# CS 3710: Hidden Markov Models (HMMs)

Presented by Paul Hoffmann

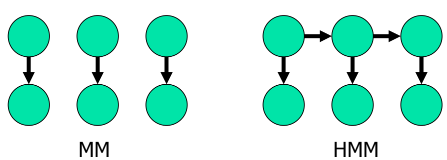
Nov 30th, 2005

## Overview (Chapter 12)

- HMMs vs MMs (Mixture Models)
- Full joint, parameterization of HMM
- Inference problems
- Inference algorithms
- Parameter estimation (EM)

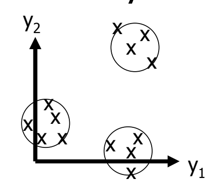
# HMMs vs. MMs

- HMMs
  - generalization of MMs
  - "dynamic" MMs
  - mixture components called states



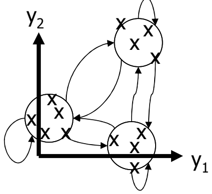

# MMs: Generation Example

- Let $y_t = (y^1_t, y^2_t)$
- 1) Randomly select mixture component according to $P(q_t)$
- 2) Randomly select y according to $P(y_t|q_t)$



# HMMs: Generation Example

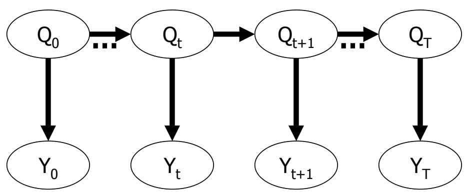
- Let $y_t = (y^1{}_t, y^2{}_t)$
- 1) Randomly select state according to $P(q_{t+1}|q_t)$
- 2) Randomly select y according to $P(y_t|q_t)$




# Some Notation

| Variable | Meaning |
|---|---|
| $t$ | time |
| $q_t$ | state at time t (from multinomial distribution |
| $q^i{}_t$ | $i^{th}$ component of state $q_t$ (1 or 0) |
| $y_t$ | observable output |
| A | state transition matrix |
| i, j | index component of state integer from 1 to M |
| M | number of components of state |
| T | number of states + 1 |

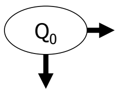
# Graphical Model




Chain Rule for BBNs:

$$p(\vec{q},\vec{y}) = p(q_0)\prod_{t=0}^{T-1} p(q_{t+1} \mid q_t)\prod_{t=0}^{T} p(y_t \mid q_t)$$

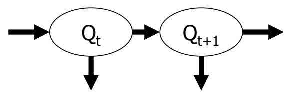
# CPT for $Q_0$




| | 1 | ... | i | ... | M |
|---|---|---|---|---|---|
| $Q_0$ | | | $\pi_i = P(Q_0 = i)$ | | |

$$\pi_t^i = 1 \text{ if } \pi_t = \text{i}, 0 \text{ otherwise} \quad \pi_{q_0} \equiv \prod_{i=1}^{M} [\pi_i]^{q_0^i}$$

# CPT for $Q_{t+1}$

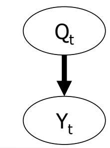



| $Q_t$ \ $Q_{t+1}$ | 1 | … | j | … | M |
|---|---|---|---|---|---|
| 1 | | | | | |
| … | | | | | |
| i | | | $a_{i,j}=P(Q_{t+1}=j \mid Q_t=i)$ | | |
| … | | | | | |
| M | | | | | |

$q_t^i = 1$ if $q_t = i$, 0 otherwise

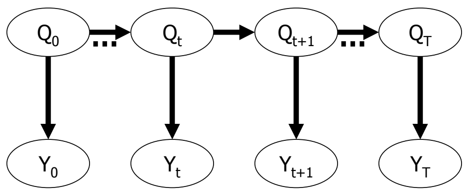
$$a_{q_t, q_{t+1}} \equiv \prod_{i=1}^{M}\prod_{j=1}^{M}[a_{ij}]^{q_t^i q_{t+1}^j}$$

# CPT for $Y_t$




- ?

