Estimation:

- Parameter Estimation
- Signal Estimation

Minimum Variance Unbiased Estimator (MVUE)

$p(x;\theta)$ ← Parameterised by θ

$x[0], \dots, x[N-1]$

$\hat{\theta} \leftarrow RV, \quad \theta \leftarrow const$

Unbiased Estimation: $E(\hat{\theta}) = \theta \quad \forall\ a \le \theta \le b$

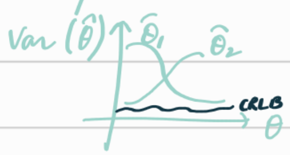
$$MSE = E\left[(\theta - \hat{\theta})^2\right]$$

$$= E\left[\left((\theta - E\hat{\theta}) + (E\hat{\theta} - \hat{\theta})\right)^2\right]$$

$$= \underset{(\theta - E\hat{\theta})}{bias^2} + \underset{E((\hat{\theta} - E\hat{\theta})^2)}{Var\ \hat{\theta}}$$

MVUE: Among all U.E. $\hat{\theta}$, $\hat{\theta}^*$ is MVUE if $Var\ \hat{\theta}^* \le Var\ \hat{\theta} \quad \forall \theta$
$\forall\ \hat{\theta} = g(x)$

The value of variance can dependent on θ itself.



Cramer-Rao Lower bound (CRLB) ~ & MVUE can be different

$x[0] = \theta + w[0]$, $w[0] \sim N(0, \sigma^2)$

$$p(x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\cdot(x[0]-\theta)^2}$$

$$\ln p(x;\theta) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x[0]-\theta)^2$$

$$\frac{\partial^2(\ln p(x;\theta))}{\partial\theta^2} = -\frac{1}{\sigma^2}$$

Theorem: (Scalar θ) Assume that the pdf $p(x,\theta)$ satisfies the regularity condition $\left[E_{p(x,\theta)}\left(\frac{\partial \ln(p(x,\theta))}{\partial\theta}\right) = 0 \quad \forall\ \theta\right]$
Then the variance of any unbiased estimator must satisfy

$$Var(\hat{\theta}) \ge \frac{1}{-E\left(\frac{\partial^2 \ln(p(x;\theta))}{\partial\theta^2}\right)}$$ ← Fischer Information metric

Further: UE can be found that attains the CRLB for $\forall\ \theta$ iff $\frac{\partial \ln p(x,\theta)}{\partial\theta} = I(\theta)\cdot[g(x) - \theta]$

for some functions g & I. Then that estimator is

$\hat{\theta} = g(x)$ is MVUE with var $= \frac{1}{I(\theta)}$

if we have a sequence $x[0] \cdots [n-1]$

$$\ln p(x,\theta) = \sim - \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (x(i) - \theta)$$

$$\frac{\partial \ln p(x,\theta)}{\partial \theta} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x(n) - \theta) \rightarrow \underbrace{\frac{N}{\sigma^2}}_{I(\theta)} \Big(\overbrace{\frac{\sum_{n=0}^{N-1} x(n)}{N}}^{g(x)} - \theta \Big)$$

$$\frac{\partial^2 \,''}{\partial \theta^2} = \frac{-N}{\sigma^2}$$

Example w.o. regularity condition:

$x[0], x[1], \cdots, x[N-1]$ iid $\sim U[0,\theta]$, Estimate θ.

$$\prod_{n=0}^{N-1} p(x(n) \mid \theta) = 1/\theta^n$$

$$E\left(\frac{\partial \ln p(x,\theta)}{\partial \theta}\right) = \frac{-n}{\theta}$$

$$-\frac{\partial^2 \ln p}{\partial \theta^2} = \frac{-n}{\theta^2} \qquad \frac{\partial^2 \ln p}{\partial \theta^2} = \frac{-n}{\theta^2}$$

$$E\left(\frac{\partial \ln}{\partial \theta}\right)^2 = \frac{n^2}{\theta^2}$$

Exercise

$$\left\{ \begin{aligned} & -E\left[\frac{\partial^2 \ln p}{\partial \theta^2}\right] \\ & = E\left(\left(\frac{\partial \ln p}{\partial \theta}\right)^2\right) \end{aligned} \right\}$$

under regularity condition

wrong CRLB application $\mathrm{Var}(\hat{\theta}) \geq \frac{\theta^2}{n^2}$

$\hat{\theta} = \max x[n]$

Example: $Y = \max(x(n))$

$$\underbrace{P(Y \leq \alpha)}_{} \rightarrow \left(\frac{\alpha}{\theta}\right)^n$$

$$\underbrace{n\left(\frac{\alpha^{n-1}}{\theta^n}\right)}$$

$$E(Y) = \int_0^\theta n \cdot \frac{\alpha^n}{\theta^n} d\alpha = \frac{n}{n+1} \frac{\theta^{n+1}}{\theta^n} = \frac{n\theta}{n+1}$$

$$\hat{\theta} = \frac{n+1}{n} Y$$

Exercise

$$\text{☆} \quad \mathrm{Var}\, \hat{\theta} = \frac{\theta^2}{n(n+2)} \quad \{\text{smaller than CRLB}\}$$

$$E(Y^2) = \int_0^\theta \alpha^2 \cdot n \frac{\alpha^{n-1}}{\theta^n} d\alpha = \frac{n}{n+2} \theta^2$$

$h(x) = 1 \quad \{a(\theta) \leq x \leq b(x)\} \quad 0 \quad \text{o.w.}$

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f(x,t)\,dt = f(x,b(x))\frac{d}{dx}b(x) - f(x,a(x))\frac{d}{dx}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)\,dt$$

$\hat{\theta} = g(x)$

unbiased $\Rightarrow \int_x \hat{\theta}\, p(x,\theta)\,dx = \theta \longrightarrow \int_x \hat{\theta}\, \frac{\partial(p(x,\theta))}{\partial\theta}\,dx = 1$

Regularity condition $\rightsquigarrow \int_x \left(\frac{\partial}{\partial\theta}\ln p(x,\theta)\right) p(x,\theta)\,d\theta = 0$

$$\int_x \frac{\partial p(x,\theta)}{\partial\theta}\,dx \stackrel{?}{=} \frac{\partial}{\partial\theta}\int_x p(x,\theta)\,dx = 0$$

X is not f^n of θ

$$\int_x (\hat{\theta}-\theta)\frac{\partial \ln p(x,\theta)}{\partial\theta}\, p(x,\theta)\,dx = 1$$

$$1 \stackrel{C.S.}{\leq} \underbrace{\int (\hat{\theta}-\theta)^2 p(x,\theta)\,dx}_{Var(\hat{\theta})} \int \left[\frac{\partial \ln p(x,\theta)}{\partial\theta}\right]^2 p(x,\theta)\,dx$$

Claim : $$E\left[\left(\frac{\partial \ln(p(x,\theta))}{\partial\theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln(p(x,\theta))}{\partial\theta^2}\right]$$ under regularity

From Cauchy Schwarz, they must be a scalar multiple.

$$\Rightarrow \frac{\partial \ln p(x,\theta)}{\partial\theta} = \frac{1}{c(\theta)}(\hat{\theta}-\theta)$$

$$E\left(\frac{\partial^2 \ln p(x,\theta)}{\partial\theta^2}\right) = \frac{-1}{c(\theta)} + E\left(\frac{\partial\left(\frac{1}{c(\theta)}\right)}{\partial\theta}(\hat{\theta}-\theta)\right)$$

$Cov(\hat{\theta}_*) - Cov(\hat{\theta}_{BAD}) \preceq 0$ for vectors.

$\left\{\begin{array}{l}\hat{\theta} \text{ is M.V.U.E. for any unbiased estimator } \bar{\theta} \\ \quad \text{iff} \quad Var(\hat{\theta}_i) \leq Var(\bar{\theta}_i);\ i = 1, \dots, p\end{array}\right.$

Relaxed definition

CRLB

Regularity Condition $E\left(\frac{\partial \ln p(x,\theta)}{\partial\theta}\right)_{p\times 1} = 0 \quad \forall\theta$

Cov of any unbiased estimator satisfies

$$\underbrace{E\left([\hat{\theta} - E(\hat{\theta})][\hat{\theta} - E(\hat{\theta})]^T\right)}_{cov(\hat{\theta})} - I^{-1}(\theta) \geq 0$$

$$[I(\theta)]_{ij} = -E\left[\frac{\partial^2 \ln p(x,\theta)}{\partial\theta_i \, \partial\theta_j}\right]$$

An unbiased estimator can be found that attains CRLB iff

$$\left.\frac{\partial \ln p(x,\theta)}{\partial\theta}\right|_{p\times 1} = I(\theta)\,[\underbrace{g(x)}_{\hat{\theta}} - \theta]$$

MVUE $\rightarrow \hat{\theta} = g(x)$ with variance $I^{-1}(\theta)$

Example :

$x[n] = A + Bn + w[n]$; $n = 0, \ldots, N-1$

$\theta = [A \quad B]^T$

$$p(x,\theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x(n) - A - Bn)^2\right\}$$

$$I(\theta) = \frac{1}{\sigma^2}\begin{bmatrix} N & \frac{N(N-1)}{2} \\ \frac{N(N-1)}{2} & \frac{N(N-1)(2N-1)}{6} \end{bmatrix}$$

$$I^{-1}(\theta) = \sigma^2 \left[\quad \sim \quad \right]$$

$$Var(\hat{A}) \geq \frac{2(2N-1)}{N(N+1)}\sigma^2$$

$$Var(\hat{B}) \geq \frac{12\sigma^2}{N(N^2-1)}$$

$\hat{A}$ & $\hat{B}$ can be computed, but hard to compute.

MLE is efficient in the limit

→ Objective: Synthesize MVUE even when MVUE ≠ efficient

↳ Two properties are required to be assumed

- ↳ sufficient statistics
- ↳ Complete statistics

Rao - Blackwell - Lehman - Scheffe Theorem (RBLS Thm.)

↳ Using sufficient, we can construct unique MVUE

To find the sufficient stats, we have Neyman - Fischer Factorisazation Thm.

Sufficiency Principle: If $T(x)$ is a sufficient statistic for θ, then MLE estimate of θ should depend only on $T(X)$

$\Leftrightarrow$ If for x, y, $T(x) = T(y)$ then $\hat{\theta}_{MLE}$ should be same for both $X = x$ & $X = y$

Definition of sufficient statistic: A statistic $T(x)$ is a sufficient statistic for θ if the conditioned distribution of the sample X given the value of $T(x)$ doesnot depend on θ.

$$P(X = x \mid T(X) = T(x)) = \frac{P(X = x \ \& \ T(X) = T(x))}{P(T(X) = T(x))}$$

$$= \frac{P(X = x)}{P(T(X) = T(x))}$$

Defn: Let $P(x, \theta)$ be the pdf / pmf of X & $q(t, \theta)$ be the pmf / pdf of $T(X)$, then $T(X)$ is sufficient for θ if $\forall x, \forall \theta$

$$\frac{P(x, \theta)}{q(T(x), \theta)} \text{ is ind. of } \theta$$

$$P(X = x) = P_\theta(X = x \mid T(X) = T(x)) \cdot P_\theta(T(X) = T(x))$$

$$\theta^1_{MLE} = \arg\max \left(\ln P_\theta (X = x \mid T(X) = T(x)) + \ln P_\theta (T(X) = T(x)) \right)$$

$$\theta^2_{MLE} = \arg\max \left(\ln P_\theta (X = y \mid T(X) = T(y))'' + \ln P_\theta (T(X) = T(y)) \right)$$

① → if dependence is removed, then $\theta'_{MLE} = \theta^2_{MLE}$

Ex $x_1 \dots x_n$ iid $\mathcal{N}(\mu, \sigma^2)$

unknown (μ) known (σ^2)

$$T(x) = \bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i$$

$$p(x;\mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\frac{p(x,\theta)}{q(T(x),\theta)} = \text{is Ind. of } \theta.$$

way to check, not construct. yet

Neyman Fischer Factorisation Theorem

Let $p(x,\theta)$ be the pdf / pmf of X, a statistic $T(x)$ is sufficient statistic for θ iff $\exists$ functions $g(t,\theta)$ & $h(x)$ s.t. $\forall x, \forall \theta$

$$p(x,\theta) = g(T(x),\theta)\, h(x)$$

Proof: 1 way: if $\exists$ functions $\Rightarrow$ sufficient statistic

$$\frac{p(x,\theta)}{q(T(x),\theta)} = \frac{g(T(x),\theta)\, h(x)}{q(T(x),\theta)}$$

$$A_{T(x)} = \{y : T(y) = T(x)\}$$

$$= \frac{g(T(x),\theta)\cdot h(x)}{\sum_{A_{T(x)}} p(i,\theta)} = \frac{\cancel{g(T(x),\theta)}\cdot h(x)}{\sum \cancel{g(T(i),\theta)}\cdot h(i)} = \frac{h(x)}{\sum_{i \in A_{T(x)}} h(i)}$$

Example: $x[n] = w[n] \rightarrow$ i.i.d. with pdf

Estimate the mean $\mu\ [0,\beta]$

$$p(x,\beta) = \begin{cases} \frac{1}{\beta^n} & 0 < \min(x[n])\ ;\ \max(x[n]) < \beta \\ 0 & \text{otherwise} \end{cases}$$

$$= \underbrace{\frac{1}{\beta^N}\, u(\beta - \underbrace{\max(x[n])}_{T(x)})}_{g(T(x),\beta)} \cdot \underbrace{u(\min(x[n]))}_{h(x)} \qquad u(t) = \begin{cases} 0 & t<0 \\ 1 & t \geq 0 \end{cases}$$

1) MVUE is unique

2) Sufficiency of $T(x)$ $\Rightarrow$ $\phi(T) = E(W|T)$ is also an unbiased estimator

Let W be any unbiased estimator

$$\text{Var}(\phi(T)) \leq \text{Var}(W)$$

Identities

$$E(X) = E[E(X|Y)]$$

$$Var(X) = Var(E(X|Y)) + E(Var(X|Y))$$

def

$$Var(X|Y) = E((X - E(X|Y))^2 | Y)$$

$$Var(E(X|Y)) = E((E(X|Y) - E(X))^2)$$

Thm: If W is MVUE of θ, it is unique

Proof: W' be another MVUE

$$Var\left(\frac{W+W'}{2}\right) = \frac{1}{4}Var(W) + \frac{1}{4}Var(W') + \frac{1}{2}Cov(W, W')$$

(Cauchy Schwarz)

$$\leq \frac{1}{4}Var(W) + \frac{1}{4}Var(W') + \frac{1}{2}\sqrt{Var W'}\sqrt{Var W}$$

$$= Var W' = Var W$$

Equality $\Rightarrow$ $W' = \alpha(\theta) W + \beta(\theta)$

$$Cov(W, W') = Cov(W, \alpha(\theta)W + \beta(\theta))$$

$$= \alpha(\theta) Var_\theta(W)$$

$$\Rightarrow \alpha(\theta) = 1$$

Rao - Blackwell Theorem

Let W be any unbiased estimator of θ. Let T be a sufficient statistic. Define $\phi(T) = E(W|T)$

Then (i) $\phi(T)$ is an estimator of θ

(ii) $E_\theta[\phi(T)] = \theta$

(iii) $Var_\theta[\phi(T)] \leq Var_\theta(W)$ $\forall \theta$

$$Var(W) = Var(\underbrace{E(W|T)}_{\phi(T)}) + E(Var(W|T))$$

Example: Let X_1, X_2 be i.i.d. $N(\theta, 1)$

$$\bar{X} = \frac{X_1 + X_2}{2} \qquad E(\bar{X}) = \theta$$

$$Var(\bar{X}) = 1/2$$

$T = X_1$ is not a sufficient statistic

$$Var(E(\bar{X}/X_1)) \leq 1/2$$

But it is not an estimator

Thm: If $EW = \theta$, W is MVUE of θ iff W is uncorrelated with all unbiased estimator of θ.

