2.b) $L(\theta) = \sum_{i=1}^{I} \mathbb{E}\left[-\log|\tilde{\Sigma}| - (x_i - \tilde{\mu} - \tilde{\Phi} h_i)^T \tilde{\Sigma}^{-1} (x_i - \tilde{\mu} - \tilde{\Phi} h_i)\right]$

$$= \sum_{i=1}^{I} \mathbb{E}\Big[-\log|\tilde{\Sigma}| - x_i^T \tilde{\Sigma}^{-1} x_i - \tilde{\mu}^T \tilde{\Sigma}^{-1}\tilde{\mu} - (\tilde{\Phi} h_i)^T \tilde{\Sigma}^{-1} \tilde{\Phi} h_i + x_i^T \tilde{\Sigma}^{-1}\tilde{\mu} + \tilde{\mu}^T \tilde{\Sigma}^{-1} x_i - (\tilde{\Phi} h_i)^T \tilde{\Sigma}^{-1} \tilde{\mu} - \tilde{\mu}^T \tilde{\Sigma}^{-1} \tilde{\Phi} h_i + x_i^T \tilde{\Sigma}^{-1} \tilde{\Phi} h_i + (\tilde{\Phi} h_i)^T \tilde{\Sigma}^{-1} x_i\Big]$$

$$= \sum_{i=1}^{I} \mathbb{E}\Big[-\log|\tilde{\Sigma}| - x_i^T \tilde{\Sigma}^{-1} x_i - \tilde{\mu}^T \tilde{\Sigma}^{-1}\tilde{\mu} - h_i^T \tilde{\Phi}^T \tilde{\Sigma}^{-1} \tilde{\Phi} h_i + 2\tilde{\mu}^T \tilde{\Sigma}^{-1} x_i - 2\tilde{\mu}^T \tilde{\Sigma}^{-1} \tilde{\Phi} h_i + 2 x_i^T \tilde{\Sigma}^{-1} \tilde{\Phi} h_i\Big]$$

$$= \sum_{i=1}^{I} \Big\{-\log|\tilde{\Sigma}| - x_i^T \tilde{\Sigma}^{-1} x_i - \tilde{\mu}^T \tilde{\Sigma}^{-1}\tilde{\mu} - \mathbb{E}[h_i^T \tilde{\Phi}^T \tilde{\Sigma}^{-1} \tilde{\Phi} h_i] + 2\tilde{\mu}^T \tilde{\Sigma}^{-1} x_i - 2\tilde{\mu}^T \tilde{\Sigma}^{-1} \tilde{\Phi}\, \mathbb{E}(h_i) + 2 x_i^T \tilde{\Sigma}^{-1} \tilde{\Phi}\, \mathbb{E}(h_i)\Big\}$$

To get $\tilde{\theta} = \underset{\tilde{\theta}}{\operatorname{argmax}}\ L(\theta)$ we find $\frac{\partial L}{\partial \tilde{\mu}}$, $\frac{\partial L}{\partial \tilde{\Sigma}^{-1}}$ and $\frac{\partial L}{\partial \tilde{\Phi}}$ and set them to zero to estimate the parameters.

$$\therefore \frac{\partial L}{\partial \tilde{\mu}} = \sum_{i=1}^{I}\left[-2\tilde{\Sigma}^{-1}\tilde{\mu} + 2\tilde{\Sigma}^{-1} x_i - 2\tilde{\Sigma}^{-1}\tilde{\Phi}\,\mathbb{E}(h_i)\right] \stackrel{!}{=} 0 \quad \text{— (1)}$$

$$\frac{\partial L}{\partial \tilde{\Sigma}^{-1}} = \sum_{i=1}^{I} \operatorname{diag}\left[\tilde{\Sigma} - x_i x_i^T - \tilde{\mu}\tilde{\mu}^T - \tilde{\Phi}\, E[h_i h_i^T]\, \tilde{\Phi}^T + 2 x_i \tilde{\mu}^T - 2\mu\, E(h_i)^T \tilde{\Phi}^T + 2 x_i\, E(h_i)^T \tilde{\Phi}^T\right]$$

$$\stackrel{!}{=} 0 \quad \text{— (2).}$$

$$\frac{\partial L}{\partial \tilde{\Phi}} = \sum_{i=1}^{I}\left\{-2\tilde{\Sigma}^{-1}\tilde{\Phi}\,\mathbb{E}(h_i h_i^T) - 2\tilde{\Sigma}^{-1}\tilde{\mu}\, E(h_i)^T + 2\tilde{\Sigma}^{-1} x_i\, E(h_i)^T\right\} \stackrel{!}{=} 0 \quad \text{— (3).}$$

Solving (1) we get,

$$-\tilde{\mu} I + \sum_{i=1}^{I} x_i - \tilde{\Phi}\sum_{i=1}^{I} E(h_i) = 0$$

$$\Rightarrow \mu = \frac{1}{I}\sum_{i=1}^{I}\left(x_i - \tilde{\Phi}\, E(h_i)\right) \quad \text{— (4)}$$

Solving ③,

$$\tilde{\Phi}\left\{\sum_{i=1}^{I} \mathbb{E}(h_i h_i^T)\right\} - \sum_{i=1}^{I} (x_i-\tilde{\mu})\, E(h_i)^T = 0.$$

$$\Rightarrow \tilde{\Phi} = \left(\sum_{i=1}^{I} (x_i-\tilde{\mu})\, E(h_i)^T\right)\left(\sum_{i=1}^{I} \mathbb{E}(h_i h_i^T)\right)^{-1} \quad \text{— ⑤}$$

