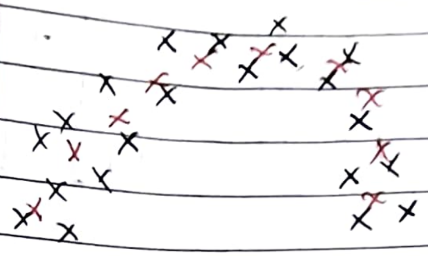
ISSP tut notes

Discrete Kalman Filter

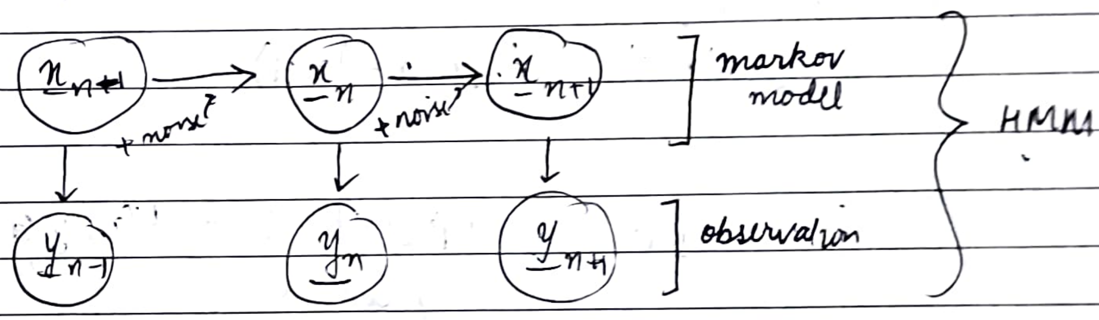


Given distribution
Kalman filter
"provides a smooth trajectory for this"
$\underline{x} \rightarrow$ vector of states

Transition model $\underline{x}_{n+1} = f(\underline{x}_n, \underline{w}_n)$ → noise

Observation model $\underline{y}_n = g(\underline{x}_n, \underline{v}_n)$

state evolution through time: follows a markov model
↳ state spaced model
$x_n \rightarrow$ true state



so lets formulate this...

State} $\underline{x}_{n+1} = f(\underline{x}_n, \underline{w}_n)$ $\quad \underline{x}_{n+1} = A_n \underline{x}_n + \underline{w}_n$ } linear & gaussian

Observation} $\underline{y}_n = g(\underline{x}_n, \underline{v}_n)$

$\underline{y}_n = C_n \underline{x}_n + \underline{v}_n$ } linear & gaussian

KF only works when f, g is linear & gaussian

$\underline{w}_n \sim \mathcal{N}(0, Q_w)$

$\underline{v}_n \sim \mathcal{N}(0, Q_v)$

Example AR(1) process
state $x[n] = a_1 x[n-1] + w[n]$ → $\mathcal{N}(0, \sigma^2)$
Observe $y[n] = x[n] + v[n]$ → $\mathcal{N}(0, \sigma_v^2)$

AR(p) process
state $x[n] = \sum_{i=1} a_i x[n-i] + w[n]$
observed $y[n] = x[n] + v[n]$

AR(p) process

$$\underline{x}_n = \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-p] \end{bmatrix} + \begin{bmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & \\ \vdots & & & \\ 0 & 0 & \cdots\ 1 & 0 \end{bmatrix} \begin{bmatrix} x[n-1] \\ x[n-2] \\ \vdots \\ x[n-p-1] \end{bmatrix} + \begin{bmatrix} w(n) \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\underline{x_n} = A\,\underline{x_{n-1}} + \underline{w_n}$$

kalman filters

1] $\underline{x_n} = A_n \underline{x_{n-1}} + \underline{w_n}$

2] $\underline{y_n} = C_n \underline{x_n} + \underline{v_n}$

⇒ Estimate $\underline{x}_n$ & $\underline{y}_n$

Objective: Estimate ~~xxx~~ hidden $\hat{\underline{x}}_n$ given observation $\underline{y}_{1:n}$

Solution

Let state estimate of a time $n-1$ be given by $\hat{x}(n-1|n-1)$

① Prediction step

$$\hat{\underline{x}}(n|n-1) = A_n \hat{\underline{x}}(n-1|n-1)$$

② Update step

$$\hat{\underline{x}}(n|n) = K_1(n) \hat{\underline{x}}(n|n-1) + K(n)\,\underline{y}_n$$

some process of ~~depends on~~ prediction

some process of y_n

" like a weighted average "

finding $K_1[n]$ and $K[n]$ get $\hat{x}(n|n)$ which is the optimal linear estimate of $\underline{x}_n$.

