4.10 The likelihood function is given by

$$p(\mathbf{t}|\pi_1, \pi_2, \ldots, \pi_K, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}) = \prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k p(\boldsymbol{\phi}_n|C_k))^{t_{nk}}$$

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} (\pi_k \mathcal{N}(\boldsymbol{\phi}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}))^{t_{nk}}$$

The log-likelihood function is given by

$$\sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln(\pi_k \mathcal{N}(\boldsymbol{\phi}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}))$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln(\pi_k) + t_{nk} \ln(\mathcal{N}(\boldsymbol{\phi}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}))$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln(\pi_k) + \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln(\mathcal{N}(\boldsymbol{\phi}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}))$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln(\pi_k) + \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \left(-\frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln(|\boldsymbol{\Sigma}|) - \frac{1}{2} (\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\phi}_n - \boldsymbol{\mu}_k) \right)$$

To maximize the log-likelihood w.r.t $\boldsymbol{\mu}_k$, we take the derivative w.r.t $\boldsymbol{\mu}_k$ and set it to 0:

$$\implies \frac{d}{d\boldsymbol{\mu}_k} p(\mathbf{t}|\pi_1, \pi_2, \ldots, \pi_K, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_k, \boldsymbol{\Sigma})$$

$$= 0 + \sum_{n=1}^{N} t_{nk} \left(-\frac{1}{2} \right) \left(-2\boldsymbol{\Sigma}^{-1} (\boldsymbol{\phi}_n - \boldsymbol{\mu}_k) \right) \quad \text{Using Eqn. 86 of matrix cookbook}$$

$$\implies \mathbf{0} = \sum_{n=1}^{N} t_{nk} \left(\boldsymbol{\Sigma}^{-1} (\boldsymbol{\phi}_n - \boldsymbol{\mu}_k) \right)$$

$$\implies \mathbf{0} = \boldsymbol{\Sigma}^{-1} \sum_{n=1}^{N} t_{nk} (\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)$$

$$\implies \mathbf{0} = \sum_{n=1}^{N} t_{nk} (\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)$$

$$\implies \sum_{n=1}^{N} t_{nk}\boldsymbol{\phi}_n = \sum_{n=1}^{N} t_{nk}\boldsymbol{\mu}_k$$

$$\implies \sum_{n=1}^{N} t_{nk}\boldsymbol{\phi}_n = \boldsymbol{\mu}_k \sum_{n=1}^{N} t_{nk}$$

$$\implies \sum_{n=1}^{N} t_{nk}\boldsymbol{\phi}_n = \boldsymbol{\mu}_k N_k$$

$$\implies \boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} t_{nk}\boldsymbol{\phi}_n.$$

To maximize the log-likelihood w.r.t $\boldsymbol{\Sigma}$, we take the derivative w.r.t $\boldsymbol{\Sigma}$ and set it to 0:

$$\implies \frac{d}{d\boldsymbol{\Sigma}} p(\mathbf{t}|\pi_1, \pi_2, \ldots, \pi_K, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_k, \boldsymbol{\Sigma})$$

$$= 0 + \frac{d}{d\boldsymbol{\Sigma}} \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \left(-\frac{D}{2}\ln(2\pi) - \frac{1}{2}\ln(|\boldsymbol{\Sigma}|) - \frac{1}{2}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\phi}_n - \boldsymbol{\mu}_k) \right)$$

Using equations 49 and 61 of Matrix cookbook, we get:

$$\implies \mathbf{0} = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \left(-\frac{1}{2}(\boldsymbol{\Sigma})^{-T} - \frac{1}{2}(-\boldsymbol{\Sigma}^{-T})(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-T} \right)$$

$$\implies \mathbf{0} = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1} - \frac{1}{2}(-\boldsymbol{\Sigma}^{-1})(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} \right)$$

$$\implies \mathbf{0} = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1} + \frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} \right)$$

$$\implies \mathbf{0} = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \left(-\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} \right)$$

$$\implies \mathbf{0} = \left(\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} \left(-\mathbf{I} + \boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T \right) \right) \boldsymbol{\Sigma}^{-1}$$

$$\implies \mathbf{0} = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\left(-\mathbf{I} + \boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T\right)$$

$$\implies \mathbf{0} = \sum_{n=1}^{N}\sum_{k=1}^{K} -t_{nk}\mathbf{I} + \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T$$

$$\implies \mathbf{0} = -\mathbf{I}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} + \boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T$$

$$\implies \mathbf{I}N = \boldsymbol{\Sigma}^{-1}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T$$

$$\implies \boldsymbol{\Sigma}N = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T$$

$$\implies \boldsymbol{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T$$

$$\implies \boldsymbol{\Sigma} = \frac{1}{N}\sum_{k=1}^{K}\sum_{n=1}^{N} t_{nk}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T$$

$$\implies \boldsymbol{\Sigma} = \sum_{k=1}^{K}\frac{1}{N}\left(\frac{N_k}{N_k}\right)\sum_{n=1}^{N} t_{nk}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T$$

$$\implies \boldsymbol{\Sigma} = \sum_{k=1}^{K}\left(\frac{N_k}{N}\right)\frac{1}{N_k}\sum_{n=1}^{N} t_{nk}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T$$

$$\text{Let } \mathbf{S}_k = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk}(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)(\boldsymbol{\phi}_n - \boldsymbol{\mu}_k)^T$$

This gives us:

$$\boldsymbol{\Sigma} = \sum_{k=1}^{K}\left(\frac{N_k}{N}\right)\mathbf{S}_k.$$

Thus proved.