Lecture 11
Dynamic Bayesian Networks
Linear Dynamical Systems

Lei Li and Yu-xiang Wang

UCSB

Recap

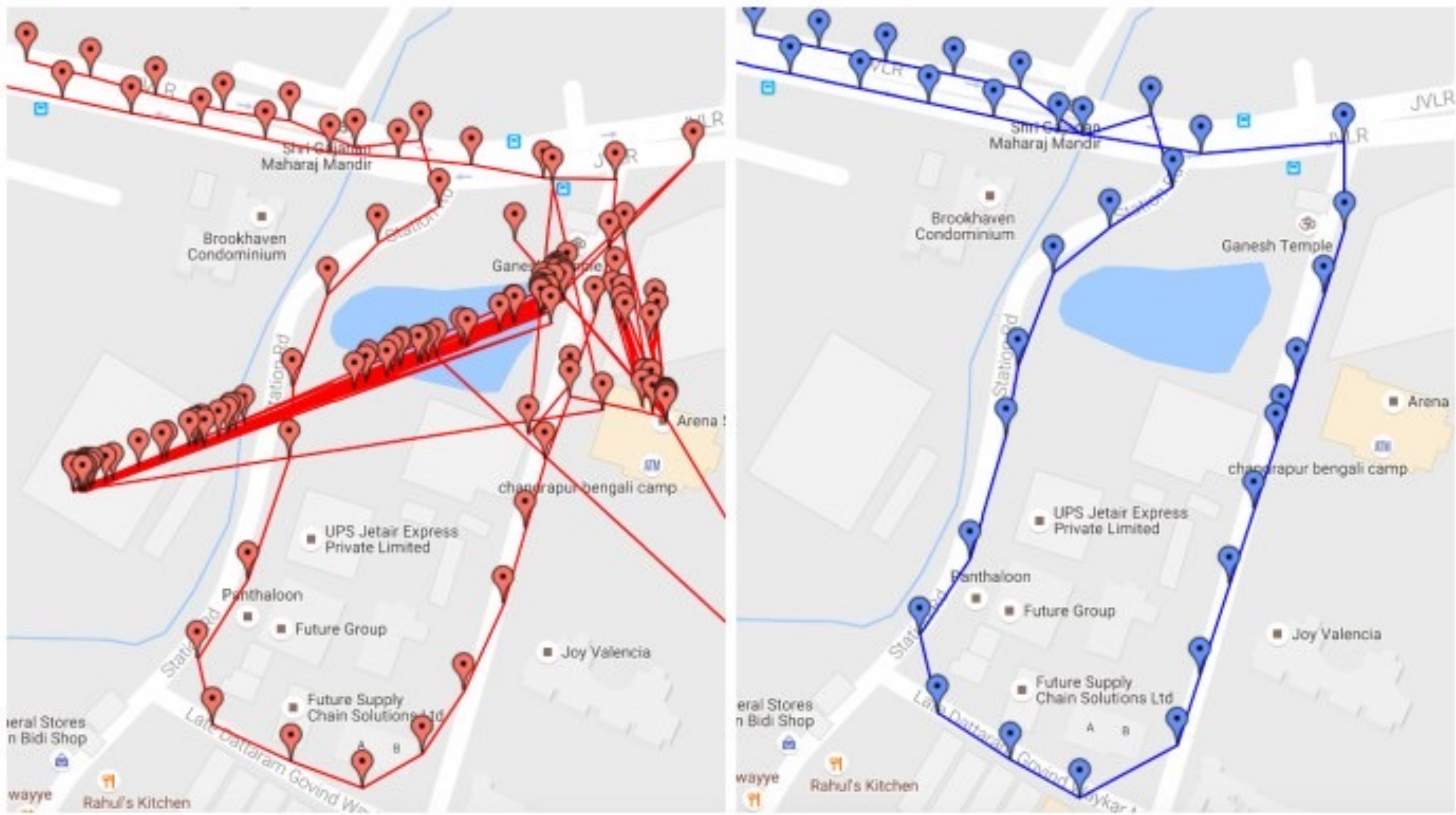
- Bayesian networks:
  - Directed acyclic graph
  - Nodes are random variables
  - arcs are probabilistic dependencies
- Mixture of Gaussian Model
- Expectation-Maximization