Proof: $W + aU$ $\qquad E(U) = 0$

$\Rightarrow \mathrm{Var}(W + aU) = \mathrm{Var}\,W + a^2 \mathrm{Var}(U) + 2a\,\mathrm{Cov}(W, U)$

$\neq 0$

$\left\{\begin{matrix} \text{MVUE} \\ \Downarrow \\ \mathrm{cov} = 0 \end{matrix}\right\}$

$\Rightarrow$ minima $\quad a = \dfrac{-\mathrm{cov}(W, U)}{\mathrm{Var}(U)}$

$\left\{\begin{matrix} \text{MVUE} \\ \Uparrow \\ \mathrm{cov} = 0 \end{matrix}\right\}$

$\mathrm{Cov}(W, U) = 0 \quad \forall U \quad$ s.t. $E(U) = 0$

Let W' be an unbiased estimator

$E(W' - W) = 0$

$\Rightarrow \mathrm{Var}\,W' = \mathrm{Var}\,W + \mathrm{Var}(W' - W) + 2\,\cancel{\mathrm{cov}(W, W' - W)}$ $\quad = 0$

Def Let $f(t, \theta)$ be a family of pdfs / pmfs for a statistic $T(x)$. This set of pdfs / pmfs is called complete

if $\quad E_\theta\, g(T) = 0 \quad \forall \theta$

$\Rightarrow \quad P_\theta(g(T) = 0) = 1 \quad \forall \theta$

Ex $\quad x(n) = A + w(n)$

$T = \sum_{n=0}^{N-1} x(n)$

$$\int_{-\infty}^{\infty} g(T) \frac{1}{2\pi N\sigma^2} \exp\left(\frac{-1}{2N\sigma^2} (T - NA)^2 \right) dt = 0 \quad \forall A$$

$\rightarrow g(T) \equiv 0$

Fact: If $T(x)$ is complete there is exactly one function of $T(x)$ which is an unbiased estimator.

Incomplete sufficient statistic

$x[0] = A + w[0] \qquad w[0] \sim U(-1/2, 1/2)$

$T = x[0]$

$E[v(T)] = 0 \Rightarrow v(T) = 0 \quad \forall A$

$\int v(t)\, P(t, A)\, dt = 0 \qquad \int_{A-1/2}^{A+1/2} v(t)\, dt = 0 \not\Rightarrow v(t) \equiv 0$ $\quad$ periodic

Rao - Blackwell - Lehmann - Scheffe

Thm: Let T be a complete sufficient statistic for θ and let $\phi(T)$ be any unbiased estimator based on T. Then $\phi(T)$ is the unique MVUE.

Algorithm:

1) Find a sufficient statistic $T(x)$ using Neyman-Fischer Theorem
2) Determine if T is complete. If not find another T (sufficient + complete)
3) Find $\hat{\theta} = E(w / T(x))$ where w is any unbiased est.

3') Find a f^n $g(T)$ which is an unbiased est $\longrightarrow \hat{\theta} = g(T(x))$

Example: $x(n) = w(n) \overset{i.i.d}{\rightarrow} \mathcal{U}[0, \beta]$ $\quad n = 0, 1, \dots, N-1$ $\quad \beta > 0$

Estimate the mean $\longrightarrow \beta/2$

1) $$p(x, \beta) = \underbrace{\frac{1}{\beta^N} u(\beta - \max x[n])}_{g(T(x), \beta)} \underbrace{u(\min x[n])}_{h(x)}$$

From N-F

$$T(x) = \max x[n]$$

2) Check completeness

$$p(t, \beta) = \begin{cases} \frac{N}{\beta^N} t^{N-1} & 0 < t < \beta \\ 0 & \text{otherwise} \end{cases}$$

Let $E(g(t)) = 0 \quad \forall \beta$ $\rightarrow$ using this

$$\frac{\partial}{\partial \beta} E(g(T)) = 0 \qquad \frac{\partial}{\partial \beta}\left(\beta^{-N} \int_0^\beta g(t) N t^{N-1} dt\right)$$

$$= \beta^{-1} N g(\beta) = 0$$

$$\Rightarrow g(\beta) \equiv 0$$

Hence a complete sufficient statistic

3) Now we can construct an unbiased estimator

And it will be the MVUE

$p(x;\ \theta)$

$\uparrow$ θ also has a random variation in Bayesian Estimation

$p(x, \theta)$

$$p(\theta) = \int p(x,\theta)\, dx$$

$$p(x) = \int p(x,\theta)\, d\theta$$

Bayesian MSE $= \iint (\theta - \hat{\theta})^2\, p(x,\theta)\, dx\, d\theta$

$$= \int \underbrace{\left[\int (\theta-\hat{\theta})\, p(\theta/x)\, d\theta\right]}_{I(x)} p(x)\, dx$$

Need to minimize for each x

$$\frac{\partial}{\partial \hat{\theta}} \int (\theta - \hat{\theta})\, p(\theta/x)\, d\theta$$

$$= -2\int \theta\, p(\theta/x)\, d\theta + 2\hat{\theta}\int p(\theta/x)\, d\theta = 0$$

$$\Rightarrow \quad \hat{\theta} = E(\theta/x)$$

$$p(\theta/x) = p(x/\theta)\frac{p(\theta)}{p(x)} = \frac{p(x/\theta)\, p(\theta)}{\int p(x/\theta)\, p(\theta)\, d\theta}$$

posterior ($p(\theta/x)$), prior ($p(\theta)$)

$x(n) = A + w(n)$

$A \rightarrow p(A) \sim N(\mu_A, \sigma_A^2)$

$w(n) \rightarrow N(0, \sigma^2)$

$$p(A) = k_1 \exp\left(-\frac{(A-\mu_A)^2}{2\sigma_A^2}\right)$$

$$p(x/A) = k_2 \exp\left(\frac{-\sum (x(n) - A)^2}{2\sigma^2}\right)$$

$$\Rightarrow p(A/x) = k_3 \exp\left(\frac{-1}{2\sigma^2_{A/x}}(A - \mu_{A/x})^2\right)$$

$$\sigma^2_{A/x} = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}}, \quad \mu_{A/x} = \left(\frac{N}{\sigma^2}\bar{x} + \frac{\mu_A}{\sigma_A^2}\right)\sigma^2_{A/x}$$

$$\hat{A} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}\bar{x} + \frac{\sigma^2/N}{\sigma_A^2 + \sigma^2/N}\mu_A$$

~~Example:~~

$p(A) \sim u(-A_0, A_0)$

BLUE

Best Linear Unbiased Estimator

Assume $\hat{\theta} = \sum_{n=0}^{N-1} a_n \, x[n]$; $E\,\hat{\theta} = \theta \quad \forall \theta$

$x[n] \sim N(A, \sigma^2)$ — σ^2: known, A: parameter

$\hat{\theta} = \underbrace{\sum a_n x[n]}_{a^T x}$, $x[n] = s[n]\,\theta + w[n]$; $x = s\theta + w$, $E(w) = 0$

$E(\hat{\theta}) = a^T s\,\theta = \theta$
$\Rightarrow a^T s = 1$

Gauss Markov Theorem: If the data is generated by linear model

$x = H\theta + W$ where $H \in \mathbb{R}^{N\times P}$, $\theta \in \mathbb{R}^{P\times 1}$

& W is $N\times 1$ noise vector & $E(W) = 0$.

and $\mathrm{cov} = C$, then BLUE of θ is

$$\boxed{\hat{\theta} = (H^T C^{-1} H)^{-1} H^T C^{-1} x}$$

$$\boxed{\mathrm{cov}(\hat{\theta}) = (H^T C^{-1} H)^{-1}}$$

If in addition $W \sim N(0, C)$ then $\hat{\theta}$ given above is also MVUE
(also attains CRLB)

$\mathrm{Var}(\hat{\theta}) = E\left[\left(\sum a_n x_n - E\left(\sum a_n x_n\right)\right)^2\right]$

$a = [a_0 \cdots a_{N-1}]$

$\mathrm{Var}\,\hat{\theta} = E\left(\left(a^T x - a^T E(x)\right)^2\right)$

$= E\left(a^T (x - Ex)(x - Ex)^T a\right)$

$= a^T \mathrm{cov}(x)\, a$ & $a^T s = 1$

$\mathcal{L} = a^T C a + \lambda(a^T s - 1)$

$\frac{\partial \mathcal{L}}{\partial a} = 2Ca + \lambda s = 0$ & $a^T s = 1$

$\hookrightarrow \quad a = \frac{-C^{-1}\lambda s}{2}$

$a_{opt} = \frac{C^{-1} s}{s^T C^{-1} s}$ $\quad \hat{\theta} = a_{opt}^T x = \frac{s^T C^{-1} x}{s^T C^{-1} s}$

Ex $x[n] = A + w[n]$, $\mathrm{Var}\, w[n] = \sigma_n^2$, $E(w(n)) = 0$

$\hat{A} = \frac{\mathbb{1}^T C^{-1} x}{\mathbb{1}^T C^{-1} \mathbb{1}}$ $\quad C = \begin{bmatrix} \sigma_0^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_N^2 \end{bmatrix}$

$\hat{A} = \frac{\sum \frac{x(n)}{\sigma_n^2}}{\sum \frac{1}{\sigma_n^2}}$, $\mathrm{Var} = \frac{1}{\sum 1/\sigma_n^2}$

Least Square

$J = \underline{\|y - Hx\|_2^2}$ $\qquad x \in \mathbb{R}^{n\times 1}$, $\quad y \in \mathbb{R}^{N\times 1}$

$$\frac{\partial J}{\partial x} = \hat{x} H^T H - y^T H = 0 \qquad H^T H \hat{x} = H^T y$$

for $\{H^T H > 0\}$ minima

Fact: A vector $\hat{x}$ is the minimizer of $f(x)$ iff it satisfies the normal eq^{n}s $H^T H x = H^T y$

The minimum is $\|y\|^2 - \|H\hat{x}\|^2$

$R \to$ col span, $N \to$ Null space

$$N(H) = (R(H^T))^{\perp} \qquad N(H) \oplus R(H^T) = \mathbb{R}^n$$

$$N(H^T) = (R(H))^{\perp} \qquad N(H^T) \oplus R(H) = \mathbb{R}^n$$

$$R(H) = N(H^T)^{\perp}$$

Exercise: $R(H^T H) = R(H^T)$

$$R(H^T) = N(H)^{\perp}$$

$N(H^T H) \oplus R(H^T H) = \mathbb{R}^n$

$R(H^T H) \subseteq R(H^T)$

$v \in R(H^T)$ & $v \in N(H^T H)$

$v = H^T b \qquad H^T \underbrace{H H^T b}_{Hu \in N(H^T)} = 0$

Fact: Consider $H^T H \hat{x} = H^T y$

(a) when H is full col. rank, the unique sol. is $\hat{x} = (H^T H)^{-1} H^T y$

(b) When H is not full rank, the normal equations have more than one solution where any two solutions $\hat{x}_1$ & $\hat{x}_2$ satisfy $H(\hat{x}_1 - \hat{x}_2) = 0$

(c) The projection of y onto $R(H)$ is unique $\hat{y} := H\hat{x}$ where $\hat{x}$ is any solution of the normal eqns.

if H is full rank $\underline{\hat{y} = H(H^T H)^{-1} H^T y}$

$H^T H x = H^T y$

QR decomposition

$A = Q R$ (square)

Q: Ortho-normal or unitary; R: upper triangular

$H^T y = H^T H x$

$(H^T H^{-1}) H^T y = x$

For full column rank $A = [a_1 \cdots a_n]$

GSOP : $(e_1 \cdots e_n)$

$Q = [e_1 \; e_2 \; \cdots \; e_n]$

$$R = \begin{bmatrix} \langle e_1, a_1\rangle & \langle e_1, a_2\rangle & \langle e_1, a_3\rangle & \cdots & \langle e_1, a_n\rangle \\ 0 & \langle e_2, a_2\rangle & \approx & & \vdots \\ \vdots & 0 & 0 & & \vdots \\ 0 & \vdots & \vdots & & \vdots \\ & 0 & 0 & & \end{bmatrix}$$

$\begin{matrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{matrix} = a_1$

Cholesky Decomposition

$A = L L^*$

A: Hermitian Positive Definite matrix

L: Lower Triangular matrix

Unique if PD

Non unique if PSD

Singular Value Decomposition

$$\underset{m\times n}{M} = \underset{m\times m}{U} \underset{m\times n}{\Sigma} \underset{n\times n}{V^*}$$

Σ: Rect. diagonal → values are called singular values

U, V^*: Complex unitary

$H = QR$

$$H^T H \hat{x} = H^T y$$

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

$$(Q^T Q) R \hat{x} = (R^T)^{-1} R^T Q^T y$$

$$R \hat{x} = Q^T y$$

Projection Definition:

Let L be a subspace of an inner product space V and let $y \in V$, The projection of y into L is defined to be the unique $\hat{y}_L \in L$

$$\langle y - \hat{y}_L, a \rangle = 0 \quad \forall\ a \in L$$

[Uniqueness & Existence] is skipped

Fact : $\|y - \hat{y}_L\|^2 < \|y - a\|^2 \quad \forall\ a \in L$

Regularized Least Square

$$J(x) = (x - x_0)^T \Pi_0^{-1} (x - x_0) + \| y - Hx \|^2$$

$\Pi_0 > 0$ $\qquad$ $\Pi_0 = \Pi_0^{1/2} \cdot \Pi_0^{1/2}$

$x' = x - x_0$, $y' = y - Hx_0$

$$\min \left\| \begin{bmatrix} 0 \\ y' \end{bmatrix} - \begin{bmatrix} \Pi_0^{1/2} \\ H \end{bmatrix} x' \right\|^2$$

Solution

$$\hat{x} = x_0 + [\Pi_0^{-1} + H^T H]^{-1} H^T (y - Hx_0)$$

{ Condition number }

$$(A + BCD)^{-1} = A^{-1} - A^{-1} B (C^{-1} + D A^{-1} B)^{-1} D A^{-1}$$

$$[A + BCD]^{-1} = A^{-1} - A^{-1} B (C^{-1} + D A^{-1} B)^{-1} D A^{-1}$$

$$\underbrace{[\pi_0^{-1} + H^T H}_{P_i^{-1}} + h_i^T h_i]^{-1} [H y + h_i y_i]$$

$$H_i = \begin{bmatrix} H \\ h_i \end{bmatrix}$$

$$Y_i = \begin{bmatrix} y \\ y_i \end{bmatrix}$$

$$P_{i+1}^{-1} = P_i^{-1} + h_i^T h_i$$

$$P_{i+1} = [\underbrace{P_i^{-1}}_{A} + \underbrace{h_i^T h_i}_{B\,C\,D}]^{-1}$$

$$P_{i+1} = P_i - \frac{P_i h_i^T h_i P_i}{1 + h_i P_{i-1} h_i^T}$$

$$\rightsquigarrow \quad \hat{x}_i = \hat{x}_{i-1} + K_{p,i} (y(i) - h_i \hat{x}_{i-1})$$

$$K_{p,i} = P_{i-1} h_p^T r_e^{-1}(i)$$

$$r_{e,i} = 1 + h_i P_{i-1} h_i^T$$

$$P_i = P_{i-1} - P_{i-1} h_i^T r_{e,i}^{-1} h_i P_{i-1}$$

$x_{j+1} = x_j$

$y_j = h_j x_j + v(j)$

$E(x_0 x_0^T) = \pi_0$

$E(v(i) v(j)) = \delta_{ij}$

Difference Ricatti Equation

Order Recursion: when x also increases in size.

