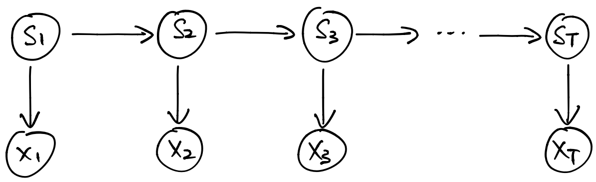
State Space Model:

1. HMM: (Just for notation).



$S_t \in \{1, \cdots, k\}$. $\pi_j = P(S_1 = j)$. $\phi_{ij} = P(S_{t+1} = j \mid S_t = i)$

$A_j(\underset{\sim}{x}) = P(\underset{\sim}{x}_t = \underset{\sim}{x} \mid S_t = j)$ for continuous $\underset{\sim}{x}_t$

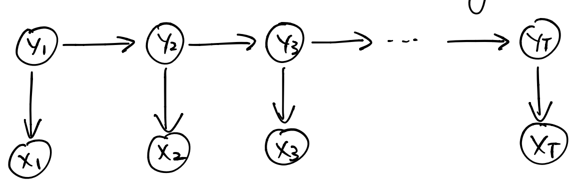
$A_{jk} = P(\underset{\sim}{x}_t = k \mid S_t = j)$ for discrete $\underset{\sim}{x}_t$.

i.e. $S_1 \sim \underset{\sim}{\pi}$

$S_t \mid S_{t-1} \sim \underset{\approx}{\phi}_{S_{t-1}}$

$\underset{\sim}{x}_t \mid S_t \sim \underset{\approx}{A}_{S_t}$

2. LGSSM (Kalman Filtering)



$\underset{\sim}{y}_1 \sim \mathcal{N}(\mu_0, Q_0)$. $\leftarrow$ Initial State.

$\underset{\sim}{y}_t \mid \underset{\sim}{y}_{t-1} \sim \mathcal{N}(A\underset{\sim}{y}_{t-1}, Q)$. $\leftarrow$ transition probability.

$\underset{\sim}{x}_t \mid \underset{\sim}{y}_t \sim \mathcal{N}(C\underset{\sim}{y}_t, R)$ $\leftarrow$ emmission probability.

General Structure

$$P(x_{1:T}, y_{1:T}) = P(y_1) \cdot \underbrace{\prod_{t=2}^{T} P(y_t \mid y_{t-1}) \cdot \prod_{t=1}^{T} P(x_t \mid y_t)}_{\prod_{t=2}^{T} f_t(y_t, y_{t-1}) \cdot \prod_{t=1}^{T} g_t(x_t, y_t)}. \quad \text{Joint pdf}$$

$\Rightarrow$ Comparison with factor graph

Learning: (M-step):

$$\operatorname{argmax} \langle \log P(x_{1:T}, y_{1:T}) \rangle_{q(y_{1:T})}$$

$$= \operatorname{argmax} \left\{ \langle \log P(y_1) \rangle_{q(y_1)} + \sum_{t=2}^{T} \langle \log P(y_t \mid y_{t-1}) \rangle_{q(y_t, y_{t-1})} + \sum_{t=1}^{T} \langle \log P(x_t \mid y_t) \rangle_{q(y_t)} \right\}.$$

E-step: $q(y_1) = P(y_1 | x_{1:T}, \theta)$ ← Hard to compute by brutal force.

We need $P(y_{1:T} | x_{1:T}, \theta)$ to calculate the marginal $P(y_1 | x_{1:T}, \theta)$.

But if each y has k state, $O(K^{T-1})$

Instead: $P(y_t | x_{1:T}, \theta) = P(x_{t+1:T} | y_t, \cancel{x_{1:t}}) \cdot P(y_t | x_{1:t})$

$$= \frac{1}{Z} \underbrace{P(x_{t+1:T} | y_t)}_{\text{② Backward}} \cdot \underbrace{P(y_t | x_{1:t})}_{\text{① Forward}}.$$

DERIVE FORWARD-BACKWARD FROM BELIEF PROPAGATION

① : $P(y_t | x_{1:t}) = \frac{1}{Z} P(y_t, x_{1:t})$

FORWARD.

