LEAST SQUARE METHOD

$$Y = C X$$

$Y \downarrow$ known, $C \downarrow$ known, $X \rightarrow$ unknown

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} c_{11}x_1 + c_{12}x_2 + \cdots + c_{1n}x_n \\ c_{21}x_1 + c_{22}x_2 - - - + c_{2n}x_n \\ \vdots \\ c_{m1}x_1 + \cdots + c_{mn}x_n \end{bmatrix}$$

$$y_1 = c_{11}x_1 + c_{12}x_2 + \cdots + c_{1n}x_n$$

$$y_2 = c_{21}x_1 + \cdots + c_{2n}x_n$$

$$\vdots$$

$$y_m = c_{m1}x_1 + \cdots + c_{mn}x_n$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}_{m\times 1} = \begin{bmatrix} c_{11} & c_{12} & & c_{1n} \\ & & & \\ c_{m1} & \cdots & & c_{mn} \end{bmatrix}_{m\times n} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}$$

$Y = CX$ (Linear relationship)

But when error is present then

$Y = CX + \text{error}$

$$\boxed{\therefore \text{error} = Y - CX}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = C X$$

$$\boxed{Y = CX}$$

Error vector $E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix} = \begin{bmatrix} y_1 - c_1 X \\ y_2 - c_2 X \\ \vdots \\ y_m - c_m X \end{bmatrix}$

$$\boxed{E = Y - CX}$$

$$P(X) = e_1^2 + e_2^2 + \cdots + e_m^2$$

$$= (y_1 - c_1 X)^2 + (y_2 - c_2 X)^2 + \cdots (y_m - c_m X)^2$$

We have to minimize $P(X)$:

$$P(X) = e^T \cdot e$$

$$= (Y - CX)^T (Y - CX)$$

$$= (Y^T - X^T C^T) \cdot (Y - CX)$$

$$= Y^T Y - X^T C^T Y - Y^T C X + X^T C^T C X$$

$$\frac{dP}{dX} = ??$$

(For what value of X our error^2 will be minimum i.e P will be minimum)

$$\frac{dP}{dX} = 0 - C^T Y - C^T Y + 2\, C^T C X$$

$$\frac{dP}{dX} = 0$$

$$-C^T Y - C^T Y + 2 C^T C X = 0$$

$$2 C^T Y = 2 C^T C X$$

$$(C^T C) X = C^T Y$$

$$(C^T C)^{-1} (C^T C) X = (C^T C)^{-1} C^T Y$$

$$\boxed{X = (C^T C)^{-1} C^T Y}$$

$$\boxed{X_{\text{least square}} = (C^T C)^{-1} C^T Y}$$

Thus if we have m number of samples of C and Y then we can find the unknown variable X by using least square method.

KALMAN FILTER

- Discrete time Kalmon filter
- Continuous time Kalman filter

* In the Kalman filter, we aim to estimate the **state** x (which could represent position, velocity or any other system variable) based on the **output** y (which is the noisy measurement from sensors or observations).

* x represents the true but hidden state of the system.

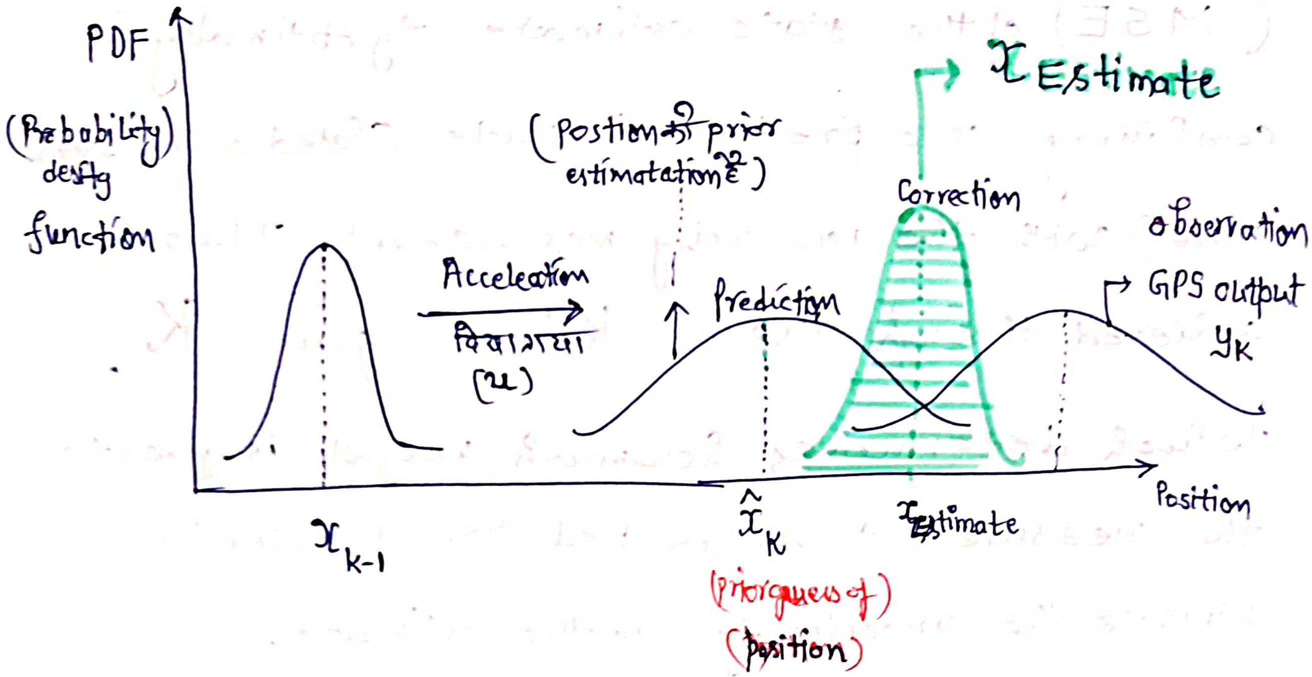
* y represents the observable measurement that are influenced by noise.

* The parameter that is minimized is the **estimation error covariance**, often denoted as P.
This covariance represents the uncertainty in the estimate of system's state.

* The kalman filter minimizes the mean square error (MSE) of the state estimate by optimally combining the predicted state (based on the system model) with the new noisy measurement. This is achieved through the Kalman gain K, which determines how much weight is given to the measurement verses prediction in order to minimize the uncertainty in the estimate.

* Thus key parameter being minimized is the uncertainty (variance) of the state estimate, ensuring that the estimate is as close as possible to the true state.

Theory



Initially हम x_{k-1} position पर है। जैसे ही हम Input को बढ़ाते हैं (यहां Input acceleration है) हमारी गाड़ी का position बदल जाता है।

यहां पर नई Position की Predicted value $\hat{x}_k$ है यह prior guess है।

हमारा model ऐसा है कि गाड़ी में sensor लगा है (मतलब GPS लगा है) जो Postion को measure कर रहा है इसे y से लिखते हैं।

GPS से Position की मिली हुई value $= [y_K]_{GPS}$

or y_K

गाड़ी हिलती रहती है अत: $[y_K]_{GPS}$ मिल तो रहा है पर हम Kalman के method से इसकी True value को estimate करते हैं और ये $x_{estimate}$ कहलाता है।

$$\therefore\ x_{estimate} = \hat{x}_K + K\cdot(\text{Error in measurement}).$$

$$= \hat{x}_K + K \cdot (\text{Correction term})$$

$$= \hat{x}_K + K\left([y_K]_{GPS} - [\hat{y}_K]\right)$$

$$\boxed{x_{estimate} = \hat{x}_K + K\,[\,y_K - \hat{y}_K\,]} \quad \text{—— (1)}$$

$\hat{x}_K$ → prior guess

y_K → GPS output

$\hat{y}_K$ → Predicted position (it linearly depends on $\hat{x}_K$) $\hat{y}_K = C\,\hat{x}_K$

K → Kalman gain

हम K की value को बदलते रहते हैं और हम Position को measure करते रहते हैं एक time पर हमें $x_{estimate}$ की सबसे true value मिल जाती है।

Note :-

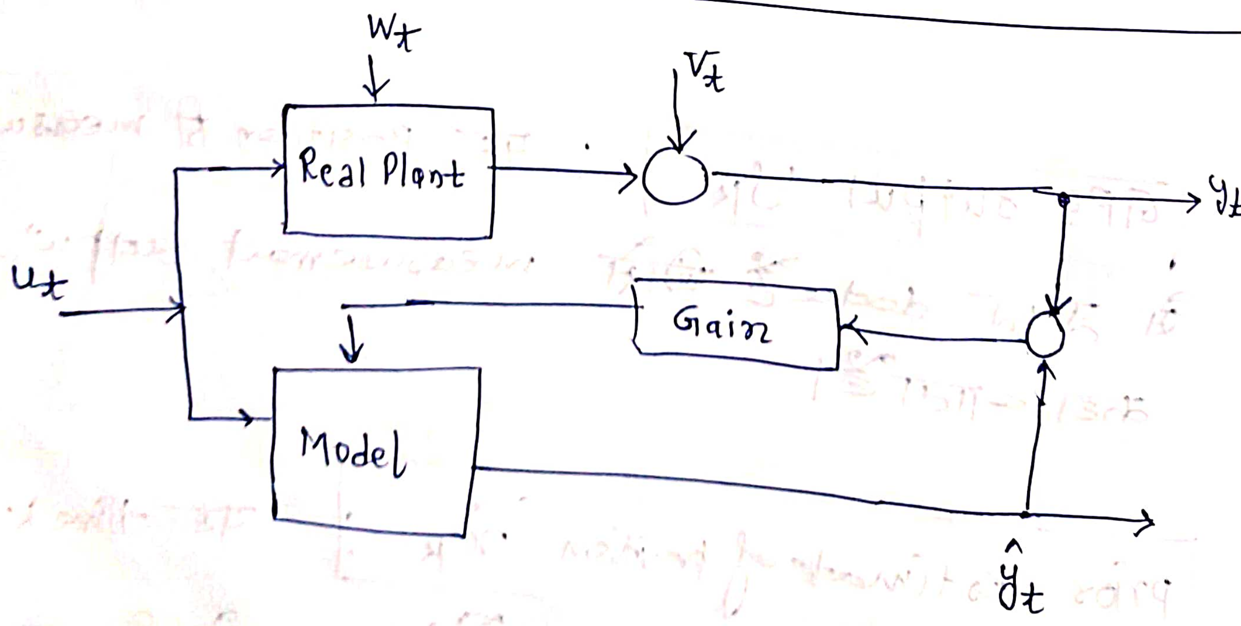
$\boxed{\text{GPS output } y_K}$: यह Position की measurement से प्राप्त data है जिसे measurement output कहा जाता है।

