Lecture 1: Introduction to Optimal State Estimation

Today, I am going to build on univariate sensor fusion to derive a 1D Kalman Filter. Towards the end of class, I will then show how this extends to the multivariate case. On Wednesday, we will dicuss the limitations of the Kalman Filter and how we can avoid or mitigate those limitations. From the syllabus, we will be covering:

- Developing a Kalman Filter (Lecture 1, Homework, Lab)
- Difference between KF, EKF, and UKF (Lecture 2, Lab)
- Tuning a KF (Lecture 2, Lab)

Motivating Example: Pedestrian Distance Detection

- You are driving in dense fog at night coming home to Houghton. You only have one radar on your vehicle. Little do you know, someone is waiting for a ride on the highway shoulder, but they had a rough night at the Mosquito Inn and are teetering a little. How do you **combine** what we know about drunk people and the radar detection to **best** determine how far away the pedestrian is from the road?
  - How would you combine these two sources of information?
    - Linear function
  - What should be our definition of **best**?
    - Unbiased
    - Minimize Squared-Error/ Maximum *a posteriori*
    - Iterative (e.g. efficient)

Problem Setup

First, how can we express our measurements mathematically?

$$\begin{aligned} x_p &= \mu + w \\ z &= h\mu + \nu \end{aligned} \tag{1}$$

where $x_p \sim \mathcal{N}(\mu, q)$ is what we know about the **process** and $z \sim \mathcal{N}(\mu, r)$ is what we know from a **measurement** or observation. Let q be the variance of x_p and r the variance of z.

Bayes' Theorem or BLUE?

$$\begin{aligned} P(x|z) &= \frac{P(x,z)}{P(z)} \text{ Cond Prob} \\ &= \frac{P(z|x)P(x)}{P(z)} \text{ Cond Prob on numerator} \\ &= \frac{P(z|x)P(x)}{\int P(x,z)dx} \text{ Marginalization} \\ &= \frac{P(z|x)P(x)}{\int P(z|x)P(x)dx} \text{ Cond Prob again} \end{aligned}$$

The KF is typically derived using Bayes' Theorem and Bayes' Theorem is more generalizable to non-well-behaving distributions. However, in the case of Gaussians, using BLUE or Bayes' Theorem is identical! Since we only have a couple days to discuss the KF, I would rather help you achieve a more intuitive understanding of how all the pieces of the KF work together. I have found this understanding is more obvious using a BLUE derivation.

For Gaussians, minimizing the mean-square error is identical to finding the maximum a-posteriori estimate from Bayes' Theorem!

Linear ("L" of BLUE)

$$\hat{x} = lhx_p + kz$$

Remember, we are fusing our measurement z with what we think the measurement is hx_p. The variables k and l are weights. Let the error be

$$x^* - \hat{x} = e \tag{2}$$

What is our goal now? -> Find a and b!

Unbiased ("U" of BLUE)

Assume what we know is unbiased:

$$E[x^*] = E[\hat{x}] = lhE[x_p] + kE[z] = lh\mu + kh\mu$$

If we take the expected value of (1) and apply the unbiased condition, we have

$$1 = lh + kh \tag{3}$$

This is a constraint we must satisfy!

Minimum Mean Squared Error ("B" of BLUE)

The problem we are trying to solve is

$$\arg\min_{l,k} E[e^2] : lh + kh = 1$$

Remember, x^* is a constant...

$$\begin{aligned} E[e^2] &= E[(x^*)^2 - 2x^* \hat{x} + \hat{x}^2] \\ &= (x^*)^2 - 2x^* E[\hat{x}] + E[\hat{x}^2] \\ &= (x^*)^2 - x^* E[\hat{x}] \underbrace{- x^* E[\hat{x}] + E[\hat{x}^2]}_{Variance!} \end{aligned}$$

Since $E[x_f]$ is constant, we can rewrite the optimization as

$$\arg\min_{l,k} Var[\hat{x}] : lh + kh = 1$$

Currently, this is a constrained optimization problem, where we need to minimize our objective while ensuring (3) is satisfied. Is there a way to make this an unconstrained optimization problem?

$$\arg\min_{k} \left\{ (1 - kh)^2 q + k^2 r \right\}$$

Any ideas on how to find the minimum of this expression?

$$k^* = \frac{hq}{h^2 q + r}$$

Iterative or Recursive

How do we produce an estimate as we get closer to the pedestrian?
How can we use what we have just derived to produce estimates at each time step?

