

Estimation Theory

Behzad Samadi • 27 Oct 2024

- An **observation** is defined as:

$$y = h(x) + w$$

where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ denote the unknown vector and the measurement vector. $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a function of x and w is the observation noise with the power density $f_w(w)$.

- It is assumed that x is a random variable with an a priori power density $f_x(x)$ before the observation.
- The goal is to compute the “best” estimation of x using the observation.

Optimal Estimation

- The optimal estimation \hat{x} is defined based on a cost function J :

- $$\hat{x}_{opt} = \arg \min_{\hat{x}} E[J(x - \hat{x})]$$

Some typical cost functions:

- Minimum Mean Square Error (\hat{x}_{MMSE}):

- $$J(x - \hat{x}) = (x - \hat{x})^T W (x - \hat{x}), \quad W > 0$$

Absolute Value (\hat{x}_{ABS}):

- $$J(x - \hat{x}) = |x - \hat{x}|$$

Maximum a Posteriori (\hat{x}_{MAP}):

- $$J(x - \hat{x}) = \begin{cases} 0 & \text{if } |x - \hat{x}| \leq \epsilon \\ 1 & \text{if } |x - \hat{x}| > \epsilon \end{cases}, \quad \epsilon \rightarrow 0$$

It can be shown that:

- $$\hat{x}_{MMSE}(y) = E[x|y]$$
$$\int_{-\infty}^{\hat{x}_{ABS}(y)} f_{x|y}(x|y) dx = \int_{\hat{x}_{ABS}(y)}^{+\infty} f_{x|y}(x|y) dx$$
$$\hat{x}_{MAP} = \arg \max_x f_{x|y}(x|y)$$

If the a posteriori density function $f_{x|y}(x|y)$ has only one maximum and it is symmetric with respect to $E[x|y]$ then all the above estimates are equal to $E[x|y]$.

- In fact, assuming these conditions for $f_{x|y}(x|y)$, $E(x|y)$ is the optimal estimation for any cost function J if $J(0) = 0$ and $J(x - \hat{x})$ is nondecreasing with distance (**Sherman's Theorem**).
- **Maximum Likelihood Estimation:** \hat{x}_{ML} is the value of x that maximizes the probability of observing y :

$$\hat{x}_{ML}(y) = \arg \max_x f_{y|x}(y|x)$$

It can be shown that $\hat{x}_{ML} = \hat{x}_{MAP}$ if there is no a priori information about x .

Linear Gaussian Observation

- Consider the following observation: $y = Ax + Bw$ where $w \sim \mathcal{N}(0, I_r)$ is a Gaussian random vector and matrices $A_{m \times n}$ and $B_{m \times r}$ are known.
- In this observation, x is estimable if A has full column rank otherwise there will be infinite solutions for the problem.
- If BB^T is invertible, then:

$$f_{y|x}(y|x) = \frac{1}{(2\pi)^{n/2} |BB^T|^{1/2}} \times \exp\left(-\frac{1}{2} [(y - Ax)^T (BB^T)^{-1} (y - Ax)]\right)$$

The maximum likelihood estimation can be computed as:

$$\begin{aligned} \hat{x}_{ML} &= \arg \min_x (y - Ax)^T (BB^T)^{-1} (y - Ax) \\ &= (A^T (BB^T)^{-1} A)^{-1} A^T (BB^T)^{-1} y \end{aligned}$$

It is very interesting that \hat{x}_{ML} is the Weighted Least Square (WLS) solution to the following equation: $y = Ax$ with the weight matrix $W = BB^T$ i.e.

$$\hat{x}_{WLS} = \arg \min_x (y - Ax)^T W (y - Ax)$$

\hat{x}_{ML} is an unbiased estimation:

$$\begin{aligned} b(x) &= E[x - \hat{x}_{ML}|x] \\ &= E\left[(A^T (BB^T)^{-1} A)^{-1} A^T (BB^T)^{-1} y - x \mid x\right] = 0 \end{aligned}$$

The covariance of the estimation error is:

- $P_e(X) = E[(x - \hat{x}_{ML})(x - \hat{x}_{ML})^T | x] = (A^T (BB^T)^{-1} A)^{-1}$
 \hat{x}_{ML} is *efficient* in the sense of Cramér Rao bound.

- **Example:** Consider the following linear Gaussian observation:
 $y = ax + w$ where a is a nonzero real number and $w \sim \mathcal{N}(0, r)$ is the observation noise.

