Kalman Filter

$$y_k = a_k x_k + n_k$$

y_k: observed signal; a_k: gain term; x_k: info bearing signal; n_k: additive noise

Deriving MSE as the loss fn:

$$\underbrace{f}_{\text{objective function}}(\underbrace{e_k}_{\text{error}}) = f(x_k - \underbrace{\hat{x}_k}_{\text{estimate}})$$

let $f(\cdot)$ be the squared error function then, $f(e_k) = e_k^2 = (x_k - \hat{x}_k)^2$

This was just for 1 instance. we extend this over a time period by taking the mean

$$\text{loss fn} = \underbrace{E}_{\text{Mean squared error}}(f(e_k)) = E(e_k^2) = E(t)$$

Deriving MSE from MLE:

Goal is to find $\hat{x}$ which maximizes the probability of y. That is;

$$\max[P(y|\hat{x})]$$

Assume $n_k = \mathcal{N}(0, \sigma_k^2)$ this implies that given x_k, the observed value y_k is also Gaussian as

$$y_k|\hat{x}_k \sim \mathcal{N}(a_k x_k, \sigma_k^2)$$

$$P(y_k|\hat{x}_k) = \frac{1}{2\pi\sigma^2} e^{\left(-\frac{(y_k - \overset{\mu}{a_k x_k})^2}{2\sigma^2}\right)}$$

For the conditional distribution $P(y_k|x_k)$, you are simply using the conditional mean and conditional variance of y_k given x_k to construct the Gaussian formula.

Proof that $\mu = E[y_k|x_k] = a_k x_k$:

$$E[y_k|x_k] = E[a_k x_k + n_k|x_k] = \underbrace{E[a_k x_k|x_k]}_{a_k x_k \text{ as it is a constant due to } x_k \text{ being given/known}} + \underbrace{E[n_k|x_k]}_{\text{this equal } E[n_k] \text{ as } n_k \text{ and } x_k \text{ are independent random variables and } E[n_k]=0 \text{ as } n_k \sim \mathcal{N}(0,\sigma^2)} = a_k x_k$$

Proof that $var(y_k|x_k) = \sigma^2$: used, x_k and y_k are independent

$$var(y_k|x_k) = var(a_k x_k + n_k|x_k) = \underbrace{var(a_k x_k|x_k)}_{0} + \underbrace{var(n_k|x_k)}_{\sigma^2} = \sigma^2$$

0 as deterministic variables have zero variance. So for a given x_k, $a_k x_k$ is known and there is no randomness.

σ^2 as n_k and x_k are indep and $var(n_k|x_k) = var(n_k) = \sigma^2$ as $n_k \sim \mathcal{N}(0, \sigma^2)$

②

then $P(y|\hat{x}) = P(y_1, y_2, \ldots | x_1, x_2, \ldots) = P(y_1|x_1) \cdot P(y_2|x_2) \ldots$

$= \prod_k \frac{1}{2\pi\sigma^2} \cdot e^{\left(-\frac{(y_k - a_k x_k)^2}{2\sigma^2}\right)}$ then take log of both sides;

$\log P(y|\hat{x}) = \log(P(y_1|x_1) \cdot P(y_2|x_2) \ldots) = \log(P(y_1|x_1)) + \log(P(y_2|x_2)) + \ldots$

$= \sum_k \log\left[\frac{1}{2\pi\sigma^2} \cdot e^{\left(-\frac{(y_k - a_k x_k)^2}{2\sigma^2}\right)}\right] = \sum_k \left[\log\left(\frac{1}{2\pi\sigma^2}\right) + \log\left(e^{\left(\frac{-(y_k - a_k x_k)^2}{2\sigma^2}\right)}\right)\right]$

$= \Bigg[\underbrace{\sum_k \log\left(\frac{1}{2\pi\sigma^2}\right)}_{\text{some constant value}}\Bigg] + \left[\sum_k \frac{-(y_k - a_k x_k)^2}{2\sigma^2}\right]$

as maximizing $\log(P(\cdot))$ also maximizes $P(\cdot)$ too.