$$= \frac{1}{Z} P(x_t | y_t) \cdot P(y_t, x_{1:t-1})$$

$$= \frac{1}{Z} P(x_t | y_t) \cdot \int P(y_t, y_{t-1}, x_{1:t-1}) dy_{t-1}$$

$$= \frac{1}{Z} P(x_t | y_t) \cdot \int P(y_t | y_{t-1}, x_{1:t-1}) \cdot P(y_{t-1}, x_{t-1}) dy_{t-1}$$

Define $\underline{M_{(t-1)\to t}(y_t)} = P(x_{1:t}, y_t)$

Message from the past

$$= P(x_t | y_t) \cdot \int P(y_t | y_{t-1}) \cdot P(x_{1:t-1}, y_{t-1}) dy_{t-1}$$

$$= \underbrace{P(x_t | y_t)}_{\text{Emission}} \cdot \int \underbrace{P(y_t | y_{t-1})}_{\text{Transition}} \underbrace{M_{t-2 \to t-1}(y_{t-1})}_{\text{Belief Message.}} dy_{t-1}$$

$$\alpha_t(i) = A_i(x_t) \cdot \left(\sum_{j=1}^{K} \phi_{ij} \cdot \alpha_{t-1}(j) \right).$$ Discrete case in HMM.

$M_{0\to 1}(y_t) = P(y_1, x_1) = P(y_1) \cdot P(x_1 | y_1)$.

$$P(y_t | x_{1:t}) = \frac{1}{Z} M_{(t-1)\to t}(y_t) = \frac{M_{(t-1)\to t}(y_t)}{\int M_{(t-1)\to t}(y) dy}.$$

② BACKWARD

Define $M_{t+1\to t}(y_t) = P(x_{t+1:T} \mid y_t)$

$$\underbrace{P(x_{t+1:T} \mid y_t)}_{\text{Message from future}} = \int P(x_{t+1:T}, y_{t+1} \mid y_t)\, dy_{t+1}$$

$$= \int P(y_{t+1} \mid y_t) \cdot P(x_{t+1} \mid y_t, y_{t+1}) \cdot P(x_{t+2:T} \mid y_{t+1})\, dy_{t+1}$$

↓ Markov property

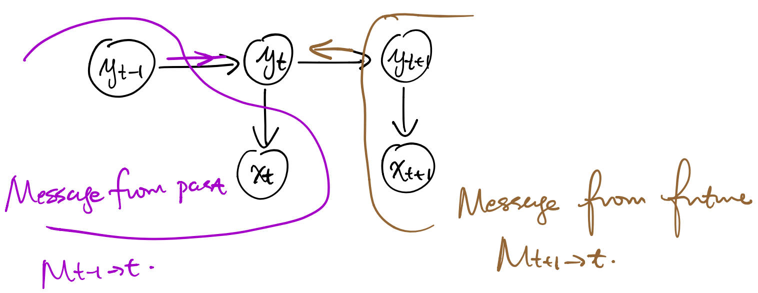
$$= \int P(y_{t+1} \mid y_t) \cdot P(x_{t+1} \mid y_{t+1}) \cdot P(x_{t+2:T} \mid y_{t+1})\, dy_{t+1}$$

$$= \int \underbrace{P(y_{t+1} \mid y_t)}_{\text{Transition}} \cdot \underbrace{P(x_{t+1} \mid y_{t+1})}_{\text{Emission}} \cdot \underbrace{M_{t+2\to t+1}(y_{t+1})}_{\text{Belief Message}}\, dy_{t+1}$$

$$\beta_t(i) = \sum_{j=1}^{K} \phi_{ij}\, A_j(x_{t+1})\, \beta_{t+1}(j).$$

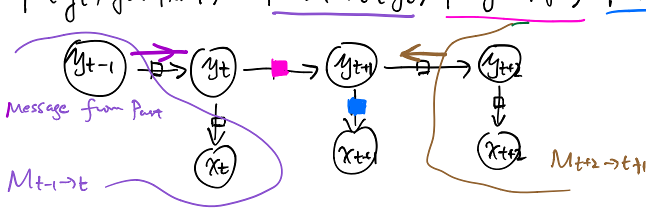
BACKWARD – FORWARD :

$$P(y_t \mid x_{1:T}) \propto \underline{M_{t-1\to t}(y_t)} \cdot P(x_{t+1} \mid y_{t+1}) \cdot \underline{M_{t+1\to t}(y_t)}.$$



Example

$$P(y_t, y_{t+1} \mid x_{1:T}) \propto \underline{M_{t-1\to t}(y_t)} \cdot \underline{P(y_{t+1} \mid y_t)} \cdot \underline{P(x_{t+1} \mid y_{t+1})} \cdot \underline{M_{t+2\to t+1}(y_{t+1})}$$ Message from future



Three Basic Problems in HMM & LGSSM.

