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

So after it,

$$egin{aligned} & \ln p(oldsymbol{X},oldsymbol{Z}|oldsymbol{ heta}) \ &= \sum_{n=1}^{N} \sum_{j=1}^{K} \sum_{t=1}^{T} z_{nj} x_n(t) \ln \phi_j(t) + \sum_{n=2}^{N} \sum_{k=1}^{K} \sum_{j=1}^{K} z_{n-1,j} z_{n,k} \ln A_{kj} + \sum_{k=1}^{K} z_{1k} \ln \pi_k \ &= \sum_{n=1}^{N} oldsymbol{z}_{oldsymbol{n}}^{ op} \langle \ln oldsymbol{\phi}, oldsymbol{x}_{oldsymbol{n}}
angle + \sum_{n=2}^{N} oldsymbol{z}_{n-1}^{ op} \ln oldsymbol{A}^{ op} \cdot oldsymbol{z}_n + \ln oldsymbol{\pi}^{ op} \cdot oldsymbol{z}_1 \end{aligned}$$

if t is continuous

$$egin{aligned} \langle \ln oldsymbol{\phi}, oldsymbol{x_n}
angle &= \int_{-\infty}^{+\infty} \ln oldsymbol{\phi}(oldsymbol{x})^ op \cdot x_n(x) dx \ &= \int_{-\infty}^{+\infty} \ln oldsymbol{\phi}(oldsymbol{x})^ op \cdot \delta(x-x_n) dx \ &= \ln oldsymbol{\phi}(oldsymbol{x_n})^ op \end{aligned}$$

then

$$egin{aligned} \max_{oldsymbol{ heta}} & oldsymbol{Q}\left(oldsymbol{ heta}, oldsymbol{ heta} ^{ ext{old}}
ight) \equiv \sum_{oldsymbol{z}}^{oldsymbol{p}} oldsymbol{Z} oldsymbol{Z} oldsymbol{Z} \left(oldsymbol{eta}, oldsymbol{ heta} ^{ ext{old}} \ln oldsymbol{p} oldsymbol{Q} \left(oldsymbol{X}, oldsymbol{Z} oldsymbol{ heta} ^{ ext{old}} \ln oldsymbol{A} ig) + \sum_{n=2}^{N} \operatorname{tr}\left(oldsymbol{\xi}_{oldsymbol{n}}^{ ext{old}} \ln oldsymbol{\phi}, oldsymbol{x}_{oldsymbol{n}} ig) + \sum_{n=2}^{N} \operatorname{tr}\left(oldsymbol{\xi}_{oldsymbol{n}}^{ ext{old}} \ln oldsymbol{A} ig) + oldsymbol{\gamma}_{oldsymbol{1}}^{ ext{old}} + oldsymbol{\gamma}_{oldsymbol{1}}^{ ext{old}} + oldsymbol{\gamma}_{oldsymbol{1}}^{ ext{old}} + oldsymbol{\gamma}_{oldsymbol{1}}^{ ext{old}} + oldsymbol{\gamma}_{oldsymbol{1}} + oldsymbol{\gamma}_{oldsymbol{1}}^{ ext{old}} + old$$

subject to the following conditions

$$\mathbf{s.t.}$$
 $\boldsymbol{\gamma}_n^{ op} \cdot 1. = 1$ $\mathrm{tr}\left(\boldsymbol{\xi}_n^{ op} \cdot 1.\right) = 1$ $1.^{ op} \cdot \boldsymbol{\phi} = 1.^{ op} \ 1.^{ op} \cdot \boldsymbol{A} = 1.^{ op} \ 1.^{ op} \cdot \boldsymbol{\pi} = 1$ $\boldsymbol{\gamma}_n \equiv \left[\sum_{\boldsymbol{z}_n} p\left(\boldsymbol{z}_n | \boldsymbol{X}, \boldsymbol{\theta}^{\mathrm{old}}\right) \boldsymbol{z}_n\right]$ 为常数,是期望 $E[\boldsymbol{z}_n | \boldsymbol{X}, \boldsymbol{\theta}^{\mathrm{old}}] = \vec{p}(\boldsymbol{z}_n | \boldsymbol{X}, \boldsymbol{\theta}^{\mathrm{old}})$,大小 $K \times 1$ $\boldsymbol{\xi}_n \equiv \left[\sum_{\boldsymbol{z}_n} \boldsymbol{z}_n p\left(\boldsymbol{z}_{n-1}, \boldsymbol{z}_n | \boldsymbol{X}, \boldsymbol{\theta}^{\mathrm{old}}\right) \boldsymbol{z}_{n-1}^{ op}\right]$ 为常数,是期望

