Intro to Nonlinear Filters

The Kalman filter we have described this semester only works on linear systems:

$$\bar{x}_k = F_{k-1}\bar{x}_{k-1} + G_{k-1}\bar{u}_{k-1} + \bar{w}_{k-1}$$

$$\bar{z}_k = H_k\bar{x}_k + \bar{v}_k$$

→ State transition function f must be linear (a matrix F)

→ Measurement function h must be linear (a matrix H)

→ Control function g technically not required to be linear IF no process noise exists on the control input... but if it does, it must be linear (a matrix G)

→ g should be linear from a linear control systems standpoint, but not required in estimation

→ If nonlinear, control component $g(\bar{u}_{k-1})$ must still be additive

→ Noise $\bar{w}_{k-1}$ and $\bar{v}_{k-1}$ must be additive

Take a step back... what does a Kalman filter actually do?

Remember, $\bar{x}_k$ is an RV* because of process noise:

$$\bar{x}_k = F_{k-1}\bar{x}_{k-1} + G_{k-1}\bar{u}_{k-1} + \bar{w}_k$$

* Two ways to think of this. One is that $\bar{x}_k$ is deterministic and our estimate has some error w/ zero-mean and some covariance P_k. Another is that $\bar{x}_k$ is an RV whose mean and covariance we are updating in posterior as we observe measurements. Both are valid and useful in different cases.

Suppose we know $w_k \sim (0, Q_k)$ and $\bar{x}_{k-1} \sim (\hat{x}^+_{k-1}, P^+_{k-1})$

$$\rightarrow E\{\bar{x}_k\} = E\{F_{k-1}\bar{x}_{k-1} + G_{k-1}\bar{u}_{k-1} + \bar{w}_k\}$$

$$= F_{k-1}\underbrace{E\{\bar{x}_{k-1}\}}_{\hat{x}^+_{k-1}} + G_{k-1}\underbrace{E\{\bar{u}_{k-1}\}}_{\bar{u}_{k-1}} + \underbrace{E\{\bar{w}_k\}}_{0}$$

$$= \boxed{F_{k-1}\hat{x}^+_{k-1} + G_{k-1}\bar{u}_{k-1} = \hat{x}^-_k}$$

→ This is our state prediction equation! So, our state prediction is the mean of the PDF of $\bar{x}_k$ conditioned on measurements $\bar{z}_1, \dots, \bar{z}_{k-1}$

Covariance of $\bar{x}_k$:

$$\text{cov}\{\bar{x}_k\} = E\{(\bar{x}_k - E\{\bar{x}_k\})(\bar{x}_k - E\{\bar{x}_k\})^T\}$$ (we did this derivation earlier this semester)

$$= F_{k-1}\,\text{cov}\{\bar{x}_{k-1}\}\,F_{k-1}^T + \text{cov}\{\bar{w}_k\}$$

$$= \boxed{F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_k = P_k^-}$$

→ This is our covariance prediction equation! Absent our measurements, our covariance prediction is the covariance of the PDF of $\bar{x}_k$ conditioned on $\bar{z}_1, \dots, \bar{z}_{k-1}$

How about the measurement?

→ $\bar{z}_k = H_k \bar{x}_k + \bar{v}_k$ and we know $\bar{x}_k \sim (\hat{x}_k^-, P_k^-)$ and $\bar{v}_k \sim (0, R_k)$

→ $E\{\bar{z}_k\} = E\{H_k \bar{x}_k + \bar{v}_k\} = H_k E\{\bar{x}_k\} + E\{\bar{v}_k\} = \underline{H_k \hat{x}_k^-}$

→ We use this in our measurement update equation! This is the mean of the PDF of $\bar{z}_k$ conditioned on $\bar{z}_1, \dots, \bar{z}_{k-1}$

Covariance of $\bar{z}_k$:

$$\text{cov}\{\bar{z}_k\} = E\{(\bar{z}_k - E\{\bar{z}_k\})(\bar{z}_k - E\{\bar{z}_k\})^T\}$$ (skip some steps...)

