Discrete Kalman Filter

Andersen Ang

First created: 2014. Last update: 2017-Feb-1

1. A discrete system

A discrete linear system in terms of difference equation

$$\begin{cases} x_k = Ax_{k-1} + Bu_{k-1} \\ y_k = Cx_k \end{cases}$$

where A, B, C are all matrix and x, u, z are all vector.
Matrix A tells about how the system current internal state x_k related to its previous state x_{k-1}
Matrix B tells about how the system current internal state x_k related to its previous input u_{k-1}
Matrix C tells about how the system output y_k related to its current internal state x_k

In the real world situation, the environment, the equipment, the system, are all full of noises. Thus there are *process noise* w and *measurement noise* v .
Noise can be characterized by its *average magnitude* : the standard deviation σ
The noise distribution can be characterized by frequency spectrum. For *white noise*, the noise has Gaussian distribution $\mathcal{N}\left(\mu, \sigma^2\right)$.
Thus the discrete linear system becomes

$$\begin{cases} x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \\ y_k = Cx_k + v_k \end{cases}$$

with probability distribution of the noises as

$$\begin{cases} \mathcal{P}(w) \sim \mathcal{N}(0, Q) \\ \mathcal{P}(v) \sim \mathcal{N}(0, R) \end{cases}$$

The two distributions are assumed $zero - mean$ $(E[w_k] = E[v_k] = 0)$, *independent* $E[w_i w_j] = 0$ for $i \neq j$, and the constant Q , R here tells the covariance of the noises.

2. The problem

With such system

$$\begin{cases} x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \\ y_k = Cx_k + v_k \end{cases} \qquad \begin{cases} \mathcal{P}(w) \sim \mathcal{N}(0, Q) \\ \mathcal{P}(v) \sim \mathcal{N}(0, R) \end{cases}$$

We usually only have the *external input output information*. By using known input u , the output of the system y is obtained. And the *internal* state variables x are hidden to the external world. Since understanding the internal process of the system provides many useful information, thus it is useful to *estimate the state* x_k. In the *observer* , the state variables are being estimated *asymptotically*. The Kalman filter, also estimate the state variable x , but it is based on *minimization of squared error*.

To estimate the state variable x_k (at stage k), there are two estimates : the *priori* estimate $\hat{x}_k^-$ and the *posteriori* $\hat{x}_k$.

- The priori state estimate is obtained based on the information given / known of the whole system before go to the stage k.
- The posteriori state estimate is obtained based on the measurement carried out exactly at stage k.

Thus, the priori state estimate $\hat{x}^-$ at stage k is estimated using the information at stage $k-1$, that is , using the posteriori state estimate $\hat{x}$, the system information A, B and the inputs u at stage $k-1$

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}$$

Thus, the posteriori state estimate $\hat{x}$ at stage k is estimated using the priori state estimate $\hat{x}^-$ at stage k and measurement y at stage k. And the aims is to "correct" the priori state estimate using the measurement.

$$\text{posteriori state estimate} = \text{priori state estimate} + \text{correction}$$

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

The "correction" should be *based on the difference between* the *real measurement* and the *predicted measurement*

$$\text{correction} \propto y_k - C\hat{x}_k^-$$

Thus , a proportional constant, called the *Kalman Gain* K_k is proposed

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

And therefore

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

Then, since the goal is to obtain best estimate based on minimum squared error, thus the errors have to be defined :

$$\begin{gathered} \text{priori error : } e_k^- = x_k - \hat{x}_k^- \\ \text{posteriori error : } e_k = x_k - \hat{x}_k \end{gathered}$$

And the errors can be characterised by the error covariance

$$\begin{gathered} \text{priori error covariance : } P_k^- = E\left[e_k^- e_k^{-T}\right] \\ \text{posteriori error covariance : } P_k = E\left[e_k e_k^T\right] \end{gathered}$$