 $E[\boldsymbol{z}_{n}\boldsymbol{z}_{n-1}^{\top}|\boldsymbol{X},\boldsymbol{\theta}^{\mathrm{old}}] = \vec{p}(\boldsymbol{z}_{n}\boldsymbol{z}_{n-1}^{\top}|\boldsymbol{X},\boldsymbol{\theta}^{\mathrm{old}}), \boldsymbol{\pm}/|\boldsymbol{X} \times K|$

with Lagrange method:

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

where
$$\sigma_k = egin{bmatrix} 0 \ dots \ 1 \ dots \ 0 \end{bmatrix}$$
 , $1 imes K$

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

$$\delta(t) = egin{bmatrix} 0 \ dots \ 1 \ dots \ 0 \end{bmatrix}$$
 , $1 imes T$,

$$\langle \ln oldsymbol{\phi}, oldsymbol{x_n}
angle = \ln oldsymbol{\phi}^ op \cdot oldsymbol{x_n}$$

$$\langle oldsymbol{\phi}, \delta(t)
angle = oldsymbol{\phi}^ op \cdot \delta(t)$$

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\phi}} = \left\{ \sum_{n=1}^{N} [\boldsymbol{\gamma}_{n} \cdot \boldsymbol{x}_{n}^{\top}] \right\}^{\top} \odot \frac{1}{\boldsymbol{\phi}} - \sum_{t=1}^{M} u_{t} \cdot 1.\delta(t)^{\top}$$

$$= \left\{ \sum_{n=1}^{N} [\boldsymbol{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{\top}] \right\} \odot \frac{1}{\boldsymbol{\phi}} - \begin{bmatrix} u_{1} & \cdots & u_{t} \cdots & u_{T} \\ u_{1} & \cdots & u_{t} \cdots & u_{T} \\ \vdots & & & \vdots \\ u_{1} & \cdots & u_{t} \cdots & u_{T} \end{bmatrix} = 0$$

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{A}} = [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}] \odot \frac{1}{\boldsymbol{A}} - \sum_{k=1}^{K} v_{k} \cdot 1.\sigma_{k}^{\top}$$

$$= [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}] \odot \frac{1}{\boldsymbol{A}} - \begin{bmatrix} v_{1} & \cdots & v_{k} \cdots & v_{K} \\ v_{1} & \cdots & v_{k} \cdots & v_{K} \\ \vdots & & & \vdots \\ v_{1} & \cdots & v_{k} \cdots & v_{K} \end{bmatrix} = 0$$

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\pi}} = \boldsymbol{\gamma}_{1} \odot \frac{1}{\boldsymbol{\pi}} - w_{1} \cdot 1. = 0$$

