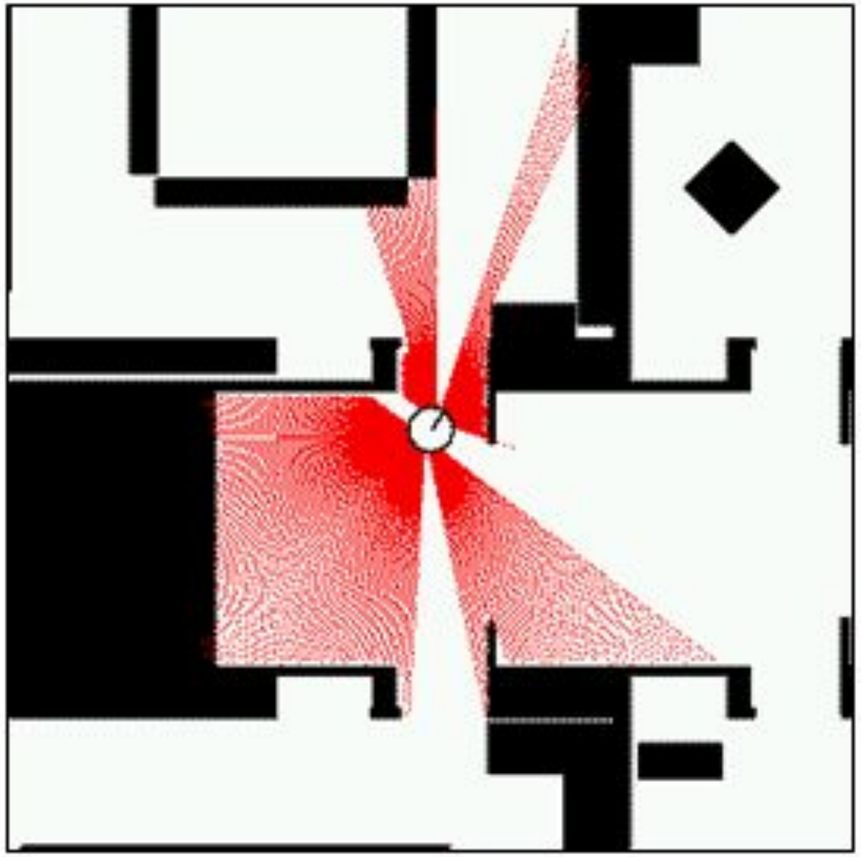
2.160 Identification, Estimation, and Learning

Part 2 Estimation



Lecture 10

Bayes Filter

H. Harry Asada
Department of Mechanical Engineering
MIT

Recap

Kalman Filter and Extended Kalman Filter

Given a State Equation: $$x_t = A_{t-1}x_{t-1} + B_{t-1}u_{t-1} + w_{t-1}$$

$$x_t = f(x_{t-1}, u_{t-1}, t) + w_{t-1}$$

$$\hat{x}_{t|t-1} = A_{t-1}\hat{x}_{t-1} + B_{t-1}u_{t-1}$$

$$\hat{x}_{t|t-1} = f(\hat{x}_{t-1}, u_{t-1}, t)$$

Simulation **Propagate**

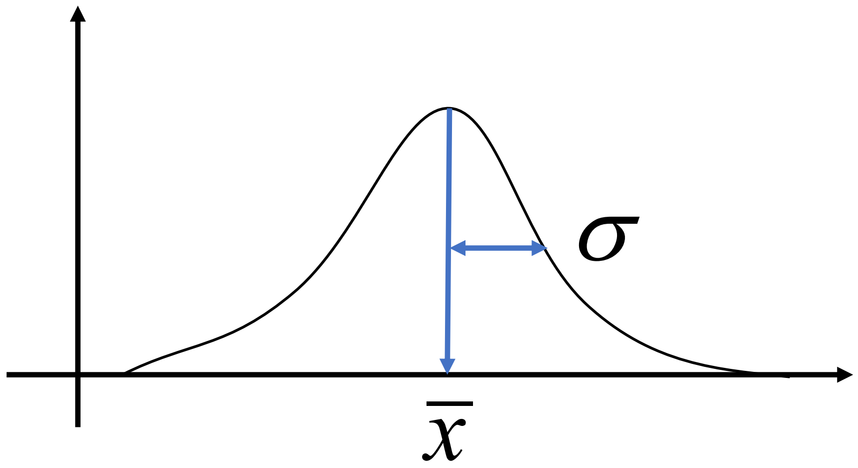
Update Measurement

$$\hat{x}_t = \hat{x}_{t|t-1} + K_t[y_t - \hat{y}_t]$$

$$\hat{x}_t = \hat{x}_{t|t-1} + K_t[y_t - h(\hat{x}_{t|t-1}, t)]$$

- KF, EKF, and UKF provide a particular value of the state as estimate.

- Giving a particular value as estimate makes sense when the state distribution is Gaussian or unimodal.

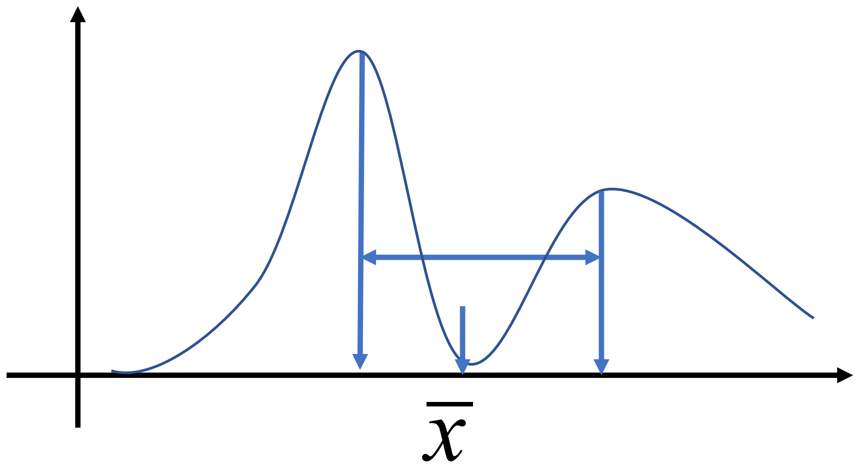


Gaussian Distribution

Mean $\bar{x}$ and standard deviation σ

Mean represents the estimated value very well, with Standard Deviation being Accuracy of prediction.

