

why it could be defined as below:

 $z_n = [0,0,\ldots,1,\ldots,0]^T$ , represent probability of length = K states, only when  $z_{nk} = 1$ , it would be counted into p, random variable

$$egin{aligned} p\left(oldsymbol{z_1} \middle| oldsymbol{\pi}
ight) &= \prod_{k=1}^K \pi_k^{z_{1k}} \ p\left(oldsymbol{z_n} \middle| oldsymbol{z_{n-1}}, oldsymbol{A}
ight) &= \prod_{j=1}^K \prod_{k=1}^K A_{j,k}^{z_{n-1,j}z_{nk}} \ p\left(oldsymbol{x_n} \middle| oldsymbol{z_n}, oldsymbol{\phi}
ight) &= \prod_{j=1}^K \prod_{t=1}^{any\ dimension} \phi_j(t)^{z_{nj}x_n(t)} \end{aligned}$$

Define  $\vec{p}$  constant matrix

$$ec{p}(oldsymbol{z}_{n}|oldsymbol{ heta}) \equiv egin{bmatrix} p(oldsymbol{z}_{n}|oldsymbol{ heta}) oldsymbol{z}_{n} = \operatorname{state} 1 \ dots oldsymbol{p}(oldsymbol{z}_{n}|oldsymbol{ heta}) oldsymbol{z}_{n} = \operatorname{state} K \end{bmatrix} = egin{bmatrix} \sum_{oldsymbol{z}_{n}} p\left(oldsymbol{z}_{n}|oldsymbol{X}, oldsymbol{ heta} & \operatorname{old} \end{array}\right) oldsymbol{z}_{n} \end{bmatrix} \quad \operatorname{state} k : oldsymbol{z}_{n} = egin{bmatrix} 0 \ dots \ D \ oldsymbol{z}_{n} = oldsymbol{z} \\ D \ oldsymbol{z}_{n} = oldsymbo$$

where, for short: write  $\vec{p}(\pmb{z_n}) \equiv \vec{p}(\pmb{z_n}|\pmb{\theta}), \vec{p}(\pmb{x_n}) \equiv \vec{p}(\pmb{x_n}|\pmb{\theta})$ 

$$egin{aligned} ec{p}(oldsymbol{z_1}) &= oldsymbol{\pi} \ ec{p}(oldsymbol{z_n}) &= oldsymbol{A} ec{p}(oldsymbol{z_{n-1}}) \ ec{p}(oldsymbol{z_n}) &= oldsymbol{\phi} \cdot ec{p}(oldsymbol{z_n}) \end{aligned}$$

Size $(p(m{x_n}|m{\phi}))$  =  $any\ dimension imes K$  ,  $m{\phi} = [\phi_1,\phi_2,\ldots,\phi_j,\ldots\phi_K]$ 

 $\phi_k$  is probability distribution sequence of k th state,  $\emph{any dimension} imes 1$ 

 $oldsymbol{x_n}$  is probability distribution sequence,  $\emph{any dimension} imes 1$ 

 $\boldsymbol{z_n}$  is probability distribution sequence,  $K \times 1$ 

In P(z,theta), z is observation value;

In  $z_n = A z_{n-1}$ , z is random variable;

Thus

$$p\left(oldsymbol{x_n} | oldsymbol{z_n}, oldsymbol{\phi}
ight) = \prod_{j=1}^{K} \prod_{t=1}^{any\ dimension} \phi_j(t)^{z_{nj}x_n(t)}$$

## Goal

$$egin{aligned} p(x_1,\dots,x_n,z_1,\dots,z_n|m{\pi},m{A},m{\phi}) &= p(x_1,\dots,x_n,z_1,\dots,z_n|m{ heta}) \ &= p(x_n|x_1,\dots,x_{n-1},z_1,\dots,z_n,m{ heta}) \cdot p(x_1,\dots,x_{n-1},z_1,\dots,z_n|m{ heta}) \ &= [p(x_n|z_n,m{\phi})] \cdot p(x_1,\dots,x_{n-1},z_1,\dots,z_n|m{ heta}) \ &= [p(x_n|z_n,m{\phi})] \cdot p(z_n|x_1,\dots,x_{n-1},z_1,\dots,z_{n-1},m{ heta}) \cdot p(x_1,\dots,x_{n-1},z_1,\dots,z_{n-1}|m{ heta}) \ &= [p(x_n|z_n,m{\phi})] \cdot [p(z_n|z_{n-1},m{A})] \cdot p(x_1,\dots,x_{n-1},z_1,\dots,z_{n-1}|m{ heta}) \end{aligned}$$

thus we get

$$p(x_1, \ldots, x_n, z_1, \ldots, z_n | \boldsymbol{\theta}) = [p(x_n | z_n, \boldsymbol{\phi})] \cdot [p(z_n | z_{n-1}, \boldsymbol{A})] \cdot p(x_1, \ldots, x_{n-1}, z_1, \ldots, z_{n-1} | \boldsymbol{\theta})$$

So

$$p(oldsymbol{X},oldsymbol{Z}|oldsymbol{ heta}) \equiv p(oldsymbol{x_1},\ldots,oldsymbol{x_N},oldsymbol{z_1},\ldots,oldsymbol{z_N}|oldsymbol{ heta}) = [\prod_{n=1}^N p(oldsymbol{x_n}|oldsymbol{z_n},oldsymbol{\phi})] \cdot [\prod_{n=2}^N p(oldsymbol{z_n}|oldsymbol{z_{n-1}},oldsymbol{A})] \cdot p\left(oldsymbol{z_1}|oldsymbol{\pi}
ight)$$

