Kalman Filter

Uri Shaham

October 11, 2022

1 Introduction: a simple case

Suppose we observe samples from a stationary process modeled by $z_i = \mu + \epsilon_i$, where z_i are observation, μ can be thought of (constant) system state, and ϵ_i is a measurement error, which is normally distributed with zero mean and standard deviation σ. We know that a good estimator for μ is $\hat{\mu} = \bar{z}_n = \frac{1}{k}\sum_{i=1}^{k} z_i$. When a new data point arrives, we can (and should) update our estimation. Rather than computing the average all over again, we can just update via

$$\begin{aligned}
\bar{z}_{n+1} &= \frac{1}{k+1}\sum_{i=1}^{k+1} z_i \\
&= \frac{1}{k+1}\left(\sum_{i=1}^{k} z_i + z_{k+1}\right) \\
&= \frac{1}{k+1}\left(k\frac{1}{k}\sum_{i=1}^{k} z_i + z_{k+1}\right) \\
&= \frac{k}{k+1}\bar{z}_k + \frac{1}{k+1}z_{k+1} \\
&= \bar{z}_k - \frac{1}{k+1}\bar{z}_k + \frac{1}{k+1}z_{k+1} \\
&= \bar{z}_k + \frac{1}{k+1}\left(z_{k+1} - \bar{z}_k\right)
\end{aligned}$$

This means that the new estimation is a multiple of the previous estimation, plus a factor (decaying in time) resulting from the current error. Kalman filter generalizes this idea, and is a highly used model in prediction and control.

2 The Kalman Filter

2.1 Model

The Kalman filter model is

$$x_k = F_k x_{k-1} + B_k u_k + w_k,$$

where:

- $x_k\, \mathbb{R}^d$ is the state of the system at time k, (unknown)

- F_k is a linear state transition model, applied to the previous state (known)
- u_k is optional external control input (known)
- B_k is the control-input model (known)
- w_k is the process noise, which is a sample from multivariate Gaussian with zero mean and covariance Q_k

In addition, and time k we observe a sample $z_k = H_k x_k + v_k$, where

- H_k is the state-observation model (known)
- v_k is measurement noise, drawn from a multivariate Gaussian with zero mean and covariance R_k

2.2 Example

A truck drives along a straight road, starting at position 0. Time indices k refer to Δt intervals. We want to keep track of the truck position y and velocity $\dot{y}$, i.e.,

$$x_k = \begin{bmatrix} y_k \\ \dot{y}_k \end{bmatrix}.$$

We consider constant F, Q, R, H (hence time indices are omitted, and $B = 0$, as no external inputs are involved. We set the matrices as follows. We assume that at the k'th time interval there is a constant acceleration given by a_k, which is normally distributed with zero mean and σ_a standard deviation. Then applying Newton's laws, we can write

$$x_k = F x_{k-1} + a_k G,$$

where

$$F = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix},$$

and

$$G = \begin{bmatrix} \frac{1}{2}\Delta t^2 \\ \Delta t \end{bmatrix}.$$

This means that we can write

$$x_k = F x_{k-1} + w_k,$$

with

$$w_k \sim G \cdot \mathcal{N}(0, \sigma_a).$$

At time k we measure the position of the truck, with measurement noise thus

$$z_k = H x_k + v_k,$$

where $H = (1, 0)^T$ and v_k is zero mean with variance σ_z^2.

