Lecture 11 Dynamic Bayesian Networks Linear Dynamical Systems

Lei Li and Yu-xiang Wang
UCSB

Recap

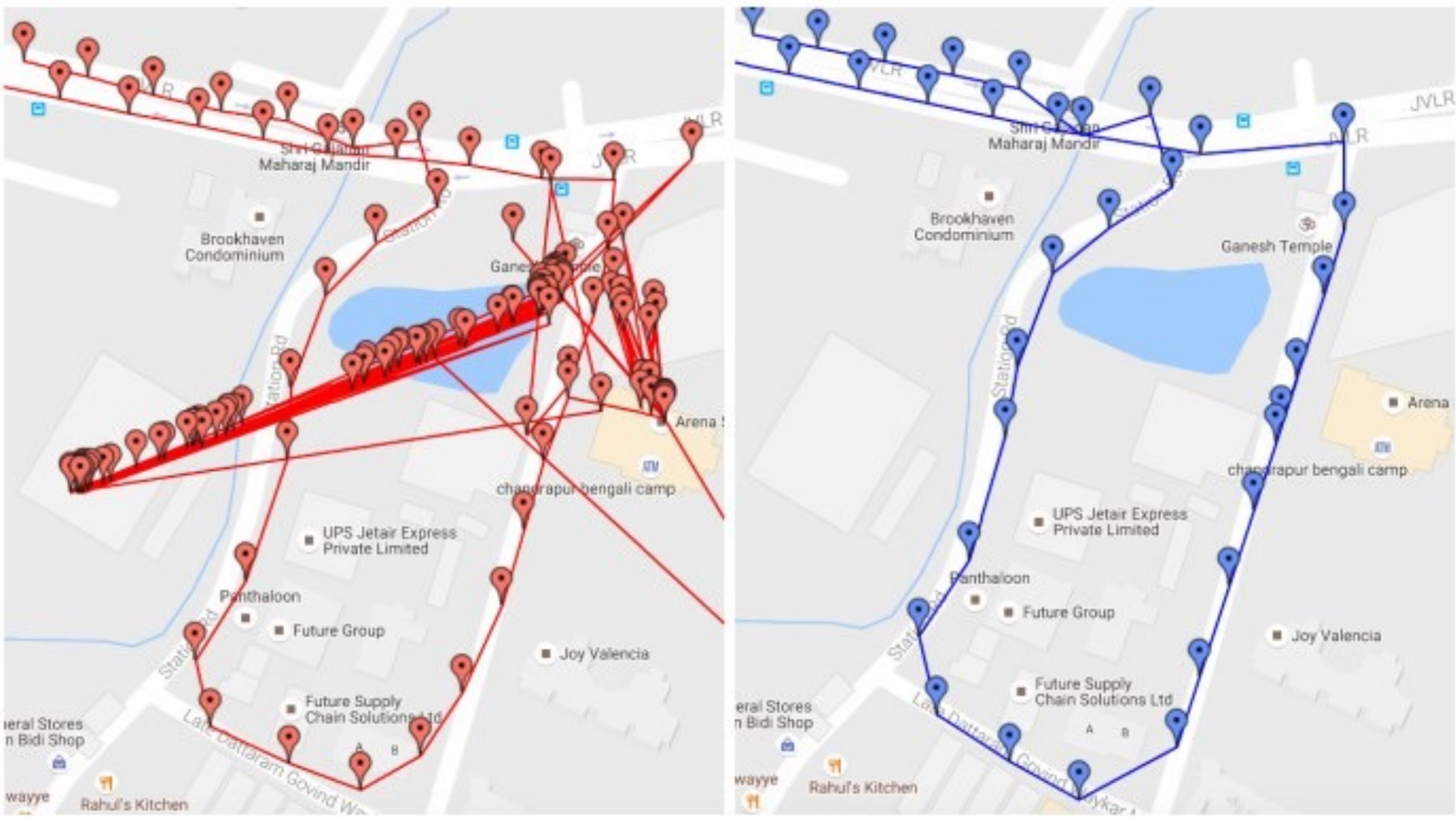
- Bayesian networks:
  - Directed acyclic graph
  - Nodes are random variables
  - arcs are probabilistic dependencies
- Mixture of Gaussian Model
- Expectation-Maximization

