

# Homework 1

Tong Thi Van Anh - 11200365

August 2022

## 1 Exercise 1

### (a) Gaussian distribution is normalized

The Gaussian distribution is:

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

With  $\sigma > 0$  and  $x, \mu \in \mathbb{R}$ , we have  $\mathcal{N}(x|\mu, \sigma^2) > 0$ .

To prove that Gaussian distribution is normalized, we prove  $\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) dx = 1$ , equivalent to:

$$\int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) dx = 1$$

Or,

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right) dx = 1$$

Let  $u = \frac{x - \mu}{\sigma}$ , then  $du = \frac{1}{\sigma} dx$ .

$$\Rightarrow LHS = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du$$

Squaring both sides of above equation:

$$\begin{aligned} LHS^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}v^2\right) dv \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{u^2 + v^2}{2}\right) dudv \end{aligned}$$

We take the transformations of variables:  $u = r\cos(\theta)$  and  $v = r\sin(\theta)$ .

Then,  $u^2 + v^2 = r^2$  and  $dudv = \left| \frac{\partial(u,v)}{\partial(r,\theta)} \right| drd\theta$ . Applying Jacobian, we get:

$$\begin{aligned} \frac{\partial(u,v)}{\partial(r,\theta)} &= \begin{vmatrix} \frac{\partial u}{\partial r} & \frac{\partial u}{\partial \theta} \\ \frac{\partial v}{\partial r} & \frac{\partial v}{\partial \theta} \end{vmatrix} \\ &= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} \\ &= r\cos^2\theta + r\sin^2\theta \\ &= r \end{aligned}$$

So,  $dudv = r dr d\theta$  and:

$$\begin{aligned} LHS^2 &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty \exp\left(-\frac{r^2}{2}\right) r dr d\theta \\ &= -\frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty \exp\left(-\frac{r^2}{2}\right) d\left(-\frac{r^2}{2}\right) d\theta \\ &= -\frac{1}{2\pi} \int_0^{2\pi} \left[ \exp\left(-\frac{r^2}{2}\right) \right]_{r=0}^{r=\infty} d\theta \\ &= -\frac{1}{2\pi} \int_0^{2\pi} (0 - 1) d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\theta \\ &= \frac{1}{2\pi} (2\pi - 0) \\ &= 1. \end{aligned}$$

Because  $\mathcal{N}(x|\mu, \sigma^2) > 0$ , LHS = 1 = RHS. Therefore, Gaussian distribution is normalized.

### (b) Expectation of Gaussian distribution is $\mu$ (mean)

From the definition of expected value for continuous random value:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

We have:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx \end{aligned}$$

Substituting  $t = \frac{x - \mu}{\sigma}$ , we get:

$$x = t\sigma + \mu$$

$$dx = \sigma dt$$

Then,

$$\begin{aligned} E(X) &= \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + \mu) \exp\left(-\frac{t^2}{2}\right) dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma t \cdot \exp\left(-\frac{t^2}{2}\right) dt + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu \cdot \exp\left(-\frac{t^2}{2}\right) dt \\ &= \frac{-\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) d\left(\frac{-t^2}{2}\right) + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt \end{aligned}$$

We have:

$$\begin{aligned} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) d\left(\frac{-t^2}{2}\right) &= \left[\exp\left(-\frac{t^2}{2}\right)\right]_{-\infty}^{\infty} \\ &= 0 - 0 = 0 \end{aligned}$$

From part (a),

$$\begin{aligned} LHS &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du = 1 \\ \Rightarrow \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2}\right) du &= \sqrt{2\pi} \end{aligned}$$

Apply with variable  $t$ , we get

$$\int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt = \sqrt{2\pi}$$

Replacing with  $E(X)$ :

$$E(X) = 0 + \frac{\mu}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = \mu$$

### (c) Variance of Gaussian distribution is $\sigma^2$ (variance)

We have  $V(X) = E(X^2) - [E(X)]^2$ . In part (b), we proved that  $E(X) = \mu$ .