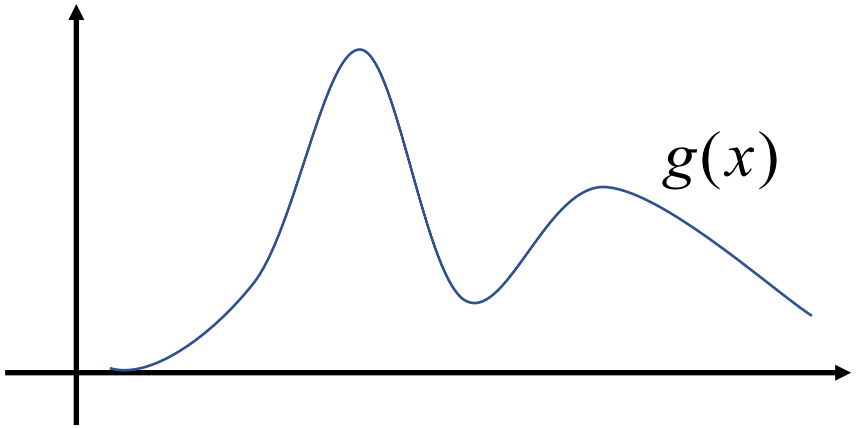
- However, if the distribution is not Gaussian and multimodal, the single value is a poor representation.



Non-Gaussian Distribution

A mean value does not represent the overall distribution of the random variables.
The mean is the least likely value in this example.

Belief



Non-Gaussian Distribution

A mean value does not represent the overall distribution of the random variables.

Belief $g(x)$: the entire pdf distribution rather than a single value.

Bayes Filter predicts the pdf distribution of a random variable.

$$\hat{x}_{t-1} \quad \mapsto \quad \hat{x}_{t|t-1} \quad \mapsto \quad \hat{x}_t \qquad \text{Kalman Filter}$$

$$g_{t-1}(x_{t-1}) \quad \mapsto \quad g_{t|t-1}(x_t) \quad \mapsto \quad g_{t|t}(x_t) \qquad \text{Bayes Filter}$$

Propagate Update

at time t-1 a priori estimate. a posteriori estimate

Markov Process

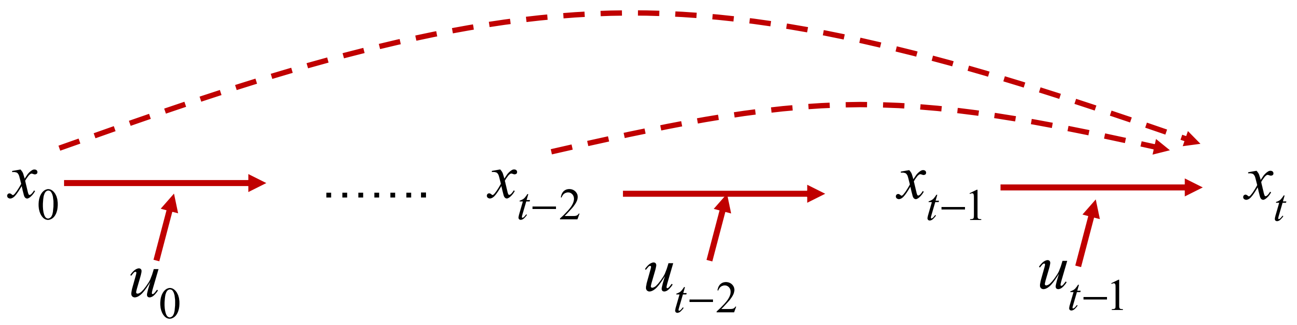
Discrete-time stochastic state transition, in general:

$$\Pr(x_t \mid x_0, \cdots, x_{t-1}, u_0, \cdots, u_{t-1})$$

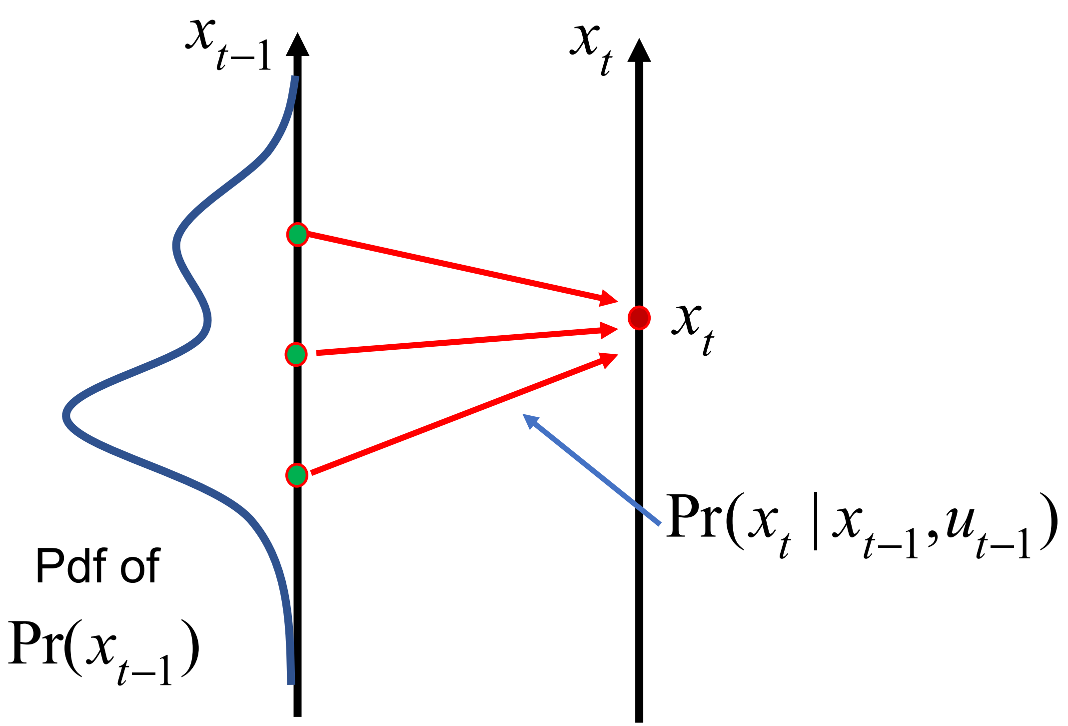
The probability of random variable $X_t = x_t$, given previous states and inputs $x_0, \cdots, x_{t-1}, u_0, \cdots, u_{t-1}$.

