LECTURE 2

* Kalman Filter

The recursive LS formulation from last time was:

$$P_i^{-1} = P_{i-1}^{-1} + A_i^T R_i^{-1} A_i$$

$$K_i = P_i A_i^T R_i^{-1}$$

$$\hat{X}_i = \hat{X}_{i-1} + K_i \left(Y_i - A_i \hat{X}_{i-1} \right)$$

→ The fundamental assumption is that X is constant and that its estimation $\hat{X}$ must be updated as new data points arrive

Let's assume now that X can actually evolve in time:

$$X_i = F_{i-1} X_{i-1} + \varepsilon_{i-1} \qquad \text{or} \qquad X_i = F_{i-1} X_{i-1} + G_{i-1} u_{i-1} + \varepsilon_{i-1}$$

$$Y_i = A_i X_i + \eta_i \qquad\qquad Y_i = A_i X_i + \eta_i$$

$\varepsilon_i \triangleq$ noise in the model
$\eta_i \triangleq$ noise in the measurements ($\Rightarrow$ it's the "ε" in lecture 1)
$u_i \triangleq$ control vector (if any)

Let's also assume:

$\varepsilon_i \sim N(0, Q_i)$, $\eta_i \sim N(0, R_i)$ Gaussian noise

and F_i, G_i, A_i known for every i

Because of ε_i, η_i, X_i is a random variable $\Rightarrow$ Our estimation will be on the mean and we can write:

$$E(X_i) = E\left(F_{i-1} X_{i-1} + G_{i-1} u_{i-1} + \varepsilon_{i-1} \right)$$

$$= F_{i-1} E(X_{i-1}) + G_{i-1} u_{i-1}$$

$\Rightarrow$ The mean of X_i is propagated by the formula:

$$\bar{x}_i = F_{i-1}\bar{x}_{i-1} + G_{i-1}u_{i-1} \quad \text{where } \bar{x}_i \triangleq E(X_i) \qquad (1)$$

$$E\left[(X_i-\bar{x}_i)(X_i-\bar{x}_i)^T\right] =$$

$$= E\left[\left(F_{i-1}X_{i-1} + G_{i-1}u_{i-1} + \varepsilon_{i-1} - F_{i-1}\bar{x}_{i-1} - G_{i-1}u_{i-1}\right)(\dots)^T\right]$$

$$= E\left[\left(F_{i-1}(X_{i-1}-\bar{x}_{i-1}) + \varepsilon_{i-1}\right)\left(F_{i-1}(X_{i-1}-\bar{x}_{i-1}) + \varepsilon_{i-1}\right)^T\right]$$

$$= F_{i-1}\, E\left[(X_{i-1}-\bar{x}_{i-1})(X_{i-1}-\bar{x}_{i-1})^T\right] F_{i-1}^T + \underbrace{E\left[\varepsilon_{i-1}\varepsilon_{i-1}^T\right]}_{\triangleq Q_{i-1}}$$

$$+ \underbrace{E\left[\varepsilon_{i-1}\left(F_{i-1}(X_{i-1}-\bar{x}_{i-1})\right)^T\right]}_{=0} + \underbrace{E\left[(X_{i-1}-\bar{x}_{i-1})^T F_{i-1}^T \varepsilon_{i-1}\right]}_{=0}$$

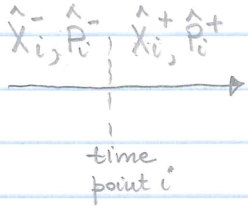
By definition, the noise is uncorrelated with the variable X_i

Because $E\left[(X_i-\bar{x}_i)(X_i-\bar{x}_i)^T\right] = P_i$ - covariance of the estimation,

we can write:

$$P_i = F_{i-1}P_{i-1}F_{i-1}^T + Q_{i-1} \qquad (2)$$

Note: In eq. (1) and (2), we have just used the evolution model of X_i, not the estimations $\hat{X}_i$, $\hat{P}_i$ $\Rightarrow$ when it comes about estimation and we are at time i, two options are viable:



a) We use data collected up to i, EXCLUDED the data collected at time i $\Rightarrow$ Let's use the notation: $\hat{X}_i^-$, $\hat{P}_i^-$

b) We use data collected up to i, INCLUDED the data collected at time i $\Rightarrow$ Let's use the notation: $\hat{X}_i^+$, $\hat{P}_i^+$