Now we have formulated the initial problem

The problem statement
Given the noisy system

$$\begin{cases} x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \\ y_k = Cx_k + v_k \end{cases} \qquad \begin{cases} \mathcal{P}(w) \sim \mathcal{N}(0, Q) \\ \mathcal{P}(v) \sim \mathcal{N}(0, R) \end{cases}$$

With state estimations

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \qquad \hat{x}_k = \hat{x}_k^- + K_k \left(y_k - C\hat{x}_k^-\right)$$

With errors and the errors covariances defined as

$$\begin{gathered} \text{priori error : } e_k^- = x_k - \hat{x}_k^- \\ \text{posteriori error : } e_k = x_k - \hat{x}_k \end{gathered} \qquad \begin{gathered} \text{priori error covariance : } P_k^- = E\left[e_k^- e_k^{-T}\right] \\ \text{posteriori error covariance : } P_k = E\left[e_k e_k^T\right] \end{gathered}$$

Find the K to minimize the P

$$find\, K_k \text{ s.t. } P_k \text{ is minimized}$$

Kalman proved that , the K_k, called the Kalman gain is

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

3. Understanding the Kalman Gain

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

where C is the matrix relating the internal state variable x and the output. And R is a constant of the covariance of the measurement error, and P^- is the priori error covariance.

If $R \to 0$.
When $R \to 0$, it means the covariance magnitude of the measurement noise is small. Thus it means the *measurement is more accurate*. Thus, we should trust the measurement more.

$$\lim_{R \to 0} K_k = \lim_{R \to 0} P_k^- C^T \left(C P_k^- C^T + R \right)^{-1} = C^{-1}$$

And then

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

$$\hat{x}_k = \hat{x}_k^- + C^{-1} \left(y_k - C \hat{x}_k^- \right)$$

$$\hat{x}_k = C^{-1} y_k$$

Since $y_k = C x_k + v_k$, and if $R \to 0$, that means v_k has very small magnitude. And thus $y_k = C x_k$, and therefore $C^{-1} y_k = x_k$

$$\hat{x}_k = x_k$$

Which means we should take the estimate as the current measured internal state.

If $P_k^- \to 0$
When $P_k^- \to 0$, that means the priori estimatation error is small. Thus it means *priori estimate is more accurate*. Thus we should trust the estimate more.

$$\lim_{P_k^- \to 0} K_k = \lim_{P_k^- \to 0} P_k^- C^T \left(C P_k^- C^T + R \right)^{-1} = 0$$

And then

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

$$\hat{x}_k = \hat{x}_k^- + 0 \left(y_k - C \hat{x}_k^- \right)$$

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

Which means we should take the estimate as the current estimated state.

General case
In general case, $P_k^- > 0$ and $R > 0$. Thus the posteriori state estimate is the *mix* of the priori estimate and the measurement.

4. The Kalman Filter Algorithm

Apart from the state estimated being "updated" , the "Kalman Gain", "error covariance" also need to be updated. The Kalman filter algorithm has 2 groups of equations : the prediction and the correction

The Kalman Filter Algorithm
1. Initial estimates of x_{k-1} and P_{k-1}
Prediction Eqaution
2. Predict the state $\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}$
3. Predict the error covariance $P_k^- = AP_{k-1}A^T + Q$
Correction
4. Obtain the measurement $y_k = Cx_k + v_k$
5. Compute the Kalman Gain $K_k = P_k^- C^T \left(CP_k^- C^T + R\right)^{-1}$
6. Correct the estimate with measurement $\hat{x}_k = \hat{x}_k^- + K_k\left(y_k - C\hat{x}_k^-\right)$
7. Correct the error covariance $P_k = \left(I - K_k C\right) P_k^-$
8. Repeat step 2 to 7 until end condition is satisfied

