Maximum Likelihood Estimation

$\tilde{y}$: measurements

x : parameters

Model : linear Gaussian $\tilde{y} = Hx + v \quad v \sim N(0,R)$

nonlinear non Gaussian $\tilde{y} = f(x,v) \quad v \sim p(v)$

$\tilde{y} \sim p(\tilde{y}|x)$

$p(\tilde{y}|x)$: probability of seeing meas $\tilde{y}$ given parameters x.

$$\max_{x} \; p(\tilde{y}|x) = L(\tilde{y}|x) \Leftarrow$$

multiple ind. meas. $\tilde{y}_i$

$$L(\tilde{y}_i|x) = \prod_{i=1}^{q} p(\tilde{y}_i|x)$$

Ex.

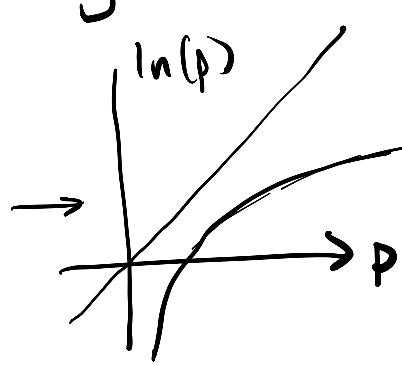
$$\tilde{y} = Hx + v \qquad v \sim N(0,R) \Leftarrow$$

$$p(\tilde{y}|x) = \frac{1}{\sim} e^{-\frac{1}{2}(\tilde{y}-Hx)^T R^{-1}(\tilde{y}-Hx)} \Big\} \text{ ugly}$$

take $\ln p(\tilde{y}|x)$

$$\ln(\tilde{p}|x) = \ln L(\tilde{y}|x)$$ log-likelihood.

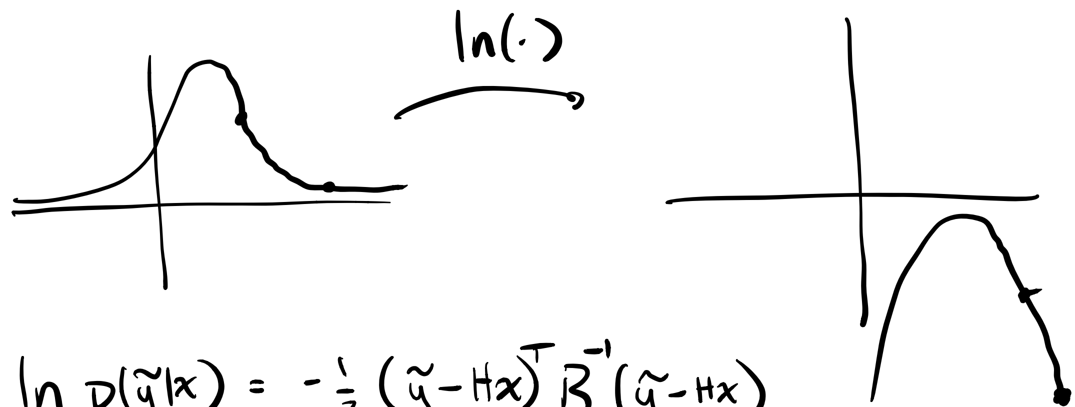
$p(\tilde{y}|x) \geq 0$



ln : monotonically increasing on $\mathbb{R}_+$

$h(p)$ $\quad$ $h(Ep) = p$

Ex.



$$\max_{x} \ln p(\tilde{y}|x) = -\frac{1}{2}(\tilde{y} - \underline{Hx})^T R^{-1} (\tilde{y} - \underline{Hx})$$

$$\hat{x} = (H^T R^{-1} H)^{-1} H^T R^{-1} \tilde{y} \;] \Longleftarrow$$

<u>Practical:</u>

Model: $\tilde{y} = Hx + v \qquad v \sim \mathcal{N}(0, R) \;]$

Nonlinear nonGaussian model: $\tilde{y} = h(x) + v \quad v \sim p(v)$

$$\max_{x} \ln p(\tilde{y}|x)$$

Maximum A-Posteriori Estimation
(MAP Estimation)

