

# Derivation of Multivariate Normal marginal and conditional distributions

Mazin Abdelghany, MD, MS

3 December 2025

## 1 Background

Gaussian process regression relies on several special properties of multivariate normal (MVN) distributions. The two most important of those properties are that:

1. The conditional distribution of a multivariate normal is normal
2. The marginal distribution of a multivariate normal is normal

Moreover, these two above items can be calculated in closed form. Suppose that we have an  $n$ -dimensional Gaussian<sup>1</sup> random vector<sup>2</sup>  $\mathbf{x} \in \mathbb{R}^n$ . That is,

$$\mathbf{x} = \begin{bmatrix} x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ x_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \\ \vdots \\ x_n \sim \mathcal{N}(\mu_n, \sigma_n^2) \end{bmatrix}$$

Using vector-matrix notation, we can rewrite the distribution statement as:

$$\mathbf{x} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with the subscript  $n$  indicating that vector  $\mathbf{x}$  is  $n$ -dimensional.

In this case,  $\boldsymbol{\mu}$  has become a mean **vector** of (possibly different) means for each of our random variables and  $\boldsymbol{\Sigma}$  has become a covariance **matrix**:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

---

<sup>1</sup>Gaussian and normal will be interchanged throughout this document and used as synonyms.

<sup>2</sup>Throughout this document, vectors will be denoted using lowercased bold letters while matrices will be denoted as uppercase bold letters.

We can make a few observations about this formulation:

1.  $\boldsymbol{\mu} \in \mathbb{R}^n$  and  $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$  (i.e., the mean is  $n$ -dimensional and the covariance matrix is  $n \times n$ -dimensional).
2.  $\boldsymbol{\mu}$  is the same number of dimensions as  $\mathbf{x}$ .
3.  $\boldsymbol{\mu}$  can assign a different  $\mu_i$  for each  $x_i \in \mathbf{x}$ .
4.  $\boldsymbol{\Sigma}$  is symmetric and all its entries are positive or zero (more strictly, it must be positive semidefinite).
5. The diagonal of  $\boldsymbol{\Sigma}$  is the covariance between  $x_i$  and itself, which is the same as the variance of  $x_i$  (i.e.,  $\sigma_{11} = \sigma_1^2$ ).

The pdf of a multivariate Gaussian is:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

Now, let us partition our multivariate normal random vector  $\mathbf{x} \in \mathbb{R}^n$  into two multivariate normal random vectors  $\mathbf{x}_a \in \mathbb{R}^d$  and  $\mathbf{x}_b \in \mathbb{R}^{(n-d)}$ .<sup>3</sup>

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix}$$

This new partitioned vector should have a multivariate normal distribution denoted by

$$\begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix} = \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix} \right) \quad (1)$$

With this background in mind, we can proceed to show that

1. The conditional mean vector and covariance matrix of a multivariate normal for the vector  $\mathbf{x}_a$  given a realization of the vector  $\mathbf{x}_b$  are

- $\boldsymbol{\mu}_{\mathbf{x}_a | \mathbf{x}_b} = \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b)$
- $\boldsymbol{\Sigma}_{\mathbf{x}_a | \mathbf{x}_b} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba}$

2. The marginal mean vector and covariance matrix of a multivariate normal for  $\mathbf{x}_a$  are

- $\boldsymbol{\mu}_{\mathbf{x}_a} = \boldsymbol{\mu}_a$
- $\boldsymbol{\Sigma}_{\mathbf{x}_a} = \boldsymbol{\Sigma}_{aa}$

---

<sup>3</sup>Note the dimensions of each vector!

## 2 Marginal of multivariate normal

Let us start with proof of the marginal mean vector and covariance matrix. The marginal pdf of  $\mathbf{x}_a$  is:

$$f(\mathbf{x}_a) = \int_{\mathbb{R}^{(n-d)}} f(\mathbf{x}_a, \mathbf{x}_b) d\mathbf{x}_b$$

because we are integrating out  $\mathbf{x}_b$  leaving only the pdf of  $\mathbf{x}_a$ . Plugging in the values of the mean vectors and covariance matrices from equation (11), we can write

$$\begin{aligned} f(\mathbf{x}_a) &= \int_{\mathbb{R}^{(n-d)}} \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp \left\{ -\frac{1}{2} \left( \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix} - \begin{bmatrix} \mathbf{u}_a \\ \mathbf{u}_b \end{bmatrix} \right)^\top \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}^{-1} \left( \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix} - \begin{bmatrix} \mathbf{u}_a \\ \mathbf{u}_b \end{bmatrix} \right) \right\} d\mathbf{x}_b \\ &= \int_{\mathbb{R}^{(n-d)}} \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp \left\{ -\frac{1}{2} \left( \begin{bmatrix} \mathbf{x}_a - \mathbf{u}_a \\ \mathbf{x}_b - \mathbf{u}_b \end{bmatrix} \right)^\top \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}^{-1} \left( \begin{bmatrix} \mathbf{x}_a - \mathbf{u}_a \\ \mathbf{x}_b - \mathbf{u}_b \end{bmatrix} \right) \right\} d\mathbf{x}_b \end{aligned}$$

First, to simplify notation, let us define the normalizing constant as

$$Q = (2\pi)^{n/2} \det(\Sigma)^{1/2}$$

Now, we need to introduce the concept of the precision matrix  $\Lambda$ . The precision matrix is defined as the inverse of the covariance matrix:  $\Lambda = \Sigma^{-1}$ . Using this new definition, we note that

$$\Lambda = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix} = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}^{-1} = \Sigma^{-1}$$

Now we can rewrite the above as:

$$f(\mathbf{x}_a) = \int_{\mathbb{R}^{(n-d)}} \frac{1}{Q} \exp \left\{ -\frac{1}{2} \begin{bmatrix} \mathbf{x}_a - \mathbf{u}_a \\ \mathbf{x}_b - \mathbf{u}_b \end{bmatrix}^\top \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix} \begin{bmatrix} \mathbf{x}_a - \mathbf{u}_a \\ \mathbf{x}_b - \mathbf{u}_b \end{bmatrix} \right\} d\mathbf{x}_b \quad (2)$$

It is important to note that  $\Lambda_{aa} \neq \Sigma_{aa}^{-1}$  but rather that  $\Lambda = \Sigma^{-1}$ !

