Derivation of Multivariate Normal marginal and conditional distributions

Mazin Abdelghany, MD, MS

3 December 2025

Contents

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:

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

---

[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.

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**:

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

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

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

   - $\boldsymbol{\Sigma}_{\boldsymbol{x}_a \mid \boldsymbol{x}_b} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab} \boldsymbol{\Sigma}_{bb}^{-1} \boldsymbol{\Sigma}_{ba}$

---

[3]Note the dimensions of each vector!

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

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

2 Marginal distribution 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 (11), we can write

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

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

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:

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

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:

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

3

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

$$+ \frac{1}{2}(\boldsymbol{x}_b - \boldsymbol{u}_b)^\top \boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{u}_a) + \frac{1}{2}(\boldsymbol{x}_b - \boldsymbol{u}_b)^\top \boldsymbol{\Lambda}_{bb}(\boldsymbol{x}_b - \boldsymbol{u}_b) \quad (3)$$

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*

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

$$= b_1 z_1 + b_2 z_1 + \dots + b_n z_n \qquad = z_1 b_1 + z_1 b_2 + \dots + z_n b_n$$

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 $\frac{1}{2}$.

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

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:

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

4

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

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

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

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

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

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

Pull out A from the first expression.

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

Remove a transpose from the first expression.

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

$\square$

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

$$\frac{1}{2}(x_a - u_a)^\top \Lambda_{aa}(x_a - u_a) + \frac{1}{2}(x_a - u_a)^\top \Lambda_{ab}(x_b - u_b) +$$

$$+ \frac{1}{2}(x_b - u_b)^\top \Lambda_{ba}(x_a - u_a) + \frac{1}{2}(x_b - u_b)^\top \Lambda_{bb}(x_b - u_b)$$

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

$$z = x_b - \mu_b$$

$$A = \Lambda_{bb} \qquad b = \Lambda_{ba}(x_b - \mu_b)$$

$$c = \frac{1}{2}(x_a - \mu_a)^\top \Lambda_{aa}(x_a - \mu_a)$$

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

$$b = \Lambda_{ba}(x_a - \mu_a) \implies b^\top = (x_a - \mu_a)^\top \Lambda_{ab}$$

We note,

$$\frac{1}{2}z^\top Az = \frac{1}{2}(x_b - \mu_b)^\top \Lambda_{bb}(x_b - \mu_b)$$

$$b^\top z = (x_a - \mu_a)^\top \Lambda_{ab}(x_b - \mu_b)$$

$$c = \frac{1}{2}(x_a - \mu_a)^\top \Lambda_{aa}(x_a - \mu_a)$$

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:

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

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 $e^{ab} = e^a e^b$,

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

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

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

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

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

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

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

$$
\exp\left\{-\frac{1}{2}\left[(\boldsymbol{x}_a - \boldsymbol{\mu}_a)^\top \boldsymbol{\Lambda}_{aa} - \boldsymbol{\Lambda}_{ab}\boldsymbol{\Lambda}_{bb}^{-1}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)\right]\right\} \quad (6)
$$

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

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

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

Reading across, we see that

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

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,

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

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

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

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

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

Rearranging terms.

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

Combining exponents over π.

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

And if

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

then,

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

and likewise

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

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

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

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

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

We can now rewrite equation (6) as

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

(9)

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

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 $\boldsymbol{\Sigma}_{aa}$. Thus,

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

and likewise

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

3 Conditional distribution 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

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

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

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

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,

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

Clean up the fraction and combine the exponents.

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

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

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

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

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

Note that

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

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,

$$\left[(\boldsymbol{x}_b-\boldsymbol{u}_b)^{\top}\boldsymbol{\Lambda}_{ba}(\boldsymbol{x}_a-\boldsymbol{u}_a)\right]^{\top} = (\boldsymbol{x}_a-\boldsymbol{u}_a)^{\top}\boldsymbol{\Lambda}_{ba}^{\top}(\boldsymbol{x}_b-\boldsymbol{u}_b)$$
$$\text{Using } \boldsymbol{\Lambda}_{ba}^{\top} = \boldsymbol{\Lambda}_{ab} \text{ by symmetry}$$
$$= (\boldsymbol{x}_a-\boldsymbol{u}_a)^{\top}\boldsymbol{\Lambda}_{ab}(\boldsymbol{x}_b-\boldsymbol{u}_b)$$

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

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

Again using Theorem 3,

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

Plugging in,

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

Multiplying out the third term in the exponent,

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

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

left with

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

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

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

We also note that

$$
\begin{aligned}
\boldsymbol{C} &= \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1} \\
\boldsymbol{C}^\top &= \big(\boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\big)^\top \\
&= \boldsymbol{\Sigma}_{bb}^{-\top}\boldsymbol{\Sigma}_{ab}^{-1}
\end{aligned}
$$

Transpose of symmetric matrix is that same matrix.

