Homework Assignment 2

Matthew Tiger

September 9, 2015

Problem 1. Suppose that $\boldsymbol{X_1} \sim N(\boldsymbol{\mu}, \Sigma)$. Show that $\boldsymbol{Y} = \boldsymbol{a} + B\boldsymbol{X}$ is also a multivariate normal random vector and specify the mean and covariance matrix of Y.

Solution. Note that a vector $\boldsymbol{X}$ is a multivariate normal random vector if and only if every linear combination of its components is a univariate normal random variable. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\intercal$. Then we have that

$$
\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{a} + B\boldsymbol{X} \\
&= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \\
&= \begin{pmatrix} a_1 + b_{11}X_1 + b_{12}X_2 + \cdots + b_{1n}X_n \\ a_2 + b_{21}X_1 + b_{22}X_2 + \cdots + b_{2n}X_n \\ \vdots \\ a_m + b_{m1}X_1 + b_{m2}X_2 + \cdots + b_{mn}X_n \end{pmatrix}
\end{aligned}
$$

From the above it is clear that every linear combination of the components of $\boldsymbol{Y}$ is some linear combination of $\boldsymbol{X}$. Therefore, it follows that since $\boldsymbol{X}$ is a multivariate random vector, so must $\boldsymbol{Y} = \boldsymbol{a} + B\boldsymbol{X}$.

Now all that is left is to describe the mean $\boldsymbol{\mu_Y}$ and the covariance matrix $\Sigma_{\boldsymbol{YY}}$. We begin with $\boldsymbol{\mu}_{\boldsymbol{Y}}$, where due to the linearity of the expectation operator

$$\begin{aligned}
\mathrm{E}(\boldsymbol{Y}) &= \mathrm{E}(\boldsymbol{a} + B\boldsymbol{X}) \\
&= \boldsymbol{a} + \mathrm{E}(B\boldsymbol{X}) \\
&= \boldsymbol{a} + \mathrm{E}\left(\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}\right) \\
&= \boldsymbol{a} + \mathrm{E}\left(\begin{pmatrix} b_{11}X_1 + b_{12}X_2 + \cdots + b_{1n}X_n \\ b_{21}X_1 + b_{22}X_2 + \cdots + b_{2n}X_n \\ \vdots \\ b_{m1}X_1 + b_{m2}X_2 + \cdots + b_{mn}X_n \end{pmatrix}\right) \\
&= \boldsymbol{a} + \begin{pmatrix} b_{11}\mathrm{E}(X_1) + b_{12}\mathrm{E}(X_2) + \cdots + b_{1n}\mathrm{E}(X_n) \\ b_{21}\mathrm{E}(X_1) + b_{22}\mathrm{E}(X_2) + \cdots + b_{2n}\mathrm{E}(X_n) \\ \vdots \\ b_{m1}\mathrm{E}(X_1) + b_{m2}\mathrm{E}(X_2) + \cdots + b_{mn}\mathrm{E}(X_n) \end{pmatrix} \\
&= \boldsymbol{a} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} \begin{pmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_n) \end{pmatrix} \\
&= \boldsymbol{a} + B\mathrm{E}(\boldsymbol{X}) = \boldsymbol{a} + B\boldsymbol{\mu}.
\end{aligned}$$

Now, knowing the linearity of the expectation operator for random vectors, we see that

