Problem 1 Optimization over the simplex

1.1 Lagrangian: $\Sigma_{k'=1}^{K} a_{k'} ln(q_{k'}) - (\Sigma_{k'=1}^{K} \mu_{k'}(-q_{k'})) + \lambda((\Sigma_{k'=1}^{K} q_{k'}) - 1)$
Stationary condition:

$$\frac{d}{d(q_k)}\Sigma_{k'=1}^{K} a_{k'} ln(q_{k'}) + (\Sigma_{k'=1}^{K} \mu_{k'} q_{k'}) + \lambda((\Sigma_{k'=1}^{K} q_{k'}) - 1) = 0$$
$$\frac{a_k}{q_k} + \mu_k + \lambda = 0$$
$$q_k = -\frac{a_k}{\mu_k + \lambda} \tag{1}$$

Complimentary slackness and feasibility:

$$\forall k' \quad \mu_{k'} q_{k'} = 0 \tag{2}$$
$$\Sigma_{k'=1}^{K} q_{k'} = 1 \tag{3}$$

Since $ln(0)$ is undefined, $q_{k'}$ cannot be 0; thus from (2), $\forall k \quad \mu_{k'} = 0$. Substituting $\mu_k = 0$ into (1), we get $q_k = -\frac{a_k}{\lambda}$; combining this with (3), we get $\lambda = -\Sigma_{k'=1}^{K} a_{k'}$. Therefore $q_k^* = \frac{a_k}{\Sigma_{k'=1}^{K} a_{k'}}$.

1.2 Lagrangian: $\Sigma_{k'=1}^{K}(q_{k'} b_{k'} - q_{k'} ln(q_{k'})) - (\Sigma_{k'=1}^{K} \mu_{k'}(-q_{k'})) + \lambda((\Sigma_{k'=1}^{K} q_{k'}) - 1)$
Stationary condition:

$$\frac{d}{d(q_k)}\Sigma_{k'=1}^{K}(q_{k'} b_{k'} - q_{k'} ln(q_{k'})) - (\Sigma_{k'=1}^{K} \mu_{k'}(-q_{k'})) + \lambda((\Sigma_{k'=1}^{K} q_{k'}) - 1) = 0$$

$$\boxed{\begin{gathered} When\ y = u(x)v(x) \\ \frac{dy}{dx} = \frac{du}{dx}v + u\frac{dv}{dx} \end{gathered}}$$

$$b_k - 1 - ln(q_k) + \mu_k + \lambda = 0$$
$$ln(q_k) = b_k - 1 + \mu_k + \lambda$$
$$q_k = exp(b_k - 1 + \mu_k + \lambda) \tag{1}$$

Complimentary slackness and feasibility:

$$\forall k' \quad \mu_{k'} q_{k'} = 0 \tag{2}$$
$$\Sigma_{k'=1}^{K} q_{k'} = 1 \tag{3}$$

Since $ln(0)$ is undefined, $q_{k'}$ cannot be 0; thus from (2), $\forall k' \quad \mu_{k'} = 0$. Substituting $\mu_k = 0$ into (1), we get $q_k = exp(b_k - 1 + \lambda)$; combining this with (3), we get $exp(\lambda) = \frac{1}{\Sigma_{k'=1}^{K} exp(b_{k'} - 1)}$. Therefore $q_k^* = \frac{exp(b_k - 1)}{\Sigma_{k'=1}^{K} exp(b_{k'} - 1)} = \frac{exp(b_k)}{\Sigma_{k'=1}^{K} exp(b_{k'})}$.