Stochastic Least Squares

$$X \rightarrow [\;]_{N \times 1} \qquad Y \rightarrow [\;]_{p \times 1}$$

Fact: The estimator $\hat{x} = h(y)$ that solves $\min\limits_{h(\cdot)} E([x-\hat{x}][x-\hat{x}]^T)$ least mean squared (lms) estimator is given by $\hat{x} = E[x|y]$

$$\begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix} = \begin{bmatrix} K_0 \\ n \times p(N+1) \end{bmatrix} \begin{bmatrix} y_0 \\ \vdots \\ \hline y_1 \\ \hline \vdots \\ \hline y_N \end{bmatrix}$$

Problem: Find K_0 s.t. for every $K \in \mathbb{R}^{n \times p(N+1)}$

$$\underbrace{E[x - Ky][x - Ky]^T}_{= p(K)} \geq \underbrace{E(x - K_0 y)(x - K_0 y)^T}_{= p(K_0)} \quad \text{—} \quad (1)$$

Thm: LLMS (Linear Least Mean Square)

Given two zero mean rvs x & y, the llms estimator of x given y satisfying (1) is given by any sol K_0 of the normal eqns $K_0 \underbrace{R_y}_{E y y^T} = \underbrace{R_{xy}}_{E x y^T}$

$$P(K_0) = E\{[x - K_0 y][x - K_0 y]^T\}$$

$$= R_x - K_0 R_{yx} - R_{xy} K_0^T + K_0 R_y K_0^T$$

$$P(K_0) \leq P(K) \quad \forall K$$

$$a\,[P(K_0) - P(K)]\,a^T \leq 0 \quad \forall a$$

$$a\,P(K)\,a^T = a R_x a^T - a R_{xy} (aK)^T - aK R_{yx} a^T + aK R_y (aK)^T \quad \forall a$$

$$\frac{\partial (a P(K) a^T)}{\partial (aK)} = 0 \quad \forall a$$

$$\Rightarrow \quad -a R_{xy} - a R_{xy} + 2 a K R_y = 0 \quad \forall a$$

$$R_{xy} = K_0 R_y$$

What if R_y is not invertible? $E(y)=0$

$$a^T (E(y y^T)) = 0$$

$$\Rightarrow \quad E(a^T y) = 0 \;\&\; Var(a^T y) = 0$$

$$a^T y = 0 \text{ a.s.}$$

One variable is purely dependent & can be dropped

Thm: Even if R_y is singular, the normal eq^ns are consistent and there will be multiple solutions no matter which solution is used $\hat{x} = K_0 y$ & $P(K_0)$ can will be unique

Modules set V, ring of "scalars" S

$$x : \Omega \to \mathbb{R}^n, \; z : \Omega \to \mathbb{R}^p$$

$$E(x z^T) \in \mathbb{R}^{n \times p} \quad \leftarrow \quad \langle x | z \rangle$$

$x, y, z \in \Omega \to \mathbb{R}^n$

Commutative, Associativity, scalar product bilinear, associativity with scalar product

Inner Product $\langle x, y \rangle$

Linear in 2^nd, conjugate symmetry

$$\langle \alpha_1 x, y \rangle = \alpha_1 E(x y^T)$$

$$x y^T \overset{?}{=} (y x^T)^T$$

Reflexivity : $\langle x, y \rangle = \langle y, x \rangle^T$

$$\langle x, x \rangle \geq 0 \quad = 0 \text{ iff } x = 0$$

Idea of Projection: $\langle x - K_0 y \,|\, K y \rangle = 0 \quad \forall K$

$$\Rightarrow \quad E((x - K_0 y) y^T) K^T = 0 \quad \forall K$$

$$\Rightarrow \quad E((x - K_0 y) y^T) = 0$$

$$y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix} \quad \Rightarrow \quad E((x - K_0 y) y_i^T) = 0 \quad \forall i = 0, 1, \cdots, N$$

Example: Find LLMS: $\{y(t)\} \longrightarrow$ zero mean stationary

$$\langle y(t), y(t-\tau) \rangle = R_y(\tau)$$

Estimate $\int_0^T y(t)\,dt$ from $y(0), y(T)$

$$z = a\,y(0) + b\,y(T)$$

$$\Rightarrow \left\langle \int_0^T y(t)\,dt - a\,y(0) - b\,y(T) \,\middle|\, y(0) \right\rangle = 0$$

$$| y(T) \rangle = 0$$

$$E \int_0^T y(t)\,y(0)\,dt = \int_0^T R_y(t)\,dt$$

$$\Rightarrow \int_0^T R_y(t)\,dt - a\,R_y(0) - b\,R_y(T) = 0$$

$$\int_0^T R_y(t)\,dt - a\,R_y(T) - b\,R_y(0) = 0$$

$$a = b = \frac{\int R_y(t)\,dt}{R_y(0) + R_y(T)}$$

$$\hat{z} = a\,(y(0) + y(T))$$

Linear Model $y = Hx + v$ → x & v are uncorrelated

$$y \in \mathbb{R}^p, \quad H \in \mathbb{R}^{p \times n}, \quad x \in \mathbb{R}^n$$

$$R_y = H R_x H^T + R_v$$

$$R_{xy} = R_x H^T \qquad E\,xy^T = x x^T H^T + \cancel{x v^T}$$

$$K_0 = R_{xy} R_y^{-1}$$

$$= R_x H^T [H R_x H^T + R_v]^{-1}$$

$$= [R_x^{-1} + H^T R_v^{-1} H]^{-1} H^T R_v^{-1}$$

$$y = H\theta + v \longrightarrow \hat{\theta} = (H^T C^{-1} H)^{-1} H^T C^{-1} y$$

Stochastic Process Estimation

$\{s_i\}_{i=0}^N$ ← Signal process, $i \in \{0, \dots, N\}$ $\qquad \{y_i\}_{i=0}^N$ ← Estimate

zero mean discrete time process

Finite Horizon filter [FIR]

Finite Impulse response

Three problems: 1) Smoothing

For each i, find $\hat{s}_{i|N} = \sum_{j=0}^{N} k_{s,ij}\, y_j$

$$\min_k E\| s_i - \hat{s}_{i|N} \|^2$$

2) Causal Filtering

$$\hat{s}_{i|i} = \sum_{j=0}^{i} k_{f,ij}\, y_j$$

$$\min_{K_f} E\|s_i - \hat{s}_{i|i}\|^2 \qquad \{K_{f,ij}=0 \; ; \; j>i\}$$

3) Prediction

$$\hat{s}_{i+\lambda|i} = \sum_{j=0}^{i} K_{\lambda,ij}\, y_i$$

$$\min_{K_\lambda} E\| s_{i+\lambda} - \hat{s}_{i+\lambda|i}\|^2$$

Smoothening:

$$(s_i - \hat{s}_{i/N}) \perp y_e \quad , i=0, \dots N$$

$$(s_i - \sum_{j=0}^{N} K_{s,ij}\, y_j) \perp y_e \quad , i=0,\dots,N$$

$$[R_{sy}(i,0), \dots, R_{sy}(i,N)] = [K_{s,i0}, \dots\dots K_{s,iN}]$$

$$R_{sy} = K_s R_y \qquad \begin{bmatrix} R_y(0,0) & \cdots & R_y(0,N) \\ \vdots & & \vdots \\ R_y(N,0) & \cdots & R_y(N,N) \end{bmatrix}$$

Causal filtering

For each i $\quad (s_i - \sum_{j=0}^{i} K_{f,ij}\, y_i) \perp y_e \quad , i=0,\dots i$

$i=0$, $\quad R_{sy}(0,0) = K_{f00}\, R_y(0,0)$

$i=1$, $\quad R_{sy}(1,0) = K_{f10}\, R_y(00) + K_{f11}\, R_y(1,0)$

N, $\quad R_{sy}(1,1) = K_{f10}\, R_y(0,1) + K_{f11}\, R_y(1,1)$

$$\begin{bmatrix} R_{sy}(0,0) & ? \\ R_{sy}(1,0) & R_{sy}(1,1) \end{bmatrix} = \begin{bmatrix} K_{f,00} & 0 \\ K_{f,10} & K_{f,11} \end{bmatrix} \begin{bmatrix} R_y(0,0) & R_y(0,1) \\ R_y(1,0) & R_y(1,1) \end{bmatrix}$$

Wiener Hopf Technique

Cholesky Decomposition $\rightarrow$ 1) $\quad A \in \mathbb{R}^{n\times n} > 0$

$$A = \begin{bmatrix} A_{11} & A^T_{2:n,1} \\ A_{2:n,1} & A_{2:n,2:n} \end{bmatrix}$$

Schur's complement: $\quad S = A_{2:n,2:n} - \frac{1}{A_{11}} A_{2:n,1} A^T_{2:n,1}$

Fact: if A is PD $\Leftrightarrow S>0$

Proof: Any $x \neq 0$, $\quad y = -(A^T_{2:n,1}\, x)/A_{11}$

$$x^T S x = \begin{bmatrix} y \\ x \end{bmatrix}^T \begin{bmatrix} A_{11} & A^T_{2:n,1} \\ A_{2:n,1} & A_{2:n,2:n} \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} > 0$$

Fact: $A = R^T R$ $\quad R \to$ upper triangular with the diagonal elements

LDL^T ← Not doing sqrt can do better wrt time

$$\begin{bmatrix} A_{11} & A^T_{2:n,1} \\ A_{2:n,1} & A_{2:n,2:n} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R^T_{1,2:n} & R^T_{2:n,2:n} \end{bmatrix} \begin{bmatrix} R_{11} & R_{1,2:n} \\ 0 & R_{2:n,2:n} \end{bmatrix}$$

$R_{11} = \sqrt{A_{11}}$ ← LDL^T can get rid of it.

$R_{1,2:n} = \frac{1}{R_{11}} A_{1,2:n}$

$$R^T_{2:n,2:n} R_{2:n,2:n} = \underbrace{A_{2:n,2:n} - R^T_{1,2:n} R_{1,2:n}}_{\text{Schur Complement}}$$

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 8 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & R_{22} & 0 \\ -1 & R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}$$

$$\begin{bmatrix} 8 & 0 \\ 0 & 11 \end{bmatrix} - \begin{bmatrix} 9 & -3 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} R_{22} & 0 \\ R_{23} & R_{33} \end{bmatrix} \begin{bmatrix} R_{22} & R_{23} \\ 0 & R_{33} \end{bmatrix}$$

Can show uniqueness

$LL^T = MM^T$

$L^{-1} L L^T L^{-T} = L^{-1} M M^T L^{-T}$

$I = (L^{-1}M)(L^{-1}M)^T$

$LM = (L^{-1}M)^{-T} \Rightarrow L^{-1}M = \mathcal{D}$ & $\mathcal{D}^2 = I$

$L = M$ if diagonal is forced > 0.

$$\begin{bmatrix} R_{sy}(0,0) & \times \\ R_{sy}(1,0) & R_{sy}(1,1) \end{bmatrix} = \begin{bmatrix} K_{f\,00} & 0 \\ K_{f\,10} & K_{f\,11} \end{bmatrix} \begin{bmatrix} R_y(0,0) & R_y(0,1) \\ R_y(1,0) & R_y(1,1) \end{bmatrix}$$

$$[R_{sy} - K_f \cdot R_y]_{\text{lower}} = 0$$

$$R_{sy} - K_f L \underset{\substack{\uparrow \\ R_e}}{D} L^T = U^+$$

R_e {Diagonal matrix}

$$\underbrace{K_f L}_{\text{lower}\,\triangle} = \underbrace{R_{sy} L^{-T} R_e^{-1}}_{\text{mixed}} - \underbrace{U^+ L^{-T} R_e^{-1}}_{\text{strictly upper triangular}}$$

$$\boxed{K_f = \{R_{sy} L^{-T} R_e^{-1}\}_{\text{lower triangular}} L^{-1}}$$

Wiener Hopf technique solves causal filter problem

Example: $y = s + v$ $\qquad \left\langle \begin{bmatrix} s \\ v \end{bmatrix}, \begin{bmatrix} s \\ v \end{bmatrix} \right\rangle = \begin{bmatrix} R_s & 0 \\ 0 & R_v \end{bmatrix}$ ← diagonal

$$R_{sy} = R_s$$

$$R_y = R_s + R_v \qquad K_f = \left\{ [R_y - R_v] L^{-T} R_e^{-1} \right\}_{lower} L^{-1}$$

$$R_y = L R_e L^T$$

$$= I - \underbrace{\left\{ R_v L^{-T} R_e^{-1} \right\}_{lower}}_{R_v R_e^{-1}} L^{-1}$$

$$= I - R_v R_e^{-1} L^{-1}$$

$$\Rightarrow \hat{s} = y - R_v R_e^{-1} \underbrace{L^{-1} y}$$

↑ Innovations process

$$e_{N+1} = y_{N+1} - \mathrm{Proj}\left(y_{N+1} / \alpha_N\right)$$

$$= y_{N+1} - \sum_{j=0}^{N} \langle y_{N+1}, e_j \rangle \, \|e_j\|^{-2} e_j$$

e_i

↑ Innovations process

$$y_{N+1} = e_{N+1} + \sum_{i=0}^{N} \langle y_{N+1}, e_i \rangle \, \|e_i\|^{-2} e_i$$

Properties:

1) e_i is uncorrelated with e_j $\quad i \neq j$

2) $e_i \in \mathcal{L}\{y_0, \cdots, y_i\}$

$y_i \in \mathcal{L}\{e_0, \cdots, e_i\}$

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \langle y_1, e_0 \rangle \|e_0\|^{-2} & 1 & 0 \\ \langle y_2, e_0 \rangle \|e_0\|^{-2} & \langle y_2, e_1 \rangle \|e_1\|^{-2} & 0 \\ \cdots & \cdots & 1 \end{bmatrix} \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_N \end{bmatrix}$$

$$y = L_1 e \qquad \Rightarrow R_y = L_1 R_e L_1^T$$

$$\rightsquigarrow L = L_1$$

$$\xrightarrow{y} \boxed{L^{-1}} \xrightarrow{e} \boxed{L} \xrightarrow{y}$$

whitening filter $\qquad$ Modelling filter

$$\hat{s}_{i|i} = \text{llmse given } \{y_0, \cdots, y_i\}$$

$$= \text{llmse given } \{e_0, \cdots, e_i\}$$

$$= \sum_{j=0}^{i} \overbrace{\underbrace{\langle s_i, e_j \rangle}_{R_{se}(i,j)} \|e_j\|^{-2}}^{g_f(i,j)} e_j$$

$$g_f(i,j) = \begin{cases} R_{se}(i,j) \; \|e_j\|^{-2} & j \le i \\ 0 & \text{otherwise} \end{cases}$$

$$R_{se} = \langle s, e \rangle = \langle s, y \rangle L^{-T}$$
$$= R_{sy} L^{-T}$$

$$G = \{R_{sy} L^{-T} R_e^{-1}\}_{\text{lower}}$$

$$\hat{s}_{i|i} = G e = \{R_{sy} L^{-T} R_e^{-1}\}_{\text{lower}} L^{-1} y$$

Expectation Maximization

$$X \to \{x(0), \dots, x(n-1)\} \to [\theta_1, \dots, \theta_p]$$

$$\max_{\theta} \ln p_x(x, \theta) \qquad x = g(y)$$

$$x[n] = \sum_{i=1}^{p} \cos 2\pi f_i n + w(n)$$
$$\theta = [f_1 \dots f_p]$$

$$\longrightarrow \max_{f_i} J(f) = \sum_{n=0}^{N-1} \left(x(n) - \sum_{i=1}^{p} \cos 2\pi f_i n \right)^2$$

$$\left[\frac{\partial J}{\partial f_i} \; \dots \; \frac{\partial J}{\partial f_p} \right] = 0$$

$$E_{y/x}\left(\ln p_y(y, \theta) \right) = g(x, \theta)$$

X
↳ observed data

θ
↳ Parameter

Y state variable
Latent variable

Assume $p(X|Y)$ is not a function of θ

Harder : $\max_{\theta} \ln p(x;\theta)$

Easier : $\max_{\theta} \ln p(y;\theta)$

$$E_{Y/X}[\ln p(y;\theta)] = \int \ln p(y;\theta)\, p(y/x;\theta)\, dy$$

$$E \rightarrow \mathcal{U}(\theta, \theta_K) = \underbrace{\int \ln p(y;\theta)\, p(Y/x;\theta_K)\, dy}_{E^{\theta_K}_{Y/X}[\ln p(y;\theta)]}$$

$$M \rightarrow \theta_{k+1} = \arg\max_{\theta} \mathcal{U}(\theta, \theta_K)$$

Example: $x[n] = \sum_{i=1}^{p} \cos 2\pi f_i n + w[n]$; $n = 0, 1, \cdots, N-1$

$$y_i[n] = \cos 2\pi f_i n + w_i[n] \quad \begin{cases} i = 1, \cdots, p \\ n = 0, \cdots, N-1 \end{cases}$$

$$U(f_i, f_{iK}) = \sum_{n=0}^{N-1} \left[\cos 2\pi f_{iK} n + \beta_i \left[x[n] - \sum_{i=1}^{p} \cos 2\pi f_{iK} n\right]\right] \cdot \cos 2\pi f_i n$$

$$f_{i_{K+1}} = \arg\min_{f_i} \mathcal{U}(f_i, f_{iK})$$

Jensen's Inequality

Fact: Let f be a convex f^n defined on I if $x_1, x_2 \cdots \in I$ & $\lambda_1, \lambda_2, \cdots, \lambda_n \geq 0$; $\sum_{i=1}^{n} \lambda_i = 1$

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i)$$

$$f(Ex) \leq E(f(x))$$

$$L(\theta) - L(\theta_K) = \ln \frac{p(x;\theta)}{p(x;\theta_K)}$$

$$= \ln \int \frac{p(y, x;\theta)}{p(x, \theta_K)}\, dy$$

$$= \ln \int \frac{p(y, x;\theta)}{p(y, x;\theta_K)} \cdot p(y/x;\theta_K)\, dy$$

{using $p(x/y)$ is ind. of θ}

$$= \ln \int \frac{p(y;\theta)\ p(x/y)}{p(y;\theta_K)\ p(x, y)}\, p(y|x;\theta_K)\, dy$$

{using Jensen's inequality}

$$\geq \int \left[\ln\left(\frac{p(y;\theta)}{p(y;\theta_K)}\right)\right] p(y/x;\theta_K)\, dy$$

$$= E^{\theta_K}_{y/x}[\ln p(y;\theta)] - (\quad)$$

z transform — stability & causality of a filter.

