Assignment 02

Question 1

(a)

True.

(b)

First, we need to marginalize x_c . According to the marginal Gaussian distribution, $p(x_a, x_b)$ is a Gaussian distribution with arguments as follows:

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$$

Then, according to the conditional Gaussian distributions, $p(x_a|x_b)$ is a Gaussian distribution with arguments as follows:

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

Question 2

(a)

According to the marginal Gaussian distribution, $p(\mathbf{x})$ is a Gaussian distribution with arguments as follows:

$$\mu_{\mathbf{x}} = \mu, \quad \Sigma_{\mathbf{x}} = \Sigma_{\mathbf{xx}} = \Lambda^{-1}$$

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

(b)

According to the conditional Gaussian distribution, $p(\mathbf{y}|\mathbf{x})$ is a Gaussian distribution with arguments as follows:

$$\mu_{\mathbf{y}|\mathbf{x}} = \mu_{\mathbf{y}} + \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{yy}}^{-1}(\mathbf{x} - \mu) = \mathbf{A}\mu{+}\mathbf{b} + \mathbf{A}\Lambda^{-1}\Lambda(\mathbf{x} - \mu) = \mathbf{Ax} + \mathbf{b}$$

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

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

Question 3

For the term $(-\frac{N}{2}\ln|\Sigma|)$:

$$\frac{\partial}{\partial\Sigma}(-\frac{\partial}{\partial\Sigma}\frac{N}{2}\ln|\Sigma|) = -\frac{N}{2}(\Sigma^{-1})^{\mathrm{T}} = -\frac{N}{2}\Sigma^{-1}$$

For the term $(-\frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_n - \mu))$:

$$\text{take S} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n-\mu)(\mathbf{x}_n-\mu)^{\mathrm{T}}$$

$$\text{thus we have } \sum_{n=1}^{N}(\mathbf{x}_n-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_n-\mu) = N\mathrm{Tr}[\Sigma^{-1}\mathrm{S}]$$

$$\begin{aligned}
\frac{\partial}{\partial\Sigma_{ij}}N\mathrm{Tr}[\Sigma^{-1}\mathrm{S}] &= N\mathrm{Tr}[\frac{\partial}{\partial\Sigma_{ij}}\Sigma^{-1}\mathrm{S}] \\
&= N\mathrm{Tr}[\frac{\partial}{\partial\Sigma_{ij}}\Sigma^{-1}\mathrm{S}] \\
&= -N\mathrm{Tr}[\Sigma^{-1}\frac{\partial\Sigma}{\partial\Sigma_{ij}}\Sigma^{-1}\mathrm{S}] \\
&= -N\mathrm{Tr}[\frac{\partial\Sigma}{\partial\Sigma_{ij}}\Sigma^{-1}\mathrm{S}\Sigma^{-1}]
\end{aligned}$$

$$\text{So that:}$$
$$\frac{\partial}{\partial\Sigma_{ij}}N\mathrm{Tr}[\Sigma^{-1}\mathrm{S}] = -N(\Sigma^{-1}\mathrm{S}\Sigma^{-1})_{ij}$$

So that we have:

$$\frac{\partial}{\partial\Sigma}(-\frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}_n-\mu)) = \frac{N}{2}(\Sigma^{-1}\mathrm{S}\Sigma^{-1})$$

Summing up the two terms we have:

$$\begin{aligned}
\frac{\partial}{\partial\Sigma}\ln p(\mathbf{X}|\mu,\Sigma) = -\frac{N}{2}\Sigma^{-1} + \frac{N}{2}(\Sigma^{-1}\mathrm{S}\Sigma^{-1}) = 0 \\
\frac{N}{2}(\Sigma^{-1}\mathrm{S}\Sigma^{-1}) = \frac{N}{2}\Sigma^{-1} \\
\Sigma = S
\end{aligned}$$

To prove that the result is symmetric:

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

To prove that the result is positive definite:

$$\text{for } \forall\mathbf{x}\neq 0:$$
$$\mathbf{x}^{\mathrm{T}}\Sigma\mathbf{x} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}^{\mathrm{T}}(\mathbf{x}_n-\mu)(\mathbf{x}_n-\mu)^{\mathrm{T}}\mathbf{x} = \frac{1}{N}\sum_{n=1}^{N}((\mathbf{x}_n-\mu)^{\mathrm{T}}\mathbf{x})^{T}((\mathbf{x}_n-\mu)^{\mathrm{T}}\mathbf{x}) > 0$$