A special case where the probability of $X_t = x_t$ depends only on x_{t-1}, u_{t-1}, the process is called a Markov Process.

$$\Pr(x_t \mid x_{t-1}, u_{t-1})$$



Chapman-Kolmogorov Equation



State $X_t = x_t$ can be reached from x_{t-1} with conditional probability density of

$$\Pr(x_t \mid x_{t-1}, u_{t-1})$$

where $-\infty < x_{t-1} < \infty$ has a pdf of $\Pr(x_{t-1})$.

Therefore, the pdf of $\Pr(x_t)$ is given by

$$\Pr(x_t) = \int_{-\infty}^{\infty} \Pr(x_t \mid x_{t-1}, u_{t-1}) \Pr(x_{t-1})\, dx_{t-1}$$

This is called the Chapman-Kolmogorov Equation.

State Propagation Law

- In our Bayes Filter problem, we want to recursively estimate a priori belief $g_{t|t-1}(x_t)$ from a posterior belief at time $t-1$, $g_{t-1}(x_{t-1})$
- Given a state transition equation

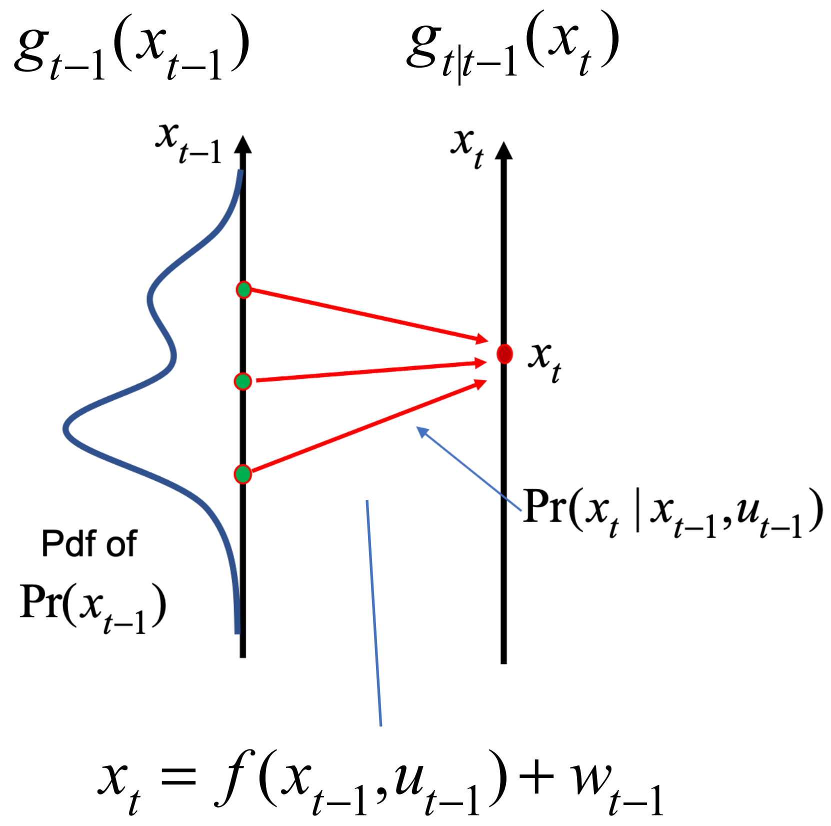
$$x_t = f(x_{t-1}, u_{t-1}) + w_{t-1}$$

Additive noise

- Applying the Chapman-Kolmogorov Equation,

$$g_{t|t-1}(x_t) = \int_{-\infty}^{\infty} \underline{\Pr(x_t \mid x_{t-1}, u_{t-1})} g_{t-1}(x_{t-1})\, dx_{t-1}$$

How can we find this probability?

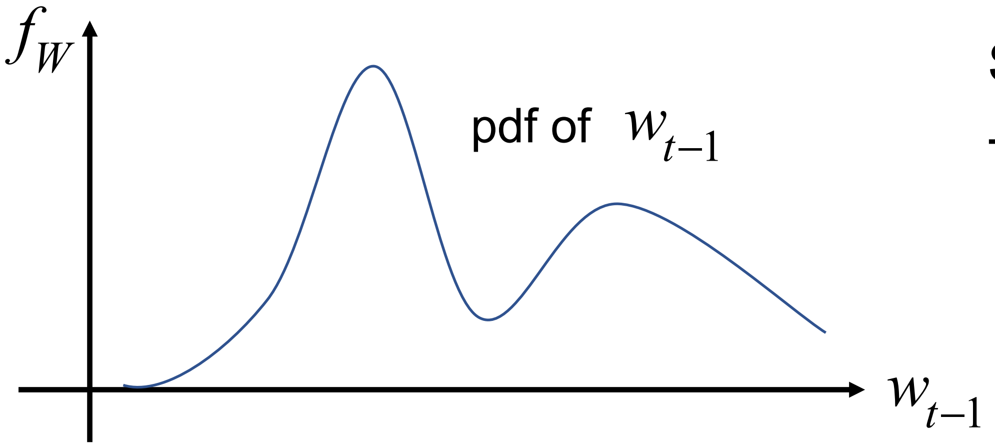


State Propagation Law (Continued)

- The State Transition Equation

$$x_t = \underbrace{f(x_{t-1}, u_{t-1})}_{\text{Deterministic}} + \underbrace{w_{t-1}}_{\text{Random}}$$

- Given x_{t-1} and u_{t-1} , the randomness of x_t comes from process noise $w_{t-1} \sim f_W(w_{t-1})$



Since $w_{t-1} = x_t - f(x_{t-1}, u_{t-1})$

The state transition pdf is given by

$$\Pr(x_t \mid x_{t-1}, u_{t-1}) = f_W(x_t - f(x_{t-1}, u_{t-1}))$$

Back to the Chapman-Kolmogorov equation:

$$g_{t|t-1}(x_t) = \int_{-\infty}^{\infty} f_W(x_t - f(x_{t-1}, u_{t-1}))\, g_{t-1}(x_{t-1})\, dx_{t-1}$$

a priori belief ….. Belief Propagation Law

Where are we ?