1. Forward Recursion (Filtering) $P(y_t|x_{1:t})$

$$P(y_t|x_{1:t}) = \int P(y_t, y_{t-1}| x_t, x_{1:t-1})\, dy_{t-1}$$ ← Continuous Version of Chapman-Kolmogorov Equation

$$= \int \frac{P(x_t, y_t, y_{t-1}|x_{1:t-1})}{P(x_t|x_{1:t-1})}\, dy_{t-1}$$

$$\propto \int P(x_t|y_t, y_{t-1}, x_{1:t-1}) \cdot P(y_t|y_{t-1}, x_{1:t-1}) \cdot P(y_{t-1}|x_{1:t-1})\, dy_{t-1}$$

$$= \int P(x_t|y_t) \cdot P(y_t|y_{t-1}) \cdot P(y_{t-1}|x_{1:t-1})\, dy_{t-1}$$ Markov Property

$$= \underbrace{P(x_t|y_t)}_{\text{Emission}} \int \underbrace{P(y_t|y_{t-1})}_{\text{transition}} \cdot \underbrace{P(y_{t-1}|x_{1:t-1})}_{\text{"}\alpha_t(i)\text{"}}\, dy_{t-1}$$

HMM.

$$\boxed{\alpha_1(i) = \pi_i A_i(x_1),}$$ Initialise

$$\alpha_t(i) = P(x_{1:t}, S_t = i|\theta)$$

$$\alpha_t(i) = \overbrace{A_i(x_t)}^{\text{Emission}} \left(\sum_{j=1}^{K} \alpha_{t-1}(j) \cdot \overbrace{\phi_{ji}}^{\text{transition}} \right)$$

$$P(x_t|y_t) \cdot \int P(y_{t-1}|x_{1:t-1}) \cdot P(y_t|y_{t-1})$$

Recursion.

Get posterior:

$$P(S_t = i|x_{1:t}, \theta) = \frac{\alpha_t(i)}{\sum_{k=1}^{K} \alpha_t(k)}.$$

Evaluation Probability:

$$P(x_{1:T}|\theta) = \sum_{S_{1:T}} P(x_{1:T}, S_{1:T}, \theta)$$

$$= \sum_{k=1}^{K} \alpha_T(k).$$

LGSSM. (Kalman Filtering)

$y_1 \sim \mathcal{N}(\mu_0, Q_0)$, $y_t|y_{t-1} \sim \mathcal{N}(A y_{t-1}, Q)$.

$x_t|y_t \sim \mathcal{N}(C y_t, R)$ transition

Let $\hat{y}_1^0 = \mu_0$ and $\hat{V}_1^0 = Q_0$.

$$P(x_1, y_1) = P(x_1|y_1) \cdot P(y_1)$$

$$\propto \exp\left\{-\frac{1}{2}(x_1 - C\mu_0)^T R^{-1} (x_1 - C\mu_0)\right\}$$

$$\cdot \exp\left\{-\frac{1}{2}\mu_0^T Q_0^{-1} \mu_0\right\}$$

Initialise

$= \cdots$ Complete Square as FA & PPCA

$$\Rightarrow P(y_1|x_1) = \mathcal{N}\Big(\underbrace{\hat{y}_1^0 + K_1(x_1 - C\hat{y}_1^0)}_{\hat{y}_1},\ \underbrace{\hat{V}_1^0 - K_1 C \hat{V}_1^0}_{\hat{V}_1}\Big).$$

where $K_1 = \hat{V}_1^0 C^T (C \hat{V}_1^0 C^T + R)^{-1}$

Let $\hat{y}_t^\tau \equiv \mathbb{E}[y_t|x_{1:\tau}]$, $\hat{V}_t^\tau \equiv \mathrm{Var}[y_t|x_{1:\tau}]$.

