ASEN 5044, Fall 2018

Statistical Estimation for Dynamical Systems

Lecture 28 [Special Topic #6]: Bayesian Derivation of the Kalman Filter

Prof. Nisar Ahmed (Nisar.Ahmed@Colorado.edu)

Friday 11/9/2018

University of Colorado Boulder

Aerospace Engineering

Overview

- Recap the general Bayes filter
- Probabilistic derivation of Kalman filter prediction step
  - **via the Chapman-Kolmogorov equation**
- Probabilistic derivation of Kalman filter measurement update step
  - **Via Bayes' rule and conditional Gaussian pdfs**
- Implications
  - KF as a special case for linear-Gaussian systems
  - Using Bayes' rule + KF to go beyond simple linear-Gaussian models

Recap: The General Bayes Filter (Lec 25)

- Recall: the general probabilistic filtering problem:

given measurements (online) $y_1, \dots, y_{k+1}$

& initial state pdf $p(x_0)$ & state transition pdf $p(x_{k+1} | x_k)$

$\rightarrow$ find: estimate of x_{k+1} as $\hat{x}^+_{k+1} = \underset{\hat{x}^+_{k+1} \in \mathbb{R}^n}{\arg\min} E[C(x_{k+1}, \hat{x}^+_{k+1})]_{p(x_{k+1} | y_{1:k+1})}$

- General solution: the recursive Bayes filter:

For $k \geq 0$, we have following recursions to get $p(x_{k+1} | y_{1:k+1})$:

I. Dynamics/Prediction step: Chapman-Kolmogorov Eq.

$$p(x_{k+1} | y_{1:k}) = \int_{-\infty}^{\infty} p(x_{k+1} | x_k) \cdot p(x_k | y_{1:k}) \, dx_k$$

repeat for next time...

II. Bayesian Meas. update step: receive new y_{k+1} data & compute

$$p(x_{k+1} | y_{1:k+1}) = \frac{p(x_{k+1} | y_{1:k}) \cdot p(y_{k+1} | x_{k+1})}{\int_{-\infty}^{\infty} p(x_{k+1} | y_{1:k}) \cdot p(y_{k+1} | x_{k+1}) \, dx_{k+1}}$$

$p(y_{k+1} | y_{1:k})$

The Bayes Filter for Linear-Gaussian Systems

- Sketch of derivation: to get the KF out of the Bayes Filter

Start w/ $p(x_k | y_{1:k}) \sim \mathcal{N}(\hat{x}_k^+, P_k^+)$ for $k \geq 0$ (where $y_{1:0} = \phi$ & $\hat{x}_0^+$ & P_0^+ are given)

& knowing $x_{k+1} = F x_k + G u_k + w_k$, $w_k \sim \mathcal{N}(0, Q)$ [AWGN]

& $y_{k+1} = H x_{k+1} + v_{k+1}$, $v_{k+1} \sim \mathcal{N}(0, R)$ ["]

I. For Pred. Step: to show Chapman-Kolmogorov eq. reduces to

$$p(x_{k+1} | y_{1:k}) = \mathcal{N}(m_{k+1}^-, P_{k+1}^-)$$

[b/c $p(x_{k+1} | x_k) = \mathcal{N}(F x_k + G u_k, Q)$ can be easily shown]

II. For Bayes update: first find $p(x_{k+1}, y_{k+1} | y_{1:k}) \sim \mathcal{N}(\cdots)$

& then use Bayes' rule for Gaussian pdfs to get $p(x_{k+1} | y_{1:k+1}) \sim \mathcal{N}(\ldots)$

[conditional Gaussian pdfs]

III. Bayesian est. of x_{k+1}: $\hat{x}_{k+1}^+ = \text{argmin}\; \mathbb{E}\left[C(x_{k+1}, \hat{x}_{k+1}^+) \right]_{p(x_{k+1} | y_{1:k+1})}$

Linear-Gaussian Dynamics Prediction Step

- At time k+1, the Chapman-Kolmogorov equation can be expressed as:

$$p(x_{k+1} \mid y_{1:k}) = \int_{-\infty}^{\infty} p(x_{k+1} \mid x_k)\, p(x_k \mid y_{1:k})\, dx_k = \int_{-\infty}^{\infty} \underbrace{p(x_{k+1}, x_k \mid y_{1:k})}_{\text{joint pdf of } x_{k+1} \,\&\, x_k \text{ given } y_{1:k}} dx_k$$

