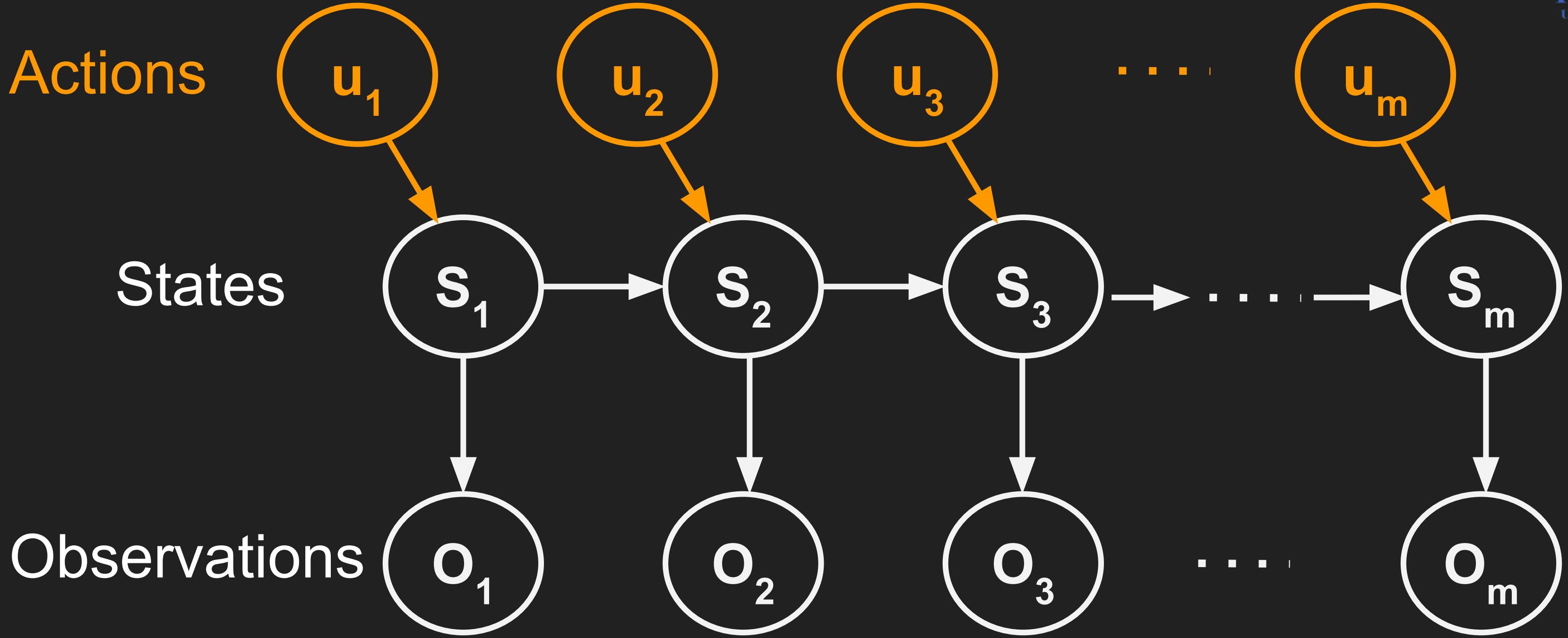
Probabilistic Graphical Models

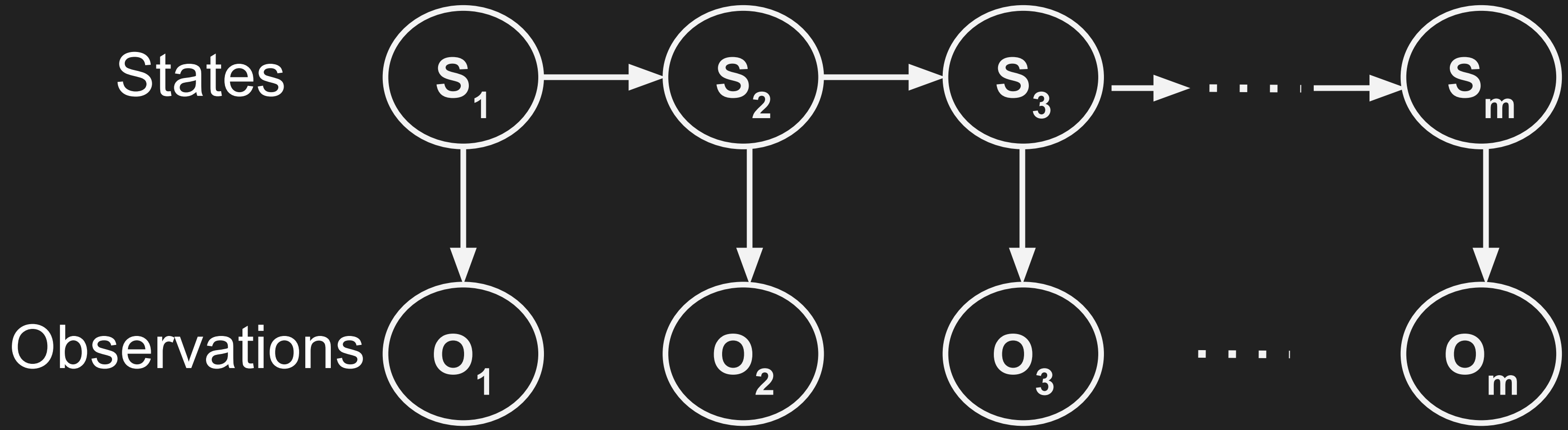
Lectures 17

Inference in Temporal Models
Recursive Bayesian Filtering
Intro to Kalman Filter

Remember: Temporal Models

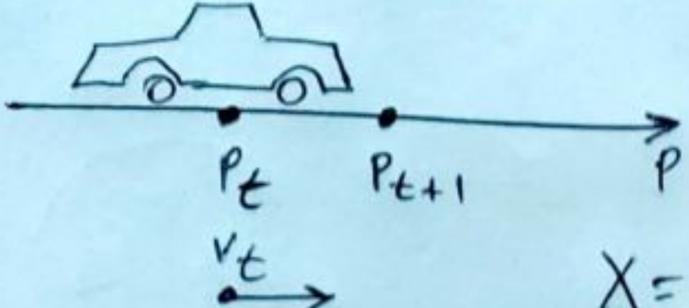


Remember: Temporal Models



Remember: Temporal Models

pgm17 (I)



$$X_t = \begin{bmatrix} p_t \\ v_t \end{bmatrix}$$

$$X_t = \begin{bmatrix} p_t \\ v_t \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_p \\ \varepsilon_v \end{bmatrix}$$

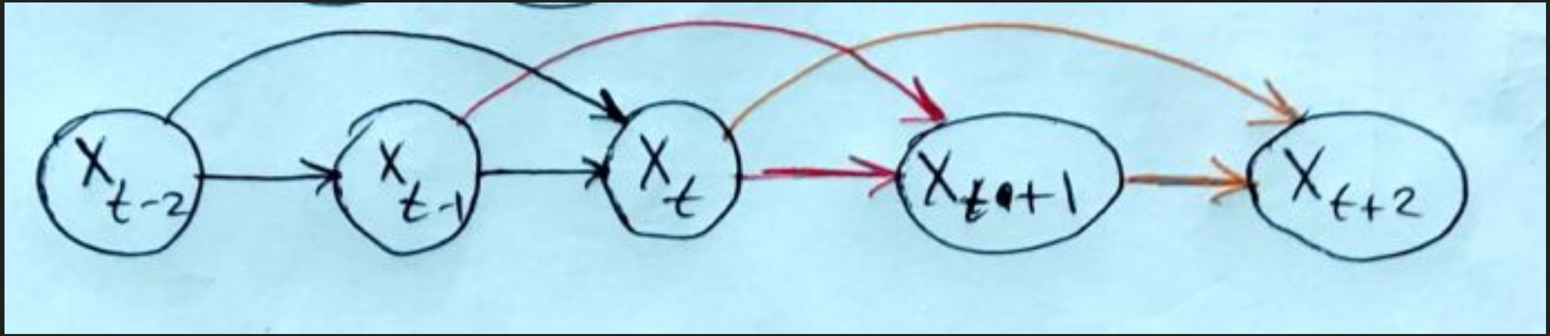
$$X_t = A X_{t-1} + \varepsilon_X$$

$$\varepsilon_X \sim P_{\varepsilon_X}$$

$$P(X_t \mid X_{t-1}) = P_\varepsilon (X_t - A X_{t-1})$$

What if ε_v cannot model approximation?

Solution 1: create more dependencies.



What if ε_v cannot model approximation?

K. N. Toosi University of Technology