Let us expand the positive of the expression within the exponent  $\exp\{\cdot\}$  in two steps:

$$\begin{aligned} &\begin{bmatrix} \mathbf{x}_a - \mathbf{u}_a \\ \mathbf{x}_b - \mathbf{u}_b \end{bmatrix}^\top \begin{bmatrix} \Lambda_{aa}(\mathbf{x}_a - \mathbf{u}_a) + \Lambda_{ab}(\mathbf{x}_b - \mathbf{u}_b) \\ \Lambda_{ba}(\mathbf{x}_a - \mathbf{u}_a) + \Lambda_{bb}(\mathbf{x}_b - \mathbf{u}_b) \end{bmatrix} \\ &\frac{1}{2}(\mathbf{x}_a - \mathbf{u}_a)^\top \Lambda_{aa}(\mathbf{x}_a - \mathbf{u}_a) + \frac{1}{2}(\mathbf{x}_a - \mathbf{u}_a)^\top \Lambda_{ab}(\mathbf{x}_b - \mathbf{u}_b) + \\ &\quad + \frac{1}{2}(\mathbf{x}_b - \mathbf{u}_b)^\top \Lambda_{ba}(\mathbf{x}_a - \mathbf{u}_a) + \frac{1}{2}(\mathbf{x}_b - \mathbf{u}_b)^\top \Lambda_{bb}(\mathbf{x}_b - \mathbf{u}_b) \quad (3) \end{aligned}$$

At this point we need to take advantage of the concept of completing the square.

**Theorem 1.** Let  $\mathbf{A} \in \mathbb{R}^{n \times n}$  be a symmetric, positive definite matrix and  $\{\mathbf{z}, \mathbf{b}, \mathbf{c}\} \in \mathbb{R}^n$ , then

$$\frac{1}{2} \mathbf{z}^\top \mathbf{A} \mathbf{z} + \mathbf{b}^\top \mathbf{z} + \mathbf{c} = \frac{1}{2} (\mathbf{z} + \mathbf{A}^{-1} \mathbf{b})^\top \mathbf{A} (\mathbf{z} + \mathbf{A}^{-1} \mathbf{b}) + \mathbf{c} - \frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}$$

**Lemma 2.** Note that  $\mathbf{b}^\top \mathbf{z}$  is a scalar ( $\mathbb{R}^{1 \times n} \cdot \mathbb{R}^{n \times 1} = 1 \times 1$ ). Note also that

$$\begin{aligned} \mathbf{b}^\top \mathbf{z} &= \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_n \end{bmatrix} & \mathbf{z}^\top \mathbf{b} &= \begin{bmatrix} z_1 & z_2 & \dots & z_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix} \\ &= b_1 z_1 + b_2 z_2 + \dots + b_n z_n & &= z_1 b_1 + z_2 b_2 + \dots + z_n b_n \end{aligned}$$

By the commutative property of multiplication of scalars,  $\mathbf{b}^\top \mathbf{z} = \mathbf{z}^\top \mathbf{b}$ .

*Proof.* Thus,

$$\frac{1}{2} \mathbf{z}^\top \mathbf{A} \mathbf{z} + \mathbf{b}^\top \mathbf{z} + \mathbf{c} = \frac{1}{2} \mathbf{z}^\top \mathbf{A} \mathbf{z} + \underbrace{\frac{1}{2} \mathbf{b}^\top \mathbf{z} + \frac{1}{2} \mathbf{z}^\top \mathbf{b}}_{\text{from lemma}} + \mathbf{c}$$

Then, add and subtract  $\frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}$

$$= \frac{1}{2} \mathbf{z}^\top \mathbf{A} \mathbf{z} + \frac{1}{2} \mathbf{b}^\top \mathbf{z} + \frac{1}{2} \mathbf{z}^\top \mathbf{b} + \mathbf{c} + \left( \frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b} - \frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b} \right)$$

Rearrange terms.

$$= \frac{1}{2} \mathbf{b}^\top \mathbf{z} + \frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b} + \frac{1}{2} \mathbf{z}^\top \mathbf{A} \mathbf{z} + \frac{1}{2} \mathbf{z}^\top \mathbf{b} + \mathbf{c} - \frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}$$

Pull out the  $\frac{1}{2}$ .

$$= \frac{1}{2} (\mathbf{b}^\top \mathbf{z} + \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b} + \mathbf{z}^\top \mathbf{A} \mathbf{z} + \mathbf{z}^\top \mathbf{b}) + \mathbf{c} - \frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}$$

Rearrange terms.

$$= \frac{1}{2} (\mathbf{z}^\top \mathbf{A} \mathbf{z} + \mathbf{b}^\top \mathbf{z} + \mathbf{z}^\top \mathbf{b} + \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}) + \mathbf{c} - \frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}$$

We can now make a comparison to the scalar version of completing the square:

$$= \frac{1}{2} (\underbrace{\mathbf{z}^\top \mathbf{A} \mathbf{z}}_{\text{like } Ax^2} + \underbrace{\mathbf{b}^\top \mathbf{z} + \mathbf{z}^\top \mathbf{b}}_{\text{like } 2zb} + \underbrace{\mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}}_{\text{like } \frac{1}{A}b^2}) + \mathbf{c} - \frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}$$

In order to factorize the above in parentheses, we need to fill in some blanks:

$$\frac{1}{2} (\mathbf{z}^\top \mathbf{A} \mathbf{z} + \mathbf{b}^\top \underline{\quad} \mathbf{z} + \mathbf{z}^\top \underline{\quad} \mathbf{b} + \mathbf{b}^\top \mathbf{A}^{-1} \underline{\quad} \mathbf{b}) + \mathbf{c} - \frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}$$

To leave the equation the same, we multiply by the identity matrix:  $I = \mathbf{A}\mathbf{A}^{-1}$ .