State transition with process noise

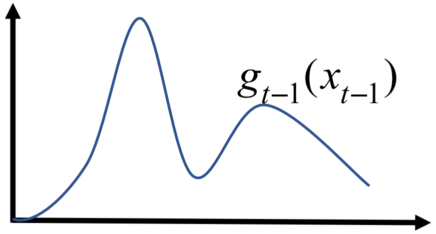
a posteriori belief $g_{t-1}(x_{t-1})$ $\longrightarrow$ $g_{t|t-1}(x_t)$ a priori belief

$t = t + 1$

$g_t(x_t)$

Assimilating a new observation y_t with measurement noise

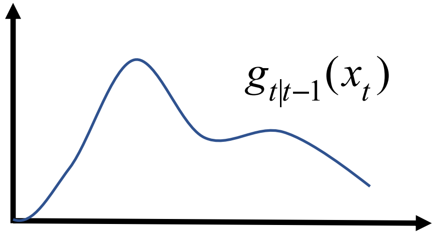
Where are we ?



Belief Propagation

State transition with process noise

a posteriori belief $g_{t-1}(x_{t-1})$ $\longrightarrow$ $g_{t|t-1}(x_t)$ a priori belief



t = *t* +1

$g_t(x_t)$

Assimilating a new observation y_t with measurement noise

Belief Update

Where are we ?

Belief Propagation

State transition with process noise

a posteriori belief $g_{t-1}(x_{t-1})$ $g_{t|t-1}(x_t)$ a priori belief

$t = t + 1$

$g_t(x_t)$

Assimilating a new observation y_t with measurement noise

Belief Update

Kalman Filter:

$$\hat{x}_t = \hat{x}_{t|t-1} + K_t[y_t - \hat{y}_t]$$

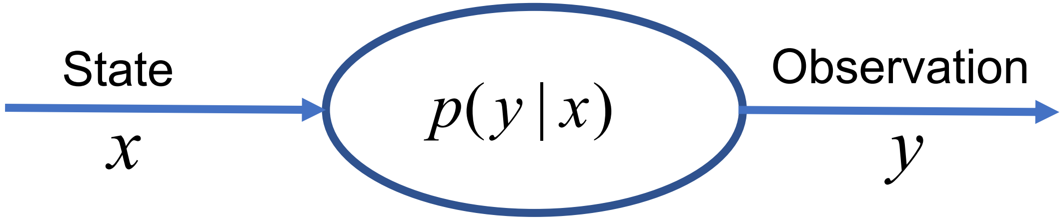
How do we construct this for Belief?

Bayes' Rule

❑ Joint probability density

$$p(x,y) = p(x\,|\,y)\,p(y) = p(y\,|\,x)\,p(x)$$

$$\therefore \quad p(x\,|\,y) = \frac{p(y\,|\,x)\,p(x)}{p(y)}$$



❑ Suppose that we know the conditional density $p(y\,|\,x)$
how can we estimate the state x from observation y: $p(x\,|\,y)$?

❑ Remark: $p(y) = \int_{-\infty}^{\infty} p(y\,|\,x)\,p(x)\,dx$

$$\left(1 = \int_{-\infty}^{\infty} p(x\,|\,y)\,dx = \int_{-\infty}^{\infty} \frac{p(y\,|\,x)\,p(x)}{p(y)}\,dx = \frac{1}{p(y)} \int_{-\infty}^{\infty} p(y\,|\,x)\,p(x)\,dx \right)$$

❑ Therefore, denominator $p(y)$ is a scaling factor.

$p(y)$ is nothing but a scaling factor that makes $\int_{-\infty}^{\infty} p(x \mid y)\, dx = 1$

- We do not need to know $p(y)$. So, let's replace it by a constant: $p(y) = \dfrac{1}{\eta}$

Then,

Measurement y

$$p(x \mid y) = \eta\, \underline{p(y \mid x)}\, \underline{p(x)}$$

a priori belief $g_{t|t-1}(x_t)$

Prior

a posteriori belief $g_t(x_t)$

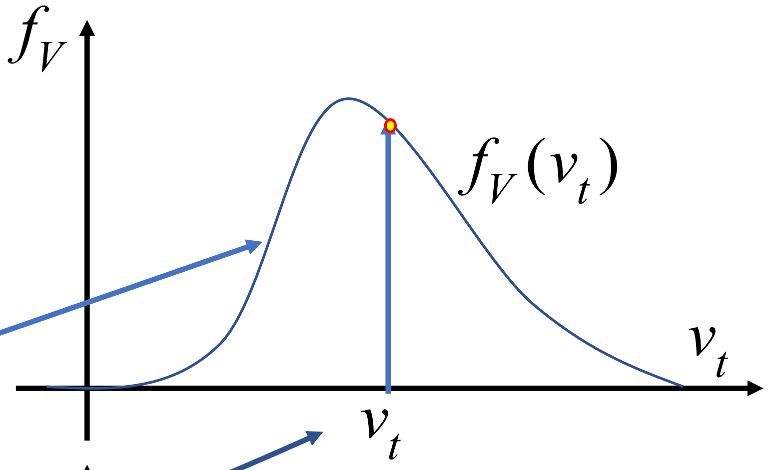
Generative Model
(Similar to Likelihood Function)

$p(y \mid x) \rightarrow f_V(y_t - h(x_t, t))$

- We can construct $p(y \mid x)$ from the measurement equation:

State Observation (Output function): $y_t = h(x_t, t) + v_t$

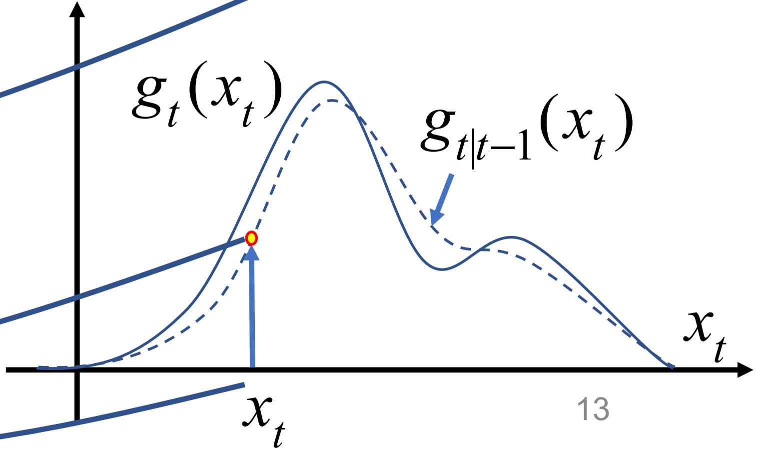
Deterministic   Random



$$\therefore \quad g_t(x_t) = \eta f_V(y_t - h(x_t, t))\, g_{t|t-1}(x_t)$$

Belief Update

Measurement y_t



The Bayes Filter Algorithm

1. Initial Conditions: $g_0(x_0)$ set $t = 1$;
2. Belief Propagation:

$$g_{t|t-1}(x_t) = \int_{-\infty}^{\infty} f_W(x_t - f(x_{t-1}, u_{t-1}))\, g_{t-1}(x_{t-1})\, dx_{t-1}$$