Here

$$egin{aligned} p\left(oldsymbol{x_n}|oldsymbol{z_n},oldsymbol{\phi}
ight) &= \prod_{j=1}^K \prod_{t=1}^{any} \phi_j(t)^{z_{nj}x_n(t)} \ p\left(oldsymbol{z_n}|oldsymbol{z_{n-1}},oldsymbol{A}
ight) &= \prod_{j=1}^K \prod_{k=1}^K A_{k,j}^{z_{n-1,j}z_{nk}} \ p\left(oldsymbol{z_1}|oldsymbol{\pi}
ight) &= \prod_{k=1}^K \pi_k^{z_{1k}} \end{aligned}$$

So

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

if  $m{t}$  is continuous

$$<\ln oldsymbol{\phi}, oldsymbol{x_n}> = \int_{-\infty}^{+\infty} \ln oldsymbol{\phi}(oldsymbol{t}) \cdot x_n(t) dt \ oldsymbol{\phi} = egin{bmatrix} \phi_1(t) \ \phi_2(t) \ dots \ \phi_K(t) \end{bmatrix}$$

if m is discrete

$$\langle \ln oldsymbol{\phi}, oldsymbol{x_n} 
angle = \ln oldsymbol{\phi}^T \cdot oldsymbol{x_n} \ oldsymbol{\phi} = egin{bmatrix} \phi_{11} & \cdots & \phi_{1m} & \cdots \phi_{1M} \ \phi_{21} & \cdots & \phi_{2m} & \cdots \phi_{2M} \ dots & & dots \ \phi_{K1} & \cdots & \phi_{Km} & \cdots \phi_{KM} \end{bmatrix}^T M imes K ext{matrix} \ oldsymbol{x_n} : M imes 1 ext{ matrix}$$

Only  $\boldsymbol{x_n}$  is known,  $n \in \{1, \dots, N\}$ 

for example X is random variable, then  $Y\equiv f(X)$  is another variable, then  $P(X=x)\equiv p(x)$  then

$$egin{aligned} P(Y=y)dy &\equiv P(X=x)dx = p(x)dx, \quad y = f(x) \ &\Longrightarrow P(Y=y) = p(x)rac{dx}{df(x)} \ &\Longrightarrow E[Y] = \int_{-\infty}^{+\infty} y\cdot P(Y=y)dy = \int_{-\infty}^{+\infty} f(x)p(x)dx \end{aligned}$$

期望最大化算法,或者EM算法

$$\max \quad \ln p(\boldsymbol{X}|\boldsymbol{\theta}) = \mathcal{L}(q,\boldsymbol{\theta}) + \mathrm{KL}(q||p)$$

where

$$egin{aligned} \mathcal{L}(q,m{ heta}) &= \sum_{m{Z}} q(m{Z}) \ln \Big\{ rac{p(m{X},m{Z}|m{ heta})}{q(m{Z})} \Big\} \ \mathrm{KL}(q\|p) &= -\sum_{m{Z}} q(m{Z}) \ln \Big\{ rac{p(m{Z}|m{X},m{ heta})}{q(m{Z})} \Big\} \end{aligned}$$

Where

- q(Z) is  $\forall$  arbitrary distribution for Z
- $\mathcal{L}(q, \boldsymbol{\theta})$  概率分布 $q(\mathbf{Z})$ 的一个泛函,  $\boldsymbol{\theta}$  函数
- $\mathrm{KL}(q||p)$  概率分布q(Z)的一个泛函,  $\theta$  函数,KL散度 of  $p(Z|X,\theta), q(Z), \geq 0$

So

$$egin{aligned} & \ln p(m{X}|m{ heta}) \geq \mathcal{L}(q,m{ heta}) = \sum_{m{Z}} q(m{Z}) \ln iggl\{ rac{p(m{X},m{Z}|m{ heta})}{q(m{Z})} iggr\}, \quad orall q(m{Z}) \ & ext{if } q(m{Z}) \Leftarrow p(m{Z}|m{X},m{ heta}), ext{here } \mathrm{KL}(q||p) = 0 \ & \ln p(m{X}|m{ heta}) = \mathcal{L}(q,m{ heta}) = \sum_{m{Z}} p(m{Z}|m{X},m{ heta}) \ln iggl\{ rac{p(m{X},m{Z}|m{ heta})}{p(m{Z}|m{X},m{ heta})} iggr\} \end{aligned}$$

**notice:** X is fixed, Z,  $\theta$  are variables, 转而用q(Z),  $\theta$  替代;

调整法

$$\ln p(\boldsymbol{X}|\boldsymbol{ heta}) = \mathcal{L}(q, \boldsymbol{ heta}) + \mathrm{KL}(q|p)$$

1. fix heta, change q(Z) [E步骤]

$$egin{aligned} q(oldsymbol{Z}) &\Leftarrow p(oldsymbol{Z}|oldsymbol{X},oldsymbol{ heta}) \ \ln p(oldsymbol{X}|oldsymbol{ heta}) &= \mathcal{L}(q,oldsymbol{ heta}) + \mathrm{KL}(q|p) \downarrow \Leftarrow 0, \quad \mathrm{update} \ \mathcal{L}(q,oldsymbol{ heta}) \uparrow \ \ln p(oldsymbol{X}|oldsymbol{ heta}) &== \mathcal{L}(q,oldsymbol{ heta}) \ oldsymbol{ heta} \ oldsymbol{ heta} \end{array}$$