Solution 2: add acceleration to state vector.

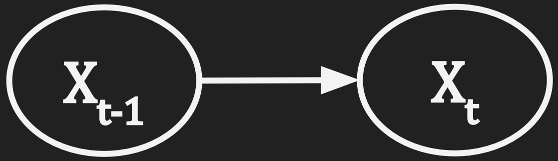
$$\vec{X}_t = \begin{bmatrix} p_t \\ v_t \end{bmatrix} \Longrightarrow \vec{X}_t = \begin{bmatrix} p_t \\ v_t \\ a_t \end{bmatrix}$$

$$\vec{X}_t = \begin{bmatrix} p_t \\ v_t \\ a_t \end{bmatrix} = \begin{bmatrix} p_{t-1} + \Delta t\, v_{t-1} + \frac{1}{2}\Delta t^2 a_{t-1} + \varepsilon_p \\ v_{t-1} + \Delta t\, a_{t-1} + \varepsilon_v \\ a_{t-1} + \varepsilon_a \end{bmatrix}$$

$$\begin{bmatrix} p_t \\ v_t \\ a_t \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ \emptyset & 1 & \Delta t \\ \emptyset & \emptyset & 1 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ v_{t-1} \\ a_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_p \\ \varepsilon_v \\ \varepsilon_a \end{bmatrix}$$

$$\vec{X}_t = A\,\vec{X}_{t-1} + \vec{\varepsilon}_x$$

Remember: Temporal Models (nonlinear case)



Linear:

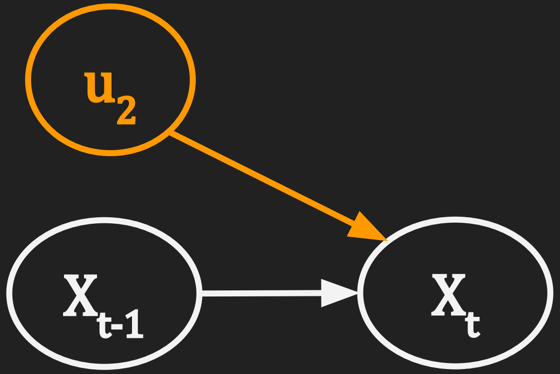
$$X_t = \mathbf{A}\, X_{t-1} + \varepsilon$$

Nonlinear:

$$X_t = f(X_{t-1}) + \varepsilon$$

$$X_t = f(X_{t-1}, \varepsilon)$$

Remember: Temporal Models (nonlinear case with input)



Linear:

$$X_t = \mathbf{A}\,X_{t-1} + \mathbf{B}\,u_t + \varepsilon$$

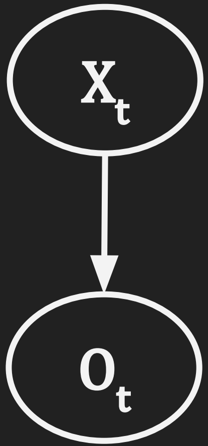
Nonlinear:

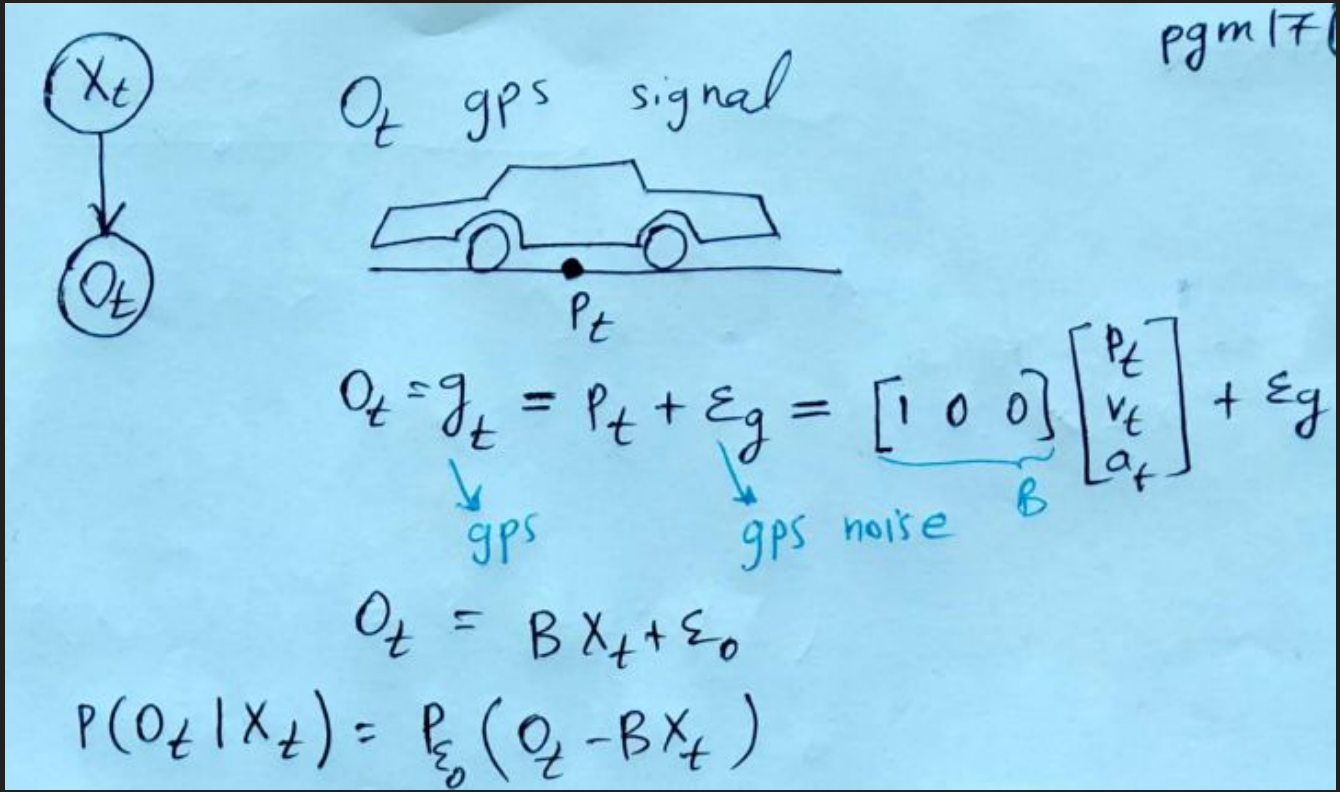
$$X_t = f(X_{t-1},\ u_t) + \varepsilon$$

$$X_t = f(X_{t-1},\ u_t, \varepsilon)$$

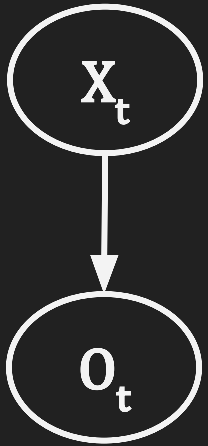
Remember: Observation Model Example

K. N. Toosi
University of Technology





Remember: Observation Model Example



$$O_t = \begin{bmatrix} gps \\ acc \end{bmatrix} = \begin{bmatrix} g_t \\ a'_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_t \\ v_t \\ a_t \end{bmatrix} + \begin{bmatrix} \varepsilon_g \\ \varepsilon_{a'} \end{bmatrix}$$

