ELEG 5633 Detection and Estimation
Least Sqauares Estimation

Jingxian Wu

Department of Electrical Engineering
University of Arkansas

Outline

- Least Squares Estimation
- Application: Linear Regression

Motivations

- Bayesian Estimators
  - MAP or MMSE: require knowledge of the likelihood function $p(x|\theta)$ and prior distribution $p(\theta)$
  - LMMSE: require 1st and 2nd order statistics: μ_x, μ_θ, $\mathbf{C}_x$, $\mathbf{C}_\theta$, and $\mathbf{C}_{\theta x}$
- Classical Estimator
  - ML and MVUE: require likelihood function $p(x|\theta)$
  - BLUE: require a linear model and the noise covariance $\mathbf{C}_w$.

What if none of the statistical information is unavailable:
that is, $p(x|\theta)$, $\mathbf{C}_x$, $\mathbf{C}_\theta$, $\mathbf{C}_w$ are not known?

- Least squares estimation (LSE) does not require a statistical model.

Least Squares Estimation (LSE)

- Model
  - Signal model (known): $\mathbf{s}(\boldsymbol{\theta})$, which describes the relationship between the desired signal and parameter
  - Observed model: $\mathbf{x} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{w}$
- LSE: minimize the following objective function

$$
\begin{aligned}
J(\boldsymbol{\theta}) &= \sum_{i=1}^{N} |x_n - s_n(\boldsymbol{\theta})|^2 = \|\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})\|_2^2 \\
&= \mathrm{tr}\left\{[\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})][\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})]^T\right\}
\end{aligned}
$$

$$
\hat{\boldsymbol{\theta}}_{\mathrm{LSE}} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}}\ \|\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})\|_2^2
$$

There is no $\mathbb{E}$ in the objective function!

Least Squares Estimation (LSE)

Some examples of signal models

- Non-linear signal model: $s_n = \cos(2\pi f n)$, $n = 1, \cdots, N$. Estimate f.
  - It is usually difficult to find the explict solution for non-linear signal models, and it needs to be solved with numerical methods.
- Linear signal model: $\mathbf{s}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta}$
  - Linear LSE. Easy to solve.
- Affine signal model: $\mathbf{s}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta} + \mathbf{b}$
  - Can be easily converted to linear models.

Linear LSE

- Signal Model $\mathbf{s}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta}$
- Problem formulation (there is no assumption about the distribution of noise)

$$\min_{\boldsymbol{\theta}} \|\mathbf{x} - \mathbf{H}\boldsymbol{\theta}\|^2$$

Solutions

- Objective function

$$J(\boldsymbol{\theta}) = [\mathbf{x} - \mathbf{H}\boldsymbol{\theta}]^T[\mathbf{x} - \mathbf{H}\boldsymbol{\theta}] = \mathbf{x}^T\mathbf{x} - \mathbf{x}^T\mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{x} + \boldsymbol{\theta}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\theta}$$

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T\mathbf{x} + 2\mathbf{H}^T\mathbf{H}\boldsymbol{\theta} = 0$$

- If $\mathbf{H}^T\mathbf{H}$ is invertible

$$\hat{\boldsymbol{\theta}}_{\text{LSE}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

Linear LSE

- Minimum objective function

$$
\begin{aligned}
\min J(\boldsymbol{\theta}) &= [\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}]^T[\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}] \\
&= \mathbf{x}^T\mathbf{x} - \mathbf{x}^T\mathbf{H}\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^T\mathbf{H}^T\mathbf{x} + \hat{\boldsymbol{\theta}}^T\mathbf{H}^T\mathbf{H}\hat{\boldsymbol{\theta}} \\
&= \mathbf{x}^T\mathbf{x} - \mathbf{x}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} \\
&= \mathbf{x}^T\left[\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right]\mathbf{x}
\end{aligned}
$$

Performance of Linear LSE

Consider the case that the noise is zero-mean $\mathbb{E}[\mathbf{w}] = 0$ with covariance matrix $\mathbf{C}$.

- Unbiased estimator

$$\mathbb{E}[\hat{\boldsymbol{\theta}}] = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{H}\boldsymbol{\theta} + \mathbf{w}) = \boldsymbol{\theta}$$

- Covariance matrix of $\hat{\boldsymbol{\theta}}$

$$\begin{aligned}
\mathbf{C}_{\hat{\boldsymbol{\theta}}} &= \mathbb{E}[\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^T] - \boldsymbol{\theta}\boldsymbol{\theta}^T \\
&= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbb{E}[\mathbf{x}\mathbf{x}^T]\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1} - \boldsymbol{\theta}\boldsymbol{\theta}^T \\
&= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{H}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{H}^T + \mathbf{C}_w)\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1} - \boldsymbol{\theta}\boldsymbol{\theta}^T \\
&= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}
\end{aligned}$$

  - If the noise is white $\mathbf{C} = \sigma^2\mathbf{I}$

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2(\mathbf{H}^T\mathbf{H})^{-1}$$

Linear LSE v.s. BLUE

- BLUE:

$$\hat{\boldsymbol{\theta}}_{\text{BLUE}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$$

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}_{\text{BLUE}}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}^T)^{-1}$$

- Linear LSE:

$$\hat{\boldsymbol{\theta}}_{\text{LSE}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}_{\text{LSE}}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}$$

- The linear LSE is the same as BLUE if

$$\mathbf{C} = \sigma^2\mathbf{I}$$

that is, white noise (distribution unknown).

Applications: Linear Regression

suppose we make noisy observations of an unknown function f according to

$$y_i = f(\mathbf{x}_i) + w_i, \quad i = 1, \cdots, n$$

where $\mathbf{x}_i \in \mathbb{R}^p$ is a p-dimensional vector with $p < n$, and w_i are observation noise.

- Regression (function estimation): estimate f by using the n observatons $(\mathbf{x}_i, y_i)$, $i = 1, \cdots, n$.

- Linear Regression: assume the function is linear (affine)

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}$$

Estimate $\boldsymbol{\beta} = [\beta_0, \beta_1, \cdots, \beta_p]^T$ by using the n observations $(\mathbf{x}_i, y_i)$, $i = 1, \cdots, n$.

Linear Regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{w}$$

- observation vector $\mathbf{y}$, coefficient vector $\boldsymbol{\beta}$, noise vector $\mathbf{w}$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^{n\times 1}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \in \mathbb{R}^{p\times 1} \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \in \mathbb{R}^{n\times 1}$$

- data matrix

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \ddots & \ddots & \ldots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \in \mathbb{R}^{n\times p}$$

Linear Regression

- Solution:

$$\hat{\boldsymbol{\beta}}_{\mathsf{LSE}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

- Covariance matrix of $\hat{\boldsymbol{\beta}}_{\mathsf{LSE}}$ If the noise is zero-mean and white with covariance matrix $\mathbf{C} = \sigma^2\mathbf{I}$

$$\mathbf{C}_{\hat{\boldsymbol{\beta}}} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

$$\mathsf{var}(\hat{\beta}_i) = (\mathbf{C}_{\hat{\boldsymbol{\beta}}})_{ii}$$

- The impact of x_i on the output y can be evaluated by using the combination $\hat{\beta}_i$ and var($\hat{\beta}_i$).