Dynamic Bayesian Networks

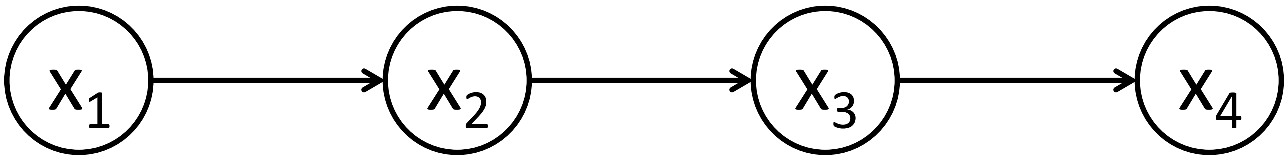
- What about non-IID data / sequential data
- Markov assumption

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

Estimating the true trajectory

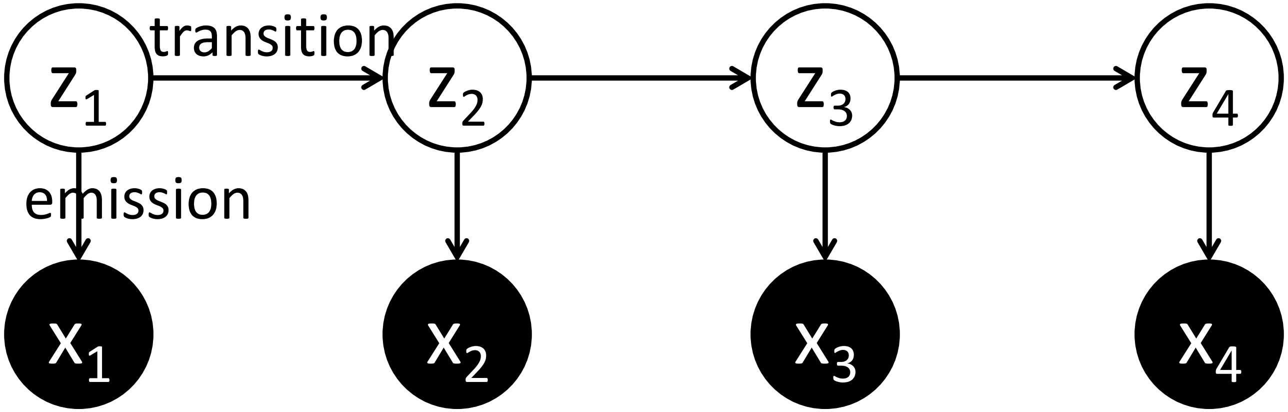


Markov Process



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

Linear Dynamical Systems



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

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

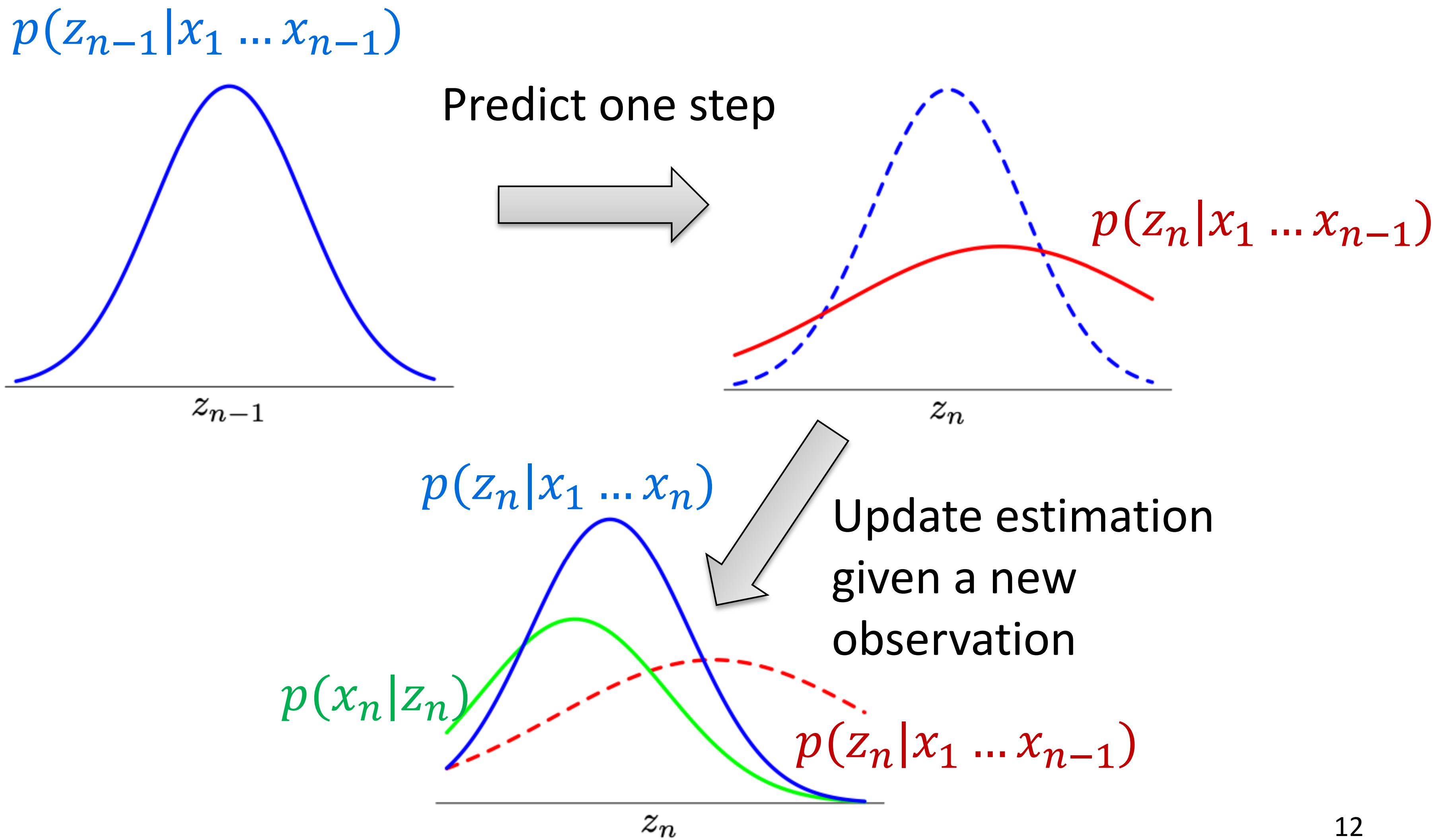