Prediction error

$$e(n|n-1) = \underline{x_n} - \hat{\underline{x}}(n|n-1)$$

Update error

$$e(n|n) = \underline{x_n} - \hat{\underline{x}}(n|n) \quad \Big\}\ \text{want to minimize error}$$

classmate
Date
Page

Statistics of error

$$E(\underline{e}[n|n-1]) = E(\underline{x}_n - \hat{x}(n|n-1))$$
$$= E(A_n x_{n-1} + w_n - A_n \hat{x}(n-1|n-1))$$
$$E(A_n(x_{n-1} - \hat{x}(n-1|n-1)))$$

tends to zero

① $E(e(n|n-1)) = A_n E(e(n-1|n-1)) \to 0$

② $E(e(n|n)) =$
$$E(\underline{x}_n - k_1(n)\hat{x}(n|n-1) - k(n)y_n)$$

$$E(x_n - k_1(n)[\underline{x}_n - e(n|n-1)] - k(n)(C_n \underline{x}_n + \underline{v}_n))$$

tends to 0

$$E[(I - k_1(n) - k(n)C_n)\underline{x}_n] - k_1(n)\cancel{E(e(n|n-1))} + k(n)\cancel{E[V_n]} \to 0$$

want to set this to zero

$$I - k_1[n] - k(n)C_n = 0$$
$$\boxed{K_1[n] = I - k[n]C_n}$$

If we want to set $e(n|n)$ to zero

statistic error:

$$P(n|n-1) = E(\underline{e}(n|n-1)\,e^T(n|n-1))$$
$$P(n|n) = E(\underline{e}(n|n)\,\underline{e}^T(n|n))$$

Error covariance matrices

$$P(n|n-1) = \text{cov}(e(n|n-1)) = \text{cov}(\tilde{x}(n|n-1))$$
$$P(n|n) = \text{cov}(e(n|n)) = \text{cov}(\tilde{x}(n|n))$$

second order statistics

$$E[\hat{x}(n|n)] = \underline{x}_n$$

comes from

Show that

$$P(n|n-1) = A_n P(n-1|n-1) A_n^T + Q_w(t)$$
$$P(n|n) = (I - K(n)C_n) P(n|n-1)$$

P represents uncertainity

~~$\hat{x}(n|n) = (I$~~

~~want to minimize $E[(x_n - \hat{x}(n|n))^2]$ chose a $k(n)$ such that the cross terms of error becomes 0~~

~~$E[(I - k(n)C_n)x_n + k(n)v_n]$~~

How?

Derivation of $P(n|n-1)$

$$x_n = A_n x_{n-1} + w_n \qquad w_n \sim N(0, Q_w)$$

$$\hat{x}(n|n-1) = A_n \hat{x}(n-1|n-1)$$

so

$$e(n|n-1) = x_n - \hat{x}(n|n-1) = A_n(e(n-1|n-1)) + w_n$$

$$\boxed{P(n|n-1) = \mathbb{E}(e(n|n-1)e^T(n|n-1)) = A_n P(n-1|n-1)A_n^T + Q_w(n)}$$

~~x_n is deterministic~~

Similarly

~~$P(n|n-1) = \mathbb{E}(e(n-1|n-1)e^T(n-1|n-1))$~~

~~$\mathbb{E}([x_n - k_1(n)\hat{x}(n|n-1) - k(n)C_n(C_n x_n + v_n)][\;]^T)$~~

~~$\mathbb{E}([1 - k(n)C_n)$~~

$$e(n|n) = x_n - \hat{x}(n|n) =$$

$$x_n - (1 - k(n)C_n)\hat{x}(n|n-1) - k(n)C_n(C_n x_n) - k(n)v_n$$

$$\boxed{e(n|n) = \underbrace{(1 - k(n)C_n)}_{a(n)}e(n|n-1) - k(n)v_n}$$

$$P(n|n) = \mathbb{E}(e(n|n)e^T(n|n))$$

~~$\mathbb{E}(1 - k(n)C_n\, e(n|n-1)$~~

$$\mathbb{E}[a(n)e(n|n-1)e^T(n|n-1)a^T(n)]$$

$$+ 0 + 0 + k(n)\mathbb{E}[v_n v_n^T]k^T(n)$$

this is not what sir gave.

$$P(n|n) = (1 - k(n)C_n)P(n|n-1)(1 - k(n)C_n)^T + k(n)Q_v(n)k^T(n)$$

how is it the same then?