$\boxed{\text{prior estimate of position } \hat{x}_K}$: यह time K पर Position का अनुमानित data है। मान लो कि कार के wheel का Circumference हमें पता है फिर हम इसे time और number of rotation से multiply करेंगे

नई position का अनुमान लगा सकते हैं। अत: एक guess value होगी क्योंकि नई position को गाड़ी के और भी variables प्रभावित कर रहे हैं।

Note: किसी भी model के सारे variables को हम measure नहीं कर सकते हैं।

- जिनको हम measure कर लेते हैं उनको output y कहा जाता है
- इस output y की मदद से model के internal सारे variables हम measure कर सकते हैं। यही इस Technique का फायदा है।



Kalman Filter

Example :-

Position, velocity, and acceleration estimation in a kalman filter :

We will now define the state vector to include position(p) velocity(v) and acceleration(a).

State vector $X(k)$

$$X(k) = \begin{bmatrix} p(k) \\ v(k) \\ a(k) \end{bmatrix}$$

$p(k)$ is the position at time step k.
$v(k)$ is the velocity at time step k.
$a(k)$ is the acceleration at time step k.

State equation (Dynamics of the system): Now we need to describe how position, velocity and acceleration evolve overtime.

1. Position update :- Position at time k+1 is given as:

$$p(k+1) = p(k) + v(k)\,\Delta t + \frac{1}{2} a(k)\,\Delta t^2$$

2. Velocity update :- The velocity at time k+1 is given as:

$$v(k+1) = v(k) + a(k)\,\Delta t$$

3. Acceleration update : We assume the acceleration might change slowly, but it remains constant over small time interval

or it evolves with some noises.

$$a(k+1) = a(k) + w_a(k)$$

where $w_a(k)$ is the process noise for the acceleration.

these updates can be combined into a state space model. the state equation becomes:.

$$\boxed{x(k+1) = A\, x(k) + W(k)}$$

where $A = \begin{bmatrix} 1 & \Delta t & \frac{1}{2}\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}$ = state transition matrix

$W(k) = \begin{bmatrix} w_p(k) \\ w_v(k) \\ w_a(k) \end{bmatrix}$ = Process noise vector representing uncertainties in the position velocity & acceleration

Output Equation (Measurement Model): Now, suppose we still only measuring the position of the object using a sensor (e.g. a GPS). The output vector is

$$\boxed{y(k) = p(k) + v_p(k)}$$

$v_p(k)$ is the measurement noise affecting the position measurement

In matrix form, the measurement equation becomes :

$$y(k) = C\,x(k) + v(k)$$

where $C = [1 \quad 0 \quad 0]$

This means we are only measuring the position $p(k)$, and not the velocity or acceleration.

<u>State space equation for Kalman filter</u> :

$$x(k+1) = \begin{bmatrix} 1 & \Delta t & \frac{1}{2}\Delta t^2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} x(k) + \underbrace{W(k)}_{\text{Process noise}}$$

$$y(k) = [1 \quad 0 \quad 0]\, x(k) + \underbrace{V_p(k)}_{\text{Measurement noise}}$$

Why include position in the state vector (despite measuring it)??

- Position is included in the state vector because:

* The state vector models the entire system dynamics which requires knowing how position, velocity and acceleration evolve overtime.

* Even though position is measured, it is noisy, and the kalman filter estimates the true state (including positions) by combining the prediction and the measurement.

* Position, velocity and acceleration are linked, and measuring position helps the filter (kalman) update estimates of velocity and acceleration as well.

Note In kalman filter, state estimation is not just about observing the measured variable (position), it is about estimating the true internal state by integrating information from both the prediction (dynamics) and noisy measurements.

Imp Note

let we want to estimate position, velocity and acceleration using a kalman filter.

- $\hat{x}_k$ would include these three variables.
- P_k will represent uncertainties (variances) of these estimates and their covariances.

State Vector:

$$\hat{x}_k = \begin{pmatrix} x_k \\ v_k \\ a_k \end{pmatrix}$$

x_k : Position

v_k : Velocity

a_k : Acceleration

Covariance matrix P_k :

$$P_k = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}$$

P_{11} : Variance of the position estimate $[Var(x_k)]$

P_{22} : Variance of velocity estimate $[Var(v_k)]$

P_{33} : Variance of acceleration estimate $[Var(a_k)]$

$$\left.\begin{matrix} P_{12} = P_{21} \\ P_{13} = P_{31} \\ P_{32} = P_{23} \end{matrix}\right\} \text{Covariances between position velocity and acceleration}$$

Trace of covariance matrix:

$$\boxed{\text{Trace } P_k = P_{11} + P_{22} + P_{33}}$$

- The trace can be interpreted as the ~~total~~ total uncertainty in the state estimates.
- Trace (P_k) is the sum of error variances (squared errors) of position, velocity and acceleration

Cost function:

$$\boxed{\text{Cost function } J = P_{11} + P_{22} + P_{33}}$$

P_{11} = Variance (x_k) = uncertainty in position estimate

P_{22} = Variance (v_k) = uncertainty in velocity estimate

P_{33} = Variance (a_k) = uncertainty in acceleration estimate.

By minimizing J, the kalman filter attempts to reduce the total uncertainty in the estimates of position, velocity, and acceleration.

Recursive least square estimation

$$X_K = A_{K-1} x_{K-1} + B_{K-1} u_{K-1} + W_{K-1}$$

$$y_K = C_K X_K + V_K$$

$$\hat{x}_K = \hat{x}_{K-1} + \underset{\substack{\downarrow \\ \text{Kalman gain}}}{K_K} [y_K - C_K \hat{x}_{K-1}]$$

$$= \hat{x}_{K-1} + K_K [C_K X_K + V_K - C_K \hat{X}_{K-1}]$$

$$= \hat{X}_{K-1} + K_K C_K (X - \hat{X}_{K-1}) + K_K V_K$$

$$\text{Error} = \mathcal{E}_{K-1} = X - \hat{X}_{K-1}$$

$$\text{Error} = \mathcal{E}_K = X - \hat{X}_K$$

<u>Note</u>

$$P_K = E[\mathcal{E}_K \cdot \mathcal{E}_K^T]$$

$$P_K = E[(x - \hat{x}_K)(x - \hat{x}_K)^T] \quad \text{③}$$

$$P_{K-1} = E[(x - \hat{x}_{K-1})(x - \hat{x}_{K-1})^T] \quad \text{④}$$

Recursive Least Square Estimation

Gain matrix update

$$K_k = P_{k-1} C_k^T \left(R_k + C_k P_{k-1} C_k^T\right)^{-1}$$

Estimate update

$$\hat{x}_k = \hat{x}_{k-1} + K_k \left(y_k - C_k \hat{x}_{k-1}\right)$$

Error covariance :

$$P_k = (I - K_k C_k) P_{k-1} (I - K_k C_k)^T + K_k R_k K_k^T$$

or

$$P_k = (I - K_k C_k) P_{k-1}$$

$$\boxed{y_k = C_k x + v_k}$$

$y_k \rightarrow$ observation

C_k = Measurement matrix

$x \rightarrow$ Parameter vector that we want to estimate

$v_k \rightarrow$ measurement noise

$E[v_k] = 0$

$E[v_k v_k^T] = R_k$

= Covariance of the measurement noise

$$\boxed{\hat{x}_k = \hat{x}_{k-1} + K_k (y_k - C_k \hat{x}_{k-1})}$$

$K_k \rightarrow$ Gain matrix

$k \rightarrow$ Discrete time instant (Written here lower suffix)

$$\hat{x}_k = \begin{bmatrix} \hat{x}_{1k} \\ \hat{x}_{2k} \\ \vdots \\ \hat{x}_{nk} \end{bmatrix} \cong \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \\ x_n \end{bmatrix}$$

$$\varepsilon_k = \begin{bmatrix} \varepsilon_{1k} \\ \varepsilon_{2k} \\ \varepsilon_{nk} \end{bmatrix} = x - \hat{x}_k = \begin{bmatrix} x_1 - \hat{x}_{1k} \\ x_2 - \hat{x}_{2k} \\ \\ x_n - \hat{x}_{nk} \end{bmatrix}$$

$$\varepsilon_{ik} = x_i - \hat{x}_{ik} \qquad i = 1, 2, \dots, n$$

Gain K_k is computed by minimizing the sum of variances of estimation errors.