- **Maximum a Posteriori Estimation:** To compute \hat{x}_{MAP} , it is assumed that the a priori density of x is Gaussian with mean m_x and variance p_x :

- $$x \sim \mathcal{N}(m_x, p_x)$$

The conditions of Sherman's Theorem is satisfied and therefore:

- $$\begin{aligned}\hat{x}_{MAP} &= E[x|y] \\ &= m_x + \frac{p_{xy}}{p_y}(y - m_y) \\ &= m_x + \frac{ap_x}{a^2p_x + r}(y - am_x) \\ &= \frac{ap_x y + m_x r}{a^2p_x + r}\end{aligned}$$

Estimation bias:

- $$b_{MAP} = E[x - \hat{x}_{MAP}] = m_x - \frac{ap_x E[y] + rm_x}{a^2p_x + r} = m_x - \frac{a^2p_x m_x + rm_x}{a^2p_x + r} = 0$$

Estimation error covariance:

- $$p_{MAP} = E[(x - \hat{x}_{MAP})^2] = p_x - \frac{a^2 p_x^2}{a^2 p_x + r} = \frac{p_x r}{a^2 p_x + r}$$

Maximum Likelihood Estimation: For this example, we have:

- $$f_{y|x}(y|x) = f_w(y - ax) = \frac{1}{\sqrt{2\pi r}} \exp\left(-\frac{(y - ax)^2}{2r}\right)$$

With this information:

- $$\hat{x}_{ML} = \arg \max_x f_{y|x}(y|x) = \frac{y}{a}$$

Estimation bias:

- $$b_{ML} = E[x - \hat{x}_{ML}|x] = x - \frac{ax}{x} = 0$$

Estimation error covariance:

- $$p_{ML} = E[(x - \hat{x}_{ML})^2|x] = E\left[\left(x - \frac{ax + w}{a}\right)^2\right] = \frac{r}{a^2}$$

Comparing x_{MAP} and x_{ML} , we have:

$$\lim_{p_x \rightarrow +\infty} \hat{x}_{MAP} = \hat{x}_{ML}$$

It means that if there is no a priori information about x , the two estimations are equal.

- For the error covariance, we have:

$$\frac{1}{p_{MAP}} = \frac{1}{p_{ML}} + \frac{1}{p_x}$$

- In other words, information after observation is the sum of information of the observation and information before the observation.

- Estimation error covariance:

$$\lim_{p_x \rightarrow +\infty} \hat{p}_{MAP} = \hat{p}_{ML}$$

It is possible to include a priori information in maximum likelihood estimation.

- A priori distribution of x , $\mathcal{N}(m_x, p_x)$, can be rewritten as the following observation: $m_x = x + v$ where $v \sim \mathcal{N}(0, p_x)$ is the observation noise.

- **Combined observation:** $z = Ax + u$ where:

$$z = \begin{bmatrix} m_x \\ y \end{bmatrix}, A = \begin{bmatrix} 1 & 0 \\ 0 & a \end{bmatrix}, u = \begin{bmatrix} v \\ w \end{bmatrix}$$

The assumption is that v and w are *independent*. Therefore:

$$u \sim \mathcal{N}\left(0, \begin{bmatrix} p_x & 0 \\ 0 & r \end{bmatrix}\right)$$

Maximum likelihood estimation:

$$\begin{aligned} \hat{x}_{MLp}(z) &= \arg \max_x f_{z|x}(z|x) \\ &= \arg \min_x \left(\frac{(m_x - x)^2}{p_x} + \frac{(y - ax)^2}{r} \right) \\ &= \frac{ap_x y + m_x r}{a^2 p_x + r} = \hat{x}_{MAP} \end{aligned}$$

\hat{x}_{MLp} is unbiased and has the same error covariance as \hat{x}_{MAP} .

- Therefore \hat{x}_{MLp} and \hat{x}_{MAP} are equivalent.

Standard Kalman Filter

- Consider the following linear system:

$$\begin{cases} x(k+1) &= A(k)x(k) + w(k) \\ y(k) &= C(k)x(k) + v(k) \end{cases}$$

where $x(k) \in \mathbb{R}^n$, $y(k) \in \mathbb{R}^m$ denote the state vector and measurement vector at time t_k .

