# Linear Dynamical Systems

Sargur N. Srihari

srihari@cedar.buffalo.edu

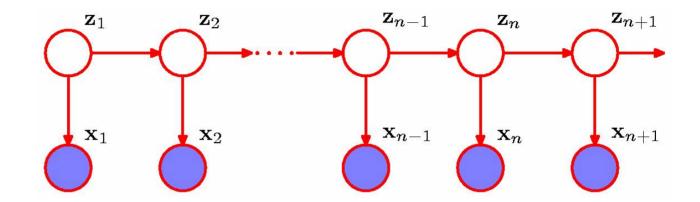
Machine Learning Course:

http://www.cedar.buffalo.edu/~srihari/CSE574/index.html

## Two Models Described by Same Graph

Latent variables

**Observations** 



- Hidden Markov Model: If latent variables are discrete:
   Observed variables in a HMM may be discrete or continuous
- Linear Dynamical Systems: If both latent and observed variables are Gaussian

### **Motivation**

- Simple problem in practical settings
  - Measure a value of unknown quantity z using a noisy sensor that returns a value x
  - x is z plus zero mean Gaussian noise
- Given single measurement, best guess for z is to assume that z = x
- We can improve estimate for z by taking lots of measurements and averaging them
  - Random noise terms tend to cancel each other

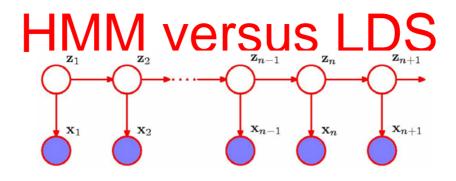
### **Time Varying Data**

- Take regular measurements of x
- We have obtained  $x_1, ..., x_N$
- We wish to find corresponding values  $z_1,..., z_N$
- Simply averaging them will reduce error due to random noise but we only get a single average estimate
- Can do better by averaging only last few measurements x<sub>N-L</sub>,...,x<sub>N</sub>
  - If z is changing slowly and random noise error is high then use a relatively long window to average
  - If z is varying rapidly and noise is small then use  $z_N = x_N$
  - Can do better by weighted averaging where more recent measurements make a greater contribution than less recent ones

#### Formalization of Intuitive Idea

- How to form weighted average?
- LDS are a systematic approach
  - Probabilistic model that captures time evolution and measurement processes
  - Apply inference and learning methods





# Corresponding Factorization

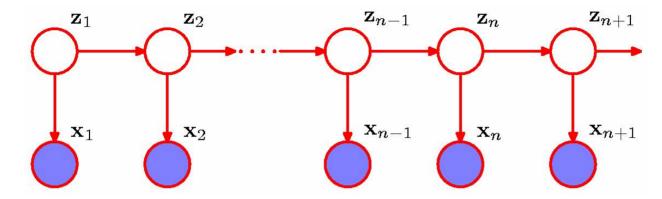
$$p(X, Z \mid \theta) = p(z_1 \mid \pi) \left[ \prod_{n=2}^{N} p(z_n \mid z_{n-1}, A) \right] \prod_{m=1}^{N} p(x_m \mid z_m, \phi)$$
where  $X = \{x_1, ..., x_N\}, Z = \{z_1, ..., z_N\}, \theta = \{\pi, A, \phi\}$ 

- LDS has continuous latent variables
  - In Sum-Product Algorithm summations become integrals
- General form of inference algorithms are same as for HMM
- Two models were developed independently

### **Need for Gaussian Distribution**

- Key requirement is to retain efficient algorithm which is linear in length of chain
- Posterior probability of  $z_n$  given  $x_1,...x_n$  is denoted  $\alpha(z_{n-1})$
- It is multiplied by the transition probability  $p(z_n|z_{n-1})$  and emission probability  $p(x_n|z_n)$
- Then marginalize over z<sub>n-1</sub> to obtain a distribution over z<sub>n</sub>
- So that distribution does not become more complex at each stage, we need the exponential form, e.g., Gaussian

#### **Linear Gaussian Units**



- Both latent variables and observed variables are multivariate Gaussian
- Their means are linear functions of the states of their parents in the graph
- Linear Gaussian units are equivalent to joint Gaussian distribution of the variables

#### Kalman Filter

- Since model is a tree-structured graph inference solved using sum-product
- Forward recursions, analogous to alpha messages in HMM, are known as Kalman filter equations
- Backward messages are known as Kalman smoother equations

### Kalman Filter Applications

- Used widely in real-time tracking applications
  - Used originally in the Apollo program navigation computer
  - Also in computer vision

### Joint Distribution over Latent and Observed variables

$$p(X, Z \mid \theta) = p(z_1 \mid \pi) \left[ \prod_{n=2}^{N} p(z_n \mid z_{n-1}, A) \right] \prod_{m=1}^{N} p(x_m \mid z_m, \phi)$$
where  $X = \{x_1, ..., x_N\}, Z = \{z_1, ..., z_N\}, \theta = \{\pi, A, \phi\}$ 

## Linear Dynamical System

- It is a linear-Gaussian model
- Joint distribution over all variables, as well as marginals and conditionals, is Gaussian
- Therefore sequence of individually most probable latent variable values is same as most probable latent sequence
  - Thus there is no need to consider analog of Viterbi algorithm

#### LDS Formulation

 Transition and Emission probabilities are written in the general form:

$$p(\mathbf{z}_n \mid \mathbf{z}_{n-1}) = N(\mathbf{z}_n \mid A\mathbf{z}_{n-1}, \Gamma)$$
$$p(\mathbf{x}_n \mid \mathbf{z}_{n-1}) = N(\mathbf{x}_n \mid C\mathbf{z}_n, \Sigma)$$

