Introduction to Artificial Intelligence

Lecture 15 – Temporal models

CS/CNS/EE 154

Andreas Krause

Announcements

- Homework 3 out, due Wed Nov 24
  - Questions 1 & 3 already covered
  - MDPs (question 2) coming up Wednesday
- Code for project final released later today

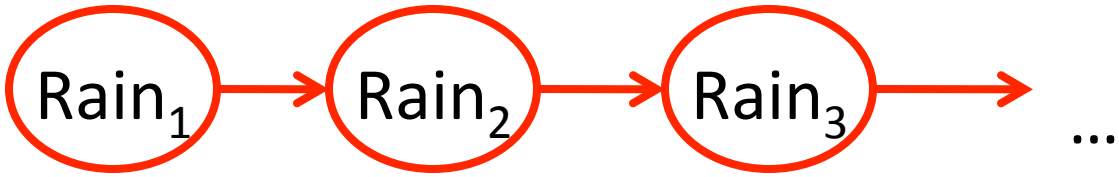
Temporal models

- So far: "Static" models (no notion of time)
  - Variables don't change values
- In practice:
  - World changes over time
  - Want to "keep track" of change by using probabilistic inference
- *Basic idea*: Create "copies" of variables, one per time step

Dynamical models

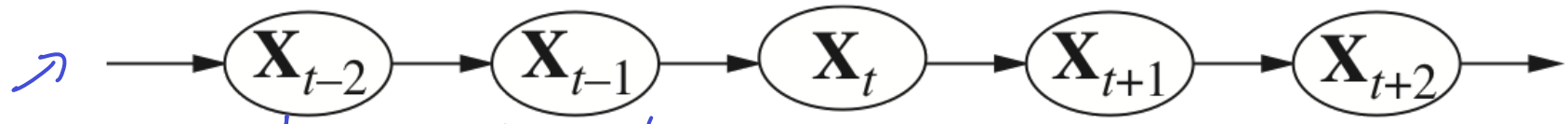
- Static model

- Dynamic model



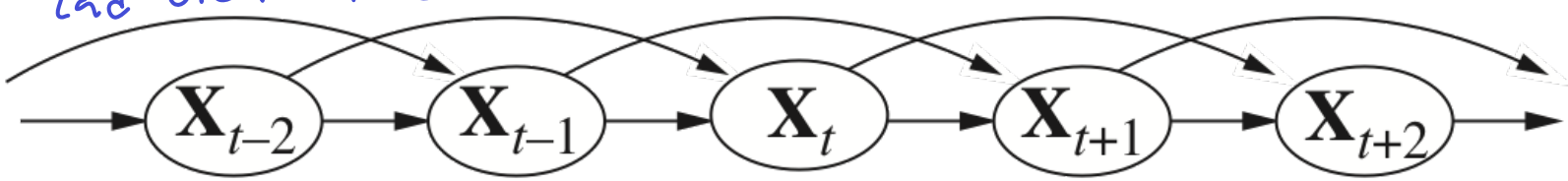
- Assumes discrete, unit-length time steps!

Markov chains



This lecture 1st order Markov chain

2nd order MC



- **Markov assumption:** 1st order MC: $X_t \perp X_{1:t-2} \mid X_{t-1}$

  k-th order MC: $X_t \perp X_{1:t-k-1} \mid X_{t-k:t-1}$

- **Stationarity assumption:** 1st order $P(X_t \mid X_{t-1})$ indep. of t

  k-th order $P(X_t \mid X_{t-k:t-1})$ —"—

Prediction in Markov Chains

- E.g.: Given that it rains now, how likely is it to rain a week from now?

Sps. I know $P(X_1)$, $P(X_t|X_{t-1})$

What is $P(X_t)$

$$P(X_t) = \sum_{x_{1:t-1}} P(x_{1:t}) = \sum_{x_{1:t-1}} \underbrace{P(X_t|X_{1:t-1})}_{P(X_t|X_{t-1})} \underbrace{P(X_{1:t-1})}_{P(X_{t-1})\cdot P(X_{1:t-2}|X_{t-1})}$$

$$= \sum_{x_{t-1}} P(X_t|X_{t-1}) P(X_{t-1}) \underbrace{\sum_{x_{1:t-2}} P(X_{1:t-2}|X_{t-1})}_{1}$$

$$= \sum_{x_{t-1}} P(X_t|X_{t-1}) P(X_{t-1})$$

Prediction in Markov Chains

$$P_t = P(X_t) = \sum_{x_{t-1}} \underbrace{P(x_{t-1})}_{P_{t-1}} P(x_t \mid x_{t-1})$$

$$P_t = P_{t-1} T$$

$$P_t = P_{t-2} T^2$$

$$P_t = P_0 T^t$$

$$T = \begin{pmatrix} P(1|1) & \cdots & P(k|1) \\ \vdots & & \vdots \\ P(1|k) & \cdots & P(k|k) \end{pmatrix}$$

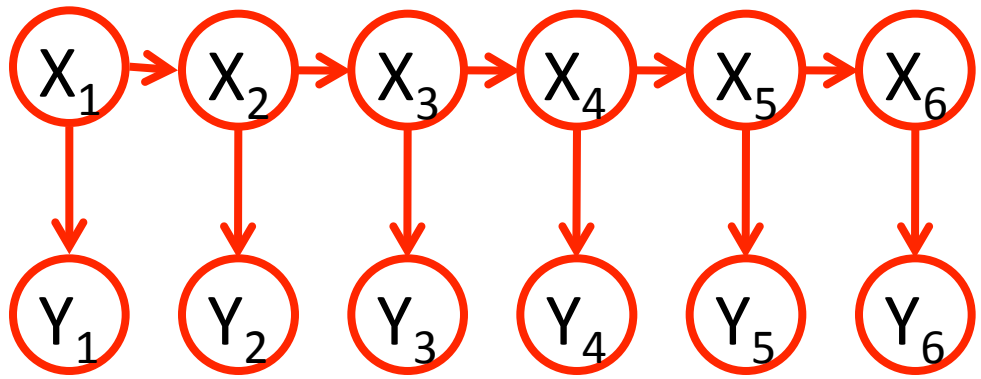
"transition matrix"

HMMs / Kalman Filters

- Most famous Bayesian networks:
  - Naïve Bayes model
  - Hidden Markov model
  - Kalman Filter

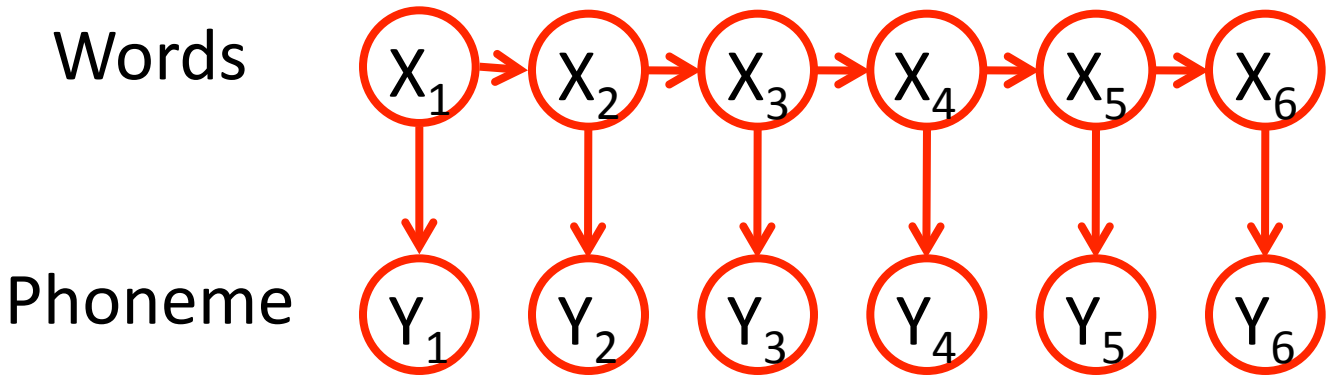
- Hidden Markov models
  - Speech recognition
  - Sequence analysis in comp. bio
- Kalman Filters control
  - Cruise control in cars
  - GPS navigation devices
  - Tracking missiles..
- Very simple models but very powerful!!

