8. The Extended Kalman Filter (EKF)

- There exist many versions - we will cover one that appears often, the *hybrid* EKF

$$
\begin{aligned}
\dot{x}(t) &= q\left(x(t), u(t), v(t), t\right) && \text{continuous process model} \\
z(kT) &= h_k\left(x(kT), w(kT)\right) && \text{discrete measurement model}
\end{aligned}
$$

- We will use $[\cdot]$ notation in this lecture to denote discrete time indices for continuous time signals:

$$z[k] := z(kT), \quad T = \text{sampling time, a constant}$$

  Now we can rewrite the measurement equation:

$$z[k] = h_k\left(x[k], w[k]\right)\ .$$

- Basic idea: linearize the equations about the current operating conditions, then use the standard Kalman Filter equations.

8.1 Sensor noise and measurement update (Step 2)

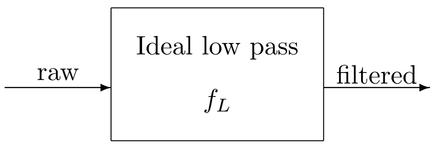
We first consider the measurement update, or what we have called Step 2 in the standard Kalman Filter, with a focus on the following two issues:

1. How to convert sensor noise specifications from continuous time to discrete time.
2. How to deal with the nonlinearity.

We consider Step 2 before Step 1 because it is conceptually easier to deal with Step 2 first.

8.1.1 Sensor noise

What are typical sensor noise specifications, and in particular, how does one convert these specifications to elements of the variance matrix R? We will explore this through a specific example. Consider a rate gyroscope, measuring angular rates in deg/s. A typical noise specification for a MEMS-based rate gyroscope, which can measure $\pm 300\,\text{deg/s}$, is $0.05\,\text{deg/s}/\sqrt{\text{Hz}}$. This is the **noise density.** Notice the units: they are the same as the measurement in question, divided by the square root of frequency.



$$z[k] = \overbrace{r[k]}^{\text{rate}} + \overbrace{w[k]}^{\text{noise}}$$

$$\text{Var}[w[k]] = (0.05)^2 \cdot f_L\ (\text{deg/s})^2$$

$$f_L : \text{cut-off frequency}$$

- To obtain the variance, you simply multiple the square of the noise density by the ideal cut-off frequency. For example, if $f_L = 100\,\text{Hz}$, then $\text{Var}[w[k]] = 0.25\,(\text{deg/s})^2$, and the standard deviation (the square root of the variance) is $= 0.5\,\text{deg/s}$. Said another way, to obtain the standard deviation, you simply multiply the noise density by the square root of the ideal cut-off frequency.

Date compiled: April 27, 2012

- As a reasonable approximation, if you do not filter the signal *before* sampling, it will have a very large variance.
- Many sensors will filter the signal for you.

- Are $\{w[1], w[2], \dots\}$ independent?
  - If sampling frequency $1/T \ll 2f_L$ (equivalently, if the Nyquist frequency is significantly smaller than the ideal cut-off frequency), this is a good assumption.
  - If, on the other hand, $1/T \gg 2f_L$, this is not a good assumption, and the noise is correlated. One must use more sophisticated Kalman Filter formulas that allow for correlated noise.
- *Trade-off:* the more aggressive the filter (small f_L), the smaller the variance, but the bandwidth of the measurement is also reduced, and a phase delay is introduced in the measurement. Furthermore, the noise becomes correlated. You may have to explicitly take these facts into account. Often a good choice is $f_L \approx 1/2T$.

8.1.2 Measurement update (Step 2)

- Using simplified notation (no time indices)

$$z = h(x, w), \quad \text{where} \quad \underbrace{\begin{array}{l} E[w] = 0, \mathrm{Var}[w] = R \\ E[x] = \hat{x}, \mathrm{Var}[x] = P \end{array}}_{\text{prior}}$$

We are interested in $\mathrm{E}[x|z]$, $\mathrm{Var}[x|z]$, the posterior.

