

Probabilistic Graphical Models

Lectures 17

Inference in Temporal Models
Recursive Bayesian Filtering
Intro to Kalman Filter



Remember: Temporal Models

Actions



...



States



...



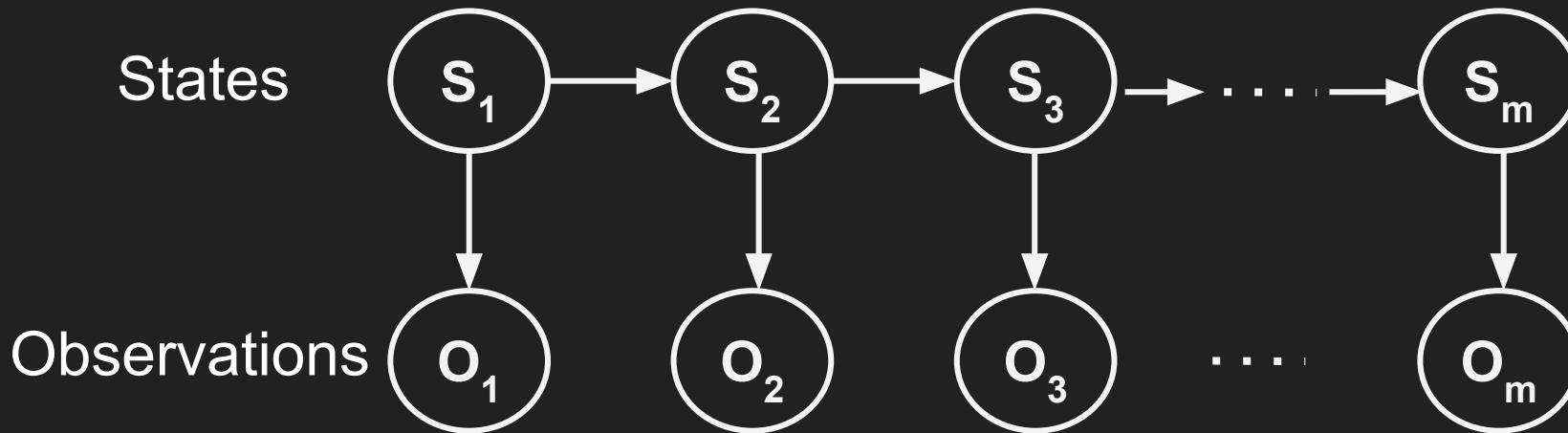
Observations



...