HMMs / Kalman Filters



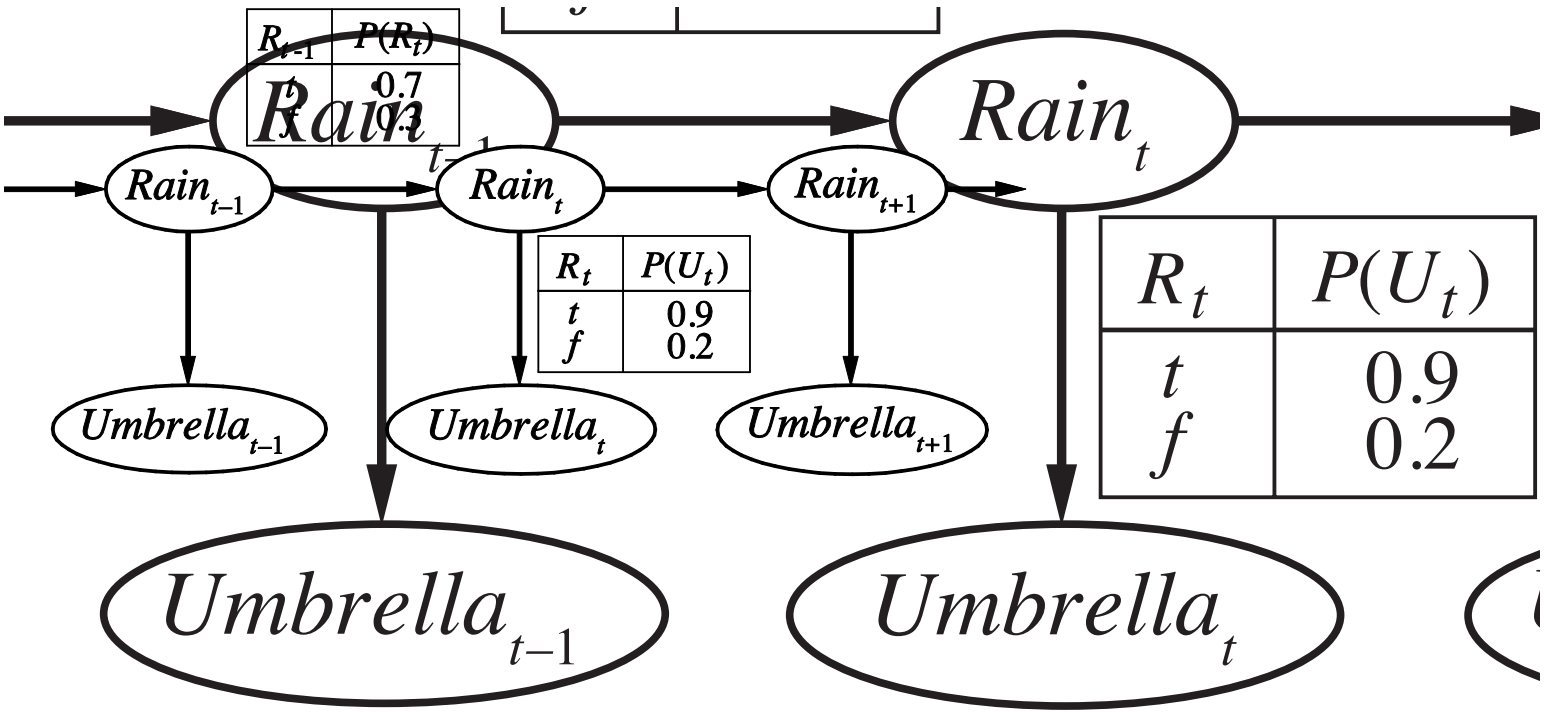
- $X_1,\ldots,X_T$: Unobserved (hidden) variables
- $Y_1,\ldots,Y_T$: Observations
- HMMs: X_i Multinomial, Y_i multinomial (or arbitrary)
- Kalman Filters: X_i, Y_i Gaussian distributions

HMMs for speech recognition



"He ate the cookies on the couch"

Example: Umbrella world



Inference tasks

Filtering $P(X_T \mid Y_{1:T})$

Prediction $P(X_{T+k} \mid y_{1:T})$

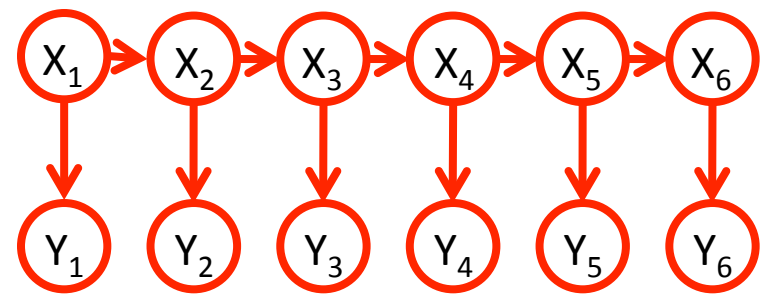
Smoothing $P(X_t \mid y_{1:T})$ $1 \le t \le T$

Most probable explanation

$$\underset{x_{1:T}}{\operatorname{argmax}} \; P(x_{1:T} \mid y_{1:T})$$

Inference in Hidden Markov Models

- Inference:
  - In principle, can use variable elimination / belief propagation
  - New variables X_t, Y_t at each time step ➔ need to rerun
  - Complexity grows with time!!

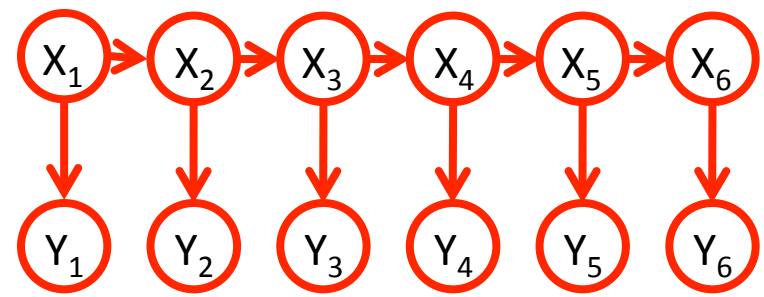


- Bayesian Filtering:
  - Suppose we already have computed $P(X_t \mid y_{1,\ldots,t})$
  - Want to efficiently (*recursively*) compute $P(X_{t+1} \mid y_{1,\ldots,t+1})$

$$P(X_{t+1} \mid y_{1:t+1}) = f\left(y_{t+1},\ P(X_t \mid y_{1:t})\right)$$

Bayesian filtering

- Start with $P(X_1)$
- At time t
  - Assume we have $P(X_t \mid y_{1\ldots t-1})$
  - Conditioning: $P(X_t \mid y_{1\ldots t})$



$$P(X_t \mid y_{1:t}) = \frac{1}{Z} P(X_t \mid y_{1:t-1}) \cdot P(y_t \mid X_t)$$

Have $P(X_1)$
Want $P(X_1 \mid Y_1)$
$= \frac{1}{Z} P(Y_1 \mid X_1) \cdot P(X_1)$
Bayes' rule

