Lecture 2: State Estimation and Kalman Filter

Yilin Mo

July 2, 2015

1 Static State Estimation

Let $x \in \mathbb{R}^n$ be the states and $y \in \mathbb{R}^m$ be the sensor measurements. The a-priori pdf of x is $f(x)$ and the conditional pdf of y given x is $f(y|x)$. Hence,

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{f(x)f(y|x)}{\int_{\mathbb{R}^n} f(x)f(y|x)\,\mathrm{d}\,x}.$$

The minimum mean square error (MMSE) estimator is given by

$$\hat{x} = \mathbb{E}(x|y) = \int_{R^n} x f(x|y)\,\mathrm{d}\,x.$$

2 State Estimation of Hidden Markov Chain

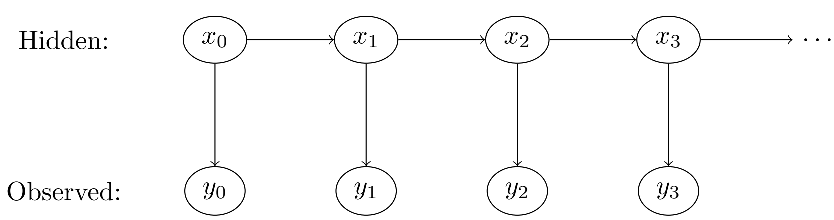

Figure 1: Hidden Markov Model

Markov Property: The joint pdf of $x_0, \dots, x_k$, $y_0, \dots, y_k$ satisfies:

$$f(x_0, \dots, x_k, y_0, \dots, y_k) = f(x_0) \prod_{i=0}^{k-1} f(x_{i+1}|x_i) \prod_{i=0}^{k} f(y_i|x_i).$$

To simplify notation, define $Y_k = (y_0, \dots, y_k)$ and $Y_{-1} = \emptyset$.

2.1 Naive Estimator

We can still use the conditional expectation to compute the MMSE estimator:

$$\hat{x}_{k|k} = \mathbb{E}(x_k|Y_k). \tag{1}$$

Drawbacks:
Estimator (1) is not recursive. We need to keep the whole history of Y_k.

2.2 Bayes Filter

Consider the conditional pdf $f(x_k|Y_k)$, by Bayes rule:

$$f(x_k|Y_k) = \frac{f(x_k, y_k|Y_{k-1})}{f(y_k|Y_{k-1})} = \frac{f(y_k|x_k, Y_{k-1})f(x_k|Y_{k-1})}{f(y_k|Y_{k-1})}$$

1. By Markov property:

$$f(y_k|x_k, Y_{k-1}) = f(y_k|x_k).$$

2. By law of total probability and Markov property:

$$\begin{aligned} f(x_k|Y_{k-1}) &= \int_{\mathbb{R}^n} f(x_k|x_{k-1}, Y_{k-1})f(x_{k-1}|Y_{k-1})\,\mathrm{d}\,x_{k-1} \\ &= \int_{\mathbb{R}^n} f(x_k|x_{k-1})f(x_{k-1}|Y_{k-1})\,\mathrm{d}\,x_{k-1} \end{aligned} \tag{2}$$

3. By law of total probability and Markov property:

$$\begin{aligned} f(y_k|Y_{k-1}) &= \int_{\mathbb{R}^n} f(y_k|x_k, Y_{k-1})f(x_k|Y_{k-1})\,\mathrm{d}\,x_k \\ &= \int_{\mathbb{R}^n} f(y_k|x_k)f(x_k|Y_{k-1})\,\mathrm{d}\,x_k. \end{aligned}$$

As a result, the Bayes filter can be written in a recursive fashion as:

1. **Initialization:**

$$f(x_0|Y_{-1}) = f(x_0).$$

2. **Correction:**

$$f(x_k|Y_k) = \alpha f(y_k|x_k)f(x_k|Y_{k-1}),$$

where

$$\alpha = \left(\int_{\mathbb{R}^n} f(y_k|x_k)f(x_k|Y_{k-1})\,\mathrm{d}\,x_k\right)^{-1}.$$