$\{ l_{-\infty} ; l_{-1}, l_0, l_1, \dots, l_{\infty} \}$

$$z \text{ transform} = \sum_{i=-\infty}^{\infty} l_i z^{-i}$$

Causal : $l_- = 0$

L is doubly infinite:

$$\Delta(z) = [z^{\infty} \dots z^2 \; z \; \boxed{1} \; z^{-1} \; z^{-2} \dots z^{-\infty}]$$

$$S_y(z) = \sum_{i=-\infty}^{\infty} R_y(i) z^{-i}$$

↑ z spectra

$$L(z) = 1 + \sum_{i=1}^{\infty} l_i z^{-i}$$

Fact: $\Delta(z) R_y = S_y(z) \Delta(z)$

$$R_y = \begin{bmatrix} \ddots & & & \\ & R_y(0) & R_y(1) & \\ R_y(1) & \boxed{R_y(0)} & & R_y(1) \\ & & & \ddots \end{bmatrix}$$

Fact: $\Delta(z) L = \underbrace{L(z)}_{\text{scalar}} \text{ re } \Delta(z) L^T$

$$L(z) = 1 + c_1 z^{-1} + c_2 z^{-2} + \dots$$
$$= L(z) \text{ re } (1/z)$$

Cannonical spectral factorization → diagonal $=1$ condition $\equiv L(\infty) = 1$

$$S_y(z) = L(z) \text{ re } L^*(1/z^*)$$

↳ stable, causal, minimum phase

↑ zeros are stable

Defⁿ (cannonical spectral Fact)

Let $S_y(z)$ be a rational z-spectrum of a finite power process and assume further that $S(e^{j\omega}) \geq 0$ $\forall \omega$,

Then the cannonical spectrum factorization of $S_y(z)$ is ($re > 0$, $L(\infty) = 1$)

$$S_y(z) = L(z) \text{ re } L^*(1/z^*)$$

where $L(z)$ is stable BIBO & minimum phase

Assume $S_y(z) = \sum_{i=-\infty}^{\infty} R_y(i) z^{-1}$

$|R_y(i)| < k \alpha^{|i|}$ $\quad 0 < \alpha < 1$ & $k > 0$

ROC of $S_y(z)$: $\alpha < |z| < 1/\alpha$

Fact: $S_y(z) = S_y^*\left(\frac{1}{z^*}\right)$

For every pole (or zero) at $z = \beta$ $\exists$ a pole (resp. zero) at $z = \frac{1}{\beta^*}$

$$S_y(z) = r_e \frac{\prod_{i=1}^{m} (z-\alpha_i)\left(\frac{1}{z} - \alpha_i^*\right)}{\prod_{i=1}^{n} (z-\beta_i)\left(\frac{1}{z} - \beta_i^*\right)}$$

with $|\alpha_i| < 1$, $|\beta_i| < 1$, $r_e > 0$

$$L(z) = z^{n-m} \frac{\prod_{i=1}^{m} (z-\alpha_i)}{\prod_{i=1}^{n} (z-\beta_i)}$$

zero mean WSS

$$u_i \longrightarrow \boxed{H(z)} \longrightarrow y_i$$

$\rightarrow$ zero mean WSS

$\rightarrow S_y(z) = H(z)\, S_u(z)\, H^*\left(1/z^*\right)$

$\rightarrow S_{yu}(z) = H(z)\, S_u(z)$

$$y_i \longrightarrow \boxed{L^{-1}(z)} \xrightarrow{e_i} \boxed{L(z)} \longrightarrow y_i$$

Canonical Spectral Factorisation:

BIBO stable + minimum phase + $L(\infty) = 1$

$$S_y(z) = L(z)\; r_e\; L^*(z^{-*})$$

rational, finite power, $S_y(e^{i\omega}) > 0 \;\forall \omega$

Example

ARMA: $y_{i+1} = a_0 y_i + a_1 y_{i-1} + u_i + b u_{i-1}$

$zY(z) = a_0 Y(z) + a_1/z\, Y(z) + U(z) + b\, U(z)/z$

$$u_i \longrightarrow \boxed{H(z)} \longrightarrow y_i$$

$(z - a_0 - a_1/z)\, Y(z) = U(z)\, (1 + b/z)$

$\Rightarrow H(z) = \dfrac{z + b}{z^2 - a_0 z - a_1}$

$\langle u_i, u_j \rangle = Q\, \delta_{ij}$

$\{u_i\} = \sum_{i=0}^{\infty} z^{-i} u_i$

$\{x_{i+1}\} = \sum_{i=0}^{\infty} z^{-i} u_{i+1}$

$= z \sum_{i=0}^{\infty} z^{-i-1} u_{i+1}$

$= z\, U(z)$

$$S_y = H(z)\, Q\, H^*(z^{-*})$$

$$= \left(\frac{z+b}{z^2 - a_0 z - a_1}\right) Q \left(\frac{1/z + b^*}{z^{-2} - a_0 z^{-1} - a_1}\right)$$

$L(z) = \dfrac{z(z+b)}{z^2 - a_0 z - a_1}$ if $|b| < 1$; $L(z) = \dfrac{z(z + 1/b)}{z^2 - a_0 z - a_1}$ if $|b| > 1$

Smoothing l.l.m.s.e. $\hat{s}_i = \sum_{m=-\infty}^{\infty} w_{im} y_m$

Estimate $\{s_i\}$ from $\{y_m\}_{m=-\infty}^{m=\infty}$

$$(s_i - \hat{s}_i) \perp y_l \quad -\infty < l < \infty$$

$$R_{sy}(i-l) = \sum_{-\infty}^{\infty} R_y(m-l) w_{im} \quad -\infty < i < \infty$$

$m - l = m'$, $i - l = i'$

$$R_{sy}(i') = \sum_{m'=-\infty}^{\infty} \underbrace{w_{i'+l, m'+l}}_{K(i'-m')} R_y(m') \quad -\infty < i < \infty$$

$$R_{sy}(i) = \sum_{m=-\infty}^{\infty} w_{i'+l_2, m'+l_2} R_y(m')$$

$$0 = \sum_{m'=-\infty} (w_{i'+l, m'+l} - w_{i'+l_2, m'+l_2}) R_y(m')$$

$$R_{sy}(i) = \sum_{m=-\infty}^{\infty} k_{i-m} R_y(m)$$

Theorem: Given two discrete time zero mean jointly stationary random processes $\{s_i, y_i\}$, l.l.m.s.e. smoother of s_i given y_i is the LTI filter

$$\begin{cases} K(z) = \dfrac{S_{sy}(z)}{S_y(z)} \\ K(e^{j\omega}) = \dfrac{S_{sy}(e^{j\omega})}{S_y(e^{j\omega})} \end{cases}$$

Discrete time wiener Hopf filter

$$\hat{s}_{i|i} = \sum_{m=-\infty}^{i} w_{im} y_m$$

$(s_i - \hat{s}_{i|i}) \perp y_l$ for $-\infty < l \le i$

$i' = i - l$, $m' = m - l$

$$R_{sy}(i') = \sum_{m=-\infty}^{i'+l} w_{i'+l, m} R_y(m-l), \quad i' \ge 0$$

$$R_{sy}(i') = \sum_{m'=-\infty}^{i'} w_{i'+l, m'+l} R_y(m'), \quad i' \ge 0$$

$l = -m'$

$$R_{sy}(i') = \sum_{m'=-\infty}^{i'} K(i'-m) R_y(m'), \quad i' \ge 0$$

$$R_{sy}(i) = \sum_{m=0}^{\infty} K(m) R_y(i-m), \quad i \ge 0$$

$$\hat{s}_{i|i} = \sum_{m=0}^{\infty} k_m y_{i-m}$$

$$g_i := R_{sy}(i) - \sum_{m=0}^{\infty} k_m R_y(i-m) \qquad -\infty < i < \infty$$

We want g_i to be strictly anticausal.

$\nwarrow g_i = 0 \quad \forall\, i \geq 0$

double sided z-transforms

$$G(z) = S_{sy}(z) - K(z)\, S_y(z)$$

$$S_y(z) = L(z)\, r_e\, L^*\left(\frac{1}{z^*}\right)$$

$$\left[\frac{\overset{\text{strictly anticausal}}{G(z)}}{\underset{\text{anticausal}}{\left(r_e L^*(z^{-*})\right)}}\right]_{\text{strictly anticausal}} = \underbrace{\frac{S_{sy}(z)}{r_e L^*(z^{-*})}} - \underbrace{K(z) L(z)}_{\text{Causal}}$$

$$K(z)\, L(z) = \left(\frac{S_{sy}(z)}{r_e L^*(z^{-*})}\right)_{\text{causal}}$$

$$\Rightarrow \quad K(z) = \left(\frac{S_{sy}(z)}{r_e L^*(z^{-*})}\right)_{\text{causal part}} L^{-1}(z)$$

Thm (Wiener Filter):

Consider two zero mean jointly WSS scalar random process $\{s_i\}$, $\{y_i\}$ with rational z-spectra & z-cross spectra $S_s(z)$, $S_y(z)$, $S_{sy}(z)$

Assuming Further, $S_y(z)$ has no unit circle zeros, then the llms filter is

$$K(z) = \left\{\frac{S_{sy}(z)}{r_e L^*(1/z^*)}\right\}_+ \frac{1}{L(z)}$$

Three Rules: (a) $\{c\}_+ = c$ {constant}

(b) $$\left\{\frac{1}{z+\alpha}\right\}_+ = \begin{cases} 1/(z+\alpha) & |\alpha| < 1 \\ 1/\alpha & |\alpha| > 1 \end{cases}$$

(c) $$\left\{\frac{1}{(z+\alpha)^i}\right\}_+ = \begin{cases} 1/(z+\alpha)^i & |\alpha| < 1 \\ 1/\alpha^i & |\alpha| > 1 \end{cases}$$

$$F(z) = \{r_0\} + \sum_{i=1}^{m} \sum_{k=1}^{l} \frac{r_{ik}}{(z-p_i)^k}$$

$$\{F(z)\}_+$$

$$F(z) = \frac{4z+3}{z^2 + 7/3\, z + 2/3} = \frac{3}{z+2} + \frac{1}{z+1/3}$$

$$\{F(z)\}_+ = \frac{3}{2} + \frac{1}{(z+1/3)}$$

$$u_i \longrightarrow \boxed{H(z)} \longrightarrow y_i$$

$\{x_i\}$ another zero mean process WSS with x_i, y_i

$$S_{xy}(z) = S_{xu}(z)\; H^*(1/z^*)$$

$$S_{yx}(z) = H(z)\; S_{ux}(z)$$

$$S_{xy}(z) = \sum_{-\infty}^{\infty} R_{xy}(i)\, z^{-i}$$

$$Y(z) = H(z)\; U(z)$$

$$Y^*(z^{-*}) = U^*(z^{-*})\; H^*(z^{-*})$$

$$= \sum_{-\infty}^{\infty} E(x_0 y_i^*)\, z^{-i}$$

$$= E(x_0\, Y^*(1/z^*))$$

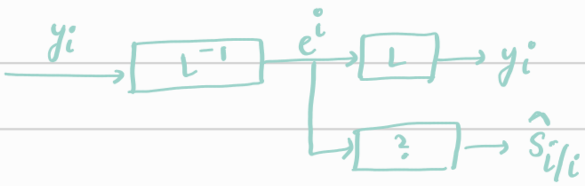
$$= E(x_0\, U^*(z^{-*}))\, H^*(z^{-*})$$

$$= \sum_{i=-\infty}^{\infty} E(x_0 u^*(i))\, H^*(z^{-*})$$

$$= S_{xu}\, H^*(z^{-*})$$

1) Find $\{e_i\}$ from $\{y_i\}$

2) Find $\hat{s}_{i|i}$ based on $\{e_i\}$



$$\langle s_i - \hat{s}_{i|i},\, e_\ell \rangle = 0 \qquad -\infty < \ell \le i$$

$$R_{se}(i-\ell) = \sum_{k=-\infty}^{i} g_{i-k}\, r_e\, \delta_{k\ell} \qquad \hat{s}_{i|i} = \sum_{k=-\infty}^{i} g_{i-k}\, e_k$$

$$R_{se}(i-\ell) = g_{i-\ell}\, r_e$$

$$\Rightarrow g_i = \begin{cases} R_{se}(i)\, r_e^{-1} & \text{for } i \ge 0 \\ 0 & \text{for } i < 0 \end{cases}$$

$$e_i \longrightarrow \boxed{L(z)} \longrightarrow y_i$$

$$S_{se}(z) = \frac{S_{sy}(z)}{L^*(z^{-*})}$$

$$G(z) = \sum_{i=-\infty}^{\infty} g_i z^{-i} = \sum_{i=0}^{\infty} R_{se}(i)\, r_e^{-1} z^{-i}$$

$$= r_e^{-1} \{ S_{se}(z) \}_+$$

$$\hat{s}(z) = G(z)\, e(z) = \frac{1}{r_e} \left\{ \frac{S_{sy}(z)}{L^*(1/z^*)} \right\}_+ \frac{1}{L(z)}\, y(z)$$

Example : $y_i = s_i + v_i$

1) $\{s_i\}, \{v_i\}$ are zero mean jointly WSS & uncorrelated

2) $\langle v_i, v_j \rangle = r\, \delta_{ij}$

$\Rightarrow$ $S_v(z) = r$

$S_{vs}(z) = 0$

$S_{sy}(z) = S_s(z)$

$S_y(z) = S_s(z) + r$

Answer $\longrightarrow$ $\hat{s}(z) = y(z) - \frac{r}{r_e} L^{-1}(z)\, y(z)$

$L(z) r_e L^*(z^{-*})$

$S_y(z) - r$

$$\hat{s}(z) = \frac{1}{r_e} \left\{ \frac{S_{sy}(z)}{L^*(z^{-*})} \right\}_+ \frac{1}{L(z)}\, y(z)$$

$$= \frac{1}{r_e} \left\{ \underset{\downarrow\ r_e L(z)}{r_e L(z)} - \underset{\downarrow\ r}{\frac{r}{L^*(z^{-*})}} \right\}_+ \frac{y(z)}{L(z)}$$

$$= y(z) - \frac{r}{r_e} L^{-1}(z)\, y(z)$$

Kalman Filter

$\hat{s}_{i|i}$ $\quad$ $\hat{x}_{i|i-1}$

we know y_i and need to estimate others.

$$e_{i+1} = y_{i+1} - \hat{y}_{i+1|i}$$

$$= y_{i+1} - \sum_{j=0}^{i} \langle y_{i+1}, e_j \rangle \|e_j\|^{-2} e_j$$

$$\hat{s}_{i+1|i} = \sum_{j=0}^{i} \langle s_{i+1}, e_j \rangle \|e_j\|^{-2} e_j$$

$$= \hat{s}_{i+1|i-1} + \langle s_{i+1}, e_i \rangle \|e_i\|^{-2} e_i$$

$$x_{i+1} = F_i x_i + G_i u_i$$ ← Process noise

$$y_i = H_i x_i + v_i$$ ← Measurement noise

$u_i \in \mathbb{R}^m$, $x_i \in \mathbb{R}^n$

$v_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}^p$

m×m m×p

$$\left\langle \begin{bmatrix} u_i \\ v_i \\ x_0 \end{bmatrix}, \begin{bmatrix} u_i \\ v_i \\ x_0 \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} Q_i \delta_{ij} & S_i \delta_{ij} & 0 & 0 \\ S_i^* \delta_{ij} & R_i \delta_{ij} & 0 & 0 \\ 0 & 0 & \Pi_0 & 0 \end{bmatrix}$$

p×p

Projections are on y_is not x_is.