$$\begin{aligned}
\Sigma_{\boldsymbol{YY}} &= \mathrm{E}((\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}))(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}))^\intercal) \\
&= \mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}^\intercal) - \mathrm{E}(\boldsymbol{Y})\mathrm{E}(\boldsymbol{Y})^\intercal \\
&= \mathrm{E}((\boldsymbol{a} + B\boldsymbol{X})(\boldsymbol{a} + B\boldsymbol{X})^\intercal) - (\boldsymbol{a} + B\mathrm{E}(\boldsymbol{X}))(\boldsymbol{a} + B\mathrm{E}(\boldsymbol{X}))^\intercal \\
&= \mathrm{E}((\boldsymbol{a} + B\boldsymbol{X})(\boldsymbol{a}^\intercal + \boldsymbol{X}^\intercal B^\intercal)) - (\boldsymbol{a} + B\mathrm{E}(\boldsymbol{X}))(\boldsymbol{a}^\intercal + \mathrm{E}(\boldsymbol{X})^\intercal B^\intercal) \\
&= \mathrm{E}(\boldsymbol{a}\boldsymbol{a}^\intercal + \boldsymbol{a}\boldsymbol{X}^\intercal B^\intercal + B\boldsymbol{X}\boldsymbol{a}^\intercal + B\boldsymbol{X}\boldsymbol{X}^\intercal B^\intercal) \\
&\quad - \boldsymbol{a}\boldsymbol{a}^\intercal - \boldsymbol{a}\mathrm{E}(\boldsymbol{X})^\intercal B^\intercal - B\mathrm{E}(\boldsymbol{X})\boldsymbol{a}^\intercal - B\mathrm{E}(\boldsymbol{X})\mathrm{E}(\boldsymbol{X})^\intercal B^\intercal) \\
&= \mathrm{E}(\boldsymbol{a}\boldsymbol{a}^\intercal) + \mathrm{E}(\boldsymbol{a}\boldsymbol{X}^\intercal B^\intercal) + \mathrm{E}(B\boldsymbol{X}\boldsymbol{a}^\intercal) + \mathrm{E}(B\boldsymbol{X}\boldsymbol{X}^\intercal B^\intercal) \\
&\quad - \boldsymbol{a}\boldsymbol{a}^\intercal - \boldsymbol{a}\mathrm{E}(\boldsymbol{X})^\intercal B^\intercal - B\mathrm{E}(\boldsymbol{X})\boldsymbol{a}^\intercal - B\mathrm{E}(\boldsymbol{X})\mathrm{E}(\boldsymbol{X})^\intercal B^\intercal) \\
&= \mathrm{E}(\boldsymbol{a}\boldsymbol{a}^\intercal) + \mathrm{E}(\boldsymbol{a}\boldsymbol{X}^\intercal B^\intercal) + \mathrm{E}(B\boldsymbol{X}\boldsymbol{a}^\intercal) + \mathrm{E}(B\boldsymbol{X}\boldsymbol{X}^\intercal B^\intercal) \\
&\quad - \boldsymbol{a}\boldsymbol{a}^\intercal - \boldsymbol{a}\mathrm{E}(\boldsymbol{X})^\intercal B^\intercal - B\mathrm{E}(\boldsymbol{X})\boldsymbol{a}^\intercal - B\mathrm{E}(\boldsymbol{X})\mathrm{E}(\boldsymbol{X})^\intercal B^\intercal) \\
&= \boldsymbol{a}\boldsymbol{a}^\intercal + \boldsymbol{a}\mathrm{E}(\boldsymbol{X})^\intercal B^\intercal + B\mathrm{E}(\boldsymbol{X})\boldsymbol{a}^\intercal + B\mathrm{E}(\boldsymbol{X}\boldsymbol{X}^\intercal)B^\intercal \\
&\quad - \boldsymbol{a}\boldsymbol{a}^\intercal - \boldsymbol{a}\mathrm{E}(\boldsymbol{X})^\intercal B^\intercal - B\mathrm{E}(\boldsymbol{X})\boldsymbol{a}^\intercal - B\mathrm{E}(\boldsymbol{X})\mathrm{E}(\boldsymbol{X})^\intercal B^\intercal) \\
&= B\mathrm{E}(\boldsymbol{X}\boldsymbol{X}^\intercal)B^\intercal - B\mathrm{E}(\boldsymbol{X})\mathrm{E}(\boldsymbol{X})^\intercal B^\intercal = B\Sigma_{\boldsymbol{XX}}B^\intercal.
\end{aligned}$$

□

Problem 2. Suppose that $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X_1} \\ \boldsymbol{X_2} \end{pmatrix}$ and $\boldsymbol{X_1} \sim N(\boldsymbol{\mu_1}, \Sigma_{11})$ and $\boldsymbol{X_2} \sim N(\boldsymbol{\mu_2}, \Sigma_{22})$ so that $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Find $f_{\boldsymbol{X_2}|\boldsymbol{X_1}}(\boldsymbol{x_2}|\boldsymbol{x_1})$, the conditional distribution of $\boldsymbol{X_2}$ given $\boldsymbol{X_1} = \boldsymbol{x_1}$.

