Question 1

(a) **True.** If two sets of variables are jointly Gaussian, the conditional distribution of one set conditioned on the other is indeed Gaussian. Similarly, the marginal distribution of either set is also Gaussian.

(b) When we have a multivariate Gaussian distribution and we want to find the conditional distribution of one subset of variables given another, the formula for the conditional mean and conditional covariance are as follows:

Given

$$x = \begin{bmatrix} x_a \\ x_b \\ x_c \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_a \\ \mu_b \\ \mu_c \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} & \Sigma_{ac} \\ \Sigma_{ba} & \Sigma_{bb} & \Sigma_{bc} \\ \Sigma_{ca} & \Sigma_{cb} & \Sigma_{cc} \end{bmatrix}$$

The conditional distribution $p(x_a|x_b)$ is Gaussian with:

Mean:

$$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$$

Covariance:

$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$$

So the conditional distribution $p(x_a|x_b)$ is:

$$p(x_a|x_b) \sim \mathcal{N}(\mu_{a|b}, \Sigma_{a|b})$$

Where $\mathcal{N}(\mu, \Sigma)$ denotes the Gaussian distribution with mean μ and covariance matrix Σ.

Question 2

(a) Marginal Distribution $p(x)$

From the given joint distribution's mean and covariance matrix, we can observe:

$$\mathrm{cov}[\mathbf{z}] = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}^{-1} & \boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \\ \mathbf{A}\boldsymbol{\Lambda}^{-1} & \mathbf{L}^{-1} + \mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \end{pmatrix}$$

$$\mathbb{E}[\mathbf{x}] = \mu$$
$$\mathrm{cov}[\mathbf{x}] = \boldsymbol{\Lambda}^{-1}$$

This is the very form of $p(x)$. Hence the marginal distribution $p(x)$ is

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu, \boldsymbol{\Lambda}^{-1})$$

(b)

For the conditional distribution of a multivariate Gaussian, we can utilize the following expressions:

$$\begin{aligned}\mu_{y|x} &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(\mathbf{x}-\mu_x) \\ &= \mathbf{A}\mu + \mathbf{b} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{\Lambda}(\mathbf{x}-\mu) \\ &= \mathbf{A}\mathbf{x}+\mathbf{b}\end{aligned}$$

$$\begin{aligned}\Sigma_{y|x} &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \\ &= \mathbf{L}^{-1} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} - \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{\Lambda}\mathbf{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}} \\ &= \mathbf{L}^{-1}\end{aligned}$$

Hence, the conditional distribution $p(\mathbf{y}|\mathbf{x})$ is

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}+\mathbf{b}, \mathbf{L}^{-1})$$

Question 3

maximizes the log likelihood

We have

$$\frac{\partial}{\partial \mathbf{A}}\ln|\mathbf{A}| = (\mathbf{A}^{-1})^{\mathrm{T}}$$

$$\frac{\partial \mathrm{tr}(A\Sigma^{-1})}{\partial \Sigma} = -\Sigma^{-1}A\Sigma^{-1}$$

$$\mathbf{z}^{\mathrm{T}}\Sigma^{-1}\mathbf{z} = \mathrm{tr}(\mathbf{z}^{\mathrm{T}}\Sigma^{-1}\mathbf{z}) = \mathrm{tr}(\Sigma^{-1}\mathbf{z}\mathbf{z}^{\mathrm{T}})$$

$$\frac{\partial \mathrm{tr}(\Sigma^{-1}\mathbf{z}\mathbf{z}^{\mathrm{T}})}{\partial \Sigma} = -\Sigma^{-1}\mathbf{z}\mathbf{z}^{\mathrm{T}}\Sigma^{-1}$$

Then

$$\begin{aligned}\frac{\partial \ln(\mathbf{X}|\mu,\Sigma)}{\partial \Sigma} &= -\frac{N}{2}\Sigma^{-T} + \frac{1}{2}\sum_{n=1}^{N}\Sigma^{-1}(\mathbf{x}_n-\mu)(\mathbf{x}_n-\mu)^T\Sigma^{-1} \\ &= -\frac{N}{2}\Sigma^{-1} + \frac{1}{2}\sum_{n=1}^{N}\Sigma^{-1}(\mathbf{x}_n-\mu)(\mathbf{x}_n-\mu)^T\Sigma^{-1} \\ &= 0\end{aligned}$$