- Let $\tilde{x} = x - \hat{x}$. If we *assume* that $\tilde{x}$ and w are small (is this a good assumption?), to first order we have

$$z = h(\hat{x}, 0) + \tilde{z},$$

where

$$\begin{aligned}
&\tilde{z} := H\tilde{x} + \tilde{w}, \\
&\tilde{w} := Mw, \;\; H := \frac{\partial h}{\partial x}(\hat{x}, 0), \;\; M := \frac{\partial h}{\partial w}(\hat{x}, 0) \\
&\mathrm{E}[\tilde{w}] = 0, \mathrm{Var}[\tilde{w}] = MRM^T \\
&\mathrm{E}[\tilde{x}] = 0, \mathrm{Var}[\tilde{x}] = P
\end{aligned}$$

- We can now compute $\mathrm{E}[\tilde{x}|\tilde{z}]$: given the measurement z, we can compute $\tilde{z} = z - h(\hat{x}, 0)$, and then use Step 2 of the standard Kalman Filter.
- Once we have computed $\mathrm{E}[\tilde{x}|\tilde{z}]$, we can readily compute the following:

$$\begin{aligned}
&\mathrm{E}[x|z] = \hat{x} + \mathrm{E}[\tilde{x}|\tilde{z}] \\
&\mathrm{Var}[x|z] = \mathrm{Var}[\tilde{x}|\tilde{z}]
\end{aligned}$$

Summary – Step 2

Let $\hat{x}_p(k)$ be the prior mean, $P_p(k)$ the prior variance. Let $\mathrm{Var}[w[k]] = R(k)$. Then

$$\begin{aligned}
K(k) &:= P_p(k)H^T(k)\left(H(k)P_p(k)H^T(k) + M(k)R(k)M^T(k)\right)^{-1} \\
\hat{x}_m(k) &:= \hat{x}_p(k) + K(k)\left(z[k] - h_k(\hat{x}_p(k), 0)\right) \\
P_m(k) &:= \left(I - K(k)H(k)\right)P_p(k)\left(I - K(k)H(k)\right)^T + K(k)M(k)R(k)M^T(k)K^T(k) \\
H(k) &:= \frac{\partial h_k}{\partial x}(\hat{x}_p(k), 0), \quad M(k) := \frac{\partial h_k}{\partial w}(\hat{x}_p(k), 0).
\end{aligned}$$

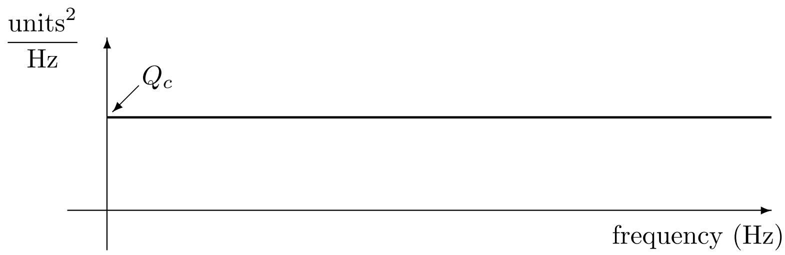
8.2 Process noise and prediction step (Step 1)

We must now deal with three issues:

1. How to directly handle continuous time process noise.
2. How to deal with a differential equation instead of a difference equation.
3. How to deal with the nonlinearity.

8.2.1 Process noise

We use the same underlying description that we used for continuous time sensor noise; instead of noise density, the power spectral density of the noise is often used, which is the square of the noise density.



- If we filtered $v(t)$ with a low pass filter at a frequency f_L, the variance, at any instant of time, would be $Q_c \cdot f_L$.
- We don't have the luxury of filtering the process noise, however. In particular, we must deal directly with continuous time white noise, which has infinite variance[1]. One can show that

$$\mathrm{E}[v(t)] = 0, \;\; \mathrm{E}[v(t)v^T(t+\tau)] = \delta(\tau)Q_c$$

8.2.2 Prediction step (Step 1)

Consider $0 \leq t < T$ for now, we will generalize this later.

$$\begin{aligned} \dot{x}(t) = q(x(t), u(t), v(t), t), \quad & \mathrm{E}[x(0)] = \hat{x}_0, \;\; \mathrm{Var}[x(0)] = P_0 \\ & \mathrm{E}[v(t)] = 0, \;\; E[v(t)v^T(t+\tau)] = \delta(\tau)Q_c \\ & v(t), \;\; x(0) \text{ are independent.} \end{aligned}$$

What are $\mathrm{E}[x(t)]$ and $\mathrm{Var}[x(t)]$?

