EE6221 Robotics and Intelligent Sensors (Part 3)

Lecture 4: Kalman Filter and Sensor Integration

Guoqiang Hu
School of EEE, NTU

Sensor Fusion by Weighted Average

- One of the simplest and most intuitive methods of **signal-level fusion** is to take a weighted average of redundant information provided by a group of sensors and use this as the fused value. The weighted average of *n* sensor measurements x, with weights is found by

$$\bar{x} = \sum_{i=1}^{n} w_i x_i \quad \text{with weights } 0 \leq w_i \leq 1$$

- The weights can be used to account for the differences in accuracy between sensors, and a moving average can be used to fuse together a sequence of measurements from a single sensor so that the more recent measurements are given a greater weight.

Sensor Fusion by Weighted Average

- Whereas this method allows for the real-time processing of dynamic low-level data, in most cases, a **Kalman filter is preferred** because it provides a method that is nearly equal in processing requirements and, in contrast to a weighted average, it results in an estimate for the fused data that is **optimal in a statistical sense.**

Background Knowledge

- Expectation
- Variance
- Covariance

Expectation

- **Expectation E**

  E (x) is the true average (mean) of x

- Example : Consider a discrete random variable X with total population consisting of $X = \{1, 1, 1, 2, 2, 2, 2, 3\}$.

  Thus any element x_i of X can take on one of three values.

$$x_i \in \{1, 2, 3\}$$

$$E(X) = (1 + 1 + 1 + 2 + 2 + 2 + 2 + 3) / 8 = 14 / 8$$

$$= (1 \times 3 + 2 \times 4 + 3 \times 1) / 8$$

$$= 1 \times 3/8 + 2 \times 4/8 + 3 \times 1/8$$

$$E(X) = 1 \times P(X=1) + 2 \times P(X=2) + 3 \times P(X=3)$$

$$E(X) = 1 \times P(X=1) + 2 \times P(X=2) + 3 \times P(X=3)$$

Therefore, in general

$$E(X) = \sum_{i=1}^{N} x_i \, P(x_i)$$

X is a discrete random variable

where $x_i \in \{x_1, x_2, \ldots, x_N\}$

and $P(x_i)$ is the probability of x_i.

If the random variable X is continuous, then

$$E(X) = \int_X x \, p(x) \, dx$$

X is a continuous random variable

where $p(x)$ is the probability density of x.

What is $E[f(X)]$? where f is any function of X.

Example : Consider a discrete random variable X with total population consisting of $X = \{1, 1, 1, 2, 2, 2, 2, 3\}$.

Thus, any element x_i of X can take on one of three values

$$x_i \in \{1, 2, 3\}$$

Similarly, $f(X)$ can take on one of three values also.

$$f(X = x_1) \text{ , } f(X = x_2) \text{ and } f(X = x_3)$$

The number of occurrence of $f(X = x_i)$ is the same as the number of occurrence of x_i .

Thus, $E(f(X))$

$= f(X = x_1) \times P(x_1) + f(X = x_2) \times P(x_2) + f(X = x_3) \times P(x_3)$

Therefore, $E[\,f(X)] = \sum_{i=1}^{N} f(x_i)P(x_i)$

If the random variable X is continuous, then

$$E\ [f(X)] = \int_X f(x)\,p(x)dx$$

where $p(x_i)$ is the probability density of x_i.

NOTE: Expectation operator uses the entire population of the random variable to give the ***true average***.
If we only use a subset of the population, then we get the ***sample average***. In general, the sample average is not equal to the true average.

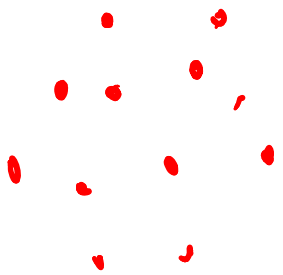
Variance

Variance of X :

The basic information conveyed by the Variance is how much of ***spread*** are the elements of X.

The greater the variance of X, the greater is the spread of the elements from the true average of X.

Set A



Set B

The variance of A is obviously larger than the variance of B

How is variance measured ?

Find the average squared distance of an element of X from the mean of X:

$$V(X) = E\{[X - E(X)]^2\}$$ Definition

Simplification

$$\begin{aligned} V[X] &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - E[2XE[X]] + E[E[X]^2] \\ &= E[X^2] - 2E[X]^2 + E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Thus the formula for Variance can be simplified to

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

If X is zero mean, that is $E(X) = 0$,
then the variance of zero-mean X is just

$$V(X) = E[X^2] \qquad \text{for zero-mean X}$$

What about the case when X is a continuous random variable ? The meaning is exactly the same.

