Kalman Filter

Daqing Yi

1 Introduction

The Kalman filter can be applied to the systems that have two properties.

- The system is linear;
- and the additive noises are white Gaussian noise.

As a Bayes filter, Kalman filter implements the update of the belief of the state by the observations. In Kalman filter, the belief of the state is by the moments, the mean μ_k and the covariance Σ_k.

1.1 Linear system

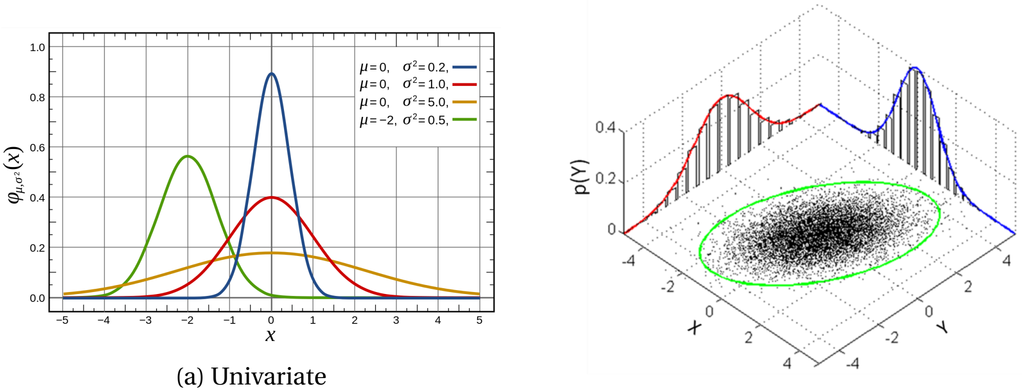
Linear system is a mathematical model to represent the system dynamics. The state of the system is presented as a vector. The operations on the state vectors are linear. Equation (1) gives an example.

$$Y = AX + BU + C \tag{1}$$

Linear abstraction is usually ideal but has significant properties and is simple in calculation.

1.2 Gaussian noise

Gaussian distribution is the most popular distribution in modeling the uncertainty in the system. By the central limit theorem, any sum or average of samples from ANY distribution (with finite mean and standard deviation) will be approximately Gaussian with the approximation better for larger samples. "Noise" or any measurement can be considered as made up of many smaller parts so that we can always assume an arbitrarily close approximation



(a) Univariate

(b) Multivariate

Figure 1: Normal distribution

to Gaussian. Gaussian distribution has some nice properties that supports the validity of the Kalman filter.

2 Properties of Gaussian distribution

We have the probability density function of univariate Gaussian distribution $N(\mu, \sigma)$ as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{2}$$

μ is the mean and σ is the variance. Figure 1a shows an illustration of the PDF, in which different mean and variance parameters are given. Similarly, the probability density function of multivariate Gaussian distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (D-dimension) as

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{3}$$

$\boldsymbol{\mu}$ is the mean, which is a D-dimension vector. $\boldsymbol{\Sigma}$ is the covariance, which is a $D \times D$ matrix. Figure 1b gives how the samples of the multivariate normal distribution look like. Notice that the center of the green ellipse is determined by the mean $\boldsymbol{\mu}$, while the shape and the size of the ellipse is determined by the covariance $\boldsymbol{\Sigma}$. If we have a vector of two random variable $[x, y]^T$, we can have the covariance matrix $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \begin{bmatrix} cov(x,x) & cov(x,y) \\ cov(y,x) & cov(y,y) \end{bmatrix}. \tag{4}$$

By definition, we know that $cov(x,x) = var(x)$ and $cov(y,y) = var(y)$.

2.1 Linear transformation of Gaussian

Let X be a $D \times 1$ multivariate normal random vector with mean μ and covariance matrix Σ. There is a linear transformation

$$Y = A + BX, \tag{5}$$

in which A is an $L \times 1$ real vector and B is an $L \times D$ full-rank real matrix. Then, Y is an $L \times 1$ multivariate normal distribution with mean

$$E[Y] = A + B\mu \tag{6}$$

and covariance matrix

$$cov[Y] = B\Sigma B^T. \tag{7}$$

2.2 Conditional Gaussian

Let x_1 and x_2 be two random variables and

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}). \tag{8}$$

we can have the conditional probability as a Gaussian distribution $N(\mu_{x_1|x_2}, \sigma_{x_1|x_2})$, in which

$$\mu_{x_1|x_2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \tag{9}$$

$$\sigma_{x_1|x_2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \tag{10}$$

2.3 Other preliminaries

- Matrix identity

$$(A + BD^{-1}C)^{-1} = A^{-1} - A{-}1B(D + CA^{-1}B)^{-1}CA^{-1} \tag{11}$$

$$(A + BD^{-1}C)^{-1}BD^{-1} = A^{-1}B(D + CA^{-1}B)^{-1} \tag{12}$$

- For two independent random variables X and Y, we have $var(X+Y) = var(X)+var(Y)$.