So the $\hat{x}$ which maximizes $\max[P(y|\hat{x})] \cong \max[\log(P(y|\hat{x}))]$

hence we want to find $\hat{x}$ which maximizes $\log(P(y|\hat{x}))$;

$\log(P(y|\hat{x})) = \left(\text{some constant value}\right) + \sum_k \frac{-(y_k - a_k x_k)^2}{2\sigma^2}$ so maximize;

$\max\left(\sum_k \frac{-(y_k - a_k x_k)^2}{2\sigma^2}\right) = \max\left(\frac{-1}{2} \sum_k \frac{(y_k - a_k x_k)^2}{\sigma^2}\right)$, the $\hat{x}$ that maximizes this function minimizes $(y_k - a_k x_k)^2$ which is the squared error function and $\sum_k (y_k - a_k x_k)^2$ is the sum of squared error fn.

as MSE is simply the sum of squared errors divided by the number of observations K.

Hence MSE is found to be applicable when the expected variation of y_k is best modelled as a Gaussian distribution. In such a case the MSE serves to provide the value of $\hat{x}_k$ which maximizes the likelihood of the signal y_k.

In short we found that for $y_k = a_k x_k + n_k$ with $n_k \sim \mathcal{N}(0, \sigma_k^2)$, the $\hat{x}$ that maximizes $P(y|\hat{x})$ can be found by finding the $\hat{x}$ that minimizes the MSE.

State Space Derivation:

MSE function $E[e_k^2] = E[(\underset{\text{real}}{x_k} - \underset{\text{predicted}}{\hat{x}_k})^2]$ is equivalent to P_k where

$$E[(x_k-\hat{x}_k)^2] = E[e_k e_k^T] = E[\underbrace{(x_k-\hat{x}_k)(x_k-\hat{x}_k)^T}] = P_k$$

$\therefore e_k = (x_k - \hat{x}_k)$ "error"

" error covariance matrix at time K $(n \times n)$

Proof that P_k is the error covariance matrix:

Remember that $cov(X,Y) = E[(X-E[X])(Y-E[Y])] = \frac{1}{N}\sum_{i=1}^{N}(X_i-\mu_x)(Y_i-\mu_y)$

Then the covariance of the error would be;

$$cov(e_a, e_b) = E[(e_a - E[e_a])(e_b - E[e_b])] = E[((x_a-\hat{x}_a) - \overbrace{E[x_a-\hat{x}_a]}^{0})((x_b-\hat{x}_b) - \overbrace{E[x_b-\hat{x}_b]}^{0})]$$

Since $\hat{x}_k$ was constructed by design so that $E[x_k - \hat{x}_k] = 0$, in other words $\hat{x}_k$ is an unbiased estimator we'll have $E[x_a - \hat{x}_a]$ and $E[x_b - \hat{x}_b] = 0$ in our covariance calculations hence;

$$cov(e_a, e_b) = E[\underbrace{(x_a-\hat{x}_a)}_{e_a}\underbrace{(x_b-\hat{x}_b)}_{e_b}] = E[e_a \cdot e_b],$$

Also covariance matrices have the form $\Rightarrow \begin{bmatrix} var(x_1) & \cdots & cov(x_n,x_1) \\ \vdots & \ddots & \vdots \\ cov(x_n,x_1) & \cdots & var(x_n) \end{bmatrix}$ where diagonal entries are variances and the off-diagonal entries are covariences.

Our error covariance matrix $P_k = E[\underbrace{(x_k-\hat{x}_k)}_{e_k}\underbrace{(x_k-\hat{x}_k)^T}_{e_k}]$

$$= E\left(\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} [e_1\ e_2 \cdots e_n]\right) = E\left(\begin{bmatrix} e_1^2 & \cdots & e_1e_n \\ \vdots & \ddots & \vdots \\ e_ne_1 & \cdots & e_n^2 \end{bmatrix}\right) = \begin{bmatrix} var(e_1) & \cdots & cov(e_1,e_n) \\ cov(e_n,e_1) & \cdots & var(e_n) \end{bmatrix}$$

By definition an error covariance matrix is defined as a matrix that represents the covariance btw estimated errors in a system, with its elements indicating the variances and covariences of these errors. In other words, the error covariance matrix (P_k) is a measure of how much uncertainty exists in the state estimation at time K. More formally, P_k quantifies the variance of the estimation error, which represents the difference between the true state x_k and the estimated state $\hat{x}_k$.

