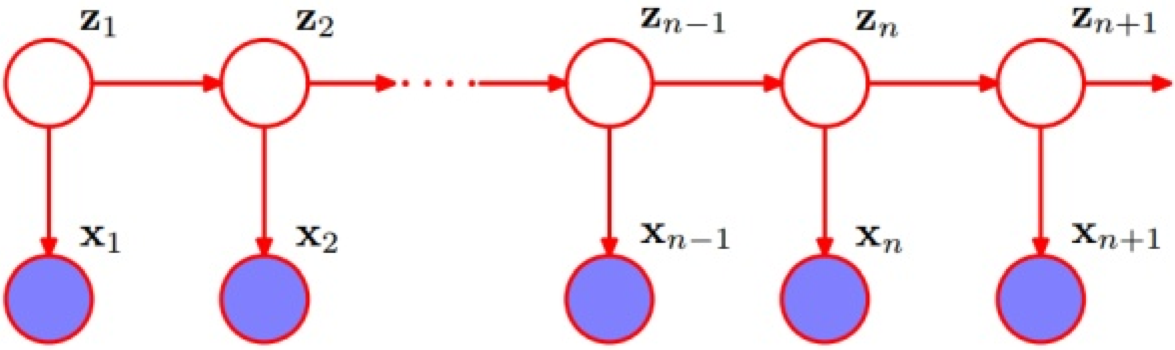

why it could be defined as below:

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

$$p\left(\boldsymbol{z}_1 \mid \boldsymbol{\pi}\right)=\prod_{k=1}^K \pi_k^{z_{1k}}$$

$$p\left(\boldsymbol{z}_n \mid \boldsymbol{z}_{n-1}, \boldsymbol{A}\right)=\prod_{j=1}^K \prod_{k=1}^K A_{j,k}^{z_{n-1,j} z_{nk}}$$

$$p\left(\boldsymbol{x}_n \mid \boldsymbol{z}_n, \boldsymbol{\phi}\right)=\prod_{j=1}^K \prod_{t=1}^{\text{any dimension}} \phi_j(t)^{z_{nj} x_n(t)}$$

Define $\vec{p}$ constant matrix

$$\vec{p}(\boldsymbol{z}_n \mid \boldsymbol{\theta}) \equiv\left[\begin{array}{c}\left.p(\boldsymbol{z}_n \mid \boldsymbol{\theta})\right|_{\boldsymbol{z}_n=\text{state } 1} \\ \vdots \\ \left.p(\boldsymbol{z}_n \mid \boldsymbol{\theta})\right|_{\boldsymbol{z}_n=\text{state } K}\end{array}\right]=\left[\sum_{\boldsymbol{z}_n} p\left(\boldsymbol{z}_n \mid \boldsymbol{X}, \boldsymbol{\theta}^{\text{ old }}\right) \boldsymbol{z}_n\right] \quad \text{state } k: \boldsymbol{z}_n=\left[\begin{array}{c}0 \\ \vdots \\ 1 \\ \vdots \\ 0\end{array}\right]$$

$$\vec{p}(\boldsymbol{z}_1 \mid \boldsymbol{\theta})=\boldsymbol{\pi}$$

$$\vec{p}\left(\boldsymbol{z}_n \mid \boldsymbol{z}_{n-1}, \boldsymbol{\theta}\right) \equiv\left[\left.p\left(\boldsymbol{z}_n \mid \boldsymbol{z}_{n-1}, \boldsymbol{A}\right)\right|_{\boldsymbol{z}_n=\text{state } i, \boldsymbol{z}_{n-1}=\text{state } j}\right]=\boldsymbol{A}$$

$$\vec{p}\left(\boldsymbol{x}_n \mid \boldsymbol{z}_n, \boldsymbol{\theta}\right) \equiv\left[\left.p\left(\boldsymbol{x}_n \mid \boldsymbol{z}_n, \boldsymbol{\phi}\right)\right|_{\boldsymbol{x}_n=\text{state } i, \boldsymbol{z}_n=\text{state } j}\right]=\boldsymbol{\phi}$$

where, for short: write $\vec{p}(\boldsymbol{z}_n) \equiv \vec{p}(\boldsymbol{z}_n \mid \boldsymbol{\theta}), \vec{p}(\boldsymbol{x}_n) \equiv \vec{p}(\boldsymbol{x}_n \mid \boldsymbol{\theta})$

$$\begin{gathered}\vec{p}(\boldsymbol{z}_1)=\boldsymbol{\pi} \\ \vec{p}(\boldsymbol{z}_n)=\boldsymbol{A} \vec{p}(\boldsymbol{z}_{n-1}) \\ \vec{p}(\boldsymbol{x}_n)=\boldsymbol{\phi} \cdot \vec{p}(\boldsymbol{z}_n)\end{gathered}$$

size($p(\boldsymbol{x}_n \mid \boldsymbol{\phi})$) = $\textit{any dimension} \times K$, $\boldsymbol{\phi}=[\phi_1, \phi_2, \ldots, \phi_j, \ldots \phi_K]$