3 Kalman filter

Now we are going to estimate the state of a system with linear transition model and linear observation model with additive white Gaussian noises.

$$x_k = A_k x_{k-1} + B_k u_k + \omega_k \tag{13}$$

$$z_k = H_k x_k + v_k \tag{14}$$

- The transition noise is $\omega_k \sim N(0, R_k)$.
- The observation noise is $v_k \sim N(0, Q_k)$.

The Kalman filter works in a recursive manner. As a Bayesian filter, there are two steps, *prediction* and *update*, in each iteration. In prediction step, $\bar{x}_k$, priori state estimate, is generated. In update step, $\hat{x}_k$, posteriori state estimate, is generated. Because the estimation is also Gaussian distribution, the Kalman filter tracks the estimated mean and covariance of the states in order to track the distribution estimation. We write

- $\bar{\mu}_k = E(\bar{x}_k)$ as the mean of the priori state estimate,
- $\bar{\Sigma}_k = cov(\bar{x}_k)$ as the covariance of the priori state estimate,
- $\hat{\mu}_k = E(\hat{x}_k)$ as the mean of the posteriori state estimate and
- $\hat{\Sigma}_k = cov(\hat{x}_k)$ as the covariance of the posteriori state estimate.

Thus, we can have the process of Kalman filter at iteration k as

- *Prediction*

$$\bar{\mu}_k = A_k\hat{\mu}_{k-1} + B_k u_k \tag{15}$$

$$\bar{\Sigma}_k = A\hat{\Sigma}_{k-1}A^T + R_k \tag{16}$$

- *Update*

$$K_k = \bar{\Sigma}_k H_k^T (H_k\bar{\Sigma}_k H_k^T + Q_k)^{-1} \tag{17}$$

$$\hat{\mu}_k = \bar{\mu}_k + K_k(z_k - H_k\bar{\mu}_k) \tag{18}$$

$$\hat{\Sigma}_k = (I - K_k H_k)\bar{\Sigma}_k \tag{19}$$

4 Derivation

The derivation of the Kalman filter relies on how $\bar{\mu}_k$, $\bar{\Sigma}_k$, K_k, $\hat{\mu}_k$ and $\hat{\Sigma}_k$ are obtained.

4.1 Mean of priori state estimate

The mean of the priori state estimate $\bar{x}_k$ is $\bar{\mu}_k$. It can be written as

$$\begin{aligned}\bar{\mu}_k =& E(x_k \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k)\\ =& E(A_k X_{k-1} + B_k u_k + \omega_k \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k)\\ =& E(A_k X_{k-1} \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k) + E(B_k u_k \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k)\\ &+ E(\omega_k \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k)\end{aligned} \tag{20}$$

Because $B_k u_k$ has no uncertainty, $E(B_k u_k \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k) = B_k u_k$. Because ω_k is white Gaussian noise and is independent with $z_1, \cdots, z_k$, $E(\omega_k \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k) = 0$. Equation (20) can be written as

$$\begin{aligned}\bar{\mu}_k =& E(x_k \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k)\\ =& E(A_k X_{k-1} \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k) + B_k u_k\\ =& A_k E(X_{k-1} \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k) + B_k u_k\end{aligned} \tag{21}$$

Notice that X_{k-1} is independent with u_k, we can have $E(X_{k-1} \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_k) = E(X_{k-1} \mid z_1, \cdots, z_{k-1}, u_1, \cdots, u_{k-1})$, which is the mean of the posteriori state estimate at iteration $k-1$. Thus, we have $\bar{\mu}_k = A_k\hat{\mu}_{k-1} + B_k u_k$, which is Equation (15).

4.2 Coariance of priori state estimate

The covariance of the priori state estimate $\bar{x}_k$ is $\bar{\Sigma}_k$. It can be written as

$$\begin{aligned}\bar{\Sigma}_k =&cov(x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\ =&cov(A_kX_{k-1}+B_ku_k+\omega_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\ =&cov(A_kX_{k-1} \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)+cov(B_ku_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\ &+cov(\omega_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\end{aligned} \tag{22}$$

Because B_ku_k has no uncertainty, $cov(B_ku_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)=0$. Because ω_k is white Gaussian noise and is independent with $z_1,\cdots,z_k$, $cov(\omega_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)=R_k$. Equation (22) can be written as

$$\begin{aligned}\bar{\Sigma}_k =&cov(x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\ =&cov(A_kX_{k-1} \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)+R_k\\ =&A_kcov(X_{k-1} \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)A_k^T+B_ku_k\end{aligned} \tag{23}$$

Notice that X_{k-1} is independent with u_k, we can have $cov(X_{k-1} \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)=cov(X_{k-1} \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_{k-1})$, which is the covariance of the posteriori state estimate at iteration $k-1$. Thus, we have $\bar{\Sigma}_k=A_k\hat{\Sigma}_{k-1}A_k^T+R_k$, which is Equation (16).