$$= H_k\,\text{cov}\{\bar{x}_k\}H_k^T + \text{cov}\{\bar{v}_k\}$$

$$= H_k P_k^- H_k^T + R_k$$

→ This shows up in the Kalman gain equation! This is the covariance of the PDF of $\bar{z}_k$ (and the PDF of the innovation $\bar{z}_k - H_k\hat{x}_k^-$) conditioned on $\bar{z}_1, \dots, \bar{z}_{k-1}$

Here's the point: One interpretation of the KF is that it is keeping track of the first two moments (mean and covariance) of the PDF of $\bar{x}_k$, and along the way also uses the first two moments of the PDF of $\bar{z}_k$.

→ If the PDF is Gaussian, it is fully defined by mean and covariance
  → Also, the mean value is the maximum likelihood estimate (the peak of the PDF)
→ If the PDF is not Gaussian, the KF will still track true mean & covariance
→ The KF is an attempt to track the PDF of $\bar{x}_k$ and extract a useful estimate from it (namely, $E\{\bar{x}_k\}$) and to exploit the information in $\bar{z}_k$ to minimize the covariance of the posterior PDF
→ Linear systems/transforms are nice because the mean and covariance can be computed in closed form

Unfortunately, many (most) practical systems are not linear, and must be written in terms of arbitrary nonlinear functions:

$$\bar{x}_k = f(\bar{x}_{k-1}, \bar{u}_{k-1}, \bar{w}_{k-1})$$
$$\bar{z}_k = h(\bar{x}_k, \bar{v}_k)$$

(Typically still assume $\bar{w}_k \sim (0, Q_k)$ and $\bar{v}_k \sim (0, R_k)$)

Suppose we have a state estimate $\hat{x}_{k-1}^+$ and want to determine $\hat{x}_k^-$ as well as $\hat{z}_{k-1}$. If noise is zero-mean, this is simple enough:

$$\hat{x}_k^- = f(\hat{x}_{k-1}^+, \bar{u}_{k-1}, 0)$$
$$\hat{z}_{k-1} = h(\hat{x}_{k-1}^+, 0)$$

→ We can pass our estimates just fine through the nonlinearities.
  → $\hat{x}_k^-$ and $\hat{z}_{k-1}$ are fine as estimates, but they are not necessarily the mean values of the PDF of $\bar{x}_k$ and $\bar{z}_k$!

Big problem: there is no closed-form expression for the covariance of $\bar{x}_k$ and $\bar{z}_k$ when $f(\cdot)$ and $h(\cdot)$ are nonlinear!

→ We can develop various nonlinear filters which attempt to approximate the PDF of $\bar{x}_k$ to form estimates of $\bar{x}_k$.

We will cover three different nonlinear filters:

- The Extended Kalman Filter (EKF)
  - → The EKF approximates the nonlinear $f(\cdot)$ and $h(\cdot)$ functions by linearization (i.e., Taylor series expansion) and then attempts to track the mean and covariance through the linearized transformations as if the system/measurement models are linear.
- The Unscented Kalman Filter (UKF)
  - → The UKF approximates the mean and covariance of $\bar{X}_{k}$ through the selection of a small set of samples called "sigma points." These points are propagated exactly through the nonlinear functions, and the new mean and covariance at the outputs are computed from the sigma points' sample mean and covariance.
- The Particle Filter (PF)
  - → The PF runs a Monte Carlo simulation of the system state in an attempt to track the entire PDF of the state. The PF is robust to strong nonlinearities as well as different kinds of noise, such as bimodal noise, but has a heavy computational requirement.

The Extended Kalman Filter (Ch. 13)

Remember that (almost) any nonlinear function may be expressed by a Taylor series expansion:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

→ This is true for any value of a at which $f(x)$ is infinitely differentiable

→ It is often useful to truncate the series to get a "nice" polynomial approximation of $f(x)$, but the approximation is only "good" for values of x close to a

→ For the EKF, we will truncate the Taylor expansion of $f(\bar{x})$ to only the constant and linear terms (a first-order expansion)

→ Partially for ease.... $f(\bar{x})$ is an $N \times 1$ vector output with an $N \times 1$ vector input, so Taylor series uses partial derivatives. 1st partial derivative of $f(\bar{x})$ is the $N \times N$ Jacobian matrix, but higher-order derivatives require tensors, etc.