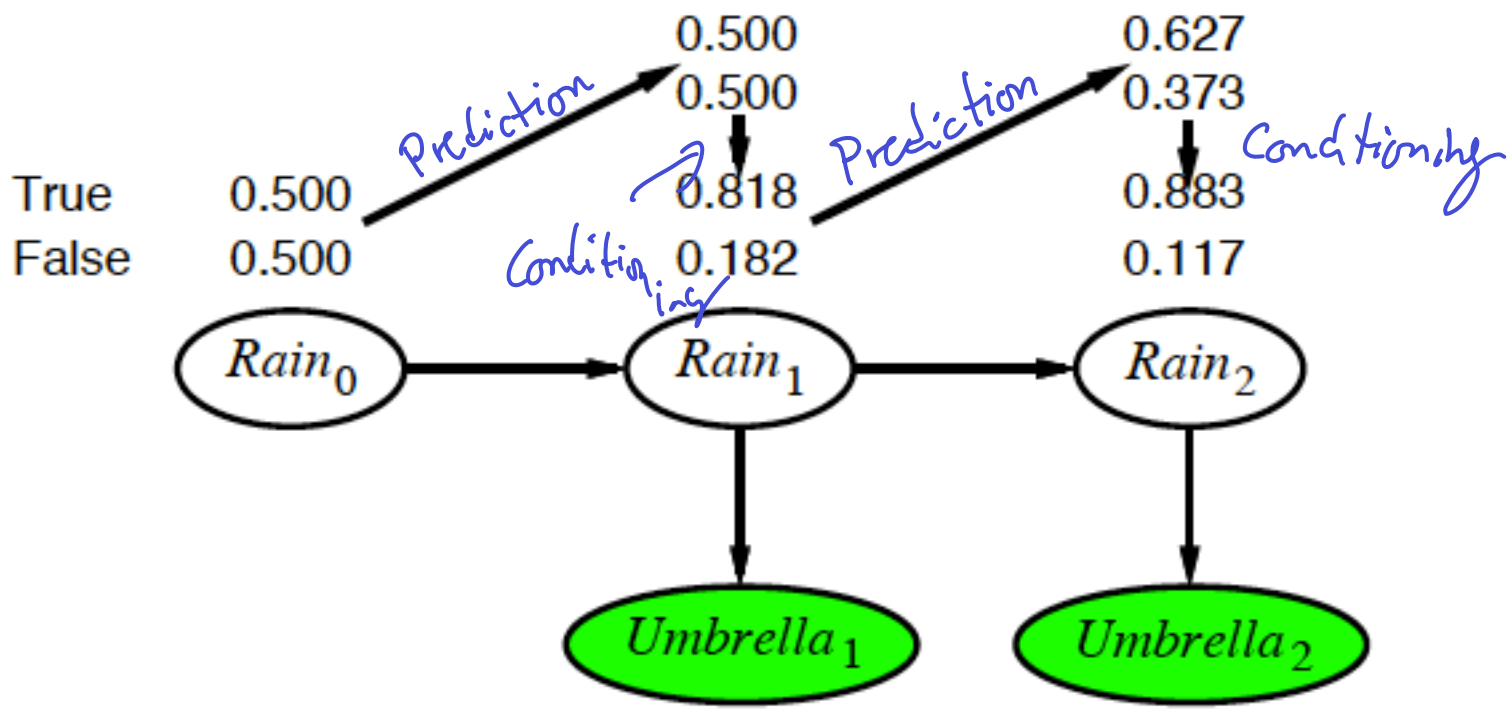
  - Prediction: $P(X_{t+1} \mid y_{1\ldots t})$

$$P(X_{t+1} \mid y_{1:t}) = \sum_{x_t} P(X_{t+1}, x_t \mid y_{1:t})$$

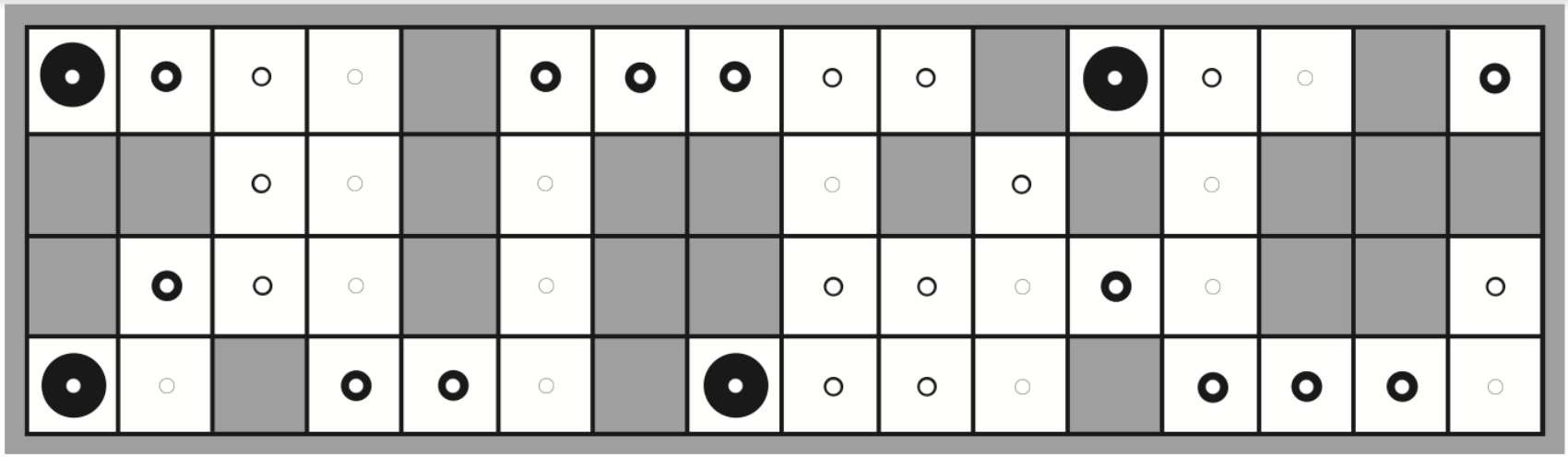
$$= \sum_{x_t} P(x_t \mid y_{1:t}) \underbrace{P(x_{t+1} \mid x_t, y_{1:t})}_{P(x_{t+1} \mid x_t)}$$

For k states, can do filtering in $O(k^2)$

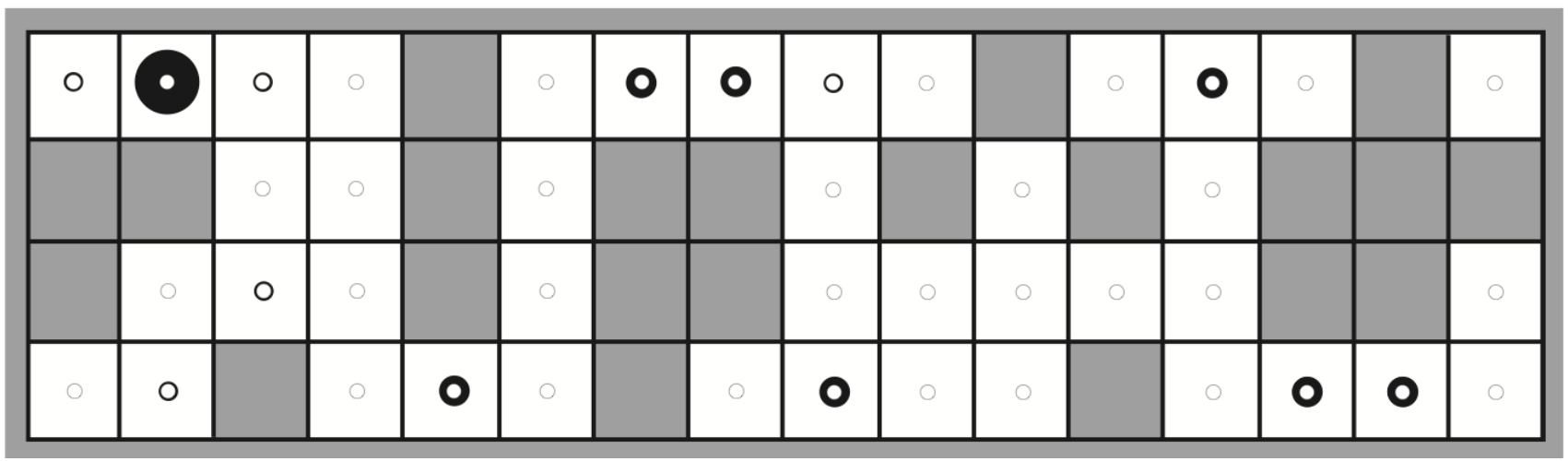
Understanding Bayesian filtering



HMM for robot localization



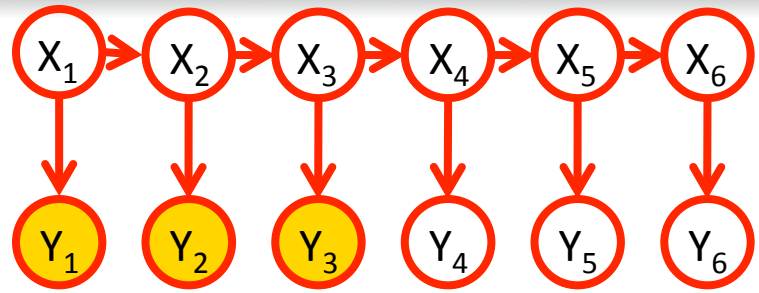
(a) Posterior distribution over robot location after $E_1 = NSW$



(b) Posterior distribution over robot location after $E_1 = NSW, E_2 = NS$

Prediction in HMMs

- Have: $P(X_t \mid y_{1:t})$
- Want: $P(X_{t+k} \mid y_{1:t})$



Just leave out conditioning!