↳ Proof that minimizing the MSE results in the condition $E[x_k - \hat{x}_k] = 0$:

$$MSE = E[(x_k-\hat{x}_k)^2] = E[(x_k \overbrace{- E[x_k] + E[x_k]}^{0} - \hat{x}_k)^2] = E[((x_k - E[x_k]) + (E[x_k]-\hat{x}_k))^2]$$

used! $E[a+b]=E[a]+E[b]$

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

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

∴ used $E[ax] = aE[x]$

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

∴ used $E[x_k]$ is constant

$$= 2\left(E[x_k E[x_k]] - E[x_k\hat{x}_k] - E[E[x_k]^2] + E[E[x_k]\hat{x}_k]\right.$$

$$= 2\left(\underbrace{(E[x_k]E[x_k])}_{\cancel{E[x_k]^2}} - \underbrace{\cancel{E[x_k\hat{x}_k]}}_{\cancel{E[x_k]\cdot E[\hat{x}_k]}} - \cancel{E[x_k]^2} + \cancel{E[x_k]E[\hat{x}_k]}\right)$$

as x_k and $\hat{x}_k$ are independent, we used $E[XY] = E[X]E[Y]$ if X and Y are independent random variables.

$$= 2(0) = 0$$ then;

$$= E[(x_k - E[x_k])^2] + \cancel{2E[\ldots]}^{0} + (E[x_k]-\hat{x}_k)^2$$

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

Then when minimizing MSE wrt $\hat{x}_k$ we minimize $E[(x_k-E[x_k])^2] + (E[x_k]-\hat{x}_k)^2$

The first term is a constant wrt $\hat{x}_k$, so the second term is what we need to minimize, we can see that $(E[x_k]-\hat{x}_k)^2$ is minimized when $\underline{\hat{x}_k = E[x_k]}$ by inspection or by taking its derivative wrt $\hat{x}_k$ and equating it to zero.

Hence we found $\hat{x}_k = E[x_k]$ which means

$E[x_k - \hat{x}_k]$ can be written as $E[x_k - \hat{x}_k] = E(x_k - E[x_k])$

$$= E[x_k] - \underbrace{E[E[x_k]]}_{E[x_k]} = E[x_k] - E[x_k] = 0$$

∴ we used the law of iterated expectations $E[E[x]] = E[x]$.

∴ For the sake of completeness, here is its proof:

$$\frac{d\left((E[x_k]-\hat{x}_k)^2\right)}{d\hat{x}_k} = 2(E[x_k]-\hat{x}_k)\frac{d(E[x_k]-\hat{x}_k)}{d\hat{x}_k} = 2(E[x_k]-\hat{x}_k)(-1)$$

$$= -2(E[x_k]-\hat{x}_k) = 0 \Rightarrow \hat{x}_k = E[x_k]$$

Now that we've established what P_k is we move on;

Assume the following:

$$x_{k+1} = \Phi x_k + w_k$$

State transition matrix of the process from the state at k to the state at $k+1$, $(n \times n)$

state vector at time k $(n \times 1)$

associated white noise process with known covariance. $(n \times 1)$

How x's are really updated/changed

∴ $\hat{x}'_k$ = prior prediction before the update from the measurement at time k

$\hat{x}_k$ = updated state which is obtained after updating $\hat{x}'_k$ using z_k measurement

$$z_k = H x_k + v_k$$

actual measurement of x at time k. $(m \times 1)$

noiseless connection btw the state vector and the measurement vector. $(m \times n)$

associated measurement error $(m \times 1)$

How the observed measurements relate to the real underlying x's at every time step.

∴ w_k and v_k are both assumed to be a white noise process with known covariance and has zero cross-correlation with each other.

$$\boxed{\text{Update Equation} \quad \hat{x}_k = \hat{x}'_k + K_k(z_k - H\hat{x}'_k)}$$ *

prior estimate of $\hat{x}_k$, and was obtained by knowledge of the system

"Kalman gain": it maps measurement residual to an update on the state estimate

"innovation" or "measurement residual"

∴ Note that $H\hat{x}'_k$ is like predicting the observation that would have corresponded to the state vector $\hat{x}'_k$, and $z_k - H\hat{x}'_k$ is like the difference btw the actual observation from the actual state x_k and the predicted (internal) state($\hat{x}'_k$)'s corresponding measurement prediction.