Dynamic Bayesian Networks

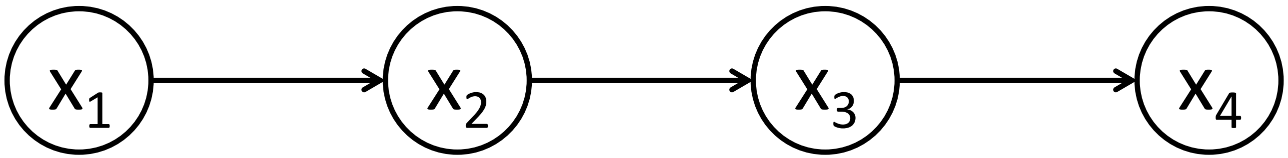
- What about non-IID data / sequential data
- Markov assumption

- GMM => Sequential => HMM
- PPCA → Sequential → LDS

Estimating the true trajectory

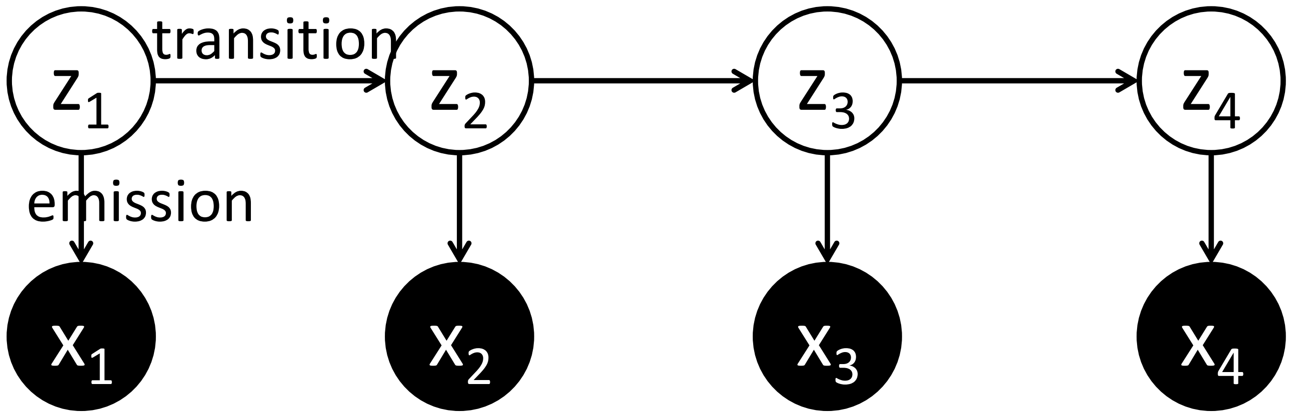


Markov Process



- Markov chain
- Current value only dependent on the previous step

Linear Dynamical Systems



$$z_1 \sim \mathcal{N}(\mu_0, Q_0)$$

$$z_{n+1} \mid z_n \sim \mathcal{N}(A \cdot z_n, Q)$$

$$x_n \mid z_n \sim \mathcal{N}(C \cdot z_n, R)$$

$$z_n = \begin{pmatrix} pos_n \\ vel_n \\ acc_n \end{pmatrix}, \quad x_n = (\text{obs\_pos})_n$$

$$z_{n+1} = \begin{pmatrix} 1 & \tau & \frac{\tau^2}{2} \\ 0 & 1 & \tau \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} pos \\ vel \\ acc \end{pmatrix}$$

$$x_n = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} pos \\ vel \\ acc \end{pmatrix}$$

Learning LDS

- EM again
- $\arg\max_{\theta} E_{p(z_{1..N}|x_{1..N};\theta_{old})} \log p(x_{1..N}, z_{1..N}|\theta)$
- E-step: estimate $p(z_n|x_{1..N})$ and $p(z_n, z_{n+1}|x_{1..N})$
- M-step: optimizing for params

Objective: Expected log-likelihood

- $E_{p(z_{1..N}|x_{1..N};\theta_{old})} \log p(x_{1..N}, z_{1..N}|\theta)$

$$= \log \prod_{n=2}^{N} p(z_n|z_{n-1}) \cdot p(z_1) \prod_{n=1}^{N} p(x_n|z_n)$$

$$= E_{z|x} \Big[-\frac{1}{2}\log|V_0| - \frac{1}{2}(z_1-\mu_0)^T V_0^{-1}(z_1-\mu_0)$$

$$-\frac{N-1}{2}\log|Q| - \frac{1}{2}\sum_{n=2}^{N}(z_n - A\cdot z_{n-1})^T Q^{-1}(z_n - A\cdot z_{n-1})$$

$$-\frac{N}{2}\log|R| - \frac{1}{2}\sum_{n=1}^{N}(x_n - C\cdot z_n)^T R^{-1}(x_n - C\cdot z_n)$$

