4.10

Intro

In this exercise, we will provide an expression for the conditional and marginal Gaussians in information form. This exercise needs a huge amount of mathematical work, involving matrix algebra and advanced sections of this chapter. If you do not want to know the mathematical details of the solution, skip the solution and go to the conclusion to see the final expressions and the major point of the exercise.

Solution

First, let's express the full MVN in the canonical/information form:

$$p(x_1, x_2) = (2\pi)^{-\frac{D}{2}} |\Lambda|^{\frac{1}{2}} exp \left[-\frac{1}{2} \left(x^T \Lambda x + \xi^T \Lambda^{-1} \xi - 2x^T \xi \right) \right] \tag{1}$$

Now, we will derive both the marginal and the conditional ($p(x_1)$ and $p(x_2|x_1)$) from Equation 1. To be able to do this, we will use the chain rule on the LHS (left hand side) of Equation 1, resulting in $p(x_1, x_2) = p(x_1)p(x_2|x_1)$. On the RHS, we will perform a BIG expansion based on the partitions of the natural parameters Λ and ξ and the random vector x. To avoid unnecessary expressions, only the argument inside brackets will be considered during the expansion. Let's begin by expressing the argument using the required partitions.

$$\begin{aligned} &\left[\begin{array}{c} x_1 \\ \hline x_2 \end{array}\right]^T \left[\begin{array}{c|c} \Lambda_{11} & \Lambda_{12} \\ \hline \Lambda_{21} & \Lambda_{22} \end{array}\right] \left[\begin{array}{c} x_1 \\ \hline x_2 \end{array}\right] + \left[\begin{array}{c} \xi_1 \\ \hline \xi_2 \end{array}\right]^T \left[\begin{array}{c|c} \Lambda_{11} & \Lambda_{12} \\ \hline \Lambda_{21} & \Lambda_{22} \end{array}\right]^{-1} \left[\begin{array}{c} \xi_1 \\ \hline \xi_2 \end{array}\right] \\ &- 2 \left[\begin{array}{c} x_1 \\ \hline x_2 \end{array}\right]^T \left[\begin{array}{c} \xi_1 \\ \hline \xi_2 \end{array}\right] \end{aligned} \tag{2}$$

From Equation 2, we have three terms we need to expand. Will start with the first and the third, making the middle the last. The first term can be expanded as follows:

$$\left[\begin{array}{c} x_1 \\ \hline x_2 \end{array}\right]^T \left[\begin{array}{c|c} \Lambda_{11} & \Lambda_{12} \\ \hline \Lambda_{21} & \Lambda_{22} \end{array}\right] \left[\begin{array}{c} x_1 \\ \hline x_2 \end{array}\right] = x_1^T \Lambda_{11} x_1 + x_2^T \Lambda_{22} x_2 + 2x_1^T \Lambda_{12} x_2 \tag{3}$$

where, we used the fact that $\Lambda_{12} = \Lambda_{21}^T$.

The third term can expanded as follows:

$$-2 \left[\begin{array}{c} x_1 \\ \hline x_2 \end{array}\right]^T \left[\begin{array}{c} \xi_1 \\ \hline \xi_2 \end{array}\right] = -2x_1^T \xi_1 - 2x_2^T \xi_2 \tag{4}$$

Now, as the last component, we need to expand the second term. In order to do this, we will need to compute an expression for the inverse of the precision matrix Λ^{-1}, using the partitions elements Λ_{11}, Λ_{12}, Λ_{21} and Λ_{22}. This will be done using the **partitioned inversion formula**, which uses Schur complement. The equation below show our expression for the inverse:

$$
\left[\begin{array}{c|c} \Lambda_{11} & \Lambda_{12} \\ \hline \Lambda_{21} & \Lambda_{22} \end{array}\right]^{-1} =
$$

$$
\left[\begin{array}{c|c} I & 0 \\ \hline -\Lambda_{22}^{-1}\Lambda_{21} & I \end{array}\right] \left[\begin{array}{c|c} (\Lambda/\Lambda_{22})^{-1} & 0 \\ \hline 0 & \Lambda_{22}^{-1} \end{array}\right] \left[\begin{array}{c|c} I & -\Lambda_{12}\Lambda_{22}^{-1} \\ \hline 0 & I \end{array}\right] \tag{5}
$$

where $\Lambda/\Lambda_{22} = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}$ is the Schur complement of Λ wrt Λ_{22}. With an appropriate expression for the inverse precision matrix, we can expand the middle term:

$$
\begin{aligned}
&\left[\begin{array}{c} \xi_1 \\ \hline \xi_2 \end{array}\right]^T \left[\begin{array}{c|c} \Lambda_{11} & \Lambda_{12} \\ \hline \Lambda_{21} & \Lambda_{22} \end{array}\right]^{-1} \left[\begin{array}{c} \xi_1 \\ \hline \xi_2 \end{array}\right] = \\
&\left[\begin{array}{c} \xi_1 \\ \hline \xi_2 \end{array}\right]^T \left[\begin{array}{c|c} I & 0 \\ \hline -\Lambda_{22}^{-1}\Lambda_{21} & I \end{array}\right] \left[\begin{array}{c|c} (\Lambda/\Lambda_{22})^{-1} & 0 \\ \hline 0 & \Lambda_{22}^{-1} \end{array}\right] \left[\begin{array}{c|c} I & -\Lambda_{12}\Lambda_{22}^{-1} \\ \hline 0 & I \end{array}\right] \left[\begin{array}{c} \xi_1 \\ \hline \xi_2 \end{array}\right] = \\
&(\xi_1 - \Lambda_{12}\Lambda_{22}^{-1}\xi_2)^T(\Lambda/\Lambda_{22})^{-1}(\xi_1 - \Lambda_{12}\Lambda_{22}^{-1}\xi_2) + \xi_2^T\Lambda_{22}^{-1}\xi_2
\end{aligned} \tag{6}
$$

