Man-Wai MAK

Dept. of Electronic and Information Engineering,
The Hong Kong Polytechnic University

enmwmak@polyu.edu.hk
http://www.eie.polyu.edu.hk/~mwmak

References:

November 22, 2018
Overview

1. Fundamental of Kalman Filter
   - What is Kalman Filter
   - Areas of Applications
   - 1-Dimensional Example
   - Kalman Filter’s Output and Optimal Linear Estimate
What is Kalman Filter

- Named after Rudolf E. Kalman, the Kalman filter is one of the most important and common data fusion algorithms in use today.
- The most famous early use of the Kalman filter was in the Apollo navigation computer that took Neil Armstrong to the moon.
- The Kalman filter has found numerous applications in fields related to control of dynamical systems, e.g., predicting the trajectory of celestial bodies, missiles and microscopic particles.
- Kalman filters are at work in every satellite navigation devices, every smart phone, UAV, and many computer games.
- The Kalman filter is a Bayesian model similar to a hidden Markov model but where the state space of the latent variables is continuous and where all latent and observed variables have a Gaussian distribution.
Kalman filter is suitable for applications where:
- the variables of interest can only be measured *indirectly*;
- measurements are available from various sensors, but they might be subject to noise.
- you have uncertain information about some dynamical systems but you can make an educated guess about what the system is going to do next.

https://www.youtube.com/watch?v=mwn8xhgNpFY
The Kalman filter assumes that the state of a system at a time $t$ evolved from the prior state at time $t - 1$ according to the equation:

$$x_t = F_t x_{t-1} + B_t u_t + w_t$$  \hspace{1cm} (1)

where

- $x_t$ is the state vector containing the terms of interest for the system at time $t$, e.g., position and velocity;
- $u_t$ is the vector containing any control inputs, e.g., steering angle, throttle setting, and braking force;
- $F_t$ is the state transition matrix which applies the effect of each system state parameter at time $t - 1$ on the system state at time $t$, e.g., the position and velocity at time $t - 1$ both affect the position at time $t$;
- $B_t$ is the control input matrix which applies the effect of each control input parameter in the vector $u_t$ on the state vector, e.g., applies the effect of the throttle setting on the system velocity and position;
- $w_t$ is the vector containing the process noise with covariance matrix $Q_t$ for each parameter in the state vector. Typically, $w_t \sim \mathcal{N}(0, Q_t)$. 

The system is measured according to the model

\[ z_t = H_t x_t + v_t \] (2)

where

- \( z_t \) is the vector of measurements, e.g., position;
- \( H_t \) is the transformation matrix that maps the state vector parameters into the measurement domain;
- \( v_t \) is a vector containing the measurement noise with covariance matrix \( R_t \). Typically, \( v_t \sim \mathcal{N}(0, R_t) \).

The true state \( x_t \) cannot be directly observed.

The Kalman filter gives an estimate \( \hat{x}_t \) by combining models of the system (\( F_t \) and \( B_t \)) and noisy measurements \( z_t \).

The covariance of estimation error at time \( t \) is denoted as \( P_t \).
The Kalman filter algorithm involves two stages: Prediction and Measurement update.

**Prediction stage** (estimate $\hat{x}_{t|t-1}$ from previous measurements up to $z_{t-1}$):

\[
\hat{x}_{t|t-1} = F_t \hat{x}_{t-1|t-1} + B_t u_t \\
P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q_t
\]  

Eq. 3 means that the new best estimate is a prediction made from previous best estimate, plus a correction for known external influences.

Eq. 4 means that the new uncertainty is predicted from the old uncertainty, with some additional uncertainty from the environment.
**Measurement update** (obtain $\hat{x}_{t|t}$ from current measurement $z_t$):

$$
\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t (z_t - H_t \hat{x}_{t|t-1}) 
$$  \hfill (5)

$$
P_{t|t} = P_{t|t-1} - K_t H_t P_{t|t-1} 
$$  \hfill (6)

$$
K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1} 
$$  \hfill (7)

In Eq. 5, if the measurement $z_t$ totally agrees with the prediction, then the current prediction is good enough and update is not necessary.
Kalman Filter Formulations

Prediction:
\[
\hat{x}_{t|t-1} = F_t \hat{x}_{t-1|t-1} + B_t u_t \\
P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q_t
\]