Recall $E[f(X)] = \int_X f(x)\, p(x)\, dx$

Let $f(X) = [(X - E(X)]^2$

$$E\left\{[X - E(X)]^2\right\} = \int_X [x - E(X)]^2\, p(x)\, dx$$

It can be simplified (derivation on next page) to

$$E\{\ [X - E(X)]^2\ \} = \int_X x^2\, p(x) dx\ -\ [E(X)]^2 \qquad (\textit{general formula})$$

$$= \int_X x^2\, p(x)\, dx \qquad (\textit{if } X \textit{ is zero mean})$$

where as before $[E(X)]^2 = [\int_X x\, p(x) dx\ ]^2$

Derivation :

$$E\{\ [X - E(X)]^2\ \} = \int_X [x - E(X)]^2\ p(x)dx$$

$$= \int_X x^2\ p(x)dx + [E(X)]^2 \int_X p(x)dx - 2\int_X x\ E(X)\ p(x)dx$$

$$= \int_X x^2\ p(x)dx + [E(X)]^2 \int_X p(x)dx - 2\,E(X)\int_X x\ p(x)dx$$

$$= \int_X x^2\ p(x)dx + [E(X)]^2 \int_X p(x)dx - 2\,E(X)\,E(X)$$

$$= \int_X x^2\ p(x)dx - [E(X)]^2 \int_X p(x)dx$$

$$= E[X^2] - [E(X)]^2$$

total probability = 1

Covariance

Independent Random Variables

The covariance between two signals X and Y is defined as $V(X,Y) = E\{ (X - E(X))\,(Y - E(Y)) \}$ [covariance]

$$E\{ (X - E(X))\,(Y - E(Y)) \} = E[XY] - E[E(X)Y] - E[XE(Y)] + E[E(X)E(Y)]$$

$$= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y]$$

$$= E[XY] - E[X]E[Y]$$

If two random variables, X and Y, are (uncorrelated) independent, then $V(X,Y) = E\{ (X - E(X))\,(Y - E(Y)) \} = E[XY] - E(X)E(Y) = 0$

And if their mean values are zero, i.e. $E(X) = 0$ and $E(Y) = 0$, then

$$E[X\,Y] = 0 \qquad (X, Y \text{ are zero mean})$$

$V(X + Y) = V(X) + V(Y)$ for X and Y uncorrelated

Proof :

from page 12

$V(X + Y)$

$= E[(X + Y)^2] - (E[X + Y])^2$

$= E[(X^2 + Y^2 + 2XY] - (E[X] + E[Y])^2$

$= E[X^2] + E[Y^2] + 2\,E[XY] - (E[X])^2 - (E[Y])^2 - 2E[X]E[Y]$

$= E[X^2] + E[Y^2] - (E[X])^2 - (E[Y])^2 + 2E[XY] - 2E[X]E[Y]$

$= E[X^2] - (E[X])^2 + E[(Y)^2] - (E[Y])^2 + 2Cov(XY)$

$= V(X) + V(Y)$

since $Cov(XY) = 0$ if X and Y are uncorrelated.

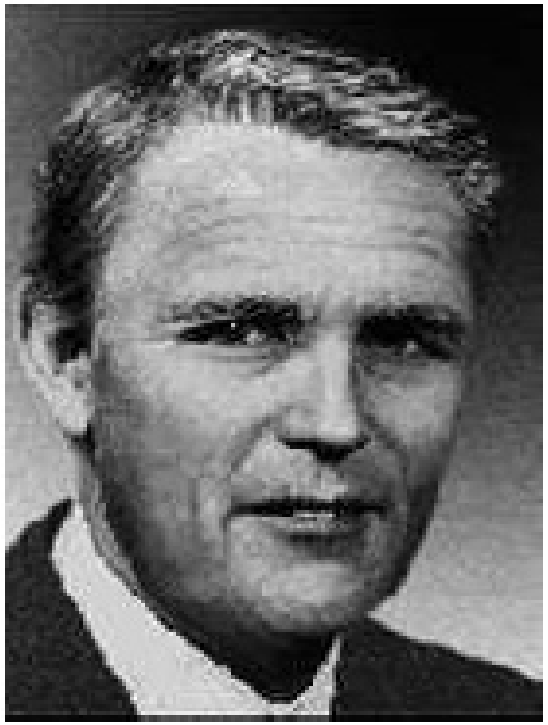
Kalman Filter for Sensor Fusion / Integration

Intro of Kalman Filter

- Originally, **Kalman** proposed his method for state estimation. It has been extended to other applications such as tracking. Here, we study how the Kalman filter can be applied to sensor fusion.
- The Kalman filter is used in a number of multi-sensor systems when it is necessary to fuse dynamic low-level redundant data in real-time. The filter uses the statistical characteristics of a measurement model to recursively determine estimates for the fused data that are optimal in a statistical sense. *If the system can be described with a linear model and both the system and sensor error can be modeled as white Gaussian noise, the Kalman filter will provide* ***unique statistically optimal estimates*** *for the fused data.*
- The recursive nature of the filter makes it appropriate for use in systems without large data storage capabilities.