$$W_k = E[\varepsilon_{1k}^2] + E[\varepsilon_{2k}^2] \dots + E[\varepsilon_{nk}^2]$$

$$W_k = \text{trace}(P_k)$$

where P_k = estimation error covariance matrix

$$\boxed{P_k = E[\varepsilon_k \cdot \varepsilon_k^T]}$$

$$= E[(\text{error}) \cdot (\text{error})^T]$$

Note :-

$$\varepsilon_k \cdot \varepsilon_k^T = \begin{bmatrix} \varepsilon_{1k} \\ \varepsilon_{2k} \\ \vdots \\ \varepsilon_{nk} \end{bmatrix} \begin{bmatrix} \varepsilon_{1k} & \varepsilon_{2k} & \dots & \varepsilon_{nk} \end{bmatrix}$$

$$= \begin{bmatrix} \varepsilon_{1k}^2 & \varepsilon_{1k}\varepsilon_{2k} & & \varepsilon_{1k}\varepsilon_{nk} \\ \varepsilon_{2k}\varepsilon_{1k} & \varepsilon_{2k}^2 & & \varepsilon_{2k}\varepsilon_{nk} \\ & & \ddots & \varepsilon_{nk}^2 \\ \varepsilon_{1k}\varepsilon_{nk} & & & \end{bmatrix}$$

$$P_k = \underbrace{E}_{\text{expectation}}[\varepsilon_k \cdot \varepsilon_k^T]$$

$\varepsilon_k = x - \hat{x}_k$ | $\varepsilon_{k-1} = x - \hat{x}_{k-1}$

P_k = Estimation error covariance matrix.

$$P_k = E\left[\varepsilon_k \cdot \varepsilon_k^T\right]$$

$$trace(P_k) = E[\varepsilon_{1k}^2] + E[\varepsilon_{2k}^2] + \cdots + E[\varepsilon_{nk}^2]$$

Estimation error:.

$$\varepsilon_k = x - \hat{x}_k$$

$$= x - [\hat{x}_{k-1} + K_k(y_k - C_k\hat{x}_{k-1})]$$

$$= x - \hat{x}_{k-1} - K_k(y_k - C_k\hat{x}_{k-1})$$

$$= x - \hat{x}_{k-1} - K_k(C_k x + v_k - C_k\hat{x}_{k-1})$$

$$= x - \hat{x}_{k-1} - K_k C_k x - K_k v_k + K_k C_k \hat{x}_{k-1}$$

$$= (I - K_k C_k)(x - \hat{x}_{k-1}) - K_k v_k$$

$$\boxed{\varepsilon_k = (I - K_k C_k)\,\varepsilon_{k-1} - K_k v_k}$$

$$\therefore \varepsilon_k \cdot \varepsilon_k^T = [(I - K_k C_k)\varepsilon_{k-1} - K_k v_k][(I - K_k C_k)\varepsilon_{k-1} - K_k v_k]^T$$

$$= ((I - K_k C_k)\varepsilon_{k-1} - K_k v_k)(\varepsilon_{k-1}^T (I - K_k C_k)^T - v_k^T K_k^T)$$

$$= (I - K_k C_k)\varepsilon_{k-1} \cdot \varepsilon_{k-1}^T (I - K_k C_k)^T - (I - K_k C_k)\varepsilon_{k-1} v_k^T K_k^T$$

$$- K_k v_k \varepsilon_{k-1}^T (I - K_k C_k)^T + K_k v_k v_k^T K_k^T$$

$$\therefore \boxed{P_k = \mathbf{E}[\varepsilon_k \cdot \varepsilon_k^T]}$$

(**E** means estimation) ;

$$P_k = (I - K_k C_k)\, E[\varepsilon_{k-1}\varepsilon_{k-1}^T](I - K_k C_k)^T - (I - K_k C_k)\, E[\underbrace{\varepsilon_{k-1} v_k^T}_{\to 0}] K_k^T$$

$$- K_k E[\underbrace{v_k \varepsilon_{k-1}^T}_{\to 0}](I - K_k C_k)^T + K_k E[v_k \cdot v_k^T] K_k^T$$

$$P_{k-1} = E[\varepsilon_{k-1} \varepsilon_{k-1}^T]$$

$$E[\varepsilon_{k-1} V_k^T] = E[\varepsilon_{k-1}] \cdot E[V_k^T] = 0$$

$$E[V_k \varepsilon_{k-1}^T] = E[V_k] \cdot E[\varepsilon_{k-1}^T] = 0$$

$$E[V_k V_k^T] = R_k$$

$$\boxed{P_k = (I - K_k C_k) P_{k-1} (I - K_k C_k)^T + K_k R_k K_k^T}$$

$$= (P_{k-1} - K_k C_k P_{k-1})(I - C_k^T K_k^T) + K_k R_k K_k^T$$

$$= P_{k-1} - P_{k-1} C_k^T K_k^T - K_k C_k P_{k-1} + K_k C_k P_{k-1} C_k^T K_k^T + K_k R_k K_k^T \quad \text{——— (1)}$$

Cost function W_k :-

$$\boxed{W_k = trace(P_k)}$$

$$W_k = tr(P_{k-1}) - tr(P_{k-1} C_k^T K_k^T) - tr(K_k C_k P_{k-1}) + tr(K_k C_k P_{k-1} C_k^T K_k^T)$$
$$+ tr(K_k R_k K_k^T) \quad \text{——— (2)}$$

Now we want to optimize cost function W_k.

$$\therefore \frac{\partial W_k}{\partial K_k} = 0 \quad \text{——— (3)}$$

$$\therefore \frac{\partial}{\partial K_k}[tr(P_k)] = 0$$

Note : Reference : Matrix CookBook

$$\frac{\partial}{\partial X} \operatorname{Tr}(AX^T) = A$$

$$\frac{\partial}{\partial X} \operatorname{Tr}(XA) = A^T$$

$$\frac{\partial}{\partial X} \operatorname{Tr}(XBX^T) = XB^T + XB$$

$$= X(B^T + B)$$

$$= 2XB \qquad (\text{if } B^T = B)$$

$$\frac{\partial}{\partial K_K}\left[tr(P_{K-1})\right] = 0$$

$$\frac{\partial\left[tr\left(P_{K-1}\, C_K^T\, K_K^T\right)\right]}{\partial K_K} = P_{K-1}\, C_K^T$$

$$\frac{\partial}{\partial K_K}\left[tr\left(K_K\, C_K\, P_{K-1}\right)\right] = P_{K-1}\, C_K^T$$

$$\frac{\partial}{\partial K_K}\left[tr\left(K_K\, C_K\, P_{K-1}\, C_K^T\, K_K^T\right)\right] = 2 K_K\, C_K\, P_{K-1}\, C_K^T$$

$$\cancel{\frac{\partial}{\partial K_K}\, tr\left(K_K\, C_K\, P_{K-1}\right) = P_{K-1}\, C_K^T}$$

$$\frac{\partial}{\partial K_K}\left[tr\left(K_K\, R_K\, K_K^T\right)\right] = 2\, K_K\, R_K$$

(4)

from (2) (3) and (4) :

equation (3) will become :

$$0 - P_{K-1}\, C_K^T - P_{K-1}\, C_K^T + 2 K_K\, C_K\, P_{K-1} C_K^T + 2 K_K R_K = 0$$

$$-2 P_{K-1}\, C_K^T + 2\, K_K\, C_K\, P_{K-1}\, C_K^T + 2 K_K R_K = 0$$

$$P_{K-1}\, C_K^T = K_K\, C_K\, P_{K-1} C_K^T + K_K R_K$$

$$K_K\left[R_K + C_K\, P_{K-1}\, C_K^T\right] = P_{K-1}\, C_K^T$$

$$\boxed{K_K = P_{K-1}\, C_K^T\, \left[R_K + C_K\, P_{K-1} C_K^T\right]^{-1}} \quad (5)$$

Note

$$\left\{\left(R_k + C_k P_{k-1} C_k^T\right)^{-1}\right\}^T = \left(R_k + C_k P_{k-1} C_k^T\right)^{-1}$$

$\because$ $R_k + C_k P_{k-1} C_k^T$ is symmetric in nature.

Now put the value of K_K into equation (1)

$$P_K = P_{K-1} - P_{K-1} C_K^T \{P_{K-1} C_K^T [R_K + C_K P_{K-1} C_K^T]^{-1}\}^T$$
$$- P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} \cdot C_K P_{K-1}$$
$$+ \{P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1}\} C_K P_{K-1} C_K^T \cdot \{P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)\}^{-1 T}$$
$$+ \{P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1}\} \cdot R_K \cdot \{P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1}\}^T$$

$$= P_{K-1} - P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} C_K P_{K-1}$$
$$- P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} C_K P_{K-1}$$
$$+ P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} C_K P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} C_K P_{K-1}$$
$$+ P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} \cdot R_K \cdot (R_K + C_K P_{K-1} C_K^T)^{-1} C_K P_{K-1}$$

$$= P_{K-1} - 2 P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} C_K P_{K-1}$$
$$P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} \cancel{[C_K P_{K-1} C_K^T + R_K]} \cancel{(R_K + C_K P_{K-1} C_K^T)^{-1}} C_K P_{K-1}$$

$$= P_{K-1} - 2 P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} C_K P_{K-1}$$
$$+ P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} C_K P_{K-1}$$
$$= P_{K-1} - P_{K-1} C_K^T (R_K + C_K P_{K-1} C_K^T)^{-1} C_K P_{K-1}$$
$$= P_{K-1} - K_K C_K P_{K-1}$$

$$P_k = P_{k-1} - K_k C_k P_{k-1}$$

$$\boxed{P_k = (I - K_k C_k) P_{k-1}}$$

Thus the recursive least square method consists of the following three equations.

