$$\boldsymbol{\phi} \equiv [\phi_1(x) \quad \phi_2(x) \quad \cdots \quad \phi_K(x)]$$

So after it,

$$\begin{aligned}
&\ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta}) \\
&= \sum_{n=1}^{N}\sum_{j=1}^{K}\sum_{t=1}^{T} z_{nj} x_n(t) \ln \phi_j(t) + \sum_{n=2}^{N}\sum_{k=1}^{K}\sum_{j=1}^{K} z_{n-1,j} z_{n,k} \ln A_{kj} + \sum_{k=1}^{K} z_{1k} \ln \pi_k \\
&= \sum_{n=1}^{N} \boldsymbol{z_n}^\top \langle \ln \boldsymbol{\phi}, \boldsymbol{x_n} \rangle + \sum_{n=2}^{N} \boldsymbol{z_{n-1}}^\top \ln \boldsymbol{A}^\top \cdot \boldsymbol{z_n} + \ln \boldsymbol{\pi}^\top \cdot \boldsymbol{z_1}
\end{aligned}$$

if t is continuous

$$\begin{aligned}
\langle \ln \boldsymbol{\phi}, \boldsymbol{x_n} \rangle &= \int_{-\infty}^{+\infty} \ln \boldsymbol{\phi(x)}^\top \cdot x_n(x) dx \\
&= \int_{-\infty}^{+\infty} \ln \boldsymbol{\phi(x)}^\top \cdot \delta(x - x_n) dx \\
&= \ln \boldsymbol{\phi(x_n)}^\top
\end{aligned}$$

then

$$\begin{aligned}
\max_{\boldsymbol{\theta}} \quad \boldsymbol{Q}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{ old}}\right) &\equiv \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}\right) \ln p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta}) \\
&= \sum_{n=1}^{N} \boldsymbol{\gamma_n}^\top \langle \ln \boldsymbol{\phi}, \boldsymbol{x_n} \rangle + \sum_{n=2}^{N} \operatorname{tr}\left(\boldsymbol{\xi_n}^\top \ln \boldsymbol{A}\right) + \boldsymbol{\gamma_1}^\top \ln \boldsymbol{\pi} \\
&= \sum_{n=1}^{N} \boldsymbol{\gamma_n}^\top [-(x_n 1. - \mu) \odot (x_n 1. - \mu) \odot 0.5 \frac{1.}{\sigma^2} - 0.5 \cdot \ln \sigma^2 - (0.5 \cdot \ln 2\pi) \\
&+ \sum_{n=2}^{N} \operatorname{tr}\left(\boldsymbol{\xi_n}^\top \ln \boldsymbol{A}\right) + \boldsymbol{\gamma_1}^\top \ln \boldsymbol{\pi}
\end{aligned}$$

subject to the following conditions

$$\begin{gathered}
\text{s.t.} \quad \boldsymbol{\gamma_n}^\top \cdot 1. = 1 \quad \operatorname{tr}\left(\boldsymbol{\xi_n}^\top \cdot 1.\right) = 1 \\
1.^\top \cdot \boldsymbol{\phi} = 1.^\top \quad 1.^\top \cdot \boldsymbol{A} = 1.^\top \quad 1.^\top \cdot \boldsymbol{\pi} = 1
\end{gathered}$$

$\boldsymbol{\gamma_n} \equiv \left[\sum_{\boldsymbol{z_n}} p\left(\boldsymbol{z_n}|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}\right) \boldsymbol{z_n}\right]$ 为常数，是期望$E[\boldsymbol{z_n}|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}] = \vec{p}(\boldsymbol{z_n}|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}})$, 大小 $K \times 1$

$\boldsymbol{\xi_n} \equiv [\sum_{\boldsymbol{z_{n-1}}, \boldsymbol{z_n}} \boldsymbol{z_n} p\left(\boldsymbol{z_{n-1}}, \boldsymbol{z_n}|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}\right) \boldsymbol{z_{n-1}}^\top]$ 为常数，是期望 $E[\boldsymbol{z_n}\boldsymbol{z_{n-1}}^\top|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}] = \vec{p}(\boldsymbol{z_n}\boldsymbol{z_{n-1}}^\top|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}})$, 大小$K \times K$

with Lagrange method:

$$
\begin{aligned}
L \equiv & \left\{ \sum_{n=1}^{N} \boldsymbol{\gamma_n}^\top \langle \ln \boldsymbol{\phi}, \boldsymbol{x_n} \rangle + \sum_{n=2}^{N} \operatorname{tr}\left(\boldsymbol{\xi_n}^\top \ln \boldsymbol{A} \right) + \boldsymbol{\gamma_1}^\top \ln \boldsymbol{\pi} \right\} \\
& - \sum_{t=1}^{T} u_t \left\{ 1.^\top \cdot \langle \boldsymbol{\phi}, \delta(t) \rangle - 1 \right\} \\
& - \sum_{k=1}^{K} v_k \left\{ 1.^\top \boldsymbol{A} \cdot \sigma_k - 1 \right\} \\
& - w_1 \left\{ 1.^\top \boldsymbol{\pi} - 1 \right\}
\end{aligned}
$$

where $\sigma_k = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}, 1 \times K$

if t is discrete, $t = 1, \cdots, T$

$$
\delta(t) = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}, 1 \times T,
$$

$$
\langle \ln \boldsymbol{\phi}, \boldsymbol{x_n} \rangle = \ln \boldsymbol{\phi}^\top \cdot \boldsymbol{x_n}
$$

$$
\langle \boldsymbol{\phi}, \delta(t) \rangle = \boldsymbol{\phi}^\top \cdot \delta(t)
$$

$$
\begin{aligned}
\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\phi}} &= \left\{\sum_{n=1}^{N} [\boldsymbol{\gamma_n} \cdot \boldsymbol{x_n}^\top]\right\}^\top \odot \frac{\mathbf{1}}{\boldsymbol{\phi}} - \sum_{t=1}^{M} u_t \cdot 1.\delta(t)^\top \\
&= \left\{\sum_{n=1}^{N} [\boldsymbol{x_n} \cdot \boldsymbol{\gamma_n}^\top]\right\} \odot \frac{\mathbf{1}}{\boldsymbol{\phi}} - \begin{bmatrix} u_1 & \cdots & u_t \cdots & u_T \\ u_1 & \cdots & u_t \cdots & u_T \\ \vdots & & & \vdots \\ u_1 & \cdots & u_t \cdots & u_T \end{bmatrix} = 0 \\
\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{A}} &= [\sum_{n=2}^{N} \boldsymbol{\xi_n}] \odot \frac{\mathbf{1}}{\boldsymbol{A}} - \sum_{k=1}^{K} v_k \cdot 1.\sigma_k^\top \\
&= [\sum_{n=2}^{N} \boldsymbol{\xi_n}] \odot \frac{\mathbf{1}}{\boldsymbol{A}} - \begin{bmatrix} v_1 & \cdots & v_k \cdots & v_K \\ v_1 & \cdots & v_k \cdots & v_K \\ \vdots & & & \vdots \\ v_1 & \cdots & v_k \cdots & v_K \end{bmatrix} = 0 \\
\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\pi}} &= \boldsymbol{\gamma_1} \odot \frac{1.}{\boldsymbol{\pi}} - w_1 \cdot 1. = 0
\end{aligned}
$$

Thus, we obtain

$$\frac{\left\{\sum_{n=1}^{N}[\boldsymbol{x_n}\cdot\boldsymbol{\gamma_n}^\top]\right\}_{ij}}{\boldsymbol{\phi}_{ij}}=u_j$$

$$\Longrightarrow\sum_{i=1}^{K}\frac{\left\{\sum_{n=1}^{N}[\boldsymbol{x_n}\cdot\boldsymbol{\gamma_n}^\top]\right\}_{ij}}{u_j}=\sum_{i=1}^{K}\boldsymbol{\phi}_{ij}=1$$

$$\Longrightarrow u_j=\sum_{i=1}^{K}\left\{\sum_{n=1}^{N}[\boldsymbol{x_n}\cdot\boldsymbol{\gamma_n}^\top]\right\}_{ij}$$

$$\Longrightarrow\boldsymbol{\phi}_{ij}=\frac{\left\{\sum_{n=1}^{N}[\boldsymbol{x_n}\cdot\boldsymbol{\gamma_n}^\top]\right\}_{ij}}{\sum_{i=1}^{K}\left\{\sum_{n=1}^{N}[\boldsymbol{x_n}\cdot\boldsymbol{\gamma_n}^\top]\right\}_{ij}}$$

$$\Longrightarrow\boldsymbol{\phi}=\left\{\sum_{n=1}^{N}[\boldsymbol{x_n}\cdot\boldsymbol{\gamma_n}^\top]\right\}\odot\frac{1.}{1.\cdot1.^\top\left\{\sum_{n=1}^{N}[\boldsymbol{x_n}\cdot\boldsymbol{\gamma_n}^\top]\right\}}$$

$$\Longrightarrow\boldsymbol{\phi}=\left\{\sum_{n=1}^{N}[\boldsymbol{x_n}\cdot\boldsymbol{\gamma_n}^\top]\right\}\odot\frac{1.}{1.\left[\sum_{n=1}^{N}\boldsymbol{\gamma_n}\right]^\top}$$