ϕ_k is probability distribution sequence of k th state, $\textit{any dimension} \times 1$

$\boldsymbol{x}_n$ is probability distribution sequence, $\textit{any dimension} \times 1$

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

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

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

Thus

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

Goal

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

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

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

Here

$$\begin{gathered}
p\left(\boldsymbol{x_n} | \boldsymbol{z_n}, \boldsymbol{\phi}\right) = \prod_{j=1}^{K} \prod_{t=1}^{any\ dimension} \phi_j(t)^{z_{nj} x_n(t)} \\
p\left(\boldsymbol{z_n} | \boldsymbol{z_{n-1}}, \boldsymbol{A}\right) = \prod_{j=1}^{K} \prod_{k=1}^{K} A_{k,j}^{z_{n-1,j} z_{nk}} \\
p\left(\boldsymbol{z_1} | \, \boldsymbol{\pi}\right) = \prod_{k=1}^{K} \pi_k^{z_{1k}}
\end{gathered}$$

So

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

if t is continuous

$$< \ln \boldsymbol{\phi}, \boldsymbol{x_n} >= \int_{-\infty}^{+\infty} \ln \boldsymbol{\phi(t)} \cdot x_n(t) dt$$

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_1(t) \\ \phi_2(t) \\ \vdots \\ \phi_K(t) \end{bmatrix}$$

if m is discrete

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

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_{11} & \cdots & \phi_{1m} & \cdots \phi_{1M} \\ \phi_{21} & \cdots & \phi_{2m} & \cdots \phi_{2M} \\ \vdots & & & \vdots \\ \phi_{K1} & \cdots & \phi_{Km} & \cdots \phi_{KM} \end{bmatrix}^T M \times K \text{matrix}$$

$$\boldsymbol{x_n} : M \times 1 \text{ matrix}$$

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

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

then

$$P(Y = y)dy \equiv P(X = x)dx = p(x)dx, \quad y = f(x)$$

$$\implies P(Y = y) = p(x) \frac{dx}{df(x)}$$

$$\implies E[Y] = \int_{-\infty}^{+\infty} y \cdot P(Y = y)dy = \int_{-\infty}^{+\infty} f(x)p(x)dx$$

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

$$\mathcal{L}(q, \boldsymbol{\theta}) = \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln \left\{ \frac{p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta})}{q(\boldsymbol{Z})} \right\}$$

$$\mathrm{KL}(q\|p) = -\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln \left\{ \frac{p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta})}{q(\boldsymbol{Z})} \right\}$$

Where

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

So

$$
\begin{gathered}
\ln p(\boldsymbol{X}|\boldsymbol{\theta}) \geq \mathcal{L}(q,\boldsymbol{\theta}) = \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln\left\{\frac{p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})}{q(\boldsymbol{Z})}\right\}, \quad \forall q(\boldsymbol{Z}) \\
\text{if } q(\boldsymbol{Z}) \Leftarrow p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}), \text{here } \mathrm{KL}(q\|p) = 0 \\
\ln p(\boldsymbol{X}|\boldsymbol{\theta}) = \mathcal{L}(q,\boldsymbol{\theta}) = \sum_{\boldsymbol{Z}} p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}) \ln\left\{\frac{p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})}{p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta})}\right\}
\end{gathered}
$$

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

调整法

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

1. fix θ, change $q(Z)$ [E步骤]

$$
\begin{gathered}
q(\boldsymbol{Z}) \Leftarrow p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}) \\
\ln p(\boldsymbol{X}|\boldsymbol{\theta}) = \mathcal{L}(q,\boldsymbol{\theta}) + \mathrm{KL}(q|p) = \text{const与q无关} \\
\mathrm{KL}(q|p) \downarrow \Leftarrow 0, \quad \text{update } \mathcal{L}(q,\boldsymbol{\theta}) \uparrow \\
\ln p(\boldsymbol{X}|\boldsymbol{\theta}) == \mathcal{L}(q,\boldsymbol{\theta}) \\
\boldsymbol{\theta}^{\text{ old }} \Leftarrow \boldsymbol{\theta}
\end{gathered}
$$

