1 Performing integrals over normal distributes takes practice.

(a) Consider the integral

$$I_1 = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2} \mathrm{d}x.$$

Directly evaluating this is difficult, but there is a trick. Consider instead

$$I_1^2 = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{e}^{-y^2/2} \mathrm{d}y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-(x^2+y^2)/2} \mathrm{d}x \mathrm{d}y.$$

By making the change of variables to polar coordinates where $r = \sqrt{x^2+y^2}$ and $\theta = \arctan(y/x)$ (so that $x = r\cos(\theta)$, $y = r\sin(\theta)$) then $\mathrm{d}x\mathrm{d}y = r\mathrm{d}r\mathrm{d}\theta$. Note that to integrate over all space we let θ vary from 0 to 2π and r to vary from 0 to ∞. Write down the integral in polar coordinate, make a further the change of variables $u = r^2/2$ to evaluate I_1^2 hence compute I_1 [5 marks]

$$I_1^2 = \int_0^{2\pi} \mathrm{d}\theta \int_0^{\infty} r e^{-r^2/2} \mathrm{d}r = 2\pi \int_0^{\infty} r e^{-r^2/2} \mathrm{d}r.$$

Making the change of variables $u = r^2/2$ so that $\mathrm{d}u/\mathrm{d}r = r$ or $\mathrm{d}u = r\mathrm{d}r$ we get

$$I_1^2 = 2\pi \int_0^{\infty} e^{-u} \mathrm{d}u = 2\pi$$

so $I_1 = \sqrt{2\pi}$.

(b) By making a change of variables compute

$$I_2 = \int_{-\infty}^{\infty} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)} \mathrm{d}x$$

[5 marks]

The trick is to let $z = (x-\mu)/\sigma$ so that $\mathrm{d}z = \mathrm{d}x/\sigma$ or $\mathrm{d}x = \sigma\mathrm{d}z$. Then

$$I_2 = \sigma \int_{-\infty}^{\infty} \mathrm{e}^{-z^2/2} \mathrm{d}z = \sigma I_1 = \sqrt{2\pi}\sigma$$

Note that the *probability density function* (PDF) for a normally distributed random variable is given by

$$\mathcal{N}(x \mid \mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}.$$

(Observe that $\mathcal{N}(0 \mid \mu,\sigma^2) = 1/(\sqrt{2\pi}\sigma)$ so when $\sigma < \sqrt{2\pi}$ then $\mathcal{N}(0 \mid \mu,\sigma^2) > 1$, showing that PDFs are not probabilities.)

(c) By using the identity $\mathrm{e}^{a+b} = \mathrm{e}^a \mathrm{e}^b$, or more generally

$$\mathrm{e}^{\sum_i a_i} = \prod_i \mathrm{e}^{a_i}$$

compute

$$I_3 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{x}\|_2^2} \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathsf{T}}$. [3 marks]

We note that $\|\boldsymbol{x}\|_2^2 = \sum_{i=1}^{n} x_i^2$

$$\begin{aligned} I_3 &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2} \mathrm{d}x_1 \cdots \mathrm{d}x_n \\ &= \prod_{i=1}^{n} \int_{-\infty}^{\infty} \mathrm{e}^{-x_i^2/2} \mathrm{d}x_i = \prod_{i=1}^{n} \sqrt{2\pi} = (2\pi)^{n/2}. \end{aligned}$$

(d) By using the fact that for a positive semi-definite matrix, $\boldsymbol{\Xi}$, we can use the eigenvector decomposition $\boldsymbol{\Xi}^{-1} = \mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{\mathsf{T}}$ where $\mathbf{V}$ is an orthogonal matrix with determinant $\det(\mathbf{V}) = \pm 1$ and $\boldsymbol{\Lambda}^{-1}$ is a diagonal matrix with elements λ_i^{-1} compute

$$I_4 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Xi}^{-1}(\boldsymbol{x}-\mu)} \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

[6 marks]

This needs some confidence to push through to the end.