Problem 2 Gaussian Mixture Model and EM

2.1 Please distinguish between $\boldsymbol{\Sigma}$, which represents covariance matrix, and Σ, which represents summation.
Lagrangian: $\Sigma_n \Sigma_{k'} \gamma_{nk'} ln(\omega_{k'}) + \Sigma_n \Sigma_{k'} \gamma_{nk'} ln(N(x_n \mid \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})) - \Sigma_{k'} \alpha_{k'}(-\omega_{k'}) + \lambda((\Sigma_{k'} \omega_{k'}) - 1)$
Taking derivative w.r.t ω_k:

$$\frac{d}{d(\omega_k)}\Sigma_n \Sigma_{k'} \gamma_{nk'} ln(\omega_{k'}) + \Sigma_n \Sigma_{k'} \gamma_{nk'} ln(N(x_n \mid \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})) - \Sigma_{k'} \alpha_{k'}(-\omega_{k'}) + \lambda((\Sigma_{k'} \omega_{k'}) - 1) = 0$$
$$\frac{d}{d(\omega_k)}\Sigma_n \Sigma_{k'} \gamma_{nk'} ln(\omega_{k'}) + \Sigma_{k'} \alpha_{k'} \omega_{k'} + \lambda((\Sigma_{k'} \omega_{k'}) - 1) = 0$$
$$\frac{\Sigma_n \gamma_{nk}}{\omega_k} + \alpha_k + \lambda = 0$$

Since all γ_{nk} are given, $\Sigma_n \gamma_{nk}$ can be treated as a constant exchangeable with $a_{k'}$ in 1.1, and the solution just becomes $\omega_k^* = \frac{\Sigma_n \gamma_{nk}}{\Sigma_n \Sigma_{k'} \gamma_{nk'}}$.

Taking derivative w.r.t $\boldsymbol{\mu}_k$:

$$\frac{d}{d(\boldsymbol{\mu}_k)} \Sigma_n \Sigma_{k'} \gamma_{nk'} ln(\omega_{k'}) + \Sigma_n \Sigma_{k'} \gamma_{nk'} ln(N(x_n \mid \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})) - \Sigma_{k'} \alpha_{k'}(-\omega_{k'}) + \lambda((\Sigma_{k'} \omega_{k'}) - 1) = 0$$

$$\frac{d}{d(\boldsymbol{\mu}_k)} \Sigma_n \Sigma_{k'} \gamma_{nk'} ln(\frac{exp(-0.5(\mathbf{x}_n - \boldsymbol{\mu}_{k'})^T \boldsymbol{\Sigma}_{k'}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_{k'}))}{(\sqrt{2\pi})^D \|\boldsymbol{\Sigma}_{k'}\|^{0.5}}) = 0$$

$$\frac{d}{d(\boldsymbol{\mu}_k)} \Sigma_n \Sigma_{k'} \gamma_{nk'} (-0.5(\mathbf{x}_n - \boldsymbol{\mu}_{k'})^T \boldsymbol{\Sigma}_{k'}^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_{k'})) = 0$$

(denominator doesn't contain $\boldsymbol{\mu}$; ln and exp on the numerator cancel out)

$$\Sigma_n \gamma_{nk} \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_k) = 0$$

(all $k' \neq k$ in $\Sigma_{k'}$ won't be in derivative)

$$\boldsymbol{\mu}_k = \frac{\Sigma_n \gamma_{nk} \mathbf{x}_n}{\Sigma_n \gamma_{nk}}$$

Taking derivative w.r.t $\boldsymbol{\Sigma}_k$:

$$\frac{d}{d(\boldsymbol{\Sigma}_k)} \Sigma_n \Sigma_{k'} \gamma_{nk'} ln(\omega_{k'}) + \Sigma_n \Sigma_{k'} \gamma_{nk'} ln(N(x_n \mid \boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})) - \Sigma_{k'} \alpha_{k'}(-\omega_{k'}) + \lambda((\Sigma_{k'} \omega_{k'}) - 1) = 0$$

$$\frac{d}{d(\boldsymbol{\Sigma}_k)} \Sigma_n \gamma_{nk} ln(\frac{exp(-0.5(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_k))}{(\sqrt{2\pi})^D \|\boldsymbol{\Sigma}_k\|^{0.5}}) = 0$$

$$\frac{d}{d(\boldsymbol{\Sigma}_k)} \Sigma_n \gamma_{nk} (-0.5 ln(\|\boldsymbol{\Sigma}_k\|) - 0.5(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_k)) = 0$$

$$\Sigma_n \gamma_{nk} (-0.5 \boldsymbol{\Sigma}_k^{-1} + 0.5 \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}) = 0$$