2. fix $q(Z)$, change θ [M步骤]

ignore influence of θ in $\mathrm{KL}(q|p) = -\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln\left\{\frac{p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta})}{q(\boldsymbol{Z})}\right\}$

$$
\begin{aligned}
\mathcal{L}(q,\boldsymbol{\theta}) &= \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln\left\{\frac{p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})}{q(\boldsymbol{Z})}\right\} \\
&= \sum_{\boldsymbol{Z}} p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}) \ln\left\{\frac{p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})}{p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }})}\right\} \\
&= \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \ln p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta}) - \text{const}
\end{aligned}
$$

set

$$
\boldsymbol{Q}\left(\boldsymbol{\theta},\boldsymbol{\theta}^{\text{ old }}\right) \equiv \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \ln p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})
$$

get

$$
\begin{gathered}
\boldsymbol{\theta} \Leftarrow \underset{\overline{\boldsymbol{\theta}}}{\operatorname{argmax}} \mathcal{L}(q,\overline{\boldsymbol{\theta}}) \\
= \underset{\overline{\boldsymbol{\theta}}}{\operatorname{argmax}} \boldsymbol{Q}\left(\overline{\boldsymbol{\theta}},\boldsymbol{\theta}^{\text{ old }}\right) \\
\text{update } \mathcal{L}(q,\boldsymbol{\theta}) \uparrow
\end{gathered}
$$

update

$$
\begin{gathered}
\mathrm{KL}(q|p) \uparrow = -\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln\left\{\frac{p(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta})}{q(\boldsymbol{Z})}\right\} > 0 \\
\ln p(\boldsymbol{X}|\boldsymbol{\theta}) \uparrow = \mathcal{L}(q,\boldsymbol{\theta}) \uparrow + \mathrm{KL}(q|p) \uparrow
\end{gathered}
$$

Conclusion

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

且1中$\ln p(\boldsymbol{X}|\boldsymbol{\theta})=\mathcal{L}(q,\boldsymbol{\theta})\uparrow+0\downarrow=\text{const}$

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

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

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

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

X is fix observation value;

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

$\sum_{\boldsymbol{Z}}p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)$ is [probability distribution] random variable **of** Z [probability distribution] random variable

$$\begin{aligned}
\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_{\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\{\sum_{n=2}^{N}\sum\left[\sum_{\boldsymbol{z_{n-1}},\boldsymbol{z_n}}\boldsymbol{z_{n-1}}p\left(\boldsymbol{z_{n-1}},\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)\boldsymbol{z_n^T}\right]\odot\ln\boldsymbol{A}^T\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{aligned}$$

where

$$
\begin{aligned}
&\sum_{\boldsymbol{z_{n-1}},\boldsymbol{z_n}} p\left(\boldsymbol{z_{n-1}},\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)\sum_{n=2}^{N}\boldsymbol{z_{n-1}^T}\ln\boldsymbol{A}^T\cdot\boldsymbol{z_n}\\
&=\sum\left\{p\left(\boldsymbol{z_{n-1}},\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)\odot\sum_{n=2}^{N}\left[\boldsymbol{z_{n-1}^T}\ln\boldsymbol{A}^T\cdot\boldsymbol{z_n}\right]\right\}\\
&=\sum_{n=2}^{N}\operatorname{tr}\left(p\left(\boldsymbol{z_{n-1}},\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)\left[\boldsymbol{z_{n-1}^T}\ln\boldsymbol{A}^T\cdot\boldsymbol{z_n}\right]^T\right)\\
&=\sum_{n=2}^{N}\operatorname{tr}\left(\left[\boldsymbol{z_{n-1}}p\left(\boldsymbol{z_{n-1}},\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)\boldsymbol{z_n^T}\right]\cdot\ln\boldsymbol{A}\right)\\
&=\sum_{n=2}^{N}\sum\left\{\left[\boldsymbol{z_n}p\left(\boldsymbol{z_{n-1}},\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)\boldsymbol{z_{n-1}^T}\right]\odot\ln\boldsymbol{A}\right\}
\end{aligned}
$$

其中, 因为X,θ^{old} 分别为已知、固定 在 M步骤

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

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

下面求解

$$
\begin{aligned}
\max_{\boldsymbol{\theta}}\quad &\boldsymbol{Q}\left(\boldsymbol{\theta},\boldsymbol{\theta}^{\text{ old}}\right)\equiv\sum_{\boldsymbol{Z}}p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)\ln p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta})\\
&\qquad=\sum_{n=1}^{N}\boldsymbol{\gamma_n^T}<\ln\boldsymbol{\phi},\boldsymbol{x_n}>+\sum_{n=2}^{N}\operatorname{tr}\left(\boldsymbol{\xi_n^T}\ln\boldsymbol{A}\right)+\boldsymbol{\gamma_1^T}\ln\boldsymbol{\pi}\\
&\text{s.t.}\quad\boldsymbol{\gamma_n^T}\cdot 1.=1\quad\operatorname{tr}\left(\boldsymbol{\xi_n^T}\cdot 1.\right)=1\\
&1.^T\cdot\boldsymbol{\phi}=1.^T\quad 1.^T\cdot\boldsymbol{A}=1.^T\quad 1.^T\cdot\boldsymbol{\pi}=1
\end{aligned}
$$

with Lagrange method:

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

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

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

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

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

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

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

So

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

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

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

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

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

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

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

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

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

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

$$\implies \boldsymbol{A}=[\sum_{n=2}^{N}\boldsymbol{\xi_n}]\odot\frac{1.}{1.\cdot 1.^T[\sum_{n=2}^{N}\boldsymbol{\xi_n}]}$$

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

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

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

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

$$\implies \boldsymbol{\pi}=\boldsymbol{\gamma_1}\odot\frac{1.}{1.\cdot 1.^T\boldsymbol{\gamma_1}}$$

To sum up

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

How to calculate $\boldsymbol{\gamma_n}, \boldsymbol{\xi_n}$

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

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

similarly get：

$$
\boldsymbol{\gamma_n} \equiv \vec{p}(\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}) \equiv \begin{bmatrix} p(\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}})\Big|\boldsymbol{z_n}=\text{state }1\\ \vdots \\ p(\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}})\Big|\boldsymbol{z_n}=\text{state }K\end{bmatrix}
$$

$$
\boldsymbol{\xi_n} \equiv \vec{p}\left(\boldsymbol{z_{n-1}},\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right) \equiv \left[p\left(\boldsymbol{z_{n-1}},\boldsymbol{z_n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old}}\right)\Big|\boldsymbol{z_n}=\text{state }i,\boldsymbol{z_{n-1}}=\text{state }j\right]
$$

$$
\text{the same size as}\quad \boldsymbol{z_n}\boldsymbol{z_{n-1}^T}
$$

so

$$
\begin{aligned}
\boldsymbol{\gamma}_{\boldsymbol{n}} &\equiv \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old }}) \\
&= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) \odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{n}}|\boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}+\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{N}}|\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{n}}|\boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}+\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{N}}|\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= \boldsymbol{\alpha}_{\boldsymbol{n}} \odot \boldsymbol{\beta}_{\boldsymbol{n}} \odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
\boldsymbol{\xi}_{\boldsymbol{n}} &\equiv \vec{p}\left(\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{z}_{\boldsymbol{n}}|\boldsymbol{X}, \boldsymbol{\theta}^{\text{ old }}\right) \\
&= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) \odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{n}-\mathbf{1}}|\boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{\boldsymbol{n}}, \ldots, \boldsymbol{x}_{\boldsymbol{N}}|\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{n}-\mathbf{1}}|\boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{\boldsymbol{n}}, \ldots, \boldsymbol{x}_{\boldsymbol{N}}|\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{n}-\mathbf{1}}|\boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}|\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}}, \ldots, \boldsymbol{x}_{\boldsymbol{N}}|\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{n}-\mathbf{1}}|\boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}|\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}}, \ldots, \boldsymbol{x}_{\boldsymbol{N}}|\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{n}-\mathbf{1}}|\boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}|\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}}|\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}+\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{N}}|\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{n}-\mathbf{1}}|\boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}|\boldsymbol{z}_{\boldsymbol{n}-\mathbf{1}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}}|\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}+\mathbf{1}}, \ldots, \boldsymbol{x}_{\boldsymbol{N}}|\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \\
&\odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= [1. \cdot \boldsymbol{\alpha}_{\boldsymbol{n}-\mathbf{1}}^{T}] \odot \boldsymbol{A} \odot \left[[\boldsymbol{\phi}^{T} \cdot \boldsymbol{x}_{\boldsymbol{n}}] \odot \boldsymbol{\beta}_{\boldsymbol{n}} \cdot 1.^{T}\right] \odot \frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \\
&= \frac{1}{p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }})} \left\{ \left[[\boldsymbol{\phi}^{T} \cdot \boldsymbol{x}_{\boldsymbol{n}}] \odot \boldsymbol{\beta}_{\boldsymbol{n}}\right] \cdot \boldsymbol{\alpha}_{\boldsymbol{n}-\mathbf{1}}^{T} \right\} \odot \boldsymbol{A}
\end{aligned}
$$

