# Maximum likelihood cost function

Avneet Kaur

May 2025

State-space models are systems of ordinary differential equations used for estimating state of a system equipped with measurements. A continuous-time state-space model is given as follows:

$$\begin{aligned} \dot{x}(t) &= f(t, x) \\ y(t) &= h(t, x) \\ x(0) &= x_0 \end{aligned} \tag{1}$$

where $x(t) \in \mathcal{X} \sim \mathbb{R}^{N_x}$ is called the state of the system and $y(t) \in \mathcal{Y} \sim \mathbb{R}^{N_{obs}}$ is called the measurement(observation) vector. $f$ and $h$ are functions that define the state equation and measurement equation respectively and $x(0) = x^0$ gives the initial condition. It is usually assumed that both $f$ and $h$ are sufficiently smooth , for example, continuously differentiable.

The aim is to develop an estimator which is in itself a dynamical system, for example

$$\begin{aligned} \dot{\hat{x}}(t) &= f(t, \hat{x}) + F(t)(y(t) - \hat{y}(t)) \\ \hat{y}(t) &= h(t, \hat{x}). \end{aligned} \tag{2}$$

Its solution will be the estimated state $\hat{x}$ and it uses the update based on new measurement $y(t)$ to correct the state. It does so by adding the update term $F(t)(y(t) - \hat{y}(t))$ where $F(t)$ is the filter value, into the differential equation for the estimated state.

Solving the estimator system (2) may not seem difficult but in practical applications this is a challenging task due to various reasons. The complexity of high-order systems, availability of finite number of measurements, inaccurate measurements, incomplete information about the initial state and errors in model make it difficult to estimate the state of the system. To account for these errors, noises are added to the system to get:

$$\begin{aligned} \dot{x}(t) &= f(t, x) + w(t) \\ y(t) &= h(t, x) + v(t) \\ x(0) &= x_0 + \bar{x}_0 \end{aligned} \tag{3}$$

where $w(t) \in \mathcal{W}$ is the process noise vector, $v(t) \in \mathcal{V}$ is the measurement noise vector and $\bar{x}_0 \in \mathcal{X}$ is the initial condition noise.

This work considers the following discretized version of the system (3):

$$\begin{aligned} x_{t+1} &= f(x_t) + w_t \\ y_t &= h(x_t) + v_t \\ x(0) &= x_0 + \bar{x}_0 \end{aligned} \tag{4}$$

where $x_{t+1}$ is the state at time $t+1$ given the state at time $t$ and $y_t$ is the measurement at time $t$. $w_t$ and $v_t$ are process and measurement noise at time $t$.

The least square estimator(LSE), also called the minimum variance estimator which uses variance as a cost function is one of the most commonly used optimal cost function for state estimation. In this chapter, a particular type of cost function called the maximum likelihood cost function for discrete-time state-space systems is discussed and is shown equivalent to the minimum variance cost function. The maximum likelihood cost function will thus be used in the implementation of the proposed research problems.

The goal of maximum likelihood estimation is to make inferences about the population that is most likely to have generated the sample. To formulate the likelihood function assume that a given set of independent and identically distributed Gaussian random variables $\{z_1, z_2, \ldots, z_n\}$ are sampled from an unknown population at random. Specifically, the aim is to find the joint probability distribution $\mathbf{p}(z_1, z_2, \ldots, z_n)$ of these random variables. Associated with each probability distribution is a unique vector $\theta = [\theta_1, \theta_2, \ldots, \theta_k]^T$ of parameters that index the probability distribution within a parametric family $\{\mathbf{p}(\cdot;\theta) | \theta \in \Theta\}$, where $\Theta$ is called the parameter space, a finite-dimensional subset of Euclidean space. Evaluating the joint density at the observed data sample $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ gives a real-valued function,

$$L_n(\theta) = L_n(\theta; \mathbf{z}) = \mathbf{p}(\mathbf{z})$$

which is called the likelihood function. Now, let $\{z_1, z_2, \ldots, z_n\}$ be such that $z_i \sim \mathcal{N}(\mu, \sigma), \forall i \in 1, 2, \ldots, n$, i.e., the mean of each of these variables is $\mu$ and their covariance is $\sigma$ then the associated likelihood function is

$$L_n(\mu, \sigma; \mathbf{z}) = \prod_{i=1}^{n} \mathbf{p}(z_i) \tag{5}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(z_i - \mu)^2}{2\sigma^2}\right). \tag{6}$$

The log-likelihood function for (6) will be:

$$\log L_n(\mu, \sigma; \mathbf{z}) = \log\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} - \sum_{i=1}^{n} \frac{(z_i - \mu)^2}{2\sigma^2}. \tag{7}$$

Maximizing the above function is the same as minimizing

$$J(\mu, \sigma; \mathbf{z}) = \sum_{i=1}^{n} \frac{(z_i - \mu)^2}{\sigma^2}$$

which is a cost function for the variance of $\mathbf{z}$. Thus, maximizing log likelihood implies minimizing variance. This can be easily extended to the case where the variables are non-i.i.d. Hence, in statistics, both minimum variance and maximum likelihood have the same solution.