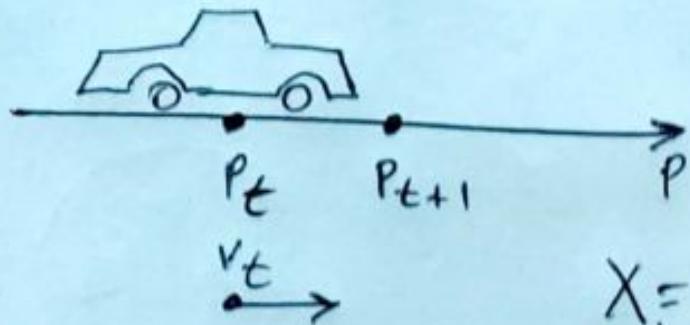
Remember: Temporal Models





Remember: Temporal Models

pgm17 ①



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

$$\varepsilon_x \sim \mathcal{N}(0, P_{\varepsilon_x})$$

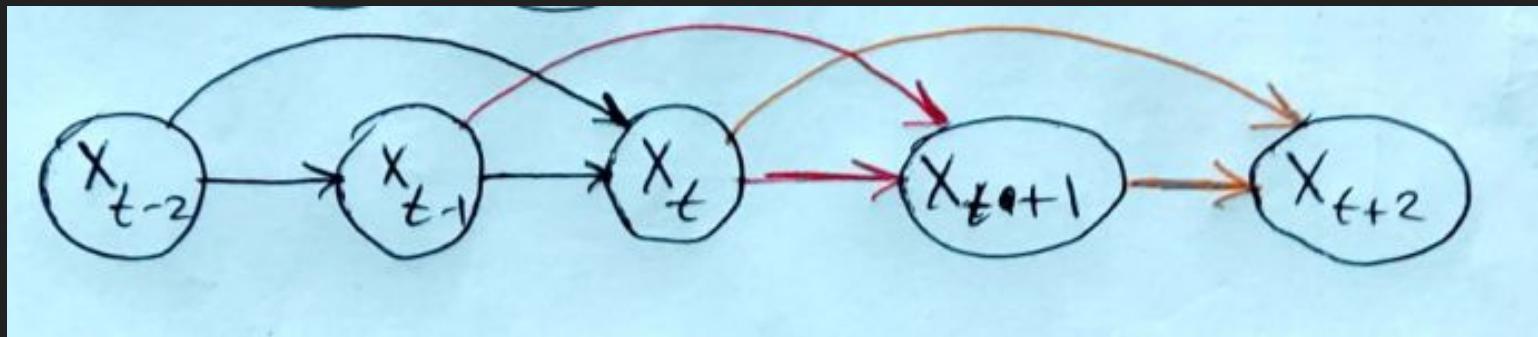
$$X_t = A X_{t-1} + \varepsilon_x$$

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



What if ε_v cannot model approximation?

Solution 1: create more dependencies.



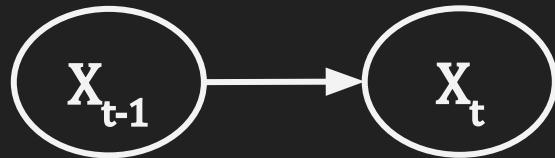


What if ε_v cannot model approximation?

Solution 2: add acceleration to state vector.

$$\vec{x}_t = \begin{bmatrix} p_t \\ v_t \end{bmatrix} \Rightarrow \vec{x}_t = \begin{bmatrix} p_t \\ v_t \\ a_t \end{bmatrix}$$
$$\vec{x}_t = \begin{bmatrix} p_t \\ v_t \\ a_t \end{bmatrix} = \begin{bmatrix} p_{t-1} + \Delta t v_{t-1} + \frac{1}{2} \Delta t^2 a_{t-1} + \varepsilon_p \\ v_{t-1} + \Delta t a_{t-1} + \varepsilon_v \\ a_{t-1} + \varepsilon_a \end{bmatrix}$$
$$\begin{bmatrix} p_t \\ v_t \\ a_t \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ v_{t-1} \\ a_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_p \\ \varepsilon_v \\ \varepsilon_a \end{bmatrix}$$
$$\vec{x}_t = A \vec{x}_{t-1} + \vec{\varepsilon}_x$$

Remember: Temporal Models (nonlinear case)



Linear:

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

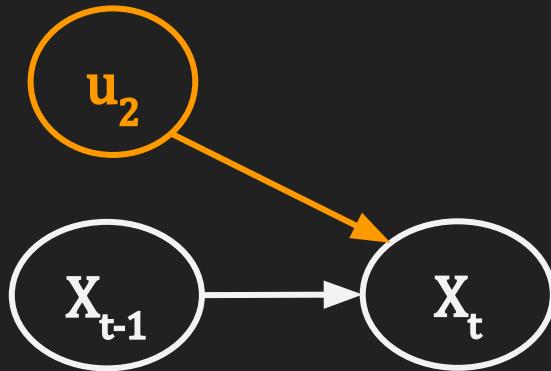
Nonlinear:

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

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



Remember: Temporal Models (nonlinear case with input)



Linear:

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

Nonlinear:

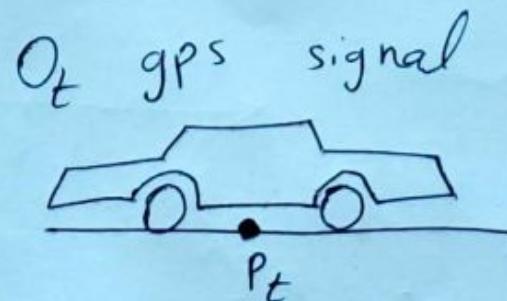
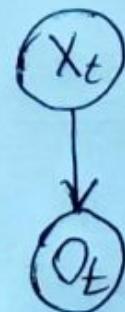
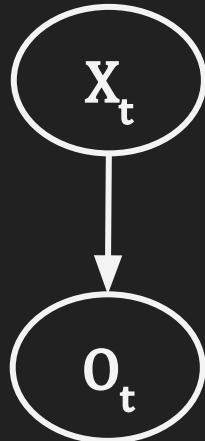
$$X_t = f(X_{t-1}, u_t) + \varepsilon$$