Gain Matrix Update:

$$K_k = P_{k-1} C_k^T (R_k + C_k P_{k-1} C_k^T)^{-1}$$

Estimation update:

$$\hat{x}_k = \hat{x}_{k-1} + K_k (y_k - C_k \hat{x}_{k-1})$$

Error covariance matrix:

$$P_k = (I - K_k C_k) P_{k-1} (I - K_k C_k)^T + K_k R_k K_k^T$$

or

$$P_k = (I - K_k C_k) P_{k-1}$$

$k=0$ $\quad \hat{x}_0 , \hat{P}_0$

$k=1$ $\quad K_1 = P_0 C_1^T (R_1 + C_1 P_0 C_1^T)^{-1}$

$$\hat{x}_1 = \hat{x}_0 + K_1 (y_1 - C_1 \hat{x}_0)$$

$$P_1 = (I - K_1 C_1) P_0 (I - K_1 C_1)^T + K_1 R_1 K_1^T$$

or $\quad P_1 = (I - K_1 C_1) P_0$

KALMAN FILTER

$$x_k = A_{k-1}\, x_{k-1} + B_{k-1}\, u_{k-1} + w_{k-1}$$

$$y_k = C_k\, x_k + v_k$$

$x_{k-1} \rightarrow$ state vector at the discrete time instant k

$u_{k-1} \rightarrow$ control input vector

$w_{k-1} \rightarrow$ process noise vector

$$\boxed{Q_k = E[w_k \cdot w_k^T]}$$

Q_k is the covariance matrix

$E[\cdot]$ is the mathematical expectation operator

$\left.A_{k-1},\ B_{k-1}\right\}$ state matrix & input matrix

$C_k \rightarrow$ output matrix

$y_k \rightarrow$ output vector (observed measurements)

$v_k \rightarrow$ Measurement noise vector

$$\boxed{R_k = E[v_k\, v_k^T]}$$

R_k is the covariance matrix

Note: this is a time varying system. However if the system is time invariant then all the system matrices are constant. We assume that A_k, B_k, C_k, Q_k and R_k are known. Our objective is to design an observer estimator that will estimate the state vector x_k

Real Plant $\Rightarrow$

$$\left[\begin{aligned} x_k &= A_{k-1}\, x_{k-1} + B_{k-1}\, u_{k-1} + w_{k-1} \\ y_k &= C_k\, x_k + v_k \end{aligned}\right]$$

A priori estimate $\left\{ \hat{x}_k^- = A_{k-1}\, \hat{x}_{k-1}^+ + B_{k-1}\, u_{k-1} \right.$ (A priori estimate equation)

In kalman filter :-

$$\varepsilon_k^- = x_k - \hat{x}_k^- = \left[A_{k-1}\, x_{k-1} + \cancel{B_{k-1} u_{k-1}} + w_{k-1}\right] - \left[A_{k-1} \hat{x}_{k-1}^+ + \cancel{B_{k-1} u_{k-1}}\right]$$

$$= A_{k-1}\,(x_{k-1} - \hat{x}_{k-1}^+) + w_{k-1} = A_{k-1}\, \varepsilon_{k-1}^+ + w_{k-1}$$

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

$$P_k^- = E\left[\varepsilon_k^- \varepsilon_k^{-T}\right]$$

$$P_k^+ = E\left[\varepsilon_k^+ \varepsilon_k^{+T}\right]$$

एप्रायोराई

A priori estimate :

$$\hat{x}_k^- = E[\, x_k \mid y_1, y_2, y_3 \cdots \cdots\, y_{k-1}]$$

= estimate of x_k before we process the measurement y_k at time k

ए- पोस्टी-रियोराई

A posteriori estimate

$$\hat{x}_k^+ = E[\, x_k \mid y_1, y_2, \ldots\ldots, y_k]$$

= estimate of x_k after we process the measurement y_k at time k

Covariance matrix :

$$P_k^- = E[(x_k - x_k^-)(x_k - x_k^-)^T]$$

$$P_k^+ = E[(x_k - x_k^+)(x_k - x_k^+)^T]$$

$$\varepsilon_k^- = x_k - \hat{x}_k^-$$

$$= (A_{k-1} x_{k-1} + \cancel{B_{k-1} u_{k-1}} + w_{k-1}) - (A_{k-1} \hat{x}_{k-1}^+ + \cancel{B_{k-1} u_{k-1}})$$

$$= A_{k-1} (x_{k-1} - x_{k-1}^+) + w_{k-1}$$

$$= A_{k-1} \varepsilon_{k-1}^+ + w_{k-1}$$

Note

In kalman filter :-

$$\varepsilon_k^- = x_k - \hat{x}_k^- \quad \text{or} \quad x - \hat{x}_k^-$$

$$\varepsilon_k^+ = x_k - \hat{x}_k^+ \quad \text{or} \quad x - \hat{x}_k^+$$

A posteriori :-

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

$$= x - \hat{x}_k^+ \quad \text{(General form)}$$

$$= x - [\hat{x}_k^- + K_k (y_k - C_k \hat{x}_k^-)]$$

$$= x - \hat{x}_k^- - K_k (y_k - C_k \hat{x}_k^-)$$

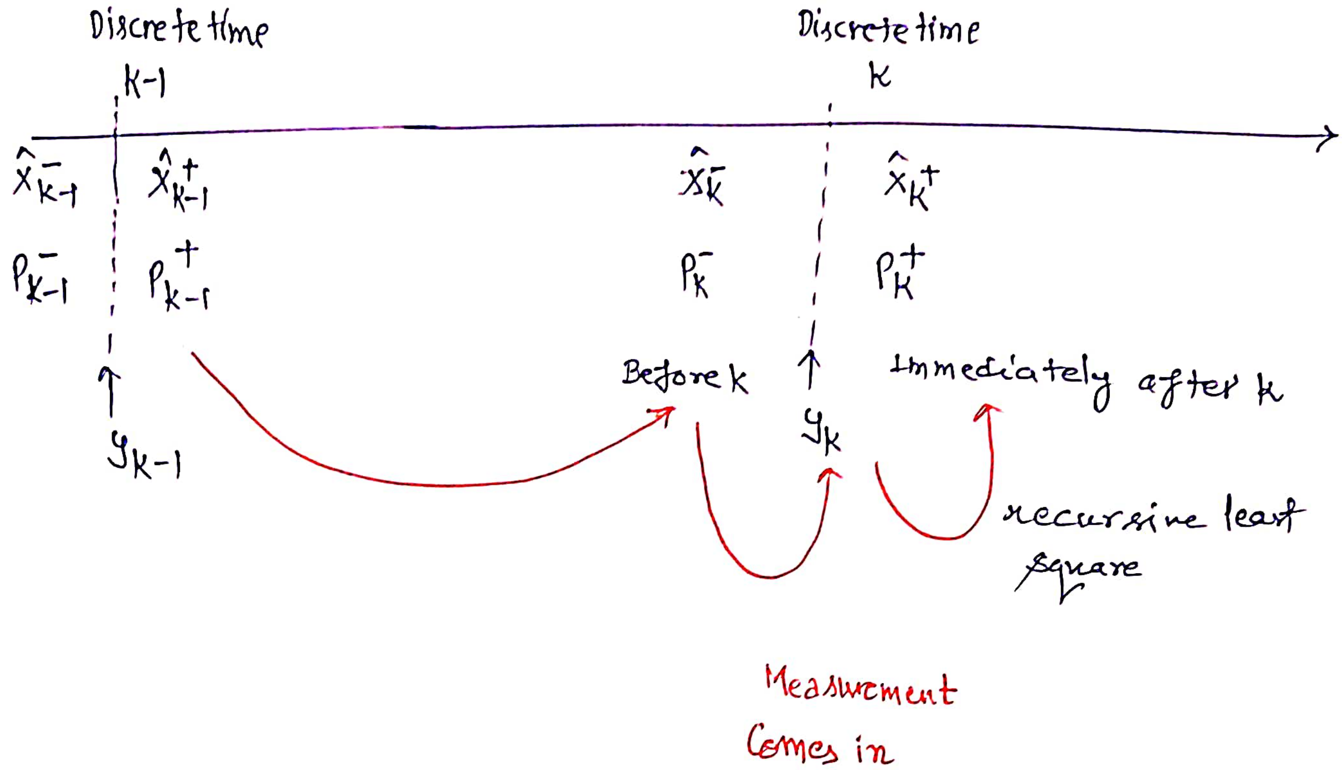
$$= x - \hat{x}_k^- - K_k (C_k x + V_k - C_k \hat{x}_k^-)$$

$$= x - \hat{x}_k^- - K_k C_k x - K_k V_k + K_k C_k \hat{x}_k^-$$

$$= (x - \hat{x}_k^-) - K_k C_k (x - \hat{x}_k^-) - K_k V_k$$

$$= (I - K_k C_k)(x - \hat{x}_k^-) - K_k V_k$$

$$\varepsilon_k^+ = (I - K_k C_k)\, \varepsilon_k^- - K_k V_k$$



<u>Recursive least square equations:</u>

$$P_k = A_{k-1} P_{k-1} A_{k-1}^T + Q_{k-1}$$

$$K_k = P_{k-1} C_k^T \left(R_k + C_k P_{k-1} C_k^T\right)^{-1}$$

$$\hat{x}_k = \hat{x}_{k-1} + K_k \left(y_k - C_k \hat{x}_{k-1}\right)$$

$$P_k = (I - K_k C_k) P_{k-1} (I - K_k C_k)^T + K_k R_k K_k^T$$

or

$$P_k = (I - K_k C_k) P_{k-1}$$

Recursive least square method variable	Kalman filter variable	
P_{k-1}	P_k^-	= a priori covariance
P_k	P_k^+	= a posteriori covariance
$\hat{x}_{k-1}$	$\hat{x}_k^-$	= a priori estimate
$\hat{x}_k$	$\hat{x}_k^+$	= a posteriori estimate