Examples of the use of the filter for multi-sensor fusion include object recognition using sequences of images from a sensor, robot navigation, multi-target tracking, inertial navigation, and remote sensing.

The measurements from a group of *n* sensors can be fused using a Kalman filter to provide both an estimate of the current state of a system and a prediction of the future state of the system.



The state being estimated may, for example, correspond to the current location of a mobile robot, the position and velocity of an object in the environment, features extracted from sensory data (e.g., edges in an image), or the actual measurements themselves.

Kalman Estimation

KALMAN Estimation: Given a system represented as a linear discrete Markov process, the state-space model

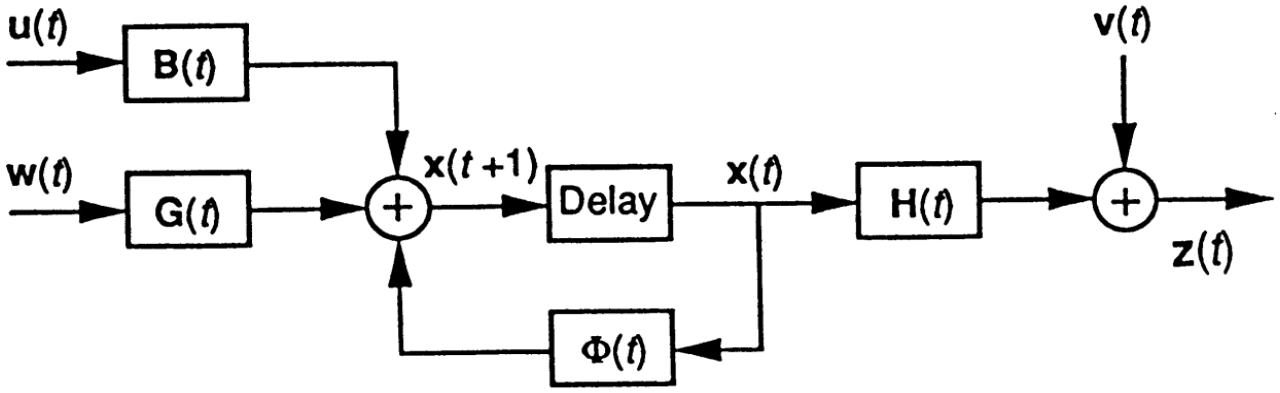
$$x(t+1) = \Phi(t)x(t) + B(t)u(t) + G(t)w(t) \qquad (2)$$

Zero-mean Gaussian noise

and the measurement model

$$z(t) = H(t)x(t) + v(t) \qquad (3)$$

can be used to describe the system below, where



x : m state vector

Φ : m × m state transition matrix

B : m × p input transition matrix

u : p input vector (e.g., position of sensor platform)

G : m × q process noise transmission matrix

w : q process noise vector

z : n measurement vector

H : n × m measurement matrix

v : n measurement noise vector

The *w* and *v* are assumed uncorrelated discrete-time ***zero-mean*** white Gaussian noise sequences with covariance kernels:

$$E\{w(t_i)w^T(t_j)\} = Q(t_i)\delta_{ij} \quad (4)$$

$$E\{v(t_i)v^T(t_j)) = R(t_i)\delta ij \quad (5)$$

where E{· · ·} denotes the expectation operator and the Kronecker delta function.

Kalman Filtering Equations

When all of the parameters (the matrices Φ, B, G, H, Q and R) of the models are known, the optimal *Kalman filtering* equations are:

$$\hat{X}(t|t) = \hat{x}(t|t-1) + K(t)[z(t) - H(t)\hat{x}(t|t-1)] \quad (6)$$

$$\hat{x}(t+1|t) = \Phi(t)\hat{x}(t|t) + B(t)u(t) \quad (7)$$

where $\hat{x}(t|t)$ is the estimate of $x(t)$ based on the measurements $\{z(0), \cdots, z(t)\}$,

and $\hat{x}(t+1|t)$ is the prediction of $x(t+1)$ based on the measurements $\{z(0), \cdots, z(t)\}$.