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

Remember: Observation Model Example



K. N. Toosi
University of Technology



pgm 17

$$O_t = g_t = P_t + \varepsilon_g = \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_B \begin{bmatrix} P_t \\ v_t \\ a_t \end{bmatrix} + \varepsilon_g$$

gps noise

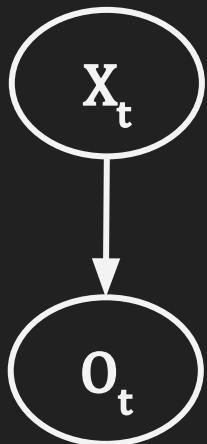
gps

The equation shows the observation O_t as a function of position P_t , velocity v_t , and acceleration a_t , plus GPS noise ε_g . The matrix B is shown underlined, and arrows point from the g_t term to both P_t and the noise term, with the label "gps noise" pointing to the noise term.

$$O_t = B X_t + \varepsilon_o$$

$$P(O_t | X_t) = P_{\varepsilon_o}(O_t - BX_t)$$

Remember: Observation Model Example



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

Handwritten notes in blue:

- An arrow points from "GPS" to the first element g_t of the observation vector.
- An arrow points from "acc" to the second element a'_t of the observation vector.
- An arrow points from "نحوه مدل" (Observation Model) to the equation $O_t = B X_t + \varepsilon_o$.
- An arrow points from "نحوه مدل" (Observation Model) to the error term $\begin{bmatrix} e_g \\ \varepsilon_a' \end{bmatrix}$.

$$O_t = B X_t + \varepsilon_o$$

Remember: Transition and Observation Models

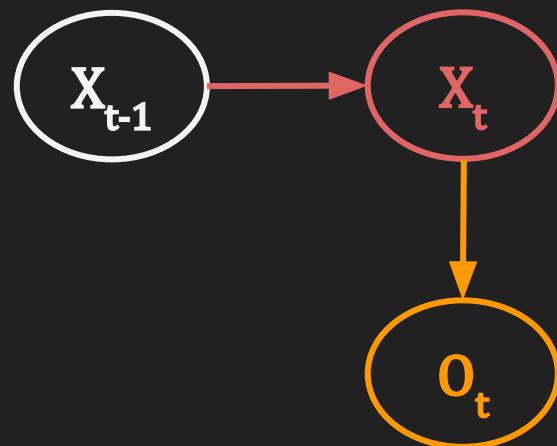
(No input)