Recursion.

$$P(y_t|x_{1:t-1}) = \int P(y_t|y_{t-1}) \cdot P(y_{t-1}|x_{1:t-1})\, dy_{t-1}$$

$$= \mathcal{N}\Big(\underbrace{A\hat{y}_{t-1}^{t-1}}_{\hat{y}_t^{t-1}},\ \underbrace{A\hat{V}_{t-1}^{t-1}A^T + Q}_{\hat{V}_t^{t-1}}\Big)$$

$$P(y_t|x_{1:t}) = \mathcal{N}\Big(\underbrace{\hat{y}_t^{t-1} + K_t(x_t - C\hat{y}_t^{t-1})}_{\hat{y}_t^t},\ \underbrace{\hat{V}_t^{t-1} - K_t C \hat{V}_t^{t-1}}_{\hat{V}_t^t}\Big)$$

$$\text{where } K_t = \overbrace{\hat{V}_t^{t-1} C^T}^{\langle yx^T\rangle} \cdot \overbrace{(C\hat{V}_t^{t-1}C^T + R)}^{\langle xx^T\rangle}{}^{-1}$$

2. BackWard Recursion (Smoothing) $P(y_t | x_{1:T})$.

$$P(y_t|x_{1:T}) = P(y_t|x_{1:t}, x_{t+1:T}) = \frac{P(y_t, x_{t+1:T}|x_{1:t})}{P(x_{t+1:T}|x_{1:t})}$$

$$= \frac{P(x_{t+1:T}|y_t)\, P(y_t|x_{1:t})}{P(x_{t+1:T}|x_{1:t})}.$$

HMM

$$\beta_t(i) \equiv P(x_{t+1:T}|S_i = i)$$

$$= \sum_{j=1}^{K} P(S_{t+1}=j, x_{t+1}, x_{t+2:T}|S_t=i)$$

$$= \sum_{j=1}^{K} P(S_{t+1}=j|S_t=i)\cdot P(x_{t+1}|S_{t+1}=j) \cdot P(x_{t+2:T}|S_{t+1}=j)$$

$$= \sum_{j=1}^{K} \phi_{ij} A_j(x_{t+1})\, \beta_{t+1}(j).$$

$$\gamma_t(i) \equiv P(S_t=i|x_{1:T})$$

$$= \frac{P(S_t=i, x_{1:t})\cdot P(x_{t+1:T}|S_i=i)}{P(x_{1:T})}$$

$$= \frac{\alpha_t(i)\cdot\beta_t(i)}{\sum_j \alpha_j(i)\,\beta_j(i)}.$$

LGSMM (Kalman Smoothing)

$$P(y_t|x_{1:T}) = \int P(y_t, y_{t+1}|x_{1:T})\, dy_{t+1}$$

$$= \int P(y_t|y_{t+1}, \underline{x_{1:T}})\cdot P(y_{t+1}|x_{1:T})\, dy_{t+1}$$

Markov Property

$$= \int P(y_t|y_{t+1}, \underline{x_{1:t}})\cdot P(y_{t+1}|x_{1:T})\, dy_{t+1}$$

Backward Recursion:

$$J_t = \hat{V}_t^t A^T (\hat{V}_{t+1}^t)^{-1}$$

$$\hat{y}_t^T = \hat{y}_t^t + J_t(\hat{y}_{t+1}^T - A\hat{y}_t^t).$$

$$\hat{V}_t^T = \hat{V}_t^t + J_t(\hat{V}_{t+1}^T - \hat{V}_{t+1}^t) J_t^T$$

EM learning for SMM.

$y_1 \sim \mathcal{N}(\mu_0, Q_0)$, $y_t | y_{t-1} \sim \mathcal{N}(Ay_{t-1}, Q)$, $x_t | y_t \sim \mathcal{N}(Cy_t, R)$

Parameters : $\theta = \{\mu_0, Q_0, A, Q, C, R\}$

Free Energy : $\mathcal{F}(q, \theta) = \int \log\left(\frac{P(x_{1:T}, y_{1:T}|\theta)}{q(y_{1:T})}\right) \cdot q(y_{1:T}) \cdot dy_{1:T}$.

E - step : max $\mathcal{F}$ w.r.t q with θ fixed.