Therefore, we can have a two-tiered approach:

- If we are waiting for the data to be collected at time i, we can enhance our estimation from time $i-1$ by using the model:

$$(*) \quad \begin{aligned} \hat{X}_i^- &= F_{i-1}\hat{X}_{i-1}^+ + G_{i-1}u_{i-1} \\ \hat{P}_i^- &= F_{i-1}\hat{P}_{i-1}^+ F_{i-1}^T + Q_{i-1} \end{aligned}$$

- If we have collected the data at time i, we can refine the estimation by using the recursive Least square formula:

$$\hat{P}_i^+ = \left[(\hat{P}_i^-)^{-1} + A_i^T R_i^{-1} A_i \right]^{-1}$$

$$(**) \quad K_i = \hat{P}_i^+ A_i^T R_i^{-1}$$

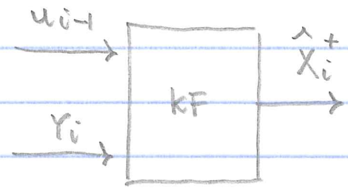
$$\hat{X}_i^+ = \hat{X}_i^- + K_i (Y_i - A_i \hat{X}_i^-)$$

$(*)$ is a prediction step and $(**)$ is a correction step $\Rightarrow$ The combination is the Kalman filter

Note a few things:

$$\hat{X}_i^+ = F_{i-1}\hat{X}_{i-1}^+ + G_{i-1}u_{i-1} + K_iY_i - K_iA_iF_{i-1}\hat{X}_{i-i}^+ - K_iA_iG_{i-1}u_{i-1}$$

$$= \left(F_{i-1} - K_i A_i F_{i-1}\right)\hat{X}_{i-1}^{+} + \left(G_{i-1} - K_i A_i G_{i-1}\right) u_{i-1} + K_i Y_i$$



$\Rightarrow$ KF is actually a time-varying filter that operates on the input variables u and Y

In (*), we provide an estimation for X in the absence of new data $\Rightarrow$ If no further data were provided, our estimation would evolve simply based on the model $\Rightarrow$ Uncertainty would grow in time because of the term Q_{i-1} $\Rightarrow$ Estimation worsens between measures

In (**), a large $\hat{P}_i^-$ is overshadowed by $A_i^T R_i^{-1} A_i$, since it contributes as $(\hat{P}_i^-)^{-1}$ $\Rightarrow$ A small uncertainty on the measurements (i.e., R_i small) will likely make $\hat{P}_i^+$ small, i.e.:

$$P_i^+ \cong \left(A_i^T R_i^{-1} A_i\right)^{-1}$$

However, in $K_i = P_i^+ A_i^T R_i^{-1}$, a small P_i^+ is bad for updating the estimation $\hat{X}_i^+$ $\Rightarrow$ That's why the presence of R_i^{-1} compensates for P_i^+ and makes K_i sizable enough to use Y_i in the update of $\hat{X}_i^+$

Moreover, note: R_i - large $\Rightarrow$ $A_i^T R_i^{-1} A_i$ - small $\Rightarrow$ $P_i^+ \cong P_i^-$ and K_i - small $\Rightarrow$ The uncertainty of the estimation is let evolve as in the absence of new measuments and new measurements are trusted very little when it comes about updating the estomate of X

Finally note: while a small R_i is generally useful (i.e., we can trust the measurements), a small Q_i may eventually lead to undesired conditions:

Q_i - small $\Rightarrow$ P_i^- small $\Rightarrow$ $P_i^+ \cong P_i^-$ small $\Rightarrow$ K_i small, i.e., if we trust the model too much, the measurements will be eventually ignored. □

* Kalman Filter for Nonlinear Models

$$\left.\begin{aligned} X_i &= F_{i-1} X_{i-1} + G_{i-1} u_{i-1} + \varepsilon_{i-1} \\ Y_i &= A_i X_i + \eta_i \end{aligned}\right\}$$

If the model is accurate, then the error $(Y_i - A_i \hat{X}_i^-) \to$ Gaussian RV