Maximization

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

What does Kalman filter (forward-pass) do?



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

EM for LDS

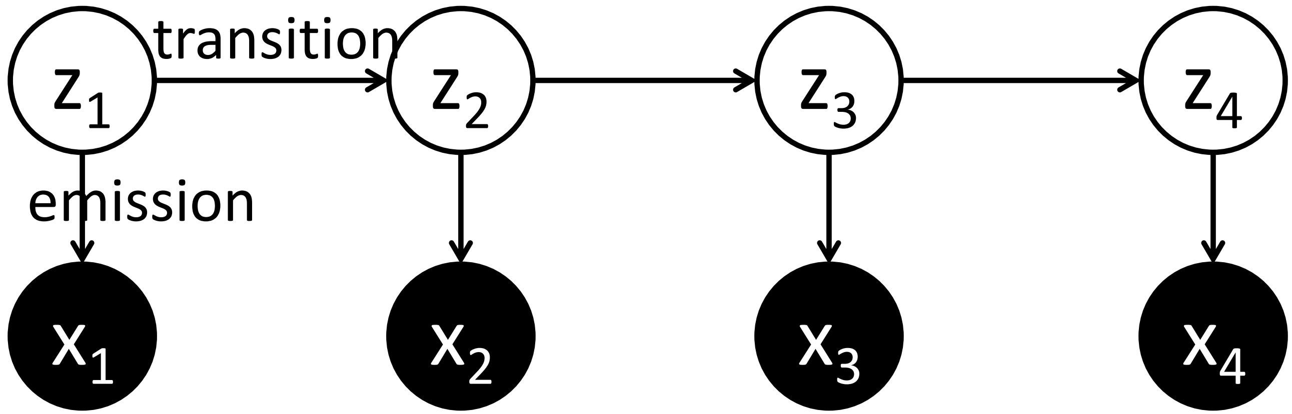
- Observation: $x_{1..N}$
- $\theta = \{\mu_0, Q_0, A, Q, C, R\}$
- Iterate until convergence
  1. E step: use X and current θ to calculate marginal posterior mean E[z|x] and co-variance Cov[z|x]
     - Using forward (Kalman filtering) and backward (Kalman smoothing)
  2. M step:

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

Application of LDS

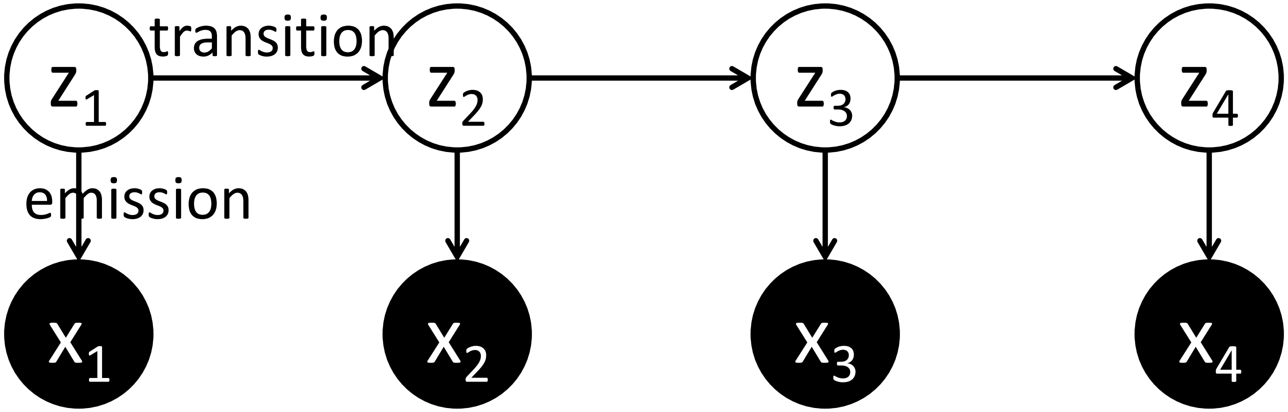
- Kalman filter: Tracking object movement
- Time series forecasting