$q^*(y) = P(y|x, \theta)$.

Two-state extension of Kalman smoother.

M - step : max $\mathcal{F}$ w.r.t. θ with q fixed.

$$P(y_{1:T}, x_{1:T}|\theta) = P(y_1) \cdot P(x_1|y_1) \cdot \prod_{t=2}^{T} P(y_t|y_{t-1}) \cdot P(x_t|y_t).$$

* M - step for C and R

$$P(x_t|y_t) \propto |R|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x_t - Cy_t)^T R^{-1}(x_t - Cy_t)\right\}.$$

$$\Rightarrow \quad C_{new} = \underset{C}{\operatorname{argmax}} \left\langle \sum_t \log P(x_t|y_t) \right\rangle_q$$

$$= \underset{C}{\operatorname{argmax}} \left\langle -\frac{1}{2}\sum_t (x_t - Cy_t)^T \cdot R^{-1}(x_t - Cy_t) \right\rangle_q + \text{constant}.$$

$$= \underset{C}{\operatorname{argmax}} \left\{ -\frac{1}{2} \sum_t x_t^T R^{-1} x_t - 2x_t^T R^{-1} C\langle y_t \rangle + \langle y_t^T C^T R^{-1} C y_t \rangle \right\} + \text{constant}$$

$$= \underset{C}{\operatorname{argmax}} \left\{ \underline{\operatorname{Tr}\left[C \sum_t \langle y_t \rangle x_t^T R^{-1}\right]} - \frac{1}{2}\underline{\operatorname{Tr}\left[C^T R^{-1} C \left\langle \sum_t y_t y_t^T \right\rangle\right]} \right\} + \text{constant}'.$$

$$\frac{\partial\{\cdot\}}{\partial C} = \underline{R^{-1}\sum_t x_t \langle y_t \rangle^T} - \underline{R^{-1} C \left\langle \sum_t y_t y_t^T \right\rangle} = 0 \Rightarrow C_{new} = \frac{\sum_t x_t \langle y_t \rangle^T}{\sum_t \langle y_t y_t^T \rangle}$$

$$\Rightarrow R_{new} = \underset{R}{\operatorname{argmax}} \left\langle \sum_t \log P(x_t|y_t) \right\rangle_q$$

$$= \underset{R}{\operatorname{argmax}} \left\langle -\frac{1}{2}\sum_t \left((x_t - Cy_t)^T \cdot R^{-1}(x_t - Cy_t)\right) - \sum_t \frac{1}{2}\log|R| \right\rangle_q + \text{constant}.$$

$$= \underset{R}{\operatorname{argmax}} \left\{ -\frac{1}{2} \sum_t \left(x_t^T R^{-1} x_t - 2x_t^T R^{-1} C\langle y_t \rangle + \langle y_t^T C^T R^{-1} C y_t \rangle\right) - \frac{T}{2}\log|R| \right\} + \text{constant}$$

$$= \underset{R}{\operatorname{argmax}} \left\{ -\frac{1}{2}\operatorname{Tr}\left[R^{-1}\sum_t x_t^T x_t\right] + \operatorname{Tr}\left[R^{-1} C \sum_t \langle y_t \rangle \cdot x_t^T\right] - \frac{1}{2}\operatorname{Tr}\left[R^{-1} C \left\langle \sum_t y_t y_t^T \right\rangle C^T\right] + \frac{T}{2}\log|R^{-1}| \right\}$$

$$\frac{\partial\{\cdot\}}{\partial R^{-1}} = -\frac{1}{2}\sum_t x_t x_t^T + \sum_t x_t \langle y_t\rangle^T \cdot C^T - \frac{1}{2} C \left\langle \sum_t y_t y_t^T \right\rangle C^T + \frac{T}{2} R = 0 .$$

$$\Rightarrow R_{new} = \frac{1}{T}\left[\sum_t x_t x_t^T - 2\sum_t x_t \langle y_t\rangle^T C_{new}^T + C_{new} \left\langle \sum_t y_t y_t^T \right\rangle C_{new}^T\right]$$

Since $C_{new} = \dfrac{\sum_t x_t \langle y_t\rangle^T}{\sum_t \langle y_t y_t^T\rangle}$ $\Rightarrow \sum_t \langle y_t y_t^T\rangle = \dfrac{1}{C_{new}} \sum_t x_t \langle y_t\rangle^T$

