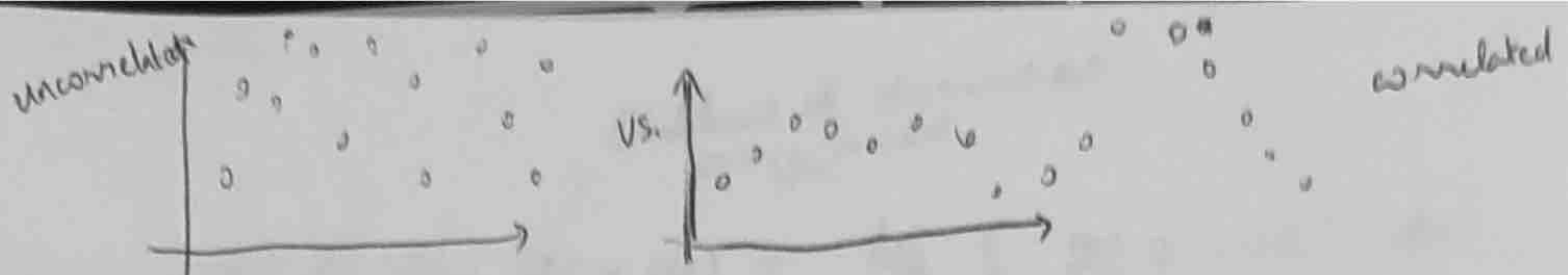
Last time:



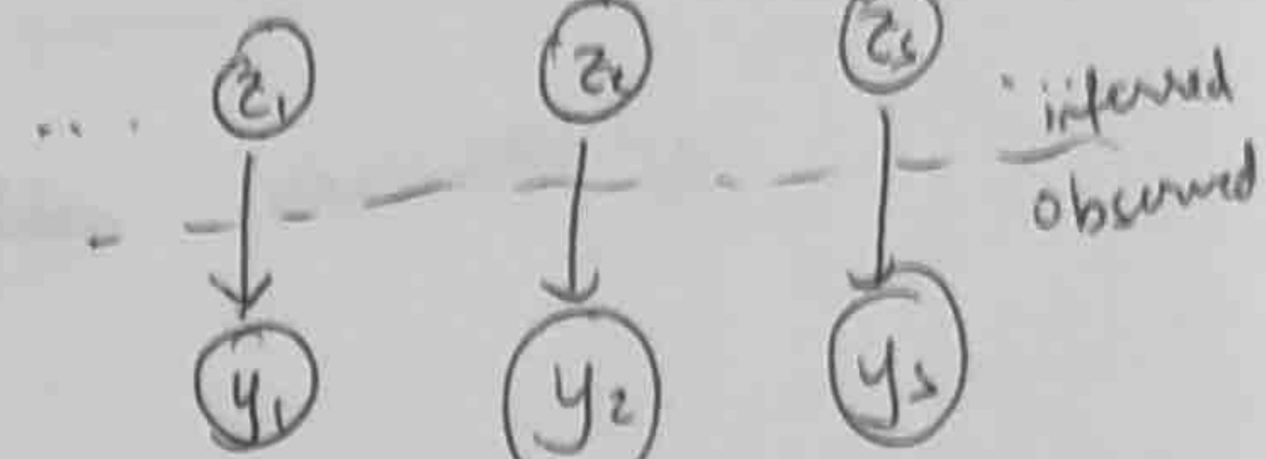
$$y_n = C z_n + \varepsilon_n$$

$\varepsilon_n \sim \mathcal{N}(0, R)$

$z_n \in \mathbb{R}^D$ $\quad D << K$

$y_n \in \mathbb{R}^K$

y_n iid



$$z_t = \underbrace{A}_{\text{transition matrix}} z_{t-1} + \underbrace{\eta_t}_{\text{state evolution noise}}$$

$$y_t = \underbrace{C}_{\text{observation matrix}} z_t + \underbrace{\varepsilon_t}_{\text{observation noise}}$$