$$= \frac{1}{2}(\mathbf{z}^\top \mathbf{A}\mathbf{z} + \mathbf{z}^\top \mathbf{A}\mathbf{A}^{-1}\mathbf{b} + \mathbf{b}^\top \mathbf{A}^{-1}\mathbf{A}\mathbf{z} + \mathbf{b}^\top \mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\mathbf{b}) + \mathbf{c} - \frac{1}{2}\mathbf{b}^\top \mathbf{A}^{-1}\mathbf{b}$$

Pull out  $(\mathbf{z} + \mathbf{A}^{-1}\mathbf{b})$  on the right.

$$= \frac{1}{2}(\mathbf{z}^\top \mathbf{A} + \mathbf{b}^\top \mathbf{A}^{-1}\mathbf{A})(\mathbf{z} + \mathbf{A}^{-1}\mathbf{b}) + \mathbf{c} - \frac{1}{2}\mathbf{b}^\top \mathbf{A}^{-1}\mathbf{b}$$

Pull out  $\mathbf{A}$  from the first expression.

$$= \frac{1}{2}(\mathbf{z}^\top + \mathbf{b}^\top \mathbf{A}^{-1})\mathbf{A}(\mathbf{z} + \mathbf{A}^{-1}\mathbf{b}) + \mathbf{c} - \frac{1}{2}\mathbf{b}^\top \mathbf{A}^{-1}\mathbf{b}$$

Remove a transpose from the first expression.

$$\frac{1}{2}\mathbf{z}^\top \mathbf{A}\mathbf{z} + \mathbf{b}^\top \mathbf{z} + \mathbf{c} = \frac{1}{2}(\mathbf{z} + \mathbf{A}^{-1}\mathbf{b})^\top \mathbf{A}(\mathbf{z} + \mathbf{A}^{-1}\mathbf{b}) + \mathbf{c} - \frac{1}{2}\mathbf{b}^\top \mathbf{A}^{-1}\mathbf{b}$$

□

Back to the problem at hand. Let us complete the square from equation (3):

$$\begin{aligned} & \frac{1}{2}(\mathbf{x}_a - \mathbf{u}_a)^\top \mathbf{\Lambda}_{aa}(\mathbf{x}_a - \mathbf{u}_a) + \frac{1}{2}(\mathbf{x}_a - \mathbf{u}_a)^\top \mathbf{\Lambda}_{ab}(\mathbf{x}_b - \mathbf{u}_b) + \\ & + \frac{1}{2}(\mathbf{x}_b - \mathbf{u}_b)^\top \mathbf{\Lambda}_{ba}(\mathbf{x}_a - \mathbf{u}_a) + \frac{1}{2}(\mathbf{x}_b - \mathbf{u}_b)^\top \mathbf{\Lambda}_{bb}(\mathbf{x}_b - \mathbf{u}_b) \end{aligned}$$

Define the following recalling that all precision and covariance matrices are positive semidefinite:

$$\begin{aligned} \mathbf{z} &= \mathbf{x}_b - \boldsymbol{\mu}_b \\ \mathbf{b} &= \mathbf{\Lambda}_{ba}(\mathbf{x}_b - \boldsymbol{\mu}_b) \\ \mathbf{A} &= \mathbf{\Lambda}_{bb} \\ \mathbf{c} &= \frac{1}{2}(\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \mathbf{\Lambda}_{aa}(\mathbf{x}_a - \boldsymbol{\mu}_a) \end{aligned}$$

Note also that, using the fact that  $\mathbf{\Lambda}_{ba}^\top = \mathbf{\Lambda}_{ab}$ ,

$$\mathbf{b} = \mathbf{\Lambda}_{ba}(\mathbf{x}_b - \boldsymbol{\mu}_b) \implies \mathbf{b}^\top = (\mathbf{x}_b - \boldsymbol{\mu}_b)^\top \mathbf{\Lambda}_{ab}$$

We note,

$$\begin{aligned} \frac{1}{2}\mathbf{z}^\top \mathbf{A}\mathbf{z} &= \frac{1}{2}(\mathbf{x}_b - \boldsymbol{\mu}_b)^\top \mathbf{\Lambda}_{bb}(\mathbf{x}_b - \boldsymbol{\mu}_b) \\ \mathbf{b}^\top \mathbf{z} &= (\mathbf{x}_b - \boldsymbol{\mu}_b)^\top \mathbf{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b) \\ \mathbf{c} &= \frac{1}{2}(\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \mathbf{\Lambda}_{aa}(\mathbf{x}_a - \boldsymbol{\mu}_a) \end{aligned}$$

And applying Theorem 1,

$$\begin{aligned} & \frac{1}{2} \left[ ((\mathbf{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a))^\top \boldsymbol{\Lambda}_{bb} ((\mathbf{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a)) \right] \\ & + \frac{1}{2} (\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{aa} (\mathbf{x}_a - \boldsymbol{\mu}_a) - \frac{1}{2} (\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{ab} \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba} (\mathbf{x}_a - \boldsymbol{\mu}_a) \end{aligned} \quad (4)$$

Keeping track of the  $-1$  that we left out when completing the square, and combining (4) with (2), the full expression becomes:

$$\begin{aligned} f(\mathbf{x}_a) = & \frac{1}{Q} \int_{\mathbb{R}^{n-d}} \exp \left\{ -\frac{1}{2} \left[ ((\mathbf{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a))^\top \boldsymbol{\Lambda}_{bb} ((\mathbf{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a)) \right] \right. \\ & \left. - \frac{1}{2} (\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{aa} (\mathbf{x}_a - \boldsymbol{\mu}_a) + \frac{1}{2} (\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{ab} \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba} (\mathbf{x}_a - \boldsymbol{\mu}_a) \right\} d\mathbf{x}_b \end{aligned}$$

Since we are integrating over  $\mathbf{x}_b$ , all of the terms that contain  $\mathbf{x}_a$  can be taken outside of the integral. Using  $e^{ab} = e^a e^b$ ,

$$\begin{aligned} f(\mathbf{x}_a) = & \frac{1}{Q} \exp \left\{ -\frac{1}{2} (\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{aa} (\mathbf{x}_a - \boldsymbol{\mu}_a) + \frac{1}{2} (\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{ab} \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba} (\mathbf{x}_a - \boldsymbol{\mu}_a) \right\} \\ & \int_{\mathbb{R}^{n-d}} \exp \left\{ -\frac{1}{2} \left[ ((\mathbf{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a))^\top \boldsymbol{\Lambda}_{bb} ((\mathbf{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a)) \right] \right\} d\mathbf{x}_b \end{aligned} \quad (5)$$

We can remove the inner parentheses  $(\mathbf{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a)$  and instead write  $\mathbf{x}_b - \boldsymbol{\mu}_b + \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a)$ .