It is solved that

$$\Sigma_{ML} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n-\mu)(\mathbf{x}_n-\mu)^T$$

Symmetric

The transpose of Σ_{ML} is:

$$\Sigma_{ML}^{T} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n-\mu)^T(\mathbf{x}_n-\mu)$$

Because $(\mathbf{x}_n-\mu)^T(\mathbf{x}_n-\mu)$ is scalar, the transpose is itself. Thus we have

$$\Sigma_{ML} = \Sigma_{ML}^{T}$$

Positive definition

If the sample covariance is nonsingular, then it is invertible, which implies it is positive definite.

Question 4 (a)

$$
\begin{aligned}
(\sigma_{ML}^2)^{(N)} &= \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2 \\
&= \frac{1}{N}(x_N-\mu)^2 + \frac{1}{N}\sum_{n=1}^{N-1}(x_n-\mu)^2 \\
&= \frac{1}{N}(x_N-\mu)^2 + \frac{N-1}{N}(\sigma_{ML}^2)^{(N-1)} \\
&= (\sigma_{ML}^2)^{(N-1)} + \frac{1}{N}\left((x_N-\mu)^2 - (\sigma_{ML}^2)^{(N-1)}\right)
\end{aligned}
$$

$$
\theta^{(N)} = \theta^{(N-1)} - \alpha_{N-1}\frac{\partial}{\partial\theta^{(N-1)}}[-\mathrm{ln}p(x_N|\theta^{(N-1)})]
$$

$$
\begin{aligned}
(\sigma_{ML}^2)^{(N)} &= (\sigma_{ML}^2)^{(N-1)} - \alpha_{N-1}\frac{\partial}{\partial(\sigma_{ML}^2)^{(N-1)}}\left[-\mathrm{ln}p(x_N|(\sigma_{ML}^2)^{(N-1)})\right] \\
&= (\sigma_{ML}^2)^{(N-1)} - \alpha_{N-1}\frac{\partial}{\partial(\sigma_{ML}^2)^{(N-1)}}\left[\frac{1}{2}\mathrm{ln}(2\pi) + \frac{1}{2}\mathrm{ln}(\sigma_{ML}^2)^{(N-1)} + \frac{(x_N-\mu)^2}{2(\sigma_{ML}^2)^{(N-1)}}\right] \\
&= (\sigma_{ML}^2)^{(N-1)} - \alpha_{N-1}\left[\frac{1}{2(\sigma_{ML}^2)^{(N-1)}} - \frac{(x_N-\mu)^2}{2(\sigma_{ML}^4)^{(N-1)}}\right] \\
&= (\sigma_{ML}^2)^{(N-1)} + \frac{\alpha_{N-1}}{2(\sigma_{ML}^4)^{(N-1)}}\left[(x_N-\mu)^2 - (\sigma_{ML}^2)^{(N-1)}\right]
\end{aligned}
$$

Hence, take

$$
\alpha_N = \frac{2(\sigma_{ML}^4)^{(N)}}{N+1}
$$

to match

$$
(\sigma_{ML}^2)^{(N-1)} + \frac{1}{N}\left((x_N-\mu)^2 - (\sigma_{ML}^2)^{(N-1)}\right)
$$

