Kalman Filter

Uri Shaham

January 28, 2025

1 Introduction: a simple case

Suppose we observe samples from a stationary process modeled by $z_i = \mu + \epsilon_i$, where z_i is an 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 $\hat{\mu} := \bar{z}_k := \frac{1}{k}\sum_{i=1}^{k} z_i$. 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

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

i.e., the updated estimate is the current estimate + a term in the direction of the prediction error, with a small magnitude.

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:

- $X_k \in \mathbb{R}^d$ is random variable representing 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)
- W_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.,

$$x_k = \begin{pmatrix} y_t \\ \dot{y}_t \end{pmatrix}.$$

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

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

and

$$G = \begin{pmatrix} \frac{1}{2}(\Delta t)^2 \\ \Delta t \end{pmatrix}.$$

This means that we can write

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

with

$$w_k \sim GG^T \cdot \mathcal{N}(0, \sigma_a^2)$$

i.e., w_k is a sample from a multivariate normal distribution, with zero mean and covariance $\sigma_a^2 GG^T$. At time k we measure the position of the truck

$$z_k = H x_k + v_k,$$

where $H = (1, 0)^T$ and $v_k \sim \mathcal{N}(0, \sigma^2)$ is measurement noise.

2.3 The conditional distributions

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

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

and

$$P_{k|k} = \mathbb{E}\left[(X_k - \hat{x}_{k|k})(X_k - \hat{x}_{k|k})^T|\tilde{z}_k\right].$$

Recall from the previous lesson that if $\begin{pmatrix} X \\ Z \end{pmatrix}$ 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 $X_k|\tilde{z}_{k-1} \sim \mathcal{N}\left(\hat{x}_{k|k-1}, P_{k|k-1}\right)$ and $X_k|\tilde{z}_k \sim \mathcal{N}\left(\hat{x}_{k|k}, P_{k|k}\right)$.

2.4 Prediction and update

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

$$\hat{x}_{k|k-1} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

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

The algorithm works in two alternating steps:
Prediction: Since $X_k = F_k X_{k-1} + B_k u_k + W_k$ we have

$$\begin{aligned} \hat{x}_{k|k-1} &= \mathbb{E}[X_k|\tilde{z}_{k-1}] = F_k\,\mathbb{E}[X_{k-1}|\tilde{z}_{k-1}] = F_k\hat{x}_{k-1|k-1} + B_k u_k, \\ P_{k|k-1} &= \mathrm{Cov}(X_k|\tilde{z}_{k-1}) = F_k P_{k-1|k-1} F_k^T + Q_k \end{aligned}$$

where the latter equation holds as given $\tilde{z}_{k-1}$, X_{k-1} and W_k are independent.
Update: Since $Z_k = H_k X_k + V_k$,

$$\begin{bmatrix} X_k \\ Z_k \end{bmatrix} \Big| \ \tilde{z}_{k-1} \sim \mathcal{N}\left(\begin{bmatrix} \hat{x}_{k|k-1} \\ H_k\hat{x}_{k|k-1} \end{bmatrix}, \begin{bmatrix} P_{k|k-1} & P_{k|k-1}H_k^T \\ H_k P_{k|k-1} & H_k P_{k|k-1} H_k^T + R_k \end{bmatrix}\right).$$

Since $\begin{bmatrix} X_k \\ Z_k \end{bmatrix} \Big| \ \tilde{z}_{k-1}$ is Gaussian, conditioning on Z_k (i.e., on $\tilde{z}_k$), we have

$$\begin{aligned} \hat{x}_{k|k} &= \mathbb{E}[X_k|\tilde{z}_k] \\ &= \mathbb{E}[X_k|\tilde{z}_{k-1}] + P_{k|k-1}H_k^T\left(H_k P_{k|k-1} H_k^T + R_k\right)^{-1}\left(Z_k - H_k\hat{x}_{k|k-1}\right) \end{aligned}$$

and

$$\begin{aligned} P_{k|k} &= \mathrm{Cov}[X_k|\tilde{z}_k] \\ &= P_{k|k-1} - P_{k|k-1}H_k^T\left(H_k P_{k|k-1} H_k^T + R_k\right)^{-1} H_k P_{k|k-1}. \end{aligned}$$

Remark 2.1. *The factor* $Z_k - H_k\hat{x}_{k|k-1} = z_k - \mathbb{E}[Z_k|Zk-1]$ *is called innovation, whose covariance is* $H_k P_{k|k-1} H_k^T + Rk$.

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.*