Solution. Note that $f_{\boldsymbol{X_2}|\boldsymbol{X_1}}(\boldsymbol{x_2}|\boldsymbol{x_1})$, the conditional distribution of $\boldsymbol{X_2}$ given $\boldsymbol{X_1} = \boldsymbol{x_1}$, is

$$f_{\boldsymbol{X_2}|\boldsymbol{X_1}}(\boldsymbol{x_2}|\boldsymbol{x_1}) = \frac{f_{\boldsymbol{X_1},\boldsymbol{X_2}}(\boldsymbol{x_1}, \boldsymbol{x_2})}{f_{\boldsymbol{X_1}}(\boldsymbol{x_1})} \tag{1}$$

where $f_{\boldsymbol{X_1},\boldsymbol{X_2}}(\boldsymbol{x_1}, \boldsymbol{x_2})$ is the joint distribution of $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ and $f_{\boldsymbol{X_1}}(\boldsymbol{x_1})$ is the marginal distribution of $\boldsymbol{X_1}$ given by

$$f_{\boldsymbol{X_1}}(\boldsymbol{x_1}) = \int_{-\infty}^{\infty} f_{\boldsymbol{X_1},\boldsymbol{X_2}}(\boldsymbol{x_1}, \boldsymbol{x_2})\, \mathrm{d}\boldsymbol{x_2}. \tag{2}$$

Note that the joint distribution of $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ is the same as the distribution of $\boldsymbol{X}$ since $\boldsymbol{X}$ is a partition of $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$. Since we know $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$, it is clear that

$$f_{\boldsymbol{X_1},\boldsymbol{X_2}}(\boldsymbol{x_1}, \boldsymbol{x_2}) = f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\intercal \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\} \tag{3}$$

where n is the length of $\boldsymbol{X}$. Using the above partition of $\boldsymbol{X}$ and $\boldsymbol{\mu}$ as stated in the problem, we can rewrite (3) as

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-\frac{(n_1+n_2)}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \begin{pmatrix} (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal \\ (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal \end{pmatrix}^\intercal \Sigma^{-1} \begin{pmatrix} \boldsymbol{x_1} - \boldsymbol{\mu_1} \\ \boldsymbol{x_2} - \boldsymbol{\mu_2} \end{pmatrix} \right\} \tag{4}$$

where n_1 is the length of $\boldsymbol{X_1}$ and n_2 is the length of $\boldsymbol{X_2}$.

It is clear that the partitioned matrix Σ is symmetric since

$$\begin{aligned} \Sigma &= \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11}^\intercal & \Sigma_{21}^\intercal \\ \Sigma_{12}^\intercal & \Sigma_{22}^\intercal \end{pmatrix} = \Sigma^\intercal \end{aligned}$$

due to the symmetry of Σ_{11} and Σ_{22} and the fact that $\Sigma_{12}^\intercal = \Sigma_{21}$ and $\Sigma_{21}^\intercal = \Sigma_{12}$.

Using this partitioned matrix's symmetric property, we can find the determinant as such

$$\begin{aligned} |\Sigma| &= \begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^\intercal & \Sigma_{22} \end{vmatrix} \\ &= \begin{vmatrix} \Sigma_{11} & 0 \\ \Sigma_{12}^\intercal & I \end{vmatrix} \begin{vmatrix} I & \Sigma_{11}^{-1}\Sigma_{12} \\ 0 & \Sigma_{22} - \Sigma_{12}^\intercal \Sigma_{11}^{-1} \Sigma_{12} \end{vmatrix} \\ &= |\Sigma_{11}| \left| \Sigma_{22} - \Sigma_{12}^\intercal \Sigma_{11}^{-1} \Sigma_{12} \right| \end{aligned} \tag{5}$$

using the property of determinants of block matrices where one entry is 0.

Since Σ is symmetric it must also follow that Σ^{-1} is symmetric. Say $\Sigma^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$. Then the symmetric property of Σ^{-1} tells us that $B_{12}^\intercal = B_{21}$, meaning that to find Σ^{-1} we only need to find B_{11}, B_{12}, and B_{22}.

Using the formula for the inverse of a block matrix and the symmetric property of Σ, we have that

$$\begin{aligned} B_{11} &= (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^\intercal)^{-1} \\ B_{12} &= -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1} \\ B_{22} &= (\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1}. \end{aligned} \tag{6}$$

The formula $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} - DA^{-1}B)^{-1}DA^{-1}$ informs us that

$$\begin{aligned} B_{11} &= (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^\intercal)^{-1} \\ &= \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}^\intercal\Sigma_{11}^{-1}. \end{aligned} \tag{7}$$