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

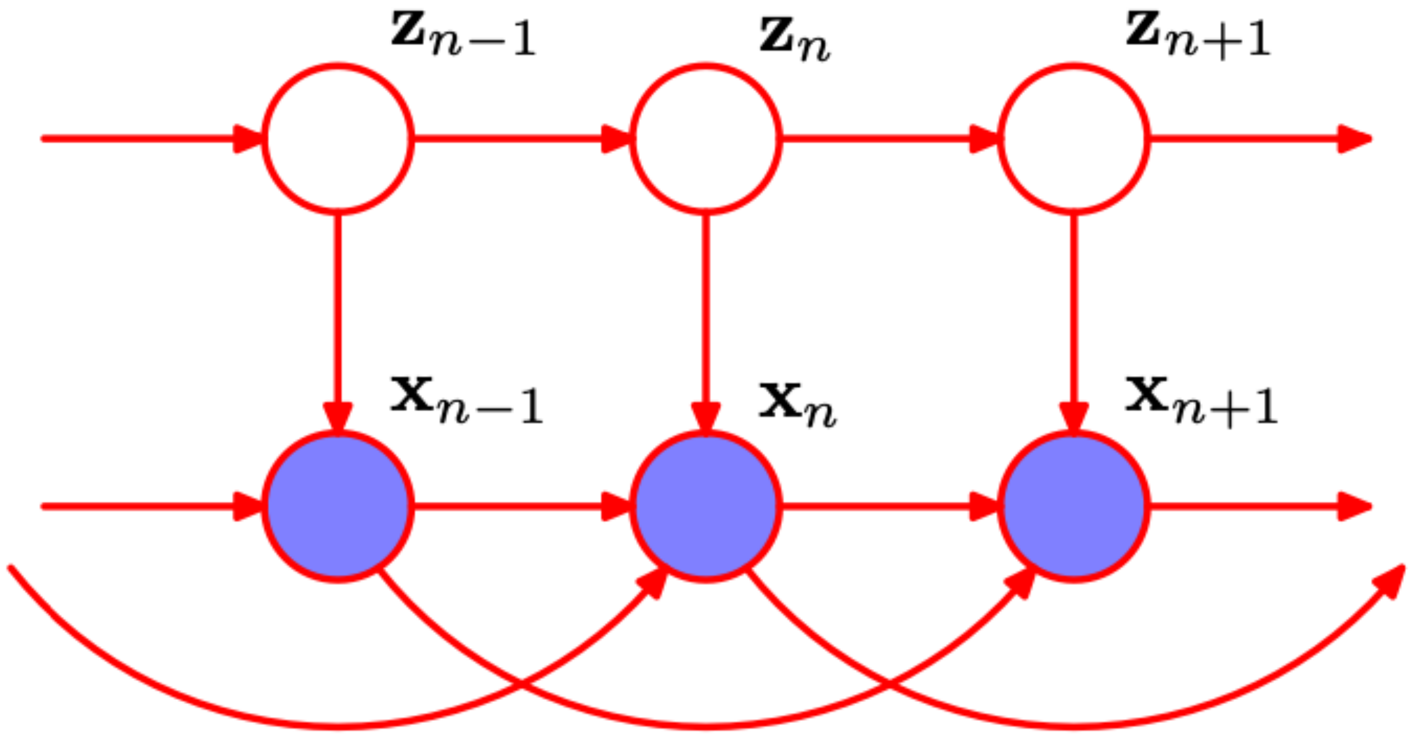
Hidden Markov Model



- 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.