# Graphical Model




Chain Rule for BBNs:

$$p(\vec{q},\vec{y}) = p(q_0)\prod_{t=0}^{T-1} p(q_{t+1} \mid q_t)\prod_{t=0}^{T} p(y_t \mid q_t)$$

$$p(\vec{q},\vec{y}) = \pi_{q_0}\prod_{t=0}^{T-1} a_{q_t,q_{t+1}}\prod_{t=0}^{T} p(y_t \mid q_t)$$

# Inference

Filtering: $P(q_z \mid y_0,\ldots,y_w), z = w$

Prediction: $P(q_z \mid y_0,\ldots,y_w), z > w$

Smoothing: $P(q_z \mid y_0,\ldots,y_w), z < w$

$P(\vec{q} \mid \vec{y})$

$P(q_t \mid \vec{y})$

## Computing the Posterior

$$P(\vec{q} \mid \vec{y}) = \frac{P(\vec{q}, \vec{y})}{P(\vec{y})}$$

$$P(\vec{q}, \vec{y}) = \pi_{q_0} \prod_{t=0}^{T-1} a_{q_t, q_{t+1}} \prod_{t=0}^{T} P(y_t \mid q_t)$$

$$P(\vec{y}) = \underbrace{\sum_{q_0} \sum_{q_1} \cdots \sum_{q_T}}_{\text{T+1 summations, each summing M variables}} \prod_{t=0}^{T-1} a_{q_t, q_{t+1}} \prod_{t=0}^{T} P(y_t \mid q_t, \eta)$$

T+1 summations, each summing M variables

## Computing $P(q_t|y)$

- Goal: compute $P(q_t|y)$
  - Problem: conditioning on y for HMMs doesn't result in conditional independences
  - Solution: use Bayes Rule so can condition on $q_t$

$$P(q_t \mid \vec{y}) = \frac{P(\vec{y} \mid q_t) P(q_t)}{P(\vec{y})}$$

## Computing $P(q_t|y)$

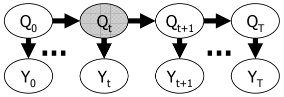
- Goal: compute $P(q_t, y)$ and $P(y)$
  - Step 1: obtain $P(y)$ from $P(q_t, y)$

$$P(\vec{y}) = \sum_{q_t} P(q_t, \vec{y})$$

  - Step 2: obtain $P(q_t, y)$ by using conditional independences and recursion
    - recursion allows us to use dynamic programming

## α-β Recursion

# Computing $P(q_t, y)$




- Goal: use conditional independences to split up $P(q_t, y)$
  - Solution: use $q_t$ to split up

$$P(\vec{y} \mid q_t)P(q_t) = P(y_0, \ldots, y_t \mid q_t)P(y_{t+1}, \ldots, y_T \mid q_t)P(q_t)$$

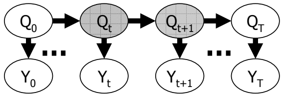
  - Solution: regroup and combine terms, use recursion on each result

$$P(\vec{y} \mid q_t)P(q_t) = \alpha(q_t)\beta(q_t) \text{ where}$$

$$\alpha(q_t) = P(y_0, \ldots, y_t, q_t)$$

$$\beta(q_t) = P(y_{t+1}, \ldots, y_T \mid q_t)$$

# α Recursion




- Goal: define α as a recursive function

$\alpha(q_{t+1}) = P(y_0, \ldots, y_{t+1}, q_{t+1})$

$= P(y_0, \ldots, y_{t+1} \mid q_{t+1})P(q_{t+1})$ #use chain rule so can condition on $q_{t+1}$

$= P(y_0, \ldots, y_t \mid q_{t+1})P(y_{t+1} \mid q_{t+1})P(q_{t+1})$ #use CIs from HMM graph

$= P(y_0, \ldots, y_t, q_{t+1})P(y_{t+1} \mid q_{t+1})$ #regroup & combine

$= \sum_{q_t} P(y_0, \ldots, y_t, q_t, q_{t+1})P(y_{t+1} \mid q_{t+1})$ #introduce $q_t$ so can have $\alpha(q_t)$ term