The key idea is: if $k(n)$ is chosen optimally (minimises MMSE)

classmate
Date
Page

by using this eqn

$$k(n) = P(n|n-1)\, C_n^T \left[C_n P(n|n-1) C_n^T + Q_v(n) \right]^{-1}$$

then....

$$\boxed{P(n|n) = (1 - K(n) C_n) P(n|n-1)}$$

okay ill derive this seriously

$$P(n|n) = (1-KC)\, P(n|n-1)\, (1-KC)^T + KQK^T,$$

$$K = \underbrace{PC^T (CPC^T + Q)^{-1}}_{S}$$

$$K = PC^T S$$

$$P_u(n|n) = (1-KC)\, \cancel{P(n|n-1)}\, P \,(1-KC)^T + KQK^T$$

$$\cancel{1-KC \Rightarrow \quad 1 - PC^T SC}$$

for thoroughness

$$\cancel{P(n|n) = P(n|n-1) - KCP(n|n-1) - PC^T K^T + KCPC^T K^T}$$

$$P_u = P - KCP - PC^T K^T + KCPC^T K^T + KQK^T$$

$$P_u = P - PC^T SCP - PC^T S^T C P^T + \cancel{PC^T SCPC^T S^T CP + PC^T SQS^T CP}$$

$$KCPC^T K^T + KQK^T$$

$$P_u = P - PC^T SCP - PC^T S^T C P^T + K(CPC^T + Q)K^T$$

$$P_u = P - KCP - \cancel{PC^T K^T} + \cancel{PC^T K^T}$$

$$P_u = P - PC^T SCP$$

$$P_u = P - KCP$$

$$\boxed{P_u = (1 - K(n) C_n)\, P(n|n-1)}$$

can ignore

okay... but deriving $K(n)$ equation??

$K(n)$? We want to minimize $\mathbb{E}[\|x_n - \hat{x}(n|n)\|^2]$

$$\hat{x}(n|n) = \hat{x}(n|n-1) + K(n)\left(y_n - C_n \hat{x}(n|n-1)\right)$$

$$e(n|n-1) = x_n - \hat{x}(n|n-1)$$

$$y_n - C_n \hat{x}(n|n-1) = C_n e(n|n-1) + v_n$$

$\hat{x}(n|n) = \hat{x}(n|n-1) + k(n)(C_n e(n|n-1) + v_n)$

$e(n|n) = \underline{x_n} - \hat{x}(n|n) = \cancel{e(n|n-1)} - (1 - k(n)C_n)e(n|n-1) - k(n)\underline{v_n}$

$J = E[e(n|n)^T e(n|n)]$

$\frac{\partial J}{\partial k(n)} = 0$

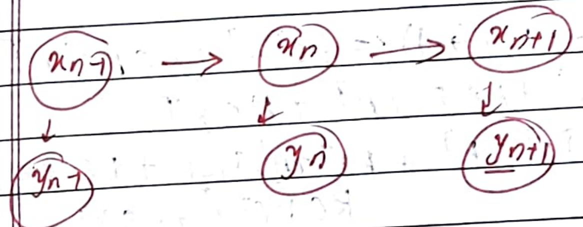
solves optimal

$k(n) = P(n|n-1)C_n^T [(C_n P(n|n-1) C_n^T) + Q_v(n)]^{-1}$

Bayesian filter $\rightarrow$ kalman filter

Kalman filters are a special case of bayesian filters

Bayesian filters: not necessarily (linear & only gaussian noise)



state $x_n = f_n(x_{n-1}, \omega_n)$

observation $= g_n(x_n, v_n)$

Given that i k transition & observation

can I find pdf $(x_{n-1}) \rightarrow$ pdf (x_n)

$P(\ ) \qquad P(\ )$

$P(\underline{x}_{n-1}) = P(\underline{x}_{n-1} | \underline{y}_1 \ldots \underline{y}_{n-1}) \rightarrow P(\underline{x}_{n-1} | y_{1:n-1})$

$\downarrow$

$P(\underline{x}_n) = P(\underline{x}_n | y_1 \ldots y_n) = P(\underline{x}_n | y_{1:n})$

Initialize $P(x_0)$ $\quad y_0 \rightarrow$ an observation

Transition

$P(\underline{x}_n | \underline{x}_{n-1})$ probability of going to nth state from n-1 th

Observation

$P(y_n | x_n)$.