- Let $\hat{x}(t)$ solve

$$\dot{\hat{x}}(t) = q(\hat{x}(t), u(t), 0, t), \;\; 0 \leq t < T, \;\; \hat{x}(0) = \hat{x}_0$$

  We will *assume* that $\hat{x}(t) \approx \mathrm{E}[x(t)]$. Equivalently, we assume that the expected value operator and function $q(\cdot)$ commute (this can be a really bad assumption in the case of strong nonlinearities). This approximation is exact if $q(\cdot)$ is linear.

- We next consider the variance. Let $\tilde{x}(t) = x(t) - \hat{x}(t)$. *Assuming* that $\tilde{x}(t)$ and $v(t)$ are small (is this a good assumption?)

$$\dot{\tilde{x}}(t) \approx A(t)\tilde{x}(t) + L(t)v(t)$$

$$A(t) := \frac{\partial q}{\partial x}(\hat{x}(t), u(t), 0, t) \quad L(t) := \frac{\partial q}{\partial v}(\hat{x}(t), u(t), 0, t)$$

[1] To handle this rigorously is well beyond the scope of this class, and one should only view what follows as a rough guide.

Solution for small t:

$$\tilde{x}(t) \approx \tilde{x}(0) + \int_0^t A(\tau)\tilde{x}(\tau) + L(\tau)v(\tau)\, d\tau$$
$$\approx \tilde{x}(0) + tA(0)\tilde{x}(0) + L(0)\int_0^t v(\tau)\, d\tau + O(t^2)\,,$$

Note that $P(t) := \mathrm{Var}[x(t)] \approx \mathrm{E}[\tilde{x}(t)\tilde{x}^T(t)]$, since $\tilde{x}$ and x have the same variance, and $\tilde{x}$ is assumed to be zero mean. We therefore have:

$$P(t) \approx P(0) + tA(0)P(0) + tP(0)A^T(0) + L(0)\underbrace{\left(\int_0^t\int_0^t \mathrm{E}[v(\tau)v^T(s)]\, d\tau\, ds\right)}_{tQ_c}L^T(0) + O(t^2)$$

Taking the limit as $t \to 0$:

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) + L(t)Q_cL^T(t), \quad P(0) = P_0$$

We can thus solve for $P(t)$ by solving the above matrix differential equation.

Summary – Step 1

- Let $\hat{x}_m(k{-}1)$, $P_m(k{-}1)$ be given from Step 2.
- Solve

$$\dot{\hat{x}}(t) = q(\hat{x}(t), u(t), 0, t) \qquad \begin{array}{l} (k{-}1)T \le t \le kT \\ \hat{x}((k{-}1)T) = \hat{x}_m(k{-}1) \\ u(t) \text{ is known} \end{array}$$

  Then $\hat{x}_p(k) := \hat{x}(kT)$.

- Solve

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) + L(t)Q_cL^T(t) \qquad \begin{array}{l} (k{-}1)T \le t \le kT \\ P((k{-}1)T) = P_m(k{-}1) \end{array}$$

$$A(t) = \frac{\partial q}{\partial x}(\hat{x}(t), u(t), 0, t), \quad L(t) = \frac{\partial q}{\partial v}(\hat{x}(t), u(t), 0, t)$$

  Set $P_p(k) := P(kT)$.

- Both the mean and variance calculations require solving a differential equation. One can make various approximations to reduce computation. For example:
  - If T is small, one can approximate the mean calculation as follows:

$$\hat{x}_p(k) = \hat{x}_m(k{-}1) + Tq(\hat{x}_m(k{-}1), u[k-1], 0, T(k{-}1)).$$

    More sophisticated approximations can be used. For example, instead of using $u[k-1]$, one could use $\frac{1}{2}(u[k-1]+u[k])$; instead of using $T(k{-}1)$, one could use $T(k-\frac{1}{2})$. If more accuracy is required, one could use Runge-Kutta methods.
  - A similar approach as above can be used to approximate the variance update.