And note $\boldsymbol{\Sigma}_{ab}^{-1} = \boldsymbol{\Sigma}_{ba}$

$$
= \boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba} = \boldsymbol{D}
$$

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

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

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

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

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

Proof.

$$\boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{u} - 2\boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} = \boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{u} - \boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{v} - \boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v}$$

$$= \boldsymbol{u}^\top \boldsymbol{A}(\boldsymbol{u} - \boldsymbol{b}) - \boldsymbol{u}^\top \boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v}$$

$$= \boldsymbol{u}^\top \boldsymbol{A}(\boldsymbol{u} - \boldsymbol{b}) - (\boldsymbol{u}^\top - \boldsymbol{v}^\top)\boldsymbol{A}\boldsymbol{v}$$

$$= \boldsymbol{u}^\top \boldsymbol{A}(\boldsymbol{u} - \boldsymbol{b}) - \underbrace{(\boldsymbol{u} - \boldsymbol{v})^\top \boldsymbol{A}\boldsymbol{v}}_{\text{transpose}}$$

$$= \boldsymbol{u}^\top \boldsymbol{A}(\boldsymbol{u} - \boldsymbol{b}) - \boldsymbol{v}^\top \boldsymbol{A}(\boldsymbol{u} - \boldsymbol{v})$$

$$= (\boldsymbol{u}^\top \boldsymbol{A} - \boldsymbol{v}^\top \boldsymbol{A})(\boldsymbol{u} - \boldsymbol{v})$$

$$= (\boldsymbol{u}^\top - \boldsymbol{v}^\top)\boldsymbol{A}(\boldsymbol{u} - \boldsymbol{v})$$

$\square$

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

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

Using Theorem 5, we have

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

Substitute back

$$\exp\left\{ -\frac{1}{2}((\boldsymbol{x}_a - \boldsymbol{\mu}_a) - \boldsymbol{C}\boldsymbol{d})^\top \boldsymbol{B}((\boldsymbol{x}_a - \boldsymbol{\mu}_a) - \boldsymbol{C}\boldsymbol{d}) \right\}$$

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

$$\left. ((\boldsymbol{x}_a - \boldsymbol{\mu}_a) - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{x}_b - \boldsymbol{\mu}_b)) \right\}$$

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

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

$$\left. ((\boldsymbol{x}_a - (\boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{x}_b - \boldsymbol{\mu}_b))) \right\} \quad (16)$$

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

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

and covariance matrix

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

13

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

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

Using Theorem 4,

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

$$= \frac{1}{(2\pi)^{\frac{d}{2}}} \cdot \frac{\det(\boldsymbol{\Sigma}_{bb})^{\frac{1}{2}}}{\det(\boldsymbol{\Sigma}_{bb})^{\frac{1}{2}}\det(\boldsymbol{\Sigma}_{bb} - \boldsymbol{\Sigma}_{aa}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{\frac{1}{2}}} \tag{17}$$

$$= \frac{1}{(2\pi)^{\frac{d}{2}}} \cdot \frac{1}{\det(\boldsymbol{\Sigma}_{bb} - \boldsymbol{\Sigma}_{aa}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{\frac{1}{2}}} \tag{18}$$

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

$$p(\boldsymbol{x}_a \,|\, \boldsymbol{x}_b) =$$

$$\frac{1}{(2\pi)^{\frac{d}{2}}\det(\boldsymbol{\Sigma}_{bb} - \boldsymbol{\Sigma}_{aa}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba})^{\frac{1}{2}}} \cdot$$

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

$$\left.((\boldsymbol{x}_a - (\boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\boldsymbol{x}_b - \boldsymbol{\mu}_b))\right\} \tag{19}$$

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

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

and

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

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. 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 (visited on 12/02/2025).

[4]   Kevin P Murphy. Machine learning: a probabilistic perspective. MIT press, 2012.