The m× n matrix *K* is the "Kalman filter gain" and is computed by

$$K(t) = P(t|t-1)H^T(t)[H(t)P(t|t-l)H^T(t) + R(t)]^{-1} \quad (8)$$

where $P(t|t-1) = E\{(x(t) - \hat{x}(t|t-1))(x(t) - \hat{x}(t|t-1))^T\}$

is the m×n conditional covariance matrix of the error in predicting *x*(*t*) and is determined using

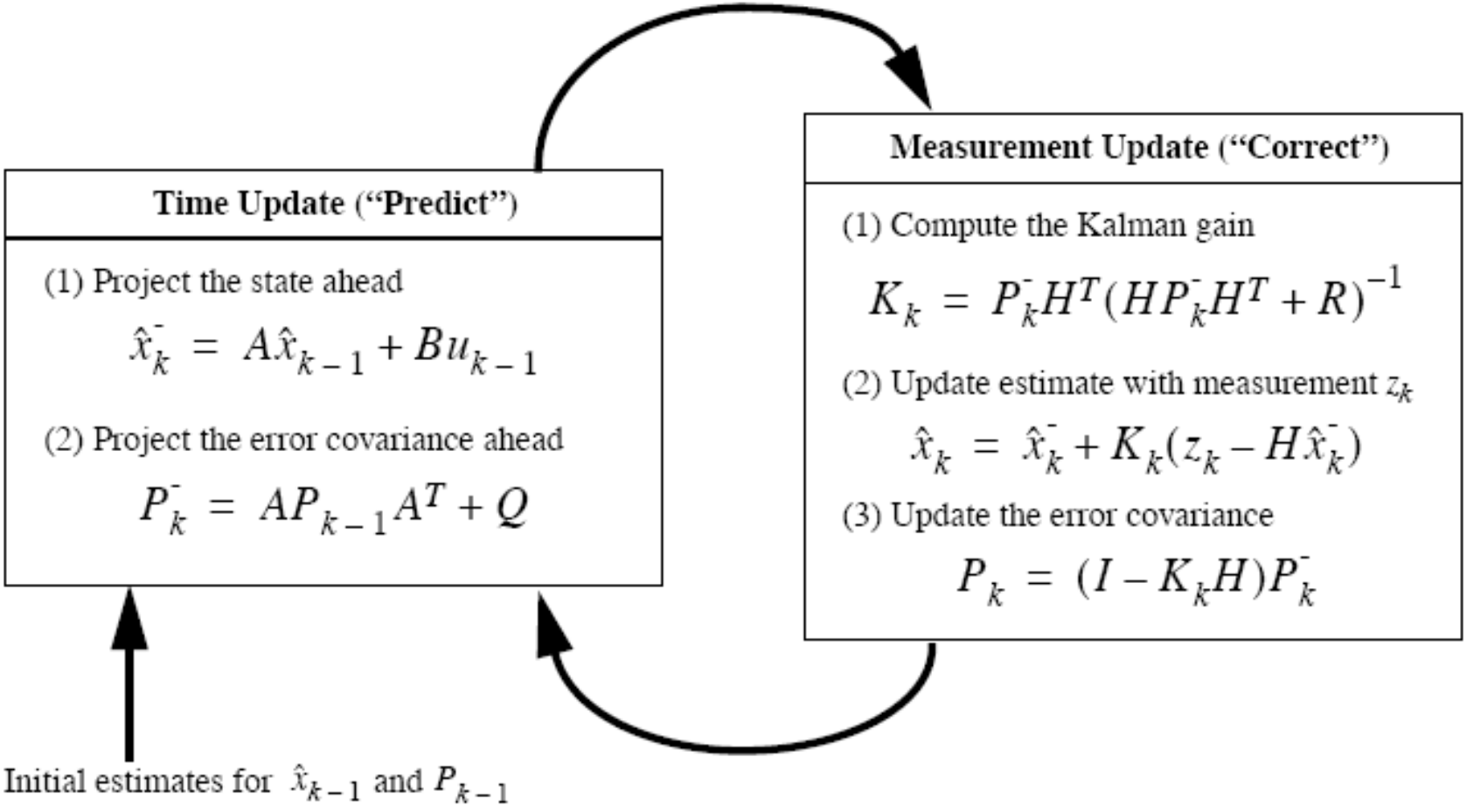
$$P(t+1|t) = \Phi(t)P(t|t)\Phi^T(t) + G(t)Q(t)G^T(t) \quad (9)$$

where $P(t|t) = P(t|t-1) - K(t)H(t)P(t|t-1)$

The initial conditions for the recursion are given by $\hat{x}(0|0) = \hat{x_0}$ and $P(0|0) = P_0$.