5. How to obtain the Kalman gain

Why $K_k = P_k^- C^T \left(CP_k^- C^T + R\right)^{-1}$? Recall that K is to minimize P, thus consider P_k

$$P_k = E\left[e_k e_k^T\right]$$

where

$$e_k = x_k - \hat{x}_k$$

Thus

$$P_k = E\left[\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T\right]$$

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

$$P_k = E\left[\left(x_k - \hat{x}_k^- - K_k\left(y_k - C\hat{x}_k^-\right)\right)\left(x_k - \hat{x}_k^- - K_k\left(y_k - C\hat{x}_k^-\right)\right)^T\right]$$

Applies $y_k = Cx_k + v_k$

$$P_k = E\left[\left(x_k - \hat{x}_k^- - K_k\left(Cx_k + v_k - C\hat{x}_k^-\right)\right)\left(x_k - \hat{x}_k^- - K_k\left(Cx_k + v_k - C\hat{x}_k^-\right)\right)^T\right]$$

Rearrange

$$P_k = E\left[\left(x_k - \hat{x}_k^- - K_k C\left(x_k - \hat{x}_k^-\right) - K_k v_k\right)\left(x_k - \hat{x}_k^- - K_k C\left(x_k - \hat{x}_k^-\right) - K_k v_k\right)^T\right]$$

$$P_k = E\left[\left(\left(I - K_k C\right)\left(x_k - \hat{x}_k^-\right) - K_k v_k\right)\left(\left(I - K_k C\right)\left(x_k - \hat{x}_k^-\right) - K_k v_k\right)^T\right]$$

where $x_k - \hat{x}_k^- = e_k^-$, the priori error

$$P_k = E\left[\left(\left(I - K_k C\right)e_k^- - K_k v_k\right)\left(\left(I - K_k C\right)e_k^- - K_k v_k\right)^T\right]$$

Expand the terms with $(A+B)^T = A^T + B^T$

$$P_k = E\left[\left(\left(I - K_k C\right)e_k^- - K_k v_k\right)\left(e_k^{-T}\left(I - K_k C\right)^T - v_k^T K_k^T\right)\right]$$

$$P_k = E\left[(I - K_k C)\, e_k^- e_k^{-T} (I - K_k C)^T - (I - K_k C)\, e_k^- v_k^T K_k^T - K_k v_k e_k^{-T} (I - K_k C)^T + K_k v_k v_k^T K_k^T\right]$$

Expectation operator is linear $E[A + B] = E[A] + E[B]$

$$P_k = E\left[(I - K_k C)\, e_k^- e_k^{-T} (I - K_k C)^T\right] - E\left[(I - K_k C)\, e_k^- v_k^T K_k^T\right] - E\left[K_k v_k e_k^{-T} (I - K_k C)^T\right] + E\left[K_k v_k v_k^T K_k^T\right]$$

Since e_k^- is the priori error, is the error obtained before making the measurement, and thus such error is uncorrelated with measurement noise

$$E\left[e_k^- v_k^T\right] \equiv 0$$

And thus the two terms in the middle are both zero

$$P_k = E\left[(I - K_k C)\, e_k^- e_k^{-T} (I - K_k C)^T\right] + E\left[K_k v_k v_k^T K_k^T\right]$$

Expectation operator is linear $E[kA] = kE[A]$

$$P_k = (I - K_k C)\, E\left[e_k^- e_k^{-T}\right] (I - K_k C)^T + K_k E\left[v_k v_k^T\right] K_k^T$$

The term $E\left[e_k^- e_k^{-T}\right] = P_k^-$ is the priori error covariance

$$P_k = (I - K_k C)\, P_k^- (I - K_k C)^T + K_k E\left[v_k v_k^T\right] K_k^T$$

The term $E\left[v_k v_k^T\right] = R$ is the measurement noise covariance (which is assumed to be constant covariance in this document)

$$P_k = (I - K_k C)\, P_k^- (I - K_k C)^T + K_k R K_k^T$$

Expand

$$P_k = (I - K_k C)\, P_k^- \left(I - C^T K_k^T\right) + K_k R K_k^T$$

$$P_k = \left(P_k^- - K_k C P_k^-\right)\left(I - C^T K_k^T\right) + K_k R K_k^T$$

$$P_k = P_k^- - P_k^- C^T K_k^T - K_k C P_k^- + K_k C P_k^- C^T K_k^T + K_k R K_k^T$$

Group the last two term $K_k(\,.\,)K_k^T$ together

$$P_k = P_k^- - P_k^- C^T K_k^T - K_k C P_k^- + K_k \left(C P_k^- C^T + R\right) K_k^T$$

Now the posteriori error covariance P_k is expressed as a function of K_k.