1) $\langle u_i, x_j \rangle = 0$, $\langle v_i, x_j \rangle = 0$ $j \le i$

2) $\langle u_i, y_j \rangle = 0$, $\langle v_i, y_j \rangle = 0$ $j \le i-1$

3) $\langle u_i, y_i \rangle = S_i$, $\langle v_i, y_i \rangle = R_i$

$$P_{i|i-1} := \langle x_i - \hat{x}_{i|i-1}, x_i - \hat{x}_{i|i-1} \rangle$$

$$\langle x_i, x_i \rangle := \Pi_i$$ $\rightsquigarrow \Pi_i = \Sigma_i + P_{i|i-1}$

$$\langle \hat{x}_{i|i-1}, \hat{x}_{i|i-1} \rangle := \Sigma_{i|i-1}$$

Iteration

$$\Pi_{i+1} = \langle x_{i+1}, x_{i+1} \rangle = \langle F_i x_i + G_i u_i, F_i x_i + G_i u_i \rangle$$
$$= F_i \Pi_i F_i^* + G_i Q_i G_i^*$$

Innovations : $e_i = y_i - \hat{y}_{i|i-1}$ → $H_i \hat{x}_{i|i-1}$

$$y_i = H_i x_i + v_i$$

$$\hat{y}_{i|i-1} = H_i \hat{x}_{i|i-1} + \hat{v}_{i|i-1}$$ $\{\hat{v}_{i|i-1} = 0\}$

$$= H_i \hat{x}_{i|i-1}$$

$$\hat{x}_{i+1|i} = \sum_{j=0}^{i} \langle x_{i+1}, e_j \rangle R_{e,j}^{-1} e_j$$

$$= \hat{x}_{i+1|i-1} + \underbrace{\langle x_{i+1}, e_i \rangle \|e_i\|^{-2}}_{K_{p,i}} e_i$$

($y_i - H_i \hat{x}_{i|i-1}$)

we know : $\hat{x}_{i+1|i-1} = \underbrace{F_i \hat{x}_{i|i-1}} + G_i \cancel{u_{i|i-1}}^{\,0}$

$$\Rightarrow \hat{x}_{i+1|i} = F_i \hat{x}_{i|i-1} + K_{p,i} (y_i - H_i \hat{x}_{i|i-1})$$

$$R_{e,i} = \langle e_i, e_i \rangle$$

$$\tilde{x}_i = x_i - \hat{x}_{i|i-1}$$

$$P_{i|i-1} = \langle \tilde{x}_i | \tilde{x}_i \rangle$$

$$e_i = H_i x_i + v_i - H_i \hat{x}_{i|i-1}$$

$$= H_i \tilde{x}_i + v_i$$

$$H_i (\tilde{x}_i \tilde{x}_i^*) H_i^* + (v_i v_i^*)$$

$$\Rightarrow R_{e,i} = H_i P_i H_i^* + R_i$$

$$\langle x_i, e_i \rangle = \langle x_i, \tilde{x}_i \rangle H_i^* + \langle x_i, v_i \rangle$$

$$= P_i H_i^*$$

$$\tilde{x}_i + \hat{x}_i \qquad \{ \tilde{x}_i \perp \hat{x}_i \}$$

error projection

$$\langle x_{i+1}, e_i \rangle = F \langle x_i, e_i \rangle + G_i \langle u_i, e_i \rangle$$

$P_i H_i^*$; $H \tilde{x}_i + v_i$

$$\langle u_i | \tilde{x}_i \rangle = 0 , \quad \langle u_i, v_i \rangle = S_i$$

$$\Rightarrow \langle x_{i+1}, e_i \rangle = F_i P_i H_i^* + G_i S_i$$

$$K_{p,i} = (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1}$$

$$R_{e,i} = H^* P_i H_i^* + R_i$$

$$\tilde{x}_{i+1} = F_i \tilde{x}_i + G_i u_i - K_{p,i} H_i \tilde{x}_i - K_{p,i} v_i$$

$$= (F_i - K_{p,i} H_i)\tilde{x}_i + \begin{bmatrix} G_i & -K_{p,i} \end{bmatrix} \begin{bmatrix} u_i \\ v_i \end{bmatrix}$$

$$P_{i+1} = [F_i - K_{p_i} H_i] P_i [F_i - K_{p_i} H_i]^* + \begin{bmatrix} G_i & -K_{p,i} \end{bmatrix} \begin{bmatrix} Q_i & S_i \\ S_i^* & R_i \end{bmatrix} \begin{bmatrix} G_i \\ -K_{p,i} \end{bmatrix}$$

$$\longrightarrow P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K_{p,i}^* \; ; \; i \geq 0$$

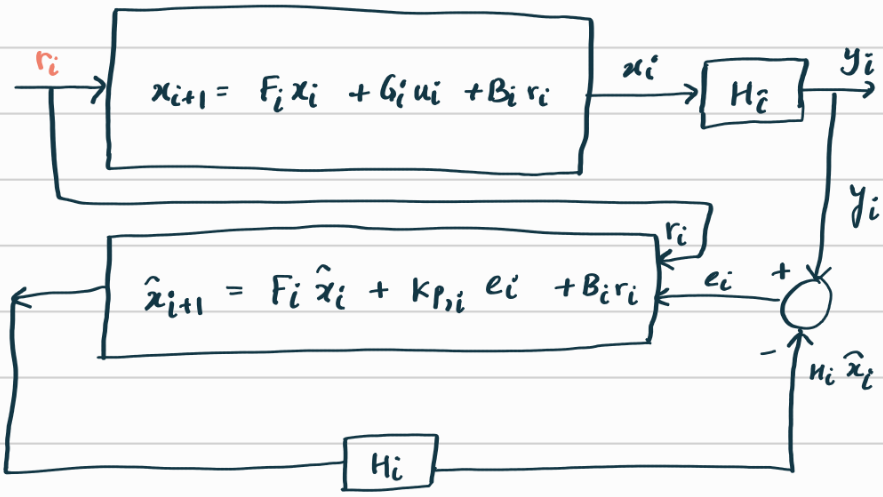
$$P_0 = \langle \tilde{x}_0, \tilde{x}_0 \rangle = \Pi_0$$

Benefits: → Ind of x → order of computation at each step.

→ Forward iterations

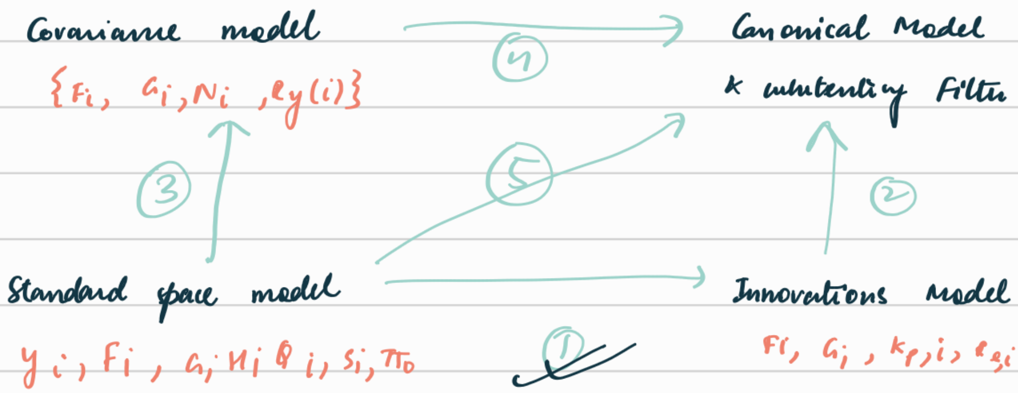
$$O(n^3) + O(n^2)$$

Inversion (n is system order) multiplication



Separation principle

$$\begin{bmatrix} x_{i+1} \\ \hat{x}_{i+1} \end{bmatrix} = \begin{bmatrix} R-BK & * \\ 0 & F-LH \end{bmatrix} \begin{bmatrix} x_i \\ \hat{x}_i \end{bmatrix}$$



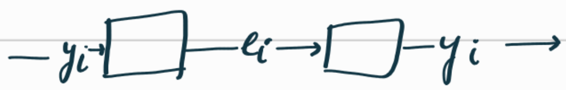
(2)

$$\hat{x}_{i+1} = F_i \hat{x}_i + K_{p,i} e_i$$

$$y_i = H_i \hat{x}_i + e_i$$

$$\rightarrow \quad \hat{x}_{i+1} = F_i \hat{x}_i + K_{p,i} (y_i - H_i \hat{x}_i)$$

$$e_i = -H_i \hat{x}_i + y_i$$



The inverse is also causal.

→ Done

$$\phi_{ij} = \begin{cases} F_{i-1} F_{i-2} \cdots F_j & \{i > j\} \\ I & \{i = j\} \end{cases}$$

$$F_{p,i} = F_i - K_{p,i} H_i$$

$$\phi_p(i,j) = \begin{cases} F_{p,i-1} \cdots F_{p,j} & \{i > j\} \\ I & \{i = j\} \end{cases}$$

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} I & 0 & \cdots & 0 \\ H_1 K_{p,0} & I & & \\ H_2 \phi(2,1) K_{p,0} & H_{2,1} K_{p,1} & \ddots & I \\ H_N \phi_{(N,1)} K_{p,0} & H_N \phi_{N,2} K_{p,2} & \cdots & I \end{bmatrix} \begin{bmatrix} e_0 \\ \vdots \\ e_N \end{bmatrix}$$

$$y = L e$$

$$e = L^{-1} y$$

$$\begin{bmatrix} e_0 \\ \vdots \\ e_N \end{bmatrix} = \begin{bmatrix} I & 0 \\ -H_1 K_{p,0} & I \\ -H_2 \phi_p(2,1) K_{p,0} & \vdots \\ \vdots & \\ -H_N \phi_p(N,1) K_{p,0} & -H_N \phi_p(N,2) K_{p,1} \cdots \end{bmatrix} \begin{bmatrix} y_0 \\ \vdots \\ y_N \end{bmatrix}$$

③: $R_y(i,j) = \langle y_i, y_j \rangle = \begin{cases} H_i \phi(i, j+1) N_j & : i > j \\ H_i \Pi_i H_i^* + R_i & : i = j \\ N_i^* \phi^*(j, i+1) H_j^* & : i < j \end{cases}$

$$N_i = F_i \Pi_i H_i^* + G_i S_i^*$$

$$\Pi_i = \langle x_i, x_i \rangle$$

$$\Pi_{i+1} = F_i \Pi_i F_i^* + G_i Q_i G_i^*$$

converges to $\bar{\Pi} = F \bar{\Pi} F^* + G Q G^*$

$V_i = V_{i+1} + x\, G Q G^*$ ⊕

$V_{i+1} < V_i$

$V_i = x^* P x$

$V_{i+1} = x^* F^* P F x$

$$\hat{x}_{i+1} = F_i \hat{x}_i + K_{p,i} (y_i - H_i \hat{x}_i)$$

$$K_{p,i} = F_i P_i H_i^* R_{e,i}^{-1}$$

$$R_{e,i} = H_i P_i H_i^* + R_i$$

$$P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_{p,i} R_{e,i} K^*_{p,i}$$

$\Rightarrow$ $\begin{cases} K_{p,i} = [N_i - F_i \Sigma_i H_i^*] R_{e,i}^{-1} \\ R_{e,i} = R_y(i,i) - H_i \Sigma_i H_i^* \\ \Sigma_{i+1} = F_i \Sigma_i F_i^* + K_{p,i} R_{e,i} K_{p,i}^* \end{cases}$

can be derived through manipulations & using $\Pi_i = \Sigma_i + P_i$

(II) when time invariant:

Algebraic Riccati Equation

$$P = FPF^* + GQG^*$$

Fact: The discrete time Lyapunov Eqn for $x_{k+1} = Ax_k$;

$A^T PA - P + Q = 0$ has a unique p.d. matrix P for any $Q = Q^T > 0$, iff $x_{k+1} = Ax_k$ is asymptotically stable.

(Sylvester Equation?)

$$x_{k+1} = Fx_k + Gu_k$$

$$y_k = Hx_k + v_k$$

Lyapunov Equation: $\bar{\Pi} = F\bar{\Pi}F^* + GQG^*$

If F is stable & $Q \geq 0$, $\exists$ a unique p.s.d. sol $\bar{\Pi}$.

$$\Pi_{i+1} = F\Pi_i F^* + GQG^*$$

$$\Rightarrow (\Pi_{i+1} - \bar{\Pi}) = F(\Pi_i - \bar{\Pi})F^*$$

$$\Rightarrow (\Pi_i - \bar{\Pi}) = F^i(\Pi_0 - \bar{\Pi})F^{*i}$$

$i \to \infty$, $F^i \to 0$ due to stability

$\longrightarrow$ y becomes WSS.

$$R_y(i-j) = \langle y_i, y_j \rangle = \begin{cases} HF^{i-j-1}\bar{N} & i>j \\ R + H\bar{\Pi}H^* & i=j \\ \bar{N}^*(F^*)^{i-j-1}H^* & i<j \end{cases}$$

$$x_{i+1} = Fx_i + [I \quad 0]\begin{bmatrix} Gu_i \\ v_i \end{bmatrix}$$

$$y_i = Hx_i + [0 \quad I]\begin{bmatrix} Gu_i \\ v_i \end{bmatrix}$$

$$Y(z) = [H(zI-F)^{-1} \quad I]\begin{bmatrix} G\,u(z) \\ v(z) \end{bmatrix}$$

$$S_y(z) = [H(zI-F)^{-1} \quad I]\begin{bmatrix} GQG^* & GS \\ S^*G^* & R \end{bmatrix}\begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix}$$

(*)

{Referenced in Thm (2)}

Use of KYP lemma, Controlability - Observability Gramian

$$\underbrace{\left(I + H(zI-F)^{-1}K_p\right)}_{L(z)} \left[R + HPH^*\right] L^*\left(1/z^*\right)$$

Riccati Iteration converges to the solution of Riccati Equation.

Difference between observability & detectibility of a discrete time system.

Thm 1: Assume $S=0$, F is stable, $\{F, GQ^{1/2}\}$ is controllable.

$$\begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \geq 0 \quad \& \quad R > 0$$

Then the discrete time Riccati Equation

Discrete Algebraic Riccati Equation

$$\begin{cases} P = FPF^* + GQG^* - K_p R_e K_p^* \\ R_e = R + HPH^*, \quad K_p = FPH^* R_e^{-1} \end{cases}$$

(Quadratic in P that's why harder)

has a unique p.s.d. sol^n P

such that $(F - K_p H)$ is stable. Also $R_e \geq 0$.