$$I_4 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{V}\boldsymbol{\Lambda}^{-1}\mathbf{V}^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{\mu})} \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

We make the change of variables $\boldsymbol{y} = \mathbf{V}^{\mathsf{T}}(\boldsymbol{x} - \boldsymbol{\mu})$. However, to make a change of variables, $\boldsymbol{x} \to \boldsymbol{y}$, where we use

$$\iiint_{\mathcal{R}(\boldsymbol{x})} f(\boldsymbol{x}) \mathrm{d}x_1 \cdots \mathrm{d}x_n = \iiint_{\mathcal{R}(\boldsymbol{y})} f(\boldsymbol{x}(\boldsymbol{y})) |\mathbf{J}| \mathrm{d}y_1 \cdots \mathrm{d}y_n$$

where $\mathcal{R}(\boldsymbol{y})$ is the same region as $\mathcal{R}(\boldsymbol{x})$ but specified in $\boldsymbol{y}$-coordinates and $\mathbf{J}$ is the Jacobian matrix with elements $J_{ij} = \partial x_i / \partial y_j$. In our case $\boldsymbol{x}(\boldsymbol{y}) = \mathbf{V}\boldsymbol{y} + \boldsymbol{\mu}$

$$J_{ij} = \frac{\partial x_i}{\partial y_j} = \frac{\partial}{\partial y_j}\left(\sum_{k=1}^{n} V_{ik} y_k - \mu_i\right) = V_{ij}$$

so that $\mathbf{J} = \mathbf{V}$ and $|\det(\mathbf{J})| = |\det(\mathbf{V})| = 1$. This makes sense as the matrix $\mathbf{V}$ corresponds to a rotation (with a possible reflection) which does not change the volume.

Thus on making the change of variables

$$I_4 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\boldsymbol{y}^\mathsf{T}\boldsymbol{\Lambda}^{-1}\boldsymbol{y}} \mathrm{d}y_1 \cdots \mathrm{d}y_n. = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \mathrm{e}^{-y_i^2/(2\lambda_i)} \mathrm{d}y_i = \prod_i \sqrt{2\pi\lambda_i}$$

where we used I_2 with $\sigma = \sqrt{\lambda_i}$. (Note if $\boldsymbol{\Xi}$ was not positive semi-definite λ_i would be negative and the integral would not converge.)

(e) Using the facts, that $\boldsymbol{\Xi} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\mathsf{T}$, for any two square matrices $\mathbf{A}$ and $\mathbf{B}$ the determinants satisfy $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$, and $\det(\mathbf{V}) = \det(\mathbf{V}^\mathsf{T}) = \pm 1$ show that $\det(\boldsymbol{\Xi}) = \prod_i \lambda_i$. [1 mark]

$$\det(\boldsymbol{\Xi}) = \det(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\mathsf{T}) = \det(\mathbf{V})\det(\boldsymbol{\Lambda})\det(\mathbf{V}^\mathsf{T}) = \det(\boldsymbol{\Lambda}) = \prod_{i=1}^{n} \lambda_i$$

We have used that $\det(\mathbf{V}) = \det(\mathbf{V}^\mathsf{T})$ which equals 1 or -1, but $\det(\mathbf{V}) \times \det(\mathbf{V}^\mathsf{T}) = 1$. Also we use that the determinant of a diagonal matrix is equal to the product of the diagonal elements. We note that

$$I_4 = (2\pi)^{n/2}\sqrt{\det(\boldsymbol{\Xi})}$$

There is another trick to simplify the notation a bit more. For an $n \times n$ matrix $\mathbf{M}$ then $\det(c\mathbf{M}) = c^n \det(\mathbf{M})$ as the determinant of the matrix involves a sum of terms where each term involves a product of n elements. Multiplying each element by c means that the determinant increases by c^n. Thus, we can write

$$I_4 = (2\pi)^{n/2}\sqrt{\det(\boldsymbol{\Xi})} = \sqrt{\det(2\pi\boldsymbol{\Xi})}$$

and the multivariate normal PDF is

$$\mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Xi}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Xi})}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Xi}^{-1}(\boldsymbol{x}-\mu)}$$

End of question 1

2 Consider a multivariate normal distribution

$$f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\mathsf{T} & \mathbf{C} \end{pmatrix} \right)$$

where $\mathbf{A}$ and $\mathbf{C}$ are symmetric (positive definite) matrices. The matrix

$$\boldsymbol{\Xi} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\mathsf{T} & \mathbf{C} \end{pmatrix}$$

is the covariance matrix.