Maximization

Estimating $p(z_n|x_{1..N})$

- Forward-backward algorithm
- Forward: also known as Kalman filter, estimate filtering density $p(z_n|x_{1..n})$
- Backward: also known as Kalman smoothing, estimate smoothing density $p(z_n|x_{1..N})$

Forward: $p(z_n|x_{1..n})$ $= \hat{\alpha}(z_n)$

$p(z_{n-1}|x_1 \ldots x_{n-1})$

$\hat{\alpha}(z_{n-1}) \longrightarrow \hat{\alpha}(z_n)$ $= N(\mu_n, V_n)$

$\Downarrow$

$N(\mu_{n-1}, V_{n-1})$

$\Downarrow$

$p(z_n|z_{n-1}) = N(A \cdot z_{n-1}, Q)$

$\Downarrow$

$p(z_n|x_1 \cdots x_{n-1}) \simeq N(A \cdot \mu_{n-1}, A V_{n-1} A^T + Q)$

$p(x_n|z_n) = N(C \cdot z_n, R)$

$\Rightarrow p(z_n|x_1 \ldots x_{n-1}, x_n) = \hat{\alpha}(z_n) =$

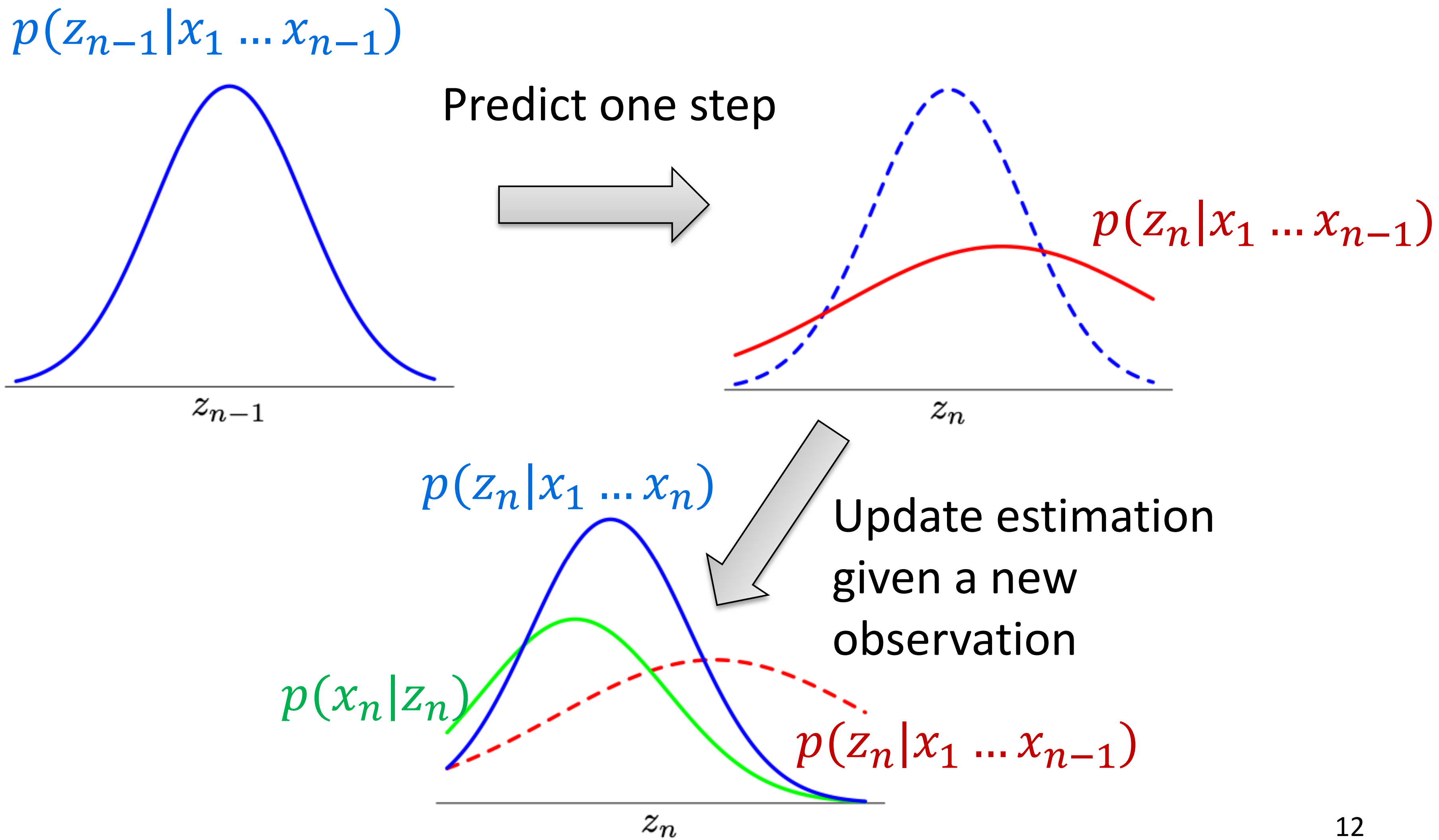