Thm 2: (Cannonical Spectral Factorization)

Consider $S_y(z)$ in ⊛ with F, G, Q, R, S satisfying assumptions of Thm 1.

Then the canonical spectral factorisation of $S_y(z)$

is $S_y(z) = L(z) R_e L^*(1/z^*)$, $L(\infty) = I$

when $L(z) = I + H(zI - F)^{-1} K_p$

$$L^{-1}(z) = I - H(zI - F + K_p H)^{-1} K_p$$

$$K_p = FPH^* R_e^{-1}, \quad R_e = R + HPH^*$$

where P is the unique PSD stabilizing sol of the DARE.

Smith form: Time update & Measurement Update form

$$\hat{x}_{i|i} = \sum_{j=0}^{i} \langle x_i, e_j \rangle R_{e_{ij}}^{-1} e_j$$

$$= \hat{x}_{i|i-1} + \underbrace{\langle x_i, e_i \rangle R_{e,i}^{-1}}_{K_{f,i}} e_i \quad \leftarrow \text{Measurement update}$$

$$\langle x_i, e_i \rangle = \langle x_i, \tilde{x}_{i|i-1} \rangle H_i^* + \langle x_i, v_i \rangle = P_{i|i-1} H_i^*$$

$$K_{f,i} = P_{i|i-1} H_i^* R_{e,i}^{-1}$$

$$x_i - \hat{x}_{i|i} \to \hat{x}_{i|i-1} + K_{f,i} e_i$$

$$P_{i|i} = \langle \tilde{x}_{i,i}, \tilde{x}_{i,i} \rangle$$

$$= P_{i|i-1} + K_{f,i} R_{e_i} K_{f,i}^*$$

$$- K_{f,i} \underbrace{\langle e_i, \tilde{x}_{i|i-1} \rangle} - \langle \tilde{x}_{i|i-1}, e_i \rangle K_{f,i}^*$$

$$\langle e_i, \tilde{x}_{i|i-1} \rangle$$

$$= H_i \langle \tilde{x}_{i|i-1}, \tilde{x}_{i|i-1} \rangle + \langle v_i, \tilde{x}_{i|i-1} \rangle$$

$$= H_i P_{i|i-1}$$

$$\Rightarrow P_{i|i} = P_{i|i-1} - \underbrace{\overbrace{P_{i|i-1} H_i^* R_{e,i}^{-1}}^{K_{f,i}} H_i P_{i|i-1}}_{= K_{f,i} R_{e,i} K_{f,i}^*}$$

$$\begin{cases} \hat{x}_{0|-1} \xrightarrow{mu} \hat{x}_{0|0} \xrightarrow{TU} \hat{x}_{1|0} \xrightarrow{mu} \hat{x}_{1|1} \xrightarrow{Tu} \hat{x}_{2|1} \\ P_{0|-1} \xrightarrow{mu} P_{0|0} \xrightarrow{TU} P_{1|0} \xrightarrow{mu} P_{1|1} \xrightarrow{Tu} P_{2|1} \end{cases}$$

Measurement Update :

$$\hat{x}_{i|i} = \hat{x}_{i|i-1} + K_{f,i} (y_i - H_i \hat{x}_{i|i-1})$$

$$K_{f,i} = P_{i|i-1} H_i^* [\underbrace{H_i P_{i|i-1} H_i^* + R_i}_{R_{e,i}}]^{-1}$$

$$P_{i|i} = P_{i|i-1} - \underbrace{K_{f,i} R_{e,i} K_{f,i}^*}_{\text{Decreases the covariance}}$$

Time update : $(S_i = 0)$

$$x_{i+1} = F_i x_i + G_i u_i$$

$$\hat{x}_{i+1|i} = F_i \hat{x}_{i|i} + G_i u_{i|i} \to 0$$

$$\tilde{x}_{i+1|i} = F_i x_i + G_i u_i - F_i \hat{x}_{i|i}$$

$$= F_i \tilde{x}_{i|i} + G_i u_i$$

$$P_{i+1|i} = \langle \tilde{x}_{i+1|i}, \tilde{x}_{i+1|i} \rangle = F_i P_{i|i} F_i^* + \underbrace{G_i Q_i G_i^*}_{\text{Increase in covariance}}$$

Continuous time

$$\dot{x} = F(t) x(t) + B(t) u(t) + G(t) w(t)$$

$$E(w(t)) = 0$$

$$E(w(t) \cdot w(t')^T) = Q(t) \delta(t - t')$$

Observations are available at $\{ t_1, t_2, t_3, \dots, t_i, \dots \}$

$$y(t_i) = H(t_i) x(t_i) + v(t_i)$$

$$E[v(t_i)] = 0$$

$$E(v(t_i) v^T(t_j)) = R(t_i) \delta_{ij}$$

(why non trivial? (Existence of ϕ))

$$x(t_i) = \phi(t_i, t_{i-1})\, x(t_{i-1}) + \int_{t_{i-1}}^{t_i} \phi(t_i, \tau)\, B(\tau)\, u(\tau)\, d\tau + \underbrace{\int_{t_{i-1}}^{t_i} \phi(t_i, \tau)\, G(\tau)\, \omega(\tau)\, d\tau}_{w_d(t_{i-1})}$$

Time update: $\hat{x}(t_i) = \phi(t_i, t_{i-1})\, \hat{x}(t_{i-1}^+) + \int_{t_{i-1}}^{t_i} \phi(t_i, t)\, B(\tau)\, u(\tau)\, d\tau$

$$P(t_i) = \phi(t_i, t_{i-1})\, P(t_{i-1}^+)\, \phi^T(t_i, t_{i-1}) + \int_{t_{i-1}}^{t_i} \phi(t_i, \tau)\, G(\tau)\, Q(\tau)\, G^T(\tau)\, \phi^T(t_i, \tau)\, d\tau$$

Measurement update: $K(t_i) = P(t_i^-)\, H^T(t_i)\left[H(t_i)\, P(t_i^-)\, H^T(t_i) + R(t_i)\right]^{-1}$

$$\hat{x}(t_i^+) = \hat{x}(t_i^-) + K(t_i)\left[y_i - H(t_i)\, \hat{x}(t_i^-)\right]$$

$$P(t_i^+) = P(t_i^-) - K(t_i)\, H(t_i)\, P(t_i^-)$$

EXAMPLE:

$$\dot{x} = \underset{\text{constant}}{u} + \omega(t) \quad \text{(Scalar)}$$

$$y(t_i) = x(t_i) + v(t_i)$$

$$\begin{cases} E(\omega(t)) = 0 \\ E\left(\omega(t)\, \omega^T(t-\tau)\right) = \sigma_\omega^2\, \delta(\tau) \\ E\left(v(t_i)\, v(t_j)\right) = \sigma_v^2 \end{cases}$$

$F = 0 \longrightarrow \phi = 1$

$\{ B = 1,\ G = 1,\ Q = \sigma_\omega^2,\ H = 1,\ R = \sigma_v^2 \}$

T.U.: $\hat{x}(t_i^-) = 1$; $\hat{x}(t_i^+) = 1\, \hat{x}(t_i^-) + \int_{t_{i-1}}^{t_i} 1 \cdot 1\, u\, dt$

$$= \hat{x}(t_i) + u\,(t_i - t_{i-1})$$

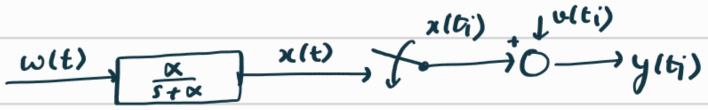
$$P(t_i^-) = P(t_{i-1}^+) + \sigma_\omega^2\,(t_i - t_{i-1})$$

M.U.: $K(t_i) = \dfrac{P(t_i^-)}{P(t_i^-) + \sigma_v^2}$

$$\hat{x}(t_i^+) = \hat{x}(t_i) + K(t_i)\left[y(t_i) - \hat{x}(t_i^-)\right]$$

$$P(t_i^+) = (1 - K(t_i))\, P(t_i^-)$$

EXAMPLE:



$\alpha = 1$ rad/hr $\qquad w(t) \longrightarrow$ deg/hr

$E(w(t)) = 0 \qquad E(w(t)\ w(t+\tau)) = Q\ \delta(\tau)$

$E(v(t_i)\ v(t_j)) = R\ \delta_{ij}$ $\qquad 2\ deg^2/hr \qquad$ ↳ $/hr$

$R = 0.5\ deg^2/hr^2$

$\dot{x} = -\alpha\ x(t) + \alpha w(t)$

$F = -\alpha = -1 \quad , \quad G = \alpha = 1$

$\phi(t_i, t_{i-1}) = e^{-1(0.25)} = 0.78$

T.U. : $\hat{x}(t_i^-) = 0.78\ \hat{x}(t_{i-1}^+)$

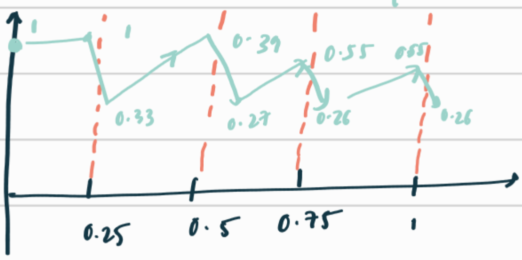
$P(t_i^-) = (0.78)^2\ P(t_{i-1}^+) + 2\int_{t_{i-1}}^{t_i} e^{-2(t_i-\tau)}\, d\tau$

$= 0.61\ P(t_{i-1}^+) + 0.39$

M.U.: $K(t_i) = \dfrac{P(t_i^-)}{P(t_i^-) + 0.5}$

$\hat{x}(t_i^+) = \hat{x}(t_i^-) + K(t_i)\left[y_i - \hat{x}(t_i^-)\right]$

$P(t_i^+) = \dfrac{0.5\ P(t_i^-)}{P(t_i^-) + 0.5}$



$0.61\ (P^+)^2 + (0.59)P^+ - 0.19 = 0$

↳ +ve root $P^+ = 0.255$

which is stabilizing $\quad P^- = 0.546$

Can refer to Stochastic models, estimation and control by Peter S. Maybeck.

Extended Kalman Filter

$x_{i+1} = f_i(x_i) + g_i(x_i)\, u_i$

$y_i = h_i(x_i) + v_i$

u_i, v_i are zero mean

$E(x(t_0)) = \bar{x}_0$

$$\left\langle \begin{bmatrix} u_i \\ v_i \\ x_0 - \bar{x}_0 \end{bmatrix}, \begin{bmatrix} u_j \\ v_j \\ x_0 - \bar{x}_0 \end{bmatrix} \right\rangle = \begin{bmatrix} Q\delta_{ij} & 0 & 0 \\ 0 & R\delta_{ij} & 0 \\ 0 & 0 & \Pi_0 \end{bmatrix}$$

$$\begin{cases} f_i(x_i) \approx f_i(\hat{x}_{i|i}) + \overbrace{\left(\left.\dfrac{\partial f_i(x)}{\partial x}\right|_{\hat{x}_{i|i}}\right)}^{F_i}(x_i - \hat{x}_{i|i}) \\ h_i(x_i) \approx h_i(\hat{x}_{i|i-1}) + \overbrace{\left(\left.\dfrac{\partial h_i(x)}{\partial x}\right|_{\hat{x}_{i|i-1}}\right)}^{H_i}(x_i - \hat{x}_{i|i-1}) \\ g_i(x_i) = g_i(\hat{x}_{i|i}) = G_i \end{cases}$$

$$\rightarrow \begin{cases} x_{i+1} = F_i x_i + \left[f_i(\hat{x}_{i|i}) - F_i \hat{x}_{i|i} \right] + G_i u_i \\ y_i - \left(h_i(\hat{x}_{i|i-1}) - H_i \hat{x}_{i|i-1}\right) = H_i x_i + v_i \end{cases}$$

$$\Rightarrow \quad \hat{x}_{i+1|i} = F_i \hat{x}_{i|i} + \left[f_i(\hat{x}_{i|i}) - F_i \hat{x}_{i|i} \right] = f_i(\hat{x}_{i|i})$$

$$\hat{x}_{i|i} = \hat{x}_{i|i-1} + K_{f,i}\left[\{ y_i - h_i(\hat{x}_{i|i-1}) + H_i \hat{x}_{i|i-1} \} - H_i \hat{x}_{i|i-1} \right] = \hat{x}_{i|i-1} + K_{f,i}\left[y_i - h_i(\hat{x}_{i|i-1}) \right]$$

$$\left.\begin{aligned} \hat{x}_{i+1|i} &= f_i(\hat{x}_{i|i}) \\ P_{i+1|i} &= F_i P_{i|i} F_i^* + G_i Q G_i^* \end{aligned}\right\} \text{Time Update}$$

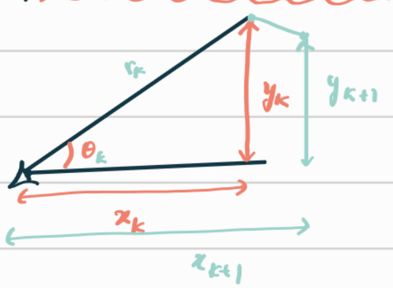
$$\begin{pmatrix} \hat{x}_{0|-1} = \bar{x}_0 \\ P_{0|-1} = \Pi_0 \end{pmatrix}$$

$$\left.\begin{aligned} \hat{x}_{i|i} &= \hat{x}_{i|i-1} + K_{f,i}\left[y_i - h_i(\hat{x}_{i|i-1}) \right] \\ K_{f,i} &= P_{i|i-1} H_i^* (H_i P_{i|i-1} H_i^* + R)^{-1} \\ P_{i|i} &= (I - K_{f,i} H_i) P_{i|i-1} \end{aligned}\right\} \text{Measurement Update}$$

F_i, G_i, H_i depend on the state

We now will see examples of applications of Kalman Filter.

Track a point in 2D with range & bearing



Measurements: $\begin{pmatrix} r_k + e_{r_k}, & \theta_k + e_{\theta_k} \end{pmatrix}$

$$\left\langle \begin{bmatrix} e_{r_k} \\ e_{\theta_k} \end{bmatrix}, \begin{bmatrix} e_{r_j} \\ e_{\theta_j} \end{bmatrix} \right\rangle = \begin{bmatrix} \sigma_r \delta_{ij} & 0 \\ 0 & \sigma_\theta \delta_{ij} \end{bmatrix}$$

$$z_k = \begin{bmatrix} r_k \sin\theta_k \\ r_k \cos\theta_k \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{x}_k \\ x_k \\ \dot{y}_k \\ y_k \end{bmatrix} + v_k$$

$$H \bar{x}_k + v_k$$

$$v_k = \begin{bmatrix} (r_k + e_{rk}) \sin(\theta_c + e_{\theta k}) - r_k \sin\theta_k \\ \cdots \end{bmatrix}$$

$$\Rightarrow \quad R_k = \langle v_k, v_k \rangle$$

$$= \begin{bmatrix} \sigma_r^2 \sin^2\theta_k + r_k^2 \sigma_\theta^2 \cos^2\theta_k & \sim \\ (\sigma_r^2 - r_k^2 \sigma_\theta^2) \sin\theta_k \cos\theta_k & \sigma_r^2 \cos^2\theta_k + r_k^2 \sigma_\theta^2 \sin^2\theta_k \end{bmatrix}$$

$$\dot{x}_{k+1} = \alpha \dot{x}_k + w_k^1 \qquad \dot{x}_k = s_k \cos\theta_k^c$$

$$\dot{y}_{k+1} = \alpha \dot{y}_k + w_k^3 \qquad \dot{y}_k = s_k \sin\theta_k^c$$

$(s_{k+1} - s_k)^2$ Mean square change in speed (s_k) $= \sigma_s^2$ sampling interval $= \Delta$

$(\theta_{k+1}^c - \theta_k^c)^2$ Mean square change in course (θ_k^c) $= \sigma_c^2$

$$\langle w_k^1, w_k^1 \rangle \approx \sigma_s^2 \sin^2\theta_k^c + \sigma_c^2 s_k^2 \cos^2\theta_k^c$$

$$\langle w_k^3, w_k^3 \rangle \approx \sigma_s^2 \cos^2\theta_k^c + \sigma_c^2 s_k^2 \sin^2\theta_k^c$$

$$\langle w_k^1, w_k^3 \rangle \approx (\sigma_s^2 - \sigma_c^2 s_k^2) \sin\theta_k^c \cos\theta_k^c$$

$$x_{k+1} = x_k + \tfrac{1}{2}\Delta (\dot{x}_k + \dot{x}_{k+1})$$

$$= x_k + \Delta (\dot{x}_k + \tfrac{1}{2} w_k^1)$$

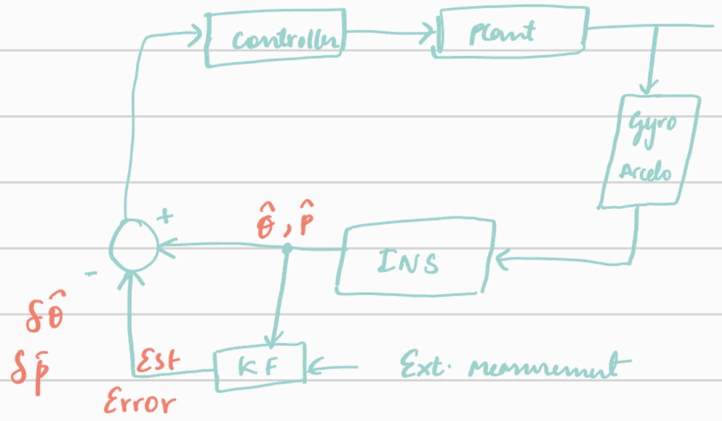
$$\begin{bmatrix} \dot{x}_{k+1} \\ x_{k+1} \\ \dot{y}_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} \alpha & 0 & 0 & 0 \\ \Delta & 1 & 0 & 0 \\ 0 & 0 & \alpha & 0 \\ 0 & 0 & \Delta & 1 \end{bmatrix} \begin{bmatrix} \dot{x}_k \\ x_k \\ \dot{y}_k \\ y_k \end{bmatrix} + \begin{bmatrix} w_k^1 \\ \frac{\Delta}{2} w_k^1 \\ w_k^3 \\ \frac{\Delta}{2} w_k^3 \end{bmatrix}$$

← Example from Anderson & Moore

↳ Optimal Filtering

Jack Kuipers : Quaternions

Indirect Feed forward



Here, INS is not modified.