(matrix cookbook (51) and (55); $\boldsymbol{\Sigma}^T = \boldsymbol{\Sigma}$)

$$\Sigma_n \gamma_{nk} \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T = \Sigma_n \gamma_{nk}$$

$$\boldsymbol{\Sigma}_k = \frac{\Sigma_n \gamma_{nk}(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T}{\Sigma_n \gamma_{nk}}$$

2.2 Substituting $b_n = ln(p(\mathbf{x}_n, z_n; \theta^{(t)}))$ into 1.2 we get $\frac{exp(ln(p(\mathbf{x}_n, z_n; \theta^{(t)})))}{\Sigma_{n'=1}^N exp(ln(p(\mathbf{x}_{n'}, z_{n'}; \theta^{(t)})))} = \frac{p(z_n|\mathbf{x}_n; \theta^{(t)}) p(\mathbf{x}_n|\theta^{(t)}) p(\theta^{(t)})}{\Sigma_{n'=1}^N p(z_{n'}|\mathbf{x}_{n'}; \theta^{(t)}) p(\mathbf{x}_{n'}|\theta^{(t)}) p(\theta^{(t)})} = \frac{p(z_n|\mathbf{x}_n; \theta^{(t)}) p(\mathbf{x}_n|\theta^{(t)})}{p(\mathbf{x}_n|\theta^{(t)})} = p(z_n \mid \mathbf{x}_n; \theta^{(t)})$

2.3 Change the Gaussian probability into an indicator hard assignment $p(\mathbf{x}_n \mid z_n = k) = \mathbb{I}(k = argmin_j(\|\mathbf{x}_n - \boldsymbol{\mu}_j\|_2^2))$.
$p(z_n = k \mid \mathbf{x}_n) = \frac{p(\mathbf{x}_n|z_n=k)p(z_n=k)}{p(\mathbf{x}_n)}$, which is essentially either $\frac{p(z_n=k)}{p(\mathbf{x}_n)}$ (We can get $p(z_n = k)$ by counting; I think we can ignore $p(\mathbf{x}_n)$) or 0 depending on whether $\mathbb{I}(k = argmin_j(\|\mathbf{x}_n - \boldsymbol{\mu}_j\|_2^2))$ is 1 or 0.

Problem 3 Hidden Markov Model

3.1

$$
\begin{aligned}
P(X_{T+1} \mid O_{1:T}) &= \Sigma_{s'} P(X_{T+1} = s, X_T = s' \mid O_{1:T}) && \text{(marginalization)} \\
&= \Sigma_{s'} P(X_{T+1} = s \mid X_T = s', O_{1:T}) P(X_T = s' \mid O_{1:T}) && \text{(chain)} \\
&= \frac{1}{P(O_{1:T})} \Sigma_{s'} P(X_{T+1} = s \mid X_T = s') P(X_T = s', O_{1:T}) && \text{(Markov property)} \\
&\propto \Sigma_{s'} a_{s',s} \alpha_{s'}(T)
\end{aligned}
$$

3.2 A HMM can be simulated by a GMM with the following parameter assignments:

$$
\begin{aligned}
\pi_s &= \omega_k \\
a_{s,s'} &= \Sigma_{k'} N(x_2 \mid x_1, \mathbf{\Sigma}_{k'}) \\
b_{s,o} &= P(z = k \mid x) = \omega_k N(x \mid \boldsymbol{\mu}_k, \mathbf{\Sigma}_k)
\end{aligned}
$$