$= \sum_{q_t} P(y_0, \ldots, y_t \mid q_{t+1}, q_t)P(q_{t+1} \mid q_t)P(q_t)P(y_{t+1} \mid q_{t+1})$ #use chain rule

$= \sum_{q_t} P(y_0, \ldots, y_t \mid q_t)P(q_{t+1} \mid q_t)P(q_t)P(y_{t+1} \mid q_{t+1})$ #use CIs from HMM graph

$= \sum_{q_t} P(y_0, \ldots, y_t, q_t)P(q_{t+1} \mid q_t)P(y_{t+1} \mid q_{t+1})$ #regroup & combine

$= \sum_{q_t} \alpha(q_t)a_{q_t, q_{t+1}}P(y_{t+1} \mid q_{t+1})$ #definitions of α, a

# α Recursion

Base Case:

$$\alpha(q_0) = P(y_0, q_0) = P(y_0 \mid q_0)P(q_0) = P(y_0 \mid q_0)\pi_{q_0}$$

Computational Complexity of Each Step: O($M^2$)

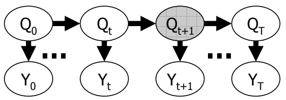
$$\alpha(q_{t+1}) = \sum_{q_t} \alpha(q_t) a_{q_t, q_{t+1}} P(y_{t+1} \mid q_{t+1})$$

$q_{t+1}$ takes on M different values

2M multiplications

Need to do these computations for t = 1 to t = T.
Computational complexity O($M^2T$)

# β Recursion




- Goal: define β as a recursive function

$$\beta(q_t) = P(y_{t+1}, \ldots, y_T \mid q_t)$$

$$= \sum_{q_{t+1}} P(y_{t+1}, \ldots, y_T, q_{t+1} \mid q_t)$$ #introduce $q_{t+1}$ so can have $\beta(q_{t+1})$ term

$$= \sum_{q_{t+1}} P(y_{t+1}, \ldots, y_T \mid q_{t+1}, q_t) P(q_{t+1} \mid q_t)$$ #use chain rule

$$= \sum_{q_{t+1}} P(y_{t+2}, \ldots, y_T \mid q_{t+1}) P(y_{t+1} \mid q_{t+1}) P(q_{t+1} \mid q_t)$$ #use CIs from HMM graph

$$= \sum_{q_{t+1}} \beta(q_{t+1}) a_{q_t, q_{t+1}} P(y_{t+1} \mid q_{t+1})$$ #definitions of β, a

## β Recursion: Base Case

$$\beta(q_{T-1}) = P(y_T \mid q_{T-1})$$

$$= \sum_{q_T} P(y_T, q_T \mid q_{T-1})$$ #introduce $q_T$

$$= \sum_{q_T} P(y_T \mid q_T, q_{T-1}) P(q_T \mid q_{T-1})$$ #use chain rule

$$= \sum_{q_T} P(y_T \mid q_T) P(q_T \mid q_{T-1})$$ #use CIs from HMM graph

$$= \sum_{q_T} 1 * a_{q_{T-1}, q_T} P(y_T \mid q_T)$$ #definitions of a

$$\beta(q_t) = \sum_{q_{t+1}} \beta(q_{t+1}) a_{q_t, q_{t+1}} P(y_{t+1} \mid q_{t+1})$$

$\beta(q_{T-1})$ has same form as other $\beta(q_t)$'s if $\beta(T)$ is set to 1

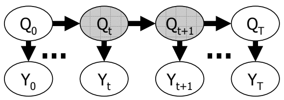
## α-γ Recursion

## α-γ Recursion

- Do alpha recursion, then gamma recursion
- Gamma function definition:

$$\gamma(q_t) \equiv P(q_t \mid \vec{y}) = \frac{\alpha(q_t)\beta(q_t)}{P(\vec{y})}$$