→ Plus, a first-order model will give a linear approximation that will fit perfectly into the existing form of the linear KF!

Suppose our system is described by:

$$\bar{x}_k = f_{k-1}(\bar{x}_{k-1}, \bar{u}_{k-1}, \bar{w}_{k-1}) \quad \text{w/} \quad \bar{w}_k \sim (0, Q_k)$$

$$\bar{z}_k = h_k(\bar{x}_k, \bar{v}_k) \quad \text{w/} \quad \bar{v}_k \sim (0, R_k)$$

We can perform a first-order Taylor expansion of f_{k-1} w.r.t. $\bar{x}_{k-1}$, $\bar{u}_{k-1}$, and $\bar{w}_{k-1}$ to put together an approximate model.

→ We don't actually need to linearize around $\bar{u}_{k-1}$, because $\bar{u}_{k-1}$ is assumed to be known (and if it isn't, we push the error into $\bar{w}_{k-1}$), and we don't actually need G_{k-1} anyway

→ We DO need to linearize around $\bar{x}_{k-1}$ and $\bar{w}_{k-1}$. What "a" values do we pick to linearize around?

→ Pick $E\{\bar{x}_{k-1}\} = \hat{x}^+_{k-1}$ and $E\{\bar{w}_{k-1}\} = 0$

$$\rightarrow \bar{x}_k \approx f_{k-1}(\hat{x}^+_{k-1}, \bar{u}_{k-1}, 0) + \underbrace{\left.\frac{\partial f_{k-1}}{\partial \bar{x}_{k-1}}\right|_{\substack{\bar{x}_{k-1}=\hat{x}^+_{k-1} \\ \bar{w}_{k-1}=0}}}_{=F_{k-1}} (\bar{x}_{k-1} - \hat{x}^+_{k-1}) + \underbrace{\left.\frac{\partial f_{k-1}}{\partial \bar{w}_{k-1}}\right|_{\substack{\bar{x}_{k-1}=\hat{x}^+_{k-1} \\ \bar{w}_{k-1}=0}}}_{=L_{k-1}} \bar{w}_{k-1}$$

$$= f_{k-1}(\hat{x}^+_{k-1}, \bar{u}_{k-1}, 0) + F_{k-1}(\bar{x}_{k-1} - \hat{x}^+_{k-1}) + L_{k-1}\bar{w}_{k-1}$$

$$= F_{k-1}\bar{x}_{k-1} + \underbrace{\left(f_{k-1}(\hat{x}^+_{k-1}, \bar{u}_{k-1}, 0) - F_{k-1}\hat{x}^+_{k-1}\right)}_{=\tilde{u}_{k-1} \rightarrow \text{Deterministic!}} + \underbrace{L_{k-1}\bar{w}_{k-1}}_{=\tilde{w}_{k-1} \sim (0,\, L_{k-1}Q_kL_{k-1}^T)}$$

$$= \boxed{F_{k-1}\bar{x}_{k-1} + \tilde{u}_{k-1} + \tilde{w}_{k-1}} \rightarrow \text{This is a perfect linear model!}$$

We can pick our prior state estimate as

$$\hat{x}^-_k = E\{\bar{x}_k\} = E\{F_{k-1}\bar{x}_{k-1} + \tilde{u}_{k-1} + \tilde{w}_{k-1}\} = F_{k-1}\hat{x}^+_{k-1} + f_{k-1}(\hat{x}^+_{k-1}, \bar{u}_{k-1}, 0) - F_{k-1}\hat{x}^+_{k-1}$$

$$= \boxed{f_{k-1}(\hat{x}^+_{k-1}, \bar{u}_{k-1}, 0)} \rightarrow \text{This state estimate is the "linearized mean"}$$

