

# 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 covariace **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 (1), we can write

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

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

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

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

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

Now we can rewrite the above as:

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

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

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

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

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

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

Plugging this into equation (5), we write

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

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

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

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

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

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

Recall,

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

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

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

We now need a result from linear algebra.

**Theorem 3.** *For a block matrix*

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

where  $M$  is symmetric, positive definite, we can define  $M^{-1}$  using the Shur complement of  $D$  in  $M$ ,  $(A - BD^{-1}C)^{-1}$ :

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

Using Theorem (3), we write

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

### 3 Conditional of multivariate normal