其中$p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) = \sum_{Z_N} p(\boldsymbol{z_N}, \boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) = \sum \boldsymbol{\alpha_N} = 1.^T \cdot \boldsymbol{\alpha_N}$ is **constant**

$$\begin{aligned}
\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) &= p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) \cdot 1. \\
p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) &= 1.^T \cdot p(\boldsymbol{z_N}, \boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) \\
&= 1.^T \cdot \boldsymbol{\alpha_N}
\end{aligned}$$

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

$$\begin{aligned}
\boldsymbol{\alpha_n} &\equiv \vec{p}(\boldsymbol{z_n}, \boldsymbol{x_1}, \ldots, \boldsymbol{x_n}|\boldsymbol{\theta}^{\text{ old }}) \\
\boldsymbol{\beta_n} &\equiv \vec{p}(\boldsymbol{x_{n+1}}, \ldots, \boldsymbol{x_N}|\boldsymbol{z_n}, \boldsymbol{\theta}^{\text{ old }})
\end{aligned}$$

How to obtain $\boldsymbol{\alpha_n}, \boldsymbol{\beta_n}$

for $\boldsymbol{\alpha_n}$, initial value:

$$\begin{aligned}
\boldsymbol{\alpha_1} &= \vec{p}(\boldsymbol{z_1}, \boldsymbol{x_1}|\boldsymbol{\theta}^{\text{ old }}) \\
&= \vec{p}(\boldsymbol{x_1}|\boldsymbol{z_1}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{z_1}|\boldsymbol{\theta}^{\text{ old }}) \\
&= [\boldsymbol{\phi}^T \cdot \boldsymbol{x_1}] \odot \boldsymbol{\pi}
\end{aligned}$$

Recurrence relation

$$\begin{aligned}
\boldsymbol{\alpha_n} &\equiv \vec{p}(\boldsymbol{z_n}, \boldsymbol{x_1}, \ldots, \boldsymbol{x_n}|\boldsymbol{\theta}^{\text{ old }}) \\
&= \vec{p}(\boldsymbol{z_n}|\boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x_1}, \ldots, \boldsymbol{x_{n-1}}|\boldsymbol{z_n}, \boldsymbol{x_n}, \boldsymbol{\theta}^{\text{ old }}) \\
&= \vec{p}(\boldsymbol{z_n}|\boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x_1}, \ldots, \boldsymbol{x_{n-1}}|\boldsymbol{z_n}, \boldsymbol{\theta}^{\text{ old }}) \\
&= \vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{z_n}, \boldsymbol{x_1}, \ldots, \boldsymbol{x_{n-1}}|\boldsymbol{\theta}^{\text{ old }}) \\
&= \vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{\vec{p}(\boldsymbol{z_{n-1}}, \boldsymbol{z_n}, \boldsymbol{x_1}, \ldots, \boldsymbol{x_{n-1}}|\boldsymbol{\theta}^{\text{ old }}) \cdot 1.\right\} \\
&= \vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{\left[\vec{p}(\boldsymbol{z_{n-1}}, \boldsymbol{z_n}|\boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x_1}, \ldots, \boldsymbol{x_{n-1}}|\boldsymbol{z_{n-1}}, \boldsymbol{z_n}, \boldsymbol{\theta}^{\text{ old }})\right] \cdot 1.\right\} \\
&= \vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{\left[\left(1. \cdot \vec{p}(\boldsymbol{z_{n-1}}|\boldsymbol{\theta}^{\text{ old }})^T\right) \odot \vec{p}(\boldsymbol{z_n}|\boldsymbol{z_{n-1}}, \boldsymbol{\theta}^{\text{ old }}) \odot \left(1. \cdot \vec{p}(\boldsymbol{x_1}, \ldots, \boldsymbol{x_{n-1}}|\boldsymbol{z_{n-1}}, \boldsymbol{\theta}^{\text{ old }})^T\right)\right] \cdot 1.\right\} \\
&= \vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{\left[\vec{p}(\boldsymbol{z_n}|\boldsymbol{z_{n-1}}, \boldsymbol{\theta}^{\text{ old }}) \odot \left(1. \cdot \vec{p}(\boldsymbol{z_{n-1}}, \boldsymbol{x_1}, \ldots, \boldsymbol{x_{n-1}}|\boldsymbol{\theta}^{\text{ old }})^T\right)\right] \cdot 1.\right\} \\
&= [\boldsymbol{\phi}^T \cdot \boldsymbol{x_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_n}] \odot [\boldsymbol{A} \cdot \boldsymbol{\alpha_{n-1}}]
\end{aligned}$$

$$\boldsymbol{x}_{\boldsymbol{n}}^T \cdot \boldsymbol{\phi} \cdot \boldsymbol{A} \cdot \boldsymbol{\alpha_{n-1}}$$

for $\boldsymbol{\beta_n}$, last initial value:

$$\boldsymbol{\beta_N} = \vec{p}(1|\boldsymbol{z_N}, \boldsymbol{\theta}^{\text{ old }}) = 1$$