Then; substitute $z_k = Hx_k + v_k$ into $\hat{x}_k = \hat{x}'_k + K_k(z_k - H\hat{x}'_k)$

$$\Rightarrow \hat{x}_k = \hat{x}'_k + K_k(\overbrace{(Hx_k + v_k)}^{z_k} - H\hat{x}'_k)$$

Then substitute this $\hat{x}_k$ into

$P_k = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$ we get;

$$P_k = E[\underbrace{(x_k - (\hat{x}'_k + K_k(Hx_k + v_k - H\hat{x}'_k))}\,(x_k - \hat{x}'_k + K_k(Hx_k + v_k - H\hat{x}'_k))^T]$$

$$\hookrightarrow (x_k - \hat{x}'_k - K_k H x_k - K_k v_k + K_k H\hat{x}'_k) = (\underbrace{x_k - \hat{x}'_k}_{(I)(x_k - \hat{x}'_k)}\, \underbrace{- K_k H x_k + K_k H\hat{x}'_k}_{(-K_k H)(x_k - \hat{x}'_k)}) - K_k v_k$$

$$= (I - K_k H)(x_k - \hat{x}'_k) - K_k v_k \quad \text{then}$$

$$= E[((I - K_k H)(x_k - \hat{x}'_k) - K_k v_k)((I - K_k H)(x_k - \hat{x}'_k) - K_k v_k)^T]$$

multiply the 2 matrices

∴ use $E[x+y] = E[x] + E[y]$

$$= E[(I - K_k H)(x_k - \hat{x}'_k)(x_k - \hat{x}'_k)^T(I - K_k H)^T - (I - K_k H)(x_k - \hat{x}'_k)(K_k v_k)^T - (K_k v_k)(x_k - \hat{x}'_k)^T(I - K_k H)^T + (K_k v_k)(K_k v_k)^T$$

$$\hookrightarrow = E\left[(I-K_kH)(x_k-\hat{x}'_k)(x_k-\hat{x}'_k)^T(I-K_kH)^T\right] - E\left[(I-K_kH)(x_k-\hat{x}'_k)(K_k v_k)^T\right] - E\left[(K_k v_k)(x_k-\hat{x}'_k)^T (I-K_kH)^T\right] + E\left[(K_k v_k)(K_k v_k)^T\right]$$

so we used the fact that $E[AX] = AE[X]$

∴ Also note that we are taking Expectation w.r.t (x_k, v_k)

$$= (I-K_kH)\, E[(x_k-\hat{x}'_k)(x_k-\hat{x}'_k)^T]\,(I-K_kH)^T - \underbrace{(I-K_kH)\, E[(x_k-\hat{x}'_k) v_k^T](K_k)}_{0}$$

$$- \underbrace{(K_k)\, E[v_k (x_k-\hat{x}'_k)^T]\,(I-K_kH)^T}_{0} + (K_k)\, E[v_k v_k^T](K_k^T)$$

both the 2^{nd} and the 3^{rd} term equals zero as $E[(x_k-\hat{x}'_k)v_k^T]=0$. since; $(x_k-\hat{x}'_k)$ and v_k are ucorrelated which means $cov((x_k-\hat{x}'_k), v_k) = 0$, using the fact $cov(A,B) = E[AB]-E[A]E[B]$ yields $cov((x_k-\hat{x}'_k), v_k) = E[(x_k-\hat{x}'_k)(v_k)^T] - E[x_k-\hat{x}'_k]\,E[v_k] = 0 \Rightarrow E[(x_k-\hat{x}'_k)(v_k)^T] = E[x_k-\hat{x}'_k]E[v_k]$
As $E[x_k-\hat{x}'_k]=0$ due to $\hat{x}_k$ being an unbiased estimator as we've proven earlier, we get: $E[(x_k-\hat{x}'_k)v_k^T] = 0$.

$$= (I-K_kH)\underbrace{E[(x_k-\hat{x}'_k)(x_k-\hat{x}'_k)^T]}_{P'_k}(I-K_kH)^T + (K_k)\underbrace{E[v_k v_k^T]}_{R}(K_k^T)$$

$$= (I-K_kH)\, P'_k (I-K_kH)^T + K_k R K_k^T \quad \text{where} \quad R = E[v_k v_k^T]$$