2. fix q(Z), change heta [M步骤]

ignore influence of 
$$heta$$
 in  $\mathrm{KL}(q|p) = -\sum_{m{Z}} q(m{Z}) \ln \Big\{ rac{p(m{Z}|m{X},m{ heta})}{q(m{Z})} \Big\}$  
$$\mathcal{L}(q,m{ heta}) = \sum_{m{Z}} q(m{Z}) \ln \Big\{ rac{p(m{X},m{Z}|m{ heta})}{q(m{Z})} \Big\}$$
 
$$= \sum_{m{Z}} p(m{Z}|m{X},m{ heta}^{\mathrm{old}}) \ln \Big\{ rac{p(m{X},m{Z}|m{ heta})}{p(m{Z}|m{X},m{ heta}^{\mathrm{old}})} \Big\}$$
 
$$= \sum_{m{Z}} p\left(m{Z}|m{X},m{ heta}^{\mathrm{old}}\right) \ln p(m{X},m{Z}|m{ heta}) - \mathrm{const}$$

set

$$oldsymbol{Q}\left(oldsymbol{ heta},oldsymbol{ heta}^{
m old}
ight) \equiv \sum_{oldsymbol{Z}} p\left(oldsymbol{Z} | oldsymbol{X},oldsymbol{ heta}^{
m old}
ight) \ln p(oldsymbol{X},oldsymbol{Z} | oldsymbol{ heta})$$

get

$$egin{aligned} oldsymbol{ heta} & lpha rgmax \mathcal{L}(q, \overline{oldsymbol{ heta}}) \ & = rgmax oldsymbol{Q} \left( \overline{oldsymbol{ heta}}, oldsymbol{ heta} 
ight. ^{
m old} 
ight. \ & ext{update } \mathcal{L}(q, oldsymbol{ heta}) \uparrow \end{aligned}$$

update

$$egin{aligned} \operatorname{KL}(q|p) \uparrow &= -\sum_{m{Z}} q(m{Z}) \ln \left\{ rac{p(m{Z}|m{X},m{ heta})}{q(m{Z})} 
ight\} > 0 \ & \ln p(m{X}|m{ heta}) \uparrow &= \mathcal{L}(q,m{ heta}) \uparrow + \operatorname{KL}(q|p) \uparrow \end{aligned}$$

## Conclusion

因为1,2中  $\mathcal{L}(q,\theta)$  个均上升;

$$2 \oplus \ln p(\boldsymbol{X}|\boldsymbol{\theta}) \uparrow = \mathcal{L}(q,\boldsymbol{\theta}) \uparrow + \mathrm{KL}(q|p) \uparrow$$

• 所以在一个循环1,2中 $\ln p(\boldsymbol{X}|\boldsymbol{\theta})$  个下界  $\mathcal{L}(q,\boldsymbol{\theta})$  个

## consider $Q(\theta^{old}, \theta)$

Use 
$$\sum$$
 to delete variables,  $\sum_{z_1} p(z_1,z_2) = P(z_2)$ 

 $\boldsymbol{X}$  is fix observation value;

Z is [probability distribution] random variable,  $\ln p(X,Z| heta)$  is random variable

 $\sum_{m{Z}} p\left(m{Z}|m{X}, m{ heta}^{
m old}
ight.)$  is [probability distribution] random variable **of**  $m{Z}$  [probability distribution] random variable

$$\begin{split} Q\left(\boldsymbol{\theta},\boldsymbol{\theta}^{\text{ old }}\right) &\equiv \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \ln p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta}) \\ &= \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \left\{\sum_{n=1}^{N} \boldsymbol{z}_{n}^{T} < \ln \boldsymbol{\phi}, \boldsymbol{x}_{n} > + \sum_{n=2}^{N} \boldsymbol{z}_{n-1}^{T} \ln \boldsymbol{A}^{T} \cdot \boldsymbol{z}_{n} + \ln \boldsymbol{\pi}^{T} \cdot \boldsymbol{z}_{1}\right\} \\ &= \left\{\sum_{\boldsymbol{z}_{n}} p\left(\boldsymbol{z}_{n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right)\sum_{n=1}^{N} \boldsymbol{z}_{n}^{T} < \ln \boldsymbol{\phi}, \boldsymbol{x}_{n} > \right\} \\ &+ \left\{\sum_{\boldsymbol{z}_{n-1},\boldsymbol{z}_{n}} p\left(\boldsymbol{z}_{n-1},\boldsymbol{z}_{n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right)\sum_{n=2}^{N} \boldsymbol{z}_{n-1}^{T} \ln \boldsymbol{A}^{T} \cdot \boldsymbol{z}_{n}\right\} \\ &+ \left\{\sum_{\boldsymbol{z}_{1}} p\left(\boldsymbol{z}_{1}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \boldsymbol{z}_{1}^{T} \ln \boldsymbol{\pi}\right\} \\ &= \left\{\sum_{n=1}^{N} \left[\sum_{\boldsymbol{z}_{n}} p\left(\boldsymbol{z}_{n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \boldsymbol{z}_{n}^{T}\right] < \ln \boldsymbol{\phi}, \boldsymbol{x}_{n} > \right\} \\ &+ \left\{\left[\sum_{\boldsymbol{z}_{1}} p\left(\boldsymbol{z}_{1}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \boldsymbol{z}_{1}^{T}\right] \ln \boldsymbol{\pi}\right\} \\ &= \sum_{n=1}^{N} \boldsymbol{\gamma}_{n}^{T} < \ln \boldsymbol{\phi}, \boldsymbol{x}_{n} > + \sum_{n=2}^{N} \operatorname{tr}\left(\boldsymbol{\xi}_{n}^{T} \ln \boldsymbol{A}\right) + \boldsymbol{\gamma}_{1}^{T} \ln \boldsymbol{\pi} \end{split}$$