After these substitutions, we obtain the updated equations:

$$\left.\begin{aligned}
K_k &= P_k^- C_k^T \left(R_k + C_k P_k^- C_k^T\right)^{-1} \\
\hat{x}_k^+ &= \hat{x}_k^- + K_k \left(y_k - C_k \hat{x}_k^-\right) \\
P_k^+ &= \left(I - K_k C_k\right) P_k^- \left(I - K_k C_k\right)^T + K_k R_k K_k^T \\
P_k^+ &= \left(I - K_k C_k\right) P_k^-
\end{aligned}\right\} \text{A posteriori estimate}$$

$$\left.\begin{aligned}
\hat{x}_k^- &= A_{k-1} \hat{x}_{k-1}^+ + B_{k-1} u_{k-1} \\
P_k^- &= A_{k-1} P_{k-1}^+ A_{k-1}^T + Q_{k-1}
\end{aligned}\right\} \text{A priori estimate}$$

For $k=1$

$$\left.\begin{aligned}
\hat{x}_1^- &= A_0 \hat{x}_0^+ + B_0 u_0 \\
P_1^- &= A_0 P_0^+ A_0^T + Q_0
\end{aligned}\right\} \text{A priori}$$

$$\left.\begin{aligned}
K_1 &= P_1^- C_1^T \left(R_1 + C_1 P_1^- C_1^T\right)^{-1} \\
\hat{x}_1^+ &= \hat{x}_1^- + K_1 \left(y_1 - C_1 \hat{x}_1^-\right)
\end{aligned}\right\} \text{A posteriori}$$

$$P_1^+ = (I - K_1 C_1)\; P_1^- \;(I - K_1 C_1)^T + K_1 C_1 K_1^T$$

$$\text{or } P_1^+ = (I - K_1 C_1)\; P_1^-$$

A priori error covariance:

$$P_k^- = E[\varepsilon_k^- \cdot \varepsilon_k^{-T}]$$

$$= E\left\{[A_{k-1}\,\varepsilon_{k-1}^+ + w_{k-1}] \cdot [A_{k-1}\,\varepsilon_{k-1}^+ + w_{k-1}]^T\right\}$$

$$= E\left\{(A_{k-1}\,\varepsilon_{k-1}^+ + w_{k-1})\,(\varepsilon_{k-1}^{+T}\,A_{k-1}^T + w_{k-1}^T)\right.$$

$$= E\left\{A_{k-1}(\varepsilon_{k-1}^+\;\varepsilon_{k-1}^{+T})\,A_{k-1}^T + w_{k-1}\,\varepsilon_{k-1}^{+T}\,A_{k-1}^T\right.$$

$$\left. + A_{k-1}\,\varepsilon_{k-1}^+\,w_{k-1}^T + w_{k-1}\,w_{k-1}^T\right\}$$

$$= A_{k-1}\,E\{\varepsilon_{k-1}^+\;\varepsilon_{k-1}^{+T}\}A_{k-1}^T + \underbrace{E\{w_{k-1}\,\varepsilon_{k-1}^{+T}\}}_{0}$$

$$+ A_{k-1}\,\underbrace{E\{\varepsilon_{k-1}^+\,w_{k-1}^T\}}_{0} + E\{w_{k-1}\,w_{k-1}^T\}$$

$$= A_{k-1}\;P_{k-1}\;A_{k-1}^T + Q_{k-1}$$

$$\therefore \boxed{P_k^- = A_{k-1}\;P_{k-1}^+\;A_{k-1}^T + Q_{k-1}}$$

Summary of the Kalman filter

Initiate estimate of state $= \hat{x}_0^+$

initial Covariance matrix of estimation error $= P_0^+$

for $k = 1, 2, 3 \ldots$ we perform the step 1 and step 2 as given below:.

Step I

$$\hat{x}_k^- = A_{k-1} \hat{x}_{k-1}^+ + B_{k-1} u_{k-1}$$

$$P_k^- = A_{k-1} P_{k-1}^+ A_{k-1}^T + Q_{k-1}$$

Step II

$$K_k = P_k^- C_k^T \left(R_k + C_k P_k^- C_k^T\right)^{-1}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k \left(y_k - C_k \hat{x}_k^-\right)$$

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

or $$P_k^+ = (I - K_k C_k) P_k^-$$

Note

$x_k \rightarrow$ state Vector

$y_k \rightarrow$ Measurement vector

$u_k \rightarrow$ Input control vector

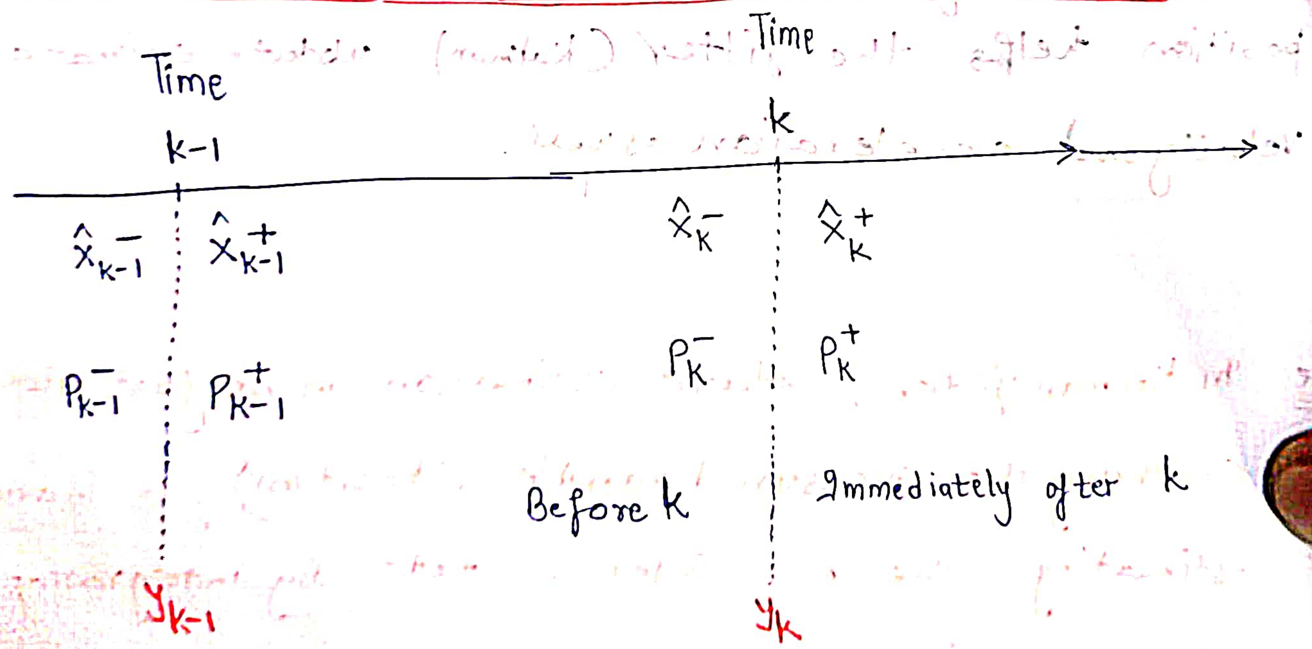
w_k : Process Noise

v_k : Measurement noise

Q_k : Symmetric, positive semidefinite, $Q \geq 0$

R_k : Symmetric, positive definite, $R > 0$

P_k : Symmetric, positive semidefinite, $P_k \geq 0$



Note: Discrete time instant $= k$

Note: The lower suffix k refers to the time step or discrete time Index.

KALMAN FILTER

$$x_k = A_{k-1} x_{k-1} + B_{k-1} u_{k-1} + w_{k-1}$$

$$y_k = C_k x_k + v_k$$

Note:

A_{k-1} , B_{k-1} , C_k , u_{k-1} , y_k } Known

$$\left.\begin{aligned} Q_k &= E\{w_k \cdot w_k^T\} \\ R_k &= E\{v_k \cdot v_k^T\} \end{aligned}\right\} \text{known}$$

state vector x_k } unknown

A priori estimation:

$$\hat{x}_k^- = A_{k-1} \hat{x}_{k-1}^+ + B_{k-1} u_{k-1}$$

$$P_k^- = E[\varepsilon_k^- \cdot \varepsilon_k^{-T}]$$

$$P_k^- = A_{k-1} P_{k-1}^+ A_{k-1}^T + Q_{k-1}$$

Proof: Let x_k is the parameter vector that we want to estimate.

$$\varepsilon_k^- = x_k - \hat{x}_k^- \quad \text{or} \quad x - \hat{x}_k^- \quad \text{(General form)}$$

$$= \{A_{k-1} x_{k-1} + \cancel{B_{k-1} u_{k-1}} + w_{k-1}\}$$

$$-\{A_{k-1} x_{k-1}^+ + \cancel{B_{k-1} u_{k-1}}\}$$

$$= A_{k-1} x_{k-1} - A_{k-1} \hat{x}_{k-1}^+ + w_{k-1}$$

$$= A_{k-1}(x_{k-1} - x_{k-1}^+) + w_{k-1} = A_{k-1} \varepsilon_{k-1}^+ + w_{k-1}$$

Note :

$$\varepsilon_k^- = x_k - \hat{x}_k^- \quad \text{or} \quad x - \hat{x}_k^-$$

$$\varepsilon_k^+ = x_k - \hat{x}_k^+ \quad \text{or} \quad x - \hat{x}_k^+$$

$$\varepsilon_{k-1}^- = x_{k-1} - \hat{x}_{k-1}^-$$

$$\varepsilon_{k-1}^+ = x_{k-1} - \hat{x}_{k-1}^+$$

A priori estimate :

$$\hat{x}_k^- = E[\, x_k \mid y_1, y_2 \,.\,.\,.\,.\,.\,.\,.\, y_{k-1}]$$

= Estimate of x_k before we process the measurement y_k at time k.