Now we calculate  $E(X^2)$ . From definition of expected value stated in part (b), we get:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \end{aligned}$$

We also substitute  $t = \frac{x-\mu}{\sigma}$  and implement as part (b):

$$\begin{aligned} E(X^2) &= \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + \mu)^2 \exp\left(-\frac{t^2}{2}\right) dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t)^2 \exp\left(-\frac{t^2}{2}\right) dt + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 2\sigma\mu t \cdot \exp\left(-\frac{t^2}{2}\right) dt + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu)^2 \cdot \exp\left(-\frac{t^2}{2}\right) dt \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \exp\left(-\frac{t^2}{2}\right) dt + \frac{2\sigma\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \cdot \exp\left(-\frac{t^2}{2}\right) dt + \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) dt \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \exp\left(-\frac{t^2}{2}\right) dt + \frac{2\sigma\mu}{\sqrt{2\pi}} \cdot 0 + \frac{\mu^2}{\sqrt{2\pi}} \cdot \sqrt{2\pi} \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \exp\left(-\frac{t^2}{2}\right) dt + \mu^2 \end{aligned}$$

Using integration by parts with  $u = t$ ,  $du = dt$  and  $v = -\exp(-\frac{t^2}{2})$ ,  $dv = t \cdot \exp(-t^2/2)dt$ :

$$\begin{aligned} I &= \int_{-\infty}^{\infty} t^2 \exp\left(-\frac{t^2}{2}\right) dt \\ &= \int_{-\infty}^{\infty} t \cdot t \cdot \exp\left(-\frac{t^2}{2}\right) dt \\ &= \left[ -t \cdot \exp\left(-\frac{t^2}{2}\right) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -\exp\left(-\frac{t^2}{2}\right) dt \\ &= 0 - (-\sqrt{2\pi}) = \sqrt{2\pi} \end{aligned}$$

Replacing  $I$  with  $E(X^2)$ :

$$E(X^2) = \frac{\sigma^2}{\sqrt{2\pi}} \cdot \sqrt{2\pi} + \mu^2 = \sigma^2 + \mu^2$$

Therefore, we get:

$$V(X) = E(X^2) - [E(X)]^2 = \sigma^2 + \mu^2 - (\mu)^2 = \sigma^2$$

#### (d) Multivariate Gaussian distribution is normalized

The multivariate Gaussian distribution takes the form:

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

where  $\boldsymbol{\mu}$  is a  $D$ -dimensional mean vector,  $\boldsymbol{\Sigma}$  is a  $D \times D$  covariance matrix, and  $|\boldsymbol{\Sigma}|$  denotes the determinant of  $\boldsymbol{\Sigma}$ . We have a quadratic form of Gaussian distribution:

$$\begin{aligned} \Delta^2 &= (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ &= \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ &= \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - 2\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ &= \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - 2\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \text{const} \end{aligned}$$

Consider the eigenvector equation for the covariance matrix,  $i = 1, \dots, D$ :

$$\boldsymbol{\Sigma} \mathbf{u}_i = \lambda_i \mathbf{u}_i$$

Because  $\boldsymbol{\Sigma}$  is a real, symmetric matrix, its eigenvalues will be real and its eigenvectors form an orthonormal set.

$$\begin{aligned} \boldsymbol{\Sigma} &= \sum_{i=1}^D \lambda_i \mathbf{u}_i \mathbf{u}_i^T \\ \Rightarrow \boldsymbol{\Sigma}^{-1} &= \sum_{i=1}^D \frac{1}{\lambda_i} \mathbf{u}_i \mathbf{u}_i^T \end{aligned}$$

Replacing  $\boldsymbol{\Sigma}^{-1}$  with  $\Delta^2$ , we get:

$$\begin{aligned} \Delta^2 &= (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \\ &= \sum_{i=1}^D \frac{1}{\lambda_i} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{u}_i \mathbf{u}_i^T (\mathbf{x} - \boldsymbol{\mu}) \end{aligned}$$

Setting  $y_i = \mathbf{u}_i^T (\mathbf{x} - \boldsymbol{\mu})$ , or vector  $\mathbf{y} = \mathbf{U}^T (\mathbf{x} - \boldsymbol{\mu})$ :

$$\Delta^2 = \sum_{i=1}^D \frac{y_i^2}{\lambda_i}$$

In going from the  $\mathbf{x}$  to the  $\mathbf{y}$  coordinate system, we have a Jacobian matrix  $\mathbf{J}$  with elements given by

$$J_{ij} = \frac{\partial x_i}{\partial y_j} = U_{ji}$$

where  $U_{ji}$  are the elements of the matrix  $\mathbf{U}^T$ . We have:

$$\begin{aligned} |\mathbf{J}|^2 &= |\mathbf{U}^T|^2 = |\mathbf{U}^T| |\mathbf{U}| = |\mathbf{U}^T \mathbf{U}| = |\mathbf{I}| = 1 \\ \Rightarrow |\mathbf{J}| &= 1 \end{aligned}$$