$X_{t:t+k} \mid y_{1:t}$ is MC!

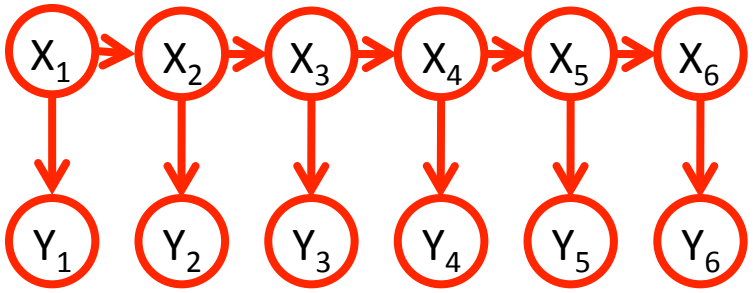
$p_t = P(X_t \mid y_{1:t})$

$p_{t+k} = p_t \cdot T^k$

$$T = \begin{pmatrix} P(1|1) & \cdots & P(k|1) \\ \vdots & & \vdots \\ P(1|k) & \cdots & P(k|k) \end{pmatrix}$$

Smoothing / MPE

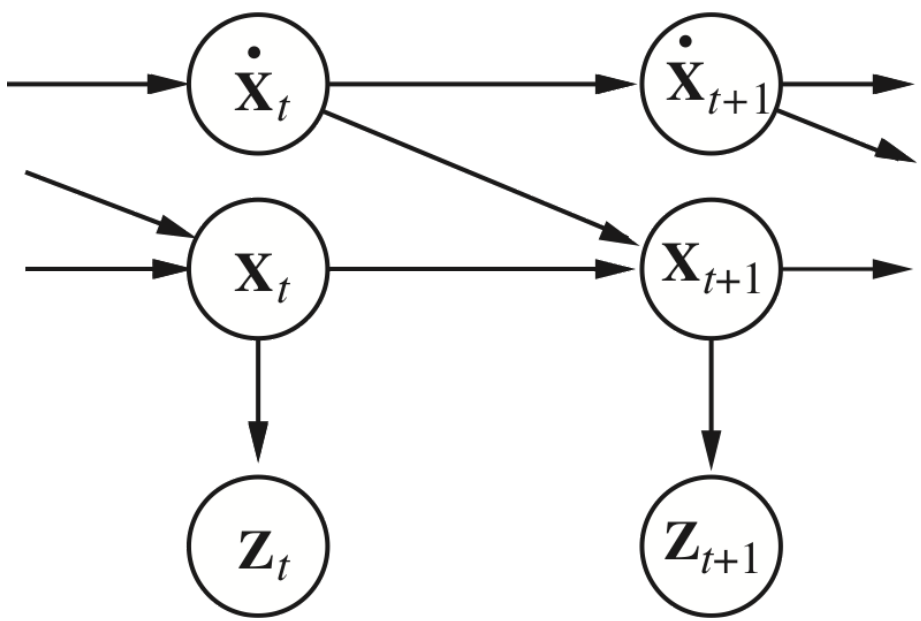
- Smoothing: $P(X_t \mid \mathbf{y_{1:T}})$
- Most probable explanation: $\arg\max_{\mathbf{x}_{1:T}} P(\mathbf{x}_{1:T} \mid \mathbf{y_{1:T}})$



- HMM is polytree Bayesian network!
- Can use *sum product* (aka forward-backward) for smoothing and *max product* (aka Viterbi algo) for MPE
- Specialized implementations using matrix algebra

Kalman filters

- Track objects in *continuous domain* using noisy measurements
  - E.g., birds flying, robots moving, chemical plants, …
- System described using Gaussian variables
  - E.g., location in X,Y,Z; velocity in X,Y,Z; acceleration in X,Y,Z,…

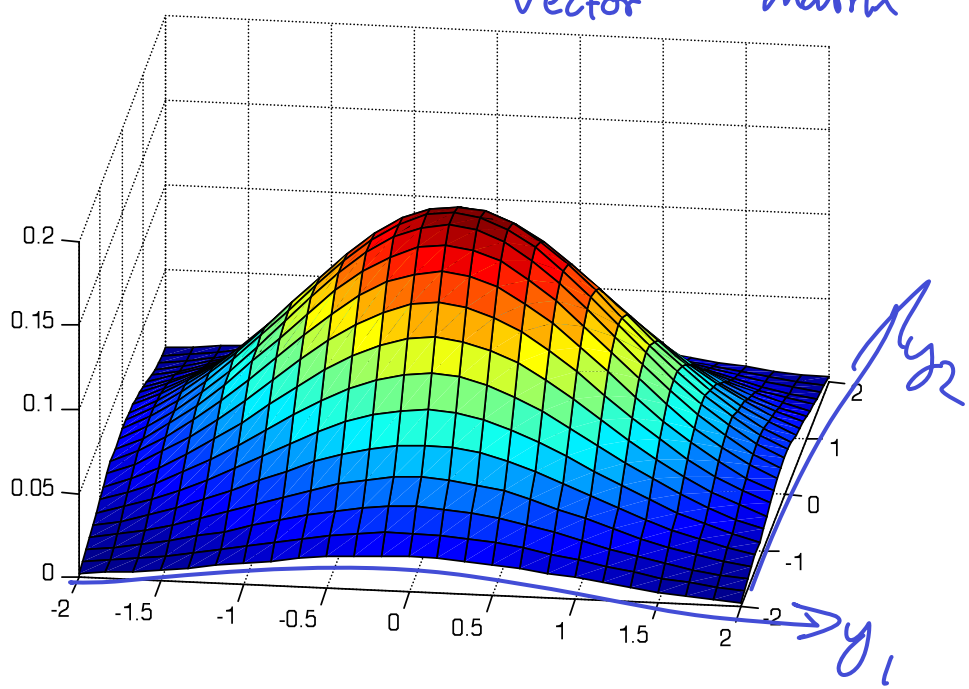


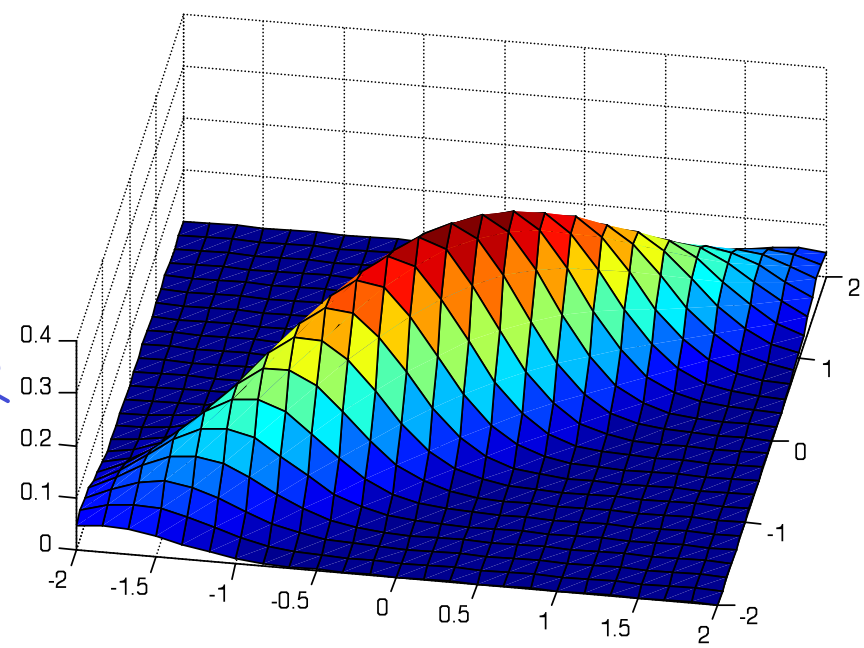
Bivariate Gaussian distribution

$$\frac{1}{2\pi\sqrt{|\Sigma|}}\exp\left(-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right)$$