$P_{n-1} = A \cdot V_{n-1} A^T + Q$

$V_n = (I - K_n \cdot C) \cdot P_{n-1}$

$K_n = P_{n-1} C^T (C \cdot P_{n-1} C^T + R)^{-1}$

$\mu_n = A \cdot \mu_{n-1} + K_n (x_n - C \cdot A \cdot \mu_{n-1})$

What does Kalman filter (forward-pass) do?



Backward: $p(z_n|x_{1..N})$

EM for LDS

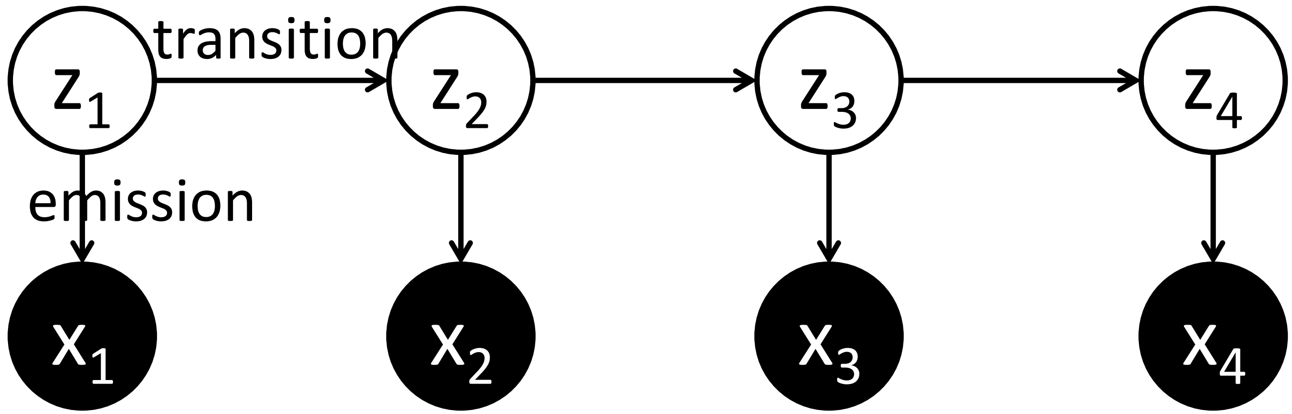
- Observation: $x_{1..N}$
- $\theta = \{\mu_0, Q_0, A, Q, C, R\}$
- Iterate until convergence
  1. E step: use X and current θ to calculate marginal posterior mean E[z|x] and co-variance Cov[z|x]
     - Using forward (Kalman filtering) and backward (Kalman smoothing)
  2. M step:

$$\theta \leftarrow \arg\max_{\theta} E_{p(z_{1..N}|x_{1..N};\theta_{old})} \log p(x_n, z_n|\theta)$$