Let $x_p = \hat{x}[t^-]$

$$\begin{aligned} \hat{x}[t^+] &= (1 - kh)\hat{x}[t^-] + kz \\ &= \hat{x}[t^-] + k(z - h\hat{x}[t^-]) \end{aligned} \tag{4}$$

What about the variance of $\hat{x}$, p?

$$
\begin{aligned}
p[t^+] &= \left(1 - kh\right)^2 p[t^-] + \left(k\right)^2 r \\
&= \left(1 - \frac{h^2 q}{h^2 q + r}\right)^2 p[t^-] + \left(\frac{hq}{h^2 q + r}\right)^2 r \\
&= \left(\frac{r}{h^2 q + r}\right)^2 p[t^-] + \left(\frac{hq}{h^2 q + r}\right)^2 r \\
&= \frac{r^2 p[t^-] + h^2 p^2 [t^-] r}{(h^2 p[t^-] + r)^2} \\
&= \frac{r}{h^2 p[t^-] + r} p[t^-] = (1 - kh) p[t^-]
\end{aligned}
$$

KF Equations

Update Equations

One-dimensional, steady-state Kalman Filter!

Mean update:

$$
\hat{x}[t^+] = \hat{x}[t^-] + k(\underbrace{z - h\hat{x}[t^-]}_{y})
$$

Variance update:

$$
p[t^+] = (1 - kh) p[t^-]
$$

where

$$
k = \frac{hp[t^-]}{h^2 p[t^-] + r}
$$

Propagate Equations

Draw a timeline with different time indices. What happens when we go from $t^+ - 1$ to t^-?

Mean propagate

$$
\hat{x}[t^-] = \hat{x}[t^+ - 1]
$$

Variance Propagate

$$
p[t^-] = p[t^+ - 1] + q
$$

Extension to Multivariate

Let m be the number of states and n the number of observations.

Propagate (or Predict)

$$
\begin{aligned}
\hat{\mathbf{x}}[t^-] &= F\hat{\mathbf{x}}[t^+ - 1] \\
P[t^-] &= FP[t^+ - 1]F^T + Q
\end{aligned}
$$

Update

$$
\begin{aligned}
\hat{\mathbf{x}}[t^+] &= \hat{\mathbf{x}}[t^-] + K(\mathbf{z} - H\hat{\mathbf{x}}[t^-]) \\
P[t^+] &= (I - KH)P[t^-]
\end{aligned}
$$

where

$$
K = P[t^-]H^T(HP[t^-]H^T + R)^{-1}
$$

Covariance Matrix

Instead of our estimate being a univariate Gaussian random variable, it is a multivariate Gaussian random variable. One of the biggest differences is the use of covariance **matrices**.

$$
f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}(\det P)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})P^{-1}(\mathbf{x} - \hat{\mathbf{x}})^T \right)
$$

Properties of Covariance Matrices

Understanding covariance matrices is paramount because a covariance matrix reports the quality of the solution obtained by the KF. We can also use the covariance matrix as a "window" into the KF to determine how parameters need to be tuned.

1. PSD: $\mathbf{a}^T P \mathbf{a} \geq 0 \forall \mathbf{a}$
2. Symmetric

Now let's gain more of an intuitive understanding by looking at the structure:

$$
P = E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T] = \begin{bmatrix} Cov(x_1, x_1) & Cov(x_1, x_2) & \dots & Cov(x_1, x_n) \\ Cov(x_2, x_1) & & & \\ \vdots & & & \\ Cov(x_n, x_1) & Cov(x_n, x_2) & \dots & Cov(x_n, x_n) \end{bmatrix}
$$

What does a diagonal covariance matrix mean?
Which (if any) of the elements in a covariance matrix can be negative?