Mean vector

Covariance matrix





$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$$

Multivariate Gaussian distribution

$$\mathcal{N}(y; \Sigma, \mu) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)\right)$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \vdots & & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix}$$

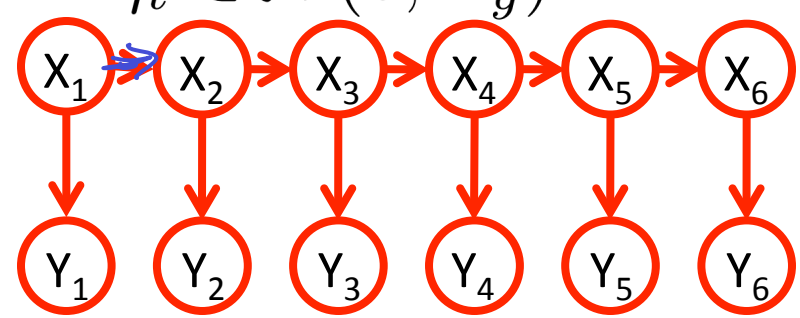
Kalman Filters (Gaussian HMMs)

- $X_1,\ldots,X_T$: Location of object being tracked $\in \mathbb{R}^d$
- $Y_1,\ldots,Y_T$: Observations $\in \mathbb{R}^{d'}$
- $P(X_1)$: Prior belief about location at time 1
- $P(X_{t+1} | X_t)$: "Motion model"
  - How do I expect my target to move in the environment?

$$\mathbf{X}_{t+1} = \mathbf{F}\mathbf{X}_t + \varepsilon_t \text{ where } \varepsilon_t \in \mathcal{N}(0, \Sigma_x)$$

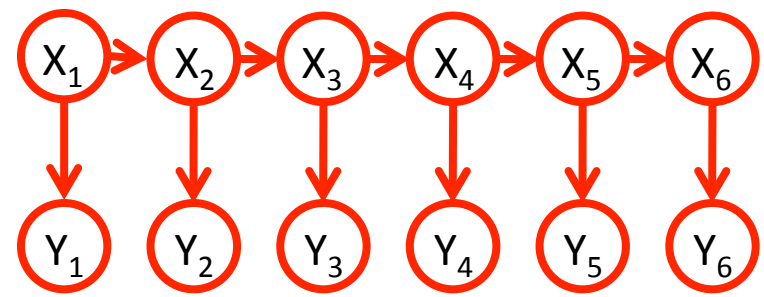
- $P(Y_t \mid X_t)$: "Sensor model"
  - What do I observe if target is at location X_t?

$$\mathbf{Y}_t = \mathbf{H}\mathbf{X}_t + \eta_t \text{ where } \eta_t \in \mathcal{N}(0, \Sigma_y)$$



Bayesian filtering in KFs (1D)

- Start with $P(X_1)$
- At time t
  - Assume we have $P(X_t \mid y_{1\ldots t-1})$
  - Conditioning: $P(X_t \mid y_{1\ldots t})$



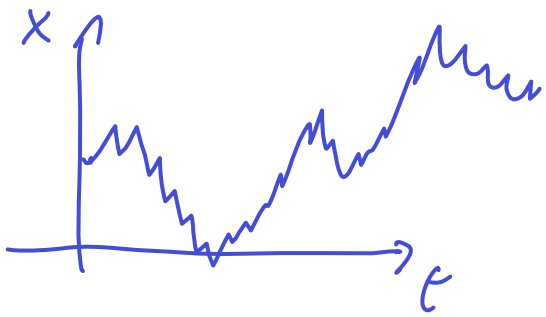
$$P(X_t \mid y_{1:t}) = \frac{1}{Z} P(X_t \mid y_{1:t-1}) P(y_t \mid X_t)$$

  - Prediction: $P(X_{t+1} \mid y_{1\ldots t})$

$$P(X_{t+1} \mid y_{1:t}) = \int P(X_{t+1} \mid X_t) P(X_t \mid y_{1:t}) \, dX_t$$

Example: Simple random walk

- Transition / motion model $P(x_{t+1} \mid x_t) = \mathcal{N}(x_t, \sigma_x^2)$



$x_{t+1} = x_t + \epsilon_t \;,\; \epsilon_t \sim N(0, \sigma_x^2)$

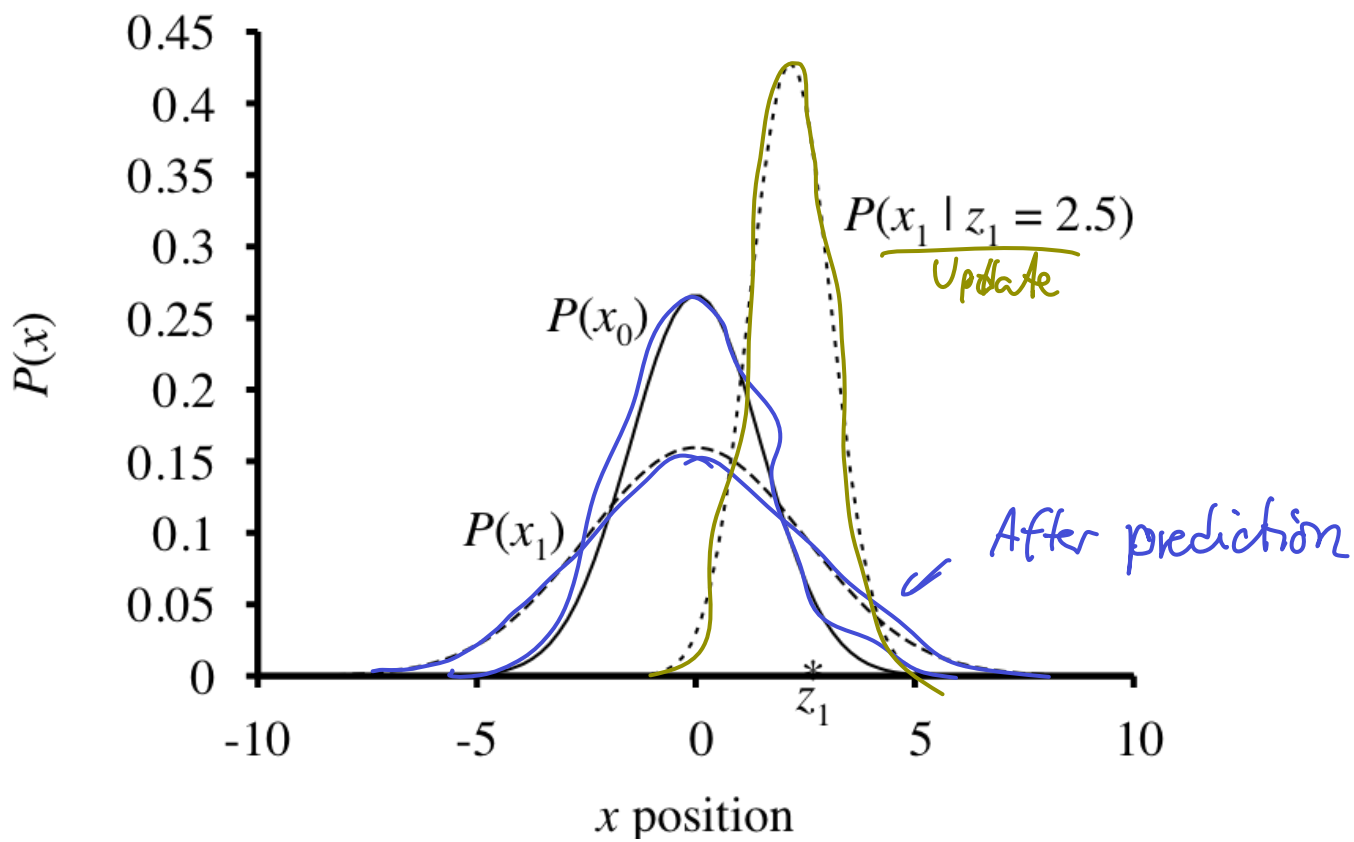
- Sensor model $P(y_t \mid x_t) = \mathcal{N}(x_t, \sigma_y^2)$