Update:
\[
\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t (z_t - H_t \hat{x}_{t|t-1}) \\
P_{t|t} = P_{t|t-1} - K_t H_t P_{t|t-1} \\
K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1}
\]
To show Eq. 4, we subtract Eq. 3 from Eq. 1 and compute the covariance of prediction error:

\[
P_{t|t-1} = \mathbb{E}\{(x_t - \hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})^T\}
\]

\[
= \mathbb{E}\left\{ (F_t x_{t-1} - F_t \hat{x}_{t-1|t-1} + w_t)(F_t x_{t-1} - F_t \hat{x}_{t-1|t-1} + w_t)^T \right\}
\]

\[
= F_t \mathbb{E}\left\{ (x_{t-1} - \hat{x}_{t-1|t-1})(x_{t-1} - \hat{x}_{t-1|t-1})^T \right\} F_t^T
\]

\[
+ F_t \mathbb{E}\{(x_{t-1} - \hat{x}_{t-1|t-1})w_t^T\} + \mathbb{E}\{w_t(x_{t-1} - \hat{x}_{t-1|t-1})^T\}F_t^T
\]

\[
+ \mathbb{E}\{w_t w_t^T\}
\]

\[
= F_t P_{t-1|t-1} F_t^T + Q_t
\]

where we have used the property that estimation error and noise are uncorrelated.
If \( w_t \) in Eq. 1 and \( v_t \) in Eq. 2 follow Gaussian distributions, we may write them as

\[
p(x_t | x_{t-1}) = \mathcal{N}(x_t | F_t x_{t-1} + B_t u_t, Q_t)
\]
\[
p(z_t | x_t) = \mathcal{N}(z_t | H_t x_t, R_t)
\]

The initial state also follows a Gaussian distribution:

\[
p(x_1) = \mathcal{N}(x_1 | B_1 u_1, Q_1)
\]
1-Dimensional Example

- We use the following one-dimensional tracking problem to derive (not so vigorously) Eq. 3–Eq. 7:

- We have the following variables and parameters:
  - $x_t$: The unknown (hidden state) distance from the pole to the train’s position at time $t$;
  - $z_t$: Noisy measure of the distance between the pole and the train’s RF antenna at time $t$ using time-of-flight techniques;
  - $H$: A proportional scale to make the time-of-flight measurement compatible with that of distance $x_t$;
  - $u_t$: The amount of throttle or brake applied by the train’s operator at time $t$;
  - $B$: A proportional scale to convert $u_t$ into distance;
  - $w_t$: Noise of $x_t$;
  - $v_t$: Measurement noise.
1-Dimensional Example

- At $t = 0$, we have an initial estimate (with error) of the train’s position:

[FIG2] The initial knowledge of the system at time $t = 0$. The red Gaussian distribution represents the pdf providing the initial confidence in the estimate of the position of the train. The arrow pointing to the right represents the known initial velocity of the train.

- At each time step $t$, we combine two information sources:
  - Predictions ($\hat{x}_{t-1|t-1}$) based on the last known position of the train.
  - Measurements ($z_t$) from a radio ranging system deployed at the track side.
1-Dimensional Example

- At $t = 0$, the location of the train is given by a Gaussian PDF.
- At $t = 1$, the new train position is predicted (by Eq. 1), which is represented by a new Gaussian PDF with a larger variance:

  ![Prediction (Estimate)](image1)

  \[ \text{[FIG3] Here, the prediction of the location of the train at time } t = 1 \text{ and the level of uncertainty in that prediction is shown. The confidence in the knowledge of the position of the train has decreased, as we are not certain if the train has undergone any accelerations or decelerations in the intervening period from } t = 0 \text{ to } t = 1. \]

- At $t = 1$, we also have a measurement $z_t$ of the train position. The measurement error is represented by another Gaussian PDF.

  ![Prediction (Estimate) & Measurement (Noisy)](image2)

  \[ \text{[FIG4] Shows the measurement of the location of the train at time } t = 1 \text{ and the level of uncertainty in that noisy measurement, represented by the blue Gaussian pdf. The combined knowledge of this system is provided by multiplying these two pdfs together.} \]
The best estimate can be obtained by combining our knowledge from the prediction and measurement.

Achieved by multiplying the two Gaussian PDFs, resulting in the green Gaussian:

![Diagram showing the prediction and measurement](FIG5)

*FIG5* Shows the new pdf (green) generated by multiplying the pdfs associated with the prediction and measurement of the train's location at time $t = 1$. This new pdf provides the best estimate of the location of the train, by fusing the data from the prediction and the measurement.