where

$$egin{aligned} &\sum_{oldsymbol{z}_{n-1},oldsymbol{z}_n} p\left(oldsymbol{z}_{n-1},oldsymbol{z}_n|oldsymbol{X},oldsymbol{ heta}^{
m old}
ight) \sum_{n=2}^N oldsymbol{z}_{n-1}^T \ln oldsymbol{A}^T \cdot oldsymbol{z}_n \ &= \sum_{n=2}^N \operatorname{tr}\left(p\left(oldsymbol{z}_{n-1},oldsymbol{z}_n|oldsymbol{X},oldsymbol{ heta}^{
m old}
ight) \left[oldsymbol{z}_{n-1}^T \ln oldsymbol{A}^T \cdot oldsymbol{z}_n
ight]^T
ight) \ &= \sum_{n=2}^N \operatorname{tr}\left(\left[oldsymbol{z}_{n-1},oldsymbol{z}_n|oldsymbol{X},oldsymbol{ heta}^{
m old}
ight) oldsymbol{z}_n^T
ight] \cdot \ln oldsymbol{A}
ight) \ &= \sum_{n=2}^N \sum_{n=2}^N \left\{\left[oldsymbol{z}_np\left(oldsymbol{z}_{n-1},oldsymbol{z}_n|oldsymbol{X},oldsymbol{ heta}^{
m old}
ight) oldsymbol{z}_{n-1}^T
ight] \odot \ln oldsymbol{A}
ight\} \end{aligned}$$

其中, 因为X,  $\theta$ <sup>old</sup> 分别为已知、固定 在 M步骤

$$m{\gamma_n} \equiv \left[\sum_{m{z_n}} p\left(m{z_n} | m{X}, m{ heta}^{
m old} 
ight) m{z_n} \right]$$
 为常数,是期望 $E[m{z_n} | m{X}, m{ heta}^{
m old} 
ight]$ ,大小 $K imes 1$   $m{\xi_n} \equiv \left[m{z_n} p\left(m{z_{n-1}}, m{z_n} | m{X}, m{ heta}^{
m old} 
ight) m{z_{n-1}}^T \right]$  为常数,是期望 $E[m{z_n} m{z_{n-1}}^T | m{X}, m{ heta}^{
m old} 
ight]$ ,大小 $K imes K$ 下面求解

$$\begin{split} \max_{\boldsymbol{\theta}} \quad \boldsymbol{Q} \left( \boldsymbol{\theta}, \boldsymbol{\theta}^{\text{ old }} \right) &\equiv \sum_{\boldsymbol{Z}} p \left( \boldsymbol{Z} | \boldsymbol{X}, \boldsymbol{\theta}^{\text{ old }} \right) \ln p(\boldsymbol{X}, \boldsymbol{Z} | \boldsymbol{\theta}) \\ &= \sum_{n=1}^{N} \boldsymbol{\gamma}_{n}^{T} < \ln \boldsymbol{\phi}, \boldsymbol{x}_{n} > + \sum_{n=2}^{N} \operatorname{tr} \left( \boldsymbol{\xi}_{n}^{T} \ln \boldsymbol{A} \right) + \boldsymbol{\gamma}_{1}^{T} \ln \boldsymbol{\pi} \\ &\text{s.t.} \quad \boldsymbol{\gamma}_{n}^{T} \cdot 1. = 1 \quad \operatorname{tr} \left( \boldsymbol{\xi}_{n}^{T} \cdot 1. \right) = 1 \\ &1.^{T} \cdot \boldsymbol{\phi} = 1.^{T} \quad 1.^{T} \cdot \boldsymbol{A} = 1.^{T} \quad 1.^{T} \cdot \boldsymbol{\pi} = 1 \end{split}$$

with Lagrange method:

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

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

if t is discrete,  $any\ dimenstion = M$ 

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

 $<\ln oldsymbol{\phi}, oldsymbol{x_n}> = \ln oldsymbol{\phi}^T \cdot oldsymbol{x_n}$ 

$$<\boldsymbol{\phi},\delta(t)>=\boldsymbol{\phi}^T\cdot\delta(t)$$

$$\begin{split} \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{\phi}} &= \left\{ \sum_{n=1}^{N} [\boldsymbol{\gamma}_{n} \cdot \boldsymbol{x}_{n}^{T}] \right\}^{T} \odot \frac{1}{\boldsymbol{\phi}} - \sum_{t=1}^{M} u_{t} \cdot 1.\delta(t)^{T} \\ &= \left\{ \sum_{n=1}^{N} [\boldsymbol{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{T}] \right\} \odot \frac{1}{\boldsymbol{\phi}} - \begin{bmatrix} u_{1} & \cdots & u_{t} \cdots & u_{M} \\ u_{1} & \cdots & u_{t} \cdots & u_{M} \\ \vdots & & & \vdots \\ u_{1} & \cdots & u_{t} \cdots & u_{M} \end{bmatrix} = 0 \\ \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{A}} &= [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}] \odot \frac{1}{\boldsymbol{A}} - \sum_{k=1}^{K} v_{k} \cdot 1.\sigma_{k}^{T} \\ &= [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}] \odot \frac{1}{\boldsymbol{A}} - \begin{bmatrix} v_{1} & \cdots & v_{k} \cdots & v_{K} \\ v_{1} & \cdots & v_{k} \cdots & v_{K} \\ \vdots & & & \vdots \\ v_{1} & \cdots & v_{k} \cdots & v_{K} \end{bmatrix} = 0 \\ \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{\pi}} &= \boldsymbol{\gamma}_{1} \odot \frac{1}{\boldsymbol{\pi}} - w_{1} \cdot 1. = 0 \end{split}$$