$$P(X_t | X_{t-1})$$

Transition Model

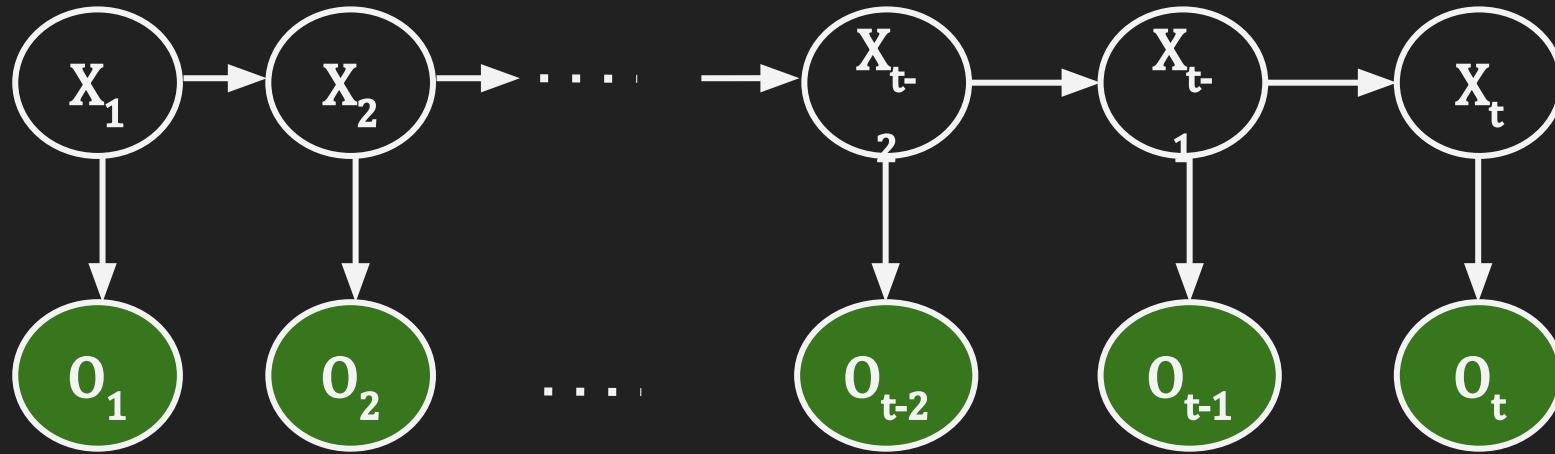
$$P(O_t | X_t)$$

Observation Model





Inference: State Estimation

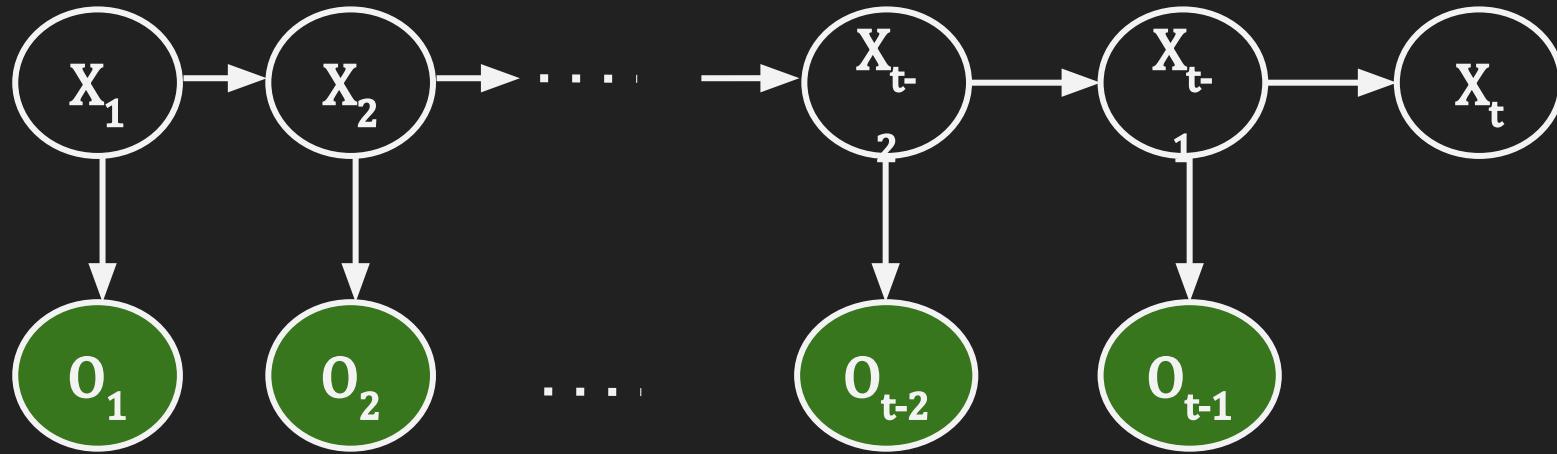


Given o_1, o_2, \dots, o_t what is X_t ?