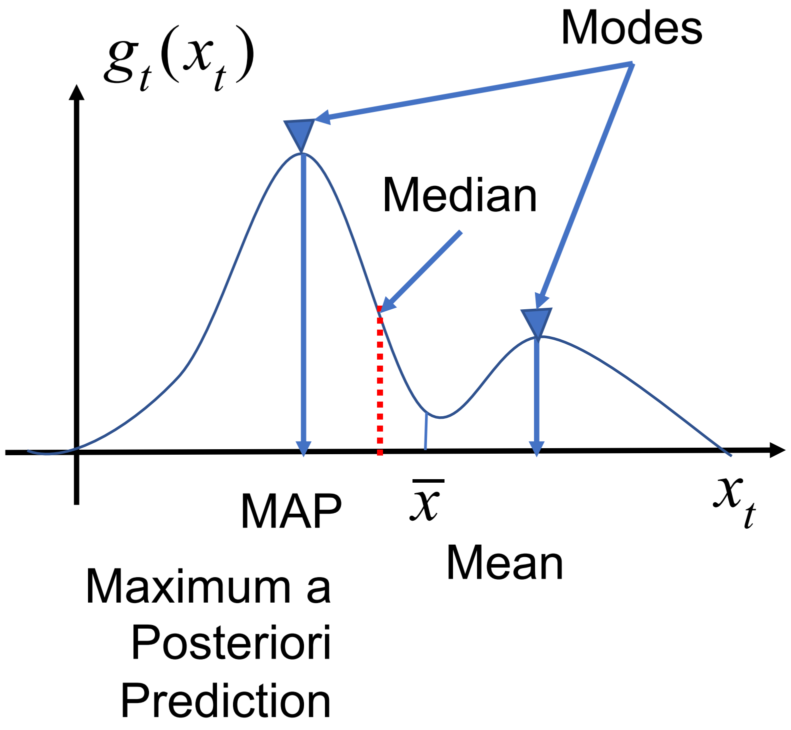
3. Assimilate y_t and update the a priori belief

$$g_t(x_t) = \eta f_V(y_t - h(x_t, t)) g_{t|t-1}(x_t)$$

4. Return $g_t(x_t)$. Set $t = t + 1$ and repeat.

❑ Interpretation of the belief $g_t(x_t)$

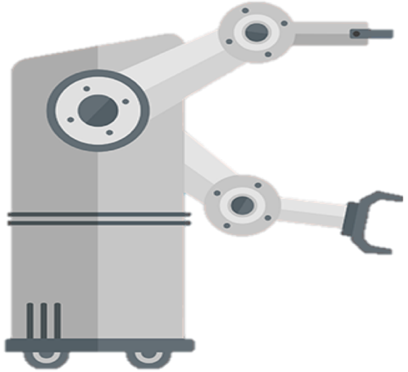
- Maximum a Posteriori Prediction (MAP)
- Modes (multiple)
- Medium
- Mean



Interpretation

Illustrative Example

A Robot at the Killian Court, MIT

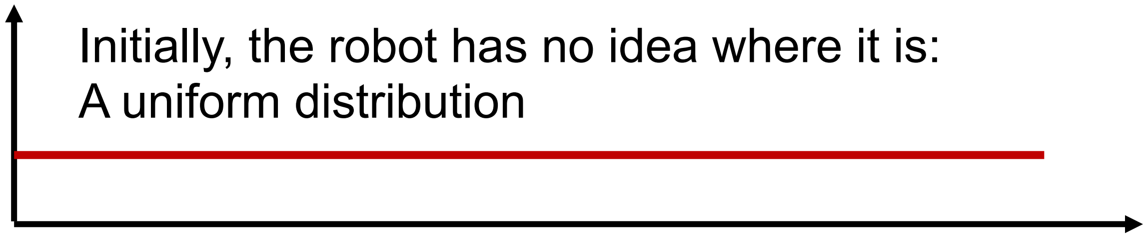


The robot has no idea where it is at the entrance of Building 10.
It has a LIDAR system to detect objects nearby, e.g. pillars.
Good thing, it has learned Bayes Filter, passing 2.160.

Illustrative Example

(A)

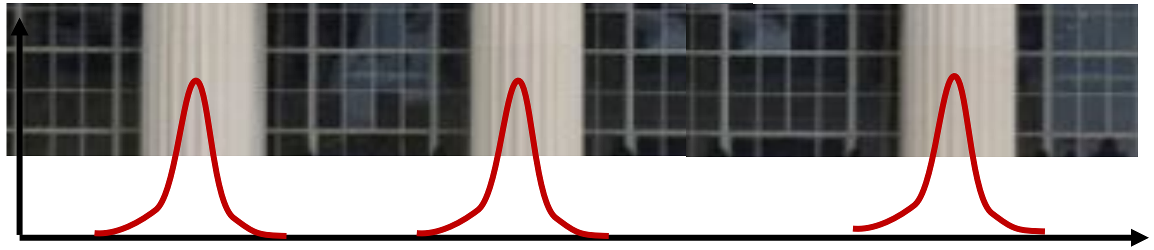
Initially, the robot has no idea where it is:
A uniform distribution



A pillar has just been detected, but the robot does not know which pillar it is.