A posteriori estimate

$$\hat{x}_k^+ = E[\, x_k \mid y_1, y_2 \,.\,.\,.\,.\,.\, y_{k-1}, y_k]$$

= Estimate of x_k after we process the measurement y_k at time k.

where $\varepsilon_{k-1}^{+} = x_{k-1} - \hat{x}_{k-1}^{+}$

A priori error covariance :- P_k^-

$$P_k^- = E\left[\varepsilon_k^- \cdot \varepsilon_k^{-T}\right]$$

$$= E\left[\left(A_{k-1}\varepsilon_{k-1}^{+} + w_{k-1}\right) \cdot \left(A_{k-1}\varepsilon_{k-1}^{+} + w_{k-1}\right)^T\right]$$

$$= E\left[\left(A_{k-1}\varepsilon_{k-1}^{+} + w_{k-1}\right) \cdot \left(\varepsilon_{k-1}^{+T} A_{k-1}^T + w_{k-1}^T\right)\right]$$

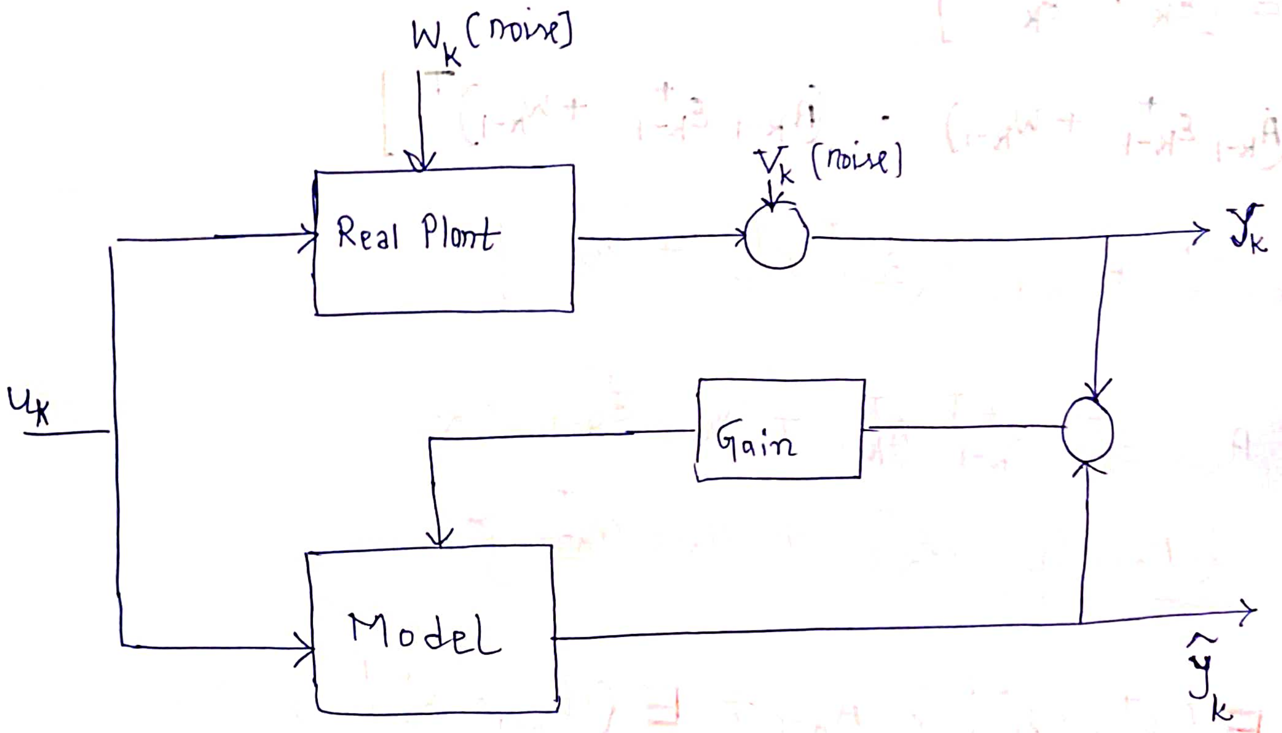
$$= E\left\{ A_{k-1}\varepsilon_{k-1}^{+}\varepsilon_{k-1}^{+T} A_{k-1}^T + w_{k-1}\varepsilon_{k-1}^{+T} A_{k-1}^T + A_{k-1}\varepsilon_{k-1}^{+} w_{k-1}^T + w_{k-1} w_{k-1}^T \right\}$$

$$= A_{k-1} E\left\{\varepsilon_{k-1}^{+} \cdot \varepsilon_{k-1}^{+T}\right\} A_{k-1} + E\left\{w_{k-1} \cdot \varepsilon_{k-1}^{+T}\right\} A_{k-1} + A_{k-1} E\left\{\varepsilon_{k-1}^{+} w_{k-1}^T\right\} + E\left\{w_{k-1} \cdot w_{k-1}^T\right\}$$

$$= A_{k-1} E\left\{\varepsilon_{k-1}^{+}\varepsilon_{k-1}^{+T}\right\} A_{k-1} + E\left\{w_{k-1} \cdot w_{k-1}^T\right\}$$

$$= A_{k-1} P_{k-1}^{+} A_{k-1} + Q_{k-1}$$

$$\therefore \boxed{P_k^- = A_{k-1} P_{k-1}^{+} A_{k-1} + Q_{k-1}}$$



Kalman filter

u_k → Input vector

W_k → Process Noise

V_k → Measurement noise

y_k → Output Vector [Observed measurement] or Sensor output
(measurement at time step k.

$\hat{y}_k$ = Predicted output vector

x_k = Parameter vector that we want to estimate
(unknown)

A posteriori estimation :

$$\hat{x}_k^+ = \hat{x}_k^- + K_k \left(y_k - C_k \hat{x}_k^- \right)$$

x_k is the parameter vector that we want to estimate

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

$$= x - \hat{x}_k^+$$

(General form)

$$= x - \left[\hat{x}_k^- + K_k \left(y_k - C_k \hat{x}_k^- \right) \right]$$

$$= x - \hat{x}_k^- - K_k \left(y_k - C_k \hat{x}_k^- \right)$$

$$= (x - \hat{x}_k^-) - K_k \left(C_k x + V_k - C_k \hat{x}_k^- \right)$$

$$= (x - \hat{x}_k^-) - K_k C_k x - K_k V_k + K_k C_k \hat{x}_k^-$$

$$= (x - \hat{x}_k^-) - K_k C_k (x - \hat{x}_k^-) - K_k V_k$$

$$= (I - K_k C_k)(x - \hat{x}_k^-) - K_k V_k$$

$$\varepsilon_k^+ = (I - K_k C_k)\, \varepsilon_k^- - K_k V_k$$

A posteriori covariance of error : P_K^+

$$P_K^+ = E\left[\varepsilon_K^+ \cdot \varepsilon_K^{+T}\right]$$

$$= E\left[\left[(I - K_K C_K)\varepsilon_K^- - K_K V_K\right] \cdot \left[(I - K_K C_K)\varepsilon_K^- - K_K V_K\right]^T\right]$$

$$= E\left[\left[(I - K_K C_K)\varepsilon_K^- - K_K V_K\right] \cdot \left[\varepsilon_K^{-T}(I - K_K C_K)^T - V_K^T K_K^T\right]\right]$$

$$= E\left[(I - K_K C_K)\varepsilon_K^- \cdot \varepsilon_K^{-T}(I - K_K C_K)^T - K_K V_K \varepsilon_K^{-T}(I - K_K C_K)^T - (I - K_K C_K)\varepsilon_K^- V_K^T K_K^T + K_K V_K \cdot V_K^T K_K^T\right]$$

$$= (I - K_K C_K)\, E\left[\varepsilon_K^- \varepsilon_K^{-T}\right](I - K_K C_K)^T - K_K\, \underbrace{E\left[V_K \varepsilon_K^{-T}\right]}_{0}(I - K_K C_K)^T - (I - K_K C_K)\, \underbrace{E\left[\varepsilon_K^- V_K^T\right]}_{0} K_K^T + K_K\, E\left[V_K \cdot V_K^T\right] K_K^T$$

Note

$\because$ $E\left[\varepsilon_K^- \varepsilon_K^{-T}\right] = P_K^-$

$E\left[V_K \varepsilon_K^{-T}\right] = E[V_K] \cdot E\left[\varepsilon_K^{-T}\right] = 0$

$E\left[\varepsilon_K^- V_K^T\right] = E\left[\varepsilon_K^-\right] \cdot E\left[V_K^T\right] = 0$

$E\left[V_K V_K^T\right] = R_K$

$$P_K^+ = (I - K_K C_K) P_K^- (I - K_K C_K)^T + K_K R_K K_K^T$$

$$= (P_K^- - K_K C_K P_K^-)(I - C_K^T K_K^T) + K_K R_K K_K^T$$

$$= P_K^- - K_K C_K P_K^- - P_K^- C_K^T K_K^T + K_K C_K P_K^- C_K^T K_K^T + K_K R_K K_K^T \quad (1)$$

Cost function W_k :

$$W_k = \text{Trace}(P_k^+)$$

$\therefore W_k = tr(P_k^-) - tr(K_k C_k P_k^-) - tr(P_k^- C_k^T K_k^T) + tr(K_k C_k P_k^- C_k^T K_k^T) + tr(K_k R_k K_k^T)$ — (2)

We want to minimize the cost W_k.