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

To sum up

$$egin{aligned} oldsymbol{\phi} &= \left\{ \sum_{n=1}^{N} [oldsymbol{x}_{oldsymbol{n}} \cdot oldsymbol{\gamma}_{oldsymbol{n}}^T] 
ight\} \odot rac{1}{1.\left[\sum_{n=1}^{N} oldsymbol{\gamma}_{oldsymbol{n}}\right]^T} \ oldsymbol{A} &= [\sum_{n=2}^{N} oldsymbol{\xi}_{oldsymbol{n}}] \odot rac{1}{1.\cdot 1.^T [\sum_{n=2}^{N} oldsymbol{\xi}_{oldsymbol{n}}]} \ oldsymbol{\pi} &= oldsymbol{\gamma}_{oldsymbol{1}} \odot rac{1}{1.\cdot 1.^T oldsymbol{\gamma}_{oldsymbol{1}}} \end{aligned}$$

How to calculate  $\gamma_n, \xi_n$ 

similarly get:

$$egin{aligned} oldsymbol{\gamma_n} &\equiv ec{p}(oldsymbol{z_n} | oldsymbol{X}, oldsymbol{ heta}^{
m old} \ ) \equiv egin{bmatrix} p(oldsymbol{z_n} | oldsymbol{X}, oldsymbol{ heta}^{
m old} \ ) & oldsymbol{z_n} = {
m state} \ K \end{bmatrix} \ oldsymbol{\xi_n} &\equiv ec{p} \left(oldsymbol{z_{n-1}}, oldsymbol{z_n} | oldsymbol{X}, oldsymbol{ heta}^{
m old} \ ) & oldsymbol{z_n} = {
m state} \ i, oldsymbol{z_{n-1}} = {
m state} \ j \end{bmatrix} \ & ext{the same size as} \quad oldsymbol{z_n} oldsymbol{z_{n-1}} \end{aligned}$$

$$\begin{split} &\gamma_n \equiv \vec{p}(z_n|X|\theta^{\text{old}}) \\ &= \vec{p}(z_n,X|\theta^{\text{old}}) \odot \frac{1}{\vec{p}(X|\theta^{\text{old}})} \\ &= \vec{p}(z_n,x_1,\ldots,x_n|\theta^{\text{old}}) \odot \vec{p}(x_{n+1},\ldots,x_N|z_n,x_1,\ldots,x_n,\theta^{\text{old}}) \odot \frac{1}{\vec{p}(X|\theta^{\text{old}})} \\ &= \vec{p}(z_n,x_1,\ldots,x_n|\theta^{\text{old}}) \odot \vec{p}(x_{n+1},\ldots,x_N|z_n,\theta^{\text{old}}) \odot \frac{1}{\vec{p}(X|\theta^{\text{old}})} \\ &= \alpha_n \odot \beta_n \odot \frac{1}{\vec{p}(X|\theta^{\text{old}})} \\ &\leq n \equiv \vec{p}(z_{n-1},z_n|X,\theta^{\text{old}}) \\ &= \vec{p}(z_{n-1},z_n|X,\theta^{\text{old}}) \odot \frac{1}{\vec{p}(X|\theta^{\text{old}})} \\ &= \vec{p}(z_{n-1},z_n,\ldots,x_{n-1}|\theta^{\text{old}}) \odot \vec{p}(z_n,x_n,\ldots,x_{n-1}|\theta^{\text{old}}) \\ &\odot \vec{p}(z_n,x_{n-1},\ldots,x_{n-1}|\theta^{\text{old}}) \\ &\odot \vec{p}(z_n,x_{n-1},\alpha^{\text{old}}) \\ &\odot \vec{p}(z_n,x_{n-1},\theta^{\text{old}}) \\ &\odot \vec{p}(z_n,x_{n-1},\theta^{\text{old}}) \\ &\odot \vec{p}(z_n,x_{n-1},\theta^{\text{old}}) \\ &\odot \vec{p}(z_n,x_n,\alpha_n|z_n,\alpha^{\text{old}}) \\ &\odot \vec{p}(z_n,x_n,\alpha^{\text{old}}) \\ &\odot \vec{p}(z_n,x_n,\theta^{\text{old}}) \\ &\odot \vec{$$

其中
$$p(m{X}|m{ heta}^{
m old}\ ) = \sum_{Z_N} p(m{z_N},m{X}|m{ heta}^{
m old}\ ) = \sum m{lpha_N} = 1.^T \cdot m{lpha_N} ext{ is constant}$$
  $ec{p}(m{X}|m{ heta}^{
m old}\ ) = p(m{X}|m{ heta}^{
m old}\ ) \cdot 1.$   $p(m{X}|m{ heta}^{
m old}\ ) = 1.^T \cdot p(m{z_N},m{X}|m{ heta}^{
m old}\ )$   $= 1.^T \cdot m{lpha_N}$ 

So if  $\alpha_n, \beta_n$  is obtained, we could get  $\gamma_n, \xi_n$ 

$$oldsymbol{lpha_n} \equiv ec{p}(oldsymbol{z_n}, oldsymbol{x_1}, \dots, oldsymbol{x_n} | oldsymbol{ heta}^{
m old} \ ) \ oldsymbol{eta_n} \equiv ec{p}(oldsymbol{x_{n+1}}, \dots, oldsymbol{x_N} | oldsymbol{z_n}, oldsymbol{ heta}^{
m old} \ )$$

How to obtain  $\alpha_n, \beta_n$ 

for  $\alpha_n$ , initial value:

$$egin{aligned} oldsymbol{lpha_1} &= ec{p}(oldsymbol{z_1}, oldsymbol{x_1} | oldsymbol{ heta} ^{ ext{old}}) &= ec{p}(oldsymbol{x_1} | oldsymbol{z_1}, oldsymbol{ heta} ^{ ext{old}}) & \odot ec{p}(oldsymbol{z_1} | oldsymbol{ heta} ^{ ext{old}}) &= [oldsymbol{\phi}^T \cdot oldsymbol{x_1}] \odot oldsymbol{\pi} \end{aligned}$$