We want to compute the conditional probability density function $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y})$. This is complicated because the normal distribution involve the inverse of the covariance matrix. Let

$$\mathbf{U} = \begin{pmatrix} \mathbf{I} & \mathbf{B}\mathbf{C}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \qquad \mathbf{D} = \begin{pmatrix} \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & \mathbf{C} \end{pmatrix}$$

where $\mathbf{I}$ is the identity matrix.

(a) Compute $\mathbf{U}\mathbf{D}$ [3 marks]

$$\mathbf{U}\mathbf{D} = \begin{pmatrix} \mathbf{I} & \mathbf{B}\mathbf{C}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & \mathbf{C} \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T} & \mathbf{B} \\ \mathbf{0}^\mathsf{T} & \mathbf{C} \end{pmatrix}$$

(b) Using the previous result compute $(\mathbf{U}\mathbf{D})\mathbf{U}^\mathsf{T}$. Hence show $\boldsymbol{\Xi} = \mathbf{U}\mathbf{D}\mathbf{U}^\mathsf{T}$. [3 marks]

$$(\mathbf{U}\mathbf{D})\mathbf{U}^\mathsf{T} = \begin{pmatrix} \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T} & \mathbf{B} \\ \mathbf{0}^\mathsf{T} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}^{-1}\mathbf{B}^\mathsf{T} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\mathsf{T} & \mathbf{C} \end{pmatrix} = \boldsymbol{\Xi}$$

(c) Given that $\Xi = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathsf{T}}$ write down Ξ^{-1} in terms of $\mathbf{U}$ and $\mathbf{D}$ [1 mark]

$$\Xi^{-1} = (\mathbf{U}^{\mathsf{T}})^{-1}\mathbf{D}^{-1}\mathbf{U}^{-1}$$

(d) Demonstrate by direct multiplication that

$$\mathbf{U}^{-1} = \begin{pmatrix} \mathbf{I} & -\mathbf{B}\mathbf{C}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \quad \mathbf{D}^{-1} = \begin{pmatrix} (\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^{\mathsf{T}})^{-1} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & \mathbf{C}^{-1} \end{pmatrix} \quad (\mathbf{U}^{\mathsf{T}})^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C}^{-1}\mathbf{B}^{\mathsf{T}} & \mathbf{I} \end{pmatrix}$$

i.e. show $\mathbf{U}^{-1}\mathbf{U} = \mathbf{I}$, $\mathbf{D}^{-1}\mathbf{D} = \mathbf{I}$ and $(\mathbf{U}^{\mathsf{T}})^{-1}\mathbf{U}^{\mathsf{T}} = \mathbf{I}$. [6 marks]

(i)

$$\mathbf{U}^{-1}\mathbf{U} = \begin{pmatrix} \mathbf{I} & -\mathbf{B}\mathbf{C}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{B}\mathbf{C}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

(ii)

$$\mathbf{D}^{-1}\mathbf{D} = \begin{pmatrix} (\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^{\mathsf{T}})^{-1} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & \mathbf{C}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^{\mathsf{T}} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & \mathbf{C} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

where we use $(\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^{\mathsf{T}})^{-1}(\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^{\mathsf{T}}) = \mathbf{I}$ and $\mathbf{C}^{-1}\mathbf{C} = \mathbf{I}$.