$$P(X_t | o_1, o_2, \dots, o_t) = ?$$



Inference: Prediction

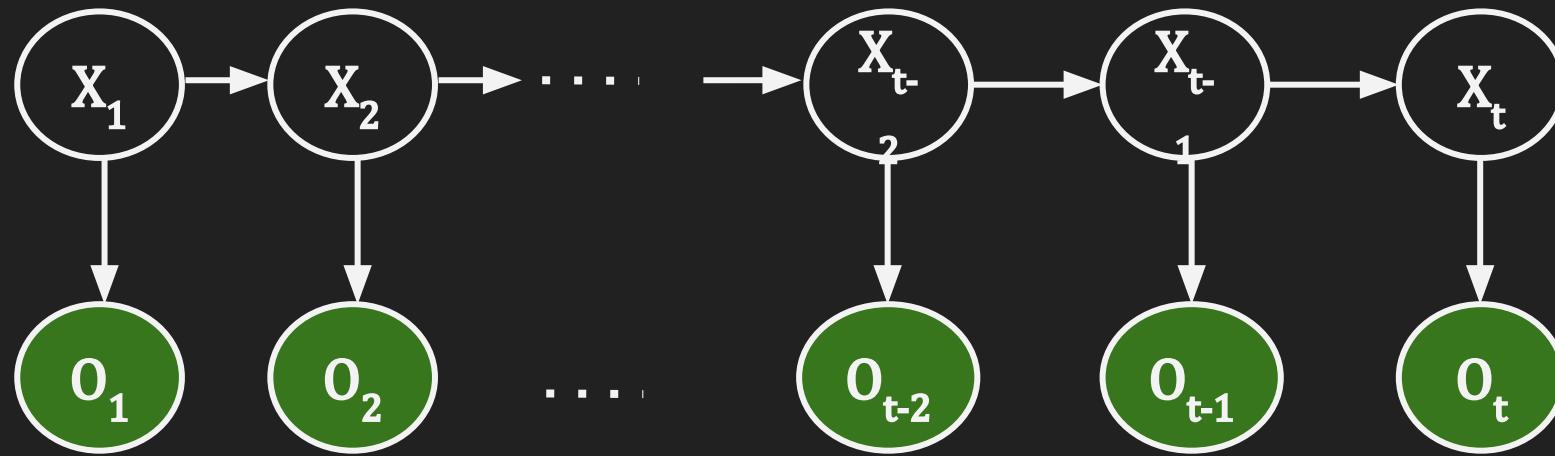


Given o_1, o_2, \dots, o_{t-1} what is X_t ?

$$P(X_t | o_1, o_2, \dots, o_{t-1}) = ?$$



Inference Problems



$\text{pred}(X_t) = P(X_t | O_1, O_2, \dots, O_{t-1})$ Prediction

$\text{corr}(X_t) = P(X_t | O_1, O_2, \dots, O_t)$ Correction (update)



Inference Problems

$$\text{pred}(X_t) = P(X_t | O_1, O_2, \dots, O_{t-1}) \quad \text{Prediction}$$

$$\text{corr}(X_t) = P(X_t | O_1, O_2, \dots, O_t) \quad \text{Correction (update)}$$

1. Variable Elimination

$$\begin{aligned} P(X_t | O_1, \dots, O_{t-1}) &= \frac{P(X_t, O_1, \dots, O_{t-1})}{\sum_{X'_t} P(X'_t, O_1, \dots, O_{t-1})} \\ &= \frac{\sum_{X_{t-1}} \sum_{X_2} \sum_{X_1} P(X_1, \dots, X_t, O_1, \dots, O_{t-1})}{\circ} \\ &= \underbrace{\sum_{X_{t-1}} \sum_{X_2} \sum_{X_1} P(X_2 | X_1) P(X_3 | X_2) \dots P(X_t | X_{t-1}) P(O_1 | X_1) \dots P(O_{t-1} | X_{t-1})}_{\circ} \end{aligned}$$



