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[1] random vector[2] $\boldsymbol{x} \in \mathbb{R}^n$. That is,

$$\boldsymbol{x} = \begin{bmatrix} x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ x_2 \sim \mathcal{N}(\mu_2, \sigma_2^2) \\ \vdots \\ x_n \sim \mathcal{N}(\mu_3, \sigma_3^2) \end{bmatrix}$$

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

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

with the subscipt n indicating that vector $\boldsymbol{x}$ 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***:

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

[1] Gaussian and normal will be interchanged throughout this document and used as synonyms.

[2] 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 $\boldsymbol{x}$.
3. $\boldsymbol{\mu}$ can assign a different μ_i for each $x_i \in \boldsymbol{x}$.
4. $\boldsymbol{\Sigma}$ is symmetric and all its entries are positive or zero (more strictly, it must be positive semidefinite).
5. The diagnoal 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:

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

Now, let us partition our multivariate normal random vector $\boldsymbol{x} \in \mathbb{R}^n$ into two multvariate normal random vectors $\boldsymbol{x}_a \in \mathbb{R}^d$ and $\boldsymbol{x}_b \in \mathbb{R}^{(n-d)}$:[3]

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

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

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

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 $\boldsymbol{x}_a$ given a realization of the vector $\boldsymbol{x}_b$ are
   - $\boldsymbol{\mu}_{\boldsymbol{x}_a \mid \boldsymbol{x}_b} = \boldsymbol{\mu}_{\boldsymbol{a}} + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{x}_b - \boldsymbol{\mu}_b)$
   - $\boldsymbol{\Sigma}_{\boldsymbol{x}_a \mid \boldsymbol{x}_b} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}$
2. The marginal mean vector and covariance matrix of a multivariate normal for $\boldsymbol{x}_a$ are
   - $\boldsymbol{\mu}_{\boldsymbol{x}_a} = \boldsymbol{\mu}_a$
   - $\boldsymbol{\Sigma}_{\boldsymbol{x}_a} = \boldsymbol{\Sigma}_{aa}$

[3] Note the dimensions of each vector!

2 Marginal of mulitvariate normal

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

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

because we are integrating out $\boldsymbol{x}_b$ leaving only the pdf of $\boldsymbol{x}_a$. Plugging in the values of the mean vectors and covariance matrices from equation (1), we can write

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

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

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

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

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

Now we can rewrite the above as:

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

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

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

$$\begin{bmatrix}\boldsymbol{x}_a - \boldsymbol{u}_a\\ \boldsymbol{x}_b - \boldsymbol{u}_b\end{bmatrix}^\top \begin{bmatrix}\boldsymbol{\Lambda}_{aa}(\boldsymbol{x}_a - \boldsymbol{u}_a) + \boldsymbol{\Lambda}_{ab}(\boldsymbol{x}_b - \boldsymbol{u}_b)\\ \boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{u}_a) + \boldsymbol{\Lambda}_{ab}(\boldsymbol{x}_b - \boldsymbol{u}_b)\end{bmatrix}$$

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

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

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

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

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

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

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

Proof. Thus,

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

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

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

Rearrange terms.

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

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

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

Rearrange terms.

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

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

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

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

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

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

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

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

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

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

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

Remove a transpose from the first expression.

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

$\square$

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

$$\frac{1}{2}(\boldsymbol{x}_a - \boldsymbol{u}_a)^\top \boldsymbol{\Lambda}_{aa}(\boldsymbol{x}_a - \boldsymbol{u}_a) + \frac{1}{2}(\boldsymbol{x}_a - \boldsymbol{u}_a)^\top \boldsymbol{\Lambda}_{ab}(\boldsymbol{x}_b - \boldsymbol{u}_b) + \\ + \frac{1}{2}(\boldsymbol{x}_b - \boldsymbol{u}_b)^\top \boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{u}_a) + \frac{1}{2}(\boldsymbol{x}_b - \boldsymbol{u}_b)^\top \boldsymbol{\Lambda}_{bb}(\boldsymbol{x}_b - \boldsymbol{u}_b)$$

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

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

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

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

We note,

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

And applying Theorem (1),

$$\frac{1}{2}\Big[\big((\boldsymbol{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\big)^\top \boldsymbol{\Lambda}_{bb}\big((\boldsymbol{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\big)\Big] + \frac{1}{2}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{aa}(\boldsymbol{x}_a - \boldsymbol{\mu}_a) - \frac{1}{2}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a) \quad (4)$$

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

$$f(\boldsymbol{x}_a) = \frac{1}{Q}\int_{\mathbb{R}^{n-d}} \exp\Big\{-\frac{1}{2}\Big[\big((\boldsymbol{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\big)^\top \boldsymbol{\Lambda}_{bb}\big((\boldsymbol{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\big)\Big] - \frac{1}{2}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{aa}(\boldsymbol{x}_a - \boldsymbol{\mu}_a) + \frac{1}{2}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\Big\}\, d\boldsymbol{x}_b$$

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

$$f(\boldsymbol{x}_a) = \frac{1}{Q}\exp\Big\{-\frac{1}{2}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{aa}(\boldsymbol{x}_a - \boldsymbol{\mu}_a) + \frac{1}{2}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\Big\} \int_{\mathbb{R}^{n-d}} \exp\Big\{-\frac{1}{2}\Big[\big((\boldsymbol{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\big)^\top \boldsymbol{\Lambda}_{bb}\big((\boldsymbol{x}_b - \boldsymbol{\mu}_b) + \boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\big)\Big]\Big\}\, d\boldsymbol{x}_b \quad (5)$$

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

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

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

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

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

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

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

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

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

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

Recall,

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

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

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

We now need a result from linear algebra.

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

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

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

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

Using Theorem (3), we write

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

Reading across, we see that

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

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

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

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

The determinant of $\boldsymbol{M}$ is

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

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

$$\frac{1}{Q}(2\pi)^{\frac{n-d}{2}}\det((\mathbf{\Lambda}_{bb}^{-1})^{1/2}) = (2\pi)^{\frac{-n}{2}}\det(\mathbf{\Sigma})^{-1/2}(2\pi)^{\frac{n-d}{2}}\det((\mathbf{\Lambda}_{bb}^{-1})^{1/2}) \tag{8}$$

Using theorem (4), the determinant of the block matrix $\mathbf{\Sigma}$ is

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

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

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

Rearranging terms.

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

Combining exponents over π.

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

And if

$$\mathbf{\Sigma}_{aa} = (\mathbf{\Lambda}_{aa} - \mathbf{\Lambda}_{ab}\mathbf{\Lambda}_{bb}^{-1}\mathbf{\Lambda}_{ba})^{-1}$$

then,

$$\mathbf{\Sigma}_{bb} = (\mathbf{\Lambda}_{bb} - \mathbf{\Lambda}_{ba}\mathbf{\Lambda}_{aa}^{-1}\mathbf{\Lambda}_{ab})^{-1}$$

and likewise

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

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

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

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

Recall that we defined $\boldsymbol{x}_a$ 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,

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

and likewise

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

3 Conditional of multivariate normal

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

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

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

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