Topics

- Max Likelihood Estimation (MLE)
- Max A posterior Estimation (MAP)
- Kalman Filter

$\left.\right\}\rightarrow$ prob perspectives on LS.

Maximum Likelihood Estimation (MLE)

unknown parameters: x

measurements: $\tilde{y} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_m \end{bmatrix}$

probability density

$p(\tilde{y} \mid x)$

prob. that meas is seen given these parameters

Gaussian example:

1D: $x = [\mu \ \sigma^2]$

$$p(\tilde{y} \mid x) = \left(\frac{1}{2\pi\sigma^2}\right)^{m/2} e^{\left[-\sum_{i=1}^{m}(\tilde{y}_i - \mu)^2 / 2\sigma^2\right]}$$

Given measurement $\tilde{y}$ the max likelihood estimation $\hat{x}$ value of x that maximizes $p(\tilde{y} \mid x)$

Likelihood Function

prob of seeing all meas.

$$L(\tilde{y} \mid x) = \prod_{i=1}^{m} \underbrace{p(\tilde{y}_i \mid x)}_{\text{prob of seeing meas } \tilde{y}_i} \longleftarrow L(\tilde{y} \mid x) \geq 0$$

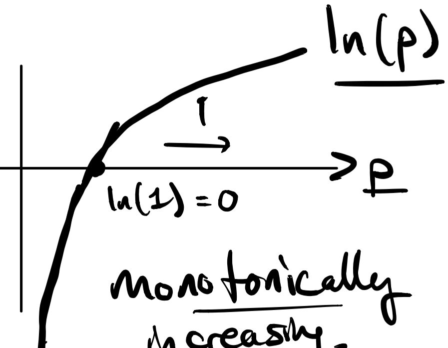
$$\hat{x} = \max_{x} L(\tilde{y}|x) \qquad (*)$$

Problems:

$$\underbrace{\prod_{i=1}^{m}}_{\text{cumbersome}} \; \underbrace{p(\tilde{y}_i|x)}_{\text{often of Gaussian form containing } e^{(\,-\,)}} \leftarrow$$

Solution: $\ln(\cdot)$ $\quad \ln(e^{(x)}) = x \quad e^{(\,-\,)}$

Log-Likelihood



$$\hat{x} = \max_{x} \ln L(\tilde{y}|x) \qquad (**)$$

monotonically increasing

optimizer of $(**)$ also optimizer of $(*)$

$$\ln L(\tilde{y}|x) = \ln\left(\prod_{i=1}^{m} p(\tilde{y}_i|x)\right) = \sum_{i=1}^{m} \ln p(\tilde{y}_i|x)$$

Necessary: $\left[\frac{\partial}{\partial x} \ln L(\tilde{y}|x)\right]_{\hat{x}} = 0 \Leftarrow$ fixed pt.

Sufficient: $\left[\frac{\partial^2}{\partial x \partial x^T} \ln L(\tilde{y}|x)\right]_{\hat{x}} < 0$ negative definite

if x is a vector $\rightarrow$ matrix

an extension:

Maximum a posteriori Estimation (MAP)

specific case of Bayesian Estimation

before: x fixed & unknown

Review:

now:

Conditional: of $\tilde{y}$ given x $\quad p(\tilde{y}|x)$

$x \sim p(x)$

how do we incorporate this "prior" knowledge of x into estimation?

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

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

Bayes Rule:

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

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

Relationship between conditional probabilities $p(\tilde{y}|x)$ & $p(x|\tilde{y})$

MLE: $\max\ p(\tilde{y}|x)$

MAP: $\max\ p(x|\tilde{y}) = \dfrac{p(\tilde{y}|x)\,p(x)}{p(\tilde{y})}$

$$\max_x\ p(x|\tilde{y}) = \left[\frac{p(\tilde{y}|x)\,p(x)}{p(\tilde{y})}\right]$$

doesn't depend on x

$$\downarrow$$

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

applying $\ln(\cdot)$

added term based on prior of x

$$\max_x\ \underbrace{\ln(p(\tilde{y}|x))}_{\text{MLE}} + \underbrace{\ln p(x)}_{\text{give a bonus for } x\text{'s } \bar{w} \text{ high probability}}$$

Ex. Gaussians.