Inference Problems

$$\text{pred}(X_t) = P(X_t | O_1, O_2, \dots, O_{t-1}) \quad \text{Prediction}$$

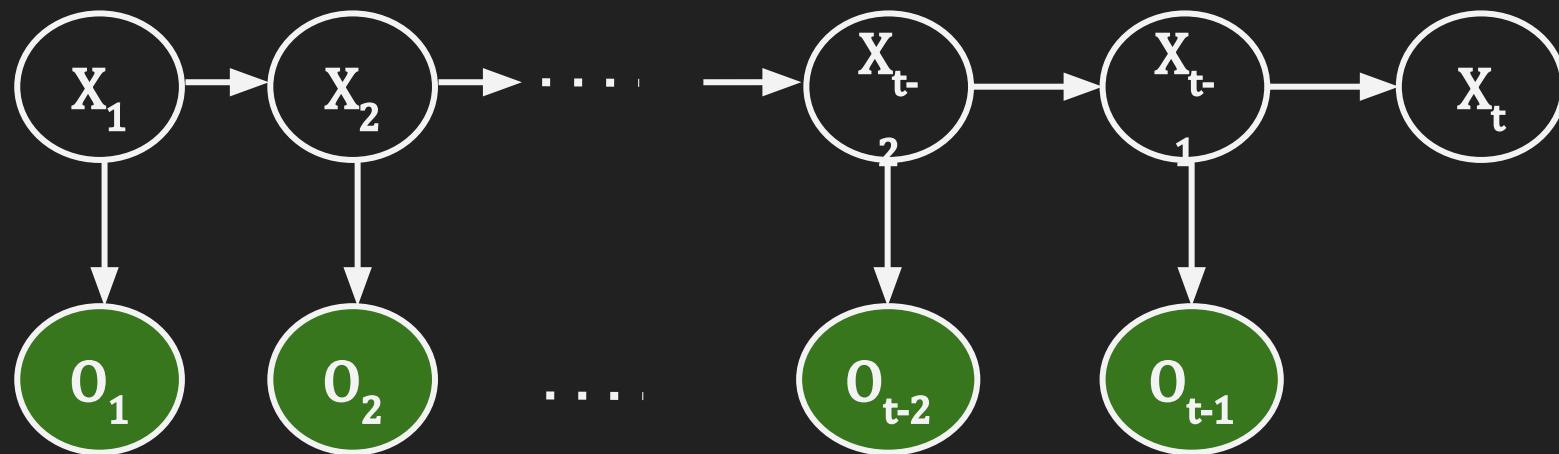
$$\text{corr}(X_t) = P(X_t | O_1, O_2, \dots, O_t) \quad \text{Correction (update)}$$

1. Variable Elimination
2. Recursive Bayesian Filtering

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



Prediction Phase



$$\text{pred}(X_t) = P(X_t | O_1, O_2, \dots, O_{t-1}) =$$



Recursive Bayesian Filtering - Prediction

K. N. Toosi
University of Technology

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

pgm 17

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

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

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

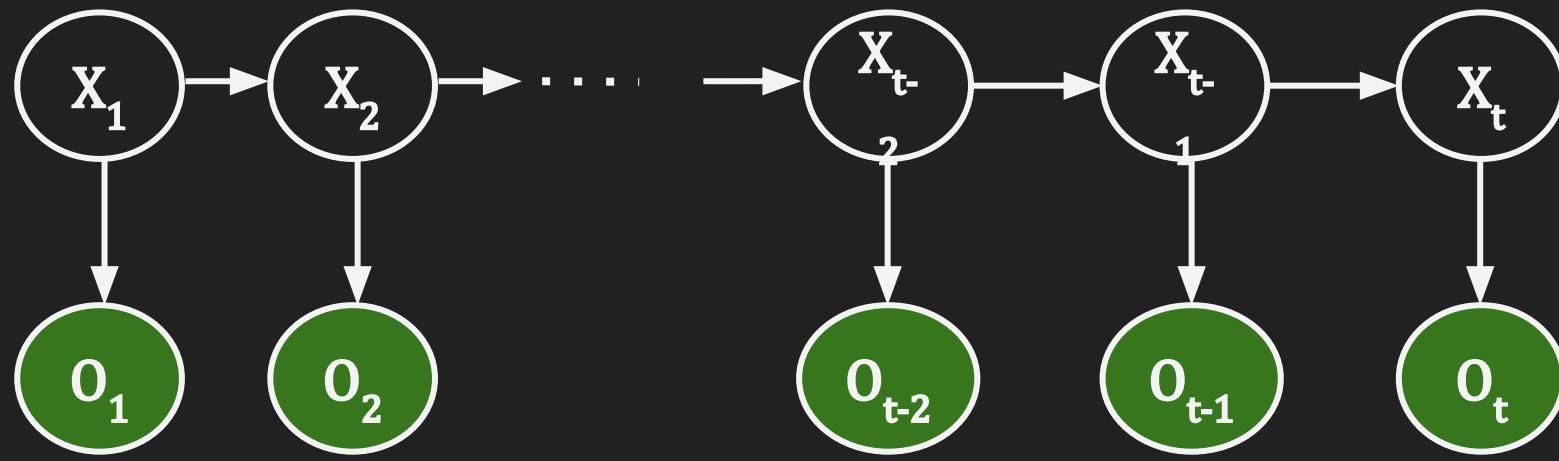
$$= \sum_{x_{t-1}} P(x_t | x_{t-1}) \text{corr}(x_{t-1})$$