Sequential State Estimation

Sequential State Estimation by Kalman Filtering



Derivation of the Kalman Gain

Derivation of the Kalman Gain for the scalar case

Given : $x_{k+1} = Ax_k + Bu_k + w_k$ --- (a)

and the measurement model

$$z_k = H x_k + v_k \quad \text{--- (b)}$$

where w_k and v_k are gaussian white noise – processes with zero mean and variances Q and R respectively. Assume also that the initial state $x(0)$ is Gaussian and $E\{x(0)\} = m_0$ and $\text{cov}\{x(0)\} = R_0$.

Q, R and R_0 are non-negative scalars.

The Kalman estimate of x is of the form :

$$\hat{x}_{k+1} = A\,\hat{x}_k + Bu_k + K_k\,[z_k - H\,\hat{x}_k\,] \qquad \text{---}(c)$$

The Kalman filter design principle is to ***minimize the variance*** of the new state estimate.

$$\tilde{x}_{k+1} = x_{k+1} - \hat{x}_{k+1} \qquad \text{---}(d)$$

$$\tilde{x}_{k+1} = \{A\,x_k + Bu_k + w_k\} - \{A\,\hat{x}_k + Bu_k + K_k\,[z_k - H\,\hat{x}_k\,]\}$$

$$= A\,x_k + Bu_k + w_k - \{A\,\hat{x}_k + Bu_k + K_k\,[H\,x_k + v_k - H\,\hat{x}_k\,]\}$$

$$= A(x_k - \hat{x}_k) + w_k - K_k\,[H\,(x_k - \hat{x}_k) + v_k\,]$$

$$= A(\tilde{x}_k) + w_k - K_k\,[H\,\tilde{x}_k + v_k\,] \qquad \text{---}(e)$$

$$P_{k+1} = V(\tilde{x}_{k+1}) = E(\tilde{x}_{k+1}^{\,2}) - E(\tilde{x}_{k+1})^2$$

$$= E(\tilde{x}_{k+1}^{\,2}) \qquad \textit{since} \quad E(\tilde{x}_{k+1}) = 0 \quad \textit{is assumed.}$$

$$P_{k+1} = E(\tilde{x}_{k+1}^2) = E\{(A\,\tilde{x}_k + w_k - K_k H\,\tilde{x}_k - K_k v_k\,])^2\}$$

$$= E\{([A - K_k H]\,\tilde{x}_k + w_k - K_k v_k)^2\}$$

$$= E\{[A - K_k H]^2\,\tilde{x}_k^2\} + E\{w_k^2\} + E\{K_k^2 v_k^2\}$$

In the above expansion, note that the expectation of terms containing ($\tilde{x}$, w), ($\tilde{x}$, v) and (w, v) are zero as they are assumed to be uncorrelated.

$$P_{k+1} = [A - K_k H]^2 \, E(\tilde{x}_k^2) + E(w_k^2) + K_k^2 \, E(v_k^2)$$

$$= [A - K_k H]^2 \, P_k + Q + K_k^2 \, R \qquad ---(f)$$

Kalman minimizes P_{k+1} with respect to K_k, the Kalman gain.

$$\frac{\partial P_{k+1}}{\partial K_k} = \frac{\partial}{\partial K_k}\left([A - K_k H]^2 P_k + Q + K_k^2 \, R \right)$$

$$= -2H[A - K_k H]P_k + 2K_k \, R = 0$$

$$HAP_k - K_k H^2 P_k = K_k R$$

$$[H^2 P_k + R] K_k = HAP_k$$

And finally, $$K_k = \frac{HAP_k}{H^2 P_k + R}$$

We could also find K in another way. From Equation (f) :

$$P_{k+1} = [A - K_k H]^2 P_k + Q + K_k^2 R$$

$$= [A - K_k H]^2 P_k + Q + K_k^2 R$$

Rearranging, $P_{k+1} = (H^2 P_k + R)[\, K_k^2 - \frac{2\,AHP_k}{H^2 P + R} K_k \,] + Q + A^2 P_k$

Form the perfect square in the square bracket :

$$P_{k+1} = (H^2 P_k + R)[\, K_k^2 - \frac{2\,AHP_k}{H^2 P_k + R} K_k + (\frac{-AHP_k}{H^2 P_k + R})^2 \,]$$

$$- (\frac{-AHP_k}{H^2 P_k + R})^2 + Q + A^2 P_k$$

$$= (H^2 P_k + R)[\, K_k - \frac{AHP_k}{H^2 P_k + R} \,]^2 - (H^2 P_k + R)(\frac{AHP_k}{H^2 P_k + R})^2 + Q + A^2 P_k$$

$$= (H^2 P_k + R)[\, K_k - \frac{AHP_k}{H^2 P_k + R} \,]^2 - \frac{A^2 H^2 P_k^2}{H^2 P_k + R} + Q + A^2 P_k \qquad \text{---(g)}$$