Question 4

(a)

From the side of σ_N^2:

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

From the side of Robbins-Monro:

$$
\begin{aligned}
\sigma_N^2 = \sigma_{N-1}^2 + a_{N-1}\frac{\partial}{\partial\sigma_{N-1}^2}(-\frac{\ln\sigma_{N-1}^2}{2}-\frac{(x_N-\mu)^2}{2\sigma_{N-1}^2}) \\
= \sigma_{N-1}^2 + a_{N-1}(-\frac{1}{2\sigma_{N-1}^2}+\frac{(x_N-\mu)^2}{2\sigma_{N-1}^4}) \\
= \sigma_{N-1}^2 + \frac{a_{N-1}}{2\sigma_{N-1}^4}((x_N-\mu)^2-\sigma_{N-1}^2)
\end{aligned}
$$

Comparing the two formulas we can get know that:

$$
\frac{a_{N-1}}{2\sigma_{N-1}^4}=\frac{1}{N}
$$

$$
a_{N-1}=\frac{2\sigma_{N-1}^4}{N}
$$

(b)

From the side of Σ_N:

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

From the side of Robbins-Monro:

$$
\Sigma_N=\Sigma_{N-1}+a_{N-1}\frac{\partial}{\partial\Sigma_{N-1}}(\ln p(\mathbf{x}_N|\mu,\Sigma_{N-1}))
$$

$$
\text{from Question3 we have:}
$$

$$
\frac{\partial}{\partial\Sigma}\ln p(\mathbf{x}|\mu,\Sigma)=-\frac{1}{2}\Sigma^{-1}+\frac{1}{2}(\Sigma^{-1}\mathrm{S}\Sigma^{-1})=\frac{1}{2}\Sigma^{-1}(S-\Sigma)\Sigma^{-1}
$$

$$
\text{So we get:}
$$

$$
\Sigma_N=\Sigma_{N-1}+\frac{a_{N-1}}{2}\Sigma_{N-1}^{-1}((\mathbf{x}_N-\mu)(\mathbf{x}_N-\mu)^T-\Sigma_{N-1})\Sigma_{N-1}^{-1}
$$

$$
\text{As } \Sigma_{N-1}^{-1} \text{ is diagnal, we can go further and get:}
$$

$$
\Sigma_N=\Sigma_{N-1}+a_{N-1}\frac{1}{2}\Sigma_{N-1}^{-2}((\mathbf{x}_N-\mu)(\mathbf{x}_N-\mu)^T-\Sigma_{N-1})
$$

Comparing the two formulas we can get know that:

$$
a_{N-1}=\frac{2}{N}\Sigma_{N-1}^{-2}
$$

Question 5

The posterior distribution should be in the form as follow:

$$p(\mu|\mathbf{X}) \propto p(\mu)\prod_{n=1}^{N} p(x_n|\mu, \sigma^2)$$

As we have:

$$p(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} exp\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\}$$

$$\prod_{n=1}^{N} p(x_n|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} exp\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\}$$

We focus on the exponential part of $p(\mu|\mathbf{X})$:

$$-\frac{(\mu-\mu_0)^2}{2\sigma_0^2} - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 = -\frac{1}{2}(\frac{1}{\sigma_0^2}+\frac{N}{\sigma^2})\mu^2 + (\frac{\mu_0}{\sigma_0^2}+\frac{1}{\sigma_0^2}\sum_{n=1}^{N}x_n)\mu + C$$

C is independent from μ, so we do not have to care about it.

From the coefficient of μ^2, we can get know that:

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$$

From the coefficient of μ, we can get know that:

$$\mu_N = \sigma_N^2(\frac{\mu_0}{\sigma_0^2}+\frac{1}{\sigma_0^2}\sum_{n=1}^{N}x_n) = \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{ML}$$

$$\text{where } \mu_{ML} = \frac{1}{N}\sum_{n=1}^{N}x_n$$