acc → شتاب سنج

$\varepsilon_{a'}$ → نویز شتاب سنج

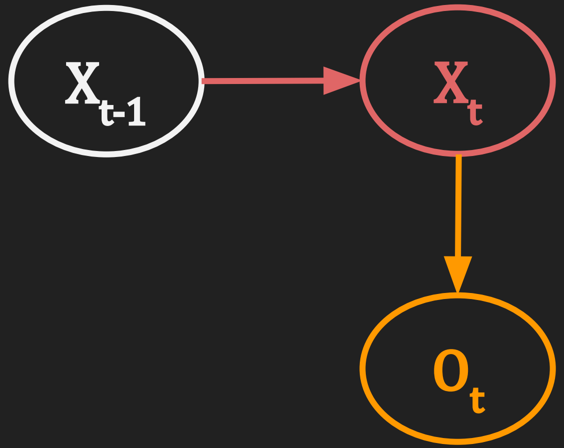
$$O_t = B X_t + \varepsilon_o$$

Remember: Transition and Observation Models

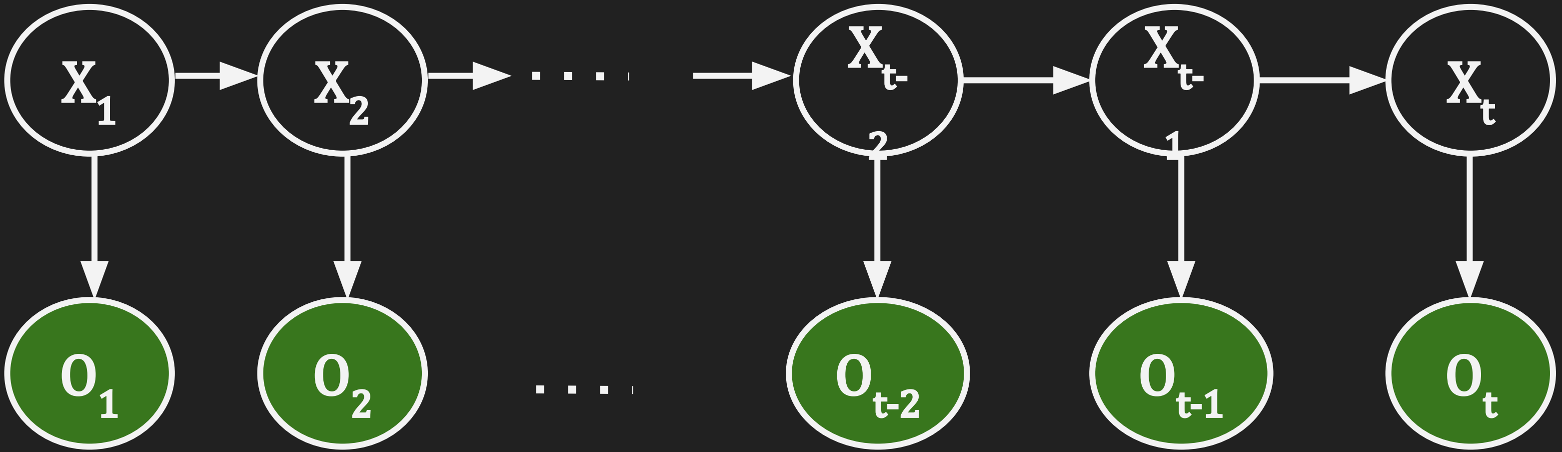
(No input)