Now that we have expanded all the three terms, we need to merge them in a way that shows the two terms of the chain rule of probability. To do this, let's display all the available terms that we have at the moment in a 'bag of terms' and take them off as we use them. We can also introduce new terms in case we do something like $a = a + b - b$.

$$
\begin{aligned}
&bag = x_1^T\Lambda_{11}x_1 + x_2^T\Lambda_{22}x_2 + 2x_1^T\Lambda_{12}x_2 + \\
&(\xi_1 - \Lambda_{12}\Lambda_{22}^{-1}\xi_2)^T(\Lambda/\Lambda_{22})^{-1}(\xi_1 - \Lambda_{12}\Lambda_{22}^{-1}\xi_2) + \xi_2^T\Lambda_{22}^{-1}\xi_2 - \\
&- 2x_1^T\xi_1 - 2x_2^T\xi_2
\end{aligned} \tag{7}
$$

Let's start by finding an expression which only depends on x_1, which will correspond to $p(x_1)$. To do this, we need to take special attention to the term $(\xi_1 - \Lambda_{12}\Lambda_{22}^{-1}\xi_2)(\Lambda/\Lambda_{22})^{-1}(\xi_1 - \Lambda_{12}\Lambda_{22}^{-1}\xi_2)$. This expression has the general form $\xi^T\Lambda^{-1}\xi$, which is one of the three compoenents of the canonical expression. So, assuming that we found our new canonical parameters, we must complete the expression with the terms $x_1^T(\Lambda/\Lambda_{22})x_1$ and $-2x_1^T(\xi_1 - \Lambda_{12}\Lambda_{22}^{-1}\xi_2)$, where we choose x_1 without loss of generality. If we take out this terms of the bag of expressions (using the $a = a + b - b$ trick for non existing terms), we arrive at:

$$
\begin{aligned}
&bag = x_1^T\Lambda_{11}x_1 + x_2^T\Lambda_{22}x_2 + 2x_1^T\Lambda_{12}x_2 \\
&- x_1^T(\Lambda/\Lambda_{22})x_1 + 2x_1^T(\xi_1 - \Lambda_{12}\Lambda_{22}^{-1}\xi_2) + \xi_2^T\Lambda_{22}^{-1}\xi_2 - \\
&- 2x_1^T\xi_1 - 2x_2^T\xi_2 = \\
&x_2^T\Lambda_{22}x_2 + x_1^T(\Lambda_{11} - \Lambda/\Lambda_{22})x_1 - 2x_2^T(\xi_2 - \Lambda_{21}x_1) + \xi_2\Lambda_{22}^{-1}\xi_2 - 2(\Lambda_{21}x_1)^T\Lambda_{22}^{-1}\xi_2 = \\
&x_2^T\Lambda_{22}x_2 + (\Lambda_{21}x_1)^T\Lambda_{22}^{-1}(\Lambda_{21}x_1) - 2x_2^T(\xi_2 - \Lambda_{21}x_1) + \xi_2\Lambda_{22}^{-1}\xi_2 - 2(\Lambda_{21}x_1)^T\Lambda_{22}^{-1}\xi_2 = \\
&x_2^T\Lambda_{22}x_2 + (\xi_2 - \Lambda_{21}x_1)^T\Lambda_{22}^{-1}(\xi_2 - \Lambda_{21}x_1) - 2x_2^T(\xi_2 - \Lambda_{21}x_1)
\end{aligned} \tag{8}
$$

As we can see, the final expression above refers to $p(x_2|x_1)$.

Finally, we can express the marginal and the conditional in the canonical form:

$$\begin{aligned}
p(x_1) &= \mathcal{N}_c(\xi^{marginal}, \Lambda^{marginal}) \\
\xi^{marginal} &= \xi_1 - \Lambda_{12}\Lambda_{22}^{-1}\xi_2 \\
\Lambda^{marginal} &= \Lambda/\Lambda_{22} = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21} \\
p(x_2|x_1) &= (N)_c(\xi^{conditional}, \Lambda^{conditional}) \\
\xi^{conditional} &= \xi_2 - \Lambda_{21}x_1 \\
\Lambda^{conditional} &= \Lambda_{22}
\end{aligned} \tag{9}$$

Conclusion

In this exercise, we derived the expression for the marginal and conditional of the Gaussian using the information form. The exercise was heavy on the math, though it is not necessary to understand the final result and its implication. From the final result, we can see that the marginal expression is very big, while the conditional is very simple. This is the opposite situation of working with the moment parametrs μ and Σ, where the marginal is simple and the conditional is big.