- Derive the update formulas of the paramerters π, μ, Λ by letting the partial derivate of the lower bound w.r.t. each parameter equal to zero.

$$\log p(X;\theta) = \log \int q(Z) \frac{p(X,Z;\theta)}{q(Z)} dZ$$

$$\geq \int q(Z) \log \frac{p(X,Z;\theta)}{q(Z)} dZ \quad (\because \text{ Jensen's inequality})$$

$$= \int q(Z) \log p(X,Z;\theta)\, dZ - \int q(Z) \log q(Z)\, dZ$$

The equality holds true when $q(Z) = p(Z|X;\theta)$ ⋯ (1)

When eq (1) is true, $\log p(X;\theta) = \int q(Z) \log p(X,Z;\theta) dZ + C$

$$\left(C := -\int q(Z) \log q(Z)\, dZ \right)$$

If there is a dataset $\{x_1, \cdots, x_N\}$,

$$q(z_{nk}=1) = \frac{p(x_n, z_{nk}=1; \pi_k, \mu_k, \Lambda_k)}{\sum_{k'=1}^{K} p(x_n, z_{nk'}=1; \pi_{k'}, \mu_{k'}, \Lambda_{k'})}$$

$$= \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Lambda_k^{-1})}{\sum_{k'=1}^{K} \pi_{k'} \mathcal{N}(x_n | \mu_{k'}, \Lambda_{k'}^{-1})} = \gamma(z_{nk})$$

$$\log p(X;\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} \underline{q(z_{nk}=1)}\, \underline{\log p(x_n, z_{nk};\theta)} + C$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \underline{\gamma(z_{nk})}\, \underline{\log \{\pi_k \mathcal{N}(x_n | \mu_k, \Lambda_k^{-1})\}} + C$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left\{ \log \pi_k - \frac{1}{2} \log |\Sigma_k| - \frac{1}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) + C' \right\} + C$$

$$(\Sigma_k := \Lambda_k^{-1})$$

$$\frac{\log P(X;\theta)}{\partial \mu_k^*} = \sum_{n=1}^{N} \gamma(z_{nk}) \left\{ -\cancel{\frac{1}{2}} \cdot \cancel{2} \cdot \Sigma_k^{-1} (x_n - \mu_k^*) \right\} = 0$$

$$\sum_{n=1}^{N} \gamma(z_{nk}) \Sigma_k^{-1} x_n = \sum_{n=1}^{N} \gamma(z_{nk}) \Sigma_k^{-1} \mu_k^*$$

$$\therefore \mu_k^* = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})} = \frac{S_k[x]}{S_k[1]} \quad \left(S_k[1] = \sum_{n=1}^{N} \gamma_{nk} ,\quad S_k[x] = \sum_{n=1}^{N} \gamma_{nk} x_n \right)$$

//

$$\frac{\log P(X;\theta)}{\partial \Sigma_k^*} = \sum_{n=1}^{N} \gamma(z_{nk}) \left\{ -\frac{1}{2\Sigma_k^*} + \frac{1}{2} \Sigma_k^{*-1} (x_n - \mu_k)(x_n - \mu_k)^T \Sigma_k^{*-1} \right\} = 0$$

$$\frac{1}{2\Sigma_k^*} \sum_{n=1}^{N} \gamma(z_{nk}) = \frac{1}{2} \Sigma_k^{*-2} \sum_{n=1}^{N} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T$$

$$\Sigma_k^* = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$= \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n x_n^T}{\sum_{n=1}^{N} \gamma(z_{nk})} - \mu_k \mu_k^T$$

$$\therefore \Sigma_k^* = \frac{S_k[x x^T]}{S_k[1]} - \mu_k \mu_k^T \quad \left(S_k[x x^T] = \sum_{n=1}^{N} \gamma(z_{nk}) x_n x_n^T \right)$$

//

By Lagrange's method of undetermined multiplier, determine the optimal value of π_k, satisfying the constraint, i.e., $\sum_{k=1}^{K} \pi_k = 1$

$$J = \log P(X;\theta) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$$

$$\frac{\partial J}{\partial \pi_k^*} = \sum_{n=1}^{N} \gamma(z_{nk}) \frac{1}{\pi_k^*} + \lambda = 0$$

$$\sum_{n=1}^{N} \gamma(z_{nk}) + \pi_k^* \lambda = 0$$

$$\sum_{n=1}^{N} \gamma(z_{nk}) \frac{1}{\pi_k^*} - \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) = 0$$

$$\pi_k^* = \frac{\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk})}{\sum_{n=1}^{N} \gamma(z_{nk})} = \frac{S_k[1]}{S_{\cdot}[1]}$$