$$\frac{[\sum_{n=2}^{N}\boldsymbol{\xi_n}]_{ij}}{\boldsymbol{A}_{ij}}=v_j$$

$$\Longrightarrow\sum_{i=1}^{K}\frac{[\sum_{n=2}^{N}\boldsymbol{\xi_n}]_{ij}}{v_j}=\sum_{i=1}^{K}\boldsymbol{A}_{ij}=1$$

$$\Longrightarrow v_j=\sum_{i=1}^{K}[\sum_{n=2}^{N}\boldsymbol{\xi_n}]_{ij}$$

$$\Longrightarrow\boldsymbol{A}_{ij}=\frac{[\sum_{n=2}^{N}\boldsymbol{\xi_n}]_{ij}}{\sum_{i=1}^{K}[\sum_{n=2}^{N}\boldsymbol{\xi_n}]_{ij}}$$

$$\Longrightarrow\boldsymbol{A}=[\sum_{n=2}^{N}\boldsymbol{\xi_n}]\odot\frac{1.}{1.\cdot1.^\top[\sum_{n=2}^{N}\boldsymbol{\xi_n}]}$$

$$\frac{[\boldsymbol{\gamma_1}]_i}{\boldsymbol{\pi}_i}=w_1$$

$$\Longrightarrow\sum_{i=1}^{K}\frac{[\boldsymbol{\gamma_1}]_i}{w_1}=\sum_{i=1}^{K}\boldsymbol{\pi}_i=1$$

$$\Longrightarrow w_1=\sum_{i=1}^{K}[\boldsymbol{\gamma_1}]_i$$

$$\Longrightarrow\boldsymbol{\pi}_i=\frac{[\boldsymbol{\gamma_1}]_i}{\sum_{i=1}^{K}[\boldsymbol{\gamma_1}]_i}$$

$$\Longrightarrow\boldsymbol{\pi}=\boldsymbol{\gamma_1}\odot\frac{1.}{1.\cdot1.^\top\boldsymbol{\gamma_1}}$$

To sum up

$$\boldsymbol{\phi} = \left\{\sum_{n=1}^{N}[\boldsymbol{x_n} \cdot \boldsymbol{\gamma_n}^\top]\right\} \odot \frac{1.}{1.\left[\sum_{n=1}^{N} \boldsymbol{\gamma_n}\right]^\top}$$
$$\boldsymbol{A} = [\sum_{n=2}^{N} \boldsymbol{\xi_n}] \odot \frac{1.}{1. \cdot 1.^\top[\sum_{n=2}^{N} \boldsymbol{\xi_n}]}$$
$$\boldsymbol{\pi} = \boldsymbol{\gamma_1} \odot \frac{1.}{1. \cdot 1.^\top \boldsymbol{\gamma_1}}$$

update $\boldsymbol{\theta}^{k+1} \equiv \{\boldsymbol{\phi}, \boldsymbol{A}, \boldsymbol{\pi}\}$

$$\begin{aligned}
\boldsymbol{\phi} &\Leftarrow \left\{\sum_{n=1}^{N}[\boldsymbol{x_n} \cdot \boldsymbol{\gamma_n}^\top]\right\} \odot \frac{1.}{1. \cdot 1.^\top \left\{\sum_{n=1}^{N}[\boldsymbol{x_n} \cdot \boldsymbol{\gamma_n}^\top]\right\}} \\
&= \left\{\sum_{n=1}^{N}[\boldsymbol{x_n} \cdot \boldsymbol{\gamma_n}^\top]\right\} \odot \frac{1.}{1.\left[\sum_{n=1}^{N} \boldsymbol{\gamma_n}\right]^\top} \\
\boldsymbol{A} &\Leftarrow [\sum_{n=2}^{N} \boldsymbol{\xi_n}] \odot \frac{1.}{1. \cdot 1.^\top[\sum_{n=2}^{N} \boldsymbol{\xi_n}]} \\
&= [\sum_{n=2}^{N} \boldsymbol{\xi_n}] \odot \frac{1.}{1. \cdot [\sum_{n=2}^{N} \boldsymbol{\gamma_{n-1}}]^\top} \\
&= [\sum_{n=2}^{N} \boldsymbol{\xi_n}] \odot \frac{1.}{1. \cdot [\sum_{n=1}^{N-1} \boldsymbol{\gamma_n}]^\top} \\
\boldsymbol{\pi} &\Leftarrow \boldsymbol{\gamma_1} \odot \frac{1.}{1. \cdot 1.^\top \boldsymbol{\gamma_1}} = \boldsymbol{\gamma_1}
\end{aligned}$$