$P(X_t \mid X_{t-1})$ Transition Model

$P(O_t \mid X_t)$ Observation Model



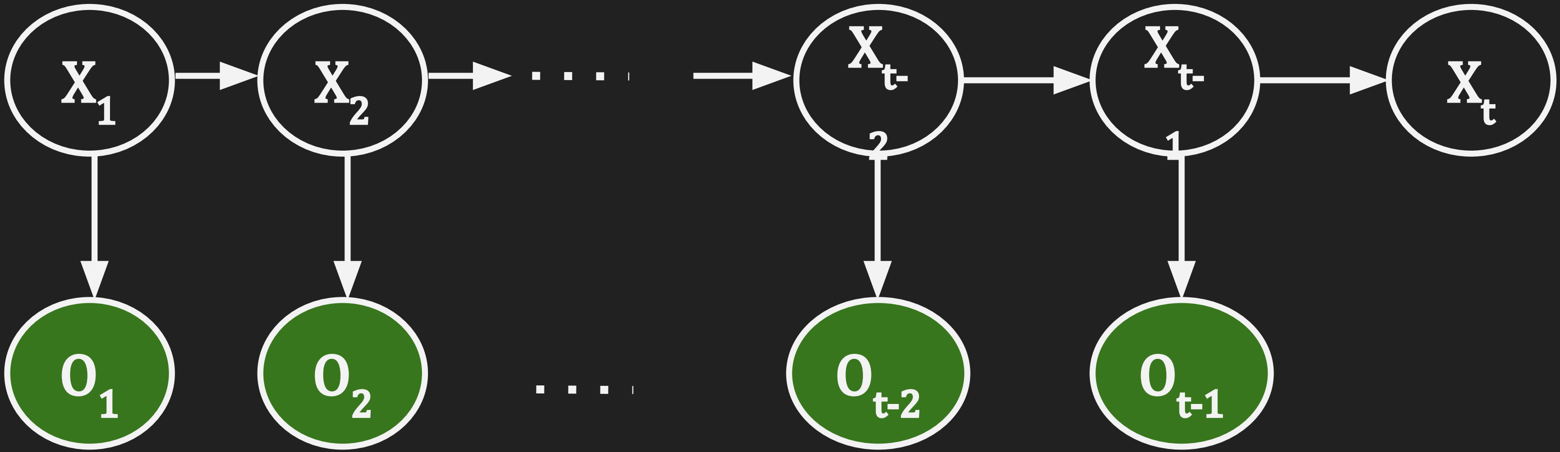
Inference: State Estimation



Given $O_{1,} O_{2}, \ldots, O_t$ what is X_t?

$P(X_t \mid O_{1,} O_{2}, \ldots, O_t) = ?$

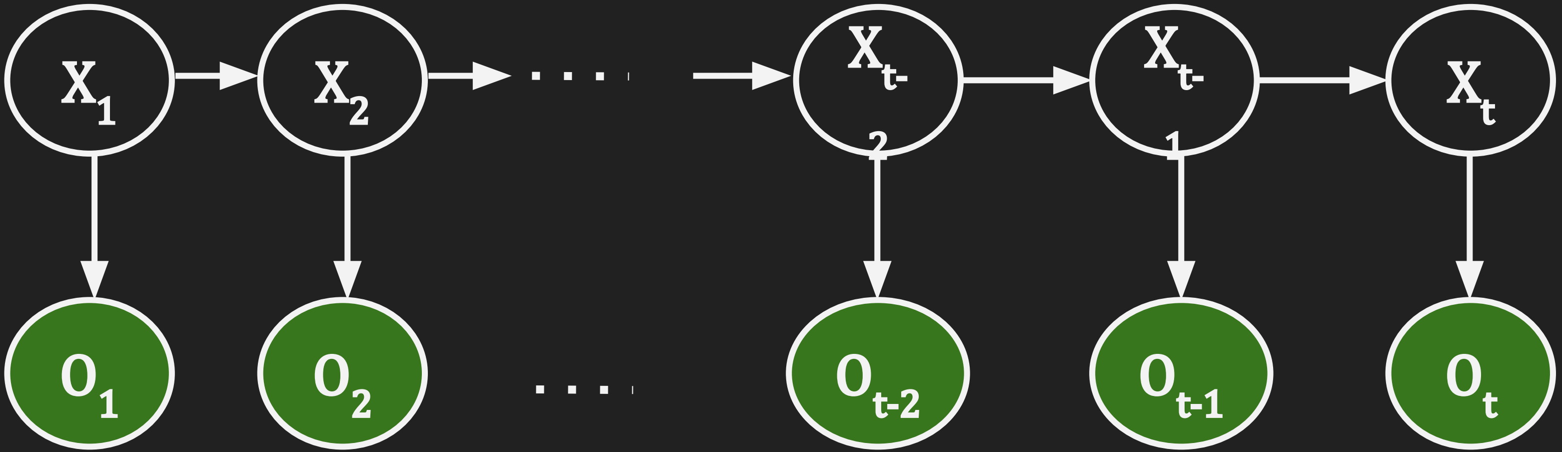
Inference: Prediction



Given $O_{1}, O_{2}, \ldots, O_{t-1}$ what is X_t?

$P(X_t \mid O_{1}, O_{2}, \ldots, O_{t-1}) = ?$

Inference Problems



$\mathrm{pred}(X_t) = P(X_t \mid O_1, O_2, \ldots, O_{t-1})$ Prediction

$\mathrm{corr}(X_t) = P(X_t \mid O_1, O_2, \ldots, O_t)$ Correction (update)

Inference Problems

$\text{pred}(X_t) = P(X_t \mid O_{1,} O_2, \ldots, O_{t-1})$ Prediction

$\text{corr}(X_t) = P(X_t \mid O_{1,} O_2, \ldots, O_t)$ Correction (update)

1. Variable Elimination

$$P(X_t \mid o_1 \text{—} o_{t-1}) = \frac{P(X_t, o_1 \text{—} o_{t-1})}{\sum_{X'_t} P(X'_t, o_1 \text{—} o_{t-1})}$$

$$= \frac{\sum_{X_{t-1}} \cdots \sum_{X_2} \sum_{X_1} P(X_1 \text{—} X_t, o_1 \text{—} o_{t-1})}{\ominus}$$

$$= \frac{\sum_{X_{t-1}} \cdots \sum_{X_2} \sum_{X_1} P(X_2 \mid X_1)\, P(X_3 \mid X_2) \cdots P(X_t \mid X_{t-1})\, P(o_1 \mid X_1) \cdots P(o_{t-1} \mid X_{t-1})}{\ominus}$$