(iii)

$$(\mathbf{U}^{\mathsf{T}})^{-1}\mathbf{U}^{\mathsf{T}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C}^{-1}\mathbf{B}^{\mathsf{T}} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C}^{-1}\mathbf{B}^{\mathsf{T}} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

(e) Letting $\boldsymbol{z} = \binom{\boldsymbol{x}}{\boldsymbol{y}}$ then we can write

$$f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = f_{\boldsymbol{Z}}(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z} \mid \boldsymbol{0}, \Xi) = \frac{1}{\sqrt{\det(2\pi\Xi)}} \mathrm{e}^{-\frac{1}{2}\boldsymbol{z}^\mathsf{T}\Xi^{-1}\boldsymbol{z}}.$$

where $\det(2\pi\Xi)$ is the determinant of the matrix $2\pi\Xi$ and is introduced to ensures that $f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})$ is normalised. From parts (c) and (d)

$$\Xi^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C}^{-1}\mathbf{B}^\mathsf{T} & \mathbf{I} \end{pmatrix} \begin{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})^{-1} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & \mathbf{C}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\mathbf{B}\mathbf{C}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}.$$

Expand out $\boldsymbol{z}^\mathsf{T}\Xi^{-1}\boldsymbol{z} = (\boldsymbol{x}^\mathsf{T}, \boldsymbol{y}^\mathsf{T})\Xi^{-1}\binom{\boldsymbol{x}}{\boldsymbol{y}}$ (start by multiplying the vectors $\boldsymbol{z}^\mathsf{T}$ by $(\mathbf{U}^\mathsf{T})^{-1}$ and $\boldsymbol{z}$ by $\mathbf{U}^{-1}$) [4 marks]

$$\begin{aligned}
\boldsymbol{z}^\mathsf{T}\Xi^{-1}\boldsymbol{z} &= (\boldsymbol{x}^\mathsf{T}, \boldsymbol{y}^\mathsf{T}) \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C}^{-1}\mathbf{B}^\mathsf{T} & \mathbf{I} \end{pmatrix} \begin{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})^{-1} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & \mathbf{C}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\mathbf{B}\mathbf{C}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \\
&= (\boldsymbol{x}^\mathsf{T} - \boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T}, \boldsymbol{y}^\mathsf{T}) \begin{pmatrix} (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})^{-1} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & \mathbf{C}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} - \mathbf{B}\mathbf{C}^{-1}\boldsymbol{y} \\ \boldsymbol{y} \end{pmatrix} \\
&= (\boldsymbol{x}^\mathsf{T} - \boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})^{-1}(\boldsymbol{x} - \mathbf{B}\mathbf{C}^{-1}\boldsymbol{y}) + y^\mathsf{T}\mathbf{C}^{-1}\boldsymbol{y}
\end{aligned}$$

In the next question we use the short-hand notation

$$\int f(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} = \int \cdots \int f(\boldsymbol{z}) \mathrm{d}z_1 \mathrm{d}z_2 \ldots \mathrm{d}z_n$$

(f) To compute $f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y}) = f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})/f_{\boldsymbol{Y}}(\boldsymbol{y})$ we need to find

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \int_{-\infty}^{\infty} f(\boldsymbol{x},\boldsymbol{y}) \mathrm{d}\boldsymbol{x} \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Xi})}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}^\mathsf{T} - \boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})(\mathbf{A}-\mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})^{-1}(\boldsymbol{x}-\mathbf{B}\,\mathbf{C}^{-1}\boldsymbol{y}) - \frac{1}{2}\boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\boldsymbol{y}} \mathrm{d}\boldsymbol{x} \\ &= \frac{\mathrm{e}^{-\frac{1}{2}\boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\boldsymbol{y}}}{\sqrt{\det(2\pi\boldsymbol{\Xi})}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}^\mathsf{T} - \boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})(\mathbf{A}-\mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})^{-1}(\boldsymbol{x}-\mathbf{B}\,\mathbf{C}^{-1}\boldsymbol{y})} \mathrm{d}\boldsymbol{x} \end{aligned}$$

By making a change of variable from $\boldsymbol{x}$ to $\boldsymbol{u} = \boldsymbol{x} - \mathbf{B}\,\mathbf{C}^{-1}\boldsymbol{y}$ rewrite the integral