Recurrence relation

$$
\begin{aligned}
\boldsymbol{\beta_n} &\equiv \vec{p}(\boldsymbol{x_{n+1}},\dots,\boldsymbol{x_N}|\boldsymbol{z_n},\boldsymbol{\theta}^{\text{ old}})\\
&=\left\{\vec{p}(\boldsymbol{z_{n+1}},\boldsymbol{x_{n+1}},\dots,\boldsymbol{x_N}|\boldsymbol{z_n},\boldsymbol{\theta}^{\text{ old}})^T\cdot 1.\right\}\\
&=\left\{\left[\vec{p}(\boldsymbol{z_{n+1}}|\boldsymbol{z_n},\boldsymbol{\theta}^{\text{ old}})\odot\vec{p}(\boldsymbol{x_{n+1}},\dots,\boldsymbol{x_N}|\boldsymbol{z_n},\boldsymbol{z_{n+1}},\boldsymbol{\theta}^{\text{ old}})\right]^T\cdot 1.\right\}\\
&=\left\{\left[\vec{p}(\boldsymbol{z_{n+1}}|\boldsymbol{z_n},\boldsymbol{\theta}^{\text{ old}})\odot\left(\vec{p}(\boldsymbol{x_{n+1}},\dots,\boldsymbol{x_N}|\boldsymbol{z_{n+1}},\boldsymbol{\theta}^{\text{ old}})\cdot 1.^T\right)\right]^T\cdot 1.\right\}\\
&=\left\{\left[\vec{p}(\boldsymbol{z_{n+1}}|\boldsymbol{z_n},\boldsymbol{\theta}^{\text{ old}})\odot\left([\vec{p}(\boldsymbol{x_{n+1}}|\boldsymbol{z_{n+1}},\boldsymbol{\theta}^{\text{ old}})\odot\vec{p}(\boldsymbol{x_{n+2}},\dots,\boldsymbol{x_N}|\boldsymbol{z_{n+1}},\boldsymbol{x_{n+1}},\boldsymbol{\theta}^{\text{ old}})]\cdot 1.^T\right)\right]^T\cdot 1.\right\}\\
&=\left\{\left[\vec{p}(\boldsymbol{z_{n+1}}|\boldsymbol{z_n},\boldsymbol{\theta}^{\text{ old}})\odot\left([\vec{p}(\boldsymbol{x_{n+1}}|\boldsymbol{z_{n+1}},\boldsymbol{\theta}^{\text{ old}})\odot\vec{p}(\boldsymbol{x_{n+2}},\dots,\boldsymbol{x_N}|\boldsymbol{z_{n+1}},\boldsymbol{\theta}^{\text{ old}})]\cdot 1.^T\right)\right]^T\cdot 1.\right\}\\
&=\left\{\left[\boldsymbol{A}\odot\left(\left[[\boldsymbol{\phi}^T\cdot\boldsymbol{x_{n+1}}]\odot\boldsymbol{\beta_{n+1}}\right]\cdot 1.^T\right)\right]^T\cdot 1.\right\}\\
&=\boldsymbol{A}^T\cdot\left([\boldsymbol{\phi}^T\cdot\boldsymbol{x_{n+1}}]\odot\boldsymbol{\beta_{n+1}}\right)
\end{aligned}
$$

To sum up

$$
\begin{aligned}
\boldsymbol{\alpha_1} &= [\boldsymbol{\phi}^T\cdot\boldsymbol{x_1}]\odot\boldsymbol{\pi}\\
\boldsymbol{\alpha_n} &= [\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot[\boldsymbol{A}\cdot\boldsymbol{\alpha_{n-1}}]\\
\boldsymbol{\beta_N} &= 1.\\
\boldsymbol{\beta_n} &= \boldsymbol{A}^T\cdot\left([\boldsymbol{\phi}^T\cdot\boldsymbol{x_{n+1}}]\odot\boldsymbol{\beta_{n+1}}\right)
\end{aligned}
$$