Inference Problems

K. N. Toosi University of Technology

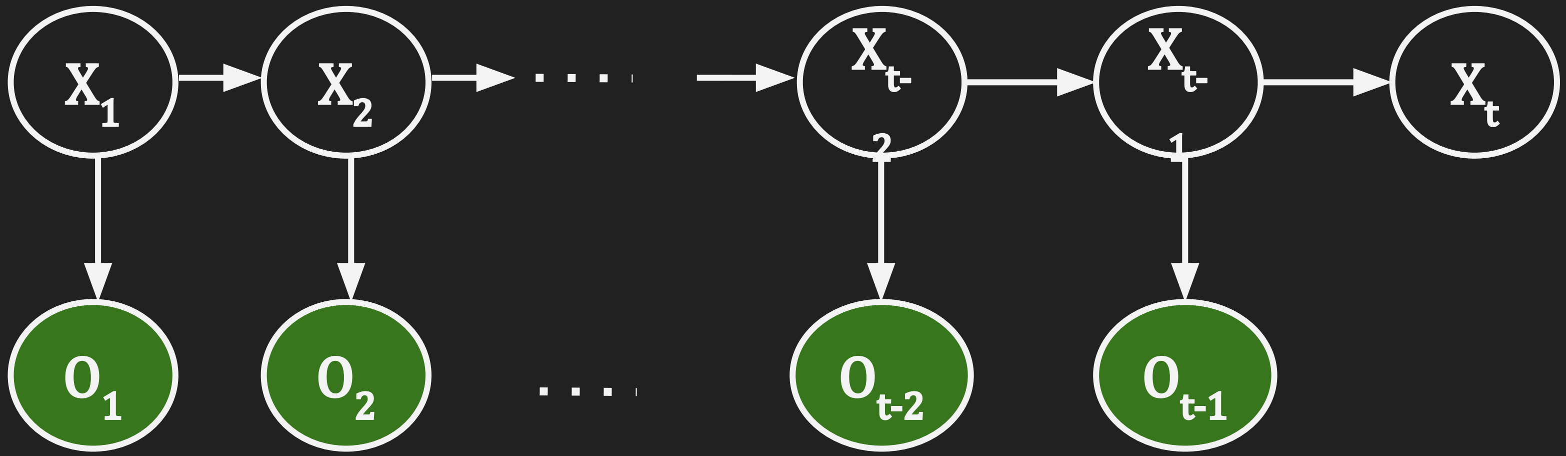
$pred(X_t) = P(X_t \mid O_{1,} O_{2}, \dots, O_{t-1})$ Prediction

$corr(X_t) = P(X_t \mid O_{1,} O_{2}, \dots, O_{t})$ Correction (update)

1. Variable Elimination
2. Recursive Bayesian Filtering

$$\cdots \rightarrow pred(X_t) \rightarrow corr(X_t) \rightarrow pred(X_{t+1}) \rightarrow corr(X_{t+1}) \rightarrow \cdots$$

Prediction Phase



$\text{pred}(X_t) = P(X_t \mid O_{1,} O_2, \dots, O_{t-1}) =$

Recursive Bayesian Filtering - Prediction

pgm 17 (II)

$$\text{pred}(x_t) = P(x_t \mid o_1, o_2, \dots, o_{t-1})$$

$$\text{corr}(x_t) = P(x_t \mid o_1, o_2, \dots, o_t)$$

$$\text{pred}(x_t) = P(x_t \mid o_1, \dots, o_{t-1}) = \sum_{x_{t-1}} P(x_t, x_{t-1} \mid o_1 — o_{t-1})$$

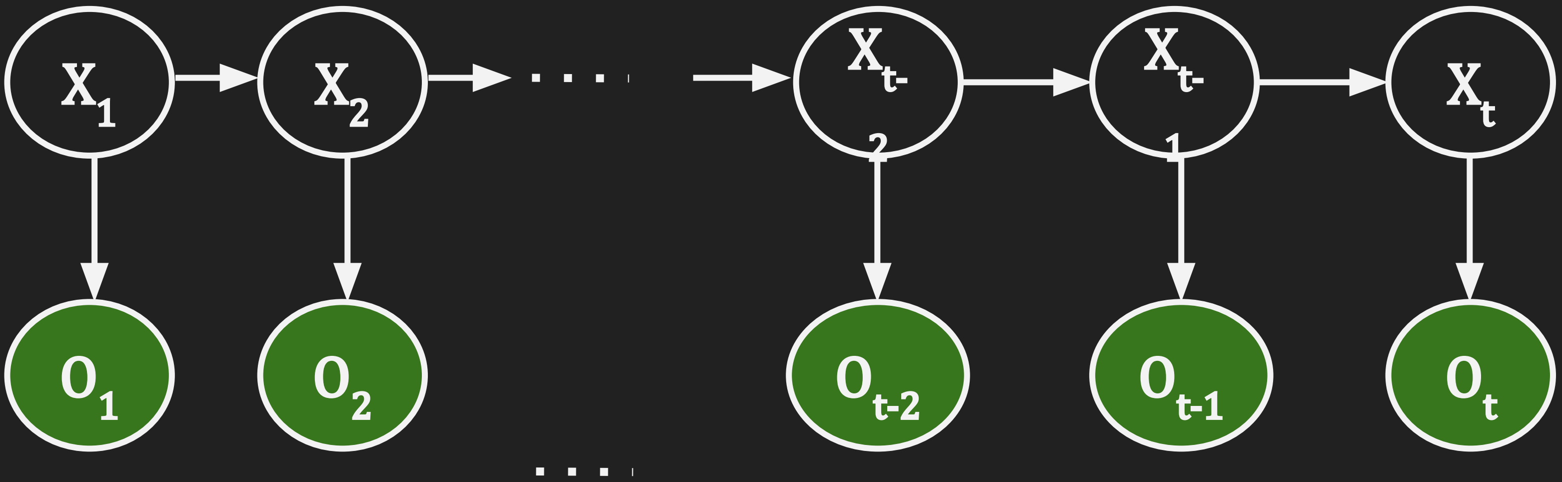
$$= \sum_{x_{t-1}} P(x_t \mid x_{t-1}, o_1, o_2, \dots, o_{t-1})\, P(x_{t-1} \mid o_1 — o_{t-1})$$