From ②,

$$I.\tilde{\Sigma} = \sum_{i=1}^{I} \operatorname{diag}\left[(x_i-\mu)(x_i-\mu)^T + \tilde{\Phi}\, E(h_i h_i^T)\, \tilde{\Phi}^T - 2(x_i-\mu)\, E(h_i)^T \tilde{\Phi}^T\right]$$

$$= \operatorname{diag}\left[\sum_{i=1}^{I}(x_i-\mu)(x_i-\mu)^T + \tilde{\Phi}\left(\sum_{i=1}^{I}\mathbb{E}(h_i h_i^T)\right)\tilde{\Phi}^T - 2\left(\sum_{i=1}^{I}(x_i-\mu)\, \mathbb{E}(h_i)^T\right)\tilde{\Phi}^T\right]$$

$$= \operatorname{diag}\left[\sum_{i=1}^{I}(x_i-\mu)(x_i-\mu)^T + \left(\sum_{i=1}^{I}(x_i-\mu)\, E(h_i)^T\right)\left(\sum_{i=1}^{I} E(h_i h_i^T)\right)^{-1}\sum_{i=1}^{I} E(h_i h_i^T)\left(\left(\sum_{i=1}^{I}(x_i-\mu)(E(h_i)^T\right)\left(\sum_{i=1}^{I} E(h_i h_i^T)\right)^{-1}\right)^T - 2\left(\sum_{i=1}^{I}(x_i-\mu)\, E(h_i)^T\right)\tilde{\Phi}^T\right]$$

$$= \operatorname{diag}\left[\sum_{i=1}^{I}(x_i-\mu)(x_i-\mu)^T + \left(\sum_{i=1}^{I}(x_i-\mu)\, E(h_i)^T\right)\tilde{\Phi}^T - 2\left(\sum_{i=1}^{I}\left((x_i-\mu)\, E(h_i)^T\right)\tilde{\Phi}^T\right]$$

$$= \operatorname{diag}\sum_{i=1}^{I}\left[(x_i-\mu)(x_i-\mu)^T - \tilde{\Phi}\, E(h_i)\,(x_i-\mu)^T\right]$$

$$= \sum_{i=1}^{I}\operatorname{diag}\left[(x_i-\mu)(x_i-\mu)^T - \tilde{\Phi}\, E(h_i)(x_i-\mu)^T\right]$$

$$\therefore \tilde{\Sigma} = \frac{1}{I}\sum_{i=1}^{I}\left[\operatorname{diag}(x_i-\mu)(x_i-\mu)^T - \tilde{\Phi}\, E(h_i)(x_i-\mu)^T\right] \quad \text{— ⑥}$$

2.c) $\mu = \mu^{(0)} = \frac{1}{I}\sum_{i=1}^{I} x_i$

Now, $\tilde{\mu} = \frac{1}{I}\sum_{i=1}^{I}\left(x_i - \tilde{\Phi} E(h_i)\right)$

and $E(h_i) = (\Phi^T \Sigma^{-1} \Phi + I)^{-1} \Phi^T \Sigma^{-1}(x_i - \mu^{(0)})$

$= (\Phi \Sigma^{-1} \Phi + I)^{-1} \Phi^T \Sigma^{-1}\left(x_i - \frac{1}{I}\sum_{j=1}^{I} x_j\right)$ [change of index to differentiate]

$\therefore \sum_{i=1}^{I} E(h_i) = (\Phi \Sigma^{-1} \Phi + I)^{-1} \Phi^{-1} \Sigma^{-1}\left(\sum_{i=1}^{I} x_i - \frac{1}{I}\sum_{i=1}^{I}\sum_{j=1}^{I} x_j\right)$

Now, $\frac{1}{I}\sum_{i=1}^{I}\left(\sum_{j=1}^{I} x_j\right) = \frac{1}{I}. I. \sum_{j=1}^{I} x_j = \sum_{j=1}^{I} x_j$

$\Rightarrow \sum_{i=1}^{I} E(h_i) = (\Phi \Sigma^{-1} \Phi + I)^{-1} \Phi^{-1} \Sigma^{-1} \underbrace{\left(\sum_{i=1}^{I} x_i - \sum_{j=1}^{I} x_j\right)}_{= 0}$

$= 0.$

$\therefore \tilde{\mu} = \frac{1}{I}\sum_{i=1}^{I} x_i - \Phi . 0. = \frac{1}{I}\sum_{i=1}^{I} x_i = \mu^{(0)}.$

So, $\tilde{\mu}$ does not change and retain the initial value.

2. d) If $\mu = \mu^{(0)}$ and $\Phi^{(0)} = 0$ then $E(h_i) = 0$ [as derived in 2.c)]. and $\tilde{\Phi} = 0$.

$\tilde{\Sigma}$ is aloso a constant $\Sigma^{(0)}$.

So, $\tilde{\mu}$ and $\tilde{\Sigma}$ never change and we cannot get an optimal estimate of the parameters needed. ~~to perform the maximi~~ Hence, the maximization step essentially does not contribute.

So the initialization is not beneficial at all as it is in conflict with our basic model.