//

- Derive the variational posteriors of the parameters π, μ Λ by using the formulas (P46)

$$\log q^*(\pi) = \log p(\pi) + \langle \log p(Z|\pi) \rangle_{q(Z)} + C$$

$$= \log \mathrm{Dir}(\pi|\alpha_0) + \left\langle \sum_{k=1}^{K}\sum_{n=1}^{N} \log \pi_k^{z_{nk}} \right\rangle_{q(Z)} + C$$

$$= \log \mathrm{Dir}(\pi|\alpha_0) + \sum_{k=1}^{K}\sum_{n=1}^{N} \langle z_{nk} \rangle_{q(Z)} \log \pi_k + C$$

$$= \underbrace{\log \mathrm{Dir}(\pi|\alpha_0)}_{= C' + \sum_{k=1}^{K} \log \pi_k^{\alpha_{0k}-1}} + \sum_{k=1}^{K} \underbrace{\sum_{n=1}^{N} \gamma_{nk}}_{N_k} \log \pi_k + C$$

$$= C'' + \sum_{k=1}^{K} \log \pi_k^{\alpha_{0k}+N_k-1}$$

$$\Sigma^{-1} = (\beta_0 \Lambda_k)^{-1}$$

$$= \log \mathrm{Dir}(\pi|\alpha) + \text{Const}$$

$$\therefore \quad q^*(\pi) = \mathrm{Dir}(\pi|\alpha)$$

$$\log q^*(\mu, \Lambda) = \log p(\mu, \Lambda) + \langle \log p(X|Z,\mu,\Lambda) \rangle_{q(Z)} + \text{Const.}$$

$$= \sum_{k=1}^{K} \log \mathcal{N}(\mu_k | m_0, (\beta_0 \Lambda_k)^{-1}) \mathcal{W}(\Lambda_k | W_0, \nu_0) + \sum_{n=1}^{N}\sum_{k=1}^{K} \langle z_{nk} \rangle_{q(Z)} \log \mathcal{N}(x_n | \mu_k, \Lambda_k^{-1}) + \text{Const.}$$

$$= \sum_{k=1}^{K} \log \mathcal{N}(\mu_k | m_0, (\beta_0 \Lambda_k)^{-1}) + \sum_{k=1}^{K} \log \mathcal{W}(\Lambda_k | W_0, \nu_0) + \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma_{nk} \log \mathcal{N}(x_n | \mu_k, \Lambda_k^{-1}) + \text{Const.}$$

$$\mathcal{W}(\Lambda_k | W_0, \nu_0) = B(W_0, \nu_0) |\Lambda_k|^{(\nu_0 - D - 1)/2} \exp\left(-\frac{1}{2}\mathrm{Tr}[W_0^{-1}\Lambda_k]\right)$$

$$= \sum_{k=1}^{K} \Big\{ -\frac{1}{2}(\mu_k - m_0)^T \beta_0 \Lambda_k (\mu_k - m_0) + \frac{1}{2}\ln|\Lambda_k| + \text{const.} - \frac{1}{2}\mathrm{Tr}[W_0^{-1}\Lambda_k] + \frac{\nu_0 - D - 1}{2}\log|\Lambda_k| + \frac{1}{2}\sum_{n=1}^{N} \gamma_{nk}\left(\log|\Sigma_k| - (x_n - \mu_k)^T \Lambda_k (x_n - \mu_k) + \text{Const.} \right) \Big\}$$

$$\log \mathcal{N}(x_n | \mu_k, \Lambda_k^{-1}) = \frac{1}{2}\log|\Lambda_k| - \frac{1}{2}(x_n - \mu_k)^T \Lambda_k (x_n - \mu_k) + \cdots$$

$$= \sum_{k=1}^{K} \Big\{ -\frac{1}{2}(\mu_k - m_0)^T \beta_0 \Lambda_k (\mu_k - m_0) + \frac{1}{2}\log|\Lambda_k| + \text{const.}$$
$$-\frac{1}{2}\mathrm{Tr}[W_0^{-1}\Lambda_k] + \frac{\nu_0 - D - 1}{2}\log|\Lambda_k|$$
$$+\frac{1}{2} N_k \log|\Sigma_k| - \frac{1}{2}\sum_{n=1}^{N} \gamma_{nk}(x_n - \mu_k)^T \Lambda_k (x_n - \mu_k) + N_k \times \text{const} \Big\} \quad \cdots ①$$

We can factorize $\log q^*(\mu, \Lambda)$ as:

$$\log q^*(\mu, \Lambda) = \log q^*(\mu|\Lambda) + \log q^*(\Lambda)$$

This means that the terms which depends on μ from ① correspond to $\log q^*(\mu|\Lambda)$.

