Lecture 2: Extensions of the Kalman Filter

This lecture will heavily cover practical knowledge regarding the Kalman Filter. We will first extend our 1D example to a multivariate filter. Then we will discuss assumptions made by the KF and different variants to mitigate those assumptions. Finally we will discuss how to tune and analyze a Kalman Filter.

Also, there was a question about the kalman gain k being adaptive, and I misspoke. The Kalman gain **is** adaptive, but does not change with respect to current measurements (i.e. can be calculated offline).

Extension to Multivariate KF

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

System Model

$$\begin{aligned}\mathbf{x}[t] &= F\mathbf{x}[t-1] + \boldsymbol{w} \\ \mathbf{z}[t] &= H\mathbf{x}[t] + \boldsymbol{\nu}\end{aligned}$$

where $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, Q)$ and $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, R)$

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

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

where

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

Can someone give me an intuitive explanation for the structure of the Kalman gain? Why does it "make sense"?

Assumptions of KF

- Markovity:

$$P(x_t|x_{t-1}) = P(x_t|x_{t-1}, x_{t-2})$$

- Linearity
- "White", zero-mean, Gaussian, additive noise

- White implies $E[x(t)x(t-\tau)] = q\delta(\tau)$
- Only aleatoric uncertainty (no epistemic uncertainty)

It's an engineering miracle the Kalman Filter works as well as it does. Any theories as to why it works so well?

Violations of Assumptions

Draw a box with two sides:

Noise/Process Kinematics	Linear	Non-Linear
Gaussian	KF	EKF/UKF
Non-Gaussian	?	Particle Filter

EKF

Show how to discretize and linearize. Example: Constant velocity unicycle model. We only observe v and θ .

$$\begin{aligned}
\dot{x} &= v\cos\theta \\
\dot{y} &= v\sin\theta \\
\dot{\theta} &= 0 \\
\dot{v} &= 0
\end{aligned}$$

First, discretize these equations using Euler integration. Then linearize! F and H now become **Jacobians** linearized about the previous estimate, which is the analogue to a first derivative for vector equations. However, the F and H used to propagated the state estimate and predict the measurement, respectively, are replaced by the nonlinear functions $f(\cdot)$ and $h(\cdot)$.

The Jacobian of the example above is

$$\frac{\partial f_i(\mathbf{x})}{\partial x_j} = F_{ij} \implies F = \begin{bmatrix} 1 & 0 & -v\Delta t\sin\theta & \Delta t\cos\theta \\ 0 & 1 & v\Delta t\cos\theta & \Delta t\sin\theta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

UKF

Since nonlinear functions of Gaussian random variables are not guaranteed to be gaussian, the EKF creates a Gaussian distribution by approximating nonlinear functions with their linear counterparts.

The UKF leverages the important observation that only mean and covariance are needed in the update and propagate equations. Instead of using linearization to simplify the process equations, the UKF uses a non-parameteric distribution over the state by using different "sigma" points $\mathbf{x}'$ such that

$$\mu = E[f(\mathbf{x})] = \sum_{i=1}^{N} w_m[i] f(\mathbf{x}'_i)$$

and

$$\Sigma = E[(f(\mathbf{x}) - \mu_x)^2] = \sum_{i=1}^{N} w_c[i](f(\mathbf{x}'_i) - \mu_x))(f(\mathbf{x}'_i) - \mu_x)^T$$

Analysis

Now that we have discussed some different ways to mitigate assumptions, to properly analyze a Kalman Filter and its variants, we need to know how well the algorithm is performing. **Covariance** matrices act like a "window" in the KF. We want to know two things:

1. Does the reported covariance match what we are actually observing?
2. Is the covariance of the state stable?

Before we answer these two questions, we need to talk about covariance matrices.

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?

Residual - Verifying the Integrity of our KF

How do we determine whether these algorithms are working properly?Do we have any "truth" available for evaluation? -> Residual

$$\mathbf{y}_t = \mathbf{z}_t - H\hat{\mathbf{x}}_t$$

We know it is Gaussian, but what is the mean, covariance, and autocorrelation? If we know what they should be, we can use the residual to tune a KF (manually or automatically).

Residual Mean

$$\begin{aligned} E[\mathbf{y}_t] &= E[\mathbf{z}_t] - HE[\hat{\mathbf{x}}_t] \\ &= HE[\hat{\mathbf{x}}_t] - HE[\hat{\mathbf{x}}_t] \\ &= 0 \end{aligned}$$

If the residual is not zero mean, something is typically wrong with H in your observation equations.

Residual Covariance

$$\begin{aligned} E[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_t - E[\mathbf{y}_t])^T] &= E[\mathbf{y}_t\mathbf{y}_t^T] \\ &= E[(\mathbf{z}_t - H\hat{\mathbf{x}}_t)(\mathbf{z}_t - H\hat{\mathbf{x}}_t)^T] \\ &= E[\mathbf{z}_t\mathbf{z}_t^T - \mathbf{z}_t\hat{\mathbf{x}}_t^T H^T - H\hat{\mathbf{x}}_t\mathbf{z}_t^T + H\hat{\mathbf{x}}_t\hat{\mathbf{x}}_t^T H^T] \\ &= Cov(\mathbf{z}_t, \mathbf{z}_t) - \cancel{E[\mathbf{z}]E[\mathbf{z}^T]} - \cancel{Cov(\mathbf{z}_t, \mathbf{x_t})H^T} + \cancel{E[\mathbf{z}_t]E[\mathbf{x}_t^T]H^T} - \\ &\cancel{HCov(\mathbf{x}_t, \mathbf{z}_t)} + \cancel{HE[\mathbf{x}_t]E[\mathbf{z}_t^T]} + \\ &HCov(\hat{\mathbf{x}}_t, \hat{\mathbf{x}}_t)H^T - \cancel{HE[\hat{\mathbf{x}}_t]E[\hat{\mathbf{x}}_t^T]H^T} \\ &= R + HPH^T = S \end{aligned}$$

If the residual's covariance does not match the predicted covariance (shown in the equation above), noise models (Q and/or R) need to be adjusted to match reality. Remember, Q is in P.

Residual Autocorrelation

If I have two vectors $\mathbf{x}$ and $\mathbf{y}$, how do I know one vector can not tell me anything about the other vector? -> Orthogonal!

$$E[(\mathbf{y}_t - E[\mathbf{y}_t])(\mathbf{y}_{t+\tau} - E[\mathbf{y}_{t+\tau}])^T] = S\delta(\tau)$$

If the Kalman Filter is using all possible information, then a previous residual should be orthogonal to the current residual. This is also known as a **white** signal. If your residual is correlated or non-white, the process equations may not be modeling all time-dependent phenomena.

Stability

We want to see how the covariance evolves over time. We start with the propagate equations:

$$P[t^-] = FP[t^+ - 1]F^T + Q$$

and then substitute the expression for $P[t^+ - 1]$:

$$P[t^-] = F(I - K[t-1]H)P[t^- - 1]F^T + Q$$

When we are looking for stability we want to find the value of P when $P[t^-] = P[t^- - 1]$. This equation is known as a discrete algebraic Ricatti equation, esentially a quadratic matric equation. These are very tricky solve analytically. Any ideas on how we determine the asymptotic value of P?