## γ Recursion




- Goal: define γ as a recursive function

$$\gamma(q_t) = P(q_t \mid y_0,\ldots,y_T)$$

$$= \sum_{q_{t+1}} P(q_t, q_{t+1} \mid y_0,\ldots,y_T)$$ #introduce $q_{t+1}$ so can have $\gamma(q_{t+1})$ term

$$= \sum_{q_{t+1}} P(q_t \mid q_{t+1}, y_0,\ldots,y_T) P(q_{t+1} \mid y_0,\ldots,y_T)$$ #use chain rule to get $\gamma(q_{t+1})$ term

$$= \sum_{q_{t+1}} P(q_t \mid q_{t+1}, y_0,\ldots,y_t) P(q_{t+1} \mid y_0,\ldots,y_T)$$ #use CIs from HMM graph

$$= \sum_{q_{t+1}} \frac{P(q_t, q_{t+1}, y_0,\ldots,y_t)}{\sum_{q_t} P(q_t, q_{t+1}, y_0,\ldots,y_t)} P(q_{t+1} \mid y_0,\ldots,y_T)$$ #use definition for P(X|Y)

$$= \sum_{q_{t+1}} \frac{P(q_t, y_0,\ldots,y_t) P(q_{t+1} \mid q_t)}{\sum_{q_t} (q_t, y_0,\ldots,y_t) P(q_{t+1} \mid q_t)} P(q_{t+1} \mid y_0,\ldots,y_T)$$ #split using CIs from HMM graph

$$= \sum_{q_{t+1}} \frac{\alpha(q_t) a_{q_t,q_{t+1}}}{\sum_{q_t} \alpha(q_t) a_{q_t,q_{t+1}}} \gamma(q_{t+1})$$ #definitions of α, a, γ

## γ Recursion: Base Case

$$\gamma(q_t) = \sum_{q_{t+1}} \frac{\alpha(q_t) a_{q_t,q_{t+1}}}{\sum_{q_t} \alpha(q_t) a_{q_t,q_{t+1}}} \gamma(q_{t+1})$$

$\gamma(q_{T-1}) = P(q_{T-1} \mid y_0,\dots,y_T)$

$= \sum_{q_T} P(q_{T-1}, q_T \mid y_0,\dots,y_T)$ #introduce $q_T$

$= \sum_{q_T} P(q_{T-1} \mid q_T, y_0,\dots,y_T) P(q_T \mid y_0,\dots,y_T)$ #use chain rule

$= \sum_{q_T} P(q_{T-1} \mid q_T, y_0,\dots,y_{T-1}) P(q_T \mid y_0,\dots,y_T)$ #use CIs from HMM graph

$= \sum_{q_T} \frac{P(q_{T-1}, q_T, y_0,\dots,y_{T-1})}{\sum_{q_{T-1}} (q_{T-1}, q_T, y_0,\dots,y_{T-1})} P(q_T \mid y_0,\dots,y_T)$ #use definition for P(X|Y)

$= \sum_{q_{t+1}} \frac{P(q_{T-1}, y_0,\dots,y_{T-1}) P(q_T \mid q_{T-1})}{\sum_{q_{T-1}} P(q_{T-1}, y_0,\dots,y_{T-1}) P(q_T \mid q_{T-1})} P(q_T \mid y_0,\dots,y_T)$ #split using CIs from HMM graph

$= \sum_{q_{t+1}} \frac{\alpha(q_{T-1}) a_{q_{T-1},q_T}}{\sum_{q_t} \alpha(q_{T-1}) a_{q_{T-1},q_T}} \alpha(q_T)$ #definitions of α, a

base case

## EM

- Initialize parameters **θ**
- do
  - Set **θ**' = **θ**
  - 1) Expectation
    - Complete all hidden and missing values with expectations given current set of parameters **θ**'
  - 2) Maximization
    - Use completed data to compute new estimates for **θ**
- while improvement possible