We now observe that if we replace  $\boldsymbol{\mu}_b + \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba}(\mathbf{x}_a - \boldsymbol{\mu}_a)$  with  $\mathbf{m}$ , the above integral in equation (5) becomes

$$\int_{\mathbb{R}^{n-d}} \exp \left\{ -\frac{1}{2} \left[ ((\mathbf{x}_b - \mathbf{m})^\top \boldsymbol{\Lambda}_{bb} ((\mathbf{x}_b - \mathbf{m})) \right] \right\} d\mathbf{x}_b$$

Note, that this integral is a multivariate normal kernel with mean vector  $\mathbf{m}$  and covariance matrix  $\boldsymbol{\Lambda}_{bb}$  and that we are integrating over the correct dimensionality of  $\mathbf{x}_b$ ! Therefore, the integral evaluates to the normalizing constant

$$(2\pi)^{\frac{n-d}{2}} \det(\boldsymbol{\Lambda}_{bb}^{-1})^{1/2}$$

Plugging this into equation (5), we write

$$\begin{aligned} f(\mathbf{x}_a) = & \frac{1}{Q} (2\pi)^{\frac{n-d}{2}} \det(\boldsymbol{\Lambda}_{bb}^{-1})^{1/2} \\ & \exp \left\{ \frac{1}{2} (\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{aa} (\mathbf{x}_a - \boldsymbol{\mu}_a) + \frac{1}{2} (\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{ab} \boldsymbol{\Lambda}_{bb}^{-1} \boldsymbol{\Lambda}_{ba} (\mathbf{x}_a - \boldsymbol{\mu}_a) \right\} \end{aligned}$$

Pulling out a  $-\frac{1}{2}$ ,

$$f(\mathbf{x}_a) = \frac{1}{Q} (2\pi)^{\frac{n-d}{2}} \det(\mathbf{\Lambda}_{bb}^{-1})^{1/2} \exp \left\{ -\frac{1}{2} [(\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \mathbf{\Lambda}_{aa} (\mathbf{x}_a - \boldsymbol{\mu}_a) - (\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \mathbf{\Lambda}_{ab} \mathbf{\Lambda}_{bb}^{-1} \mathbf{\Lambda}_{ba} (\mathbf{x}_a - \boldsymbol{\mu}_a)] \right\}$$

Then pulling out a  $(\mathbf{x}_a - \boldsymbol{\mu}_a)^\top$  from the left and an  $(\mathbf{x}_a - \boldsymbol{\mu}_a)$  from the right, we get

$$f(\mathbf{x}_a) = \frac{1}{Q} (2\pi)^{n-d/2} \det(\mathbf{\Lambda}_{bb}^{-1})^{1/2} \exp \left\{ -\frac{1}{2} [(\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab} \mathbf{\Lambda}_{bb}^{-1} \mathbf{\Lambda}_{ba} (\mathbf{x}_a - \boldsymbol{\mu}_a)] \right\} \quad (6)$$

Equation (6) appears to be a MVN with mean vector  $\boldsymbol{\mu}_a$  and covariance matrix  $\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab} \mathbf{\Lambda}_{bb}^{-1} \mathbf{\Lambda}_{ba}$ . Now, all that is left is to rewrite the covariance matrix and the normalization constant in terms of  $\boldsymbol{\Sigma}$ .

Recall,

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_{aa} & \mathbf{\Lambda}_{ab} \\ \mathbf{\Lambda}_{ba} & \mathbf{\Lambda}_{bb} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}^{-1} = \boldsymbol{\Sigma}^{-1}$$

Thus,  $\mathbf{\Lambda}^{-1} = \boldsymbol{\Sigma}$  and

$$\begin{bmatrix} \mathbf{\Lambda}_{aa} & \mathbf{\Lambda}_{ab} \\ \mathbf{\Lambda}_{ba} & \mathbf{\Lambda}_{bb} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}$$

We now need a result from linear algebra.

**Theorem 3.** *Let  $\mathbf{M}$  be a symmetric, positive definite block matrix,*

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

*We can define  $\mathbf{M}^{-1}$  using the Shur complement,  $\mathbf{D}$  in  $\mathbf{M}$ :  $(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}$ .*

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}$$

Using Theorem 3, we write

$$\begin{aligned} \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix} &= \begin{bmatrix} \mathbf{\Lambda}_{aa} & \mathbf{\Lambda}_{ab} \\ \mathbf{\Lambda}_{ba} & \mathbf{\Lambda}_{bb} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} (\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab} \mathbf{\Lambda}_{bb}^{-1} \mathbf{\Lambda}_{ba})^{-1} & -(\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab} \mathbf{\Lambda}_{bb}^{-1} \mathbf{\Lambda}_{ba})^{-1} \mathbf{\Lambda}_{ab} \mathbf{\Lambda}_{bb}^{-1} \\ -\mathbf{\Lambda}_{bb} \mathbf{\Lambda}_{ba} (\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab} \mathbf{\Lambda}_{bb}^{-1} \mathbf{\Lambda}_{ba})^{-1} & \mathbf{\Lambda}_{bb}^{-1} + \mathbf{\Lambda}_{bb}^{-1} \mathbf{\Lambda}_{ba} (\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab} \mathbf{\Lambda}_{bb}^{-1} \mathbf{\Lambda}_{ba})^{-1} \mathbf{\Lambda}_{ab} \mathbf{\Lambda}_{bb}^{-1} \end{bmatrix} \end{aligned} \quad (7)$$