Then

$$
\begin{aligned}
\boldsymbol{\gamma_n} &= \boldsymbol{\alpha_n}\odot\boldsymbol{\beta_n}\odot\frac{1.}{\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}})}\\
&=\frac{1}{1.^T\cdot\boldsymbol{\alpha_N}}[\boldsymbol{\alpha_n}\odot\boldsymbol{\beta_n}]\\
\boldsymbol{\xi_n} &= \frac{1}{p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old}})}\left\{\left[[\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot\boldsymbol{\beta_n}\right]\cdot\boldsymbol{\alpha_{n-1}^T}\right\}\odot\boldsymbol{A}\\
&=\frac{1}{1.^T\cdot\boldsymbol{\alpha_N}}\left\{\left[[\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot\boldsymbol{\beta_n}\right]\cdot\boldsymbol{\alpha_{n-1}^T}\right\}\odot\boldsymbol{A}
\end{aligned}
$$

because $\boldsymbol{\alpha_N},\boldsymbol{\beta_1}$ 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

$$
\begin{aligned}
\boldsymbol{\hat{\alpha}_n} &\equiv \frac{1}{p(\boldsymbol{x_1},\cdots,\boldsymbol{x_n}|\boldsymbol{\theta}^{\text{ old}})}\boldsymbol{\alpha_n} = \vec{p}(\boldsymbol{z_n}|\boldsymbol{x_1},\cdots,\boldsymbol{x_n},\boldsymbol{\theta}^{\text{ old}})\\
&=\left\{\prod_{k=1}^{n}p(\boldsymbol{x_k}|\boldsymbol{x_1},\cdots,\boldsymbol{x_{k-1}},\boldsymbol{\theta}^{\text{ old}})\right\}\boldsymbol{\alpha_n}\\
&=\left\{\prod_{k=1}^{n}c_k\right\}\boldsymbol{\alpha_n}\quad\text{notice}\quad 1.^T\boldsymbol{\hat{\alpha}_n}=\sum_{\boldsymbol{z_n}}p(\boldsymbol{z_n}|\boldsymbol{x_1},\cdots,\boldsymbol{x_n},\boldsymbol{\theta}^{\text{ old}})=1
\end{aligned}
$$

the other way

$$
\begin{aligned}
c_n\hat{\boldsymbol{\alpha}}_{\boldsymbol{n}} &= p(\boldsymbol{x_n}|\boldsymbol{x_1},\cdots,\boldsymbol{x_{n-1}},\boldsymbol{\theta}^{\text{ old}})\vec{p}(\boldsymbol{z_n}|\boldsymbol{x_1},\cdots,\boldsymbol{x_n},\boldsymbol{\theta}^{\text{ old}})\\
&= \vec{p}(\boldsymbol{z_n},\boldsymbol{x_n}|\boldsymbol{x_1},\cdots,\boldsymbol{x_{n-1}},\boldsymbol{\theta}^{\text{ old}})\\
&= \vec{p}(\boldsymbol{z_n}|\boldsymbol{x_1},\cdots,\boldsymbol{x_{n-1}},\boldsymbol{\theta}^{\text{ old}})\odot\vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}\boldsymbol{\theta}^{\text{ old}})\\
&= [\vec{p}(\boldsymbol{z_n},\boldsymbol{z_{n-1}}|\boldsymbol{x_1},\cdots,\boldsymbol{x_{n-1}},\boldsymbol{\theta}^{\text{ old}})\cdot 1.]\odot\vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}\boldsymbol{\theta}^{\text{ old}})\\
&= \big[[1.\cdot\vec{p}(\boldsymbol{z_{n-1}}|\boldsymbol{x_1},\cdots,\boldsymbol{x_{n-1}},\boldsymbol{\theta}^{\text{ old}})^T]\odot\vec{p}(\boldsymbol{z_n}|\boldsymbol{z_{n-1}})\cdot 1.\big]\odot\vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}\boldsymbol{\theta}^{\text{ old}})\\
&= \big[\vec{p}(\boldsymbol{z_n}|\boldsymbol{z_{n-1}})\vec{p}(\boldsymbol{z_{n-1}}|\boldsymbol{x_1},\cdots,\boldsymbol{x_{n-1}},\boldsymbol{\theta}^{\text{ old}})\big]\odot\vec{p}(\boldsymbol{x_n}|\boldsymbol{z_n}\boldsymbol{\theta}^{\text{ old}})\\
&= [\boldsymbol{A}\cdot\hat{\boldsymbol{\alpha}}_{\boldsymbol{n-1}}]\odot[\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\\
&= [\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot[\boldsymbol{A}\cdot\hat{\boldsymbol{\alpha}}_{\boldsymbol{n-1}}]\\
c_n &= 1.^T c_n\cdot\boldsymbol{\alpha_n} = 1.^T\cdot\big\{[\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot[\boldsymbol{A}\cdot\hat{\boldsymbol{\alpha}}_{\boldsymbol{n-1}}]\big\}
\end{aligned}
$$