Indirect feedback config.

$$x(t_k^+) = x(t_k^-) + K(\quad)$$

Indirect feedforward

t ← time

i ← inertia

r ← radar

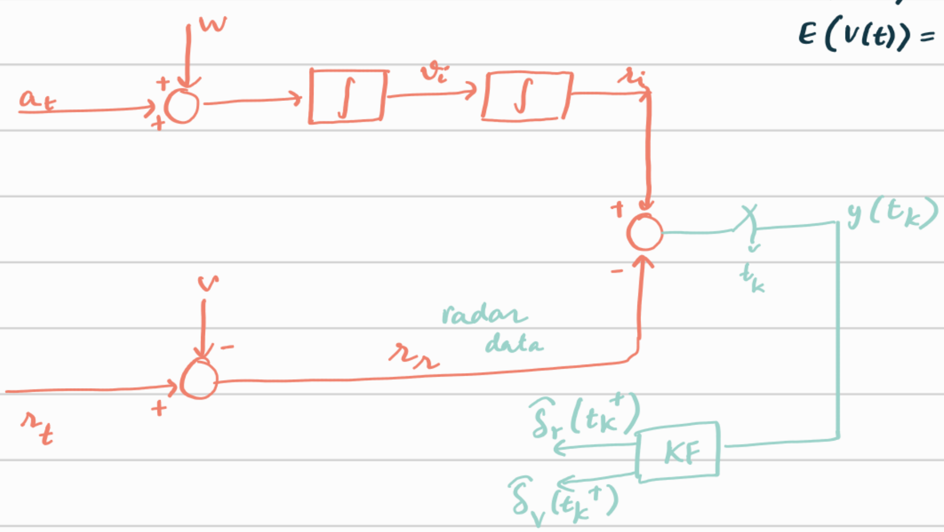
k ← time instant

$$E[w(t)\, w(t+\tau)] = Q\, \delta(\tau)$$

$$E[v(t)\, v(t+\tau)] = R_c\, \delta(\tau)$$

$$E(w(t)) = 0$$

$$E(v(t)) = 0$$



Error State

$$\delta_r(t) = r_i(t) - r_t(t)$$

$$\delta_v(t) = v_i(t) - v_t(t)$$

Observation $y(t_k) = r_i(t_k) - r_r(t_k)$

$= r_t(t_k) + \delta_r(t_k) - (r_t(t_k) - v(t_k))$

$$y(t_k) = \delta_r(t_k) + v(t_k)$$

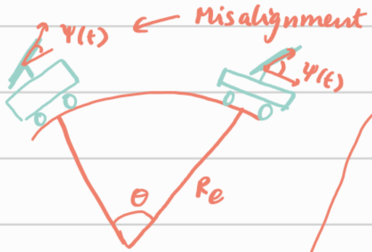
① $\begin{bmatrix} \dot{r}_i(t) \\ \dot{v}_i(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_i(t) \\ v_i(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} [a_t(t) + w(t)]$

② $\begin{bmatrix} \dot{r}_t(t) \\ \dot{v}_t(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_t(t) \\ v_t(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} a_t(t)$

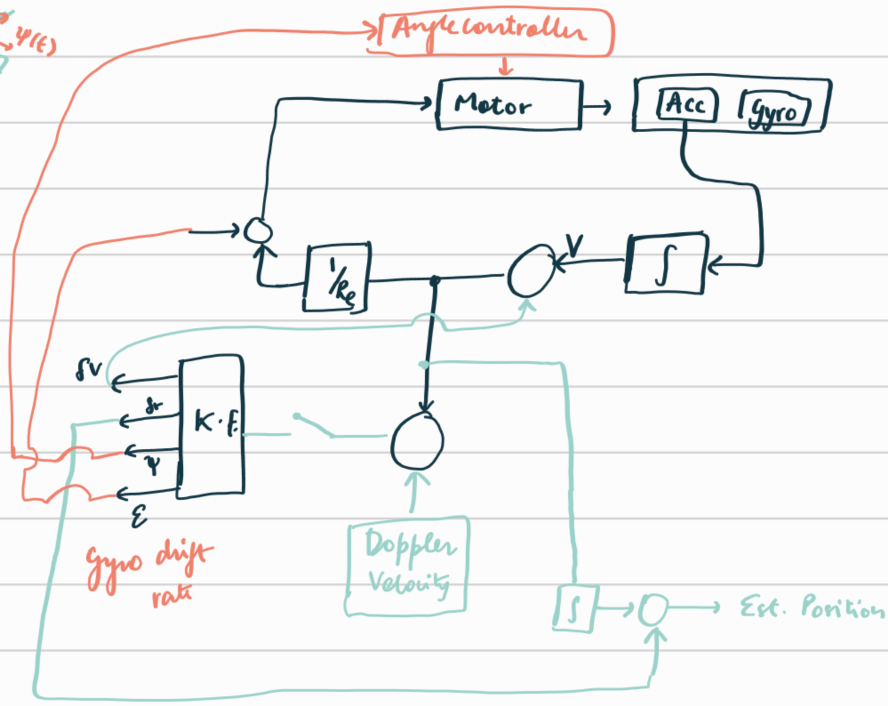
$\{① - ②\} \Rightarrow \begin{bmatrix} \delta\dot{r}(t) \\ \delta\dot{v}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \delta r(t) \\ \delta v(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t)$

$$y(t_k) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \delta r(t_k) \\ \delta v(t_k) \end{bmatrix} + v(t_k)$$

Indirect feedback configuration



(Acc. + Doppler)



$\delta r \rightarrow$ error in INS indicated position

$\delta v \rightarrow$ —— " —— velocity

$\psi \rightarrow$ Platform tilt

$\varepsilon \rightarrow$ Gyro drift rate

$$(1)\quad \delta\dot{r}(t) = \delta v(t)$$

$$(2)\quad \delta\dot{v}(t) = y\sin(\psi(t))$$

$$(3)\quad \dot{\psi}(t) = -\frac{\delta v(t)}{R_e} = \varepsilon(t)$$

$$(4)\quad \dot{\varepsilon}(t) = w(t)$$

$$\begin{pmatrix} \delta r = r_i(t) - r_t(t) \\ \delta v = v_i(t) - v_t(t) \end{pmatrix}$$

$$\begin{bmatrix} \delta\dot{r}(t) \\ \delta\dot{v}(t) \\ \dot{\psi}(t) \\ \dot{\varepsilon}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & g & 0 \\ 0 & -1/R_e & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta r(t) \\ \delta v(t) \\ \psi(t) \\ \varepsilon(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} w(t)$$

$$V_i(t_k) = V_t(t_k) + \delta_v(t_k)$$

$$V_s = V_t(t_k) - n(t_k)$$

$$y(t_k) = V_i(t_k) - V_\Delta(t_k)$$

$$= \delta v(t_k) + \eta(t_-)$$

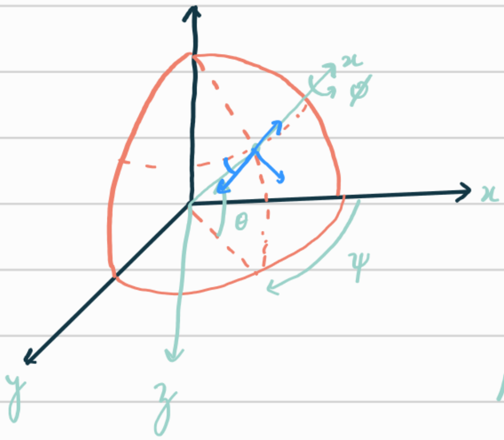
$$= \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta_r(t_k) \\ \delta_v(t_k) \\ \psi(t_k) \\ \varepsilon(t_k) \end{bmatrix} + \eta(t_k)$$

$$\hat{x}(t_{i-1}^{+c}) = \begin{bmatrix} \hat{\delta}_r(t_{i-1}^{+c}) \\ \hat{\delta}_v(t_{i-1}^{+c}) \\ \hat{\psi}(t_{i-1}^{+c}) \\ \varepsilon(\cdot) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \psi(t_{i-1}^{+}) \\ 0 \end{bmatrix}$$

After sometime we can

$$\hat{x}(t_i^-) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Smoother-based 3D attitude Estimation for mobile robot localization



Local frame $\rightarrow \psi \rightarrow \theta \rightarrow \phi \rightarrow$ Body frame

Yaw Pitch Roll

$$R_E^B = R_\phi^x \, R_\theta^y \, R_\psi^z$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & c\phi & s\phi \\ 0 & -s\phi & c\phi \end{bmatrix} \begin{bmatrix} c\theta & 0 & -s\theta \\ 0 & 1 & 0 \\ s\theta & 0 & c\theta \end{bmatrix} \begin{bmatrix} c\psi & s\psi & 0 \\ -s\psi & c\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$x_B = R_E^B \, x_E$$

$$g_B = R_E^B \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} = g \begin{bmatrix} -s\theta \\ s\phi \, c\theta \\ c\phi \, c\theta \end{bmatrix}$$

Euler's Theorem

$$q = \left[\bar{q}^T \; q_4 \right]^T$$

$q_1 \leftarrow \cos\alpha \;\; \sin\theta/_2$

$q_2 \leftarrow \cos\beta \;\; \sin\theta/_2$

$q_3 \leftarrow \cos\gamma \;\; \sin\theta/_2$

$q_4 \leftarrow \cos\theta/_2$

$R(q) \leftarrow$ transformation from q to R

$$q^a + q^b = \left[(\bar{q}^a + \bar{q}^b)^T \quad q_4^a + q_4^b \right]^T$$

$$q^a \otimes q^b = \left[(q_4^b \, \bar{q}^a + q_4^a \, \bar{q}^b + \bar{q}^a \times \bar{q}^b)^T \quad q_4^a \, q_4^b - \bar{q}^a \cdot \bar{q}^b \right]^T$$

$$q^* = \left[-\bar{q}^T \quad q_4 \right]^T$$

$$q^{-1} = q^* \left\{ \text{when } \|q\| = 1 \right\}$$

Fact $(\theta/2 = \phi)$

For any unit quaternion $q = \cos\phi + \vec{u}\sin\phi$ ($\vec{u}$: i, j, k)
and for any vector $\vec{v} \in \mathbb{R}^3$

(i) The action of $L_q(v) = q \otimes v \otimes q^*$ $\quad [v = [\vec{v}^T\ 0]]$
is equivalent to rotation of $\vec{v}$ through an angle 2ϕ around $\vec{u}$ (axis of rotation)

(ii) The action of $L_{q^*}(v) = q^* \otimes v \otimes q$ is Equivalent to the rotation of the coordinate frame through 2ϕ about $\vec{u}$ as axis.

(iii) $L_{q^*}(v) = q^* v q$ is also equivalent to an opposite rotation of V wrt the coordinate frame through an angle 2ϕ with $\vec{u}$ as axis.

$$R_e^{\,b} = R_a^b \, R_e^a$$

$$q_3 = q_2 \otimes q_1$$

$$\Rightarrow R(q_3) = R(q_2) \cdot R(q_1)$$

$$\dot{R}_E^B = \begin{bmatrix} 0 & \omega_z & -\omega_y \\ -\omega_z & 0 & \omega_x \\ \omega_y & -\omega_x & 0 \end{bmatrix} R_E^B$$

$$\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \\ \dot{q}_4 \end{bmatrix} = \dot{q} = \frac{1}{2} \begin{bmatrix} 0 & \omega_z & -\omega_y & \omega_x \\ -\omega_z & 0 & \omega_x & \omega_y \\ \omega_y & -\omega_x & 0 & \omega_z \\ -\omega_x & -\omega_y & -\omega_z & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

Gyro model: $\dot{\theta}_{true} = \omega_m + b_{true} + n_r$

gyro measures (ω_m) ; drift rate (n_r)

$$\dot{b}_{true} = n_w$$

$$\dot{\theta}_i = \omega_m + b_i \qquad \Delta b = b_{true} - b_i$$

$$\dot{b}_i = 0 \qquad \Delta\dot{b} = n_w$$

$$\delta\dot{\theta} = \Delta b + n_r$$

$$q_{true} = \delta q \otimes q_i \qquad \Bigg| \quad \begin{aligned} \dot{q}_{true} &= \frac{1}{2}\,\Omega(\dot{\theta}_{true})\,q_{true} \\ \dot{q}_i &= \frac{1}{2}\,\Omega(\dot{\hat{\theta}}_i)\,q_i \end{aligned}$$

$$\delta\dot{\bar{q}} = \begin{bmatrix} 0 & \omega_z & -\omega_y \\ -\omega_z & 0 & \omega_x \\ \omega_y & -\omega_x & 0 \end{bmatrix} \delta\bar{q} - \frac{1}{2}(\Delta b + n_\Omega)$$

$$\delta\dot{q}_4 = 0$$

$$\omega_m = \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}$$

$$\begin{bmatrix} \delta\dot{q} \\ \Delta\dot{b} \end{bmatrix} = \left[\begin{array}{c|c} \begin{matrix} 0 & \omega_z & -\omega_y \\ -\omega_z & 0 & \omega_x \\ \omega_y & -\omega_x & 0 \end{matrix} & -\frac{1}{2} I_3 \\ \hline 0 & 0 \end{array}\right] \begin{bmatrix} \delta\bar{q} \\ \Delta\bar{b} \end{bmatrix}$$

$$+ \left[\begin{array}{c|c} -\frac{1}{2} I_3 & O_3 \\ \hline O_3 & I_3 \end{array}\right] \begin{bmatrix} n_\Omega \\ n_w \end{bmatrix}$$