Reading across, we see that

$$\begin{aligned}\Sigma_{aa} &= (\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})^{-1} \\ \Sigma_{aa}^{-1} &= \Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\end{aligned}$$

Lastly, we handle the normalization constant. We need one more result from linear algebra regarding the determinant of block matrices:

**Theorem 4.** *Again, let  $\mathbf{M}$  be a symmetric, positive definite block matrix,*

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

*The determinant of  $\mathbf{M}$  is*

$$\det(\mathbf{M}) = \det(\mathbf{A}) \det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})$$

From equation (6), we pull the constant, and replace  $Q$  with its value,

$$\frac{1}{Q} (2\pi)^{\frac{n-d}{2}} \det((\Lambda_{bb}^{-1})^{1/2}) = (2\pi)^{\frac{-n}{2}} \det(\Sigma)^{-1/2} (2\pi)^{\frac{n-d}{2}} \det((\Lambda_{bb}^{-1})^{1/2}) \quad (8)$$

Using theorem 4, the determinant of the block matrix  $\Sigma$  is

$$\det(\Sigma) = \det(\Sigma_{aa}) \det(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})$$

Plugging this into equation (8) and distributing the exponent  $-1/2$ , we can write,

$$= (2\pi)^{\frac{-n}{2}} \det(\Sigma_{aa})^{-1/2} \det(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1/2} (2\pi)^{\frac{n-d}{2}} \det((\Lambda_{bb}^{-1})^{1/2})$$

Rearranging terms.

$$= (2\pi)^{\frac{-n}{2}} (2\pi)^{\frac{n-d}{2}} \det(\Sigma_{aa})^{-1/2} \det(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1/2} \det((\Lambda_{bb}^{-1})^{1/2})$$

Combining exponents over  $\pi$ .

$$= (2\pi)^{\frac{-d}{2}} \det(\Sigma_{aa})^{-1/2} \det(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1/2} \det((\Lambda_{bb}^{-1})^{1/2})$$

And if

$$\Sigma_{aa} = (\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba})^{-1}$$

then,

$$\Sigma_{bb} = (\Lambda_{bb} - \Lambda_{ba}\Lambda_{aa}^{-1}\Lambda_{ab})^{-1}$$

and likewise

$$\begin{aligned}\Lambda_{bb} &= (\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1} \\ \Lambda_{bb}^{-1} &= (\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})\end{aligned}$$



Replacing the  $\mathbf{\Lambda}_{bb}^{-1}$ , we write

$$(2\pi)^{\frac{-d}{2}} \det(\mathbf{\Sigma}_{aa})^{-1/2} \det(\mathbf{\Sigma}_{bb} - \mathbf{\Sigma}_{ba} \mathbf{\Sigma}_{aa}^{-1} \mathbf{\Sigma}_{ab})^{-1/2} \det(\mathbf{\Sigma}_{bb} - \mathbf{\Sigma}_{ba} \mathbf{\Sigma}_{aa}^{-1} \mathbf{\Sigma}_{ab})^{1/2}$$

The two determinants on the right cancel leaving the final normalizing constant as

$$(2\pi)^{\frac{-d}{2}} \det(\mathbf{\Sigma}_{aa})^{-1/2}$$

We can now rewrite equation (6) as

$$f(\mathbf{x}_a) = \frac{1}{(2\pi)^{\frac{-d}{2}} \det(\mathbf{\Sigma}_{aa})^{-1/2}} \exp \left\{ -\frac{1}{2} \left[ (\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \underbrace{\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab} \mathbf{\Lambda}_{bb}^{-1} \mathbf{\Lambda}_{ba}}_{\mathbf{\Sigma}_{aa}^{-1}} (\mathbf{x}_a - \boldsymbol{\mu}_a) \right] \right\} \quad (9)$$

$$= \frac{1}{(2\pi)^{\frac{d}{2}} \det(\mathbf{\Sigma}_{aa})^{1/2}} \exp \left\{ -\frac{1}{2} [(\mathbf{x}_a - \boldsymbol{\mu}_a)^\top \mathbf{\Sigma}_{aa}^{-1} (\mathbf{x}_a - \boldsymbol{\mu}_a)] \right\} \quad (10)$$

Recall that we defined  $\mathbf{x}_a$  as a multivariate normal random vector in  $\mathbb{R}^d$ . Therefore, miraculously, the exponent in the normalizing constant is correct! Furthermore, it is clear that the last equation (10), is the multivariate normal pdf with mean vector  $\boldsymbol{\mu}_a$  and covariance matrix  $\mathbf{\Sigma}_{aa}$ . Thus,

$$p(\mathbf{x}_a) \sim \mathcal{N}(\boldsymbol{\mu}_a, \mathbf{\Sigma}_{aa})$$

and likewise

$$p(\mathbf{x}_b) \sim \mathcal{N}(\boldsymbol{\mu}_b, \mathbf{\Sigma}_{bb}) \quad \square$$

### 3 Conditional of multivariate normal

Remaining in the same setup, our multivariate normal random vector  $\mathbf{x} \in \mathbb{R}^n$  has two multivariate normal random vector partitions  $\mathbf{x}_a \in \mathbb{R}^d$  and  $\mathbf{x}_b \in \mathbb{R}^{(n-d)}$ .