Covariance can be computed in the same way as usual:

$$\rightarrow P^-_k = F_{k-1}P^+_{k-1}F_{k-1}^T + \text{cov}\{\tilde{w}_{k-1}\} = \boxed{F_{k-1}P^+_{k-1}F_{k-1}^T + L_{k-1}Q_{k-1}L_{k-1}^T}$$

We can do the same thing to linearize the measurement model:

$$\bar{z}_k = h_k(\bar{x}_k, \bar{v}_k)$$

↳ Must linearize around both: pick $\bar{x}_k = \hat{x}_k^-$ and $\bar{v}_k = 0$

$$\rightarrow \bar{z}_k \approx h_k(\hat{x}_k^-, 0) + \underbrace{\left.\frac{\partial h_k}{\partial \bar{x}_k}\right|_{\substack{\bar{x}_k = \hat{x}_k^- \\ \bar{v}_k = 0}}}_{= H_k} (\bar{x}_k - \hat{x}_k^-) + \underbrace{\left.\frac{\partial h_k}{\partial \bar{v}_k}\right|_{\substack{\bar{x}_k = \hat{x}_k^- \\ \bar{v}_k = 0}}}_{= M_k} \bar{v}_k$$

$$= h_k(\hat{x}_k^-, 0) + H_k(\bar{x}_k - \hat{x}_k^-) + M_k \bar{v}_k$$

$$= H_k \bar{x}_k + \underbrace{\left(h_k(\hat{x}_k^-, 0) - H_k \hat{x}_k^-\right)}_{= \bar{y}_k \rightarrow \text{Deterministic!}} + \underbrace{M_k \bar{v}_k}_{= \tilde{v}_k \sim (0, M_k R_k M_k^T)}$$

$$= \boxed{H_k \bar{x}_k + \bar{y}_k + \tilde{v}_k} \rightarrow \text{Also a linear model!}$$

* We aren't used to the bias $\bar{y}_k$ in our measurement model, but just like the control input in the state model, it causes no problems since it is known deterministically!

In the innovation, we are actually computing $(\bar{z}_k - E\{\bar{z}_k\})$, which works out to $(\bar{z}_k - H_k \hat{x}_k^-)$ in the linear filter.

$\rightarrow$ But now, $E\{\bar{z}_k\} = E\{H_k \bar{x}_k + \bar{y}_k + \tilde{v}_k\} = H_k \hat{x}_k^- + \bar{y}_k$

$$= H_k \hat{x}_k^- + h_k(\hat{x}_k^-, 0) - H_k \hat{x}_k^- = \boxed{h_k(\hat{x}_k^-, 0)}$$

$\rightarrow$ So, our innovation will be $\left(\bar{z}_k - h_k(\hat{x}_k^-, 0)\right)$

At this point, we have successfully approximated both nonlinear functions w/ linear ones (i.e., matrices F_k and H_k). We can therefore compute the measurement update step using these matrices in the standard linear KF!

The Extended Kalman Filter

1. Define the system model:

$$\bar{x}_k = f_{k-1}(\bar{x}_{k-1}, \bar{u}_{k-1}, \bar{w}_{k-1}) \quad \text{w/} \quad \bar{w}_k \sim (0, Q_k)$$

$$\bar{z}_k = h_k(\bar{x}_k, \bar{v}_k) \quad \text{w/} \quad \bar{v}_k \sim (0, R_k)$$

2. Initialize the filter:

$$\hat{x}_0^+ = E\{\bar{x}_0\} \quad \text{and} \quad P_0^+ = E\{(\bar{x}_0 - \hat{x}_0^+)(\bar{x}_0 - \hat{x}_0^+)^T\}$$