Recurrence relation

$$\begin{split} & \boldsymbol{\alpha}_{\boldsymbol{n}} \equiv \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{\theta}^{\text{ old }}) \\ & = \vec{p}(\boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \\ & = \vec{p}(\boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{z}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \\ & = \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{\theta}^{\text{ old }}) \\ & = \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{ \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{\theta}^{\text{ old }}) \cdot 1. \right\} \\ & = \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{ \left[ \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \right] \cdot 1. \right\} \\ & = \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{ \left[ (1 \cdot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-1} | \boldsymbol{\theta}^{\text{ old }})^{T}) \odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{\theta}^{\text{ old }}) \odot \left( 1 \cdot \vec{p}(\boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{z}^{\text{ old }})^{T} \right) \right] \cdot 1. \right\} \\ & = \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{ \left[ \vec{p}(\boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{\theta}^{\text{ old }}) \odot \left( 1 \cdot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{\theta}^{\text{ old }})^{T} \right) \right] \cdot 1. \right\} \\ & = \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{ \left[ \vec{p}(\boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{\theta}^{\text{ old }}) \odot \left( 1 \cdot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{\theta}^{\text{ old }})^{T} \right) \right] \cdot 1. \right\} \\ & = [\boldsymbol{\phi}^{T} \cdot \boldsymbol{x}_{\boldsymbol{n}}] \odot \left\{ \left[ \boldsymbol{A} \odot \left( 1 \cdot \boldsymbol{\alpha}_{\boldsymbol{n}-1}^{T} \right) \right] \cdot 1. \right\} \\ & = [\boldsymbol{\phi}^{T} \cdot \boldsymbol{x}_{\boldsymbol{n}}] \odot [\boldsymbol{A} \cdot \boldsymbol{\alpha}_{\boldsymbol{n}-1}] \end{split}{1} \end{split}{2} . \end{split}{2} . \end{split}{2} . \end{split}{2} . \end{split}{2} . \end{split}{2} . \boldsymbol{\alpha}_{\boldsymbol{n}-1} \boldsymbol{\alpha}_{\boldsymbol$$

for  $\beta_n$ , last initial value:

$$oldsymbol{eta_N} = ec{p}(1|oldsymbol{z_N},oldsymbol{ heta}^{
m old}) = 1$$

Recurrence relation

$$\begin{split} \boldsymbol{\beta}_{n} &\equiv \vec{p}(\boldsymbol{x}_{n+1}, \dots, \boldsymbol{x}_{N} | \boldsymbol{z}_{n}, \boldsymbol{\theta}^{\text{ old }}) \\ &= \left\{ \vec{p}(\boldsymbol{z}_{n+1}, \boldsymbol{x}_{n+1}, \dots, \boldsymbol{x}_{N} | \boldsymbol{z}_{n}, \boldsymbol{\theta}^{\text{ old }})^{T} \cdot 1. \right\} \\ &= \left\{ \left[ \vec{p}(\boldsymbol{z}_{n+1} | \boldsymbol{z}_{n}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{n+1}, \dots, \boldsymbol{x}_{N} | \boldsymbol{z}_{n}, \boldsymbol{z}_{n+1}, \boldsymbol{\theta}^{\text{ old }}) \right]^{T} \cdot 1. \right\} \\ &= \left\{ \left[ \vec{p}(\boldsymbol{z}_{n+1} | \boldsymbol{z}_{n}, \boldsymbol{\theta}^{\text{ old }}) \odot \left( \vec{p}(\boldsymbol{x}_{n+1}, \dots, \boldsymbol{x}_{N} | \boldsymbol{z}_{n+1}, \boldsymbol{\theta}^{\text{ old }}) \cdot 1.^{T} \right) \right]^{T} \cdot 1. \right\} \\ &= \left\{ \left[ \vec{p}(\boldsymbol{z}_{n+1} | \boldsymbol{z}_{n}, \boldsymbol{\theta}^{\text{ old }}) \odot \left( \left[ \vec{p}(\boldsymbol{x}_{n+1} | \boldsymbol{z}_{n+1}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{n+2}, \dots, \boldsymbol{x}_{N} | \boldsymbol{z}_{n+1}, \boldsymbol{\theta}^{\text{ old }}) \right] \cdot 1.^{T} \right) \right]^{T} \cdot 1. \right\} \\ &= \left\{ \left[ \vec{p}(\boldsymbol{z}_{n+1} | \boldsymbol{z}_{n}, \boldsymbol{\theta}^{\text{ old }}) \odot \left( \left[ \vec{p}(\boldsymbol{x}_{n+1} | \boldsymbol{z}_{n+1}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{n+2}, \dots, \boldsymbol{x}_{N} | \boldsymbol{z}_{n+1}, \boldsymbol{\theta}^{\text{ old }}) \right] \cdot 1.^{T} \right) \right]^{T} \cdot 1. \right\} \\ &= \left\{ \left[ \boldsymbol{A} \odot \left( \left[ \left[ \boldsymbol{\phi}^{T} \cdot \boldsymbol{x}_{n+1} \right] \odot \boldsymbol{\beta}_{n+1} \right] \cdot 1.^{T} \right) \right]^{T} \cdot 1. \right\} \\ &= \boldsymbol{A}^{T} \cdot \left( \left[ \boldsymbol{\phi}^{T} \cdot \boldsymbol{x}_{n+1} \right] \odot \boldsymbol{\beta}_{n+1} \right) \right\} \end{split}$$

