1(c) Finally use the above results to show the following two statements

* zero entry in the i,j-th position of precision matrix Σ^{-1} implies that x_i and x_j are conditionally independent.

Ans: Without loss of generality assume $i=1$ and $j=2$

Let us partition the vector $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}$ into two parts.

$$x_{p_1} = [x_1 \ x_2]^T \quad x_{p_2} = [x_3 \ x_4 \ \cdots \ x_D]^T$$

The joint distribution

$$p(x_{p_1}, x_{p_2}) \propto \exp\left\{ -\frac{1}{2} \left[(x_{p_1}-\mu_{p_1})^T \ (x_{p_2}-\mu_{p_2})^T \right] \begin{bmatrix} \Lambda_{p_1p_1} & \Lambda_{p_1p_2} \\ \Lambda_{p_2p_1} & \Lambda_{p_2p_2} \end{bmatrix} \begin{bmatrix} x_{p_1}-\mu_{p_1} \\ x_{p_2}-\mu_{p_2} \end{bmatrix} \right\}$$

$$\propto \exp\left\{ -\frac{1}{2} \left[(x_{p_1}-\mu_{p_1})^T \Lambda_{p_1p_1} (x_{p_1}-\mu_{p_1}) + (x_{p_2}-\mu_{p_2})^T \Lambda_{p_2p_1} (x_{p_1}-\mu_{p_1}) + (x_{p_1}-\mu_{p_1})^T \Lambda_{p_1p_2} (x_{p_2}-\mu_{p_2}) + (x_{p_2}-\mu_{p_2})^T \Lambda_{p_2p_2} (x_{p_2}-\mu_{p_2}) \right] \right\}$$

Conditioning on x_{p_2} variables is equivalent to holding them constant.

We further know from the definition of conditional independence that:

$X \perp Y \mid Z$ iff there exists functions g, h such that

$$p(x,y \mid z) = g(x,z)\, h(y,z) \quad \forall\, x,y,z \text{ and } p(z) > 0$$

Thus if we can factorize

$$p(x_{p_1} \mid x_{p_2}) = g(x_1, x_{p_2})\, h(x_2, x_{p_2})$$

we will have shown the conditional independence of $x_1 \mid x_3 x_4 \dots x_D \perp x_2 \mid x_3, x_4 \dots x_D$

Starting with first term we see it factorizes.

$$(x_{p_1} - \mu_{p_1})^T \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} (x_{p_2} - \mu_p)$$

here we know $\Lambda_{12} = 0$, $\Lambda_{21} = 0$ as we know the entries in the precision matrix are zero.

$$\begin{bmatrix} (x_1 - \mu_1)^T & (x_2 - \mu_2)^T \end{bmatrix} \begin{bmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

$$\Rightarrow (x_1 - \mu_1)^T \Lambda_{11} (x_1 - \mu_1) + (x_2 - \mu_2)^T \Lambda_{22} (x_2 - \mu_2)$$

Let $g_1(x_1) = (x_1 - \mu_1)^T \Lambda_{11} (x - \mu_1)$ and $h_1(x_2) = (x_2 - \mu_2)^T \Lambda_{22} (x_2 - \mu_2)$

then the first term contains only x_1 and second term contains only x_2

The terms $(x_{p_2}-\mu_{p_2})^T \Lambda_{p_2 p_1} (x_{p_1}-\mu_{p_1})$

and $(x_{p_1}-\mu_{p_1})^T \Lambda_{p_1 p_2} (x_{p_2}-\mu_{p_2})$

are linear in x_{p_1} and can thus be written as separate additive functions of x_1, x_2

$$\Rightarrow \quad g_2(x_2, x_{p_2}) + h_2(x_2, x_{p_2})$$

The term $(x_{p_2}-\mu_{p_2})^T \Lambda_{p_2 p_2} (x_{p_2}-\mu_{p_2})$ does not contain x_{p_1} and can be considered a constat.

$$p(x_{p_1} | x_{p_2}) \propto \exp\left\{-\frac{1}{2}\left[g_1(x_1) + g_2(x_2, x_{p_2}) + h_1(x_2) + h_2(x_2, x_{p_2}) + \text{constat}\right]\right.$$

$$\propto \exp\left[\frac{-1}{2} g(x_1, x_{p_2})\right] \exp\left[\frac{-1}{2} h(x_2, x_{p_2})\right]$$

Thus the conditionon density $p(x_1, x_2 | x_3, x_4 \cdots x_p)$ factorizes as product of two functions.

$$\exp\left[\frac{-1}{2} g(x_1, x_3\, x_4 \cdots x_p)\right] \exp\left[-\frac{1}{2} h(x_2\, x_3, x_4, \cdots x_p)\right]$$

Thus $x_1 \perp x_2 \mid x_3\, x_4 \cdots x_p$ when $\Lambda_{21} = \Lambda_{12} = 0$

precision matrix entries for $(1,2)$ are zero

and this result holds for every $i \neq j$ pair by simply permuting the indices