$$= \sum_{x_{t-1}} P(x_t \mid x_{t-1})\, \text{corr}(x_{t-1})$$

continuous

$$\text{pred}(x_t) = \int p(x_t \mid x_{t-1})\, \text{corr}(x_{t-1})\, dx_{t-1}$$

Correction Phase



$$\mathrm{corr}(X_t) = P(X_t \mid O_1, O_2, \ldots, O_t)$$

Recursive Bayesian Filtering - Correction

$$\text{corr}(X_t) = P(X_t \mid o_1 \ldots o_t) = P(X_t \mid o_1, \ldots, o_{t-1}, o_t)$$

$$= \frac{P(o_t \mid X_t, o_1 \ldots o_{t-1})\, P(X_t \mid o_1 \ldots o_{t-1})}{P(o_t \mid o_1 \ldots o_{t-1})}$$

$$= \frac{P(o_t \mid X_t, o_1 \ldots o_{t-1})\, P(X_t \mid o_1 \ldots o_{t-1})}{\sum_{X'_t} P(o_t, X'_t \mid o_1 \ldots o_{t-1})}$$

$$= \frac{P(o_t \mid X_t, o_1 \ldots o_{t-1})\, P(X_t \mid o_1 \ldots o_{t-1})}{\sum_{X'_t} P(o_t \mid X'_t, o_1 \ldots o_{t-1})\, P(X'_t \mid o_1 \ldots o_{t-1})}$$

$$= \frac{P(o_t \mid X_t)\ \text{pred}(X_t)}{\sum_{X'_t} P(o_t \mid X'_t)\ \text{pred}(X'_t)}$$

Initialization Phase

K. N. Toosi University of Technology

$pred(X_t) = P(X_t \mid O_{1,} O_2, \dots, O_{t-1})$ Prediction

$corr(X_t) = P(X_t \mid O_{1,} O_2, \dots, O_t)$ Correction (update)

$pred(X_1) = P(X_1)$

$corr(X_1) = P(X_1 \mid O_1)$

Ax

Kalman Filter

K. N. Toosi University of Technology

Recursive Bayesian Filtering

- $P(X_t \mid X_{t-1})$ Transition Model
- $P(O_t \mid X_t)$ Observation Model

Simple Case:

- Continuous States
- Linear Transition and Observation Models
- Additive Gaussian Noise

Kalman Filter

K. N. Toosi
University of Technology

- Continuous States
- Linear Transition and Observation Models
  - $X_t = \mathbf{A} X_{t-1} + \varepsilon_x$
  - $O_t = \mathbf{B} X_t + \varepsilon_o$
- Additive Gaussian Noise
  - $\varepsilon_x \sim N(0, \Sigma_x)$
  - $\varepsilon_o \sim N(0, \Sigma_o)$

Kalman Filter

- Linear Transition and Observation Models
  - $X_t = \mathbf{A} X_{t-1} + \varepsilon_x$, $O_t = \mathbf{B} X_t + \varepsilon_o$
- Additive Gaussian Noise
  - $\varepsilon_x \sim N(0, \Sigma_x)$, $\varepsilon_o \sim N(0, \Sigma_o)$

$$P(X_t \mid X_{t-1}) = P_{\varepsilon x}(\mathbf{A} X_{t-1} - X_t) = \text{Normal}(\mathbf{A} X_{t-1} - X_t; 0, \Sigma_x)$$

$$P(O_t \mid X_t) = P_{\varepsilon o}(O_t - \mathbf{B} X_t) = \text{Normal}(O_t - \mathbf{B} X_t; 0, \Sigma_o)$$

Kalman Filter

K. N. Toosi
University of Technology

- Linear Transition and Observation Models
  - $X_t = \mathbf{A}\, X_{t-1} + \varepsilon_x$, $\quad O_t = \mathbf{B}\, X_t + \varepsilon_o$
- Additive Gaussian Noise
  - $\varepsilon_x \sim N(0, \Sigma_x)$, $\varepsilon_o \sim N(0, \Sigma_o)$

$$P(X_t \mid X_{t-1}) = P_{\varepsilon x}(\mathbf{A}\, X_{t-1} - X_t) = \text{Normal}(\mathbf{A}\, X_{t-1} - X_t; 0, \Sigma_x)$$

$$P(O_t \mid X_t) = P_{\varepsilon o}(O_t - \mathbf{B}\, X_t) = \text{Normal}(O_t - \mathbf{B}\, X_t; 0, \Sigma_o)$$

pred(X_t) and corr(X_t) remain Gaussian.