- $w(k) \sim \mathcal{N}(0, Q(k))$ and $v(k) \sim \mathcal{N}(0, R(k))$ are independent Gaussian white noise processes where $R(k)$ is invertible.

- It is assumed that there is an a priori estimation of x , denoted by $\hat{x}^-(k)$, which is assumed to be unbiased with a Gaussian estimation error, independent of w and v :

$$e^-(k) \sim \mathcal{N}(0, P^-(k))$$

where $P^-(k)$ is invertible.

- The Kalman filter is a recursive algorithm to compute the state estimation.
- **Output Measurement:** Information in $\hat{x}^-(k)$ and $y(k)$ can be written as the following observation:

$$\begin{bmatrix} \hat{x}^-(k) \\ y(k) \end{bmatrix} = \begin{bmatrix} I \\ C(k) \end{bmatrix} x(k) + \begin{bmatrix} e^-(k) \\ v(k) \end{bmatrix}$$

Considering the independence of $e^-(k)$ and $v(k)$, we have:

- $$\begin{bmatrix} e^-(k) \\ v(k) \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} P^-(k) & 0 \\ 0 & R(k) \end{bmatrix}\right)$$

Using the Weighted Least Square (WLS) and matrix inversion formula:

- $$(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}$$

Assuming:

- $$K(k) = P^-(k)C^T(k)[C(k)P^-(k)C^T(k) + R(k)]^{-1}$$

We have:

- $$\hat{x}(k) = \hat{x}^-(k) + K(k)(y(k) - C(k)\hat{x}^-(k))$$

State estimation is the sum of a priori estimation and a multiplicand of output prediction error. Since:

- $$\hat{y}^-(k) = C(k)\hat{x}^-(k)$$

$K(k)$ is the Kalman filter gain.

- Estimation error covariance:

- $$P(k) = (I - K(k)C(k))P^-(k)$$

Information:

$$\hat{x}(k) = x(k) + e(k)$$

where $e(k) \sim \mathcal{N}(0, P(k))$

- **State Update:** To complete a recursive algorithm, we need to compute $\hat{x}^-(k+1)$ and $P^-(k+1)$.

- Information:

$$\begin{aligned} \hat{x}(k) &= x(k) + e(k) \\ 0 &= \begin{bmatrix} -I & A(k) \end{bmatrix} \begin{bmatrix} x(k+1) \\ x(k) \end{bmatrix} + w(k) \end{aligned}$$

- By removing $x(k)$ from the above observation, we have:

- $$A(k)\hat{x}(k) = x(k+1) + A(k)e(k) - w(k)$$

It is easy to see:

- $$\hat{x}^-(k+1) = A(k)\hat{x}(k)$$

Estimation error:

- $$e^-(k+1) = A(k)e(k) - w(k)$$

Estimation covariance:

$$P^-(k+1) = A(k)P(k)A^T(k) + Q(k)$$

Summary:

- Initial Conditions: $\hat{x}^-(k)$ and its error covariance $P^-(k)$.

- Gain Calculation:

- $$K(k) = P^-(k)C^T(k)[C(k)P^-(k)C^T(k) + R(k)]^{-1}$$

$\hat{x}(k)$:

- $$\hat{x}(k) = \hat{x}^-(k) + K(k)(y(k) - C(k)\hat{x}^-(k))$$

$$P(k) = (I - K(k)C(k))P^-(k)$$

$\hat{x}^-(k+1)$:

- $$\hat{x}^-(k+1) = A(k)\hat{x}(k)$$

$$P^-(k+1) = A(k)P(k)A^T(k) + Q(k)$$

Go to gain calculation and continue the loop for $k+1$.

Remarks:

- Estimation residue:

- $$\gamma(k) = y(k) - C(k)\hat{x}^-(k)$$

Residue covariance:

- $$P_\gamma(k) = C(k)P^-(k)C^T(k) + R(k)$$

The residue signal is used for monitoring the performance of Kalman filter.

- Modeling error, round-off error, disturbance, correlation between input and measurement noise, and other factors might cause a biased and colored residue.
- The residue signal can be used in Fault Detection and Isolation (FDI).
- The standard Kalman filter is not numerically robust because it contains matrix inversion. For example, the calculated error covariance matrix might not be positive definite because of computational errors.

- There are different implementations of Kalman filter to improve the standard Kalman filter in the following aspects:
 - Computational efficiency
 - Dealing with disturbance or unknown inputs
 - Handling singular systems (difference algebraic equations)