To sum up

$$egin{aligned} oldsymbol{lpha_1} &= [oldsymbol{\phi}^T \cdot oldsymbol{x_1}] \odot oldsymbol{\pi} \ oldsymbol{lpha_n} &= [oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot [oldsymbol{A} \cdot oldsymbol{lpha_{n-1}}] \ oldsymbol{eta_N} &= oldsymbol{A}^T \cdot \left( [oldsymbol{\phi}^T \cdot oldsymbol{x_{n+1}}] \odot oldsymbol{eta_{n+1}} 
ight) \end{aligned}$$

Then

$$egin{aligned} oldsymbol{\gamma_n} &= oldsymbol{lpha_n} \odot oldsymbol{eta_n} \odot oldsymbol{eta_n} \odot rac{1}{ec{p}(oldsymbol{X} | oldsymbol{ heta} ext{ old })} \ &= rac{1}{1.^T \cdot oldsymbol{lpha_N}} igl[ oldsymbol{lpha_n} \odot oldsymbol{eta_n} igr] & oldsymbol{lpha_{n-1}} igr\} \odot oldsymbol{A} \ &= rac{1}{1.^T \cdot oldsymbol{lpha_N}} iggl\{ igl[ oldsymbol{\phi}^T \cdot oldsymbol{x_n} igr] \odot oldsymbol{eta_n} igr] \cdot oldsymbol{lpha_{n-1}} igr\} \odot oldsymbol{A} \end{aligned}$$

because  $\alpha_N, oldsymbol{eta_1}$  are so small

we divide  $p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})$  into 2 part

$$p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) = p(\boldsymbol{x_1}, \cdots, \boldsymbol{x_n}|\boldsymbol{\theta}^{\text{ old }}) p(\boldsymbol{x_{n+1}}, \cdots, \boldsymbol{x_N}|\boldsymbol{x_1}, \cdots, \boldsymbol{x_n}, \boldsymbol{\theta}^{\text{ old }})$$

define

$$egin{aligned} \hat{oldsymbol{lpha}}_{oldsymbol{n}} &\equiv rac{1}{p(oldsymbol{x_1}, \cdots, oldsymbol{x_n} | oldsymbol{ heta}^{
m old})} oldsymbol{lpha}_{oldsymbol{n}} &= egin{aligned} \prod_{k=1}^n p(oldsymbol{x_k} | oldsymbol{x_1}, \cdots, oldsymbol{x_{k-1}}, oldsymbol{ heta}^{
m old}) \end{aligned} egin{aligned} oldsymbol{lpha}_{oldsymbol{n}} &= \left\{ \prod_{k=1}^n c_k \right\} oldsymbol{lpha}_{oldsymbol{n}} &= 1.^T \hat{oldsymbol{lpha}}_{oldsymbol{n}} &= \sum_{oldsymbol{z_n}} p(oldsymbol{z_n} | oldsymbol{x_1}, \cdots, oldsymbol{x_n}, oldsymbol{ heta}^{
m old}) = 1 \end{aligned}$$

the other way

$$\begin{aligned} c_{n}\hat{\boldsymbol{\alpha}}_{n} &= p(\boldsymbol{x}_{n}|\boldsymbol{x}_{1},\cdots,\boldsymbol{x}_{n-1},\boldsymbol{\theta}^{\text{ old }})\vec{p}(\boldsymbol{z}_{n}|\boldsymbol{x}_{1},\cdots,\boldsymbol{x}_{n},\boldsymbol{\theta}^{\text{ old }}) \\ &= \vec{p}(\boldsymbol{z}_{n},\boldsymbol{x}_{n}|\boldsymbol{x}_{1},\cdots,\boldsymbol{x}_{n-1},\boldsymbol{\theta}^{\text{ old }}) \\ &= \vec{p}(\boldsymbol{z}_{n}|\boldsymbol{x}_{1},\cdots,\boldsymbol{x}_{n-1},\boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{n}|\boldsymbol{z}_{n}\boldsymbol{\theta}^{\text{ old }}) \\ &= [\vec{p}(\boldsymbol{z}_{n},\boldsymbol{z}_{n-1}|\boldsymbol{x}_{1},\cdots,\boldsymbol{x}_{n-1},\boldsymbol{\theta}^{\text{ old }}) \cdot 1.] \odot \vec{p}(\boldsymbol{x}_{n}|\boldsymbol{z}_{n}\boldsymbol{\theta}^{\text{ old }}) \\ &= [[1.\cdot\vec{p}(\boldsymbol{z}_{n-1}|\boldsymbol{x}_{1},\cdots,\boldsymbol{x}_{n-1},\boldsymbol{\theta}^{\text{ old }})^{T}] \odot \vec{p}(\boldsymbol{z}_{n}|\boldsymbol{z}_{n-1}) \cdot 1.] \odot \vec{p}(\boldsymbol{x}_{n}|\boldsymbol{z}_{n}\boldsymbol{\theta}^{\text{ old }}) \\ &= [\vec{p}(\boldsymbol{z}_{n}|\boldsymbol{z}_{n-1})\vec{p}(\boldsymbol{z}_{n-1}|\boldsymbol{x}_{1},\cdots,\boldsymbol{x}_{n-1},\boldsymbol{\theta}^{\text{ old }})^{T}] \odot \vec{p}(\boldsymbol{x}_{n}|\boldsymbol{z}_{n}\boldsymbol{\theta}^{\text{ old }}) \\ &= [\boldsymbol{A}\cdot\hat{\boldsymbol{\alpha}}_{n-1}]\vec{p}(\boldsymbol{z}_{n-1}|\boldsymbol{x}_{1},\cdots,\boldsymbol{x}_{n-1},\boldsymbol{\theta}^{\text{ old }})] \odot \vec{p}(\boldsymbol{x}_{n}|\boldsymbol{z}_{n}\boldsymbol{\theta}^{\text{ old }}) \\ &= [\boldsymbol{\phi}^{T}\cdot\boldsymbol{x}_{n}] \odot [\boldsymbol{A}\cdot\hat{\boldsymbol{\alpha}}_{n-1}] \\ &= [\boldsymbol{\phi}^{T}\cdot\boldsymbol{x}_{n}] \odot [\boldsymbol{A}\cdot\hat{\boldsymbol{\alpha}}_{n-1}] \end{aligned}$$