3 types of models → 1st Principle models
→ Grey box model
→ Black box model

Ljung

Prediction Error criterion

$$y(t) = \sum_{k=0}^{\infty} g(k)\, u(t-k) + v(t)$$

$\leftarrow t \in \{0, 1, \dots\}$

$q \cdot u(t) = u(t+1)$, $q^{-1} \cdot u(t) = u(t-1)$

↑ shift operator

$$y(t) = \sum_{k=0}^{\infty} g(k)\, u(t-k) + v(t)$$

$$= \sum_{k=0}^{\infty} g(k)\, q^{-k} u(t) + v(t) = G(q)\, u(t) + v(t)$$

$$v(t) = \sum_{k=0}^{\infty} h(k)\, e(t-k)$$

when $e(t)$ is (white, uncorrelated, zero mean)

$$y(t) = G(q)\, u(t) + H(q)\, e(t)$$

↑ stable, minimum phase, $h(0) = 1$

$v(s) = y(s) - G(q)u(s)$ $\quad s \le t-1$

One step prediction of $v(t)$ given $\{y(1), \dots, y(t-1)\}$

$$v(t) = H(q)\, e(t) = e(t) + \sum_{k=1}^{\infty} h(k)\, e(t-k)$$

$$\hat{v}(t|t-1) = \sum_{k=1}^{\infty} h(k)\, e(t-k)$$

$$= (H(q)-1)\, e(t) = [H(q)-1]\, H^{-1}(q)\, v(t)$$

$$= (1 - H^{-1}(q))\, v(t)$$

$$\hat{y}(t|t-1) = G(q)u(t) + \hat{v}(t|t-1)$$

$$= G(q)\, u(t) + [1 - H^{-1}(q)]\, v(t)$$

$$= G(q)\, u(t) + [1 - H^{-1}(q)]\, [y(t) - G(q)\, u(t)]$$

$$\hat{y}(t|t-1) = H^{-1}(q)\, G(q)\, u(t) + (1 - H^{-1}(q))\, y(t)$$

Example: ARX model:

$$y(t) + a_1 y(t-1) + \dots + a_{n_a} y(t-n_a)$$
$$= b_1 u(t-1) + \dots + b_{n_b} u(t-n_b) + e(t)$$

$\{H(q) = 1\}$

$$G(q) = \frac{b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}} = \frac{B(q)}{A(q)}$$

$$H(q) = \frac{1}{A(q)}$$

$$\hat{y}(t|t-1) = B(q)u(t) + [1 - A(q)]y(t)$$

$$\phi(t) = [-y(t-1) \quad -y(t-n_a) \quad u(t-1) \cdots u(t-n_b)]^T$$

$$\theta = [a_1 \cdots a_{n_a} \quad b_1 \cdots b_{n_b}]^T$$

$$\hat{y}(t) = \theta^T \phi(t)$$

$$\|y(t) - \theta^T \phi(t)\|_2$$

$$\begin{cases} x(t+1) = A(\theta)x(t) + B(\theta)u(t) + w(t) \\ y(t) = C(\theta)x(t) + v(t) \end{cases}$$

$$y(t) = G(q, \theta)u(t) + \boxed{\quad\quad} e(t)$$

$$\{C(\theta)[qI - A(\theta)]^{-1}B(\theta)\}$$

$$\hat{x}(t+1|\theta) = A(\theta)\hat{x}(t|\theta) + B(\theta)u(t) + K(\theta)e(t)$$

$$y(t) = C(\theta)\hat{x}(t|\theta) + e(t)$$

$$H(q, \theta) = C(\theta)[qI - A(\theta)]^{-1}K(\theta) + I$$

$$\hat{y}(t|\theta) = C(\theta)[qI - A(\theta) + K(\theta)C(\theta)]^{-1}B(\theta)u(t) + C(\theta)[qI - A(\theta) + K(\theta)C(\theta)]^{-1}K(\theta)y(t)$$

Can work it out.

$$A(\theta) = \begin{bmatrix} -a_1 & 1 & 0 \\ -a_2 & 0 & 1 \\ -a_3 & 0 & 0 \end{bmatrix} \quad B(\theta) = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$C(\theta) = [1 \quad 0 \quad 0]$$

$$K(\theta) = \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix}$$

$$G(q, \theta) = \frac{b_1 q^{-1} + b_2 q^{-2} + b_3 q^{-3}}{1 + a_1 q^{-1} + a_2 q^{-2} + a_3 q^{-3}}$$

Ho - Kalman Filter

$$C(zI - A)^{-1}B$$
$$= C\frac{1}{z}\frac{1}{(I - A/z)}B$$
$$= C\frac{1}{z}\left(I + \frac{A}{z} + \frac{A^2}{z^2} + \frac{A^3}{z^3} + \cdots\right)B$$
$$= C\left(Iz^{-1} + Az^{-2} + \cdots\right)B$$

Data : $\{G_t\} \longrightarrow n, A, B, C, D$

$$g(t) = Ce^{At}B$$
$$= C\left(I + At + \frac{A^2}{2}t^2 + \frac{A^3t^3}{3!}\cdots\right)B$$

at $t=1$, CB
at $t=2$, CAB

$\{CB, CAB, CA^2B, \ldots\}$ ← Markov parameters

$$\mathcal{O}_K = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{k-1} \end{bmatrix} \qquad \mathcal{C}_\ell = \begin{bmatrix} B & AB & & A^{\ell-1}B \end{bmatrix}$$

$$\bar{x} = T^{-1}x$$
$$\dot{\bar{x}} = T^{-1}AT\bar{x} + T^{-1}Bu$$
$$y = C^T\bar{x}$$

$$\bar{\mathcal{O}}_K = \mathcal{O}_K T$$
$$\bar{\mathcal{C}}_\ell = T^{-1}\mathcal{C}_\ell$$

$$\underset{kp \times m\ell}{\bar{\mathcal{O}}_K \bar{\mathcal{C}}_\ell} = \mathcal{O}_K \mathcal{C}_\ell = \begin{bmatrix} CB & CAB & \cdots & CA^{\ell-1}B \\ CAB & CA^2B & & \\ CA^2B & \cdots & \ddots & \end{bmatrix} = H$$

$u \in \mathbb{R}^m$
$y \in \mathbb{R}^p$

Hankel matrices

Assume $k, \ell > n$

$$H_{k,\ell} = \begin{bmatrix} G_1 & G_2 & G_3 & G_4 & & G_\ell \\ G_2 & G_3 & & G_4 & & \\ G_3 & G_4 & & \cdots & & \\ G_4 & & & & & \\ G_K & & & & & \end{bmatrix} ; \quad \bar{H}_{k\ell} = \begin{bmatrix} G_2 & G_3 & \cdots & G_{\ell+1} \\ G_3 & & & \\ \vdots & & & \\ G_{K+1} & & & \end{bmatrix}$$

Fact: Let (A, B, C) be minimal, then

(a) $\text{rank}(H) = n$

(b) $H = \mathcal{O}\mathcal{C} = \mathcal{O}TT^{-1}\mathcal{C}$

$\bar{H} = \mathcal{O}A\mathcal{C}$

(c) $H^{\uparrow} = \mathcal{O}^{\uparrow}\mathcal{C} = \mathcal{O}A\mathcal{C} = \mathcal{O}\mathcal{C}^{\leftarrow} = H^{\leftarrow}$ {Assume ∞ sized}

(1) Create $H_{k,\ell}$ & $\bar{H}_{k,\ell}$ from data

(2) get a rank n- estimate of $H_{k,\ell}$

{n is choosen when there is a decent drop in singular values}

$$H_{k,\ell} = [U_n \;\; U_s] \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma_s \end{bmatrix} \begin{bmatrix} V_n^T \\ V_s^T \end{bmatrix}$$

$$\min_{\hat{H}_n} \| \hat{H}_n - H_{k,\ell} \|_2 \qquad \hat{H}_n = U \begin{bmatrix} \Sigma_n & 0 \\ 0 & 0 \end{bmatrix} V^T$$

$$= U_n \Sigma_n V_n^T$$

Pseudo Inverse : $V_n \Sigma_n^{-1} U_n^T$

A : tall & fullrank $\to$ $\underbrace{(A^TA)^{-1}A^T}_{\text{Pseudo inverse}}$

A : wide & full rank $\to$ $A^T(AA^T)^{-1}$

$H_{k,\ell} = U_n \Sigma_n V_n^T$ ← From data

$= \mathcal{O}_k \mathcal{C}_\ell$ ← from Theory

$= \underbrace{(U_n \Sigma_n^{1/2})}_{\mathcal{O}_k} \underbrace{(\Sigma_n^{1/2} V_n^T)}_{\mathcal{C}_\ell}$

Balanced realization → If just in one, then

$\bar{H} = \mathcal{O}_k A \mathcal{C}_\ell \to \hat{A} = (\hat{\mathcal{O}}_k)^{+} \bar{H} (\hat{\mathcal{C}}_\ell)^{+}$

$\hat{B} = \hat{\mathcal{C}}_\ell(:, 1:m)$

$\hat{C} = \hat{\mathcal{O}}_k(1:p, :)$

Example: $G(z) = \dfrac{z}{z^2 - z - 1}$

$y(0) = 0$

$y(1) = 1$

$y(2) = 1$

$y(3) = 2$

$y(4) = 3$

$u(0) = 1$, $y(0) = 0$, $y(-1) = 0$

$u(t \neq 0) = 0$

$y(t+2) - y(t+1) - y(t) = u(t+1)$

$k = 5, \ell = 5$

$H_{4,4}$, $\bar{H}_{4,4}$

$$\begin{bmatrix} 1 & 1 & 2 & 3 & 5 \\ 1 & 2 & 3 & 5 & 8 \\ 2 & 3 & 5 & 8 & 13 \\ 3 & 5 & 8 & 13 & 21 \\ 5 & 8 & 13 & 21 & 34 \end{bmatrix}$$

$$\begin{array}{ccccc} u_1 & u_2 & u_3 & u_4 & u_5 \\ & & \multicolumn{3}{c}{\text{drop}} \end{array}$$

$$\begin{bmatrix} -0.18 & 0.79 & & & \\ -0.3 & & & & \\ & & & & \\ & & & & \end{bmatrix} \left[\begin{array}{cc|ccc} 20.56 & 0 & & 0 & \\ 0 & 0.434 & & & \\ \hline & & 10^{-15} & & \\ & 0 & & 10^{-16} & \\ & & & & 10^{-18} \end{array}\right]$$

$$C_\ell = \Sigma^{1/2} V^T = \begin{bmatrix} u_1 & \text{———} \\ u_2 & \text{———} \end{bmatrix}$$

$$\Theta_k = U \Sigma^{1/2} =$$

$$\hat{A} = \begin{bmatrix} 1.62 & 0.02 \\ 0.02 & -0.62 \end{bmatrix} \qquad \hat{B} = \begin{bmatrix} 0.85 \\ -0.52 \end{bmatrix}$$

$$\hat{C} = \begin{bmatrix} 0.85 & -0.52 \end{bmatrix}$$

Oblique Projection

$A \in \mathbb{R}^{p \times j}$, $B \in \mathbb{R}^{q \times j}$, $C \in \mathbb{R}^{r \times j}$

$j > p, q, r$

$$A/B = A\Pi_B = AB^T(BB^T)^+B = L_B B$$

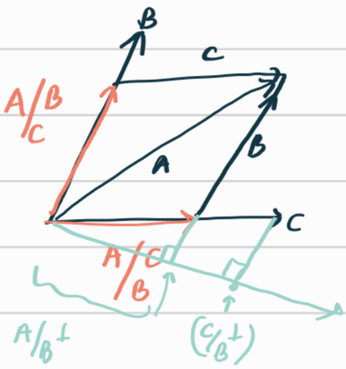
orthogonal projection of rows of A onto row span of B

$$A/B^\perp = A\,\Pi_{B^\perp} = L_{B^\perp} B$$

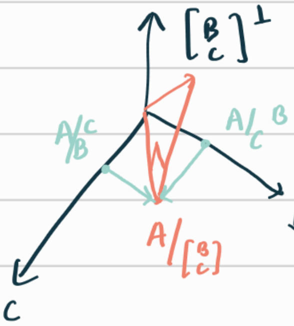
LQ decomposition:

$$\underbrace{\begin{bmatrix} A \\ B \end{bmatrix}}_{\text{wide matrix}} = \begin{array}{cc} L & Q \\ \begin{bmatrix} R_A \\ R_B \end{bmatrix} & Q^T \end{array}$$

$$A/B = R_A R_B^T \left[R_B R_B^T\right]^+ R_B Q^T$$



$$A/_B C = \left[A/B^\perp\right]\left[C/_B{}^\perp\right]^+ C$$



$$A = L_B B + L_C C + L_{B^\perp C^\perp} \begin{bmatrix} B \\ C \end{bmatrix}^\perp$$

$$A/\begin{bmatrix} B \\ C \end{bmatrix} = A\left[C^T B^T\right] \begin{bmatrix} CC^T & CB^T \\ BC^T & BB^T \end{bmatrix} \begin{bmatrix} C \\ B \end{bmatrix}$$

$$= A\left[C^T B^T\right] \begin{bmatrix} CC^T & CB^T \\ BC^T & BB^T \end{bmatrix}$$

} first r columns

} last q columns

$$B/_B C = 0 \qquad C/_B C = C$$

Problem: Given S measurements of $u_k \in \mathbb{R}^m$ & $y_k \in \mathbb{R}^l$ generated from (unknown):

$$x_{k+1} = Ax_k + Bu_k$$
$$y_k = Cx_k + Du_k$$

Find 1) order $= n$?

2) A, B, C, D (upto some transform)

$$y_0 = Cx_0 + Du_0$$
$$y_1 = CAx_0 + CBu_0 + Du_1$$
$$\vdots$$

$$y_{0|i-1} = \Gamma_i x_0 + H_i u_{0|i-1}$$

$$y_{i|2i-1} = \Gamma_i x_i + H_i u_{i|2i-1}$$

$$\begin{bmatrix} y_k & y_{k+1} & \cdots & y_{k+j-1} \\ \vdots & \vdots & & \vdots \\ y_{k+q-1} & y_{k+q} & \cdots & y_{k+j+q-2} \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ x_k & x_{k+1} & \cdots & x_{k+j-1} \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$+ \begin{bmatrix} D & & & 0 \\ CB & \ddots & & \\ CAB & \ddots & \ddots & \\ CA^{q-2}B & \cdots\, CAB & CB & D \end{bmatrix} \begin{bmatrix} u_k & u_{k+1} & \cdots & u_{k+j-1} \\ & & & \\ u_{k+q-1} & \cdots & & u_{k+q+j-2} \end{bmatrix}$$

$$y_{k|q} = \Gamma_q X_k + H_q u_{k|q}$$

Past: $k = 0,\ q = i$ $\qquad u_p = u_{0|i}$

$$y_p = y_{0|i}$$

$$X_p = X_0 = \begin{bmatrix} | & | & & | \\ x_0 & x_1 & \cdots & x_{j-1} \\ | & | & & | \end{bmatrix}$$

$$Y_p = \Gamma_i X_p + H_i u_p$$

Future: $k = i$, $\quad X_f = X_i = [x_i \cdots\cdots x_{i+j-1}]$

$$Y_f = \Gamma_i X_f + H_i U_f$$

x_f : $[x_i \quad x_{i+1} \cdots \quad x_{j+i-1}]$

x_i : $[x_0 \cdots \quad x_{j-1}]$

$$x_f = A^i x_p \quad + \Delta_i u_p$$

$$\underbrace{[A^{q-1}B \quad \cdots \quad AB \quad B]}_{\Delta_q}$$

$$X_f = A^i [\Gamma_i^+ y_p - \Gamma_i^+ H_i u_p] + \Delta_i u_p$$

$$Y_f = \Gamma_i [\Delta_i - A^i \Gamma_i^+ H_i] u_p + \Gamma_i A^i \Gamma_i^+ Y_p \quad + H_i u_f$$

$$\text{Let } W_p = \begin{bmatrix} u_p \\ y_p \end{bmatrix}$$

$$Y_f = \Gamma_i L_p W_p + H_i u_f$$

$$Y_f = \Gamma_i X_f + H_i u_f$$

$$L_p = [\Delta_i - A^i \Gamma_i^+ H_i \quad A^i \Gamma_i^+]$$