Observation

(B)



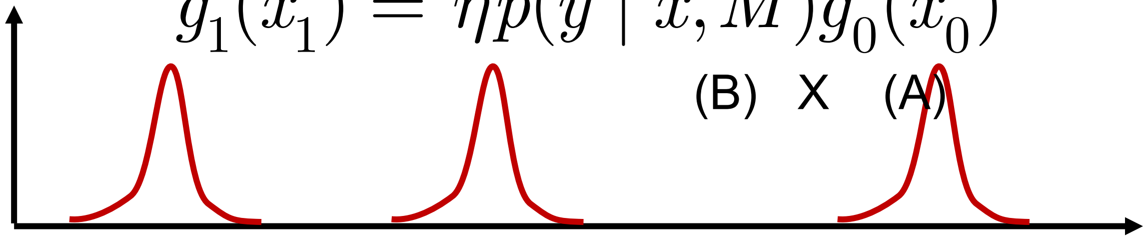
$$p(y \mid x, M)$$

Map landmarks

$$g_1(x_1) = \eta p(y \mid x, M) g_0(x_0)$$

(B) X (A)

(C)



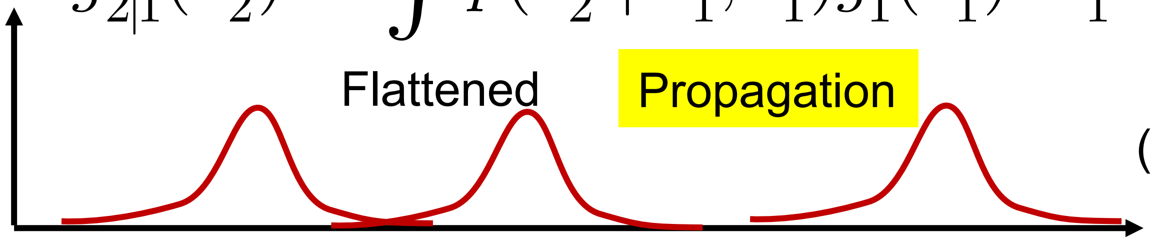
Update

$$g_{2|1}(x_2) = \int p(x_2 \mid x_1, u_1) g_1(x_1)\, dx_1$$

Flattened

Propagation

(D)

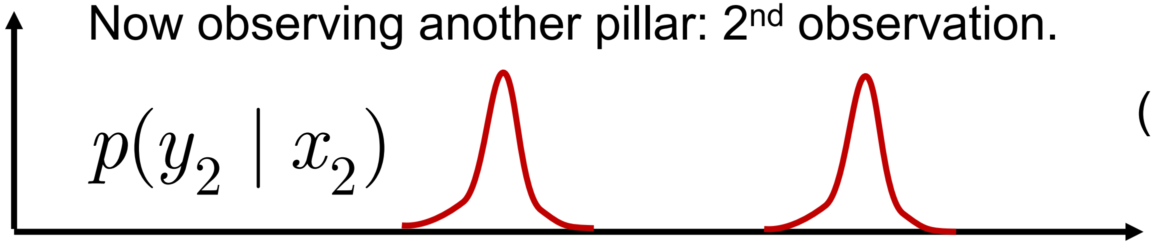


As the robot moves, the location estimate becomes more uncertain.

Observation

Now observing another pillar: 2nd observation.

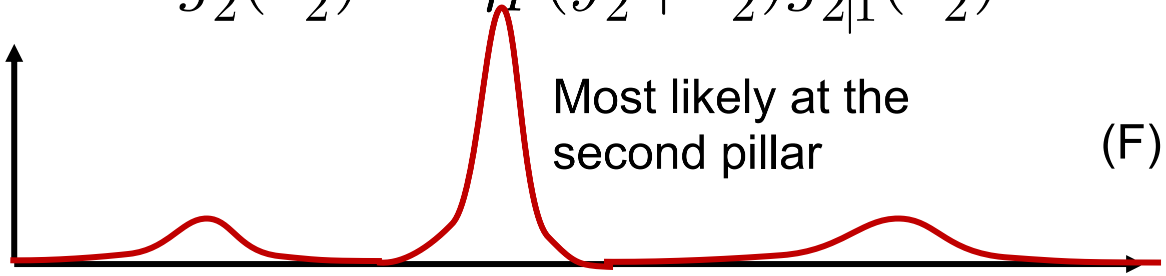
(E)



$$g_2(x_2) = \eta p(y_2 \mid x_2) g_{2|1}(x_2)$$

Most likely at the second pillar

(F)



Update

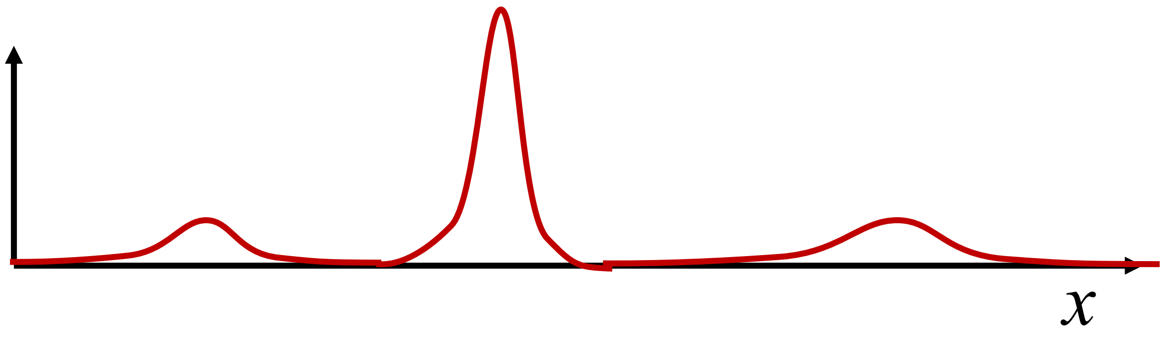
Bayes Filter

- ❑ Nonlinear dynamics, non-Gaussian distribution
- ❑ Use of “Belief”, pdf estimation rather than a single value estimation.
- ❑ Bimodal, skewed distribution of state can be treated.
- ❑ Multiple Hypothesis Tracking: all possible cases are tracked.
- ❑ Low probability cases, too, are not eliminated.