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix}$$

This new partitioned vector should have a multivariate normal distribution denoted by

$$\begin{bmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{bmatrix} = \mathcal{N} \left( \begin{bmatrix} \mathbf{u}_a \\ \mathbf{u}_b \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{aa} & \mathbf{\Sigma}_{ab} \\ \mathbf{\Sigma}_{ba} & \mathbf{\Sigma}_{bb} \end{bmatrix} \right) \quad (11)$$

We now are looking to solve for  $p(\mathbf{x}_a | \mathbf{x}_b)$ . Using Bayes' theorem, we have

$$p(\mathbf{x}_a | \mathbf{x}_b) = \frac{p(\mathbf{x}_a, \mathbf{x}_b)}{p(\mathbf{x}_b)} \propto p(\mathbf{x}_a, \mathbf{x}_b)$$

Using the definition of the pdf of a multivariate normal and the proof of the marginal distribution of a multivariate normal that was just proven, we write,

$$p(\mathbf{x}_a | \mathbf{x}_b) = \frac{\frac{1}{(2\pi)^{\frac{n}{2}} \det(\mathbf{\Sigma})^{1/2}} \exp \left\{ -\frac{1}{2} [(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})] \right\}}{\frac{1}{(2\pi)^{\frac{n-d}{2}} \det(\mathbf{\Sigma}_{bb})^{1/2}} \exp \left\{ -\frac{1}{2} [(\mathbf{x}_b - \boldsymbol{\mu}_b)^\top \mathbf{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b)] \right\}}$$

Clean up the fraction and combine the exponents.

$$= \frac{(2\pi)^{\frac{n-d}{2}} \det(\mathbf{\Sigma}_{bb})^{1/2}}{(2\pi)^{\frac{n}{2}} \det(\mathbf{\Sigma})^{1/2}} \exp \left\{ -\frac{1}{2} [(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})] \right. \\ \left. + \frac{1}{2} [(\mathbf{x}_b - \boldsymbol{\mu}_b)^\top \mathbf{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b)] \right\}$$

We can use the same method of partitioning as equation (2) to write

$$= \frac{1}{(2\pi)^{\frac{n}{2}} \det(\mathbf{\Sigma})^{1/2}} \exp \left\{ -\frac{1}{2} \begin{bmatrix} \mathbf{x}_a - \mathbf{u}_a \\ \mathbf{x}_b - \mathbf{u}_b \end{bmatrix}^\top \begin{bmatrix} \mathbf{\Lambda}_{aa} & \mathbf{\Lambda}_{ab} \\ \mathbf{\Lambda}_{ba} & \mathbf{\Lambda}_{bb} \end{bmatrix} \begin{bmatrix} \mathbf{x}_a - \mathbf{u}_a \\ \mathbf{x}_b - \mathbf{u}_b \end{bmatrix} \right. \\ \left. + \frac{1}{2} [(\mathbf{x}_b - \boldsymbol{\mu}_b)^\top \mathbf{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b)] \right\}$$

Removing the normalization constant for ease, and expanding the matrix multiplications as we did above in equation (3),

$$\exp \left\{ -\frac{1}{2} \left( (\mathbf{x}_a - \mathbf{u}_a)^\top \mathbf{\Lambda}_{aa} (\mathbf{x}_a - \mathbf{u}_a) + (\mathbf{x}_a - \mathbf{u}_a)^\top \mathbf{\Lambda}_{ab} (\mathbf{x}_b - \mathbf{u}_b) \right. \right. \\ \left. \left. + (\mathbf{x}_b - \mathbf{u}_b)^\top \mathbf{\Lambda}_{ba} (\mathbf{x}_a - \mathbf{u}_a) + (\mathbf{x}_b - \mathbf{u}_b)^\top \mathbf{\Lambda}_{bb} (\mathbf{x}_b - \mathbf{u}_b) \right. \right. \\ \left. \left. + \frac{1}{2} (\mathbf{x}_b - \boldsymbol{\mu}_b)^\top \mathbf{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b) \right) \right\} \quad (12)$$

Note that

$$\underbrace{(\mathbf{x}_b - \mathbf{u}_b)^\top}_{1 \times (n-d)} \underbrace{\mathbf{\Lambda}_{ba} (\mathbf{x}_a - \mathbf{u}_a)}_{(n-d) \times d} \underbrace{\mathbf{u}_a}_{d \times 1}$$

is a scalar quantity. Therefore, we can transpose it without concern that it will change the value of the above expression (12). Thus, we can write,

$$\begin{aligned} [(\mathbf{x}_b - \mathbf{u}_b)^\top \mathbf{\Lambda}_{ba} (\mathbf{x}_a - \mathbf{u}_a)]^\top &= (\mathbf{x}_a - \mathbf{u}_a)^\top \mathbf{\Lambda}_{ba}^\top (\mathbf{x}_b - \mathbf{u}_b) \\ \text{Using } \mathbf{\Lambda}_{ba}^\top &= \mathbf{\Lambda}_{ab} \text{ by symmetry} \\ &= (\mathbf{x}_a - \mathbf{u}_a)^\top \mathbf{\Lambda}_{ab} (\mathbf{x}_b - \mathbf{u}_b) \end{aligned}$$

There are now  $2(\mathbf{x}_a - \mathbf{u}_a)^\top \mathbf{\Lambda}_{ab}(\mathbf{x}_b - \mathbf{u}_b)$  in expression (12)

$$\exp \left\{ -\frac{1}{2} \left( (\mathbf{x}_a - \mathbf{u}_a)^\top \mathbf{\Lambda}_{aa}(\mathbf{x}_a - \mathbf{u}_a) + 2(\mathbf{x}_a - \mathbf{u}_a)^\top \mathbf{\Lambda}_{ab}(\mathbf{x}_b - \mathbf{u}_b) \right. \right. \\ \left. \left. + (\mathbf{x}_b - \mathbf{u}_b)^\top \mathbf{\Lambda}_{bb}(\mathbf{x}_b - \mathbf{u}_b) + \frac{1}{2}(\mathbf{x}_b - \boldsymbol{\mu}_b)^\top \boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b) \right) \right\}$$

Again using Theorem 3,

$$\begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{\Lambda}_{aa} & \mathbf{\Lambda}_{ab} \\ \mathbf{\Lambda}_{ba} & \mathbf{\Lambda}_{bb} \end{bmatrix} \\ = \begin{bmatrix} (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1} & -(\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1} \\ -\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}(\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1} & \boldsymbol{\Sigma}_{bb}^{-1} + \boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}(\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1} \end{bmatrix}$$