Note that the variance of the combined Gaussian is smaller, meaning that the predicted position has smaller variation.
To simplify notations, we omit subscript \( t \) and write the PDF of measurement (blue Gaussian) as:

\[
\mathcal{N}(z, \mu_z, \sigma_z) = \frac{1}{\sqrt{2\pi} \sigma_z} \exp\left\{ -\frac{(z - \mu_z)^2}{2\sigma_z^2} \right\}
\]

Because the predicted position is in meter and the measurement (time-of-flight of radio signal) is in second, we need to make the two units compatible.

This can be done by converting the prediction \( x \) to measurement \( z \) by setting \( H = \frac{1}{c} \), where \( c \) is the speed of light, i.e.,

\[
r = \frac{x}{c}
\]

The PDF of prediction becomes

\[
\mathcal{N}(r, \mu_x/c, \sigma_x/c) = \frac{1}{\sqrt{2\pi} \sigma_x/c} \exp\left\{ -\frac{(r - \mu_x/c)^2}{2(\sigma_x/c)^2} \right\}
\]
1-Dimensional Example

Note that both random variables $r$ and $z$ have the same unit (second). Their PDF can now be combined and written in terms of one random variable $s$:

$$
\mathcal{N}(s, \mu_f, \sigma_f) = \frac{\exp\left\{ -\frac{(s-(\mu_x/c))^2}{2(\sigma_x/c)^2} \right\}}{\sqrt{2\pi\sigma_x/c}} \cdot \frac{\exp\left\{ -\frac{(s-\mu_z)^2}{2\sigma_z^2} \right\}}{\sqrt{2\pi\sigma_z}}
$$

$$
= \frac{1}{\sqrt{2\pi\sigma_f}} \exp\left\{ -\frac{(s-\mu_f)^2}{2\sigma_f^2} \right\}
$$

where

$$
\mu_f = \mu_x + \left( \frac{\sigma_x^2/c}{\sigma_x^2/c^2 + \sigma_z^2} \right) \left( \mu_z - \frac{\mu_x}{c} \right)
$$

$$
\sigma_f^2 = \sigma_x^2 - \left( \frac{\sigma_x^2/c}{\sigma_x^2/c^2 + \sigma_z^2} \right) \frac{\sigma_x^2}{c}.
$$
Substituting $H = \frac{1}{c}$ and $K = \frac{H\sigma^2_x}{H^2\sigma^2_x + \sigma^2_z}$ into Eq. 8, we have

$$\begin{align*}
\mu_f &= \mu_x + K(\mu_z - H\mu_x) \\
\sigma^2_f &= \sigma^2_x - KH\sigma^2_x
\end{align*}$$

(9)

Now we set

$$\begin{align*}
\hat{x}_{t|t} &\leftarrow \mu_f \quad \text{(posterior mean)} \\
\hat{x}_{t|t-1} &\leftarrow \mu_x \\
P_{t|t} &\leftarrow \sigma^2_f \quad \text{(posterior covariance)} \\
P_{t|t-1} &\leftarrow \sigma^2_x \\
z_t &\leftarrow \mu_z \\
R_t &\leftarrow \sigma^2_z \\
H_t &\leftarrow H \\
K_t &\leftarrow K
\end{align*}$$
1-Dimensional Example

- Substituting these variables into Eq. 9, we have

\[ \mu_f = \mu_x + K(\mu_z - H\mu_x) \]

\[ \Rightarrow \hat{x}_{t|t} = \hat{x}_{t|t-1} + K(z_t - H_t\hat{x}_{t|t-1}) \]

\[ \sigma_f^2 = \sigma_x^2 - KH\sigma_x^2 \]

\[ \Rightarrow P_{t|t} = P_{t|t-1} - K_t H_t P_{t|t-1} \]

\[ K = \frac{H\sigma_x^2}{H^2\sigma_x^2 + \sigma_z^2} \]

\[ \Rightarrow K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1} \]

- Note that they are the Kalman filter equations (Eq. 5–Eq. 7).
In a Gaussian framework, the Kalman filter’s output is the optimal linear estimate:

$$\hat{x}_{t|t} = \mathbb{E}\{x_t|z_0, z_1, \ldots, z_t\}$$

$$= \mu_{f,t} \text{ in the 1-D example}$$

Given the measurement up to time $t$, the covariance of prediction error is

$$P_{t|t} = \mathbb{E}\{(\hat{x}_{t|t} - x_t)(\hat{x}_{t|t} - x_t)^T|z_0, z_1, \ldots, z_t\}$$

$$= \sigma^2_{f,t} \text{ in the 1-D example}$$