$y_t = x_t + \eta_t \;,\; \eta_t \sim N(0, \sigma_y^2)$

- State at time t: $P(x_t \mid \mathbf{y}_{1:t}) = \mathcal{N}(\mu_t, \sigma_t^2)$

Example: Bayesian filtering in KFs

$$\mu_{t+1} = \frac{\sigma_y^2 \mu_t + (\sigma_t^2 + \sigma_x^2) y_{t+1}}{\sigma_t^2 + \sigma_x^2 + \sigma_y^2} \qquad \sigma_{t+1}^2 = \frac{(\sigma_t^2 + \sigma_x^2)\sigma_y^2}{\sigma_t^2 + \sigma_x^2 + \sigma_y^2}$$



Suppose: $P(x_t \mid \mathbf{y}_{1:t}) = \mathcal{N}(\mu_t, \sigma_t^2)$

General Kalman update

- Transition model $P(\mathbf{x}_{t+1} \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t+1}; \mathbf{F}\mathbf{x}_t, \Sigma_x)$
- Sensor model $P(\mathbf{y}_t \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t; \mathbf{H}\mathbf{x}_t, \Sigma_y)$

- Kalman Update:

$$\mu_{t+1} = \mathbf{F}\mu_t + \mathbf{K}_{t+1}(\mathbf{y}_{t+1} - \mathbf{H}\mathbf{F}\mu_t)$$

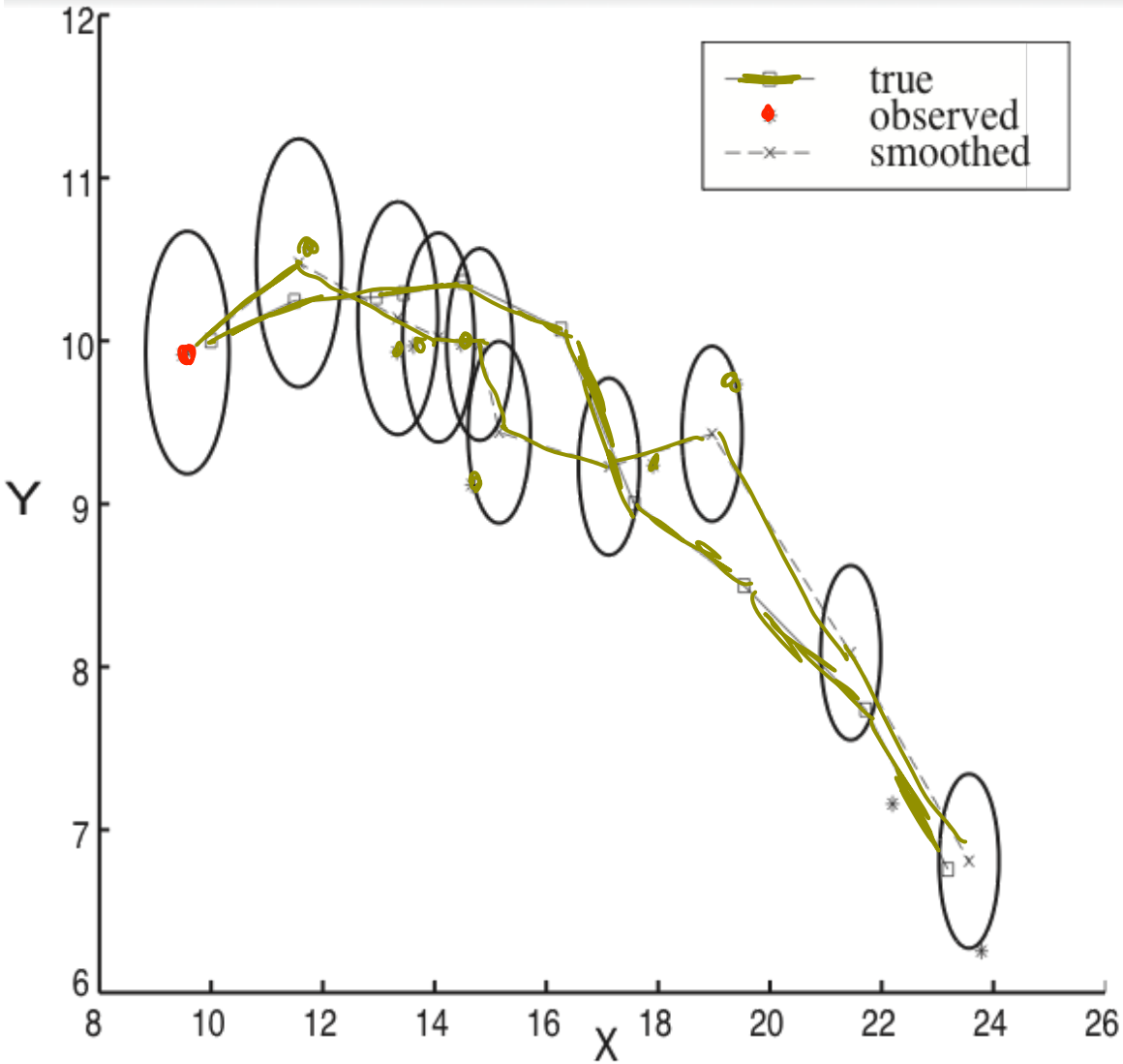
$$\Sigma_{t+1} = (\mathbf{I} - \mathbf{K}_{t+1})(\mathbf{F}\Sigma_t\mathbf{F}^T + \Sigma_x)$$

- Kalman gain:

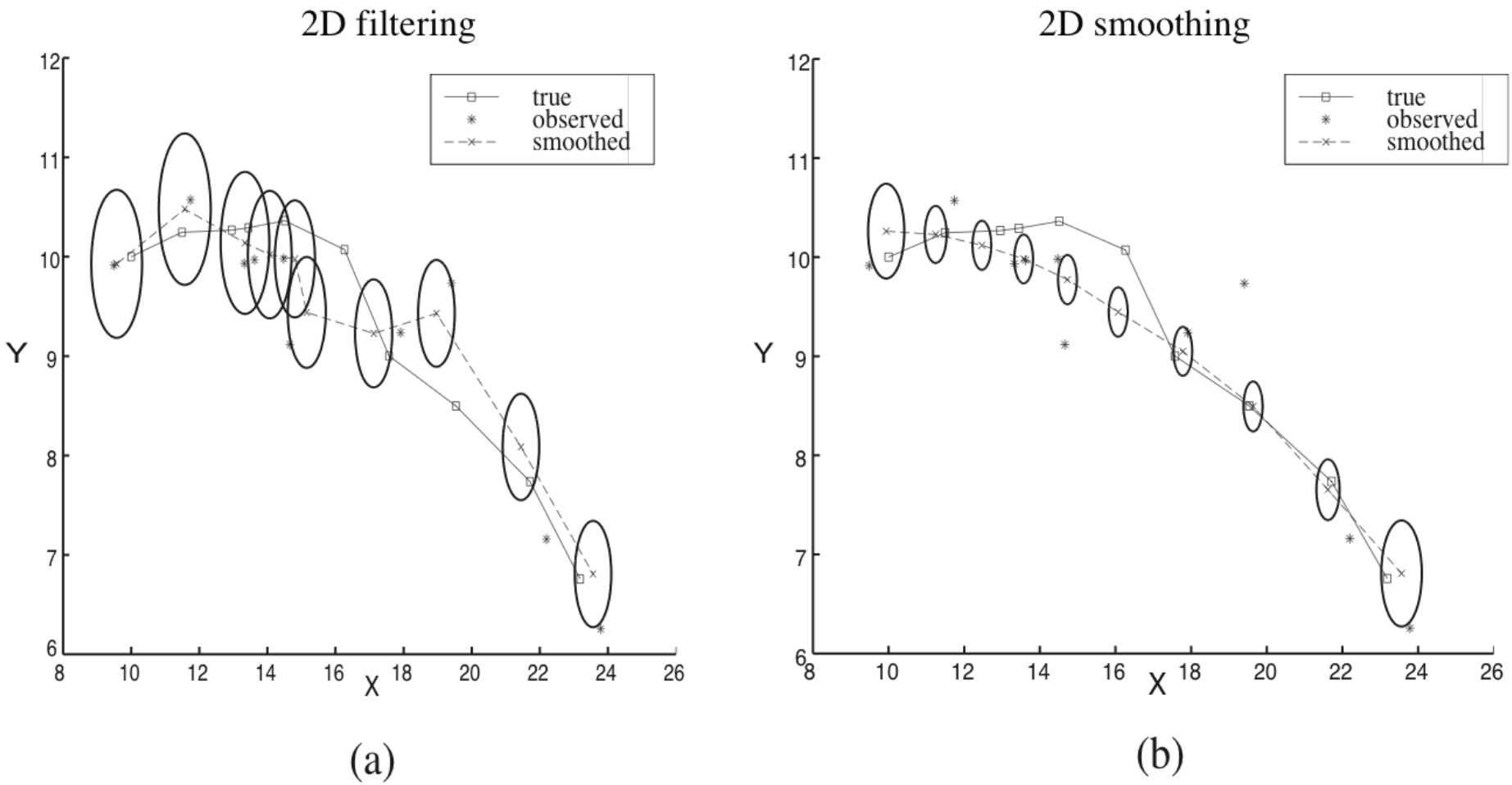
$$\mathbf{K}_{t+1} = (\mathbf{F}\Sigma_t\mathbf{F}^T + \Sigma_x)\mathbf{H}^T(\mathbf{H}(\mathbf{F}\Sigma_t\mathbf{F}^T + \Sigma_x)\mathbf{H}^T + \Sigma_y)^{-1}$$