so

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}_{\boldsymbol{n}} &\equiv \frac{1}{c_n}[\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot[\boldsymbol{A}\cdot\hat{\boldsymbol{\alpha}}_{\boldsymbol{n-1}}]\\
&= \frac{1}{1.^T\cdot\{[\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot[\boldsymbol{A}\cdot\hat{\boldsymbol{\alpha}}_{\boldsymbol{n-1}}]\}}[\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot[\boldsymbol{A}\cdot\hat{\boldsymbol{\alpha}}_{\boldsymbol{n-1}}]\\
\hat{\boldsymbol{\alpha}}_{\boldsymbol{1}} &= \frac{1}{1.^T\cdot\{[\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot\boldsymbol{\pi}\}}[\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot\boldsymbol{\pi}
\end{aligned}
$$

thesame

$$
\begin{aligned}
\frac{\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}}{\hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}}} &\equiv \frac{\boldsymbol{\beta_n}}{\boldsymbol{\beta_{n+1}}}\cdot\frac{\frac{1}{p(\boldsymbol{x_{n+1}},\cdots,\boldsymbol{x_N}|\boldsymbol{x_1},\cdots,\boldsymbol{x_n},\boldsymbol{\theta}^{\text{ old}})}}{\frac{1}{p(\boldsymbol{x_{n+2}},\cdots,\boldsymbol{x_N}|\boldsymbol{x_1},\cdots,\boldsymbol{x_{n+1}},\boldsymbol{\theta}^{\text{ old}})}}\\
&= \frac{\boldsymbol{\beta_n}}{\boldsymbol{\beta_{n+1}}}\cdot\frac{p(\boldsymbol{x_1},\cdots,\boldsymbol{x_n}|\boldsymbol{\theta}^{\text{ old}})}{p(\boldsymbol{x_1},\cdots,\boldsymbol{x_{n+1}}|\boldsymbol{\theta}^{\text{ old}})}\\
&= \frac{\boldsymbol{\beta_n}}{\boldsymbol{\beta_{n+1}}}\cdot\frac{1}{p(\boldsymbol{x_{n+1}}|\boldsymbol{x_1},\cdots,\boldsymbol{x_n},\boldsymbol{\theta}^{\text{ old}})}\\
&= \frac{\boldsymbol{\beta_n}}{\boldsymbol{\beta_{n+1}}}\cdot\frac{1}{c_{n+1}}
\end{aligned}
$$

the

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\boldsymbol{n}} &\equiv \frac{1}{p(\boldsymbol{x_{n+1}},\cdots,\boldsymbol{x_N}|\boldsymbol{x_1},\cdots,\boldsymbol{x_n},\boldsymbol{\theta}^{\text{ old}})}\boldsymbol{\beta_n}\\
&= \frac{1}{p(\boldsymbol{x_{n+1}},\cdots,\boldsymbol{x_N}|\boldsymbol{x_1},\cdots,\boldsymbol{x_n},\boldsymbol{\theta}^{\text{ old}})}\vec{p}(\boldsymbol{x_{n+1}},\ldots,\boldsymbol{x_N}|\boldsymbol{z_n},\boldsymbol{\theta}^{\text{ old}})\\
&= \frac{1}{c_{n+1}}\boldsymbol{A}^T\cdot\Big([\boldsymbol{\phi}^T\cdot\boldsymbol{x_{n+1}}]\odot\hat{\boldsymbol{\beta}}_{\boldsymbol{n+1}}\Big)\\
\hat{\boldsymbol{\beta}}_{\boldsymbol{N}} &= 1.
\end{aligned}
$$

then

$$
\begin{aligned}
\boldsymbol{\gamma_n} &= [\hat{\boldsymbol{\alpha}}_{\boldsymbol{n}}\odot\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}]\\
\boldsymbol{\xi_n} &= \frac{1}{c_n}\left\{\Big[[\boldsymbol{\phi}^T\cdot\boldsymbol{x_n}]\odot\hat{\boldsymbol{\beta}}_{\boldsymbol{n}}\Big]\cdot\hat{\boldsymbol{\alpha}}^T_{\boldsymbol{n-1}}\right\}\odot\boldsymbol{A}
\end{aligned}
$$

$$\boldsymbol{\alpha_N}$$

其中$p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) = \sum\limits_{Z_N} p(\boldsymbol{z_N}, \boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) = \sum \boldsymbol{\alpha_N} = 1.^T \cdot \boldsymbol{\alpha_N}$ is **constant**

$$\begin{aligned}
\vec{p}(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) &= p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) \cdot 1. \\
p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) &= 1.^T \cdot p(\boldsymbol{z_N}, \boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) \\
&= 1.^T \cdot \boldsymbol{\alpha_N}
\end{aligned}$$

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