{Ax} - \boldsymbol{A}^*\boldsymbol{y}) = \mathbf{0} \iff \boldsymbol{A}^*\boldsymbol{Ax} = \boldsymbol{A}^*\boldsymbol{y}
$$

Assuming that $\boldsymbol{A}^*\boldsymbol{A}$ is invertible (equivalently, $\texttt{rank}(\boldsymbol{A}) = n$), FONC yields

$$
\boldsymbol{x}_\star = (\boldsymbol{A}^*\boldsymbol{A})^{-1}\boldsymbol{A}^*\boldsymbol{y}
$$

Since the Hessian

$$
\nabla^2 e(\boldsymbol{x}) = 2\boldsymbol{A}^*\boldsymbol{A}
$$

is positive semidefinite for all $\boldsymbol{x}$, e is convex. Further, because $\boldsymbol{x}_\star$ is the unique critical point, it must also be the unique global minimum of e.

1.8

1

Consider

$$
\begin{aligned}
e(\boldsymbol{x}) &\equiv ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2 + \lambda||\boldsymbol{x}||_2^2 \\
&= (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^*(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + \lambda\boldsymbol{x}^*\boldsymbol{x} \\
&= \boldsymbol{y}^*\boldsymbol{y} - 2\boldsymbol{y}^*\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^*\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{x} + \lambda\boldsymbol{x}^*\boldsymbol{x}
\end{aligned}
$$

FONC:

$$
\nabla e(\boldsymbol{x}) = 2(\lambda\boldsymbol{x} + \boldsymbol{A}^*\boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}^*\boldsymbol{y}) = \boldsymbol{0} \iff (\boldsymbol{A}^*\boldsymbol{A} + \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{A}^*\boldsymbol{y}
$$

Given that $\boldsymbol{A}^*\boldsymbol{A} + \lambda\boldsymbol{I}$ is invertible, we obtain the critical point

$$
\boldsymbol{x}_\star = (\boldsymbol{A}^*\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^*\boldsymbol{y}
$$

Since the Hessian

$$
\nabla^2 e(\boldsymbol{x}) = 2(\boldsymbol{A}^*\boldsymbol{A} + \lambda\boldsymbol{I})
$$

is positive semidefinite for all $\boldsymbol{x}$, e is convex. Further, because $\boldsymbol{x}_\star$ is the unique critical point, it must also be the unique global minimum of e.

2

Since $\boldsymbol{A}^*\boldsymbol{A}$ is real-symmetric, it is diagonalizable. Let $\lambda_1, \cdots \lambda_n$ be the eigenvalues of $\boldsymbol{A}^*\boldsymbol{A}$, with multiplicity. The matrix is invertible if and only if the determinant

$$
|\boldsymbol{A}^*\boldsymbol{A} + \lambda\boldsymbol{I}| = \prod_{i=1}^{n}(\lambda_i + \lambda) \neq 0
$$

which holds precisely when

$$
\lambda_i \neq -\lambda, \quad \forall i
$$

If we are given $\lambda > 0$, then since $\boldsymbol{A}^*\boldsymbol{A}$ is positive semidefinite, we have

$$
\lambda_i \geq 0 > -\lambda, \quad \forall i
$$

The matrix $\boldsymbol{A}^*\boldsymbol{A} + \lambda\boldsymbol{I}$ is always invertible.