Now let us consider the problem of state estimation for the discrete-time dynamical state-space system (4). Assume $w$ and $v$ are Gaussian random variables with mean 0 and co-variance matrices

$$\mathbf{E}\{w_t w'_\tau\} = Q\delta_{t-\tau}$$
$$\mathbf{E}\{v_t v'_\tau\} = R\delta_{t-\tau}$$

where matrices $Q$ and $R$ are positive definite. Also assume that

$$\mathbf{E}\{w_t v'_\tau\} = 0, \forall t, \tau \in \mathbf{N},$$

i.e., process and measurement noise are uncorrelated. Assume that the initial state $x_0$ is also a Gaussian random variable with mean $\mu$ and covariance matrix $\Lambda$. Let us also assume that the initial state and noises are uncorrelated. Then,

$$x_{t+1}|x_t \sim \mathcal{N}(f(x_t), Q),$$

i.e., $x_{t+1}$ given $x_t$ is sampled from the Gaussian distribution centered at $f(x_t)$ with co-variance $Q$ and,

$$y_t|x_t \sim \mathcal{N}(h(x_t), R)$$

for time $t \geq 0$. The likelihood function of this problem can then be simplified using repeated application of Bayes' rule:

$$
\begin{aligned}
L(\cdot) &= \mathbf{p}(y_k, x_{k+1}, y_{k-1}, x_k, ..., y_0, x_1, x_0) \\
&= \mathbf{p}(y_k, x_{k+1}|y_{k-1}, x_k, ..., y_0, x_1, x_0) \times \mathbf{p}(y_{k-1}, x_k, ..., y_0, x_1, x_0) \\
&= \mathbf{p}(y_k, x_{k+1}|x_k) \times \mathbf{p}(y_{k-1}, x_k|y_{k-2}, x_{k-1}, ..., y_0, x_1, x_0) \times \mathbf{p}(y_{k-2}, x_{k-1}, ..., y_0, x_1, x_0) \\
&\vdots \\
&= \mathbf{p}(x_0) \times \prod_{i=1}^{k} \mathbf{p}(y_i, x_{i+1}|x_i) \\
&= \mathbf{p}(x_0) \times \prod_{i=1}^{k} \frac{\mathbf{p}(y_i, x_{i+1}, x_i)}{\mathbf{p}(x_i)} \\
&= \mathbf{p}(x_0) \times \prod_{i=1}^{k} \frac{\mathbf{p}(y_i|x_{i+1}, x_i).\mathbf{p}(x_{i+1}|x_i).\mathbf{p}(x_i)}{\mathbf{p}(x_i)} \\
&= \mathbf{p}(x_0) \times \prod_{i=1}^{k} \mathbf{p}(y_i|x_i).\mathbf{p}(x_{i+1}|x_i) \\
&= \frac{1}{\sqrt{2\pi|\Lambda|}} \exp\left(-\frac{1}{2}(x_0-\mu)'\Lambda^{-1}(x_0-\mu)\right) \times \\
&\qquad \prod_{i=1}^{k} \left\{ \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|Q|}} \exp\left(-\frac{1}{2}(x_{i+1}-f(x_i))'Q^{-1}(x_{i+1}-f(x_i))\right) \times \right. \\
&\qquad\qquad \left. \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|R|}} \exp\left(-\frac{1}{2}(y_i-h(x_i))'R^{-1}(y_i-h(x_i))\right) \right\} \\
&= \frac{1}{\sqrt{2\pi|\Lambda|}} \exp\left(-\frac{1}{2}(x_0-\mu)'\Lambda^{-1}(x_0-\mu)\right) \times \\
&\qquad \prod_{i=1}^{k} \left\{ \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|Q|}} \exp\left(-\frac{1}{2}w_i'Q^{-1}w_i\right) \times \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|R|}} \exp\left(-\frac{1}{2}v_i'R^{-1}v_i\right) \right\}.
\end{aligned}
$$

The log likelihood function then becomes

$$
\begin{aligned}
\log L(\cdot) &= \log\left(\frac{1}{\sqrt{2\pi|\Lambda|}}\right) - \left(\frac{1}{2}(x_0-\mu)'\Lambda^{-1}(x_0-\mu)\right) + \\
&\sum_{i=1}^{k} \left\{ \log\left(\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|Q|}}\right) - \left(\frac{1}{2}w_i{}'Q^{-1}w_i\right) + \log\left(\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|R|}}\right) - \left(\frac{1}{2}v_i{}'R^{-1}v_i\right) \right\}.
\end{aligned}
$$

Note that maximizing $\log L(\cdot)$ is equivalent to minimizing

$$
J(\cdot) = \frac{1}{2}(x_0-\mu)'\Lambda^{-1}(x_0-\mu) + \sum_{i=1}^{k} \frac{1}{2}\left\{w_i{}'Q^{-1}w_i + v_i{}'R^{-1}v_i\right\} \tag{8}
$$

which is the minimum variance cost function. This makes the maximum likelihood cost function an optimal choice for state estimation.