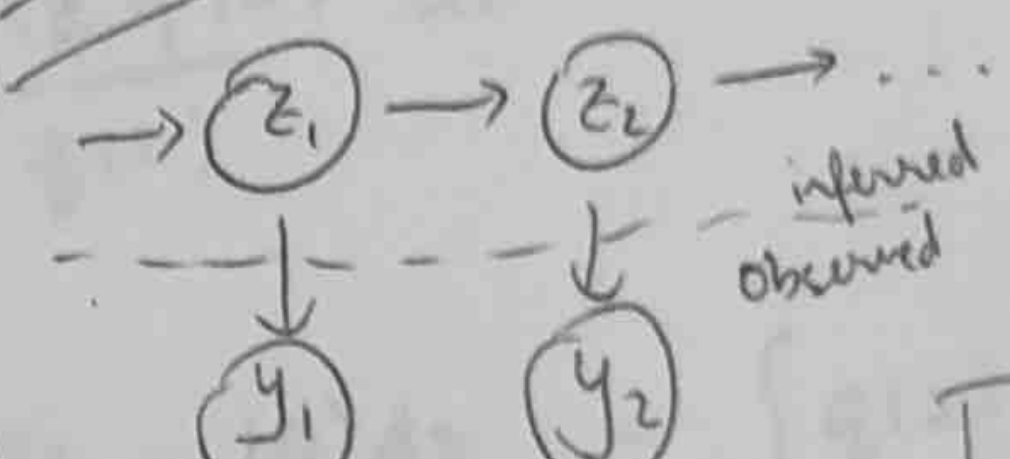
~~y_t i.i.d.~~

$\eta_t \sim \mathcal{N}(0, Q)$

$z_0 \sim \mathcal{N}(\mu_0, V_0)$

this & $\varepsilon_t \sim \mathcal{N}(0,R)$ means that:

Model Recap



Note

we assume uncorrelated noise from time step to time step.

If we think noise is correlated, expand state space to incorporate it.

$$p_\theta(z_t \mid z_{t-1}) = \mathcal{N}(A z_{t-1}, Q)$$

$$p_\theta(y_t \mid z_t) = \mathcal{N}(C z_t, R)$$

$$p_\theta(z_0) = \mathcal{N}(\mu_0, V_0)$$

$$\theta = [A, C, R, Q, \mu_0, V_0]$$

$\Rightarrow$ Jointly gaussian

$p_\theta(\{z\}, \{y\}) = \mathcal{N}(\mu, \Sigma)$

could infer these directly!

$$\Rightarrow P_\theta(\{z\}, \{y\}) = P_\theta(z_0) \prod_{t=1}^{T} P_\theta(y_t \mid z_t)\, P_\theta(z_t \mid z_{t-1})$$

$$\propto \prod_{t=1}^{T} \exp \left\{ \begin{array}{l} -\frac{1}{2}(y_t - C z_t)^T R^{-1} (y_t - C z_t) \\ -\frac{1}{2}(z_t - A z_{t-1})^T Q^{-1} (z_t - A z_{t-1}) \\ -\frac{1}{2}(z_0 - \mu_0)^T V_0^{-1} (z_0 - \mu_0) \end{array} \right\}$$

Learning & Estimation

① Inference: given θ, & $\{y_1, \dots y_T\}$ find $\{z_1, \dots z_T\}$

② Learning / System identification: given $\{y_1, \dots y_T\}$, find θ.

multiple ways of doing this. we will focus on EM!

EM (Recap)

likelihood of observed data over given model

We want to maximize $\mathcal{L}(\theta) \triangleq \log p(Y|\theta) = \underline{\log \int_z p(z, Y|\theta)\, dz}$

We hypothesize a distribution q over the hidden variables z

$$\log \int_z p(z, Y|\theta)\, dz = \log \int_z q(z) \frac{p(z, y|\theta)}{q(z)}\, dz$$

Jensen's $\log \mathbb{E}[x] \geq \mathbb{E}[\log x]$ for concave functions (log)

$$\geq \int_z q(z) \log \frac{p(z, y|\theta)}{q(z)}\, dz \quad ⊛$$

$$= \int_z q(z) \log p(z, y|\theta)\, dz - \int_z q(z) \log q(z)\, dz$$

$$= \mathbb{E}_{q(z)}\left[\log p(z, y|\theta)\right] + \underbrace{H(q)}_{\text{entropy}} \equiv \underbrace{\mathcal{F}(q, \theta)}_{\text{lower bound}}$$

E-step

$$q_{k+1} \longleftarrow \underset{q}{\operatorname{argmax}}\ \mathcal{F}(q, \theta_k)$$

set

$$q_{k+1}(z) = p(z|y, \theta_k)$$

⊛ Note: this is optimal since: want to maximize

$$\mathcal{F}(q,\theta) = \int_z q(z) \log \frac{p(z,y|\theta)}{q(z)}\, dz = \int_z q(z) \log \frac{p(z|y,\theta)\, p(y|\theta)}{q(z)}\, dz$$

$$= \int_z q(z) \log \frac{p(z|y,\theta)}{q(z)} + \underbrace{\int_z q(z) \log p(y|\theta)\, dz}_{\log p(y|\theta) \cancel{\int_z q(z) dz}}$$

$$= -KL\left(q(z) \,||\, p(z|y,\theta)\right) + \log p(y|\theta)$$

$\Rightarrow \log p(y|\theta) = \mathcal{F}(q, \theta)$ at every E-step!