→ Because $x_{k+1} = F x_k + G u_k + w_k$, $w_k \sim N(0, Q)$, it follows:

$$p(x_{k+1}, x_k \mid y_{1:k}) \sim N\left(\begin{bmatrix} m_{k+1}^- \\ \hat{x}_k^+ \end{bmatrix}, \begin{bmatrix} C_{k+1}^- & D_{k+1,k} \\ D_{k,k+1} & P_k^+ \end{bmatrix} \right)$$

where $\underline{m_{k+1}^-} = E[x_{k+1} \mid y_{1:k}] = E[F x_{k+1} + G u_k + w_k \mid y_{1:k}] = \boxed{\hat{x}_{k+1}^- = F \hat{x}_k^+ + G u_k}$

$C_{k+1}^- = E[(x_{k+1} - m_{k+1}^-)(\dots)^T \mid y_{1:k}] = E[(x_{k+1} - \hat{x}_{k+1}^-)(\dots)^T \mid y_{1:k}] = \boxed{P_{k+1}^- = F P_k^+ F^T + Q}$

pure 1 step prediction results for KF!

$\&$ $C_{k+1,k} = C_{k,k+1}^T = E[(x_{k+1} - m_{k+1}^-)(x_k - \hat{x}_k^+)^T \mid y_{1:k}] = (\dots) = F P_k^+$

→ But from Lec 13: we know that marginal Gaussian pdf is just!

$p(x_{k+1} \mid y_{1:k}) = N_{x_{k+1}}(\hat{x}_{k+1}^-, P_{k+1}^-)$ → So C-K eq. agrees w/ 1 step pure pred. result.

Linear-Gaussian Measurement Update Step

- Start with the fact that the joint pdf for x_{k+1} and y_{k+1} is also a multivariate normal:

$$p(x_{k+1}, y_{k+1} \mid y_{1:k}) \sim \mathcal{N}\left(\begin{bmatrix} \hat{x}^-_{k+1} \\ \hat{y}^-_{k+1} \, ? \end{bmatrix}, \begin{bmatrix} P^-_{k+1} & C_{xy} \, ? \\ C_{yx} \, ? & S_{k+1} \, ? \end{bmatrix} \right)$$

$\rightarrow \hat{y}^-_{k+1} = E[y_{k+1} \mid y_{1:k}] = E[H x_{k+1} + v_{k+1} \mid y_{1:k}] = H \cdot E[x_{k+1} \mid y_{1:k}] + E[\cancel{v_{k+1}} \mid y_{1:k}]$ AWGN

$= \boxed{H \cdot \hat{x}^-_{k+1} = \hat{y}^-_{k+1}}$

$\rightarrow C_{xy} = C_{yx}^T = E[(x_{k+1} - \hat{x}^-_{k+1})(y_{k+1} - \hat{y}^-_{k+1})^T \mid y_{1:k}] = E[x_{k+1} \underline{y_{k+1}^T} \mid y_{1:k}] - \hat{x}^-_{k+1} \hat{y}^{-T}_{k+1}$

$= E[x_{k+1} (H x_{k+1} + v_{k+1})^T \mid y_{1:k}] - \hat{x}^-_{k+1} \hat{y}^{-T}_{k+1}$

$= E[x_{k+1} x_{k+1}^T \mid y_{1:k}] \cdot H^T + \underbrace{E[\cancel{x_{k+1} v_{k+1}^T} \mid y_{1:k}]}_{0 \;\; \perp\!\!\!\perp, \; v_{k+1} \sim \text{AWGN}} - \hat{x}^-_{k+1} \hat{y}^{-T}_{k+1}$

$= E[x_{k+1} x_{k+1}^T \mid y_{1:k}] \cdot H^T - \underbrace{\hat{x}^-_{k+1}}_{= E[x_{k+1} \mid y_{1:k}]} \underbrace{(\hat{x}^{-T}_{k+1} H^T)}_{\text{eq. of } \hat{y}^-_{k+1}} = \{ \underbrace{E[x_{k+1} x_{k+1}^T \mid y_{1:k}] - (E[x_{k+1} \mid y_{1:k}])(\cdots)^T}_{= P^-_{k+1}} \} H^T$

$\Rightarrow \boxed{C_{xy} = C_{yx}^T = P^-_{k+1} \cdot H^T}$

$\rightarrow$ Likewise: can show that $\boxed{\overbrace{S_{k+1} = E[(y_{k+1} - \hat{y}^-_{k+1})(\cdots)^T \mid y_{1:k}]} = H P^-_{k+1} H^T + R}$