- Can compute Σ_t and $\mathbf{K}_t$ offline

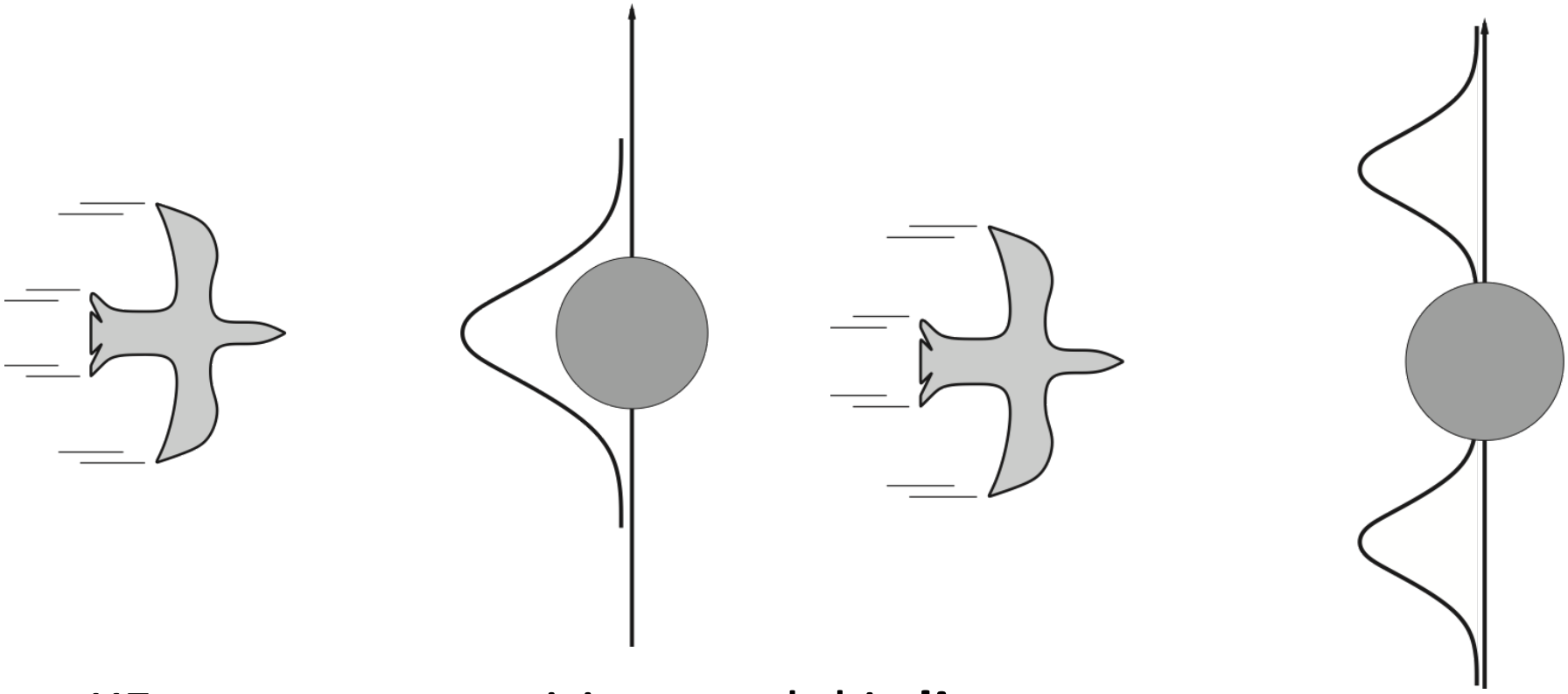
2D tracking example



Kalman smoothing



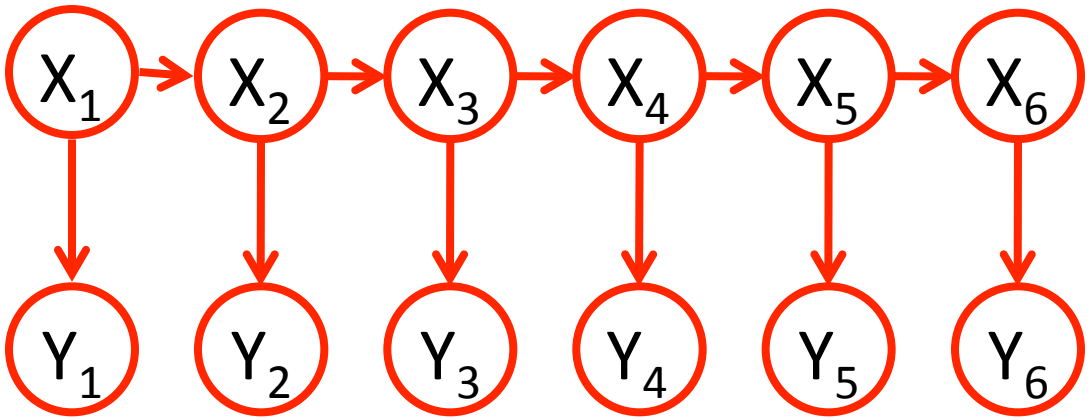
When KFs fail



- KFs assume transition model is **linear**
  - Implies that predictive distribution is Gaussian (unimodal)
- Need approximate inference to capture nonlinearities!