Application of LDS

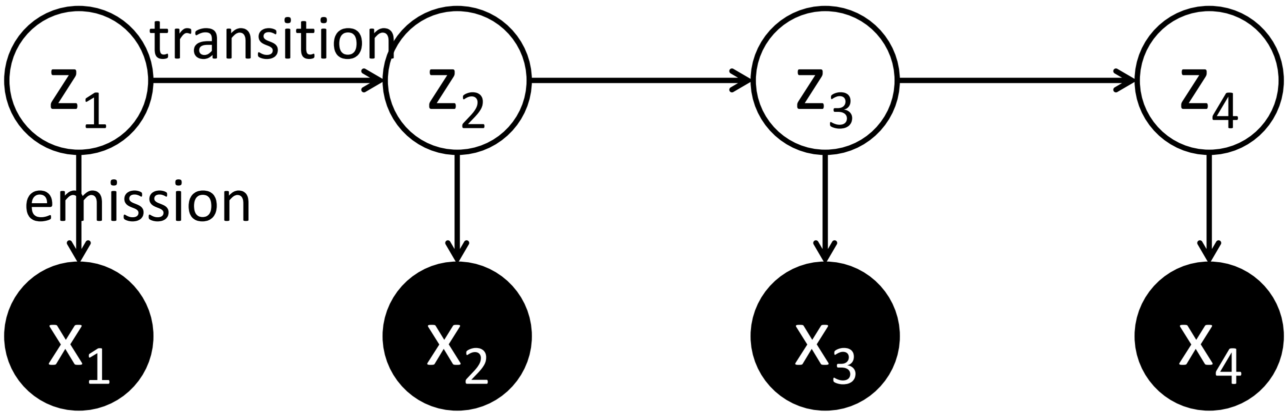
- Kalman filter: Tracking object movement
- Time series forecasting

Hidden Markov Model



- Same graph topology, but different distribution
- Sequential version of GMM
- Transition: a probability matrix
- Emission: Gaussian
- Wide applications in Speech, Communication, Genetics