In Equation g, P_{k+1} is minimized w.r.t. K_k by setting K_k as

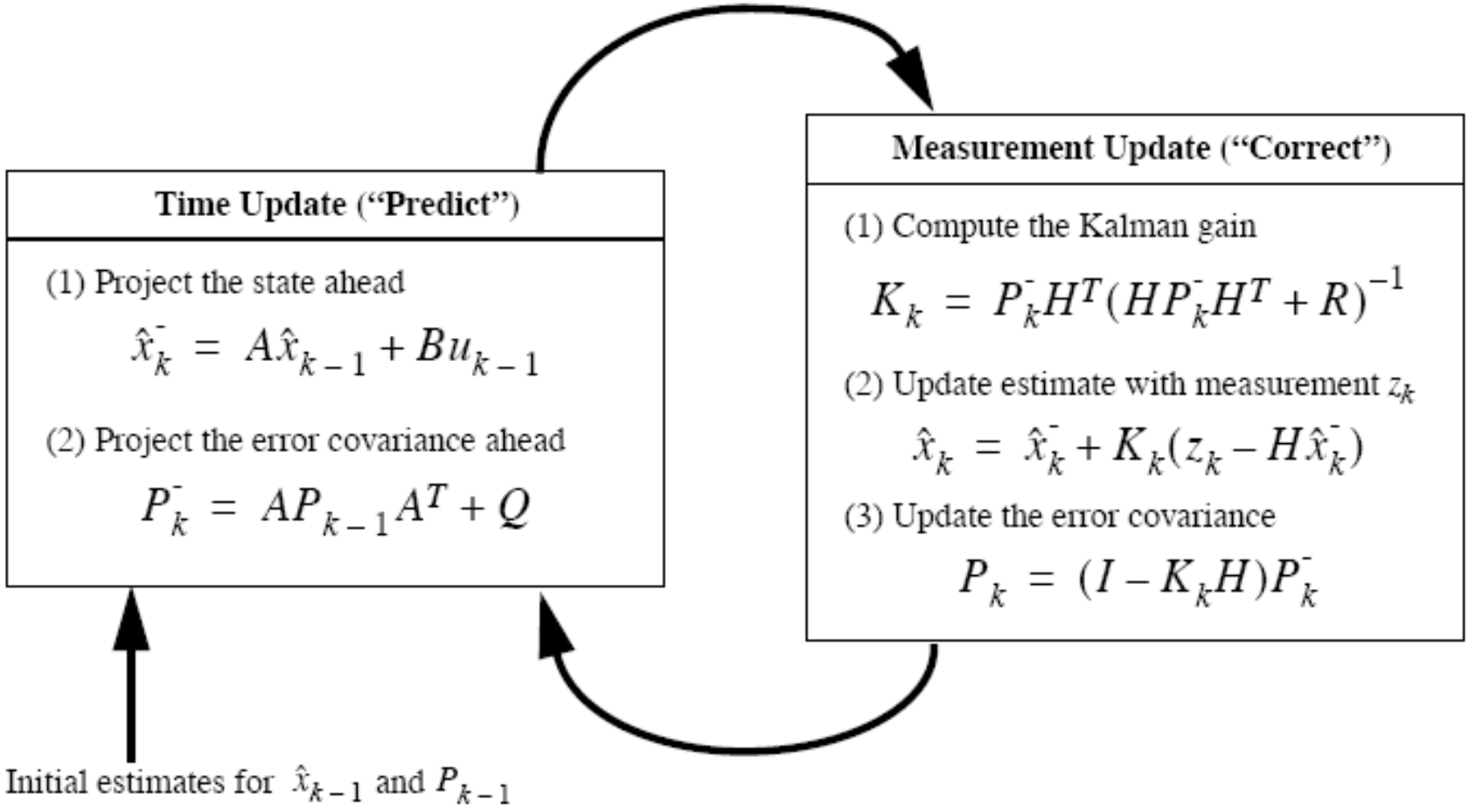
$$K_k = \frac{AHP_k}{H^2 P_k + R}$$

So, we obtain the same result as by the differentiation approach.

Now. the variance P_{k+1} becomes :

$$P_{k+1} = -\frac{(AHP_k)^2}{H^2 P_k + R} + Q + A^2 P_k$$

Sequential state estimation by Kalman Filtering



Kalman Filter for Sensor Fusion

Sensor information from different sensors can be fused by Kalman filter.