continuous

$$\text{pred}(x_t) = \int P(x_t | x_{t-1}) \text{corr}(x_{t-1}) dx_{t-1}$$



Correction Phase



$$\text{corr}(X_t) = P(X_t | O_1, O_2, \dots, O_t)$$

Recursive Bayesian Filtering - Correction



$$\begin{aligned}
 \text{corr}(X_t) &= P(X_t | o_1 - o_t) = P(X_t | \underbrace{o_1, \dots, o_{t-1}}_{\text{blue}}, o_t) \\
 &= \frac{P(o_t | X_t, o_1 - o_{t-1}) P(X_t | o_1 - o_{t-1})}{P(o_t | o_1 - o_{t-1})} \\
 &= \frac{P(o_t | X_t, o_1 - o_{t-1}) P(X_t | o_1 - o_{t-1})}{\sum_{X'_t} P(o_t, X'_t | o_1 - o_{t-1})} \\
 &= \frac{P(o_t | X_t, o_1 - o_{t-1}) P(X_t | o_1 - o_{t-1})}{\sum_{X'_t} P(o_t | X'_t, o_1 - o_{t-1}) P(X'_t | o_1 - o_{t-1})} \\
 &= \frac{P(o_t | \cancel{X_t}) \text{ pred}(X_t)}{\sum_{X'_t} P(o_t | X'_t) \text{ pred}(X'_t)}
 \end{aligned}$$



Initialization Phase

$$\text{pred}(X_t) = P(X_t | O_1, O_2, \dots, O_{t-1}) \quad \text{Prediction}$$

$$\text{corr}(X_t) = P(X_t | O_1, O_2, \dots, O_t) \quad \text{Correction (update)}$$

$$\text{pred}(X_1) = P(X_1)$$

$$\text{corr}(X_1) = P(X_1 | O_1)$$



Kalman Filter

Recursive Bayesian Filtering

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

Simple Case:

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



Kalman Filter

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



Kalman Filter

- Linear Transition and Observation Models

- $X_t = AX_{t-1} + \varepsilon_x , \quad O_t = BX_t + \varepsilon_o$

- Additive Gaussian Noise

- $\varepsilon_x \sim N(0, \Sigma_x), \quad \varepsilon_o \sim N(0, \Sigma_o)$

$$P(X_t | X_{t-1}) = P_{\varepsilon_X}(AX_{t-1} - X_t) = \text{Normal}(AX_{t-1} - X_t; 0, \Sigma_x)$$

$$P(O_t | X_t) = P_{\varepsilon_O}(O_t - BX_t) = \text{Normal}(O_t - BX_t; 0, \Sigma_o)$$