- ❑ Computationally expensive:

$$g_{t|t-1}(x_t) = \int_{-\infty}^{\infty} f_W(x_t - f(x_{t-1}, u_{t-1}))\, g_{t-1}(x_{t-1})\, dx_{t-1}$$

$$g_t(x_t) = \eta f_V(y_t - h(x_t, t)) g_{t|t-1}(x_t)$$



Most General:
Nonlinear dynamics;
Non-Gaussian noise

Bayes Filter

Assume:

- Gaussian noise
- Linear time varying system

Apply:

- Monte Carlo Approximation
- Importance sampling

Kalman Filter

Particle Filter

Kalman Filter can be derived from Bayes Filter;
Proof of Kalman Filter

Bayes Filter can be computed effectively with Particle Filter;
Implementation of Bayes Filter

8.4 Gaussian Kalman Filter (so-called Kalman Filter)

- Kalman Filter is a special case of Bayes Filter. We can derive Kalman Filter from Bayes Filter by making the following assumptions.
- Assume a Linear Time-Varying stochastic system:

$$x_{t+1} = A_t x_t + B_t u_t + G_t w_t$$
$$y_t = H_t x_t + v_t$$

Further assume for brevity $u_t \equiv 0, G_t = I$

- Assume white (uncorrelated) Gaussian noise:

$$w_t \sim \mathbb{N}(0, Q_t), \quad v_t \sim \mathbb{N}(0, R_t)$$

where

$$E[w_t w_s{}^T] = \begin{cases} Q_t; & t = s \\ 0; & t \neq s \end{cases}, \; E[v_t v_s{}^T] = \begin{cases} R_t; & t = s \\ 0; & t \neq s \end{cases}, \; E[w_t v_s{}^T] = 0, \forall t, \forall s$$

and Gaussian distribution

$$f_W = \frac{1}{\sqrt{\det(2\pi Q_t)}} \exp\left(-\frac{1}{2} w_t^T Q_t^{-1} w_t\right), \quad f_V = \frac{1}{\sqrt{\det(2\pi R_t)}} \exp\left(-\frac{1}{2} v_t^T R_t^{-1} v_t\right)$$

Deriving Kalman Filter from Bayes Filter

- Goal: Find an optimal state estimate that minimizes the mean squared prediction error conditioned by all the prior observations and inputs.

$$\hat{x}_t^o = \arg\min_{\hat{x}_t} E[|\hat{x}_t - x_t|^2 \,|\, y_1, \cdots, y_t, u_1, \cdots, u_{t-1}]$$

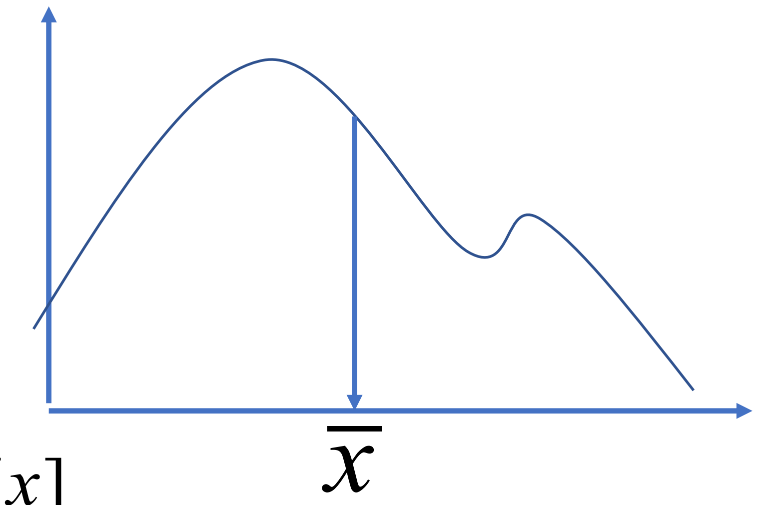
- This is equivalent to the conditional mean:

$$\hat{x}_t^o = E[\hat{x}_t \,|\, y_1, \cdots, y_t]$$

Check this for a scalar case:

$$J = E[|\hat{x} - x|^2]$$

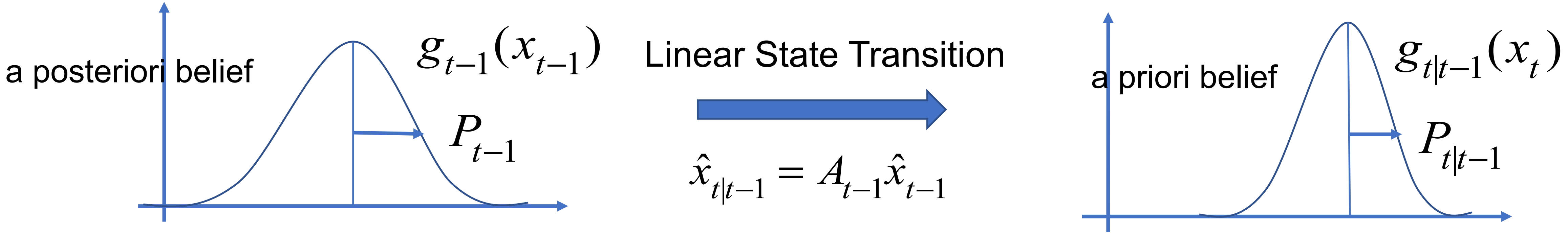
$$\frac{dJ}{d\hat{x}} = 2E[\hat{x} - x] = 2(\hat{x} - E[x]) = 0, \quad \therefore \hat{x} = E[x]$$



Proof of Gaussian Kalman Filter (Outline)

Step 1

- Given that $g_{t-1}(x_{t-1})$ is Gaussian, show that $g_{t|t-1}(x_t)$, too, is Gaussian.



- Use *Induction:* If the distribution of x_{t-1} is Gaussian with mean $\hat{x}_{t-1}$ and covariance P_{t-1}

$$g_{t-1}(x_{t-1}) = \frac{1}{\sqrt{\det(2\pi P_{t-1})}} \exp\left(-\frac{1}{2}(x_{t-1} - \hat{x}_{t-1})^T P_{t-1}^{-1}(x_{t-1} - \hat{x}_{t-1}) \right)$$

Then we can show that $g_{t|t-1}(x_t)$ is also Gaussian.

$$g_{t|t-1}(x_t) = \frac{1}{\sqrt{\det(2\pi P_{t|t-1})}} \exp\left(-\frac{1}{2}(x_t - \hat{x}_{t|t-1})^T P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1}) \right)$$

where $P_{t|t-1} = A_{t-1}P_{t-1}A_{t-1}^T + Q_{t-1}$

Linear state propagation does not distort the distribution.

- This is a highly technical derivation. See the lecture notes for details.

