Problem Setup

We have a target state $x_k \in \mathbb{R}^n$, we want to estimate it during each timestamp $k \in \mathbb{N} \cup \{0\}$. We create some measurement method to help us guess each x_k. The Kalman Filter's objective is trying to guess a theoretically perfect x_k from each measurement z_k.

TL;DR, the Kalman Filter is the process doing: get z_1, guess x_1; get z_2, guess x_2; get z_3, guess x_3, and so on.

1 Assumptions

By means of *theoretically*, we need some assumptions.

1.1 Pre-Defined Transition

Let $F_k \in \mathbb{R}^{n\times n}$ be a matrix. We assume that the next state x_{k+1} can be obtained from the current state x_k by

$$x_{k+1} = F_k x_k + u_k \tag{1}$$

where $u_k \sim \mathcal{N}(0, Q_k)$ is a white noise.

This is also called *the linear model of Kalman Filter.*

1.2 Measurement to Real State

Let $H_k \in \mathbb{R}^{m\times n}$ be a matrix. We assume that the current measurement $z_k \in \mathbb{R}^m$ is obtained from the current state x_k by

$$z_k = H_k x_k + w_k \tag{2}$$

where $w_k \sim \mathcal{N}(0, R_k)$ is a white noise.

2 Initial Setup

We never know x_k, we can only guess it, so denote the guessed value by $\hat{x}_k$. Formally, we call $\hat{x}_k$ the *estimate.*

The zero-th step of the Kalman Filter is randomly guess $\hat{x}_0$ since there's no z_{-1}. We further concern the correctness of each $\hat{x}_k$, we use the error covariance matrix $P_k = \mathrm{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$ to formally compute it. So when we start with x_0, we'll also have another randomly guess matrix P_0.

TL;DR, the initial setup is one vector $\hat{x}_0$ and one matrix P_0.

3 Preliminary Theorems

3.1 Minimum Variance Unbiased Estimate

Consider the measurement equation (2), we wanna seek a matrix $K_k \in \mathbb{R}^{n\times m}$ that can do the guessing $\hat{x}_k = K_k z_k$. Since w_k is a random vector, z_k is also a random vector, then so does $\hat{x}_k$. As a result, the error term $\hat{x}_k - x_k$ is a random vector as well.

We can define an inner product of two random vectors be

$$(x|y) = \mathrm{E}(x^T y) = \mathrm{E}\left(\sum_{i=1}^{n} x_i y_i\right),$$

and then induces a norm by

$$\|x\| = \sqrt{\mathrm{E}(x^T x)} = \sqrt{\mathrm{E}(x_1^2 + \cdots + x_n^2)} = \left\{\mathrm{Tr}\left(\mathrm{E}(xx^T)\right)\right\}^{1/2}.$$

Now, consider the l2-norm

$$\begin{aligned}
&\|\hat{x}_k - x_k\|^2 = \mathrm{E}\left[(\hat{x}_k - x_k)^T (\hat{x}_k - x_k)\right] \\
=&\,\mathrm{E}\left[(K_k z_k - x_k)^T (K_k z_k - x_k)\right] \\
=&\,\mathrm{E}\left[(K_k (H_k x_k + w_k) - x_k)^T (K_k (H_k x_k + w_k) - x_k)\right] \\
=&\,\mathrm{E}\left[\left((K_k H_k x_k)^T + (K_k w_k)^T - x_k^T\right)(K_k (H_k x_k + w_k) - x_k)\right] \\
=&\,\mathrm{E}\left[(K_k H_k x_k - x_k)^T (K_k H_k x_k - x_k)\right] + \mathrm{E}\left[(K_k w_k)^T (K_k w_k)\right] \\
=&\|K_k H_k x_k - x_k\|^2 + \mathrm{Tr}\left(\mathrm{E}\left[K_k w_k (K_k w_k)^T\right]\right) \\
=&\|K_k H_k x_k - x_k\|^2 + \mathrm{Tr}\left(K_k\, \mathrm{E}[w_k w_k^T] K_k^T\right) \\
=&\|K_k H_k x_k - x_k\|^2 + \mathrm{Tr}\left(K_k R_k K_k^T\right).
\end{aligned}$$

The final term still depends on the actual state x_k, but we never know its value. To solve this, we add one more assumption $K_k H_k = I$. Observe that if $K_k H_k = 1$, we have

$$\mathrm{E}[\hat{x}_k] = \mathrm{E}\left[K_k z_k\right] = \mathrm{E}\left[K_k H_k x_k + K_k w_k\right] = \mathrm{E}\left[K_k H_k x_k\right] + \mathrm{E}\left[K_k w_k\right] = x_k,$$

this implies that $\hat{x}_k$ is an unbiased estimate of x_k.
Then the error term becomes quite simple when we assume $K_k H_k = 1$

$$\|\hat{x}_k - x_k\|^2 = \mathrm{E}\left[(\hat{x}_k - x_k)^T (\hat{x}_k - x_k)\right] = \mathrm{Tr}\left(K_k R_k K_k^T\right)$$

Our mission now becomes

$$\begin{aligned}
&\arg\min_{K_k} \mathrm{Tr}\left(K_k R_k K_k^T\right) \\
&\mathrm{s.\,t.}\, K_k H_k = I
\end{aligned}$$

The equivalent form is

$$\begin{aligned}
&\arg\min_{k_i} k_i R_k k_i^T \\
&\mathrm{s.\,t.}\, k_i h_j = \delta_{ij}
\end{aligned}$$

where h_j is the jth column of H_k and k_i is the ith row of K_k.

4 Main Theorem

Theorem 4.1. *The optimal estimate* $\hat{x}_{k+1}$ *and* P_{k+1} *can be generated recursively as*

$$\begin{aligned}
\hat{x}_{k+1} &= F_k\hat{x}_k + F_kP_kH_k^T\left[H_kP_kH_k^T + R_k\right]^{-1}(z_k - H_k\hat{x}_k)\\
P_{k+1} &= F_kP_k\left\{I - H_k^T\left[H_kP_kH_k^T + R_k\right]^{-1}H_kP_k\right\}F_k^T + Q_k
\end{aligned}$$

Proof. Suppose we have $\hat{x}_{k-1}$ and P_{k-1}. At k, we obtain a new measurement

$$z_k = H_kx_k + w_k$$

which gives us additional information about x_k. □