Example : The application of Kalman filtering for multi-sensor fusion can be illustrated using the object recognition of a landmark. Sensors 1 and 2, respectively S1 and S2, provide redundant information relative to each other concerning the shape of the objects to be recognized. The state to be estimated is the shape x of an object and can be assumed to ***remain constant*** over time, i.e., $x(t) = x$ for all t.

$$x(t+1) = x(t) \qquad \text{i.e. } \Phi \text{ (or } A) = 1$$

x(t+1) = x(t) shape (state) of object don't change

The shape measurements z_1 and z_2 from S1 and S2, respectively, can be modeled as

$$\begin{aligned} z_1 &= x + v_1 \\ z_2 &= x + v_2, \end{aligned} \qquad (10)$$

where v_1 and v_2 are independent zero-mean Gaussian random variables with variances σ_1^2 and σ_2^2, respectively.

We can consider ***two possible cases***. The two sensor information can be processed 1) simultaneously (batch) or 2) sequentially.

1) Batch Processing

If the measurements from S1 and S2 are ***available simultaneously***, batch processing can be used for fusion, where

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = Hx + v$$

Thus we have

$$x(t+1) = x(t)$$

$$z_1 = x + v_1$$

$$z_2 = x + v_2,$$

Shape does not change, so state x is constant.

where v_1 and v_2 are independent zero-mean Gaussian random variables with variances σ_1^2 and σ_2^2, respectively.

The Kalman estimate of x is of the form :

$$\hat{x}_{k+1} = \hat{x}_k + K_k [z_k - H \hat{x}_k]$$

$$= \hat{x}_k + K_1 (z_1 - \hat{x}_k) + K_2 (z_2 - \hat{x}_k)$$

The Kalman filter design principle is to ***minimize the variance*** of the new state estimate error. The state estimate error is given by

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

error *actual* *extimated*

Now,

$$\tilde{x}_{k+1} = x_k - [\hat{x}_k + K_1 (z_1 - \hat{x}_k) + K_2 (z_2 - \hat{x}_k)]$$

$$= x_k - \hat{x}_k - [K_1 (z_1 - \hat{x}_k) + K_2 (z_2 - \hat{x}_k)]$$

Q : We could also try for a slightly less general approach with only one Gain for both residuals :

$$\hat{x}_{k+1} = \hat{x}_k + K [(z_1 - \hat{x}_k) + (z_2 - \hat{x}_k)]$$

$$\tilde{x}_{k+1} = (x_k - \hat{x}_k) - [K_1(x_k + v_1 - \hat{x}_k) + K_2(x_k + v_2 - \hat{x}_k)]$$

$$= \tilde{x}_k - [K_1(\tilde{x}_k + v_1) + K_2(\tilde{x}_k + v_2)]$$

$$= [1 - K_1 - K_2]\tilde{x}_k - K_1 v_1 - K_2 v_2$$

Now the variance of $\tilde{x}_{k+1}$ is

$$P_{k+1} = V(\tilde{x}_{k+1}) = E(\tilde{x}_{k+1}^2) - E(\tilde{x}_{k+1})^2$$

$$= E(\tilde{x}_{k+1}^2) \qquad \text{since } E(\tilde{x}_{k+1}) = 0 \text{ is assumed.}$$

$$P_{k+1} = E(\tilde{x}_{k+1}^{\,2})$$

$$= E\{ ([\, 1 - K_1 - K_2 \,]\, \tilde{x}_k - K_1 v_1 - K_2 v_2)^2 \}$$

$$= [\, 1 - K_1 - K_2 \,]^2 E(\tilde{x}_k^{\,2}) + K_1^{\,2} E(v_1^{\,2}) + K_2^{\,2} E(v_2^{\,2})$$

In the above, we have assumed that the state estimate error and the noise terms v_1 and v_2 are all uncorrelated. *Therefore all the cross-correlation terms are zero.*

$$P_{k+1} = [\, 1 - K_1 - K_2 \,]^2 P_k + K_1^{\,2} \sigma_1^{\,2} + K_2^{\,2} \sigma_2^{\,2}$$

Thus the variance of the state estimation P_{k+1} error is updated from current P_k, K_1, K_2, σ_1 and σ_2.