$$\left.\begin{aligned} &\therefore \frac{\partial W_k}{\partial K_k} = 0 \\ &\frac{\partial(\text{trace } P_k^+)}{\partial K_k} = 0 \end{aligned}\right\} \text{— (3)}$$

<u>Note</u>: $\frac{\partial}{\partial K_k} tr(P_k^-) = 0$

$$\frac{\partial}{\partial K_k} tr(K_k C_k P_k^-) = (C_k P_k^-)^T = (P_k^-)^T C_k^T = P_k^- C_k^T$$

$$\frac{\partial}{\partial K_k} tr(P_k^- C_k^T K_k^T) = P_k^- C_k^T$$

$$\frac{\partial}{\partial K_k} tr(K_k C_k P_k^- C_k^T K_k^T) = 2 K_k C_k P_k^- C_k^T$$

$$\frac{\partial}{\partial K_k} tr(K_k R_k K_k^T) = 2 K_k R_k$$

From equation (2) and (3): $\frac{\partial [W_k]}{\partial K_k} = 0$

$$0 - P_k^- C_k^T - P_k^- C_k^T + 2 K_k C_k P_k^- C_k^T + 2 K_k R_k = 0$$

$$K_k [R_k + C_k P_k^- C_k^T] = P_k^- C_k^T$$

$$\therefore \boxed{K_k = P_k^- C_k^T [R_k + C_k P_k^- C_k^T]^{-1}}$$

$\because$ $R_k + C_k P_k^- C_k^T$ is symmetric in nature.

$$\therefore \left\{ \left(R_k + C_k P_k^- C_k^T \right)^{-1} \right\}^T = \left(R_k + C_k P_k^- C_k^T \right)^{-1}$$

Now put the value of K_k in equation ①

$$
\begin{aligned}
P_k^+ = {} & P_k^- - P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1} C_k P_k^- \\
& - P_k^- C_k^T \{P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1}\}^T \\
& + \{P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1}\} \cdot C_k P_k^- C_k^T \cdot \{P_k^- C_k^T (R_k + C_k P_k^{-1} C_k^T)^{-1}\}^T \\
& + \{P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1}\} \cdot R_k \cdot \{P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1}\}^T
\end{aligned}
$$

$$
\begin{aligned}
= {} & P_k^- - P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1} C_k P_k^- \\
& - P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1} \cdot C_k P_k^- \\
& + P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1} \cdot C_k P_k^- C_k^T \cdot (R_k + C_k P_k^- C_k^T)^{-1} C_k P_k^- \\
& + P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1} \cdot R_k \cdot (R_k + C_k P_k^- C_k^T)^{-1} C_k \cdot P_k^-
\end{aligned}
$$

$$
\begin{aligned}
= {} & P_k^- - 2 P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1} C_k P_k^- \\
& + P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1} \cdot \cancel{[C_k P_k^- C_k^T + R_k] \cdot (R_k + C_k P_k^- C_k^T)^{-1}} C_k P_k^-
\end{aligned}
$$

$$
\begin{aligned}
= {} & P_k^- - 2 P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1} C_k P_k^- \\
& + P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1} C_k P_k^-
\end{aligned}
$$

$$= P_k^- - P_k^- C_k^T (R_k + C_k P_k^- C_k^T)^{-1} C_k P_k^-$$

$$= P_k^- - K_k \cdot C_k P_k^-$$

$$\boxed{P_k^+ = (I - K_k C_k) P_k^-}$$

Summary of the Kalman filter

$$X_k = A_{k-1} X_{k-1} + B_{k-1} u_{k-1} + W_{k-1}$$

$$y_k = C_k X_k + V_k$$

Assume initial estimate of state $= \hat{x}_0^+$

Assume initial covariance matrix of estimation error $= P_0^+$

For $k = 1, 2, 3, 4 \ldots\ldots$ we perform the following steps :

STEP-I : A priori (prediction) step :

A priori state estimate $(\hat{x}_k^-)$:

$$\hat{x}_k^- = A_{k-1} \hat{x}_{k-1}^+ + B_{k-1} u_{k-1}$$

A priori error covariance (P_k^-) :

$$P_k^- = A_{k-1} P_{k-1}^+ A_{k-1}^T + Q_{k-1}$$

STEP-II : A posteriori (update) step :

Kalman gain (K_k) :

$$K_k = P_k^- C_k^T \left(R_k + C_k P_k^- C_k^T \right)^{-1}$$

A posteriori state estimate $(\hat{x}_k^+)$:

$$\hat{x}_k^+ = \hat{x}_k^- + K_k \left(y_k - C_k \hat{x}_k^- \right)$$

A posteriori error covariance (P_k^+) :

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

or $$P_k^+ = (I - K_k C_k) P_k^-$$

Kalman filter reference : OCW. mit. edu ⟶ click on $\boxed{\text{Explore}}$ button.

1. In the search bar → " Identification, estimation and learning". click on this subject

You will find the selected subject (2.160 is the subject code.

It is instructed by Prof. Harry Asada.

On the webpge, on the left hand side; you can select the lecture notes by click on Lecture notes, lecture videos (If provided by MIT for access only then you can see) etc.

$\boxed{\text{or}}$

darbelofflab. mit. edu

[डी • अर्बेलोफ]

↓

at the top of this web page click on $\boxed{\text{TEACHING}}$

→ Lecture
→ slide
→ notes

By clicking on Lecture you will be directed to new website which is mit. hosted. panopto. com

Content related to this subject (Identification, estimation and learning) is available for public. Video lectures for other subjects are not for public.

MEAN SQUARED ERROR (MSE)

$\mathbf{E}(\text{Error}^2)$ or the expectation of the squared error. It is a common metric in statistical estimation control theory and machine learning.

- If you have an error term Error, then $E(\text{Error}^2)$ is the expected value (or mean) of the square of this error. It quantifies the average magnitude of the error, giving more weight to large errors due to the squaring.

In mathematical terms:

$X \rightarrow$ True value

$\hat{X} \rightarrow$ Estimate or mean value

$$\text{Error} = X - \hat{X}$$

$$\mathbf{E}(\text{Error}^2) = \int (\text{Error})^2 \, p(\text{error}) \, d(\text{error})$$

$p(\text{error})$: Probability density function of the error.

Note If the error has a mean of zero (i.e. $(\mathbf{E}(\text{Error}) = 0)$ the expectation $E(\text{Error}^2)$ simplifies to the variance of the error.

Note

Let $N = 10$

we have ten error values.

$$E(\text{Error}^2) \approx \frac{1}{N} \sum_{i=1}^{N} (\text{Error}_i)^2$$

$$E(\text{Error}^2) = \frac{1}{10}\left[(\text{Error}_1)^2 + (\text{error}_2)^2 + (\text{error}_3)^2 + \cdots (\text{error}_{10})^2\right]$$

Expected Value

- Expectedvalue, expectation, expectancy, expectation operator mean, mathematical expectation or first moment; all the same.

- It is a generalization of the weighted average.

- Expected value of a random variable X is often denoted as $E(X)$, $E[X]$ or EX with E also often styled as $\mathbb{E}$ or E.

$$E[X] = x_1 p_1 + x_2 p_2 + \dots + x_k p_k$$

$x_1\ x_2 \dots x_k \Rightarrow$ Possible outcomes

$p_1\ p_2 \dots p_k \Rightarrow$ Probability of occurs.

Expectation or Expected value

Random variable X

Discrete Case:

Expected value of discrete random variable X is found by multiplying each X-value by its probability and then summing over all values of the random variable. that is, if X is discrete

$$E(X) = \sum_{\text{all } X} x \cdot p(x) = \mu_x$$

Continuous case:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx = \mu_x$$

Variance of X :

X → Random variable

μ → Mean value of X

$$V(X) = E[(X-\mu)^2]$$

$$= E[(X-\mu)(X-\mu)]$$

$$= E[X^2 - 2\mu X + \mu^2]$$

$$= E(X^2) - 2\mu\, E(X) + \mu^2$$

$$= E(X^2) - 2\mu \cdot \mu + \mu^2$$

$$= E(X^2) - 2\mu^2 + \mu^2$$

$$V(X) = E(X^2) - \mu^2$$

$$\boxed{\sigma_x^2 = E(X^2) - \mu^2}$$

If $\mu = 0$

$$\sigma_x^2 = \text{Variance} = E(X^2)$$

फैलाव (spread)

- It is a measure of dispersion. It is a measure of how far a set of numbers is spread out from their average value. It is often represented by σ^2 or s^2.