Question 4 (b)

$$
\begin{aligned}
\Sigma_{\mathrm{ML}}^{(N)} &= \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n-\mu_{\mathrm{ML}})(\mathbf{x}_n-\mu_{\mathrm{ML}})^{\mathrm{T}} \\
&= \Sigma_{ML}^{(N-1)} + \frac{1}{N}((\mathbf{x}_N-\mu_{\mathrm{ML}})(\mathbf{x}_N-\mu_{\mathrm{ML}})^{\mathrm{T}} - \Sigma_{ML}^{(N-1)})
\end{aligned}
$$

$$
\theta^{(N)} = \theta^{(N-1)} - \alpha_{N-1}\frac{\partial}{\partial\theta^{(N-1)}}[-\mathrm{ln}p(x_N|\theta^{(N-1)})]
$$

$$
\begin{aligned}
\Sigma_{\mathrm{ML}}^{(N)} &= \Sigma_{\mathrm{ML}}^{(N-1)} - \alpha_{N-1}\frac{\partial}{\partial \Sigma_{\mathrm{ML}}^{(N-1)}}[-\ln p(\mathbf{x}_N|\Sigma_{\mathrm{ML}}^{(N-1)})] \\
&= \Sigma_{\mathrm{ML}}^{(N-1)} - \alpha_{N-1}\frac{\partial}{\partial \Sigma_{\mathrm{ML}}^{(N-1)}}[\frac{ND}{2}\ln(2\pi) + \frac{N}{2}\ln|\Sigma_{\mathrm{ML}}^{(N-1)}| + \frac{1}{2}(\mathbf{x}_N - \mu_{\mathrm{ML}})^{\mathrm{T}}(\Sigma_{\mathrm{ML}}^{(N-1)})^{-1}(\mathbf{x}_N - \mu_{\mathrm{ML}})] \\
&= \Sigma_{\mathrm{ML}}^{(N-1)} - \alpha_{N-1}[\frac{N}{2}((\Sigma_{\mathrm{ML}}^{(N-1)})^{-1})^{\mathrm{T}} - \frac{1}{2}(\Sigma_{\mathrm{ML}}^{(N-1)})^{-1}(\mathbf{x}_N - \mu_{\mathrm{ML}})(\mathbf{x}_N - \mu_{\mathrm{ML}})^{\mathrm{T}}(\Sigma_{\mathrm{ML}}^{(N-1)})^{-1}] \\
&= \Sigma_{\mathrm{ML}}^{(N-1)} - \alpha_{N-1}[\frac{N}{2}((\Sigma_{\mathrm{ML}}^{(N-1)})^{-1}) - \frac{1}{2}(\Sigma_{\mathrm{ML}}^{(N-1)})^{-1}(\mathbf{x}_N - \mu_{\mathrm{ML}})(\mathbf{x}_N - \mu_{\mathrm{ML}})^{\mathrm{T}}(\Sigma_{\mathrm{ML}}^{(N-1)})^{-1}] \\
&= \Sigma_{\mathrm{ML}}^{(N-1)} + \alpha_{N-1}\frac{N}{2}(\Sigma_{\mathrm{ML}}^{(N-1)})^{-1}[\frac{1}{N}((\mathbf{x}_N - \mu_{\mathrm{ML}})(\mathbf{x}_N - \mu_{\mathrm{ML}})^{\mathrm{T}} - \Sigma_{ML}^{(N-1)})](\Sigma_{\mathrm{ML}}^{(N-1)})^{-1}
\end{aligned}
$$

Hence, take

$$
\alpha_N = \frac{2\Sigma_{ML}^{2(N)}}{N+1}
$$

to match.

Question 5

Assume

$$
p(\mu|\mathbf{X}) = N(\mu|\mu_N, \Sigma_N)
$$

By Bayesian Inference,

$$
p(\mu|\mathbf{X}) \propto p(\mathbf{X}|\mu)p(\mu)
$$

$$
-\frac{1}{2}(\mu-\mu_N)^T\Sigma_N^{-1}(\mu-\mu_N) =
$$

$$
\sum_{n=1}^{N} -\frac{1}{2}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu) - \frac{1}{2}(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)
$$

Then

$$
\begin{aligned}
&\Sigma_N = N\Sigma^{-1} + \Sigma_0^{-1} \\
&\mu_N = \Sigma_N(\Sigma^{-1}\sum_{n=1}^{N}\mathbf{x_n} + \Sigma_0^{-1}\mu_0) = (N\Sigma^{-1} + \Sigma_0^{-1})(\Sigma^{-1}\sum_{n=1}^{N}\mathbf{x_n} + \Sigma_0^{-1}\mu_0)
\end{aligned}
$$