4.3 Mean and covariance of posteriori state estimate

Construct a vector that contains the state and the observation at iteration k, which is $[x_k,z_k]^T$. The mean of the vector $[x_k,z_k]^T$ is

$$E\left(\begin{bmatrix}x_k\\ z_k\end{bmatrix} \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k\right)=\begin{bmatrix}E(x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\ E(z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\end{bmatrix}. \tag{24}$$

The covariance of the vector $[x_k,z_k]^T$ is

$$\begin{aligned}&cov\left(\begin{bmatrix}x_k\\ z_k\end{bmatrix} \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k\right)\\ =&\begin{bmatrix}cov(x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) & cov(x_k,z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\ cov(z_k,x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) & cov(z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\end{bmatrix}.\end{aligned} \tag{25}$$

We then need to calculate each element of the mean and the covariance of the vector.

$$\begin{aligned}&E(z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\ =&E(H_kx_k+v_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\ =&E(H_kx_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)+E(v_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\ =&H_kE(x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\ =&H_k\bar{\mu}_k.\end{aligned} \tag{26}$$

$$
\begin{aligned}
&cov(z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\
=&cov(H_k x_k + v_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\
=&cov(H_k x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) + cov(v_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\
=&H_k cov(x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) H_k^T + Q_k\\
=&H_k \bar{\Sigma}_k H_k^T + Q_k
\end{aligned} \tag{27}
$$

We have $cov(x_k, z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) = cov(z_k, x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)$.

$$
\begin{aligned}
&cov(x_k, z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\
=&E((x_k-\bar{x}_k)(z_k - H_k\bar{x}_k)^T \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\
=&E((x_k-\bar{x}_k)(H_k x_k + v(k) - H_k\bar{x}_k)^T \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\
=&E((x_k-\bar{x}_k)(H_k(x_k-\bar{x}_k))^T \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) + E((x_k-\bar{x}_k)v(k) \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)
\end{aligned} \tag{28}
$$

Because $v(k)$ and $x_k - \bar{x}_k$ are independent, we have $E((x_k-\bar{x}_k)v(k) \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) = 0$. Equation (28) is written into

$$
\begin{aligned}
&cov(x_k, z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\
=&E((x_k-\bar{x}_k)(x_k-\bar{x}_k)^T H_k^T \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k)\\
=&E((x_k-\bar{x}_k)(x_k-\bar{x}_k)^T \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) H_k^T\\
=&\bar{\Sigma}_k H_k^T
\end{aligned} \tag{29}
$$

Let

$$
E\left(\begin{bmatrix} x_k \\ z_k \end{bmatrix} \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k\right) = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}. \tag{30}
$$

and

$$
\begin{aligned}
&cov\left(\begin{bmatrix} x_k \\ z_k \end{bmatrix} \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k\right)\\
=&\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}
\end{aligned} \tag{31}
$$

, we can have

- $\mu_1 = E(x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) = \bar{\mu}_k$,
- $\mu_2 = E(z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) = H_k\bar{\mu}_k$,
- $\Sigma_{11} = cov(x_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) = \bar{\Sigma}_k$,
- $\Sigma_{12} = \Sigma_{21} = cov(x_k, z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) = \bar{\Sigma}_k H_k^T$ and
- $\Sigma_{22} = cov(z_k \mid z_1,\cdots,z_{k-1},u_1,\cdots,u_k) = H_k\bar{\Sigma}_k H_k^T + Q_k$.

By conditional Gaussian, we can have the distribution of $x_k \mid z_k$ as

$$x_k \mid z_k, z_1, \cdots, z_{k-1}, u_1, \cdots, u_k \sim N(\mu_1 + \Sigma_{11}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}). \tag{32}$$

It means that

$$E(x_k \mid z_1, \cdots, z_k, u_1, \cdots, u_k) = \bar{\mu}_k + \bar{\Sigma}_k H_k (H_k \bar{\Sigma}_k H_k^T + Q_k)^{-1}(z_k - H_k \bar{\mu}_k) \tag{33}$$

and

$$cov(x_k \mid z_1, \cdots, z_k, u_1, \cdots, u_k) = \bar{\Sigma}_k - \bar{\Sigma}_k H_k^T (H_k \hat{\Sigma}_k H_k^T + Q_k)^{-1} H_k \bar{\Sigma}_k. \tag{34}$$

Notice that $\hat{\mu}_k = E(x_k \mid z_1, \cdots, z_k, u_1, \cdots, u_k)$ and $\hat{\Sigma}_k = cov(x_k \mid z_1, \cdots, z_k, u_1, \cdots, u_k)$. Define the Kalman gain K_k as Equation (17), we can write Equation (33) as Equation (18) and Equation (34) as Equation (19).