Now recall that $P_k = E[\underbrace{(x_k-\hat{x}_k)}_{e_k}\underbrace{(x_k-\hat{x}_k)^T}_{e_k^T}] = \begin{bmatrix} E[e_1e_1^T] & \cdots & E[e_1e_n^T] \\ \vdots & \ddots & \\ E[e_ne_1^T] & \cdots & E[e_ne_n^T] \end{bmatrix}$ ∴ $E[e_ne_n^T] = E[(x_n-\hat{x}_n)(x_n-\hat{x}_n)^T]$ which is MSE

where the trace (sum of the diagonal elements) of the matrix is the sum of the mean squared errors. Therefore the MSE may be minimized by minimizing the trace of P_k.

Expand by multiplying the terms

$$= (P'_k - K_kHP'_k)(I-K_kH)^T + K_kRK_k^T = (P'_k - K_kHP'_k)(I - H^TK_k^T) + K_kRK_k^T$$

$$= P'_k - P'_kH^TK_k^T - K_kHP'_k + \underbrace{K_kHP'_kH^TK_k^T + K_kRK_k^T}_{\substack{K_k(HP'_kH^TK_k^T + RK_k^T) \\ (HP'_kH^T+R)(K_k^T)}}$$

$$= P'_k - P'_kH^TK_k^T - K_kHP'_k + K_k(HP'_kH^T+R)K_k^T$$ } we want to minimize its (P_k's) trace to minimize the MSE. So we focus on its trace

Let $T[A]$ be the trace of A.

$$T[P_k] = T[P'_k] - \underbrace{T[P'_kH^TK_k^T]}_{\text{this equals}} - T[K_kHP'_k] + T[K_k(HP'H^T+R)K_k^T]$$

this equals $T[(P'_kH^TK_k^T)^T] = T[K_kHP'_k]$ as trace does not change by the transpose operation.

$$= T[P'_k] - 2T[K_kHP'_k] + T[K_k(HP'H^T+R)K_k^T]$$

we want to find the optimal K_k which minimizes MSE so we differentiate $T[P_k]$ wrt K_k and equate it to zero and solve for the optimal K_k

$$\frac{d(T[P_k])}{d K_k} = \underbrace{\frac{d(T[P'_k])}{d K_k}}_{0} - 2\underbrace{\frac{d(T[K_k H P'_k])}{d K_k}}_{-2(HP'_k)^T} + \underbrace{\frac{d(T[K_k(HP'H^T+R)K_k^T])}{d K_k}}_{2K_k(HP'_kH^T+R)}$$

0 as P'_k does not depend on K_k since P'_k is from the previous time step which used the K_k from the previous iteration which is different from the K_k of the current time step which we are taking the derivative wrt.

its proof

its proof

let $A = K_k H P'_k$ then $A_{i,j} = \sum_k (K_k)_{i,k}(HP'_k)_{k,j} = \sum_k \left[(K_k)_{i,k}\left(\sum_l (H)_{k,l}(P'_k)_{l,j}\right)\right]$

then for the trace we have

$$A_{i,i} = \sum_k \left[(K_k)_{i,k}\left(\sum_l (H)_{k,l}(P'_k)_{l,i}\right)\right]$$

then $\dfrac{d(A_{i,i})}{d(K_k)_{a,b}} = \sum_k \left[\dfrac{d((K_k)_{i,k})}{d(K_k)_{a,b}} \cdot \left(\sum_l (H)_{k,l}(P'_k)_{l,i}\right)\right]$ this is nonzero only if $i=a$ and $k=b$ then;

if $a \neq i$ $\quad \dfrac{d(A_{i,i})}{d(K_k)_{a,b}} = 0$

if $a = i$ $\quad \dfrac{d(A_{i,i})}{d(K_k)_{i,b}} = \sum_l (H)_{b,l}(P'_k)_{l,i} = (HP'_k)_{b,i} = ((HP'_k)^T)_{i,b}$

derivative of trace would equal sum of the derivatives of the diagonal elements hence;