Linear-Gaussian Measurement Update Step

- So from Lec 17, we know that $p(x_{k+1} | y_{1:k+1})$ is a conditional Gaussian pdf:

[Slides 5, 8, 9]

$$p(x_{k+1} \mid y_{k+1}, y_{1:k}) = \mathcal{N}_{x_{k+1}}\left(m_{x_{k+1}|y_{1:k+1}}, C_{x_{k+1}|y_{1:k+1}}\right) = \frac{p(x_{k+1}|y_{1:k}) \cdot p(y_{k+1}|x_{k+1})}{p(y_{k+1}|y_{1:k})}$$

where (in this case)

$$m_{x_{k+1}|y_{1:k+1}} = \underbrace{\hat{x}^-_{k+1}}_{m_{k+1}} + \boxed{\underbrace{P^-_{k+1} H^T}_{C_{xy}} \cdot \underbrace{(S_{k+1})^{-1}}_{C_{yy}^{-1}}} \cdot \underbrace{(y_{k+1} - \hat{y}^-_{k+1})}_{a - m_2}$$

KF gain!

$$C_{x_{k+1}|y_{1:k+1}} = \underbrace{P^-_{k+1}}_{C_{xx}} - \boxed{\underbrace{P^-_{k+1} H^T}_{C_{xy}} \cdot \underbrace{(S_{k+1})^{-1}}_{C_{yy}^{-1}}} \cdot \underbrace{H P^-_{k+1}}_{C_{yx}}$$

(Note: $C_{x_{k+1}, y_{k+1} | y_{1:k}} = \left[\begin{array}{c|c} \overset{P^-_{k+1}}{C_{xx}} & C_{xy} \\ \hline C_{yx} & \underset{S_{k+1}}{C_{yy}} \end{array}\right]$)

Linear-Gaussian Filtering Estimate

- Recall from Lec 20: to get the MMSE estimate, we need

$$\hat{x}_k^+ = \arg\min E\left[(x_{k+1} - \hat{x}_{k+1}^+)^T (\cdots) \mid y_{1:k+1} \right]_{p(x_{k+1} \mid y_{1:k+1})}$$

$$\Rightarrow \hat{x}_k^+ = E\left[x_{k+1} \mid y_{1:k+1} \right] = \text{cond. mean of } p(x_{k+1} \mid y_{1:k+1}) \sim \mathcal{N}(\ldots)$$

$\rightarrow$ <u>So</u>:

$$\begin{cases} \hat{x}_{k+1}^+ = \hat{x}_{k+1}^- + P_{k+1}^- H^T \cdot \left[H P_{k+1}^- H^T + R \right]^{-1} \cdot (y_{k+1} - H \hat{x}_{k+1}^-) \\ P_{k+1}^+ = P_{k+1}^- - P_{k+1}^- H^T (H P_{k+1}^- H^T + R)^{-1} H P_{k+1}^- = (I - K_{k+1} H) P_{k+1}^- \end{cases}$$

exactly the KF updates!

KF gain: K_{k+1} !

(likewise: there are also the MAP & MMAE !!)

A Closer Look at the Observation Likelihood

- The denominator from Bayes' rule $p(y_{k+1}|y_{1:k})$ is actually quite interesting for linear-Gaussian models in context of non-linear filtering
- Provides a means of comparing models or KFs that run on the same data – i.e. which one does a better job given the same $y_{1:k+1}$?
- Allows for development of **multiple hypothesis KFs**

$$p(y_{k+1}|y_{1:k}) = \mathcal{N}_{y_{k+1}}(0, S_{k+1})$$

e.g. suppose 2 models: A: $p_A(y_{k+1}|y_{1:k})$ given $(F_A, G_A, H_A, M_A, Q_A, R_A)$

B: $p_B(y_{k+1}|y_{1:k})$ given $(F_B, G_B, H_B, M_B, Q_B, R_B)$

→ Could use KF estimates based "blending" estimates from A & B:

("multiple model filter") $\hat{x}^+_k = w_A \cdot \hat{x}^{+,A}_k + (1-w_A)\cdot \hat{x}^{+,B}_k$, where $w_A = \dfrac{p_A(y_k|y_{1:k-1})}{p_A(y_k|y_{1:k-1}) + p_B(y_k|y_{1:k-1})}$