$\Downarrow$

The KF is able to extract all the available information buried in noisy data

What if the model is <u>nonlinear</u>?

$$X_i = f_i(X_{i-1}, u_{i-1}) + \varepsilon_{i-1} \qquad \varepsilon_i \sim N(0, Q_i)$$
$$Y_i = g_i(X_i) + \eta_i \qquad \eta_i \sim N(0, R_i)$$

EKF) One approach focuses on the variations of X_i from its expected value μ_i and uses the first order Taylor series expansion:

$$f_i(X_{i-1}, u_{i-1}) \cong f_i(\mu_{i-1}, u_{i-1}) + \frac{df_i}{dX}(\mu_{i-1}, u_{i-1})(X_{i-1} - \mu_{i-1})$$

$$g_i(X_i) \cong g_i(\mu_i) + \frac{dg_i}{dX}(\mu_i)(X_i - \mu_i)$$

$$\Downarrow$$

$$\mu_i = E(X_i) = E(f_i(X_{i-1}, u_{i-1})) \cong f_i(\mu_{i-1}, u_{i-1})$$

Define: $Z_i \triangleq X_i - \mu_i$; $V_i \triangleq Y_i - g_i(\mu_i)$; $F_{i-1} \triangleq \frac{df_i}{dX}(\mu_{i-1}, u_{i-1})$; $A_i \triangleq \frac{dg_i}{dX}(\mu_i)$

and consider the linearized model:

$$\begin{cases} Z_i = F_{i-1} Z_{i-1} + \varepsilon_{i-1} \\ V_i = A_i Z_i + \eta_i \end{cases} \qquad (1)$$

Now, in order to estimate X_i, one can observe:

Our estimation focuses on the expected value of X_i (so it was in the linear case) $\Rightarrow$ Denoted with $\hat{X}_{i-1}^{+}$ the best estimate of μ_{i-1} after collecting the data at time $i-1$, we can write:

$$\hat{X}_i^{-} = f_i(\hat{X}_{i-1}^{+}, u_{i-1})$$

$$F_{i-1} = \frac{df_i}{dX}(\hat{X}_{i-1}^{+}, u_{i-1})$$

Also note: $P_i \triangleq E\left((X_i-\mu_i)(X_i-\mu_i)^T\right) \underset{\substack{\uparrow \\ \text{def of } Z_i}}{=} E(Z_i Z_i^T) \Rightarrow$ We can apply the KF to (1) and estimate P_i:

$$\left.\begin{array}{l} \hat{P}_i^{-} = F_{i-1}\hat{P}_{i-1}^{+}F_{i-1}^{T} + Q_{i-1} \\ \hat{X}_i^{-} = f_i(\hat{X}_{i-1}^{+}, u_{i-1}) \end{array}\right] (E1)$$

where:

- $\hat{P}_{i-1}^{+}$ is the best estimate of P_{i-1} after collecting the data at time $i-1$
- $F_{i-1} = \frac{df_i}{dX}(\hat{X}_{i-1}^{+}, u_{i-1})$

Similarly note:

$$\hat{Z}_i^{+} = \hat{Z}_i^{-} + K_i(V_i - A_i\hat{Z}_i^{-}) \underset{\substack{\uparrow \\ Z_i \triangleq X_i - \mu_i}}{\Rightarrow} \hat{X}_i^{+} = \hat{X}_i^{-} + K_i\Big(Y_i - \underbrace{g_i(\mu_i) - A_i(\hat{X}_i^{-} - \mu_i)}_{\underset{\substack{\uparrow \\ \text{Taylor series expansion}}}{\cong} g_i(\hat{X}_i^{-})}\Big)$$

$$\Rightarrow \hat{X}_i^{+} = \hat{X}_i^{-} + K_i\left(Y_i - g_i(\hat{X}_i^{-})\right)$$

Remember: we need to update the estimate of P_i in order to compute K_i:

$$\hat{P}_i^{+} = E\left((\hat{X}_i^{+}-\mu_i)(\hat{X}_i^{+}-\mu_i)^T\right) = E\left(\hat{Z}_i^{+}\hat{Z}_i^{+T}\right) \Rightarrow \text{By using KF on (1):}$$

$$\left.\begin{aligned}
\hat{P}_i^+ &= \left[(\hat{P}_i^-)^{-1} + A_i^T R_i^{-1} A_i\right]^{-1} \\
K_i &= \hat{P}_i^+ A_i^T R_i^{-1} \\
\hat{X}_i^+ &= \hat{X}_i^- + K_i\left(Y_i - g_i(\hat{X}_i^-)\right)
\end{aligned}\right\} \quad (E2)$$

where:

- $\hat{X}_i^-$ is the best estimate of X_i before collecting the data at time i
- $A_i = \dfrac{dg_i}{dX}(\hat{X}_i^-)$

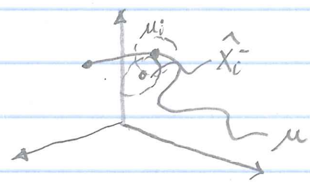
The combination (E1)–(E2) is called "Extended Kalman Filter" (EKF)

Problems:

- F_{i-1} and A_i are obtained by linearizing around the estimated $\hat{X}_{i-1}^+$ and $\hat{X}_i^-$, respectively, not the actual mean values μ_{i-1}, μ_i, respectively $\Rightarrow$ If the estimates are bad, the filter may drift away

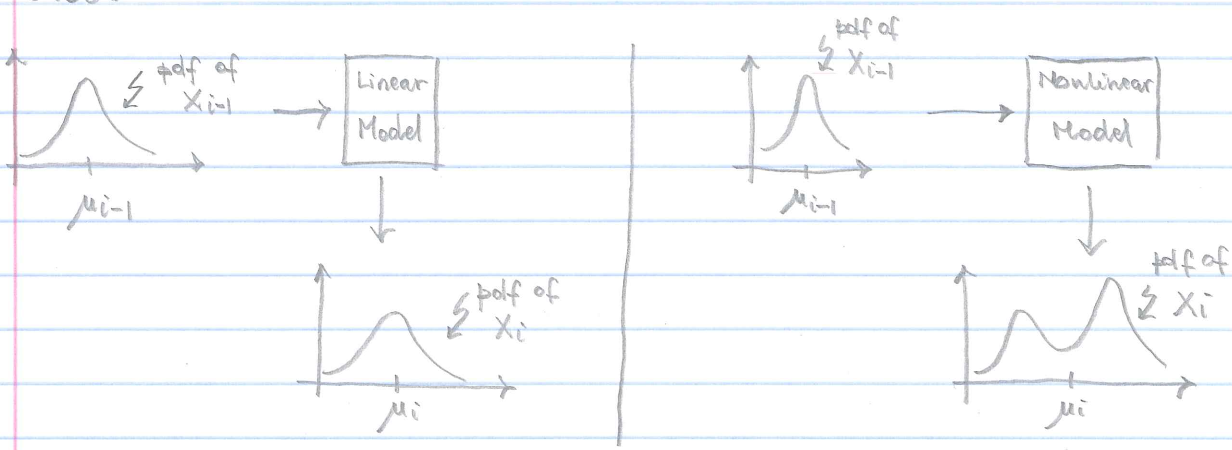
$\Downarrow$

EKF can give acceptable results as long as the nonlinearity is weak (i.e., the linearization introduces a small error)