$E(x^2)$: Expectation of (x^2)

For a discrete random variables :

If x is a discrete random variable with possible values x_1, x_2 ... ... x_n and corresponding probabilities p_1, p_2 ...... p_n, then the expectetaion of x^2 is calculated as:

$$E[x^2] = \sum_{i=1}^{n} x_i^2 \cdot p_i$$

For Continuous random variable :

$$E[x^2] = \int_{-\infty}^{\infty} x^2 \underbrace{p(x)}_{\text{Probability desity function}} dx$$

Covariance (X, Y)

$$\text{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

$$= E[XY - \mu_Y X - \mu_X Y - \mu_X \mu_Y]$$

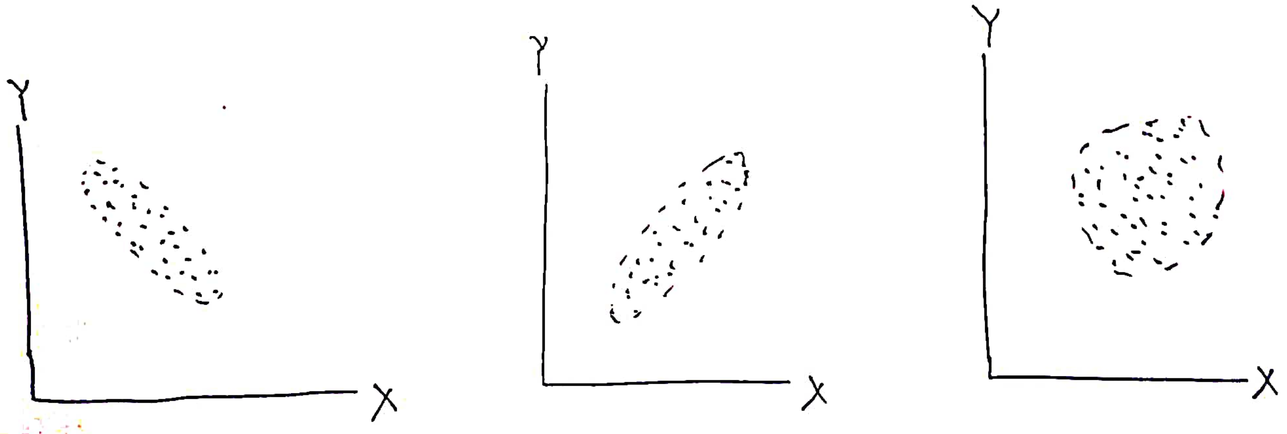
$$= E(XY) - \mu_Y E(X) - \mu_X E(Y) - E(\mu_X \mu_Y)$$

Constant term

$$= E(XY) - \mu_Y \cdot \mu_X - \mu_X \mu_Y + \mu_X \mu_Y$$

$$= E(XY) - \mu_X \mu_Y \cancel{- \mu_X \mu_Y} \cancel{+ \mu_X \mu_Y}$$

$$= E(XY) - \mu_X \mu_Y$$



$\text{COV}(X,Y) < 0$ $\qquad$ $\text{COV}(X,Y) > 0$ $\qquad$ $\text{COV}(X,Y) \approx 0$

• Covariance only measures whether (कि क्या) the two data points are directly or inversely related to each other. <u>It does not provide the strength of relationship</u> between two data points.

Correlation (X, Y) : It quantifies the strength of strength of relationship between two variables.

$$\rho_{X,Y} = \text{Correlation}(X, Y)$$

$$= \frac{\text{Cov}(X, Y)}{\sigma_X \; \sigma_Y}$$

$$= \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \; \sigma_Y}$$

where Cov is the covariance

σ_X is the standard deviation of X

σ_Y is the standard deviation of Y

$$\mu_X = E[X]$$

$$\mu_Y = E[Y]$$

$$\sigma_X^2 = E(X-\mu_X)^2 = E[X^2] - \mu_X^2$$

$$\sigma_Y^2 = E(Y-\mu_Y)^2 = E[Y^2] - \mu_Y^2$$

$$E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X \cdot \mu_Y$$

$$\rho_{XY} = \frac{E[XY] - \mu_X \cdot \mu_Y}{\sqrt{E[X^2] - \mu_X^2} \cdot \sqrt{E[Y^2] - \mu_Y^2}}$$

$\boxed{MSE = \text{Variance} + \text{Bias}^2}$

Proof :-

True value $= x$

Estimated value $= \hat{x}$

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

$$\text{Bias} = \mathbb{E}[\hat{x}] - x$$

$$\text{Variance} = \mathbb{E}\left[(\hat{x} - \mathbb{E}[\hat{x}])^2\right]$$

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

$$\text{Expand } (\hat{x}-x)^2 = \big(\underbrace{\hat{x} - \mathbb{E}[\hat{x}]}_{\text{Deviates from its expected value}} + \underbrace{\mathbb{E}[\hat{x}] - x}_{\text{Account for bias}}\big)^2$$

Expanding the square :

$$(\hat{x}-x)^2 = (\hat{x} - \mathbb{E}[\hat{x}])^2 + 2(\hat{x} - \mathbb{E}[\hat{x}])\cdot(\mathbb{E}[\hat{x}] - x) + (\mathbb{E}[\hat{x}] - x)^2$$

Step I : Taking expectation on both sides :

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

$$+ 2\,\mathbb{E}\left[(\hat{x} - \mathbb{E}[\hat{x}])\cdot(\mathbb{E}[\hat{x}] - x)\right]$$

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

Expectation of $[\hat{x} - \mathbb{E}[\hat{x}]] = 0$

This is constant term does not depend on $\hat{x}$

$\because \quad E[\hat{x} - E[\hat{x}]] = 0$ **STEP II:**

$$\therefore E[(\hat{x}-x)^2] = E[(\hat{x} - E[\hat{x}])^2] + 0 + E[(\underbrace{E[\hat{x}] - x}_{\text{It is constant term}})^2]$$

$\because$ expectation of constant = constant

$$E[(\hat{x}-x)^2] = E[(\hat{x} - E[\hat{x}])^2] + (E[\hat{x}] - x)^2$$

$$\boxed{MSE = \text{Variance} + \text{Bias}^2}$$

Note I: more explanation of middle term in step I.

$$E[(\hat{x} - E[\hat{x}]) \cdot (E[\hat{x}] - x)]$$

This is constant so it will come outside from the expectation.

$$= (E[\hat{x}] - x) * E[(\hat{x} - E[\hat{x}])]$$

$$= (E[\hat{x}] - x) * (E[\hat{x}] - E[\hat{x}])$$

$$= (E[\hat{x}] - x) * 0$$

$$= 0$$

Note $\quad E(E[\hat{x}] - x)^2 = E[(\text{Constant})^2]$

$$= (\text{Constant})^2$$

Note:

Unbiased Estimator

Case: when $E[\text{Error}] = 0$

when the expected error is zero, it means the estimator is unbiased and thus the bias is zero.

$$\text{Bias} = E[\hat{x}] - x = 0$$

In other words $E[\hat{x}] = x$

$\because$ $MSE = \text{Bias}^2 + \text{Variance}$

$$MSE = 0^2 + \text{Variance}$$

$$\boxed{MSE = \text{Variance}}$$

Note

estimation of error

$E(\text{error}) = E(\text{Expected value} - \text{True value})$

$= E(\hat{x} - x)$

$= E(\hat{x}) - E(x)$

$= x - x$

$= 0$

for unbiased estimator

$E[\hat{x}] = x$

$E[x]$ is a constant. that is equal to x.

$\mathbb{E}[\text{error}] = \text{Bias}$

Proof :- True value $= x$

Estimated value $= \hat{x}$

Error $= \hat{x} - x$

Expected error or mean error $\Rightarrow$

$$\mathbb{E}[\text{Error}] = \mathbb{E}[\hat{x} - x]$$

$$= \mathbb{E}[\hat{x}] - x$$

<u>Bias</u> : $\because$ $\text{Bias} = \mathbb{E}[\hat{x}] - x$ (By definition of Bias)

$$\therefore \mathbb{E}[\text{Error}] = \mathbb{E}[\hat{x} - x] = \mathbb{E}[\hat{x}] - x = \text{Bias}$$

$$\boxed{\text{Variance} = \text{MSE} - \text{Bias}^2}$$

Proof:

$$\text{Variance} = E\left[\left(\hat{x} - E[\hat{x}]\right)^2\right]$$

$$\hat{x} - E[\hat{x}] = \hat{x} - x + x - E[\hat{x}]$$

$$= \{(\hat{x} - x)\} - \{E[\hat{x}] - x\}$$

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

$$\because (a-b)^2 = a^2 - 2ab + b^2$$

$$= (\hat{x} - x)^2 - 2(\hat{x} - x)\cdot\{E[\hat{x}] - x\} + \{E[\hat{x}] - x\}^2$$

Taking expectation on both sides of above equation :-

$$\therefore E\{\hat{x} - E[\hat{x}]\} = E[(\hat{x} - x)^2]$$

$$-2E\left[(\hat{x} - x)\cdot\{\underbrace{E[\hat{x}] - x}_{\text{Constant term}}\}\right]$$

$$+ E\{\underbrace{E[\hat{x}] - x}_{\text{constant term}}\}^2$$

$$= \text{MSE} - 2\{E[\hat{x}] - x\}\cdot\{E[\hat{x}] - \underbrace{E[x]}_{=x}\}$$

$$+ \{E[\hat{x}] - x\}^2$$

$\because$ True value $= x$

$$\text{Variance} = \text{MSE} - 2\{E[\hat{x}] - x\}\{E[\hat{x}] - x\}$$

$$+ \{E[\hat{x}] - x\}^2$$

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

$$\boxed{\text{Variance} = MSE - Bias^2}$$

Note Variance $= E[(\hat{x} - E[\hat{x}])^2]$

$$= E\{[\hat{x}]^2 - 2\hat{x}\, E[\hat{x}] + (E[\hat{x}])^2\}$$

$$= E\{[\hat{x}]^2\} - 2E[\hat{x}] \cdot E[\hat{x}] + (E[\hat{x}])^2\}$$

$$= E\{[\hat{x}]^2\} - 2(E(\hat{x}))^2 + (E[\hat{x}])^2$$

$$\text{Variance} = E\{[\hat{x}]^2\} - (E(\hat{x}))^2$$

$$\boxed{\text{Variance} = E\{[\hat{x}]^2\} - \mu^2}$$ where $\mu = E(\hat{x})$ (2)

We also have

$$\boxed{\text{Variance} = MSE - Bias^2}$$ (3)

In equation (2) and (3) L.H.S is variance. therm on the right hand side looks similar in fashion. But variables are totally different.

$$\because \mu = E[\hat{x}]$$

$$\& \quad Bias = E[\hat{x}] - x$$

$$\& \quad MSE = E[(\hat{x} - x)^2]$$