Initial latent variable has the form:

$$p(\mathbf{x}_n | \mathbf{z}_n) = N(\mathbf{z}_1 | \mu_0, V_0)$$

• Parameters of the model  $\theta = \{A, \Gamma, C, \Sigma, \mu_0, V_0\}$  are determined using maximum likelihood through EM algorithm

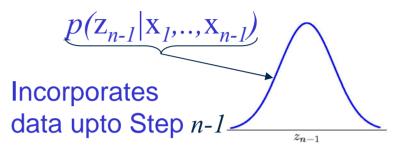
### Inference in LDS

- Finding marginal distributions for latent variables conditioned on observation sequence
- For given parameter settings wish to make predictions of next latent state  $z_n$  and next observation  $x_n$  conditioned on observed data  $x_1,...x_{n-1}$
- Done efficiently using sum-product algorithm which are Kalman filter and smoother equations
- Joint distribution over all latent and observed variables is simply Gaussian
  - Inference problem solved using marginals and conditionals of multivariate Gaussian

## **Uncertainty Reduction**

Kalman filter is a process of making successive predictions and then correcting predictions in light of new observations

Before Observing  $x_n$ 



 $p(\mathbf{z}_n|\mathbf{x}_1,..,\mathbf{x}_{n-1})$  Diffusion arising from non-zero variance of transition probability  $p(\mathbf{z}_n|\mathbf{z}_{n-1})$ 

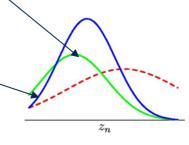
rom e of lity  $p(\mathbf{z}_n|\mathbf{z}_{n-1})$ 

After Observing  $x_n$ 

$$p(\mathbf{z}_n|\mathbf{x}_1,...,\mathbf{x}_n)$$

Not a density wrt  $p(\mathbf{X}_n | \mathbf{Z}_n) \mathbf{z}_n$  and not normalized to one

Inclusion of new data leads to revised distribution of state density—shifted and narrowed compared to  $p(\mathbf{z}_n|\mathbf{x}_1,...,\mathbf{x}_{n-1})$ 



### Illustration: LDS to track moving object

#### Blue points:

True position of object in 2-D space at successive time steps

#### Green points:

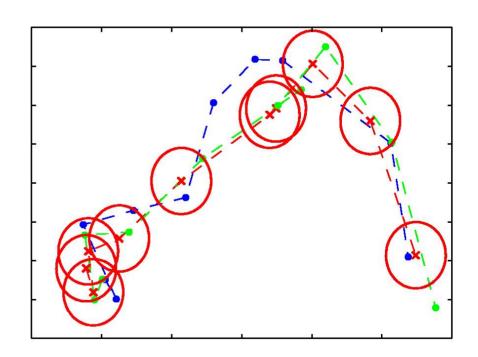
Noisy measurements of the positions

#### Red crosses:

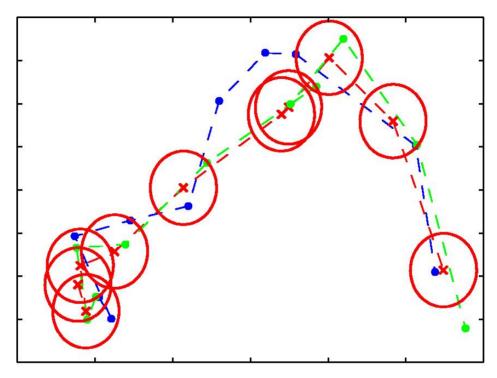
Means of inferred posterior distributions obtained by running Kalman filtering equations

#### Red ellipses:

Covariances of inferred positions— correspond to contours of one std deviation



### **Estimated Path**



 Red dashed line is closer to blue dashed line than green dashed line

### Learning in LDS

- Determination of parameters  $\theta = \{A, \Gamma, C, \Sigma, \mu_0, V_0\}$  using EM algorithm
- Denote the estimated parameter values at some particular cycle of the algorithm as  $\theta^{old}$
- For these parameter values we run the inference algorithm to determine the posterior distribution of the latent variables  $p(\mathbf{Z}|\mathbf{X}, \theta^{old})$
- In the M step, function  $Q(\theta, \theta^{old})$  is maximized with respect to the components of  $\theta$

## Further Topics in LDS

- Extensions of LDS
  - Using mixture of K Gaussians
  - Expanding graphical representation similar to HMMs
    - Switching state-space model combines HMM and LDS
- Particle Filters
  - If a non-Gaussian emission density is used, use sampling methods to find a tractable inference algorithm

## Other Topics on Sequential Data

Sequential Data and Markov Models:
 <a href="http://www.cedar.buffalo.edu/~srihari/CSE574/Chap11/Ch11.1-MarkovModels.pdf">http://www.cedar.buffalo.edu/~srihari/CSE574/Chap11/Ch11.1-MarkovModels.pdf</a>

Hidden Markov Models:

http://www.cedar.buffalo.edu/~srihari/CSE574/Chap11/Ch11.2-HiddenMarkovModels.pdf

Extensions of HMMs:

http://www.cedar.buffalo.edu/~srihari/CSE574/Chap11/Ch11.3-HMMExtensions.pdf

Conditional Random Fields:

http://www.cedar.buffalo.edu/~srihari/CSE574/Chap11/Ch11.5-ConditionalRandomFields.pdf