3. For $k = 1, 2, \ldots$

a.) Compute the linearized system matrices (Jacobians):

$$F_{k-1} = \left.\frac{\partial f_{k-1}}{\partial \bar{x}_{k-1}}\right|_{\substack{\bar{x}_{k-1} = \hat{x}_{k-1}^+ \\ \bar{w}_k = 0}}$$

$$L_{k-1} = \left.\frac{\partial f_{k-1}}{\partial \bar{w}_{k-1}}\right|_{\substack{\bar{x}_{k-1} = \hat{x}_{k-1}^+ \\ \bar{w}_k = 0}}$$

b.) Perform the prediction/time-update step:

$$\hat{x}_k^- = f_{k-1}(\hat{x}_{k-1}^+, \bar{u}_{k-1}, 0)$$

$$P_k^- = F_{k-1} P_{k-1}^+ F_{k-1}^T + L_{k-1} Q_{k-1} L_{k-1}^T$$

c.) Compute the linearized measurement matrices (Jacobians):

$$H_k = \left.\frac{\partial h_k}{\partial \bar{x}_k}\right|_{\substack{\bar{x}_k = \hat{x}_k^- \\ \bar{v}_k = 0}}$$

$$M_k = \left.\frac{\partial h_k}{\partial \bar{v}_k}\right|_{\substack{\bar{x}_k = \hat{x}_k^- \\ \bar{v}_k = 0}}$$

d.) Perform the measurement-update step:

$$K_k = P_k^- H_k^T \left(H_k P_k^- H_k^T + M_k R_k M_k^T\right)^{-1} \rightarrow \text{Or alternate form}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k\left(\bar{z}_k - h_k(\hat{x}_k^-, 0)\right)$$

$$P_k^+ = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k M_k R_k M_k^T K_k^T \rightarrow \text{Or alternate forms}$$

Summary: The EKF lets us use KF principles on nonlinear systems, but has some issues:

1. Linearization is dependent on knowledge of $\hat{x}_{k-1}^+$ and $\hat{x}_k^-$. Poor estimates may lead to filter divergence if errors are high or f_{k-1}/h_k are too nonlinear
2. We must be careful in deriving (and coding) the Jacobian matrices!

3. Unlike the original KF for linear systems, we cannot make any comments about the optimality of the EKF (which will generally be true of any nonlinear filter).

4. Unlike the original KF, we cannot precompute anything. The Kalman gain, along with the F_k and H_k matrices, must be computed online as the state estimates are formed (also generally true for any nonlinear filter).

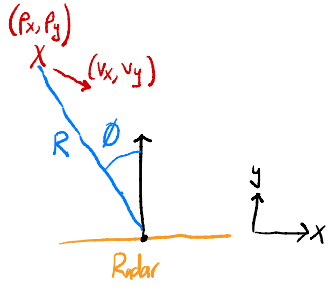
→ Despite its drawbacks, the EKF is the de facto standard for common tracking problems like navigation!

Example (related to Project 2!)

Radar array: Tracking target

→ Target moves in rectangular coordinates (x, y)

→ Radar measures in polar coordinates (R, ϕ)



$$\bar{x}_k = \begin{bmatrix} p_x \\ p_y \\ v_x \\ v_y \end{bmatrix} \rightarrow \bar{x}_k = f(\bar{x}_{k-1}, \underbrace{\bar{u}_{k-1}}_{=0}, \bar{w}_{k-1}) = \underbrace{\begin{bmatrix} p_{x,k-1} + \Delta t\, v_{x,k-1} \\ p_{y,k-1} + \Delta t\, v_{y,k-1} \\ v_{x,k-1} \\ v_{y,k-1} \end{bmatrix}}_{\bar{x}_{k-1} \text{ update}} + \underbrace{\begin{bmatrix} \frac{1}{2}\Delta t^2 \\ \frac{1}{2}\Delta t^2 \\ \Delta t \\ \Delta t \end{bmatrix}}_{\bar{w}_{k-1}} a_{n,k-1}$$

$$\rightarrow \bar{x}_k = F_{k-1}\bar{x}_{k-1} + \bar{w}_{k-1} \qquad \rightarrow F_{k-1} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

→ f is already linear, so no Jacobian necessary.