Summary of all information entered

$\uparrow$ markov assumption

Using bayes rule

$$P(\underline{x}_n | y_{1:n}) = \frac{P(y_n | \underline{x}_n, y_{1:n-1})\, P(\underline{x}_n | y_{1:n-1})}{P(y_n | y_{1:n-1})}$$

doesnt depend on x ; cannot be evaluated ;

y_n depends only on $\underline{x}_n$ if $\underline{x}_n$ is present.
If $\underline{x}_n$ was not present we could explain purpose of $y_{1:n-1}$

$$P(\underline{x}_n | y_{1:n}) = \eta\, P(y_n | \underline{x}_n)\, P(x_n | y_{1:n-1})$$

likelihood function.

so from total probability law

$$P(\underline{x}_n | y_{1:n-1}) = \int_{\underline{x}_{n-1}} P(x_n, x_{n-1} | y_{1:n-1})\, d\underline{x}_{n-1}$$

$$\int_{\underline{x}_{n-1}} P(x_n | x_{n-1})\, P(x_{n-1} | y_{1:n-1})\, d\underline{x}_{n-1}$$

So

$$P(\underline{x}_n | y_{1:n}) = \eta\, P(y_n | \underline{x}_n) \int_{x_{n-1}} P(x_n | x_{n-1})\, P(x_{n-1} | y_{1:n-1})\, dx_{n-1}$$

← posterior dis → ← prior distribution →

recursive bayesian filtering

$$\eta = \frac{1}{P(y_n | y_{1:n-1})}$$ } constant, doesnt depend on x_n } scaling it wont change shape of probability dist

Kalman is a special case of this

$P(\underline{x}_n | x_{n-1}) \rightarrow$ transition $\rightarrow$ gaussian
$P(y_n | x_n) \rightarrow$ gaussian
$P(\underline{x}_{n-1} | y_{1:n-1}) \rightarrow$ gaussian

} $P(x_n | y_{1:n}) \rightarrow$ gaussian

$\hat{x}(n|n)$ & $P(n|n) \rightarrow$ mean & covar

Kalman from bayesian

@ $T = n-1$

mean covar

$$P(\underline{x}_{n-1} / y_{1:n-1}) \sim \mathcal{N}(\underline{x}_{n-1} ; \underline{\hat{x}}(n-1|n-1) ; P(n-1|n-1))$$

$$\mathcal{N}(\underline{x}_n , \underline{u} , \Sigma)$$

$$\Rightarrow P(\underline{x})$$

transition $x_n = A_n \underline{x}_{n-1} + \underline{w}_n$

$$P(\underline{x}_n | \underline{x}_{n-1}) \sim \mathcal{N}(\underline{x}_n ; A_n \cdot \underline{x}_{n-1} ; Q_w^{n})$$

Observation $y_n = C_n x_{n-1} + \underline{w}_n$

$$P(y_n / x_n) \sim \mathcal{N}(y_n ; C_n \underline{x}_n , Q_v^{\omega})$$

So; $P(\underline{x}_n / y_{1:n})$ is also gaussian

$$P(\underline{x}_n / y_{1:n}) \sim \mathcal{N}(\underline{x}_n ; \hat{x}(n|n), P(n|n))$$

Gaussian dis is mean.

mean apriori distribution after calc using mean

Notational meanings [why use weird notation?]

Pred $\hat{x}(n|n-1) \rightarrow$ estimate of state n given observation upto time $n-1$

updation $\hat{x}(n|n) \rightarrow$ Estimate of state at time n, given observations till time n (including y_n)

$n|n-1 \rightarrow$ prediction

$n|n \rightarrow$ updation.

$P(n|n-1) \rightarrow$ uncertainty before update

$P(n|n) \rightarrow$ uncertainty after update.

we want our, (and optimize for) $\mathbb{E}(e(n|n))$ ~~for~~ to be zero. Also; $J \rightarrow \mathbb{E}(e^T(n|n) e(n|n))$ to be minimized.