Therefore,

$$\log q^*(\mu|\Lambda) = -\frac{1}{2}\sum_{k=1}^{K}\Big\{(\mu_k - m_0)^T \beta_0 \Lambda_k (\mu_k - m_0) + N_k (x_n - \mu_k)^T \Lambda_k (x_n - \mu_k)\Big\} + \text{const} \quad \cdots ②$$

We assume that $p(\mu|\Lambda)p(\Lambda)$ is a Gaussian-Wishart distribution, so $q^*(\mu|\Lambda)$ should be a multiple Gaussian distribution, and $q^*(\Lambda)$ should be a Wishert distribution.

Therefore, $q^*(\mu_k|\Lambda_k)$ should be represented as:

$$\begin{cases} \cdot\ q^*(\mu_k|\Lambda_k) = \mathcal{N}(\mu_k | m_k, (\beta_k \Lambda_k)^{-1}) \\ \cdot\ \beta_k := \beta_0 + N_k, \quad m_k = \frac{1}{\beta_k}(\beta_0 m_0 + S_k[x]) \end{cases}$$

$$\log q^*(\mu_k|\Lambda_k) = \frac{1}{2}\log|\beta_k \Lambda_k| - \frac{\beta_k}{2}(\mu_k - m_k)^T \Lambda_k (\mu_k - m_k) + \text{const}$$

$$\therefore \log q^*(\Lambda_k) = \log q^*(\mu_k, \Lambda_k) - \log q^*(\mu_k|\Lambda_k)$$
$$= -\frac{1}{2}(\mu_k - m_0)^T \beta_0 \Lambda_k (\mu_k - m_0) + \frac{1}{2}\log|\Lambda_k|$$
$$-\frac{1}{2}\mathrm{Tr}[W_0^{-1}\Lambda_k] + \frac{\nu_0 - D - 1}{2}\log|\Lambda_k|$$
$$+\frac{1}{2}N_k \log|\Lambda_k| - \frac{1}{2}\sum_{n=1}^{N}\gamma_{nk}(x_n - \mu_k)^T \Lambda_k (x_n - \mu_k)$$
$$-\frac{1}{2}\log|\Lambda_k| + \frac{\beta_k}{2}(\mu_k - m_k)^T \Lambda_k (\mu_k - m_k) + \text{const.}$$
$$= \frac{\nu_k - D - 1}{2}\log|\Lambda_k| - \frac{1}{2}\mathrm{Tr}[\Lambda_k W_k^{-1}] + \text{const.} \quad \cdots ③$$

$W_0^{-1} = Ax, \quad \Lambda_k = x^T$

$\mathrm{Tr}[Axx^T] = x^T A x = \Lambda_k W_0^{-1}$

Omit complex transformations of W_k^{-1}

$$\begin{cases} W_k^{-1} = \beta_0(\mu_k - m_0)(\mu_k - m_0)^T - W_0^{-1} + \sum_{n=1}^{N} \gamma_{nk}(x_n - \mu_k)(x_n - \mu_k)^T - \left(\beta_k(\mu_k - m_k)(\mu_k - m_k)^T\right) \\ \nu_k = \nu_0 + N_k \end{cases}$$

Comparing ③ to $\log W(\Lambda | W, \nu)$, we can get:

$$\underline{q^*(\Lambda_k) = W(\Lambda_k | W_k, \nu_k)} \quad //$$

And we can derive $q^*(\mu, \Lambda)$ as;

$$q^*(\mu, \Lambda) = q^*(\mu | \Lambda)\, q^*(\Lambda)$$

$$= \underline{\prod_{k=1}^{K} N(\mu_k | m_k, (\beta_k \Lambda_k)^{-1})\, W(\Lambda_k | W_k, \nu_k)} \quad //$$