MLE: $\max_x p(\tilde{y}|x)$ ← prob of seeing $\tilde{y}$ given x

MAP: $\max_x p(x|\tilde{y})$ ← prob x was the parameters given $\tilde{y}$

Bayes Rule:

$$p(x|\tilde{y})\,p(\tilde{y}) = p(\tilde{y}|x)\,p(x)$$

$$p(x|\tilde{y}) = \frac{p(\tilde{y}|x)\,p(x)}{\boxed{p(\tilde{y})}}$$ ← normalization factor

$$p(\tilde{y}) = \int_{-\infty}^{\infty} p(\tilde{y}|x)\,p(x)\,dx$$ ⇐ const.

$$\ln\left(p(\tilde{y}|x)\,p(x)\right)$$

$$J_{MAP}(x) = \ln \underbrace{p(\tilde{y}|x)}_{\text{same MLE}} + \ln \underbrace{p(x)}_{\text{prior on } x.}$$

$$\max_x J_{MAP}(x)$$

can use for $\left[\ \tilde{y} = h(x,v),\ v \sim p(v) \quad \underline{x \sim p(x)}\right.$

Ex. $\tilde{y} = Hx + v \qquad v \sim N(0, R)$

$x = x_a + w \qquad w \sim N(0, Q)$

$$\max_{x} J_{MAP}(x)$$

$$\Longrightarrow \hat{x} = \left(H^T R^{-1} H + Q^{-1}\right)^{-1}\left(H^T R^{-1}\tilde{y} + Q^{-1} x_a\right)$$

FAR AFIELD TANGENT:

$$e^x e^y \xrightarrow{\ln} x + y$$

$$e^{x+y} \xrightarrow{\ln} x + y$$

Tropical Algebra's

Algebra

$(\max, +)$

$\downarrow$ max. $\quad \downarrow$ add

$(+, \times)$

add multiple.

Fenchel Transform | Fourier Transform

convex optimization

DISCRETE TIME KALMAN FILTERS

Before: $\tilde{y} = Hx + v \qquad v \sim N(0, R)$

Now: $x_{k+1} = \Phi_k x_k + \Gamma_k u_k + \Upsilon_k w_k \qquad w_k \sim N(0, Q_k)$

$\tilde{y}_k = \underline{H_k} x_k + v_k \qquad v_k \sim N(0, R_k)$

LTV: system $\tilde{y} \leftarrow [\; H_k \;][x_k]$

noise: zero mean, Gaussian, white noise

$$E[v_k v_j^T] = \begin{cases} 0 & k \neq j \\ R_k & k = j \end{cases} \qquad E[v_k w_j^T] = 0$$

$$E[w_k w_j^T] = \begin{cases} 0 & k \neq j \\ Q_k & k = j \end{cases}$$

Two part estimation scheme:

$\hat{x}_k^-, \hat{x}_k^+$: two different state estimates

PREDICTION (PROPOGATION)

$$\hat{x}_{k+1}^- = \Phi_k \underline{\hat{x}_k^+} + \Gamma_k u_k$$

what we predict the estimate to be at next time step

estimate at prev time step

MEASUREMENT (UPDATE)

$$\hat{x}_k^+ = \hat{x}_k^- + \boxed{K_k}\left[\tilde{y}_k - H_k \hat{x}_k^-\right]$$

updated estimate based on meas.

Kalman gain

what we meas

what we predict the meas to be

innovation

also want to track covariance:

$$P_k^- = E\left[(\hat{x}_k^- - x_k)(\hat{x}_k^- - x_k)^T\right]$$

$$P_k^+ = E\left[(\hat{x}_k^+ - x_k)(\hat{x}_k^+ - x_k)^T\right] \leftarrow$$

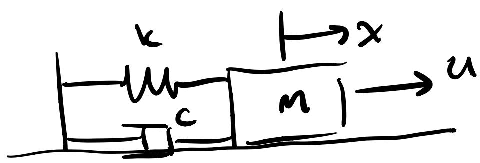
Need a way to update covariance.

Note: want to track density of $\hat{x}$

if $\hat{x}_0$ nomally distributed
dynamics linear w Gaussian noise w
meas. eqn linear w Gaussian noise v

$\hat{x}_k \rightarrow$ will be Gaussian for all time

Ex.