Combining the formulas found in (6) and (7), we can see that the expression in the exponential of (4) can be simplified as

$$\begin{aligned} -\frac{1}{2}\begin{pmatrix} (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal \\ (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal \end{pmatrix}^\intercal \Sigma^{-1} \begin{pmatrix} \boldsymbol{x_1} - \boldsymbol{\mu_1} \\ \boldsymbol{x_2} - \boldsymbol{\mu_2} \end{pmatrix} &= -\frac{1}{2}\begin{pmatrix} (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal \\ (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal \end{pmatrix}^\intercal \begin{pmatrix} B_{11} & B_{12} \\ B_{12}^\intercal & B_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{x_1} - \boldsymbol{\mu_1} \\ \boldsymbol{x_2} - \boldsymbol{\mu_2} \end{pmatrix} \\ &= -\frac{1}{2}\begin{pmatrix} (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal B_{11} + (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal B_{12}^\intercal \\ (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal B_{12} + (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal B_{22} \end{pmatrix}^\intercal \begin{pmatrix} \boldsymbol{x_1} - \boldsymbol{\mu_1} \\ \boldsymbol{x_2} - \boldsymbol{\mu_2} \end{pmatrix} \\ &= -\frac{1}{2}((\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal B_{11}(\boldsymbol{x_1} - \boldsymbol{\mu_1}) + (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal B_{12}^\intercal(\boldsymbol{x_1} - \boldsymbol{\mu_1}) \\ &\quad + (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal B_{12}(\boldsymbol{x_2} - \boldsymbol{\mu_2}) + (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal B_{22}(\boldsymbol{x_2} - \boldsymbol{\mu_2})) \\ &= -\frac{1}{2}((\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal B_{11}(\boldsymbol{x_1} - \boldsymbol{\mu_1}) + 2(\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal B_{12}(\boldsymbol{x_2} - \boldsymbol{\mu_2}) \\ &\quad + (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal B_{22}(\boldsymbol{x_2} - \boldsymbol{\mu_2})) \\ &= -\frac{1}{2}G(\boldsymbol{x_1}, \boldsymbol{x_2}) \end{aligned} \tag{8}$$

where we make use of the fact $u^\intercal A v = v^\intercal A^\intercal u$ to show that

$$(\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal B_{12}^\intercal(\boldsymbol{x_1} - \boldsymbol{\mu_1}) = (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal B_{12}(\boldsymbol{x_2} - \boldsymbol{\mu_2})$$

to arrive at the above.

Substituting B_{ij} with the derivations in (6) and (7) into $G(\boldsymbol{x_1}, \boldsymbol{x_2})$ we can see that

$$\begin{aligned}
G(\boldsymbol{x_1}, \boldsymbol{x_2}) &= (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal B_{11}(\boldsymbol{x_1} - \boldsymbol{\mu_1}) + 2(\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal B_{12}(\boldsymbol{x_2} - \boldsymbol{\mu_2}) \\
&\quad + (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal B_{22}(\boldsymbol{x_2} - \boldsymbol{\mu_2}) \\
&= (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal(\Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}^\intercal\Sigma_{11}^{-1})(\boldsymbol{x_1} - \boldsymbol{\mu_1}) \\
&\quad - 2(\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal(\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1})(\boldsymbol{x_2} - \boldsymbol{\mu_2}) \\
&\quad + (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal((\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1})(\boldsymbol{x_2} - \boldsymbol{\mu_2}) \\
&= (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal\Sigma_{11}^{-1}(\boldsymbol{x_1} - \boldsymbol{\mu_1}) \\
&\quad + (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal(\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}^\intercal\Sigma_{11}^{-1})(\boldsymbol{x_1} - \boldsymbol{\mu_1}) \\
&\quad - (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal(\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1})(\boldsymbol{x_2} - \boldsymbol{\mu_2}) \\
&\quad - (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal(\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1})(\boldsymbol{x_2} - \boldsymbol{\mu_2}) \\
&\quad + (\boldsymbol{x_2} - \boldsymbol{\mu_2})^\intercal((\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1})(\boldsymbol{x_2} - \boldsymbol{\mu_2}) \\
&= (\boldsymbol{x_1} - \boldsymbol{\mu_1})^\intercal\Sigma_{11}^{-1}(\boldsymbol{x_1} - \boldsymbol{\mu_1}) \\
&\quad + (\boldsymbol{x_2} - (\boldsymbol{\mu_2} + \Sigma_{12}^\intercal\Sigma_{11}^{-1}(\boldsymbol{x_1} - \boldsymbol{\mu_1})))^\intercal(\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})^{-1}(\boldsymbol{x_2} - (\boldsymbol{\mu_2} + \Sigma_{12}^\intercal\Sigma_{11}^{-1}(\boldsymbol{x_1} - \boldsymbol{\mu_1}))) \\
&= g_1(\boldsymbol{x_1}) + g_2(\boldsymbol{x_2}) \qquad (9)
\end{aligned}$$