The MMSE estimation can be derived as

$$\hat{x} = \mathbb{E}(x_k|Y_k) = \int_{\mathbb{R}^n} x_k f(x_k|Y_k)\,\mathrm{d}\,x_k.$$

3. **Prediction:**
$$f(x_{k+1}|Y_k) = \int_{\mathbb{R}^n} f(x_{k+1}|x_k) f(x_k|Y_k)\,\mathrm{d}\,x_k.$$

Drawbacks:
we need to store the conditional pdf $f(x_k|Y_k)$ or $f(x_k|Y_{k-1})$.
Possible Solutions:

- Assuming that the conditional pdf is Gaussian. Hence, only need to track the mean and covariance (Extended Kalman Filter).
- Approximating the conditional pdf by Monte Carlo sampling (Particle filter)

3 Estimation of Linear Gaussian System: Kalman Filter

Consider the following linear Gaussian system:

$$\begin{aligned} x_{k+1} &= Ax_k + w_k, \\ y_k &= Cx_k + v_k. \end{aligned}$$

where the process noise w_k is i.i.d. Gaussian noise with mean 0 and covariance Q. The measurement noise v_k is i.i.d. Gaussian noise with mean 0 and covariance R and the initial condition x_0 is Gaussian with mean 0 and covariance Σ. The random variables $\{x_0, w_0, \dots, w_k, v_0, \dots, v_k\}$ are jointly independent.

Observation: $(x_0, \dots, x_k, y_0, \dots, y_k)$ are jointly Gaussian. Hence, the conditional pdf $f(x_k|Y_k)$ and $f(x_k|Y_{k-1})$ is also Gaussian. As a result, we only need to keep track of

$$\begin{aligned} \hat{x}_{k|k} &= \mathbb{E}(x_k|Y_k),\ \hat{x}_{k|k-1} = \mathbb{E}(x_k|Y_{k-1}), \\ P_{k|k} &= \mathrm{Cov}(x_k|Y_k) = \mathbb{E}\left((x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T Y_k\right), \\ P_{k|k-1} &= \mathrm{Cov}(x_k|Y_{k-1}) = \mathbb{E}\left((x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T Y_{k-1}\right). \end{aligned}$$

1. **Initialization:**
$$\hat{x}_{0|-1} = 0,\ P_{0|-1} = \Sigma. \tag{3}$$

2. **Prediction:** Take the conditional expectation on both sides of $x_{k+1} = Ax_k + w_k$,
$$\mathbb{E}(x_{k+1}|Y_k) = A\mathbb{E}(x_k|Y_k) + \mathbb{E}(w_k|Y_k).$$
The second term on the RHS is 0 (**why?**). Hence
$$\hat{x}_{k+1|k} = A\hat{x}_{k|k}. \tag{4}$$
Therefore,
$$x_{k+1} - \hat{x}_{k+1|k} = A(x_k - \hat{x}_{k|k}) + w_k,$$

which implies that

$$(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^T = A(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T A^T + w_k w_k^T + A(x_k - \hat{x}_{k|k})w_k^T + w_k(x_k - \hat{x}_{k|k})^T A^T.$$

Take the conditional expectation on Y_k on both sides. Notice that the conditional expectation of the last two terms on the RHS is 0 (**why?**). Hence

$$P_{k+1|k} = AP_{k|k}A^T + Q. \tag{5}$$

3. **Correction:** We need the following theorem:

Theorem 1. *Assume that the joint pdf of x, y satisfies*

$$f\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{bmatrix}\right),$$

then the following equalities hold:

$$\mathbb{E}(x|y) = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y),$$
$$\mathrm{Cov}(x|y) = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T.$$

By (4) and (5), we already know that

$$\mathbb{E}(x_{k+1}|Y_k) = \hat{x}_{k+1|k},\ \mathrm{Cov}(x_{k+1}|Y_k) = P_{k+1|k}.$$