The goal is to find K_k such that the squared error is minimized. The squared error terms are the diagonal terms of the covariance matrix.

$$\text{sum of squared errors} = \text{sum of all diagonal element of the covariance matrix}$$

Thus , take the *trace*

$$\text{sum of squared errors} = \mathrm{tr} P_k = \mathrm{tr} P_k^- - \mathrm{tr}\left(P_k^- C^T K_k^T\right) - \mathrm{tr}\left(K_k C P_k^-\right) + \mathrm{tr}\Big(K_k \left(C P_k^- C^T + R\right) K_k^T\Big)$$

Since $\mathrm{tr} A = \mathrm{tr} A^T$, thus $\mathrm{tr}\left(P_k^- C^T K_k^T\right) = \mathrm{tr}\left(K_k C P_k^-\right)$

$$\text{sum of squared errors} = \mathrm{tr} P_k = \mathrm{tr} P_k^- - 2\mathrm{tr}\left(P_k^- C^T K_k^T\right) + \mathrm{tr}\Big(K_k \left(C P_k^- C^T + R\right) K_k^T\Big)$$

Now to find K_k that minimize sum of squarred errors, take the derivative w.r.t K_k

$$\frac{d}{dK_k}\text{tr}P_k = \frac{d}{dK_k}\left[\text{tr}P_k^- - 2\text{tr}\left(P_k^- C^T K_k^T\right) + \text{tr}\Bigg(K_k\left(CP_k^- C^T + R\right)K_k^T\Bigg)\right]$$

$$\frac{d}{dK_k}\text{tr}P_k = \underbrace{\frac{d}{dK_k}\text{tr}P_k^-}_{0} - 2\frac{d}{dK_k}\text{tr}\left(P_k^- C^T K_k^T\right) + \frac{d}{dK_k}\text{tr}\Bigg(K_k\left(CP_k^- C^T + R\right)K_k^T\Bigg)$$

$$\frac{d}{dK_k}\text{tr}P_k = -2\frac{d}{dK_k}\text{tr}\left(P_k^- C^T K_k^T\right) + \frac{d}{dK_k}\text{tr}\Bigg(K_k\left(CP_k^- C^T + R\right)K_k^T\Bigg)$$

Using $\frac{d}{dA}\text{tr}\left(BA^T\right) = B$

$$\frac{d}{dK_k}\text{tr}P_k = -2\left(P_k^- C^T\right) + 2K_k\left(CP_k^- C^T + R\right)$$

Set it to be zero

$$0 = -2\left(P_k^- C^T\right) + 2K_k\left(CP_k^- C^T + R\right)$$

$$P_k^- C^T = K_k\left(CP_k^- C^T + R\right)$$

And thus K_k is

$$K_k = P_k^- C^T\left(CP_k^- C^T + R\right)^{-1}$$

And recalled that since P_k is

$$P_k = P_k^- - P_k^- C^T K_k^T - K_k C P_k^- + K_k\left(CP_k^- C^T + R\right)K_k^T$$

Consider the last term , the K_k will cancel with the $\left(CP_k^- C^T + R\right)$

$$P_k = P_k^- - P_k^- C^T K_k^T - K_k C P_k^- + P_k^- C^T K_k^T$$

$$P_k = P_k^- - K_k C P_k^-$$

Thus

$$P_k = \left(I - K_k C\right)P_k^-$$

$-END-$