$$\left(\begin{array}{c}\text{derivative}\\ \text{of the}\\ \text{trace of } A\\ \text{wrt } K_k\end{array}\right)_{i,j} = \frac{d(A_{1,1})}{d(K_k)_{i,j}} + \dots + \frac{d(A_{n,n})}{d(K_k)_{i,j}} = \frac{d(A_{i,i})}{d(K_k)_{i,j}} = ((HP'_k)^T)_{i,j}$$

$$\left(\frac{d(T[A])}{d K_k}\right)_{i,j}$$

then $\dfrac{d(T[K_k H P'_k])}{d K_k} = \dfrac{d(T[A])}{d K_k} = (HP'_k)^T$ so

$$\frac{-2d(T[K_k H P'_k])}{d K_k} = -2(HP'_k)^T$$

let $A = K_k(HP'H^T+R)K_k^T$ then

$$A_{i,j} = \sum_k (K_k)_{i,k}\left((HP'H^T+R)K_k^T\right)_{k,j} = \sum_k (K_k)_{i,k}\left(\sum_\ell (HP'H^T+R)_{k,\ell}(K_k^T)_{\ell,j}\right)$$

$$= \sum_k (K_k)_{i,k}\left(\sum_\ell\left(\sum_m (H)_{k,m}(P'H^T)_{m,\ell} + (R)_{k,\ell}\right)(K_k^T)_{\ell,j}\right)$$

$$= \sum_k (K_k)_{i,k}\left(\sum_\ell\left(\sum_m\left((H)_{k,m}\left(\sum_n (P')_{m,n}(H^T)_{n,\ell}\right) + (R)_{k,\ell}\right)(K_k^T)_{\ell,j}\right)\right)$$

then

$$\frac{d(A_{i,j})}{d(K_k)_{a,b}} = \left[\sum_k \frac{d((K_k)_{i,k})}{d(K_k)_{a,b}}(\ldots)\right] + \left[\sum_k (K_k)_{i,k}\left(\sum_\ell\left(\sum_m ((H)_{k,m}(\ldots)+(R)_{k,\ell})\frac{d(K_k^T)_{\ell,j}}{d(K_k)_{a,b}}\right)\right)\right]$$

Not zero when $i=a$ and $k=b$

$\frac{d(K_k^T)_{\ell,j}}{d(K_k)_{a,b}}$ equals $\frac{d(K_k)_{j,\ell}}{d(K_k)_{a,b}}$

Not zero when $j=a$, $\ell=b$

then for the trace of A and when $i=a$

$$\frac{d(A_{i,i})}{d(K_k)_{i,b}} = \underbrace{\left((HP'H^T+R)K_k^T\right)_{b,i}}_{} + \sum_k (K_k)_{i,k}\underbrace{\left(\sum_\ell\left(\sum_m((H)_{k,m}(\ldots)+(R)_{k,b})\right)\right)}_{(HP'H^T+R)_{k,b}}$$

$$\left((HP'H^T+R)K_k^T\right)^T_{i,b}$$

$$= \left(K_k(HP'H^T+R^T)\right)_{i,b}$$

$$= \left(K_k(HP'H^T+R)\right)_{i,b}$$

$\therefore$ We used $R^T=R$ as R is symmetric due to $R=E[v_k v_k^T]$

$$\left((K_k)(HP'H^T+R)\right)_{i,b}$$

$$= \left((K_k)(HP'H^T+R)\right)_{i,b} + \left((K_k)(HP'H^T+R)\right)_{i,b} = 2\left((K_k)(HP'H^T+R)\right)_{i,b}$$

derivative of trace would equal sum of the derivatives of the diagonal elements hence;

$$\frac{d(T[K_k(HP'H^T+R)K_k^T])}{d(K_k)_{i,j}} = \frac{d(A_{1,1})}{d(K_k)_{i,j}} + \ldots + \frac{d(A_{n,n})}{d(K_k)_{i,j}} = \frac{d(A_{i,i})}{d(K_k)_{i,j}} = 2\left((K_k)(HP'H^T+R)\right)_{i,j}$$

then

$$\frac{d(T[K_k(HP'H^T+R)K_k^T])}{dK_k} = 2K_k(HP'H^T+R)$$

So we've found that $\dfrac{d(T[P_k])}{dK_k} = -2(HP'_k)^T + 2K_k(HP'_kH^T + R)$

we equate it to zero and solve for K_k;

$$-2(HP'_k)^T + 2K_k(HP'_kH^T + R) = 0$$

$$(HP'_k)^T = K_k(HP'_kH^T + R)$$

$$K_k = (HP'_k)^T(HP'_kH^T + R)^{-1}$$

∴ Think about how we know it is an investible matrix

we've just derived the:

$$\boxed{\text{"Kalman Gain" Equation} \quad K_k = P'_kH^T(HP'_kH^T + R)^{-1}}$$ *