Next step in Kalman prediction is to minimize P_{k+1} wrt K_1 and K_2.

$$\frac{\partial P_{k+1}}{\partial K_1} = \frac{\partial}{\partial K_1}\{[\,1-K_1-K_2\,]^2 P_k + K_1^2\sigma_1^2 + K_2^2\sigma_2^2\} = 0$$

$$-2[\,1-K_1-K_2\,]P_k + 2K_1\sigma_1^2 = 0$$

Rearranging, yields

$$-[\,1-K_1-K_2\,]P_k + K_1\sigma_1^2 = 0 \qquad ---(1)$$

Similarly,

$$\frac{\partial P_{k+1}}{\partial K_2} = \frac{\partial}{\partial K_2}\{[\,1-K_1-K_2\,]^2 P_k + K_1^2\sigma_1^2 + K_2^2\sigma_2^2\} = 0$$

$$-[\,1 - K_1 - K_2\,]\,P_k + K_2\,\sigma_2^{\,2} = 0 \qquad ---(2)$$

Equation 1 – equation 2 yields

$$K_1\sigma_1^{\,2} = K_2\sigma_2^{\,2} \qquad ---(3)$$

square all the sigmas

Solving,

$$K_1 = \frac{\sigma_2^{\,2} P_k}{\sigma_1^{\,2} P_k + \sigma_2^{\,2} P_k + \sigma_1^{\,2}\sigma_2^{\,2}}$$

$$K_2 = \frac{\sigma_1^{\,2} P_k}{\sigma_1^{\,2} P_k + \sigma_2^{\,2} P_k + \sigma_1^{\,2}\sigma_2^{\,2}}$$

2) Sequential Processing

If the measurements are available sequentially, recursive processing can be used to update the estimate of *x* as new measurements become available. Assuming that the measurement from S1 is available initially, $\hat{x}_0 = z_1$ and $P_0 = \sigma_1^2$ can be considered the *apriori* information available about *x* before reception of the measurement from S2. When z_2 becomes available, the optimal estimate of x is given by

Recall from page 31 : $K_k = \frac{AHP_k}{H^2 P_k + R}$

$$
\begin{aligned}
\hat{x} &= \hat{x}_0 + K[z_2 - H\hat{x}_0] \\
&= \hat{x}_0 + P_0 H^T (H P_0 H^T + R)^{-1} [z_2 - H\hat{x}_0] \\
&= z_1 + \sigma_1^2 (\sigma_1^2 + \sigma_2^2)^{-1} [z_2 - z_1] \\
&= \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2
\end{aligned}
\qquad (12)
$$

where R = diag(σ_1^2, σ_2^2).

Q: Derive the above from first principles

Interpretation The variances σ_1^2 and σ_2^2 in the estimate of $\hat{\chi}$ can be interpreted as providing a means of weighing each measurement z_1 and z_2 so that the measurement with the least variance is given the greatest weight in the fused estimate. The variance * of the estimate is

$$\sigma_1^2 \sigma_2^2 / (\sigma_1^2 + \sigma_2^2),$$

which is less than the variance of either measurement alone.

$$\sigma_1^2 \sigma_2^2 / (\sigma_1^2 + \sigma_2^2) < \sigma_1^2 \quad , \quad \sigma_1^2 \sigma_2^2 / (\sigma_1^2 + \sigma_2^2) < \sigma_2^2$$

The reduction in variance is shown in Figure 1 and represents the reduction in uncertainty due to the fusion of the measurements. The quantity $\hat{\chi}$ can be further updated as additional measurements become available from either sensor or other sources of information are made available.

* See derivation next page

$$v(x+y)=v(x)+v(y).....(I)$$

$$v(kx)=k^2v(x).....(II)$$

$$v(\hat{x})=v(\frac{{\sigma_2}^2}{{\sigma_1}^2+{\sigma_2}^2}z_1+\frac{{\sigma_1}^2}{{\sigma_1}^2+{\sigma_2}^2}z_2)$$

$$=\frac{{\sigma_2}^4}{\left({\sigma_1}^2+{\sigma_2}^2\right)^2}v(z_1)+\frac{{\sigma_1}^4}{\left({\sigma_1}^2+{\sigma_2}^2\right)^2}v(z_2)$$

$$=\frac{{\sigma_2}^4}{\left({\sigma_1}^2+{\sigma_2}^2\right)^2}{\sigma_1}^2+\frac{{\sigma_1}^4}{\left({\sigma_1}^2+{\sigma_2}^2\right)^2}{\sigma_2}^2=\frac{{\sigma_1}^2{\sigma_2}^2}{{\sigma_1}^2+{\sigma_2}^2}$$

Remarks on Kalman Filter

The Kalman filter is optimal if the following assumptions are true:

1) The system is linear;
2) The process and measurement noise are uni-modal Gaussian.

Other methods:

1) Extended Kalman filter
2) Unscented Kalman filter
3) Particle filter
4) Markov chain monte-carlo (MCMC), RANSAC
5) …