Substitute to R_{new} $\Rightarrow$

$$R_{new} = \frac{1}{T}\left[\sum_t x_t x_t^T - \sum_t x_t \langle y_t\rangle^T C_{new}^T\right]$$

* M-step for A and Q:

$$P(y_{t+1}|y_t) \propto |Q|^{-1/2} \cdot \exp\left\{-\frac{1}{2}(y_{t+1} - Ay_t)^T Q^{-1} (y_{t+1} - Ay_t)\right\}$$

$$\Rightarrow A_{new} = \arg\max_A \left\langle \sum_t \log P(y_{t+1}|y_t) \right\rangle$$

$$= \arg\max_A \left\langle -\frac{1}{2}\sum_t (y_{t+1} - Ay_t)^T Q^{-1} (y_{t+1} - Ay_t) \right\rangle_q + \text{constant} .$$

$$= \arg\max_A \left\{ -\frac{1}{2}\sum_t \left(y_{t+1}^T Q^{-1} y_{t+1} - 2\langle y_{t+1}^T Q^{-1} A y_t\rangle + \langle y_t^T A^T Q^{-1} A y_t\rangle \right)\right\} + \text{constant}$$

$$= \arg\max_A \left\{ \mathrm{Tr}\left[A \sum_t \langle y_t y_{t+1}^T\rangle \cdot Q^{-1}\right] - \frac{1}{2}\mathrm{Tr}\left[A^T Q^{-1} A \sum_t \langle y_t y_t^T\rangle\right]\right\} + \text{constant}'$$

$$\frac{\partial\{\cdot\}}{\partial A} = Q^{-1}\sum_t \langle y_{t+1} y_t^T\rangle - Q^{-1} A \sum_t \langle y_t y_t^T\rangle = 0$$

$$\Rightarrow A_{new} = \frac{\sum_t \langle y_{t+1} y_t^T\rangle}{\sum_t \langle y_t y_t^T\rangle}$$

$$\Rightarrow Q_{new} = \arg\max_Q \left\{ -\frac{1}{2}\sum_{t=2}^{T}\left(\langle y_t^T Q^{-1} y_t\rangle - 2\langle y_t^T Q^{-1} A y_{t-1}\rangle + \langle y_{t-1}^T A^T Q^{-1} A y_{t-1}\rangle - \frac{1}{2}\log|Q| \right)\right\} + \text{constan}$$

$$= \arg\max_Q \left\{ -\frac{1}{2}\mathrm{Tr}\left[Q^{-1}\sum_{t=2}^{T}\langle y_t y_t^T\rangle\right] + \mathrm{Tr}\left[Q^{-1} A \sum_{t=2}^{T}\langle y_{t-1} y_t^T\rangle\right] - \frac{1}{2}\mathrm{Tr}\left[Q^{-1} A \sum_{t=2}^{T}\langle y_{t-1} y_{t-1}^T\rangle A^T\right] + \frac{T-1}{2}\log|Q| \right\} + \text{constant}$$

$$\frac{\partial\{\cdot\}}{\partial Q^{-1}} = -\frac{1}{2}\sum_{t=2}^{T}\langle y_t y_t^T\rangle + \sum_{t=2}^{T}\langle y_t y_{t-1}^T\rangle \cdot A^T - \frac{1}{2} A \sum_{t=2}^{T}\langle y_{t-1} y_{t-1}^T\rangle A^T + \frac{T-1}{2} Q = 0 .$$

$$A_{new} = \frac{\sum_{t=1}^{T-1}\langle y_{t+1} y_t^T\rangle}{\sum_{t=1}^{T-1}\langle y_t y_t^T\rangle} = \frac{\sum_{t=2}^{T}\langle y_t y_{t-1}^T\rangle}{\sum_{t=2}^{T}\langle y_{t-1} y_{t-1}^T\rangle}$$

$$\Rightarrow Q_{new} = \frac{1}{T-1}\left[\sum_{t=2}^{T}\langle y_t y_t^T\rangle - \left(\sum_{t=2}^{T}\langle y_t y_{t-1}^T\rangle\right) \cdot A_{new}^T\right] \quad \square .$$