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}^\mathsf{T} - \boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})(\mathbf{A}-\mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})^{-1}(\boldsymbol{x}-\mathbf{B}\,\mathbf{C}^{-1}\boldsymbol{y})} \mathrm{d}\boldsymbol{x}$$

then use

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\boldsymbol{u}^\mathsf{T}\mathbf{M}^{-1}\boldsymbol{u}} \mathrm{d}\boldsymbol{u} = \sqrt{\det(2\pi\mathbf{M})}$$

to evaluate $f_{\boldsymbol{Y}}(\boldsymbol{y})$. [5 marks]

Making the change of variables $\boldsymbol{u} = \boldsymbol{x} - \mathbf{B}\,\mathbf{C}^{-1}\boldsymbol{y}$

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}\boldsymbol{u}^\mathsf{T}(\mathbf{A}-\mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})^{-1}\boldsymbol{u}} \mathrm{d}\boldsymbol{u} = \sqrt{\det(2\pi(\mathbf{A} - \mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T}))}$$

Thus

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sqrt{\frac{\det(2\pi(\mathbf{A} - \mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T}))}{\det(2\pi\boldsymbol{\Xi})}}\, \mathrm{e}^{-\frac{1}{2}\boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\boldsymbol{y}}.$$

where $\det(\boldsymbol{\Xi})$ denotes the determinant of a matrix $\boldsymbol{\Xi}$. In passing we recall that $\boldsymbol{\Xi} = \mathbf{U}^\mathsf{T}\mathbf{D}\mathbf{U}$ so $\det(\boldsymbol{\Xi}) = \det(\mathbf{U}^\mathsf{T})\det(\mathbf{D})\det(\mathbf{U})$, But it is easy to show $\det(\mathbf{U}^\mathsf{T}) = \det(\mathbf{U}) = 1$ and because $\mathbf{D}$ is block diagonal $\det(\mathbf{D}) = \det(\mathbf{A} - \mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})\det(\mathbf{C})$ so that $\det(\boldsymbol{\Xi}) = \det(\mathbf{A} - \mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})\det(\mathbf{C})$. Thus

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} \mathrm{e}^{-\frac{1}{2}\boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\boldsymbol{y}} = \mathcal{N}(\boldsymbol{y} \mid \mathbf{0}, \mathbf{C}).$$

Recall that

$$f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y}) = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \middle| \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\mathsf{T} & \mathbf{C} \end{pmatrix} \right)$$

so we see $\boldsymbol{y}$ is normally distributed with covariance $\mathbf{C}$ as might be expected.

(g) Using $f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y}) = f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})/f_{\boldsymbol{Y}}(\boldsymbol{y})$ write down $f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y})$. [3 marks]

$$
\begin{aligned}
f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \mid \boldsymbol{y}) &= \frac{f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}),\boldsymbol{y})}{f_{\boldsymbol{Y}}(\boldsymbol{y})} = \frac{\det(2\pi\mathbf{C})^{\frac{1}{2}}}{\det(2\pi\boldsymbol{\Xi})^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}^\mathsf{T}-\boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})(\mathbf{A}-\mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})^{-1}(\boldsymbol{x}-\mathbf{B}\,\mathbf{C}^{-1}\boldsymbol{y})} \\
&= \frac{1}{\det(2\pi(\mathbf{A}-\mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T}))^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}^\mathsf{T}-\boldsymbol{y}^\mathsf{T}\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})(\mathbf{A}-\mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T})^{-1}(\boldsymbol{x}-\mathbf{B}\,\mathbf{C}^{-1}\boldsymbol{y})} \\
&= \mathcal{N}(\boldsymbol{x} \mid \mathbf{B}\,\mathbf{C}^{-1}\boldsymbol{y}, \mathbf{A}-\mathbf{B}\,\mathbf{C}^{-1}\mathbf{B}^\mathsf{T}).
\end{aligned}
$$

This is the conditioning result we use in deriving the Bayesian update for a Gaussian Process.

End of question 2

END OF PAPER