Thus, we obtain

$$\begin{split} &\left\{ \sum_{n=1}^{N} [\boldsymbol{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{\top}] \right\}_{ij} = u_{j} \\ &\Rightarrow \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} \phi_{ij} = 1 \\ &\Rightarrow u_{j} = \sum_{i=1}^{K} \left\{ \sum_{n=1}^{N} [\boldsymbol{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{\top}] \right\}_{ij} \\ &\Rightarrow \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}} \\ &\Rightarrow \phi = \left\{ \sum_{n=1}^{N} [\boldsymbol{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{\top}] \right\} \odot \frac{1}{1 \cdot 1 \cdot 1^{\top} \left\{ \sum_{n=1}^{N} [\boldsymbol{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{\top}] \right\}} \\ &\Rightarrow \phi = \left\{ \sum_{n=1}^{N} [\boldsymbol{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{\top}] \right\} \odot \frac{1}{1 \cdot \left[\sum_{n=1}^{N} \boldsymbol{\gamma}_{n} \right]^{\top}} \\ &\Rightarrow \phi = \left\{ \sum_{n=1}^{N} [\boldsymbol{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{\top}] \right\} \odot \frac{1}{1 \cdot \left[\sum_{n=1}^{N} \boldsymbol{\gamma}_{n} \right]^{\top}} \\ &\Rightarrow \phi = \left\{ \sum_{n=1}^{N} [\boldsymbol{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{\top}] \right\} \odot \frac{1}{1 \cdot \left[\sum_{n=1}^{N} \boldsymbol{\gamma}_{n} \right]^{\top}} \\ &\Rightarrow \sum_{i=1}^{K} \frac{[\boldsymbol{\Sigma}_{n=2}^{N} \boldsymbol{\xi}_{n}]_{ij}}{v_{j}} = \sum_{i=1}^{K} \boldsymbol{A}_{ij} = 1 \\ &\Rightarrow v_{j} = \sum_{i=1}^{K} [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}]_{ij} \\ &\Rightarrow \boldsymbol{A} = [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}] \odot \frac{1}{1 \cdot 1 \cdot 1^{\top} [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}]} \\ &\Rightarrow \boldsymbol{A} = [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}] \odot \frac{1}{1 \cdot 1 \cdot 1^{\top} [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}]} \\ &\Rightarrow \boldsymbol{\chi} = \sum_{i=1}^{K} [\boldsymbol{\gamma}_{1}]_{i} \\ &\Rightarrow \boldsymbol{\chi} = \sum_{i=1}^{K} [\boldsymbol{\gamma}_{1}]_{i} \\ &\Rightarrow \boldsymbol{\pi}_{i} = \boldsymbol{\gamma}_{1} \odot \frac{1}{1 \cdot 1 \cdot 1^{\top} \boldsymbol{\gamma}_{1}} \end{aligned}$$

$$oldsymbol{\phi} = \left\{ \sum_{n=1}^{N} [oldsymbol{x}_{oldsymbol{n}} \cdot oldsymbol{\gamma}_{oldsymbol{n}}^{ op}] iggl\} \odot rac{1.}{1.\left[\sum_{n=1}^{N} oldsymbol{\gamma}_{oldsymbol{n}}]^{ op}} \ oldsymbol{A} = [\sum_{n=2}^{N} oldsymbol{\xi}_{oldsymbol{n}}] \odot rac{1.}{1.\cdot 1.^{ op} oldsymbol{\gamma}_{oldsymbol{1}}} \ oldsymbol{\pi} = oldsymbol{\gamma}_{oldsymbol{1}} \odot rac{1.}{1.\cdot 1.^{ op} oldsymbol{\gamma}_{oldsymbol{1}}}$$

update $oldsymbol{ heta}^{k+1} \equiv \{oldsymbol{\phi}, oldsymbol{A}, oldsymbol{\pi}\}$

$$egin{aligned} oldsymbol{\phi} &\Leftarrow \left\{ \sum_{n=1}^{N} [oldsymbol{x}_{n} \cdot oldsymbol{\gamma}_{n}^{ op}]
ight\} \odot rac{1.}{1. \cdot 1.^{ op} \left\{ \sum_{n=1}^{N} [oldsymbol{x}_{n} \cdot oldsymbol{\gamma}_{n}^{ op}]
ight\}} &= \left\{ \sum_{n=1}^{N} [oldsymbol{x}_{n} \cdot oldsymbol{\gamma}_{n}^{ op}]
ight\} \odot rac{1.}{1. \cdot [\sum_{n=2}^{N} oldsymbol{\xi}_{n}]} &= [\sum_{n=2}^{N} oldsymbol{\xi}_{n}] \odot rac{1.}{1. \cdot [\sum_{n=2}^{N} oldsymbol{\gamma}_{n-1}]^{ op}} &= [\sum_{n=2}^{N} oldsymbol{\xi}_{n}] \odot rac{1.}{1. \cdot [\sum_{n=1}^{N-1} oldsymbol{\gamma}_{n}]^{ op}} &= oldsymbol{\gamma}_{1} \odot rac{1.}{1. \cdot 1.^{ op} oldsymbol{\gamma}_{1}} &= oldsymbol{\gamma}_{1} &= oldsymbol{\gamma}_{1} \odot rac{1.}{1. \cdot 1.^{ op} oldsymbol{\gamma}_{1}} &= oldsymbol{\gamma}_{1} &= oldsymbol{\gamma}_{1} \odot rac{1.}{1. \cdot 1.^{ op} oldsymbol{\gamma}_{1}} &= oldsymbol{\gamma}_{1} \odot rac{1.}{1. \cdot 1.^{ op} \odot oldsymbol{\gamma}_{1}} &= oldsymbol{\gamma}_{1} \odot oldsymbol{\gamma}_{1} \odot oldsymbol{\gamma}_{1} \odot oldsymbol{\gamma}_{1} \odot oldsymbol{\gamma}_{1} &= oldsymbol{\gamma}_{1} \odot oldsymbol{\gamma}$$