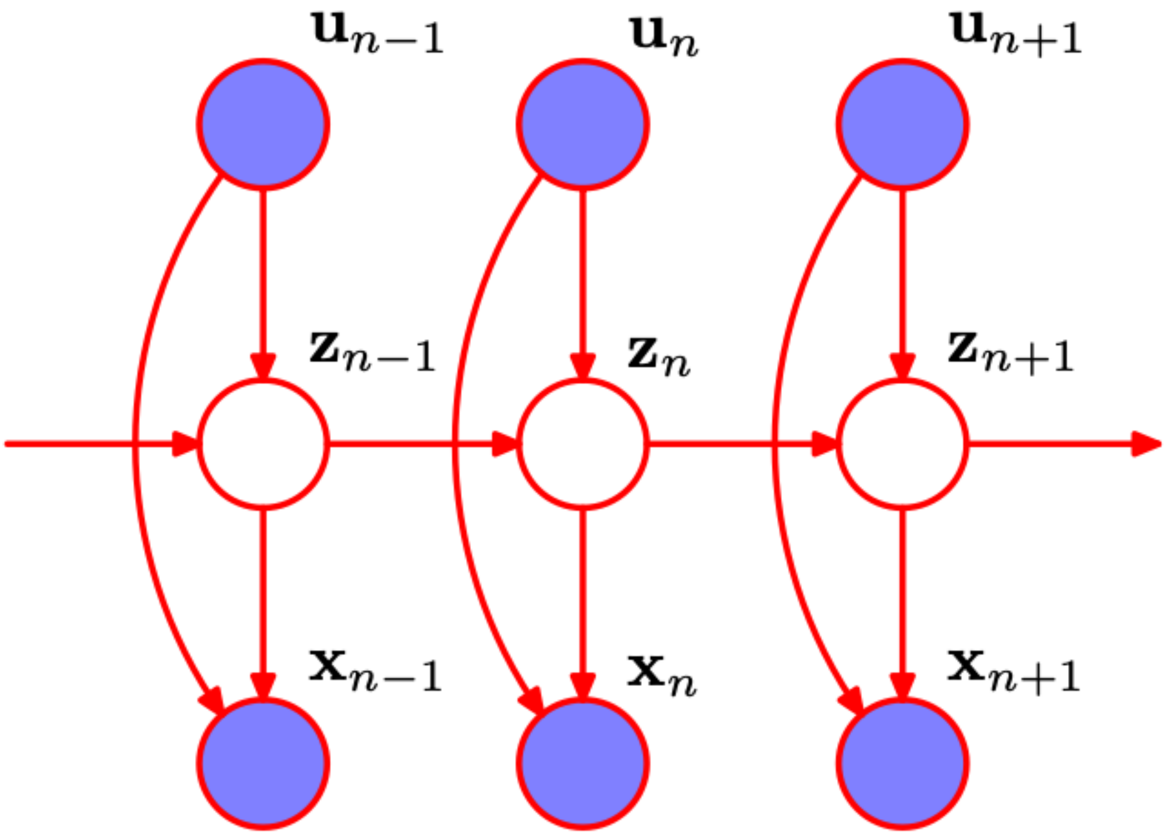
Andrew Viterbi

Other Variations



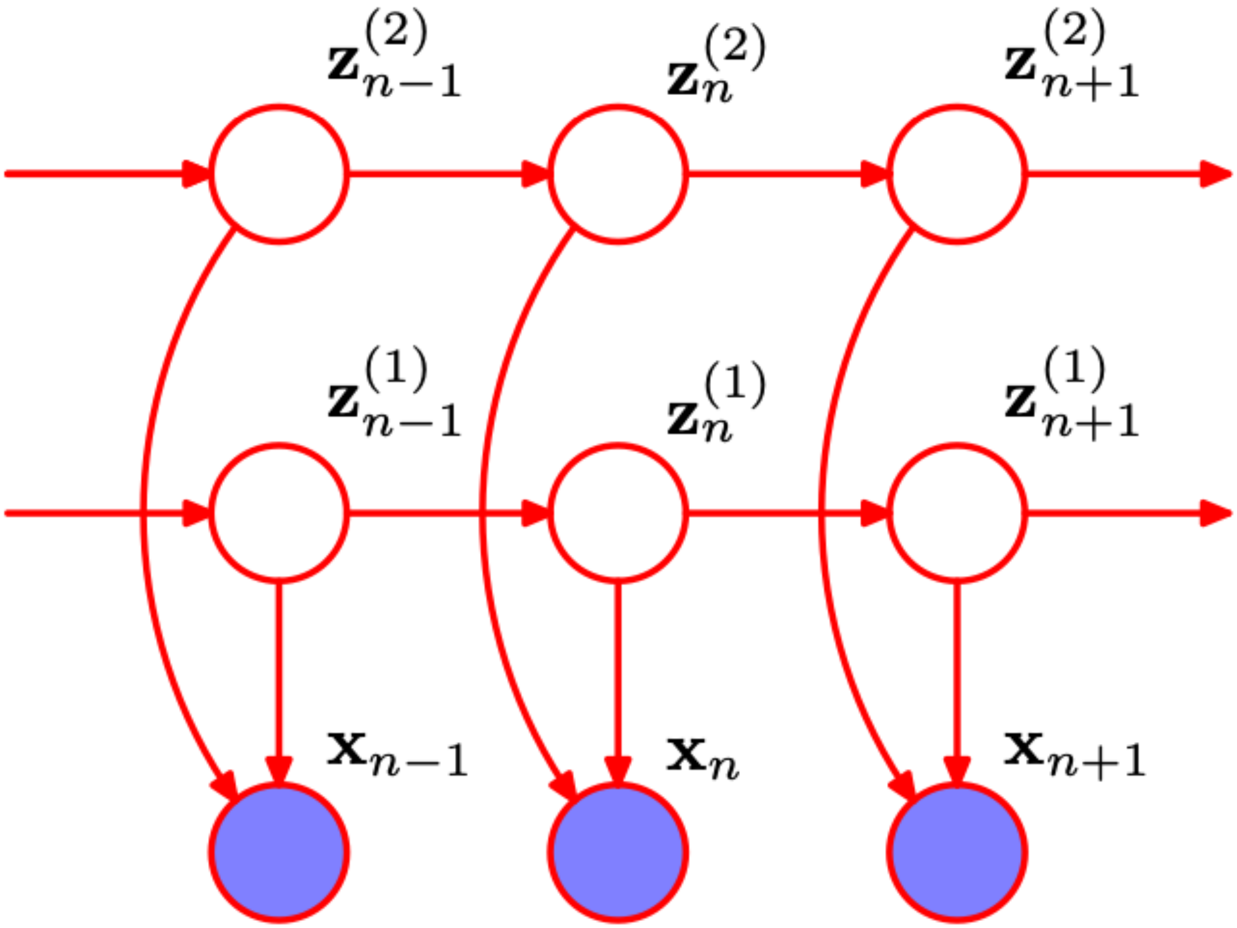
Observation also dependent on previous steps

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