Maximum Likelihood: MLE

$\tilde{y} = Hx + v \qquad v \sim N(0, R) \qquad \tilde{y} \in \mathbb{R}^m \quad H \in \mathbb{R}^{m \times n}$

$v = \tilde{y} - Hx$

$$p(\tilde{y}|x) = \frac{1}{(2\pi)^{m/2}\det(R)^{1/2}}\, e^{-\frac{1}{2}[\tilde{y}-Hx]^T R^{-1}[\tilde{y}-Hx]}$$

$\ln e^x = x$

$$\max_x\ \ln L(\tilde{y}|x) = \ln p(\tilde{y}|x)$$

$$= \underbrace{-\ln\left(\cancel{(2\pi)^{m/2}\det(R)^{1/2}}\right)}_{\text{constant}} - \frac{1}{2}[\tilde{y}-Hx]^T R^{-1}[\tilde{y}-Hx]$$

$$\min_{x} + \frac{1}{2}[\tilde{y}-Hx]^T R^{-1} [\tilde{y}-Hx] \longleftarrow$$

same as min variance estimation

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

Max. a posteriori (MAP)

Now a prior... $\underline{x} \sim N(\underline{x_a}, \underline{Q})$

$$p(\underline{x}) = \frac{1}{\underline{(2\pi)^{n/2} \det(Q)^{1/2}}} e^{-\frac{1}{2}[x_a - x]^T Q^{-1} [x_a - x]}$$

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

$$\min_{x} + \frac{1}{2}[\tilde{y}-Hx]^T R^{-1}[\tilde{y}-Hx] + \frac{1}{2}(x_a - x)^T Q^{-1} (x_a - x)$$

$$\frac{\partial}{\partial x} = 0 \Rightarrow \hat{x} = (\underline{H^T R^{-1} H + Q^{-1}})^{-1} (\underline{H^T R^{-1} \tilde{y} + Q^{-1} x_a})$$

minimum variance estimate w/ a prior on x

efficient

Efficiency:

MAP estimator: new term from prior distribution $p(x)$

Cramer Rao bound:

$$P = E[(\hat{x}-x)(\hat{x}-x)^T] \geq \left[F + E \frac{\partial}{\partial x}\ln(p(x))\frac{\partial}{\partial x}\ln(p(x))^T\right]^{-1}$$

---

now introduce estimation for parameters w/ dynamics...

Before: static

Reality: $\tilde{y} = Hx + v$

Estimator: $\hat{y} = H\hat{x}$

Now dynamic

$$\dot{x} = Fx + Bu$$
$$\tilde{y} = Hx + v$$

$$\dot{\hat{x}} = F\hat{x} + Bu + \overbrace{K[\tilde{y} - \underbrace{H\hat{x}}_{\hat{y}}]}^{\text{feedback}}$$
$$\hat{y} = H\hat{x}$$

K: observer feedback gain

NEW NOTATION

$$\tilde{x} = e = \hat{x} - x$$

$$\dot{e} = \dot{\hat{x}} - \dot{x} = (F - KH)e + Kv + \cancel{Bu} - \cancel{Bu}$$

LIN SYS

$A + LC$

selected char poly/ eigenvals for stability

Question: how to select char poly/ eigenvalues...