Now that we found the formula for the optimal Kalman Gain (K_k), we plug it into the P_k formula we arrived earlier.

In other words substitute $K_k = P'_kH^T(HP'_kH^T + R)^{-1}$ into

$$P_k = P'_k - P'_kH^TK_k^T - K_kHP'_k + K_k(HP'_kH^T + R)K_k^T$$

$$P_k = P'_k - P'_kH^T\left[P'_kH^T(HP'_kH^T+R)^{-1}\right]^T - \left[P'_kH^T(HP'_kH^T+R)^{-1}\right](HP'_k) + \left[P'_kH^T(HP'_kH^T+R)^{-1}\right](HP'_kH^T+R)\left[P'_kH^T(HP'_kH^T+R)^{-1}\right]^T$$

used P'_k symmetric cov matrix

$$= P'_k - P'_kH^T\left[(HP'_kH^T+R)^{-1}H(P'_k)^T\right] - P'_kH^T(HP'_kH^T+R)^{-1}(HP'_k) + P'_kH^T\underbrace{(HP'_kH^T+R)^{-1}(HP'_kH^T+R)}_{I}(HP'_kH^T+R)^{-1}H(P'_k)$$

$$= P'_k - \underbrace{P'_kH^T(HP'_kH^T+R)^{-1}}_{K_k}HP'_k$$

$$= P'_k - K_kHP'_k$$

$$= (I - K_kH)P'_k$$

we've just derived the:

$$\boxed{\text{"Covariance Update" Equation with Optimal Gain} \quad P_k = (I - K_kH)P'_k}$$ *

State Projection into $k+1$ is given by $\hat{x}'_{k+1} = \Phi \hat{x}_k$

To complete recursion we now try to find an equation which projects the error covariance matrix (P_k) into the next time interval, $k+1$.

$$e'_{k+1} = x_{k+1} - \hat{x}'_{k+1} = (\Phi x_k + w_k) - \Phi \hat{x}_k = \Phi x_k - \Phi \hat{x}_k + w_k = \Phi \underbrace{(x_k - \hat{x}_k)}_{e_k} + w_k = \Phi e_k + w_k$$

now extend the previously established $P_k = E[e_k e_k^T] = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$ from k to $k+1$

plug in

$$P'_{k+1} = E[e'_{k+1} e'^T_{k+1}] = E[(\Phi e_k + w_k)(\Phi e_k + w_k)^T]$$

$$= E[\Phi e_k (\Phi e_k)^T + (\Phi e_k)(w_k^T) + (w_k)(\Phi e_k)^T + w_k w_k^T]$$

∴ used $E[x+y] = E[x] + E[y]$

$$= E[\Phi e_k (\Phi e_k)^T] + \underbrace{E[(\Phi e_k)(w_k^T)]}_{0} + \underbrace{E[w_k (\Phi e_k)^T]}_{0} + E[w_k w_k^T]$$

e_k and w_k have zero cross-correlation. and when two random variables have $\mathrm{cov}(x,y)=0$ then $E[xy]=0$

$$= E[\Phi e_k (\Phi e_k)^T] + \underbrace{E[w_k w_k^T]}_{Q}$$

$Q \rightarrow$ "covariance matrix of the process white noise"

$$= E[\Phi e_k e_k^T \Phi^T] + Q = \Phi E[e_k e_k^T] \Phi^T + Q = \Phi P_k \Phi^T + Q$$

∴ used $E[aX] = aE[X]$

∴ recall that $P_k = E[e_k e_k^T]$

Covariance Matrix Projection into $k+1$ is given by $P_{k+1} = \Phi P_k \Phi^T + Q$