Factored dynamical models

- So far: HMMs and Kalman filters

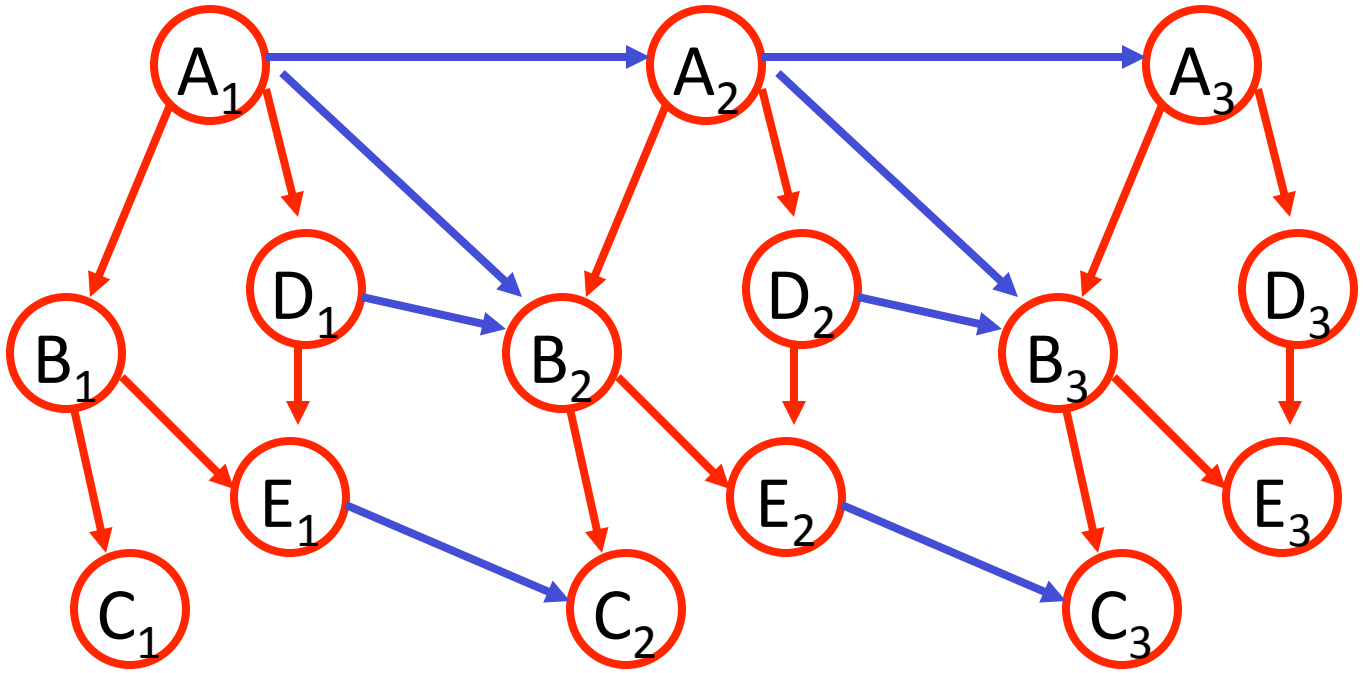


- What if we have more than one variable at each time step?
  - E.g., temperature at different locations, or road conditions in a road network?

➔ Dynamic Bayesian Networks

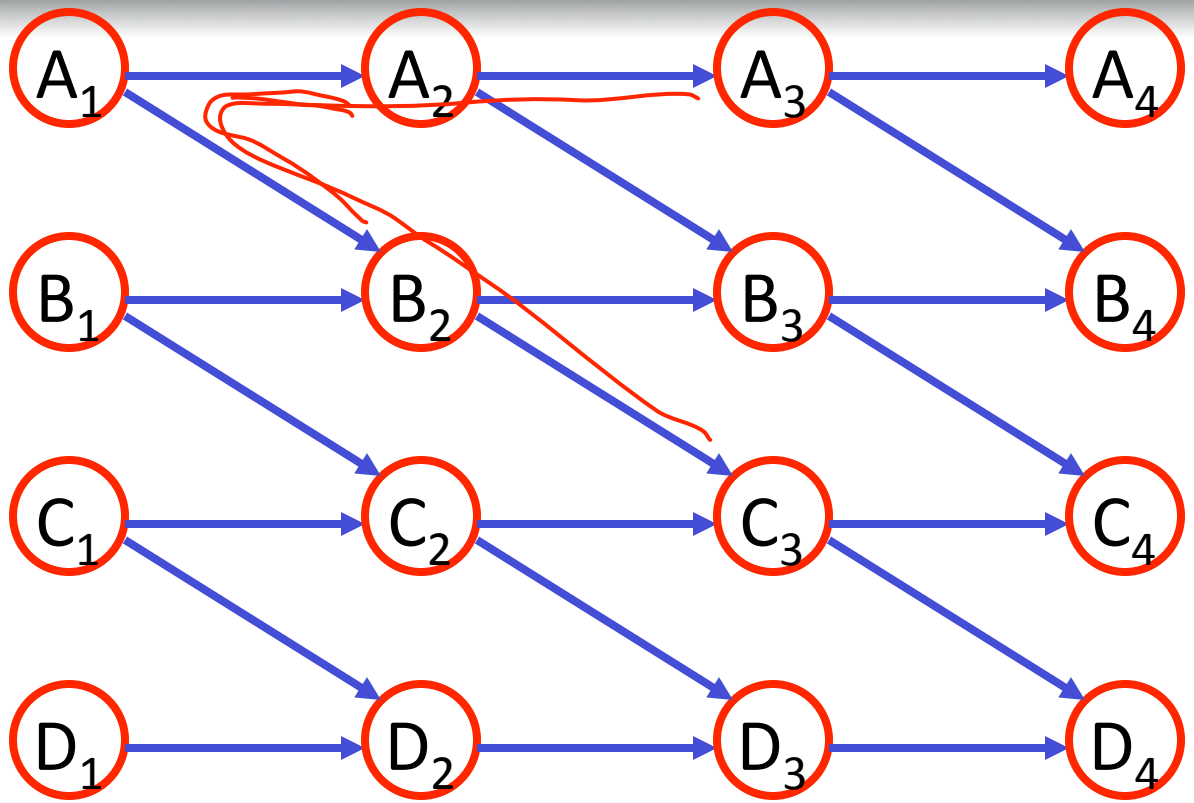
Dynamic Bayesian Networks

- At every timestep have a *Bayesian Network*



- Variables at each time step t called a slice $\mathbf{S_t}$
- "Temporal" edges connecting S_{t+1} with S_t

Flow of influence in DBNs



$A_1 \perp B_1$ ✓

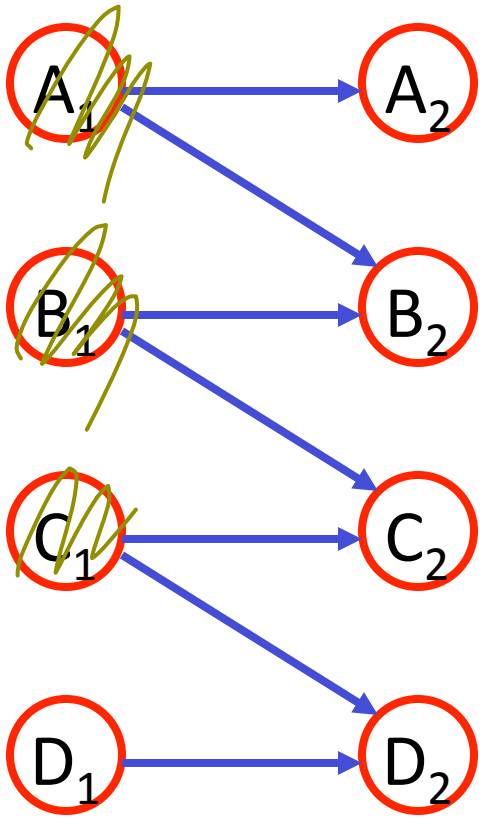
$A_2 \perp B_2$ ✗

$A_2 \perp C_2$ ✓

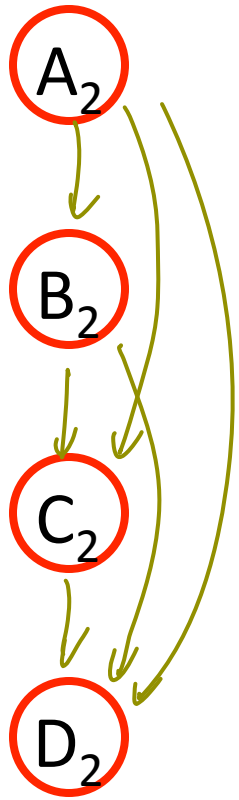
$A_3 \perp C_3$ ✗

Inference in DBNs?

DBN



Marginals at time 2



Need approximate inference!

Particle filtering

- Very useful approximate inference technique for dynamical models
  - Nonlinear Kalman filters
  - Dynamic Bayesian networks

- **Basic idea**: Approximate the posterior at each time by samples (particles), which are propagated and reweighted over time