As the determinant of an  $n \times n$  matrix is equal to the product of its eigenvalues  $\Rightarrow |\Sigma| = \prod_{j=1}^D \lambda_j$ .  
 $\Rightarrow |\Sigma|^{1/2} = \prod_{j=1}^D \lambda_j^{1/2}$   
 As a result, we get:

$$\begin{aligned} p(\mathbf{y}) = p(\mathbf{x})|\mathbf{J}| &= \frac{1}{(2\pi)^{D/2}} \frac{1}{\prod_{j=1}^D \lambda_j^{1/2}} \exp\left(\sum_{i=1}^D \frac{-y_i^2}{2\lambda_i}\right) \\ &= \prod_{i=1}^D \frac{1}{(2\pi\lambda_i)^{1/2}} \exp\left(\frac{-y_i^2}{2\lambda_i}\right) \end{aligned}$$

Integrating both sides of equation:

$$\int_{-\infty}^{\infty} p(\mathbf{y}) d\mathbf{y} = \prod_{i=1}^D \int_{-\infty}^{\infty} \frac{1}{(2\pi\lambda_i)^{1/2}} \exp\left(\frac{-y_i^2}{2\lambda_i}\right) dy_i$$

Applying the result from part (a),

$$\int_{-\infty}^{\infty} p(\mathbf{y}) d\mathbf{y} = 1$$

Therefore, the multivariate Gaussian distribution is normalized.

## 2 Exercise 2

### (a) The conditional of Gaussian distribution

Let  $\mathbf{x}$  be a D-dimensional vector,  $\mathbf{x} \sim \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \Sigma)$ :

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$$

where  $x_1$  and  $x_2$  are 2 disjoint subsets of  $\mathbf{x}$ . We also define corresponding partitions of the mean vector  $\boldsymbol{\mu}$  given by:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$$

and of the covariance matrix  $\Sigma$  given by:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Let  $A$  be the inverse of the covariance matrix:

$$A = \Sigma^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

$\Sigma$  is symmetric so  $\Sigma_{11}$  and  $\Sigma_{22}$  are symmetric while  $\Sigma_{12} = \Sigma_{21}^T$ .  
 We have:

$$\begin{aligned} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) &= (\mathbf{x} - \boldsymbol{\mu})^T A (\mathbf{x} - \boldsymbol{\mu}) \\ &= (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T A_{11} (\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_1 - \boldsymbol{\mu}_1)^T A_{12} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &\quad + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T A_{21} (\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^T A_{22} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &= \mathbf{x}_1^T A_{11} \mathbf{x}_1 - 2\mathbf{x}_1^T A_{11} \boldsymbol{\mu}_1 + \mathbf{x}_1^T A_{12} \mathbf{x}_2 - \mathbf{x}_1^T A_{12} \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1^T A_{12} \mathbf{x}_2 \\ &\quad + \mathbf{x}_2^T A_{21} \mathbf{x}_1 - \mathbf{x}_2^T A_{21} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^T A_{21} \mathbf{x}_1 + \mathbf{x}_2^T A_{22} \mathbf{x}_2 - 2\mathbf{x}_2^T A_{22} \boldsymbol{\mu}_2 + \text{const} \end{aligned}$$

We are looking for conditional distribution  $p(\mathbf{x}_1|\mathbf{x}_2)$ , so we only consider the quadratic form of  $\mathbf{x}_1$ :

$$\begin{aligned}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) &= \mathbf{x}_1^T A_{11} \mathbf{x}_1 - 2\mathbf{x}_1^T A_{11} \boldsymbol{\mu}_1 + 2\mathbf{x}_1^T A_{12} \mathbf{x}_2 - 2\mathbf{x}_1^T A_{12} \boldsymbol{\mu}_2 + \text{const} \\ &= \mathbf{x}_1^T A_{11} \mathbf{x}_1 - 2\mathbf{x}_1^T [A_{11} \boldsymbol{\mu}_1 - A_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)] + \text{const}\end{aligned}$$

Hence conditional distribution  $p(\mathbf{x}_1|\mathbf{x}_2)$  will be Gaussian, because this distribution is characterized by its mean and its variance. Compare with Gaussian distribution from Exercise 1:

$$\Delta^2 = \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - 2\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \text{const}$$

We can infer that  $\boldsymbol{\Sigma}_{1|2}^{-1} = A_{11}$  or  $\boldsymbol{\Sigma}_{1|2} = A_{11}^{-1}$ , and:

$$\boldsymbol{\Sigma}_{1|2}^{-1} \boldsymbol{\mu}_{1|2} = A_{11} \boldsymbol{\mu}_1 - A_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

$$\begin{aligned}\Rightarrow \boldsymbol{\mu}_{1|2} &= \boldsymbol{\Sigma}_{1|2} [A_{11} \boldsymbol{\mu}_1 - A_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)] \\ &= A_{11}^{-1} [A_{11} \boldsymbol{\mu}_1 - A_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)] \\ &= \boldsymbol{\mu}_1 - A_{11}^{-1} A_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\end{aligned}$$

