A Note on Kalman Filter

January 2013

last update: April 5, 2020

Kalman filter can be used to estimate model parameters in a state-space form

$$\begin{aligned} &\text{Measurement:} && \underset{n\times 1}{y_i} = \underset{n\times 1}{a_i} + \underset{n\times m}{H_i}\ \underset{m\times 1}{x_i} + \underset{n\times 1}{r_i}\,, \qquad r_i \sim N\left(0,\ \underset{n\times n}{R_i}\right) \\ &\text{State transition:} && \underset{m\times 1}{x_i} = \underset{m\times 1}{c_i} + \underset{m\times m}{F_i}\ \underset{m\times 1}{x_{i-1}} + \underset{m\times 1}{q_i}\,, \qquad q_i \sim N\left(0,\ \underset{m\times m}{Q_i}\right) \end{aligned} \tag{1}$$

where r_i and q_i are Gaussian white noise with covariance R_i and Q_i respectively. It is designed to filter out the desired true signal and the unobserved component from unwanted noises. The measurement system is observable. It describes the relationship between the observed variables y_i and the state variables x_i. The transition system is unobservable. It describes the dynamics of the state variables as formulated by vector c_i and matrix F_i. The vector r_i and q_i are innovations for measurement and transition system respectively. They are assumed to follow multivariate Gaussian distribution with zero mean and covariance matrix R_i and Q_i respectively.

The Kalman filter is a recursive estimator. This means that only the estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state. Define the mean and variance of x_i conditioning on the observed measurements $y_0, y_1, \cdots, y_h$ for $h \le i$

$$\begin{aligned} E_{x,i|h} &= \mathbb{E}_h[x_i] = \mathbb{E}[x_i | y_0, y_1, \cdots, y_h] \\ V_{x,i|h} &= \mathbb{V}\left[x_i - E_{x,i|h}\right] = \mathbb{E}\left[\left(x_i - E_{x,i|h}\right)^2\right] \end{aligned} \tag{2}$$

The procedure generally consists of four steps:

1. Initialize the state vector:

   Since we do not know anything about $E_{x,0|0}$, we will make an assumption $x_0 \sim N(\mu, \Sigma)$

$$E_{x,0|0} = \mu, \qquad V_{x,0|0} = \Sigma \tag{3}$$

2. Predict the *a priori* state vector for $h = i - 1$ and $i = 1,2,\cdots$:

$$\begin{aligned} E_{x,i|h} &= \mathbb{E}_h[x_i] = \mathbb{E}_h[c_i + F_i x_h + q_i] = c_i + F_i E_{x,h|h} \\ V_{x,i|h} &= \mathbb{V}\left[x_i - E_{x,i|h}\right] = \mathbb{V}\left[c_i + F_i x_h + q_i - c_i - F_i E_{x,h|h}\right] = F_i V_{x,h|h} F_i' + Q_i \end{aligned} \tag{4}$$

3. Forecast the measurement equation based on $E_{x,i|h}$ and $V_{x,i|h}$:

$$\begin{aligned} E_{y,i|h} &= \mathbb{E}_h[y_i] = \mathbb{E}_h[a_i + H_i x_i + r_i] = a_i + H_i E_{x,i|h} \\ V_{y,i|h} &= \mathbb{V}\left[y_i - E_{y,i|h}\right] = \mathbb{V}\left[a_i + H_i x_i + r_i - a_i - H_i E_{x,i|h}\right] = H_i V_{x,i|h} H_i' + R_i \end{aligned} \tag{5}$$

4. Update the inference to the state vector using measurement residual $z_i = y_i - E_{y,i|h}$ and Kalman gain $\underset{m\times n}{K_i}$:

$$\begin{aligned} &E_{x,i|i} = E_{x,i|h} + K_i z_i \quad \text{and} \\ &V_{x,i|i} = \mathbb{V}\left[x_i - E_{x,i|i}\right] = \mathbb{V}\left[x_i - E_{x,i|h} - K_i\left(y_i - E_{y,i|h}\right)\right] = \mathbb{V}\left[(I - K_i H_i)\left(x_i - E_{x,i|h}\right) - K_i r_i\right] \\ &= (I - K_i H_i)\, V_{x,i|h}(I - K_i H_i)' + K_i R_i K_i' = V_{x,i|h} - 2K_i H_i V_{x,i|h} + K_i\left(H_i V_{x,i|h} H_i' + R_i\right)K_i' \\ &= (I - 2K_i H_i) V_{x,i|h} + K_i V_{y,i|h} K_i' \end{aligned} \tag{6}$$

The error in the *a posteriori* state estimation is $x_i - E_{x,i|i}$. We want to minimize the expected value of the square of the magnitude of this vector, i.e. $\mathbb{E}_i\left[\left\|x_i - E_{x,i|i}\right\|^2\right]$. This is equivalent to minimizing the trace of the *a posteriori* estimate covariance matrix $V_{x,i|i}$. By setting its first derivative to zero, we can derive the optimal Kalman gain $\widehat{K}_i$

$$\frac{\partial \mathrm{tr}\left(V_{x,i|i}\right)}{\partial K_i} = -2V_{x,i|h} H_i' + 2K_i V_{y,i|h} = 0 \quad \Longrightarrow \quad \widehat{K}_i = V_{x,i|h} H_i' V_{y,i|h}^{-1} \tag{7}$$

Calculation of $V_{y,i|h}^{-1}$ involves matrix inverse, however if the R_i^{-1} is available and the $V_{x,i|h}$ has a much smaller dimension than $V_{y,i|h}$, the $V_{y,i|h}^{-1}$ can be calculated in a more efficient way using *Sherman-Morrison-Woodbury* formula. Given the optimal $\widehat{K}_i$ in (7), the $V_{x,i|i}$ can be further simplified to

$$V_{x,i|i} = \left(I - 2\widehat{K}_i H_i\right) V_{x,i|h} + \widehat{K}_i V_{y,i|h} \widehat{K}_i' = \left(I - 2\widehat{K}_i H_i\right) V_{x,i|h} + V_{x,i|h} H_i' \widehat{K}_i' = \left(I - \widehat{K}_i H_i\right) V_{x,i|h} \tag{8}$$

We recursively generate the residual z_i and its covariance $V_{y,i|h}$ by stepping through the above procedure for $i = 1, \cdots, N$. The model parameters are then estimated through Maximum Likelihood Estimation (MLE) by maximizing the log likelihood function of the z_i time series:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{N} \log\left[(2\pi)^{-\frac{n}{2}}\left|V_{y,i|h}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} z_i' V_{y,i|h}^{-1} z_i\right)\right] \\ &= -\frac{nN\log(2\pi)}{2} + \frac{1}{2}\sum_{i=1}^{N}\left(-\log\left|V_{y,i|h}\right| - z_i' V_{y,i|h}^{-1} z_i\right) \end{aligned} \tag{9}$$

We can ignore the constant term and constant multiplier in front of the sum sign, hence maximizing the log likelihood function $l(\theta)$ is equivalent to maximizing the following sum:

$$\hat{l}(\theta) = \sum_{i=1}^{N}\left(-\log\left|V_{y,i|h}\right| - z_i' V_{y,i|h}^{-1} z_i\right) \tag{10}$$