because

$$egin{aligned} 1.^ op [oldsymbol{x}_{oldsymbol{n}} \cdot oldsymbol{\gamma}_{oldsymbol{n}}^ op] &= (1.^ op oldsymbol{x}_{oldsymbol{n}}) \cdot oldsymbol{\gamma}_{oldsymbol{n}}^ op &= 1 \cdot oldsymbol{\gamma}_{oldsymbol{n}}^ op = oldsymbol{\gamma}_{oldsymbol{n}}^ op &= oldsymbol{\gamma}_{oldsymbol{n}-oldsymbol{1}}^ op &= oldsymbol{1}_{oldsymbol{n}-oldsymbol{1}}^ op &= oldsymbol{1}_{oldsymbol{n}-oldsymbol{1}_{oldsymbol{n}-oldsymbol{1}}^ op &= oldsymbol{1}_{oldsymbol{n}-old$$

Or example: Gaussian Mixed Model

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

$$egin{aligned} \ln oldsymbol{\phi}(oldsymbol{x_n})^{ op} &= egin{bmatrix} \ln \phi_1(x_n) \ \ln \phi_2(x_n) \ dots \ \ln \phi_K(x_n) \end{bmatrix} = egin{bmatrix} -rac{(x_n-\mu_1)^2}{2\sigma_1^2} - \ln(\sigma_1) - 0.5 \cdot \ln(2\pi) \ -rac{(x_n-\mu_2)^2}{2\sigma_2^2} - \ln(\sigma_2) - 0.5 \cdot \ln(2\pi) \ dots \ -rac{(x_n-\mu_K)^2}{2\sigma_K^2} - \ln(\sigma_K) - 0.5 \cdot \ln(2\pi) \end{bmatrix} \ &= -(x_n 1. - \mu) \odot (x_n 1. - \mu) \odot 0.5 rac{1.}{\sigma^2} - 0.5 \cdot \ln \sigma^2 - (0.5 \cdot \ln 2\pi) 1. \end{aligned}$$

we obtain that

for μ

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

$$= \frac{-1}{\sigma^2} \odot \sum_{n=1}^{N} \{ \boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \} = 0$$

$$\sum_{n=1}^{N} \{ \boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \}$$

$$= \{ \sum_{n=1}^{N} x_n \boldsymbol{\gamma_n} \} - \{ \sum_{n=1}^{N} \boldsymbol{\gamma_n} \} \odot \mu = 0$$

$$\Rightarrow \mu = \{ \sum_{n=1}^{N} x_n \boldsymbol{\gamma_n} \} \odot \frac{1.}{\{ \sum_{n=1}^{N} \boldsymbol{\gamma_n} \}}$$

for σ^2

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

$$= \frac{0.5.}{(\sigma^2)^2} \odot \sum_{n=1}^{N} \{ \boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \odot (x_n \cdot 1. - \mu) - \boldsymbol{\gamma_n} \odot \sigma^2 \} = 0$$

$$\sum_{n=1}^{N} \{ \boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \odot (x_n \cdot 1. - \mu) - \boldsymbol{\gamma_n} \odot \sigma^2 \}$$

$$= \{ \sum_{n=1}^{N} \boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \odot (x_n \cdot 1. - \mu) \} - \{ \sum_{n=1}^{N} \boldsymbol{\gamma_n} \} \odot \sigma^2 = 0$$

$$\Rightarrow \sigma^2 = \{ \sum_{n=1}^{N} \boldsymbol{\gamma_n} \odot (x_n \cdot 1. - \mu) \odot (x_n \cdot 1. - \mu) \} \odot \frac{1.}{\{ \sum_{n=1}^{N} \boldsymbol{\gamma_n} \}}$$

Specially

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

$$oldsymbol{\pi} \leftarrow \{\sum_{n=1}^N oldsymbol{\gamma_n}\} \odot rac{1.}{1.\cdot 1.^ op \{\sum_{n=1}^N oldsymbol{\gamma_n}\}} = rac{\{\sum_{n=1}^N oldsymbol{\gamma_n}\}}{N}$$