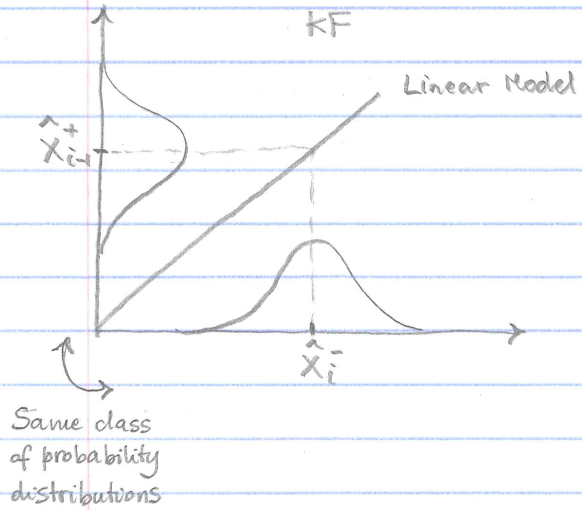


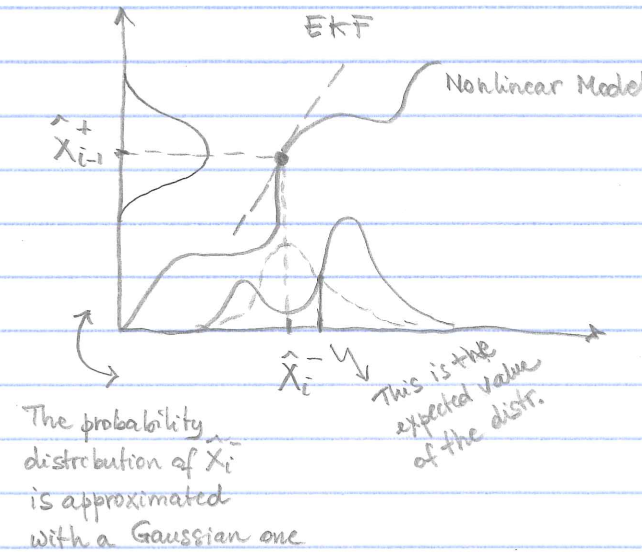
- F_{i-1} and A_i are jacobian matrices $\Rightarrow$ Their calculation can be challenging and numerical issues can rise

- By linearizing, we assume that the distribution of X propagates from time $i-1$ to time i according to a linear model $\Rightarrow$ It is not always the case:



Now, when we propagate our estimation by using KF (linear case) and EKF (nonlinear case), we have:

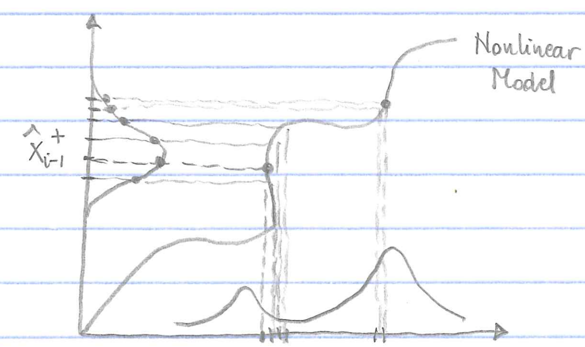




$\Downarrow$

With EKF, $\hat{X}_i^-$ may be a very bad estimation of $\mu_i \Rightarrow$ We need a different approach that aims to sample the distributions of $\hat{X}_{i-1}^+$, $\hat{X}_i^-$ rather than just assume they are Gaussian.

UKF) First, let's assume that $\hat{X}_{i-1}^+$ is Gaussian and let's propagate many values drawn from this distribution through the nonlinear model



Second, as the number of values (a.k.a. "particles") increases, a better estimation of the distribution of $\hat{X}_i^-$ can be given

$\Downarrow$

We don't need to follow a Monte Carlo approach, though. It can be proved that – if $\hat{X}_{i-1}^+$ is Gaussian and n is the size of X – we only need $n+1$ particles to estimate mean and variance of distribution determined by the nonlinear transformation of $\hat{X}^+_{i-1}$ $\Rightarrow$ We have:

$\hat{X}^+_{i-1}$ is the mean of the input distribution
$\hat{P}^+_{i-1}$ is the covariance matrix of the input distribution $\Rightarrow$ It is a $n \times n$ matrix with columns $p^+_{i-1,1}; p^+_{i-1,2}; \dots; p^+_{i-1,n}$
$\Rightarrow$ The $n+1$ points must be:

$$\sigma_0 = \hat{X}^+_{i-1}$$
$$\sigma_j = \sigma_0 + \sqrt{n\, p^+_{i-1,j}} \qquad j = 1, 2, \dots, n$$
$$\sigma_{n+j} = \sigma_0 - \sqrt{n\, p^+_{i-1,j}} \qquad j = 1, 2, \dots, n$$

$\Rightarrow$ The transformation of these $2n+1$ points is:

$$\gamma_0 = f_i(\sigma_0, u_{i-1})$$
$$\gamma_j = f_i(\sigma_j, u_{i-1}) \quad j = 1, 2, \dots, 2n$$