because

$$\begin{gathered}
1.^\top[\boldsymbol{x_n} \cdot \boldsymbol{\gamma_n}^\top] = (1.^\top \boldsymbol{x_n}) \cdot \boldsymbol{\gamma_n}^\top = 1 \cdot \boldsymbol{\gamma_n}^\top = \boldsymbol{\gamma_n}^\top \\
1.^\top \boldsymbol{\xi_n} = 1.^\top \vec{p}(\boldsymbol{z_n} \boldsymbol{z_{n-1}}^\top | \boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}}) = \vec{p}(\boldsymbol{z_{n-1}} | \boldsymbol{X}, \boldsymbol{\theta}^{\text{ old}})^\top = \boldsymbol{\gamma_{n-1}}^\top \\
1.^\top \boldsymbol{\gamma_n} = 1
\end{gathered}$$

Or example: Gaussian Mixed Model

If we assume that $\phi(x) \equiv P(X = x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$

$$\ln\boldsymbol{\phi}(\boldsymbol{x_n})^\top = \begin{bmatrix} \ln\phi_1(x_n) \\ \ln\phi_2(x_n) \\ \vdots \\ \ln\phi_K(x_n) \end{bmatrix} = \begin{bmatrix} -\frac{(x_n-\mu_1)^2}{2\sigma_1^2} - \ln(\sigma_1) - 0.5 \cdot \ln(2\pi) \\ -\frac{(x_n-\mu_2)^2}{2\sigma_2^2} - \ln(\sigma_2) - 0.5 \cdot \ln(2\pi) \\ \vdots \\ -\frac{(x_n-\mu_K)^2}{2\sigma_K^2} - \ln(\sigma_K) - 0.5 \cdot \ln(2\pi) \end{bmatrix}$$
$$= -(x_n 1. - \mu) \odot (x_n 1. - \mu) \odot 0.5 \frac{1.}{\sigma^2} - 0.5 \cdot \ln\sigma^2 - (0.5 \cdot \ln 2\pi) 1.$$

we obtain that

for μ

$$\begin{aligned}
\frac{\partial L}{\partial \mu} &= \sum_{n=1}^{N}\{\boldsymbol{\gamma_n} \odot [-0.5\frac{1.}{\sigma^2} \odot 2(x_n \cdot 1. - \mu)]\} \\
&= \frac{-1.}{\sigma^2} \odot \sum_{n=1}^{N}\{\boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu)\} = 0 \\
&\sum_{n=1}^{N}\{\boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu)\} \\
&= \{\sum_{n=1}^{N} x_n \boldsymbol{\gamma_n}\} - \{\sum_{n=1}^{N} \boldsymbol{\gamma_n}\} \odot \mu = 0 \\
&\Rightarrow \mu = \{\sum_{n=1}^{N} x_n \boldsymbol{\gamma_n}\} \odot \frac{1.}{\{\sum_{n=1}^{N} \boldsymbol{\gamma_n}\}}
\end{aligned}$$

for σ^2

$$\begin{aligned}
\frac{\partial L}{\partial \sigma^2} &= \sum_{n=1}^{N}\{\boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \odot (x_n \cdot 1. - \mu) \odot 0.5\frac{1.}{(\sigma^2)^2} - \boldsymbol{\gamma_n} \odot 0.5\frac{1.}{\sigma^2}\} \\
&= \frac{0.5.}{(\sigma^2)^2} \odot \sum_{n=1}^{N}\{\boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \odot (x_n \cdot 1. - \mu) - \boldsymbol{\gamma_n} \odot \sigma^2\} = 0 \\
&\sum_{n=1}^{N}\{\boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \odot (x_n \cdot 1. - \mu) - \boldsymbol{\gamma_n} \odot \sigma^2\} \\
&= \{\sum_{n=1}^{N} \boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \odot (x_n \cdot 1. - \mu)\} - \{\sum_{n=1}^{N} \boldsymbol{\gamma_n}\} \odot \sigma^2 = 0 \\
&\Rightarrow \sigma^2 = \{\sum_{n=1}^{N} \boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \odot (x_n \cdot 1. - \mu)\} \odot \frac{1.}{\{\sum_{n=1}^{N} \boldsymbol{\gamma_n}\}}
\end{aligned}$$

Specially

when z_n, z_{n-1} are not related, all z_n are under distribution of π

$$\boldsymbol{\pi} \Leftarrow \{\sum_{n=1}^{N} \boldsymbol{\gamma_n}\} \odot \frac{1.}{1. \cdot 1.^\top \{\sum_{n=1}^{N} \boldsymbol{\gamma_n}\}} = \frac{\{\sum_{n=1}^{N} \boldsymbol{\gamma_n}\}}{N}$$