- place cmd.
- observable/controllable canonical form
- Ackermann's formula

Discrete time version:

$$\dot{x}_t = F_t x + B_t u \qquad \Phi = e^{F\Delta t}$$

$$\tilde{y}_t = H_t x + v_t \qquad \Gamma = \int_0^{\Delta t} e^{F(\Delta t - \tau)} B \, d\tau$$

Reality

$$x_{k+1} = \Phi_k x_k + \Gamma u_k$$

$$y_k = H_k x_k + v_k$$

Estimator

$$\hat{x}_{k+1} = \Phi_k \hat{x}_k + \Gamma u_k + K_k [\tilde{y}_{k+1} - H \hat{x}_k]$$

instead

problem

① $$\hat{x}^-_{k+1} = \Phi_k \hat{x}^+_k + \Gamma u_k$$ prediction or propagation

② $$\hat{x}^+_{k+1} = \hat{x}^-_{k+1} + K[\tilde{y}_{k+1} - H\hat{x}^-_{k+1}]$$ update or correction

Full update:

$$\hat{x}^-_{k+1} = \Phi_k x^-_k + \Gamma_k u_k + \Phi_k K [\tilde{y}_k - H \hat{x}^-_k]$$

$$\tilde{x}^-_k = e^-_k = \hat{x}^-_k - x_k$$

$$\tilde{x}^+_k = e^+_k = \hat{x}^+_k - x_k$$

$$e^-_{k+1} = \Phi [I - KH] e^-_k \leftarrow$$

$$e^+_{k+1} = [I - KH] \Phi e^+_k \leftarrow$$

Note:
$\Phi[I-KH]$ &
$[I-KH]\Phi$
have the same eigenvalues
↳ square invertible A, B
eigenvalues of AB = evals of BA

$$B(AB)B^{-1} = BA$$

DISCRETE TIME KALMAN FILTER:

Reality:

$$x_{k+1} = \phi_k x_k + \Gamma_k u_k + \Upsilon_k w_k$$

state, control, process noise

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

measurement noise

NOISE v_k, w_k zero mean Gaussian white noise processes

$v_k \sim N(0, R_k)$
$w_k \sim N(0, Q_k)$

static version

white noise version

v_k & w_k are not correllated over time.

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

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

optimal estimator derivation:

$$\hat{x}_{k+1}^- = \phi_k \hat{x}_k^+ + \Gamma_k u_k \leftarrow \text{propagation}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k[\tilde{y}_k - H_k \hat{x}_k^-] \qquad \text{correction}$$

Error covariance

$$e_k^- = \hat{x}_k^- - x_k \qquad e_{k+1}^- = \hat{x}_{k+1}^- - x_{k+1}$$

$$e_k^+ = \hat{x}_k^+ - x_k \qquad e_{k+1}^+ = \hat{x}_{k+1}^+ - x_{k+1}$$

$P_k^- = E[e_k^- e_k^{-T}]$ $\quad$ $P_{k+1}^- = E[e_{k+1}^- e_{k+1}^{-T}]$

$P_k^+ = E[e_k^+ e_k^{+T}]$ $\quad$ $P_{k+1}^+ = E[e_{k+1}^+ e_{k+1}^{+T}]$

$$\underline{e_{k+1}^-} = \Phi_k e_k^+ - \Upsilon_k w_k + \cancel{\Gamma_k u_k} - \cancel{\Gamma_k u_k}$$

DISCRETE TIME KALMAN FILTER:

Reality:

$$x_{k+1} = \Phi_k \underset{\text{state}}{x_k} + \Gamma_k \underset{\text{control}}{u_k} + \Upsilon_k \underline{w_k} \leftarrow \text{process noise}$$

$$\tilde{y}_k = H_k x_k + v_k \leftarrow \text{measurement noise}$$

Covariance update:

$$P_{k+1}^- = E[e_{k+1}^- e_{k+1}^{-T}]$$

$$= E[\Phi_k e_k^+ e_k^{+T} \Phi_k^T] - E[\Phi_k \overset{0}{\cancel{e_k^+ w_k^T}} \Upsilon_k^T]$$

$$- E[\Upsilon_k \overset{0}{\cancel{w_k e_k^{+T}}} \Phi_k^T] + E[\Upsilon_k w_k w_k^T \Upsilon_k^T]$$

$$= \Phi_k E[e_k^+ e_k^{+T}] \Phi_k^T + \Upsilon_k E[w_k w_k^T] \Upsilon_k^T$$

$$\boxed{P_{k+1}^- = \Phi_k P_k^+ \Phi_k^T + \Upsilon_k Q_k \Upsilon_k^T} \longleftarrow$$