$\Rightarrow$ Sample mean and covariance of this distribution can be used for prediction:

$$\hat{X}_i^- = \frac{1}{2n+1} \sum_{j=0}^{2n} \gamma_j$$

$$\hat{P}_i^- = \mathrm{cov}\left(\{\gamma_j\}_{j=0,2n}\right) + Q_{i-1}$$

$$= \frac{1}{2n+1} \sum_{j=0}^{2n} \left(\gamma_j - \hat{X}_i^-\right)\left(\gamma_j - \hat{X}_i^-\right)^T + Q_{i-1}$$

Analogously, one can compute:

$$\xi_0 = g_i(\sigma_0)$$
$$\xi_j = g_i(\sigma_j) \quad j = 1, 2, \dots, 2n$$

$\Rightarrow$ One can estimate sample mean and covariance:

$$\tilde{y}_i = \frac{1}{2n+1} \sum_{j=0}^{2n} \xi_j$$

$$P_{yy} = \frac{1}{2n+1} \sum_{j=0}^{2n} \left(\xi_j - \tilde{y}_i\right)\left(\xi_j - \tilde{y}_i\right)^T + R_i$$

Note that we add Q_{i-1} and R_i to $\hat{P}_i^-$ and P_{YY}, respectively, because we want to compute an estimation of the covariances of samples of X and Y, respectively, but the samples have been chosen in a deterministic way

With $\{\gamma_j\}$ and $\{\xi_j\}$, one can also compute:

$$P_{XY} = \frac{1}{2n+1}\sum_{j=0}^{2n}\left(\gamma_j - \hat{X}_i^-\right)\left(\xi_j - \tilde{y}_i\right)^T$$

If X and Y have size n and m, respectively, this is a n×m matrix

It can be shown that the value:

$$\hat{X}_i^+ = \hat{X}_i^- + K_i\left(Y_i - \tilde{y}_i\right) \quad \text{with } K_i = P_{XY}P_{YY}^{-1}$$

is the unbiased estimation of X_i that minimizes the sum of the variances of the individual components of X_i ⇒ Therefore, the following filter is derived:

Sampling Step:

$$\sigma_0 = \hat{X}_{i-1}^+ \; ; \; \sigma_j = \hat{X}_{i-1}^+ + \sqrt{n}\left[\hat{P}_{i-1}^+\right]_j \; ; \; \sigma_{n+j} = \hat{X}_{i-1}^+ - \sqrt{n}\left[\hat{P}_{i-1}^+\right]_j \qquad j=1,2,\dots,n$$

(U1) with $\left[\hat{P}_{i-1}^+\right]_j \triangleq p_{i-1,j}^+$ j-th column of $\hat{P}_{i-1}^+$

$$\gamma_j = f_i\left(\sigma_j, u_{i-1}\right) \qquad j=0,1,\dots,2n$$

$$\xi_j = g_i\left(\sigma_j\right) \qquad j=0,1,\dots,2n$$

Prediction Step:

(U2) $$\hat{X}_i^- = \frac{1}{2n+1}\sum_{j=0}^{2n}\gamma_j \qquad \hat{P}_i^- = \frac{1}{2n+1}\sum_{j=0}^{2n}\left(\gamma_j - \hat{X}_i^-\right)\left(\gamma_j - \hat{X}_i^-\right)^T + Q_{i-1}$$

Correction Step:

(U3) $$\tilde{y}_i = \frac{1}{2n+1}\sum_{j=0}^{2n} \zeta_j \qquad P_{yy} = \frac{1}{2n+1}\sum_{j=0}^{2n} (\zeta_j - \tilde{y}_i)(\zeta_j - \tilde{y}_i)^T + R_i$$

$$P_{xy} = \frac{1}{2n+1}\sum_{j=0}^{2n} (\gamma_j - \hat{X}_i^-)(\zeta_j - \tilde{y}_i)^T$$

$$\hat{P}_i^+ = \hat{P}_i^- - P_{xy} P_{yy}^{-1} P_{xy}^T$$

$$\hat{X}_i^+ = \hat{X}_i^- + P_{xy} P_{yy}^{T} (Y_i - \tilde{y}_i)$$

The combination of steps (U1) – (U3) is the "Unscented Kalman Filter" (UKF) □

Slides present examples of application of Kalman filter to neural decoding of kinematic variables. □