Y_k

$$f=ma \rightarrow \begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{c}{m} \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \begin{bmatrix} 0 \\ 1/m \end{bmatrix} u + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w$$

Update Covariance:

$$P_{k+1}^- = E\left[(\hat{x}_{k+1}^- - x_{k+1})(\hat{x}_{k+1}^- - x_{k+1})^T\right]$$

$$\hat{x}_{k+1}^- = \phi_k \hat{x}_k^+ + \cancel{\Gamma_k u_k} \qquad x_{k+1} = \phi_k x_k + \cancel{\Gamma_k u_k} + \gamma_k w_k$$

$$= E\left[\phi_k(\hat{x}_k^+ - x_k)(\hat{x}_k^+ - x_k)^T \phi_k^T\right] +$$

$$E\left[\phi_k \cancel{(\hat{x}_k^+ - x_k)}^{0} w_k^T \gamma_k^T\right] +$$

$$E\left[\gamma_k w_k \cancel{(\hat{x}_k^+ - x_k)^T}^{0} \phi_k^T\right] +$$

$$E\left[\gamma_k w_k w_k^T \gamma_k^T\right]$$

Note x_{k+1} depends on w_k but x_k indep. of w_k

$$P_{k+1}^- = \phi_k P_k^+ \phi_k^T + \underbrace{\gamma_k Q_k \gamma_k^T}_{\text{affect of noise}}$$

Note $P_{k+1}^- \geq \phi_k P_k^+ \phi_k^T$ (in a PD sense)

now incorporate measurements.

$$P_k^+ = E\left[(\hat{x}_k^+ - x_k)(\hat{x}_k^+ - x_k)^T\right]$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k\left[\tilde{y}_k - H_k \hat{x}_k^-\right]$$

$$\hat{y}_k = H_k x_k + v_k$$

$$\hat{x}_k^+ - x_k = (I - K_k H_k)(\hat{x}_k^- - x_k) + K_k v_k$$

$$P_k^+ = (I - K_k H_k) E\left[(\hat{x}_k^- - x_k)(\hat{x}_k^- - x_k)^T\right](I - K_k H_k)^T$$

$$+ K_k E\left[v_k v_k^T\right] K_k^T$$

$$= (I - \underline{K_k H_k}) P_k^- (I - \underline{K_k H_k}) + K_k R_k K_k^T$$

where we used $E\left[(\underline{\hat{x}_k^- - x_k})\, v_k^T\right] = 0$

noise in meas. at time k in dep of state at k

how to pick K_k to optimize P_k^+

"how to make best use of measurements"

$$\min_{K_k} Tr(P_k^+) \Rightarrow Tr\left(E\left[(\hat{x}_k^+ - x_k)(\hat{x}_k^+ - x_k)^T\right]\right) = Tr\left(E\left[(\hat{x}_k^+ - x_k)^T(\hat{x}_k - x_k)\right]\right)$$

$$J(K_k) = Tr\left((I - \underline{K_k H_k}) P_k^- (I - \underline{K_k H_k}) + K_k R_k K_k^T \right)$$

opt cond: $\dfrac{\partial J}{\partial K_k} = 0 \Rightarrow$

$$\mathrm{Tr}\Big((I - \underline{\Delta K_k H_k}) P_k^- (I - \underline{K_k H_k}) +$$
$$(I - K_k H_k) P_k^- (I - \Delta K_k H_k) +$$
$$\Delta K_k R_k K_k^T \quad + \quad K_k R_k \Delta K_k^T \Big)$$

$$\Rightarrow \quad -2(I - K_k H_k) P_k^- H_k^T + 2 K_k R_k = 0$$

$$\boxed{K_k = P_k^- H_k^T \left[H_k P_k^- H_k^T + R_k \right]^{-1}} \Leftarrow$$

$$P_k^+ = P_k^- - \underline{K_k H_k P_k^- - P_k^- H_k^T K_k^T} + K_k \underline{\left[H_k P_k^- H_k^T + R_k \right]} K_k^T$$
$$-2 K_k H_k P_k^- \qquad + K_k H_k P_k^-$$

$$\boxed{P_k^+ = \left[I - \underline{K_k} H_k \right] P_k^-}$$

$$P_k^+ = P_k^- - P_k^- H_k^T \left[H_k P_k^- H_k^T + R_k \right]^{-1} H_k P_k^-$$

woodbury identity $\downarrow$

$$P_k^+ = \left[(P_k^-)^{-1} + H_k^T R_k^{-1} H_k \right]^{-1}$$