# EM

- For this example, it is assumed that the outputs $y_t$ are multinomial

$$P(y_t \mid q_t, \eta) = \prod_{i=1}^{M} \prod_{j=1}^{N} [\eta]^{q_t^i y_t^j}$$

# EM

- log likelihood

$$\log p(q, \vec{y}) = \log\{\pi_{q_0} \prod_{t=0}^{T-1} a_{q_t, q_{t+1}} \prod_{t=0}^{T} p(y_t \mid q_t, \eta)\}$$

$$= \log\{\prod_{i=1}^{M} [\pi_i]^{q_0^i} \prod_{t=0}^{T-1} \prod_{i=1}^{M} \prod_{j=1}^{M} [a_{i,j}]^{q_t^i q_{t+1}^j} \prod_{t=0}^{T} \prod_{i=1}^{M} \prod_{j=1}^{N} [\eta_{i,j}]^{q_t^i y_t^j}\}$$

$$= \sum_{i=1}^{M} q_0^i \log \pi_i + \sum_{t=0}^{T-1} \sum_{i=1}^{M} \sum_{j=1}^{M} q_t^i q_{t+1}^j \log a_{i,j} + \sum_{t=0}^{T} \sum_{i=1}^{M} \sum_{j=1}^{N} q_t^i y_t^j \log \eta_{i,j}$$

## EM

| Sufficient statistics | Maximum likelihood estimates |
|---|---|
| $\pi_i : q_0^i$ | $\hat{\pi}_i = q_0^i$ |
| $a_{i,j} : m_{i,j} \equiv \sum_{t=0}^{T-1} q_t^i q_{t+1}^j$ | $\hat{a}_{i,j} = \frac{m_{i,j}}{\sum_{k=1}^{M} m_{i,k}}$ |
| $\eta_{i,j} : n_{i,j} \equiv \sum_{t=0}^{T} q_t^i y_t^j$ | $\hat{\eta}_{i,j} = \frac{n_{i,j}}{\sum_{k=1}^{N} n_{i,k}}$ |

## Expectation

$$E(n_{i,j} \mid y, \theta^{(p)}) \equiv \sum_{t=0}^{T} \gamma_t^i y_t^j$$

$$E(m_{i,j} \mid y, \theta^{(p)}) \equiv \sum_{t=0}^{T-1} \xi_{t,t+1}^{i,j}$$

## Maximization (Baum-Welch updates)

$$\hat{\eta}_{i,j}^{(p+1)} = \frac{\sum_{t=0}^{T} \gamma_t^i y_t^j}{\sum_{k=1}^{N}\sum_{t=0}^{T} \gamma_t^i y_t^k} = \frac{\sum_{t=0}^{T} \gamma_t^i y_t^j}{\sum_{t=0}^{T} \gamma_t^i}$$

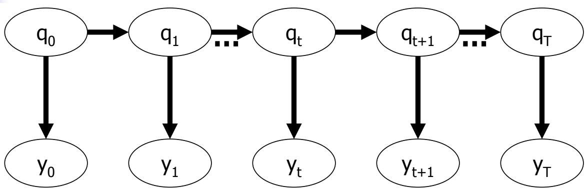
$$\hat{a}_{i,j}^{(p+1)} = \frac{\sum_{t=0}^{T-1} \xi_{t,t+1}^{i,j}}{\sum_{k=1}^{M}\sum_{t=0}^{T-1} \xi_{t,t+1}^{i,k}} = \frac{\sum_{t=0}^{T-1} \xi_{t,t+1}^{i,j}}{\sum_{k=1}^{M}\sum_{t=0}^{T-1} \gamma_t^i}$$

$$\hat{\pi}_i^{(p+1)} = \gamma_0^i$$

## Overview (Chapter 18)

- HMMs and Junction tree algorithm
- Linear Gaussian Models

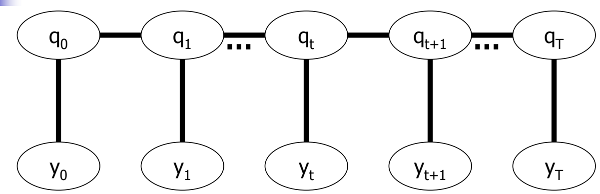
# HMM




Assume, for illustration purposes, $y_t$ is a multinomial node

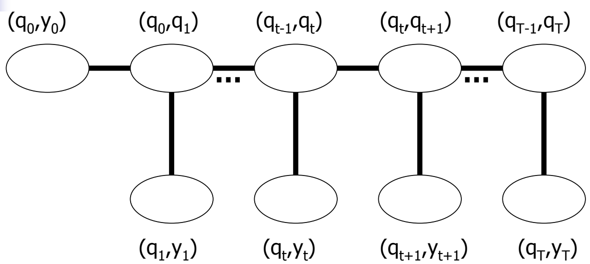
$$b_{i,j} \equiv P(y_t^i = 1 \mid q_t^i = 1)$$