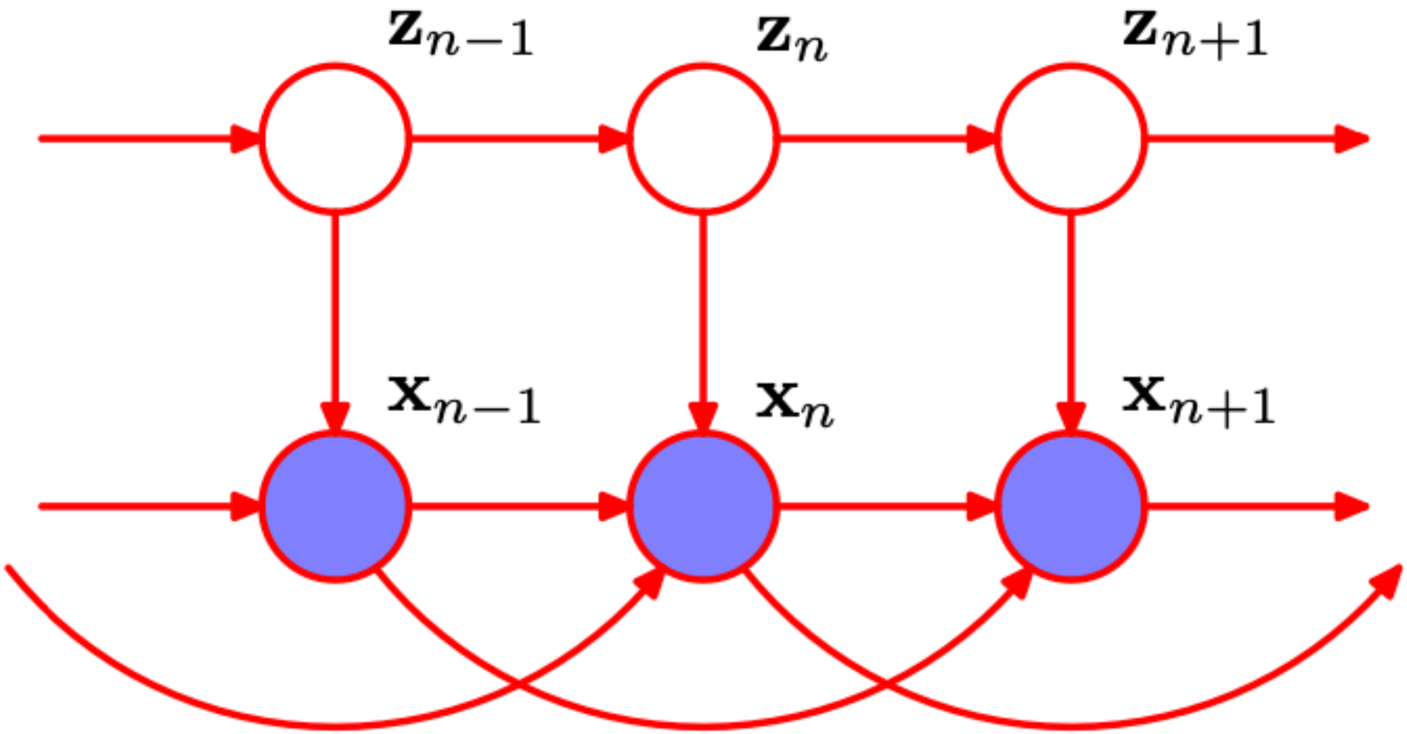
Hidden Markov Model



- Very similar algorithm
- Inference: $p(z_n|x_1,\ldots,x_N)$ using forward-backward
- Learning: same EM alg as LDS (different update eq.), also known as Baum-Welch alg.
- Decoding: finding max prob. codes for z, again forward-backward, also known as Viterbi alg.