Plugging in,

$$\begin{aligned} &= \exp \left\{ -\frac{1}{2} \left( (\mathbf{x}_a - \mathbf{u}_a)^\top (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}(\mathbf{x}_a - \mathbf{u}_a) \right. \right. \\ &\quad \left. \left. - 2(\mathbf{x}_a - \mathbf{u}_a)^\top (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \mathbf{u}_b) \right. \right. \\ &\quad \left. \left. + (\mathbf{x}_b - \mathbf{u}_b)^\top \boldsymbol{\Sigma}_{bb}^{-1} + \boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}(\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \mathbf{u}_b) \right) \right. \\ &\quad \left. + \frac{1}{2}(\mathbf{x}_b - \boldsymbol{\mu}_b)^\top \boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b) \right\} \end{aligned}$$

Multiplying out the third term in the exponent,

$$\begin{aligned} &= \exp \left\{ -\frac{1}{2} \left( (\mathbf{x}_a - \mathbf{u}_a)^\top (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}(\mathbf{x}_a - \mathbf{u}_a) \right. \right. \\ &\quad \left. \left. - 2(\mathbf{x}_a - \mathbf{u}_a)^\top (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \mathbf{u}_b) \right. \right. \\ &\quad \left. \left. + \underbrace{(\mathbf{x}_b - \mathbf{u}_b)^\top \boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \mathbf{u}_b)}_{\text{underlined}} \right. \right. \\ &\quad \left. \left. + (\mathbf{x}_b - \mathbf{u}_b)^\top \left[ \boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}(\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{-1}\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1} \right] (\mathbf{x}_b - \mathbf{u}_b) \right) \right. \\ &\quad \left. + \frac{1}{2} \underbrace{(\mathbf{x}_b - \boldsymbol{\mu}_b)^\top \boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b)}_{\text{underlined}} \right\} \end{aligned}$$

The two underlined terms cancel (when the  $-\frac{1}{2}$ ) gets distributed, and we are

left with

$$\begin{aligned}
&= \exp \left\{ -\frac{1}{2} \left( (\mathbf{x}_a - \mathbf{u}_a)^\top (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba})^{-1} (\mathbf{x}_a - \mathbf{u}_a) \right. \right. \\
&\quad \left. \left. - 2(\mathbf{x}_a - \mathbf{u}_a)^\top (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba})^{-1} \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \mathbf{u}_b) \right. \right. \\
&\quad \left. \left. + (\mathbf{x}_b - \mathbf{u}_b)^\top [\boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba} (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba})^{-1} \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1}] (\mathbf{x}_b - \mathbf{u}_b) \right) \right\} \quad (13)
\end{aligned}$$

In order apply another result from linear algebra, we need to summarize equation (13) using the following definitions. Set

$$\begin{aligned}
\mathbf{a} &= (\mathbf{x}_a - \boldsymbol{\mu}_a) & \mathbf{B} &= (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba})^{-1} \\
\mathbf{d} &= (\mathbf{x}_b - \boldsymbol{\mu}_b) & \mathbf{C} &= \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \\
& & \mathbf{D} &= \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba}
\end{aligned} \quad (14)$$

We also note that

$$\begin{aligned}
\mathbf{C} &= \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \\
\mathbf{C}^\top &= (\boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1})^\top \\
&= \boldsymbol{\Sigma}_{bb}^{-\top} \boldsymbol{\Sigma}_{ab}^{-1} \\
&\text{Transpose of symmetric matrix is that same matrix.} \\
&\text{And note } \boldsymbol{\Sigma}_{ab}^{-1} = \boldsymbol{\Sigma}_{ba} \\
&= \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba} = \mathbf{D}
\end{aligned}$$

Rewriting equation (13) with these definitions (14) and the replacement of  $\mathbf{D}$  with  $\mathbf{C}^\top$ ,

$$\exp \left\{ -\frac{1}{2} (\mathbf{a}^\top \mathbf{B} \mathbf{a} - 2\mathbf{a}^\top \mathbf{B} \mathbf{C} \mathbf{d} + \mathbf{d} \mathbf{C}^\top \mathbf{B} \mathbf{C} \mathbf{d}) \right\} \quad (15)$$

Note that  $\mathbf{C} \mathbf{d}$  is a vector. Using the following theorem from linear algebra.

**Theorem 5.** *For any vectors  $\mathbf{u}$  and  $\mathbf{v}$  and symmetric matrix  $\mathbf{A}$ :*

$$\mathbf{u}^\top \mathbf{A} \mathbf{u} - 2\mathbf{u}^\top \mathbf{A} \mathbf{v} + \mathbf{v}^\top \mathbf{A} \mathbf{v} = (\mathbf{u}^\top - \mathbf{v}^\top) \mathbf{A} (\mathbf{u} - \mathbf{v})$$

*Proof.*

$$\begin{aligned}
\mathbf{u}^\top \mathbf{A} \mathbf{u} - 2\mathbf{u}^\top \mathbf{A} \mathbf{v} + \mathbf{v}^\top \mathbf{A} \mathbf{v} &= \mathbf{u}^\top \mathbf{A} \mathbf{u} - \mathbf{u}^\top \mathbf{A} \mathbf{v} - \mathbf{u}^\top \mathbf{A} \mathbf{v} + \mathbf{v}^\top \mathbf{A} \mathbf{v} \\
&= \mathbf{u}^\top \mathbf{A} (\mathbf{u} - \mathbf{b}) - \mathbf{u}^\top \mathbf{A} \mathbf{v} + \mathbf{v}^\top \mathbf{A} \mathbf{v} \\
&= \mathbf{u}^\top \mathbf{A} (\mathbf{u} - \mathbf{b}) - (\mathbf{u}^\top - \mathbf{v}^\top) \mathbf{A} \mathbf{v} \\
&= \mathbf{u}^\top \mathbf{A} (\mathbf{u} - \mathbf{b}) - \underbrace{(\mathbf{u} - \mathbf{v})^\top \mathbf{A} \mathbf{v}}_{\text{transpose}} \\
&= \mathbf{u}^\top \mathbf{A} (\mathbf{u} - \mathbf{b}) - \mathbf{v}^\top \mathbf{A} (\mathbf{u} - \mathbf{v}) \\
&= (\mathbf{u}^\top \mathbf{A} - \mathbf{v}^\top \mathbf{A}) (\mathbf{u} - \mathbf{v}) \\
&= (\mathbf{u}^\top - \mathbf{v}^\top) \mathbf{A} (\mathbf{u} - \mathbf{v})
\end{aligned}$$