# Creating Junction Tree



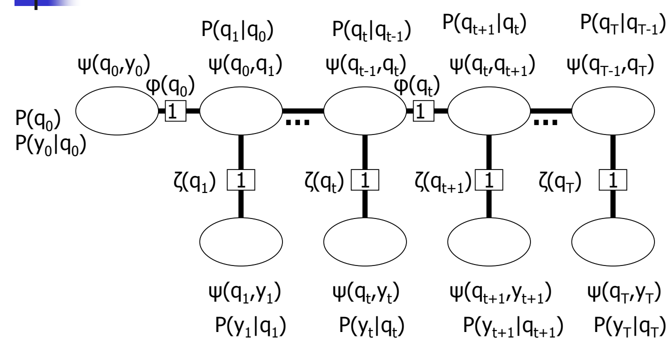

1) Moralize graph (nodes have at most 1 parent, no parents to join)
2) Triangulate graph (no cycles in graph)

# Creating Junction Tree



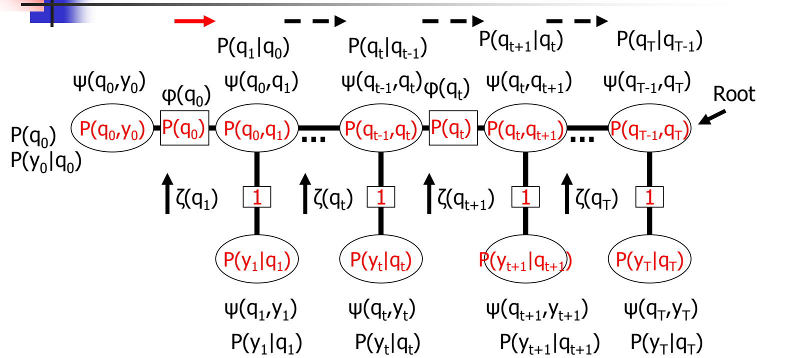

3) Create maximal spanning tree

# Creating Junction Tree



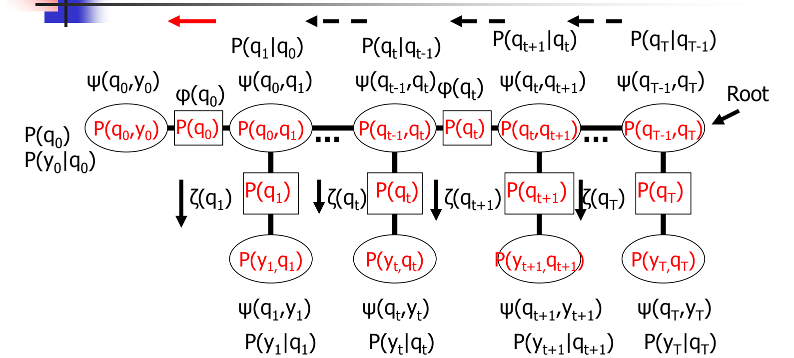

4) Make the separator set explicit
5) Assign local CPTs to potentials
6) Initialize separator potentials to 1