SO

$$egin{aligned} \hat{oldsymbol{lpha}}_{oldsymbol{n}} & = rac{1}{c_n} [oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot [oldsymbol{A} \cdot \hat{oldsymbol{lpha}}_{n-1}] \ & = rac{1}{1.^T \cdot \{ [oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot [oldsymbol{A} \cdot \hat{oldsymbol{lpha}}_{n-1}] \}} [oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot oldsymbol{\pi} \ & \hat{oldsymbol{lpha}}_1 & = rac{1}{1.^T \cdot \{ [oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot oldsymbol{\pi} \}} [oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot oldsymbol{\pi} \end{aligned}$$

thesame

$$egin{aligned} rac{\hat{oldsymbol{eta}}_{oldsymbol{n}}}{\hat{oldsymbol{eta}}_{oldsymbol{n+1}}} &\equiv rac{oldsymbol{eta}_{oldsymbol{n}}}{oldsymbol{eta}_{oldsymbol{n+1}}} \cdot rac{rac{1}{p(oldsymbol{x}_{n+1}, \cdots, oldsymbol{x}_{N} | oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n}, oldsymbol{ heta}^{ ext{ old }})}{rac{1}{p(oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n} | oldsymbol{ heta}^{ ext{ old }})}{p(oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n+1} | oldsymbol{ heta}^{ ext{ old }})} &= rac{oldsymbol{eta}_{oldsymbol{n}}}{oldsymbol{eta}_{n+1}} \cdot rac{1}{p(oldsymbol{x}_{n+1} | oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n}, oldsymbol{ heta}^{ ext{ old }})}{rac{oldsymbol{b}_{oldsymbol{n}}}{oldsymbol{eta}_{n+1}} \cdot rac{1}{c_{n+1}} &= rac{oldsymbol{h}_{oldsymbol{n}}}{oldsymbol{b}_{n+1}} \cdot rac{1}{c_{n+1}} &= rac{1}{c_{n+1}} &= rac{oldsymbol{h}_{oldsymbol{n}}}{oldsymbol{b}_{n+1}} \cdot rac{1}{c_{n+1}} &= rac{1}{c_{n+$$

the

$$egin{aligned} \hat{oldsymbol{eta}}_{oldsymbol{n}} &\equiv rac{1}{p(oldsymbol{x_{n+1}}, \cdots, oldsymbol{x_N} | oldsymbol{x_1}, \cdots, oldsymbol{x_n}, oldsymbol{ heta} ext{ old })} eta_{oldsymbol{n}} &= rac{1}{p(oldsymbol{x_{n+1}}, \cdots, oldsymbol{x_N} | oldsymbol{x_1}, \cdots, oldsymbol{x_n}, oldsymbol{ heta} ext{ old })} &= rac{1}{c_{n+1}} oldsymbol{A}^T \cdot \left( [oldsymbol{\phi}^T \cdot oldsymbol{x_{n+1}}] \odot \hat{oldsymbol{eta}}_{n+1} 
ight)} & \hat{oldsymbol{eta}}_{oldsymbol{n}} &= 1. \end{aligned}$$

then

$$egin{aligned} oldsymbol{\gamma_n} &= [\hat{oldsymbol{lpha}}_n \odot \hat{oldsymbol{eta}}_n] \ oldsymbol{\xi_n} &= rac{1}{c_n} iggl\{ \left[ [oldsymbol{\phi}^T \cdot oldsymbol{x}_n] \odot \hat{oldsymbol{eta}}_n 
ight] \cdot \hat{oldsymbol{lpha}}_{n-1}^T iggr\} \odot oldsymbol{A} \end{aligned}$$

其中
$$p(oldsymbol{X}|oldsymbol{ heta}^{
m old}) = \sum_{Z_N} p(oldsymbol{z_N}, oldsymbol{X}|oldsymbol{ heta}^{
m old}) = \sum_{Z_N} a_N = 1.^T \cdot oldsymbol{lpha_N} ext{ is constant}$$
 
$$egin{aligned} & ec{p}(oldsymbol{X}|oldsymbol{ heta}^{
m old}) = p(oldsymbol{X}|oldsymbol{ heta}^{
m old}) \cdot 1. \\ & p(oldsymbol{X}|oldsymbol{ heta}^{
m old}) = 1.^T \cdot oldsymbol{lpha_N} \\ & = 1.^T \cdot oldsymbol{lpha_N} \end{aligned}$$
 
$$egin{aligned} & oldsymbol{\phi} = \left\{ oldsymbol{x_n} \cdot [\sum_{n=2}^N oldsymbol{\gamma_n}]^T \right\} \odot \frac{1.}{1. \cdot 1.^T \left\{ oldsymbol{x_n} \cdot [\sum_{n=2}^N oldsymbol{\gamma_n}]^T \right\}} \end{aligned}$$
 
$$oldsymbol{A} = [\sum_{n=2}^N oldsymbol{\xi_n}] \odot \frac{1.}{1. \cdot 1.^T [\sum_{n=2}^N oldsymbol{\xi_n}]}$$
 
$$oldsymbol{\pi} = oldsymbol{\gamma_1} \odot \frac{1.}{1. \cdot 1.^T oldsymbol{\gamma_1}}$$
 
$$oldsymbol{\pi} = oldsymbol{\gamma_1} \odot \frac{1.}{1. \cdot 1.^T oldsymbol{\gamma_1}}$$