□

Call vector  $\mathbf{C}\mathbf{d} = \mathbf{e}$  and rewrite equation (15),

$$\exp \left\{ -\frac{1}{2} (\mathbf{a}^\top \mathbf{B} \mathbf{a} - 2\mathbf{a}^\top \mathbf{B} \mathbf{e} + \mathbf{d} \mathbf{C}^\top \mathbf{B} \mathbf{e}) \right\}$$

Using Theorem 5, we have

$$\exp \left\{ -\frac{1}{2} (\mathbf{a}^\top \mathbf{B} \mathbf{a} - 2\mathbf{a}^\top \mathbf{B} \mathbf{e} + \mathbf{d} \mathbf{C}^\top \mathbf{B} \mathbf{e}) \right\} = \exp \left\{ -\frac{1}{2} (\mathbf{a} - \mathbf{e})^\top \mathbf{B} (\mathbf{a} - \mathbf{e}) \right\}$$

Substitute back

$$\begin{aligned}
&\exp \left\{ -\frac{1}{2} ((\mathbf{x}_a - \boldsymbol{\mu}_a) - \mathbf{C}\mathbf{d})^\top \mathbf{B} ((\mathbf{x}_a - \boldsymbol{\mu}_a) - \mathbf{C}\mathbf{d}) \right\} \\
&= \exp \left\{ -\frac{1}{2} ((\mathbf{x}_a - \boldsymbol{\mu}_a) - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b))^\top (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba})^{-1} \right. \\
&\quad \left. ((\mathbf{x}_a - \boldsymbol{\mu}_a) - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b)) \right\}
\end{aligned}$$

Pull out the minus sign to get the equation into the form  $(\mathbf{x} - \boldsymbol{\mu})$ :

$$\begin{aligned}
&\exp \left\{ -\frac{1}{2} ((\mathbf{x}_a - (\boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b)))^\top (\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba})^{-1} \right. \\
&\quad \left. ((\mathbf{x}_a - (\boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b))) \right\} \quad (16)
\end{aligned}$$

Equation (16) now looks like the pdf of a multivariate normal distribution with mean vector

$$\boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b)$$

and covariance matrix

$$\boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba}$$

All that is left is to deal with the normalizing constant we left up above.

$$\frac{1}{(2\pi)^{\frac{d}{2}}} \cdot \frac{\det(\Sigma_{bb})^{\frac{1}{2}}}{\det(\Sigma)^{\frac{1}{2}}}$$

Using Theorem 4,

$$\begin{aligned} \det(\Sigma) &= \det \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} = \det(\Sigma_{bb}) \det(\Sigma_{bb} - \Sigma_{aa} \Sigma_{bb}^{-1} \Sigma_{ba}) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \cdot \frac{\det(\Sigma_{bb})^{\frac{1}{2}}}{\det(\Sigma_{bb})^{\frac{1}{2}} \det(\Sigma_{bb} - \Sigma_{aa} \Sigma_{bb}^{-1} \Sigma_{ba})^{\frac{1}{2}}} \end{aligned} \quad (17)$$

$$= \frac{1}{(2\pi)^{\frac{d}{2}}} \cdot \frac{1}{\det(\Sigma_{bb} - \Sigma_{aa} \Sigma_{bb}^{-1} \Sigma_{ba})^{\frac{1}{2}}} \quad (18)$$

Putting equations (16) and (17) together, we get

$$\begin{aligned} p(\mathbf{x}_a | \mathbf{x}_b) &= \frac{1}{(2\pi)^{\frac{d}{2}} \det(\Sigma_{bb} - \Sigma_{aa} \Sigma_{bb}^{-1} \Sigma_{ba})^{\frac{1}{2}}} \cdot \\ &\exp \left\{ -\frac{1}{2} ((\mathbf{x}_a - (\boldsymbol{\mu}_a + \Sigma_{ab} \Sigma_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b)))^\top (\Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba})^{-1} \right. \\ &\quad \left. ((\mathbf{x}_a - (\boldsymbol{\mu}_a + \Sigma_{ab} \Sigma_{bb}^{-1} (\mathbf{x}_b - \boldsymbol{\mu}_b))) \right\} \end{aligned} \quad (19)$$

We can now read off the conditional mean vector and covariance matrix as

$$\boldsymbol{\mu}_a + \Sigma_{ab} \Sigma_{bb}^{-1}$$

and

$$\Sigma_{bb} - \Sigma_{aa} \Sigma_{bb}^{-1} \Sigma_{ba} \quad \square$$

Credit for parts of the proofs above goes to the following [1], [2], [3] and [4].

## References

- [1] Felix Kohler. Marginal & Conditional for the Multivariate Normal — Full Derivation. Machine Learning & Simulation. July 1, 2021. URL: <https://www.youtube.com/watch?v=H9ETCyDx80g> (visited on 12/02/2025).
- [2] JoramSoch. Proof: Conditional distributions of the multivariate normal distribution. The Book of Statistical Proofs, Proof #P88. Mar. 20, 2020. DOI: [10.5281/zenodo.4305949](https://doi.org/10.5281/zenodo.4305949). URL: <https://statproofbook.github.io/P/mvn-cond.html> (visited on 12/02/2025).

- [3] Ruye Wang. Machine Learning Tutorial Series: Gaussian Processes. Harvey Mudd College. Nov. 14, 2006. URL: [https://www.khoury.northeastern.edu/home/vip/teach/DMcourse/5\\_topicmodel\\_summ/sampling/Marginal\\_cond\\_for\\_multivariate\\_normal.pdf](https://www.khoury.northeastern.edu/home/vip/teach/DMcourse/5_topicmodel_summ/sampling/Marginal_cond_for_multivariate_normal.pdf) (visited on 12/02/2025).
- [4] Kevin P Murphy. Machine learning: a probabilistic perspective. MIT press, 2012.