# Junction Tree, No Evidence

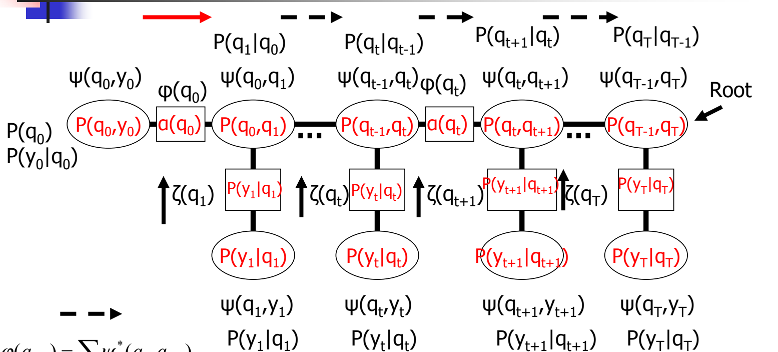



$$\sum_{y_t} P(y_t \mid q_t) = 1 \qquad \sum_{q_t} P(q_{t+1} \mid q_t) P(q_t) = P(q_{t+1}) \qquad \sum_{y_0} P(y_0 \mid q_0) P(q_0) = P(q_0)$$

# Junction Tree, No Evidence



# Junction Tree, With Evidence

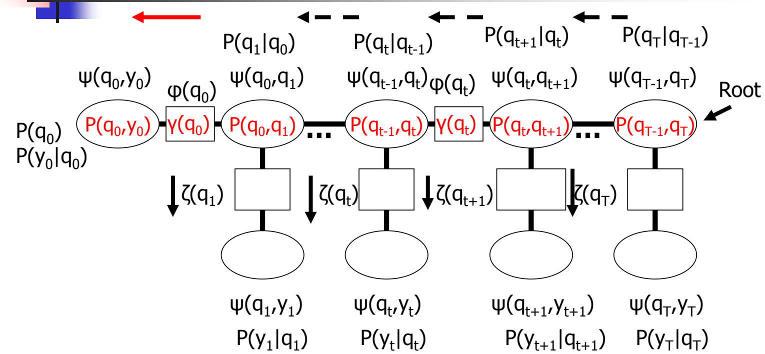



$$\varphi(q_{t+1}) = \sum_{q_t} \psi^*(q_t, q_{t+1})$$
$$= \sum_{q_t} a_{q_t, q_{t+1}} \varphi^*(q_t) P(y_{t+1} \mid q_{t+1})$$

$$\zeta(q_i) = P(y_i \mid q_i)$$

$$P(y_0 \mid q_0) P(q_0) = \alpha(q_0)$$

# Junction Tree, With Evidence



## Other algorithms

- Derived α-γ from junction tree, can also derive α-β and ρ-ζ algorithms

## Linear Gaussian Models (LG-HMMs)

- Same graphical structure as HMM
- Different node type and parameterization than HMMs
  - Nodes are linear-Gaussians
- Junction tree is the same except use linear-Gaussian potentials

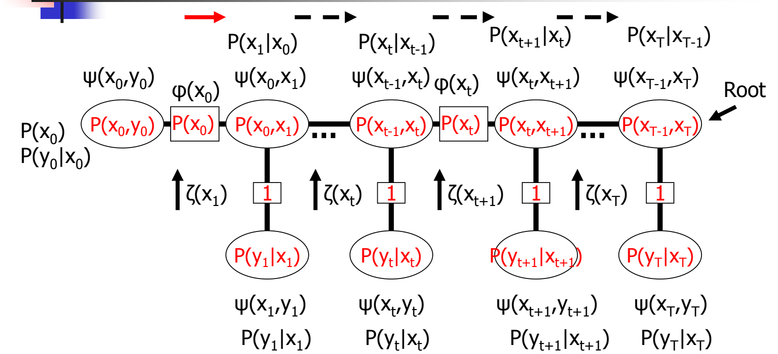
# LG-HMMs

- Gaussian CPDs (using moments)

$$P(x_{t+1} \mid x_t) \sim N(Ax_t, GQG^T)$$

$$P(y_t \mid x_t) \sim N(Cx_t, R)$$

# Junction Tree, No Evidence




$\int_{y_t} P(y_t \mid x_t) = 1$ $\quad$ $\int_{x_t} P(x_{t+1} \mid x_t) P(x_t) = P(x_{t+1})$ $\quad$ $\int_{y_0} P(y_0 \mid x_0) P(x_0) = P(x_0)$

# Junction Tree, With Evidence