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

$$= (I - K_k H_k) e_k^- + K_k v_k$$

$$P_k^+ = E(e_k^+ e_k^{+T}) = E[(I - K_k H_k) e_k^- e_k^{-T} (I - K_k H_k)^T]$$

$$+ E[(I - K_k H_k) \overset{0}{\cancel{e_k^- v_k^T}} K_k^T] + E[K_k \overset{0}{\cancel{v_k e_k^{-T}}} (I - K_k H_k)^T]$$

$$E[K_k v_k v_k^T K_k^T]$$

$$= (I - K_k H_k) E[e_k^- e_k^{-T}] (I - K_k H_k)^T + K_k E[v_k v_k^T] K_k^T$$

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

Initial cond.: $\boxed{P_0^- = E[e_0 e_0^T]} = E[(\hat{x}_0 - x_0)(\hat{x}_0 - x_0)^T]$

doesn't have to be precise

Choosing K_k:

$$\min_{K_k} J(K_k) = Tr(P_k^+) = E(Tr(e_k^+ e_k^{+T}))$$

$$= E(Tr(e_k^{+T} e_k^+))$$

sum of squared errors

$$\frac{\partial J}{\partial K_k} = 0 = -2(I - K_k H_k) P_k^- H_k^T + 2 K_k R_k$$

$$\Longrightarrow \boxed{K_k = P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1}} \leftarrow$$

plugging back into P_k^+ step...

$$P_k^+ = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T$$

$$= P_k^- - K_k H_k P_k^- - P_k^- H_k^T K_k^T + K_k R_k K_k^T$$

$$+ K_k H_k P_k^- H_k^T K_k^T$$

$$K_k [H_k P_k^- H_k^T + R_k] K_k^T$$

$$\cancel{- P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1} H_k P_k^-}$$

$$\cancel{+ P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1} H_k P_k^-}$$

$$P_k^+ = P_k^- - K_k H_k P_k^-$$

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

Summary:

MODEL: $$x_{k+1} = \Phi_k x_k + \Gamma_k u_k + \Upsilon_k w_k \qquad w_k \sim N(0, Q_k)$$
$$\tilde{y}_k = H_k x_k + v_k \qquad v_k \sim N(0, R_k)$$

Initialize: $$\hat{x}_0^- = \hat{x}(t_0)$$
$$P_0^- = E[e_0 e_0^T] = E[(\hat{x}_0 - x_0)(\hat{x}_0 - x_0)^T]$$

Gain (optimal) $$K_k = P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1}$$

Update $$\hat{x}_k^+ = \hat{x}_k^- + K_k[\tilde{y}_k - H_k \hat{x}_k^-]$$
$$P_k^+ = [I - K_k H_k] P_k^-$$

PROPAGATION $$\hat{x}_{k+1}^- = \Phi_k \hat{x}_k^+ + \Gamma_k u_k$$
$$P_{k+1}^- = \Phi_k P_k^+ \Phi_k' + \Upsilon_k Q_k \Upsilon_k^T$$

$$P_{k+1}^- = \Phi_k P_k^+ \Phi_k' + \Upsilon_k Q_k \Upsilon_k^T$$

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

small matrix

$$P_k^+ = P_k^- - P_k^- H_k^T \overbrace{[H_k P_k^- H_k^T + R_k]}^{\text{small matrix}}{}^{-1} H_k P_k^-$$

woodbury matrix identity

matrix inversion lemma

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

more computationally expensive

Wikipedia

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

pos def

pos def

sort of depending on Φ_k

dynamics

$$P_{k+1}^- = \Phi_k P_k^+ \Phi_k' + \Upsilon_k Q_k \Upsilon_k^T$$

increases size of P_k

meas. correction

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

decreases size of P_k

pos def

pos def

improve the covariance