Using Schur complement with  $M = (A - BD^{-1}C)^{-1}$ :

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{pmatrix}$$

we have:

$$A_{11} = (\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21})^{-1}$$

and

$$A_{12} = -(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21})^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1}$$

Then we get:

$$\begin{aligned}\boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ \boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}\end{aligned}$$

As a result,

$$p(\mathbf{x}_1|\mathbf{x}_2) = \mathcal{N}(\mathbf{x}_{1|2}|\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$$

## (b) The marginal of Gaussian distribution

The marginal distribution is given by:

$$p(\mathbf{x}_1) = \int p(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2$$

We have the quadratic form for the joint distribution from part (a). Because our goal is to integrate out  $\mathbf{x}_2$ , this is most easily achieved by first considering the terms involving  $\mathbf{x}_2$  and then completing the square in order to facilitate integration. Picking out just those terms that involve  $\mathbf{x}_2$ , we have:

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \mathbf{x}_2^T A_{22} \mathbf{x}_2 - 2\mathbf{x}_2^T [A_{22} \boldsymbol{\mu}_2 - A_{21}(\mathbf{x}_1 - \boldsymbol{\mu}_1)] + \text{const}$$

Setting  $\mathbf{m} = A_{22} \boldsymbol{\mu}_2 - A_{21}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ , we have:

$$\begin{aligned}\mathbf{x}_2^T A_{22} \mathbf{x}_2 - 2\mathbf{x}_2^T [A_{22} \boldsymbol{\mu}_2 - A_{21}(\mathbf{x}_1 - \boldsymbol{\mu}_1)] &= \mathbf{x}_2^T A_{22} \mathbf{x}_2 - 2\mathbf{x}_2^T \mathbf{m} \\ &= (\mathbf{x}_2 - A_{22}^{-1} \mathbf{m})^T A_{22} (\mathbf{x}_2 - A_{22}^{-1} \mathbf{m}) - \mathbf{m}^T A_{22}^{-1} \mathbf{m}\end{aligned}$$

The second term in above equation does not depend on  $\mathbf{x}_2$ . Thus, when we take the exponential of this quadratic form, we see that the integration over  $\mathbf{x}_2$  will take the form:

$$\int \exp\left(-\frac{1}{2}(\mathbf{x}_2 - A_{22}^{-1}\mathbf{m})^T A_{22}(\mathbf{x}_2 - A_{22}^{-1}\mathbf{m})\right) d\mathbf{x}_2 = (2\pi)^{n/2} |A_{22}^{-1}|^{1/2}$$

$$\left(\text{Applying } \int \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T U^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) d\mathbf{x} = (2\pi)^{n/2} |U|^{1/2}\right).$$

Combining the second term with the remaining terms from the expansion of the quadratic form  $\Delta^2$  that depend on  $\mathbf{x}_1$ , we obtain:

$$\begin{aligned} & -\mathbf{m}^T A_{22}^{-1} \mathbf{m} + \mathbf{x}_1^T A_{11} \mathbf{x}_1 - 2\mathbf{x}_1^T [A_{11} \boldsymbol{\mu}_1 + A_{12} \boldsymbol{\mu}_2] + \text{const} \\ & = -(A_{22} \boldsymbol{\mu}_2 - A_{21}(\mathbf{x}_1 - \boldsymbol{\mu}_1))^T A_{22}^{-1} (A_{22} \boldsymbol{\mu}_2 - A_{21}(\mathbf{x}_1 - \boldsymbol{\mu}_1)) \\ & \quad + \mathbf{x}_1^T A_{11} \mathbf{x}_1 - 2\mathbf{x}_1^T [A_{11} \boldsymbol{\mu}_1 + A_{12} \boldsymbol{\mu}_2] + \text{const} \\ & = \mathbf{x}_1^T (A_{11} - A_{12} A_{22}^{-1} A_{21}) \mathbf{x}_1 - 2\mathbf{x}_1^T (A_{11} - A_{12} A_{22}^{-1} A_{21}) \boldsymbol{\mu}_1 + \text{const} \end{aligned}$$

Comparing to Gaussian distribution, we get:

$$\boldsymbol{\Sigma}_1 = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1}$$

Using Schur complement again, we obtain:  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_{11}$ .

So, the mean and covariance of the marginal distribution of  $p(\mathbf{x}_1)$  are:

$$\mathbb{E}[\mathbf{x}_1] = \boldsymbol{\mu}_1$$

$$\text{cov}[\mathbf{x}_1] = \boldsymbol{\Sigma}_{11}$$

Thus, we have:

$$p(\mathbf{x}_1) = \mathcal{N}(\mathbf{x}_1 | \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$$