2.3 The Prediction and update

Let $Z_k = (z_1, \dots z_k)$ denote the first k observations collectively. At time $k-1$ we want to predict the next system state based on Z_{k-1}, denoted as $\hat{x}_{k|k-1} := \mathbb{E}[x_k|Z_{k-1}]$. In addition, once the k'th observation z_k arrives, we update our model to $\hat{x}_{k|k} := \mathbb{E}\, x_k|Z_k$. We also model the covariance matrices corresponding to the random variables $x_k|Z_{-1}k$ and $x_k|Z_k$ (whose means are $\hat{x}_{k|k-1}$ and $\hat{x}_{k|k}$, respectively) by

$$P_{k|k-1} = \mathbb{E}\left[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T | Z_k\right],$$

and

$$P_{k|k} = \mathbb{E}\left[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T | Z_k\right],$$

2.4 Derivation

Recall from the previous lesson that if $\begin{bmatrix} X \\ Z \end{bmatrix}$ is a multivariate Gaussian, then $X|Z$ is also a multivariate Gaussian with mean $\mu_X + \Sigma_{XZ}\Sigma_{ZZ}^{-1}(Z - \mu_Z)$ and covariance $\Sigma_{XX} - \Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}$. Since all noises in our case are Gaussian, and all random variables are linear combinations of Gaussian random variables, it follows that $x_k|Z_{k-1} \sim \mathcal{N}\left(\hat{x}_{k|k-1}, P_{k|k-1}\right)$ and $x_k|Z_k \sim \mathcal{N}\left(\hat{x}_{k|k}, P_{k|k}\right)$.

Suppose that at some point, we have $\hat{x}_{k-1|k-1}$ and $P_{k-1|k-1}$ IN the above example, if we know that at time 0 both the position and the velocity are zero we have

$$x_{k-1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

and

$$P_{k-1|k-1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The algorithm works in two alternating steps:

Prediction: Since $x_k = F_k x_{k-1} + B_k u_k + w_k$, and since given Z_{k-1}, x_k and w_k are independent, we have

$$\begin{aligned} \hat{x}_{k|k-1} &= \mathbb{E}[x_k|Z_{k-1}] = F_k\,\mathbb{E}[x_{k-1}|Z_{k-1}] = F_k\hat{x}_{k-1|k-1} + B_k u_k \\ P_{k|k-1} &= \mathrm{Cov}(x_k|Z_{k-1}) = F_k P_{k-1|k-1} F_k^T + Q_k. \end{aligned} \tag{1}$$

Update: Since $z_k = H_k x_k + v_k$,

$$\left(\begin{pmatrix} x_k \\ z_k \end{pmatrix} \middle| Z_{k-1}\right) \sim \mathcal{N}\left(\begin{bmatrix} \hat{x}_{k|k-1} \\ H_k\hat{x}_{k|k-1} \end{bmatrix}, \begin{bmatrix} P_{k|k-1} & P_{k|k-1}H_k^T \\ H_k P_{k|k-1} & H_k P_{k|k-1} H_k^T + R_k \end{bmatrix}\right).$$

Since $\begin{bmatrix} x_k \\ Z_k \end{bmatrix}$ is Gaussian, conditioning on Z_{k-1}, z_k (that is, on Z_k), we have

$$\begin{aligned} \hat{x}_{k|k} &= \mathbb{E}[x_k|Z_k] \\ &= \mathbb{E}[x_k|Z_{k-1}] + P_{k|k-1}H_k^T\left(H_k P_{k|k-1} H_k^T + R_k\right)^{-1}\left(z_k - H_k\hat{x}_{k|k-1}\right) \\ P_{k|k} &= \mathrm{Cov}(x_k|Z_k) \\ &= P_{k|k-1} - P_{k|k-1}H_k^T\left(H_k P_{k|k-1} H_k^T + R_k\right)^{-1} H_k P_{k|k-1}. \end{aligned}$$

Remark 2.1. *The factor* $\left(z_k - H_k\hat{x}_{k|k-1}\right) = \mathbb{E}\left[z_k - \mathbb{E}[z_k|Z_{k-1}]\right.$ *is called innovation, whose covariance is* $\left(H_k P_{k|k-1} H_k^T + R_k\right)$.

Remark 2.2. *The factor* $P_{k|k-1} H_k^T \left(H_k P_{k|k-1} H_k^T + R_k\right)^{-1}$ *is called the Kalman gain, and reflects the importance of the innovation.*