Kalman Filter

- 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

- $\varepsilon_x \sim N(0, \Sigma_x), \quad \varepsilon_o \sim N(0, \Sigma_o)$

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

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

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



Kalman Filter

- 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

- $\varepsilon_x \sim N(0, \Sigma_x)$, $\varepsilon_o \sim N(0, \Sigma_o)$

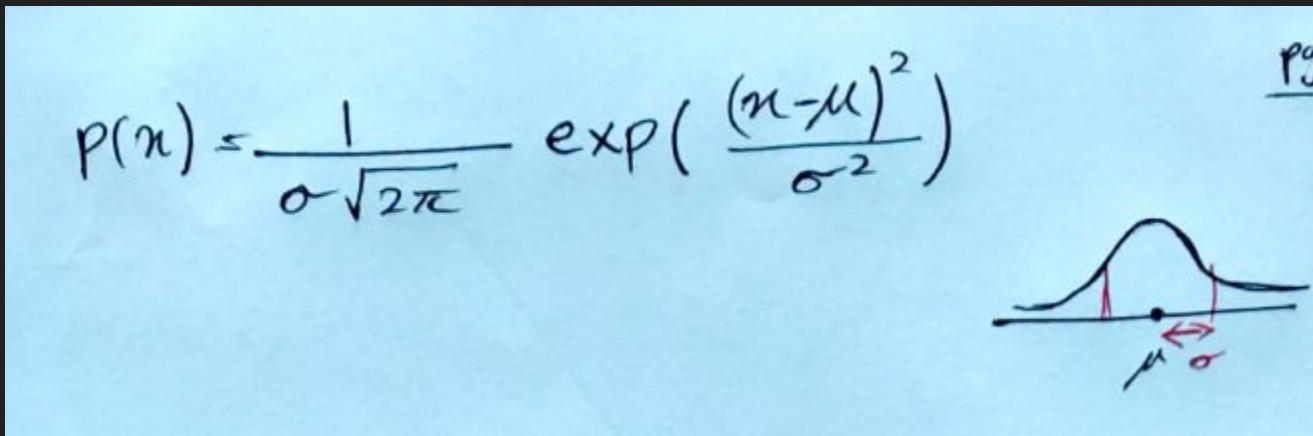
$$P(X_t | X_{t-1}) = P_{\varepsilon x}(\mathbf{A} X_{t-1} - X_t) = \text{Normal}(\mathbf{A} X_{t-1} - X_t; 0, \Sigma_x)$$

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

Recursive Bayesian Filter → Kalman Filter



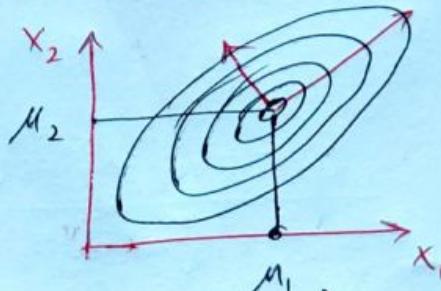
Remember: Gaussian Distribution



$p(x)$ can be represented with μ, σ



Remember: Multivariate Gaussian Distribution

$$p\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$
$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots \\ \sigma_{21} & \sigma_2^2 & & \\ \sigma_{31} & & \ddots & \\ \vdots & & & \sigma_n^2 \end{bmatrix}$$


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



Kalman Filter

$$p(x_t | X_{t-1}) = \text{Normal}(x_t - Ax_{t-1}; \emptyset, \Sigma_x)$$
$$= \frac{1}{(2\pi)^{n/2} |\Sigma_x|^{1/2}} \exp\left((x_t - Ax_{t-1})^T \Sigma_x^{-1} (x_t - Ax_{t-1})\right)$$

observation model

$$p(o_t | X_t) = \text{Normal}(o_t - Bx_t, \emptyset, \Sigma_o)$$
$$= \frac{1}{(2\pi)^{n/2} |\Sigma_o|^{1/2}} \exp\left((o_t - Bx_t)^T \Sigma_o^{-1} (o_t - Bx_t)\right)$$



Kalman Filters

vector

$$\text{Pred}(X_t) = \int_{X_{t-1}}^{\text{Gaussian}} p(X_t | X_{t-1}) \cancel{\text{corr}_{t-1}(X_{t-1})} dX_t$$
$$\mu_p^t, \Sigma_p^t \xleftarrow[\text{Gaussian}]{\text{Corr}} \text{Corr}(X_t) = \frac{p(O_t | X_t) \text{Pred}(X_t)}{\sum_{X_t}} \rightarrow \mu_p^t, \Sigma_p^t$$
$$\mu_c^t, \Sigma_c^t \xleftarrow[\text{Gaussian}]{\text{Corr}} \text{Corr}(X_t)$$