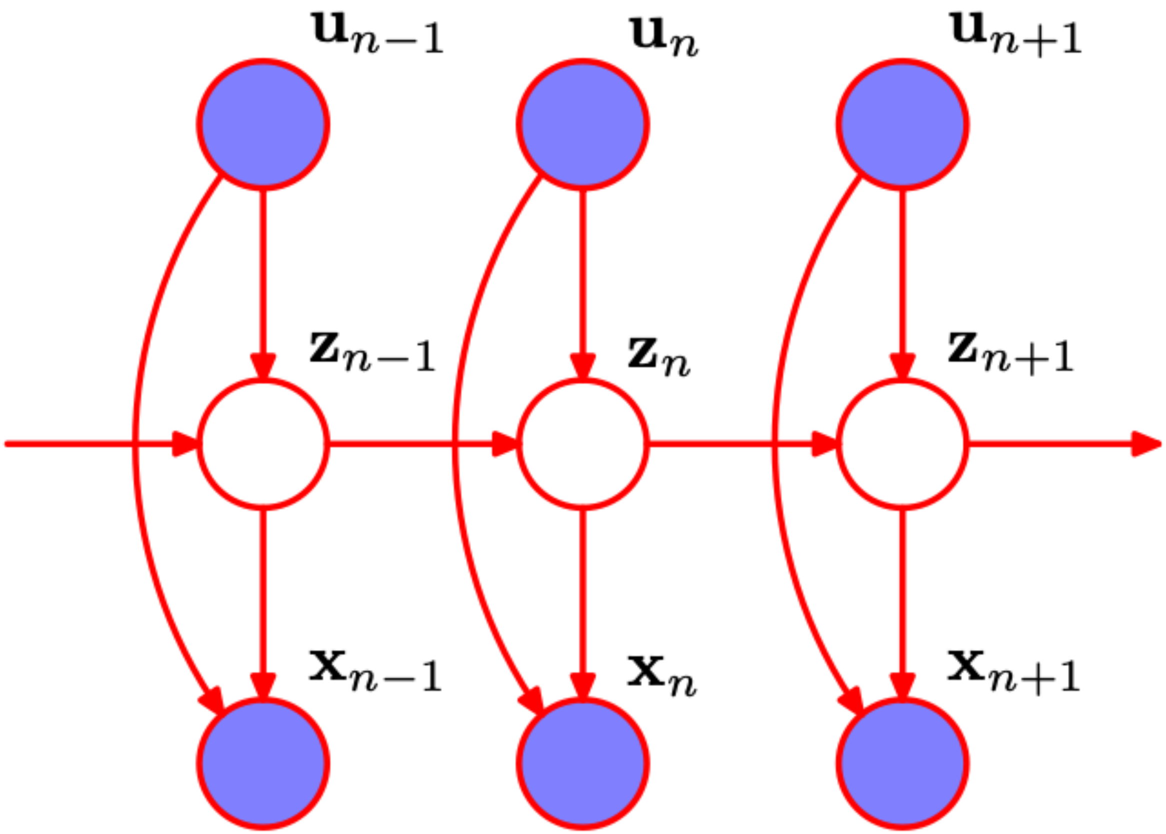
Andrew Viterbi

Other Variations



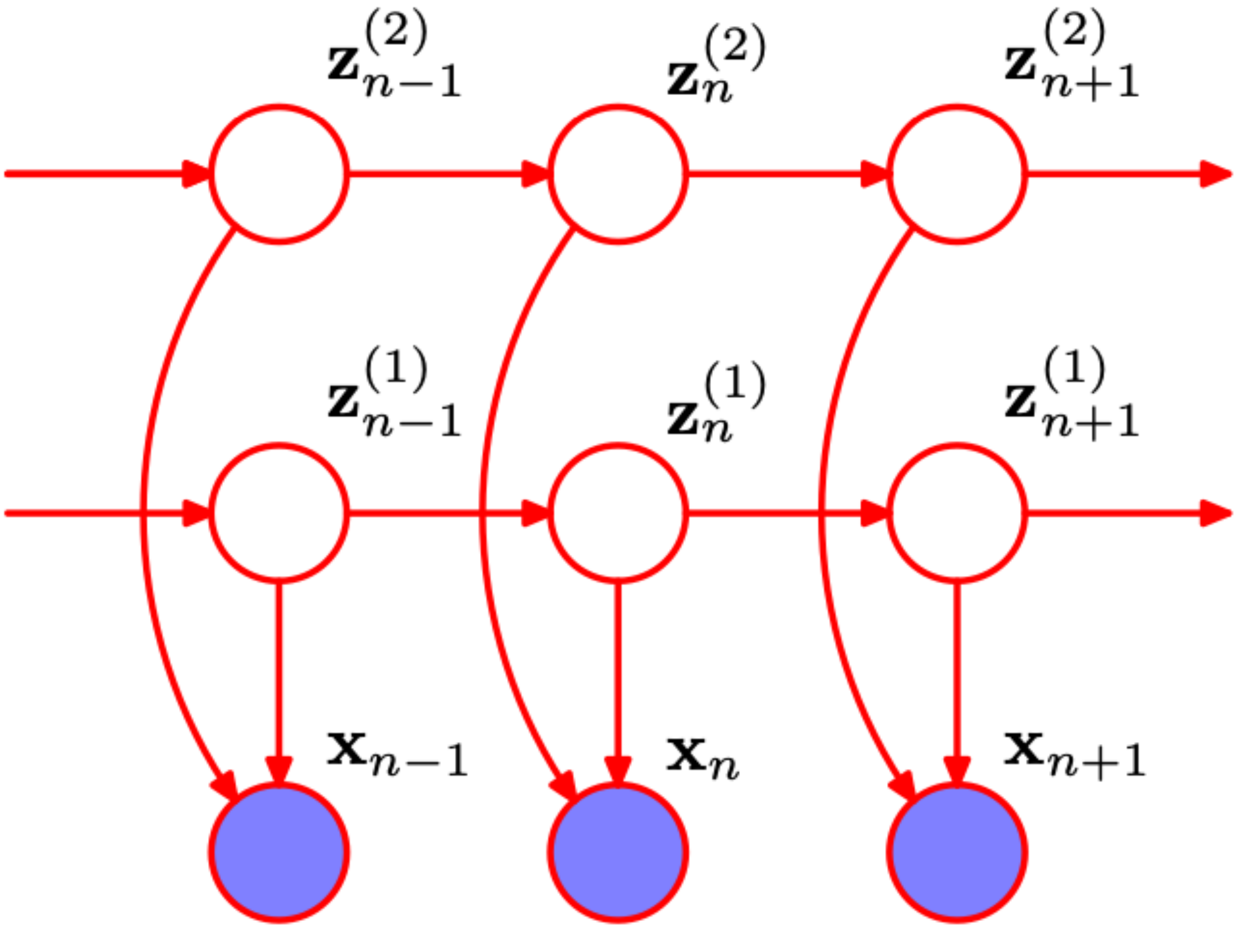
Observation also dependent on previous steps

Other Variations



Input-Output HMM/LDS

Other Variations



Factorial HMM with multiple chains

Summary

- Mixture Distribution: to build more complex distribution from simple ones
- Gaussian Mixture Model: k Gaussian components
- Expectation-Maximization: general for graphical models with latent variables
  - E-step: fix parameter, estimate posterior mean/variance
  - M-step: update parameter
- Probabilistic PCA: latent is continuous
- Linear Dynamical System:
  - E-step: Forward-backward alg.
  - M-step: update parameters

Recommended Reading

- PRML Chapter 9, 12.2, 13.3

Next up

- Undirected Graphical Models
- Approximate Inference
  - Variational Inference
  - Sampling