Now, the above and (5) show that the joint distribution (4) is

$$\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x}) &= (2\pi)^{-\frac{(n_1+n_2)}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}G(\boldsymbol{x_1}, \boldsymbol{x_2})\right\} \\
&= (2\pi)^{-\frac{n_1}{2}}(2\pi)^{-\frac{n_2}{2}} |\Sigma_{11}|^{-\frac{1}{2}} \left|\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}g(\boldsymbol{x_1})\right\} \exp\left\{-\frac{1}{2}g(\boldsymbol{x_2})\right\} \\
&= (2\pi)^{-\frac{n_1}{2}} |\Sigma_{11}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}g(\boldsymbol{x_1})\right\}(2\pi)^{-\frac{n_2}{2}} \left|\Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}g(\boldsymbol{x_2})\right\} \\
&= \mathrm{pdf}(\boldsymbol{\mu_1}, \Sigma_{11})\mathrm{pdf}(\boldsymbol{\mu_2} + \Sigma_{12}^\intercal\Sigma_{11}^{-1}(\boldsymbol{x_1} - \boldsymbol{\mu_1}), \Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12}) \qquad (10)
\end{aligned}$$

where $\mathrm{pdf}(\boldsymbol{\mu}, \Sigma)$ is the multivariate normal density function with mean $\boldsymbol{\mu}$ and covariance matrix Σ. It is clear with this definition of $f_{\boldsymbol{X_1},\boldsymbol{X_2}}(\boldsymbol{x_1}, \boldsymbol{x_2})$ that the marginal distribution of $\boldsymbol{X_1}$ is

$$\begin{aligned}
f_{\boldsymbol{X_1}}(\boldsymbol{x_1}) &= \int_{-\infty}^{\infty} f_{\boldsymbol{X_1},\boldsymbol{X_2}}(\boldsymbol{x_1}, \boldsymbol{x_2})\, \mathrm{d}\boldsymbol{x_2} \\
&= \int_{-\infty}^{\infty} \mathrm{pdf}(\boldsymbol{\mu_1}, \Sigma_{11})\mathrm{pdf}(\boldsymbol{\mu_2} + \Sigma_{12}^\intercal\Sigma_{11}^{-1}(\boldsymbol{x_1} - \boldsymbol{\mu_1}), \Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})\, \mathrm{d}\boldsymbol{x_2} \\
&= \mathrm{pdf}(\boldsymbol{\mu_1}, \Sigma_{11}) \int_{-\infty}^{\infty} \mathrm{pdf}(\boldsymbol{\mu_2} + \Sigma_{12}^\intercal\Sigma_{11}^{-1}(\boldsymbol{x_1} - \boldsymbol{\mu_1}), \Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})\, \mathrm{d}\boldsymbol{x_2} \\
&= \mathrm{pdf}(\boldsymbol{\mu_1}, \Sigma_{11}).
\end{aligned}$$

Therefore we know that the conditional distribution of $\boldsymbol{X_2}$ given $\boldsymbol{X_1} = \boldsymbol{x_1}$ is

$$\begin{aligned}
f_{\boldsymbol{X_2}|\boldsymbol{X_1}}(\boldsymbol{x_2}|\boldsymbol{x_1}) &= \frac{f_{\boldsymbol{X_1},\boldsymbol{X_2}}(\boldsymbol{x_1}, \boldsymbol{x_2})}{f_{\boldsymbol{X_1}}(\boldsymbol{x_1})} \\
&= \frac{\mathrm{pdf}(\boldsymbol{\mu_1}, \Sigma_{11})\mathrm{pdf}(\boldsymbol{\mu_2} + \Sigma_{12}^\intercal\Sigma_{11}^{-1}(\boldsymbol{x_1} - \boldsymbol{\mu_1}), \Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12})}{\mathrm{pdf}(\boldsymbol{\mu_1}, \Sigma_{11})} \\
&= \mathrm{pdf}(\boldsymbol{\mu_2} + \Sigma_{12}^\intercal\Sigma_{11}^{-1}(\boldsymbol{x_1} - \boldsymbol{\mu_1}), \Sigma_{22} - \Sigma_{12}^\intercal\Sigma_{11}^{-1}\Sigma_{12}).
\end{aligned}$$

$\square$

Problem 3. Suppose $X \sim N(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ and $X_1 \sim N(\mu_1, \sigma_1)$ and $X_2 \sim N(\mu_2, \sigma_2)$ so that $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. Show that X_1 and X_2 are independent if and only if $\rho = 0$.

Solution. Note that since X_1 and X_2 are normal random variables and $\mathrm{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2$, we have in the degenerate case that X_1 and X_2 are independent if and only if $\mathrm{Cov}(X_1, X_2) = 0$ if and only if $\rho = 0$. $\square$