Kalman Filter

K. N. Toosi University of Technology

- Linear Transition and Observation Models
  - $X_t = \mathbf{A}\, X_{t-1} + \varepsilon_x \;, \qquad O_t = \mathbf{B}\, X_t + \varepsilon_o$
- Additive Gaussian Noise
  - $\varepsilon_x \sim N(0, \Sigma_x), \;\; \varepsilon_o \sim N(0, \Sigma_o)$

$$P(X_t \mid X_{t-1}) = P_{\varepsilon x}(\mathbf{A}\, X_{t-1} - X_t) = \mathrm{Normal}(\mathbf{A}\, X_{t-1} - X_t; 0, \Sigma_x)$$

$$P(O_t \mid X_t) = P_{\varepsilon o}(\, O_t - \mathbf{B}\, X_t) = \mathrm{Normal}(O_t - \mathbf{B}\, X_t; 0, \Sigma_o)$$

Recursive Bayesian Filter → Kalman Filter

Remember: Gaussian Distribution

K. N. Toosi
University of Technology

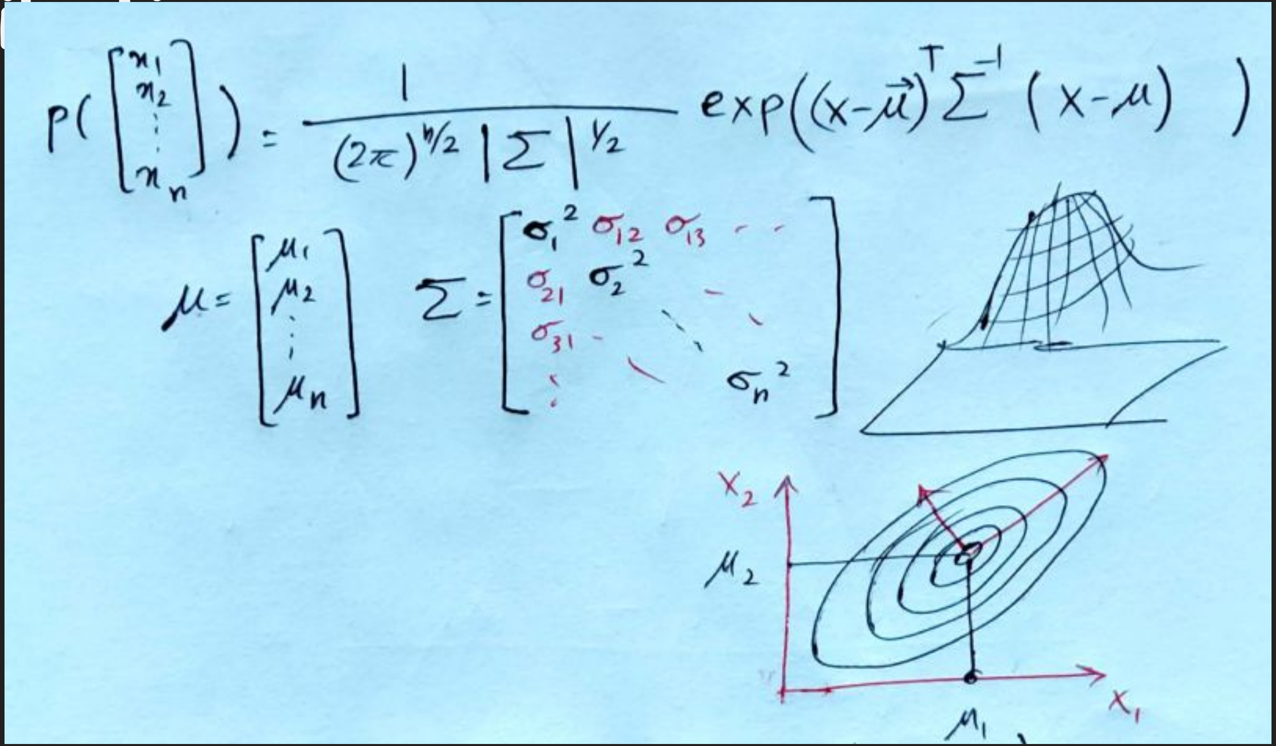
pg

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{(x-\mu)^2}{\sigma^2} \right)$$

$p(x)$ can be represented with μ, σ

Remember: Multivariate Gaussian Distri

K. N. Toosi
University of Technology



$p(\boldsymbol{x})$ can be represented with $\boldsymbol{\mu}, \boldsymbol{\Sigma}$

Kalman Filter

$$p(X_t \mid X_{t-1}) = \text{Normal}\left(X_t - AX_{t-1} ; \emptyset , \Sigma_x\right)$$

$$= \frac{1}{(2\pi)^{n/2} \, |\Sigma_x|^{1/2}} \exp\left((X_t - AX_{t-1})^T \Sigma_x^{-1} (X_t - AX_{t-1}) \right)$$

observation model

$$p(O_t \mid X_t) = \text{Normal}\left(O_t - BX_t , \emptyset , \Sigma_o\right)$$

$$= \frac{1}{(2\pi)^{n/2} \, |\Sigma_o|^{1/2}} \exp\left((O_t - BX_t)^T \Sigma_o^{-1} (O_t - BX_t) \right)$$

Kalman Filters