Measurements:

$$\bar{z}_k = h(\bar{x}_k, \bar{v}_k) = \begin{bmatrix} R_k \\ \phi_k \end{bmatrix} + \begin{bmatrix} v_{k,R} \\ v_{k,\phi} \end{bmatrix} \rightarrow \begin{matrix} R_k = \sqrt{p_x^2 + p_y^2} \\ \phi_k = \tan^{-1}\left(\frac{p_x}{p_y}\right) \end{matrix}$$

Not linear!

Will go to zero in partial diff



$$H_k = \left.\frac{\partial h}{\partial \bar{x}_k}\right|_{\substack{\bar{x}_k = \hat{x}_k^- \\ \bar{v}_k = 0}}$$

→ H_k will be a Jacobian matrix. The element in row i and column j is the partial derivative of the i^{th} element of $h(\bar{x})$ with respect to the j^{th} element of $\bar{x}$.

$$\rightarrow H = \begin{bmatrix} \frac{\partial R}{\partial p_x} & \frac{\partial R}{\partial p_y} & \frac{\partial R}{\partial v_x} & \frac{\partial R}{\partial v_y} \\ \frac{\partial \phi}{\partial p_x} & \frac{\partial \phi}{\partial p_y} & \frac{\partial \phi}{\partial v_x} & \frac{\partial R}{\partial v_y} \end{bmatrix}$$

$R = \sqrt{p_x^2 + p_y^2} \rightarrow R = u^{1/2}$ → Chain rule: $\frac{\partial R}{\partial p_x} = \frac{\partial R}{\partial u}\frac{\partial u}{\partial p_x}$

→ $u = p_x^2 + p_y^2$

$$\rightarrow \frac{\partial R}{\partial p_x} = \frac{1}{2}u^{-1/2} \cdot 2p_x = \boxed{\frac{p_x}{\sqrt{p_x^2 + p_y^2}}} \quad \text{and} \quad \boxed{\frac{\partial R}{\partial p_y} = \frac{p_y}{\sqrt{p_x^2 + p_y^2}}}$$

CHECK WORK WITH MATLAB OR WOLFRAM! (trust but verify!)

$$\rightarrow \boxed{\frac{\partial R}{\partial v_x} = \frac{\partial R}{\partial v_y} = 0}$$

$\rightarrow \phi = \tan^{-1}\left(\frac{p_x}{p_y}\right) \rightarrow \phi = \tan^{-1}(u)$ → Chain rule: $\frac{\partial \phi}{\partial p_x} = \frac{\partial \phi}{\partial u}\frac{\partial u}{\partial p_x}$

→ $u = p_x/p_y$

$$\rightarrow \frac{\partial \phi}{\partial p_x} = \frac{1}{u^2+1} \cdot \frac{1}{p_y} = \frac{1}{p_y\left(\frac{p_x^2}{p_y^2}+1\right)} = \boxed{\frac{p_y}{p_x^2 + p_y^2}}$$

→ Chain rule: $\frac{\partial \phi}{\partial \rho_y} = \frac{\partial \phi}{\partial u} \frac{\partial u}{\partial \rho_y}$

$$\rightarrow \frac{\partial \phi}{\partial \rho_y} = \frac{1}{u^2+1} \cdot \left(-\frac{\rho_x}{\rho_y^2}\right) = -\frac{\rho_x}{\rho_y^2\left(\frac{\rho_x^2}{\rho_y^2}+1\right)} = \boxed{-\frac{\rho_x}{\rho_x^2+\rho_y^2}}$$

$$\rightarrow \boxed{\frac{\partial \phi}{\partial v_x} = \frac{\partial \phi}{\partial v_y} = 0}$$

$$\rightarrow H_k = \left[\begin{array}{cccc} \frac{\rho_x}{\sqrt{\rho_x^2+\rho_y^2}} & \frac{\rho_y}{\sqrt{\rho_x^2+\rho_y^2}} & 0 & 0 \\ \frac{\rho_y}{\rho_x^2+\rho_y^2} & -\frac{\rho_x}{\rho_x^2+\rho_y^2} & 0 & 0 \end{array}\right] \Bigg|_{\substack{\rho_x = \hat{\rho}_{x,k}^- \\ \rho_y = \hat{\rho}_{y,k}^-}}$$