M-step

$$\theta_{k+1} \longleftarrow \underset{\theta}{\operatorname{argmax}}\ \mathcal{F}(q_{k+1}, \theta) = \underset{\theta}{\operatorname{argmax}} \int_z p(z|y, \theta_k) \log p(z, y|\theta)\, dz$$

(no dependence on q of θ)

$$p(z_t|x) = \frac{p(x|z_t)\,p(z_t)}{p(x)}$$

E - step

Kalman filter

want z_t given $\underbrace{z_{t-1}}_{\text{expectation}}$ and $\underbrace{y_t}_{\text{reality}}$

$\cancel{z_t \sim N(\mu_t, R_t)}$

$$p(z_t \mid \{y_1, \dots, y_t\}) \equiv N(\mu_t, \Sigma_t)$$

$$= \frac{p(y_t \mid z_t)\, p(z_t)}{p(\{y\})}$$

$$\boxed{\hat{z}_t = A \hat{z}_{t-1} + \underbrace{K_t}_{\text{Kalman gain}} \underbrace{(y_t - C A \hat{z}_{t-1})}_{\text{error from estimated value of } y_t}}$$

$$= \frac{1}{c_t} \underbrace{p(y_t \mid z_t)}_{N(Cz_t, R)} \int \underbrace{p(z_{t-1} \mid \{y_{t-1}\})}_{N(\mu_{t-1}, \Sigma_{t-1})} \underbrace{p(z_t \mid z_{t-1})}_{N(Az_{t-1}, Q)} \, dz_{t-1}$$

skip hairy math

$$\mu_t = \overbrace{A\mu_{t-1}}^{\text{expectation}} + \overbrace{K_t (y_t - CA\mu_{t-1})}^{\text{correction}} \quad \text{where } P_{t-1} = A\Sigma_{t-1}A^T + Q$$

$$\Sigma_t = \underbrace{(I - K_t C)}_{\text{correction}} \underbrace{P_{t-1}}_{\text{expectation}}$$

$$K_t = P_{t-1} C^T (C P_{t-1} C^T + R)^{-1}$$

$$K_t \sim \frac{\overbrace{\text{state evolution noise}}^{\sim P_{t-1}}}{\underbrace{\text{observation noise}}_{\sim R}}$$

Kalman smoothing
+ Backward recursion $\rightarrow p(z \mid y, \theta) = q_{k+1}(z)$

M- step

$$\theta_{k+1} \leftarrow \arg\max_\theta F(q_{k+1}, \theta) = \arg\max_\theta \mathbb{E}_{p(z|y,\theta)} [\log p(z, y \mid \theta)]$$

$$= \arg\max_\theta \mathbb{E}_{p(z|y,\theta)} \Big[-\frac{1}{2}(z_0 - \mu_0)^T V_0^{-1} (z_0 - \mu_0) - \frac{1}{2}\sum_{t=1}^{T} (z_t - Az_{t-1})^T Q^{-1} (z_t - Az_{t-1}) - \frac{1}{2}\sum_{t=1}^{T} (y_t - Cz_t)^T R^{-1} (y_t - Cz_t) \Big] + \text{const}$$

$$\mu_0^{new} = \mathbb{E}[z_0]$$
$$V_0^{new} = \mathbb{E}[z_0 z_0^T] - \mathbb{E}[z_0]\mathbb{E}[z_0^T]$$

$$\cancel{\frac{d\mu_0^T V^{-1} z_0}{d\mu_0} = 0}$$

$$\frac{d z_0^T V_0^{-1} \mu_0}{d\mu_0} = 0$$

$$A_{new} = \left(\sum_{t=1}^{T} \mathbb{E}[z_t z_{t-1}^T]\right) \left(\sum_{t=1}^{T} \mathbb{E}[z_{t-1} z_{t-1}^T]\right)^{-1}$$