Now take the conditional expectation on Y_k on both sides of $y_{k+1} = Cx_{k+1} + v_{k+1}$, we get

$$\mathbb{E}(y_{k+1}|Y_k) = C\hat{x}_{k+1|k}.$$

Similar to the proof of (5), we have

$$\mathrm{Cov}(y_{k+1}|Y_k) = CP_{k+1|k}C^T + R,$$
$$\mathrm{Cov}(x_{k+1}, y_{k+1}|Y_k) = \mathbb{E}\left((x_{k+1} - \hat{x}_{k+1|k})(y_{k+1} - \mathbb{E}(y_{k+1}|Y_k))^T|Y_k\right) = P_{k+1|k}C^T.$$

Therefore, the joint pdf of x_{k+1}, y_{k+1} satisfies

$$f\left(\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix}\middle| Y_k\right) \sim \mathcal{N}\left(\begin{bmatrix} \hat{x}_{k+1|k} \\ C\hat{x}_{k+1|k} \end{bmatrix}, \begin{bmatrix} P_{k+1|k} & P_{k+1|k}C^T \\ CP_{k+1|k} & CP_{k+1|k}C^T + R \end{bmatrix}\right).$$

Using Theorem 1, we get the correction equations:

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + P_{k+1|k}C^T(CP_{k+1|k}C^T + R)^{-1}(y_{k+1} - C\hat{x}_{k+1|k}), \tag{6}$$

$$P_{k+1|k+1} = P_{k+1|k} - P_{k+1|k}C^T(CP_{k+1|k}C^T + R)^{-1}CP_{k+1|k}. \tag{7}$$

Equation (3), (4), (5), (6) and (7) are called Kalman filter.

Observations:

- $P_{k|k}, P_{k+1|k}$ do not depend on Y_k. Hence, they can be computed off-line.
- Define
$$K_k \triangleq P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}.$$
Assume that C is invertible.
  - If $P_{k|k-1} \gg R$, then $K_k \approx C^{-1}$ and
$$\hat{x}_{k|k} \approx \hat{x}_{k|k-1} + C^{-1}(y_k - C\hat{x}_{k|k-1}) = C^{-1}y_k.$$
If the prediction is inaccurate, then we will trust the measurement.
  - If $P_{k|k-1} \ll R$, then $K_k \approx 0$ and
$$\hat{x}_{k|k} \approx \hat{x}_{k|k-1}.$$
If the measurement is inaccurate, then we will trust the prediction.

  KF can be seen as an optimal way to put weights on the current measurement and the past measurements (predicted state estimate).

Drawbacks:

We still need to compute $P_{k|k}$, which involves matrix multiplication and inversion.

However, we can avoid computing $P_{k|k}$, by the following theorems:

Theorem 2. *Assuming that (A, C) is observable and $(A, Q^{1/2})$ is controllable, then $P_{k|k-1}$ converges to a unique value P, regardless of the initial condition Σ.*

Denote $K \triangleq \lim_{k\to\infty} K_k = PC^T(CPC^T + R)^{-1}$.

Theorem 3. *Consider the following linear estimator using gain matrix K:*

$$\tilde{x}_{k+1|k} = A\tilde{x}_{k|k},\ \tilde{x}_{k+1|k+1} = \tilde{x}_{k+1|k} + K(y_{k+1} - C\tilde{x}_{k+1|k}), \tag{8}$$

with initial condition $\tilde{x}_{0|-1} = 0$. Define

$$\tilde{P}_{k|k} = \mathrm{Cov}(x_k - \tilde{x}_{k|k}|Y_k), \tilde{P}_{k+1|k} = \mathrm{Cov}(x_{k+1} - \tilde{x}_{k+1|k}|Y_k),$$

then the linear estimator achieves the same asymptotic performance as the Kalman filter, i.e.,

$$\lim_{k\to\infty} \tilde{P}_{k+1|k} = P,\ \lim_{k\to\infty} \tilde{P}_{k|k} = \lim_{k\to\infty} P_{k|k}.$$

In practice, we can compute P (remember that $P_{k|k}$ can be computed off-line) and K off-line and use the linear estimator (8) instead of the Kalman filter.