Step 2: Belief Update

- Recall $g_t(x_t) = \eta f_V(y_t - h(x_t,t)) g_{t|t-1}(x_t)$

Measurement noise pdf

exp(#) Gaussian with Covariance R_t

Gaussian with Covariance P_{t-1} exp(##)

- Therefore, the belief update should be in the following form:

$$g_t(x_t) = \eta' \exp[\#] \exp[\#\#] = \eta' \exp[\# + \#\#] = \eta' \exp[-N(x_t)]$$

- Here,

$$N(x_t) = \frac{1}{2}(y_t - H_t x_t)^T R_t^{-1} (y_t - H_t x_t) + \frac{1}{2}(x_t - \hat{x}_{t|t-1})^T P_{t|t-1}^{-1} (x_t - \hat{x}_{t|t-1})$$

- This is a quadratic function of x_t

$$N(x_t) = \frac{1}{2} x_t^T (H_t^{\ T} R_t^{-1} H_t + P_{t|t-1}^{-1}) x_t + \cdots$$

Step 2: Belief Update (Continued)

$$g_t(x_t) = \eta' \exp[-N(x_t)]$$

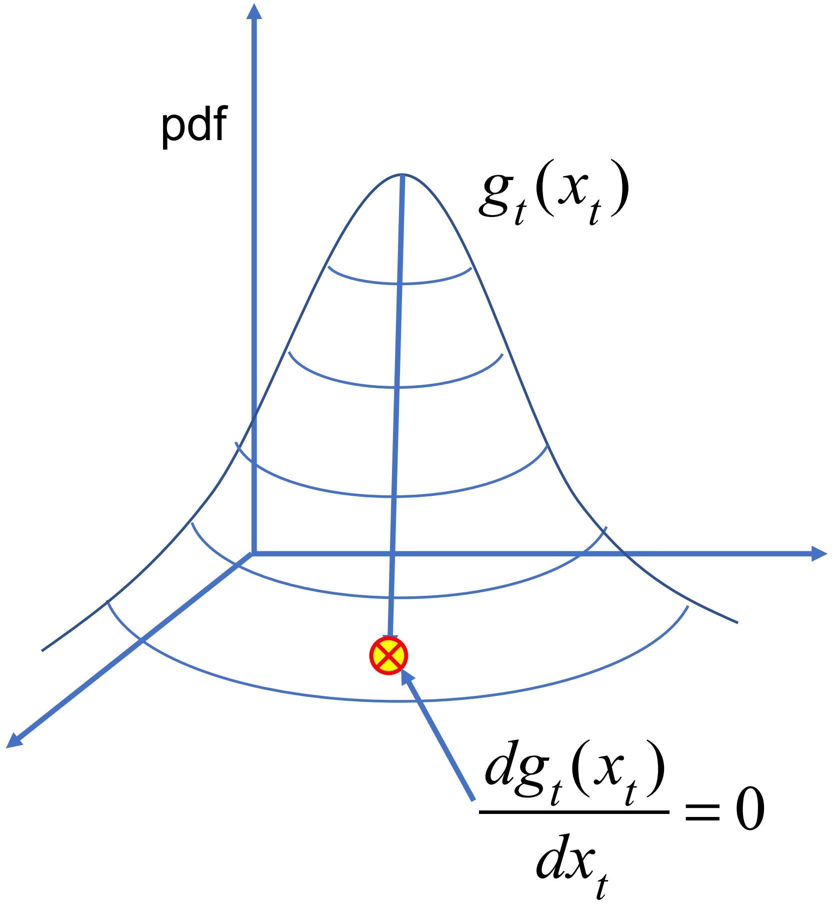
where

$$N(x_t) = \frac{1}{2} x_t^T (H_t^T R_t^{-1} H_t + P_{t|t-1}^{-1}) x_t + \cdots$$

- Recall that the optimal estimate is the conditional mean: $\hat{x}_t^o = E[\hat{x}_t | y_1, \cdots, y_t]$
- Since the pdf of x_t is Gaussian, the mean is at the peak and is unique: Convex Optimization.

$$\frac{dg_t(x_t)}{dx_t} = \eta' \exp[-N(x_t)] \frac{d}{dx_t}(-N(x_t)) = 0$$

Since $\exp[-N(x_t)] \neq 0$ $\quad \therefore \frac{dN(x_t)}{dx_t} = 0$



Step 2: Belief Update (Continued)

$$\frac{dN(x_t)}{dx_t} = 0 \quad \text{where} \quad N(x_t) = \frac{1}{2}(y_t - H_t x_t)^T R_t^{-1}(y_t - H_t x_t) + \frac{1}{2}(x_t - \hat{x}_{t|t-1})^T P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1})$$

- Recall $\frac{d}{dx}(\frac{1}{2}x^T A x + \frac{1}{2}x^T B x) = 0 \quad \rightarrow \quad Ax + Bx = 0$

  Therefore $\frac{dN(x_t)}{dx_t} = 0 \quad \rightarrow \quad -H_t{}^T R_t^{-1}(y_t - H_t x_t) + P_{t|t-1}^{-1}(x_t - \hat{x}_{t|t-1}) = 0$

- Denoting x_t that satisfies the above optimality condition by $\hat{x}_t$

$$\begin{aligned} P_{t|t-1}^{-1}(\hat{x}_t - \hat{x}_{t|t-1}) &= H_t{}^T R_t^{-1}(y_t - H_t\hat{x}_t + H_t\hat{x}_{t|t-1} - H_t\hat{x}_{t|t-1}) \\ &= H_t{}^T R_t^{-1}(y_t - H_t\hat{x}_{t|t-1}) - H_t{}^T R_t^{-1} H_t(\hat{x}_t - \hat{x}_{t|t-1}) \end{aligned}$$

$$\rightarrow (P_{t|t-1}^{-1} + H_t{}^T R_t^{-1} H_t)(\hat{x}_t - \hat{x}_{t|t-1}) = H_t{}^T R_t^{-1}(y_t - H_t\hat{x}_{t|t-1})$$

Note

- Noting that $P_t^{-1} = P_{t|t-1}^{-1} + H_t^T R_t^{-1} H_t$ and pre-multiplying P_t

$$\therefore \hat{x}_t = \hat{x}_{t|t-1} + P_t H_t{}^T R_t^{-1}(y_t - H_t\hat{x}_{t|t-1})$$

This agrees with the state update formula.

This is the Kalman Gain K_t.

Punch Line

- We have arrived at the familiar linear filter:

$$\hat{x}_t = \hat{x}_{t|t-1} + K_t(y_t - H_t\hat{x}_{t|t-1}), \quad K_t = P_t H_t^{\ T} R_t^{-1}$$

- In this proof we have never assumed that the optimal filter is linear. Instead, the linear filter has been derived from the optimality conditions.

- Kalman Filter is optimal among linear and nonlinear filters, as long as the noise, w_t and v_t, are Gaussian, and the process is linear time-varying.