$$Q^{new} = \dots$$

Model recap

$$z_{t} = A z_{t-1} + \eta_t$$

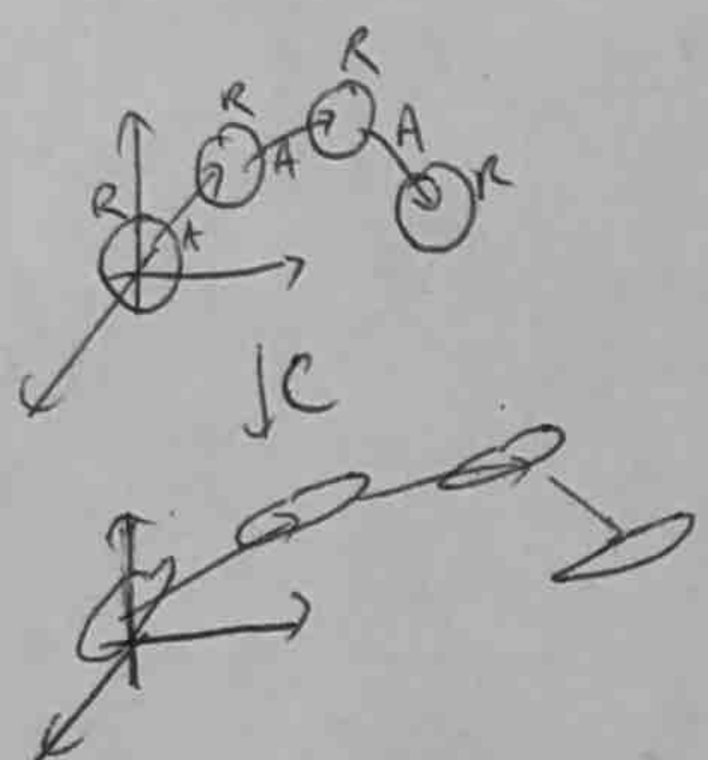
$$y_t = C z_t + \varepsilon_t$$

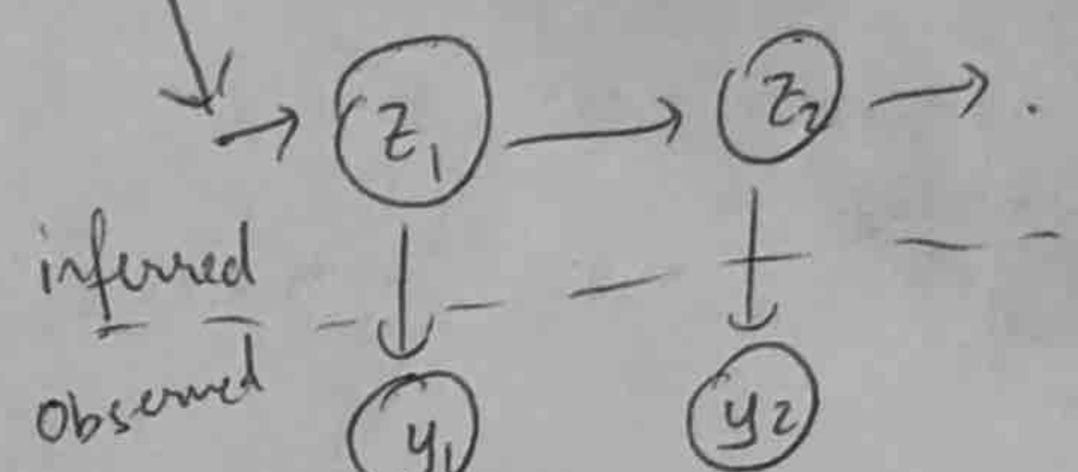
$A=0$, $Q=I$, $R \neq \mathrm{diag}$ $\Rightarrow$ Factor analysis

(R diag implies recording axes are special)

$A=0$, $Q=I$, $R = \lim_{\varepsilon \to 0} \varepsilon I$ $\Big\}$ PCA

intuition:





discrete valued z_i $\Rightarrow$ HMM

Use in neuroscience:

- Dimensionality reduction - observations coming from recordings + smoothly evolving latents
  - ⇘ way to incorporate assumptions on temporally smooth data
- Prediction (BMI applications) (Shenoy)

Extensions

- Switching LDS
  $$k \sim \Pi(P)$$
  $$z_t = A_k z_{t-1} + \eta_t + b_k$$
  $$y_t = C z_t + \varepsilon_t$$
  → EM
- "Nonlinear" observation model:
  $$z_t = A z_{t-1} + \eta_t$$
  $$y_t = f_\theta(z_t)$$
  example: $y_t \sim \text{Poisson}(f_\theta(z_t))$ ↑ EM, MLP? PPM?
- ~~Nonlinear~~ nonlinear/ Locally linear dynamics:
  $$z_t = A z_{t-1} + g_\theta(z_{t-1}) \rightarrow \text{EM?}$$
  $$y_t = f_\theta(z_t)$$