Static Linear Models and Linear Least Squares

Roland Hostettler

September 13, 2018

Intended Learning Outcomes

After this lecture, you will be able to:

- Identify and construct scalar and vector linear models;
- apply and derive (weighted and sequential) linear least squares estimators;
- investigate the properties of linear least squares estimators.

Recap

- Sensor fusion requires *sensors*, *models*, and *estimation algorithms*
- The aim is to estimate static parameters $\boldsymbol{\theta}$ or time-varying states $\boldsymbol{x}_n$
- The general measurement model is

$$y_n = g_n(\boldsymbol{\theta}) + r_n \text{ or } \boldsymbol{y}_n = g_n(\boldsymbol{\theta}) + \boldsymbol{r}_n$$

- The *quadratic cost* of the error is

$$J_{\mathsf{LS}}(\boldsymbol{\theta}) = \sum_{n=1}^{N} (y_n - g_n(\boldsymbol{\theta}))^2$$

$$J_{\mathsf{LS}}(\boldsymbol{\theta}) = \sum_{n=1}^{N} (\boldsymbol{y}_n - g_n(\boldsymbol{\theta}))^\mathsf{T} (\boldsymbol{y}_n - g_n(\boldsymbol{\theta}))$$

Scalar Model: Model & Cost Function

- Many sensors measure (a scaled) version of a single parameter θ

$$y_n = c\theta + r_n,$$

with $\mathrm{E}\{r_n\} = 0$ and $\mathrm{var}\{r_n\} = \sigma_{r,n}^2$

- The error for one measurement and cost functions are given by

$$e_n = y_n - c\theta$$

and

$$J_{\mathsf{LS}}(\theta) = \sum_{n=1}^{N} (y_n - c\theta)^2$$

Scalar Model: Minimizing the Cost

- The derivative is given by

$$\frac{\partial J_{\mathsf{LS}}(\theta)}{\partial \theta} = -2c \sum_{n=1} y_n + 2Nc^2\theta$$

- Setting the derivative to zero and solving for θ yields

$$\hat{\theta}_{\mathsf{LS}} = \frac{1}{Nc} \sum_{n=1}^{N} y_n.$$

- This is the least squares estimator for the model

$$y_n = c\theta + r_n,$$

Scalar Model: Estimator Properties

- What are the estimator's statistical properties?
- Mean:

$$\mathrm{E}\{\hat{\theta}_{LS}\} = \theta + \sum_{n=1}^{N} \mathrm{E}\{r_n\} = \theta$$

- Variance:

$$\mathrm{var}\{\hat{\theta}\} = \frac{1}{N^2 c^2} \sum_{n=1}^{N} \sigma_{r,n}^2$$

- $\hat{\theta}$ converges to $\theta \Rightarrow$ Estimator is *unbiased*

Scalar Model: Summary

- The scalar linear model is given by

$$y_n = c\theta + r_n$$

- The linear least squares estimator

$$\hat{\theta}_{\text{LS}} = \frac{1}{Nc}\sum_{n=1}^{N} y_n$$

  minimizes the cost function

$$J_{\text{LS}}(\theta) = \sum_{n=1}^{N}(y_n - c\theta)^2$$

- The mean and variance of the least squares estimator are

$$\mathrm{E}\{\hat{\theta}_{LS}\} = \theta \text{ and } \mathrm{var}\{\hat{\theta}\} = \frac{1}{N^2c^2}\sum_{n=1}^{N}\sigma_{r,n}^2$$

Scalar Model: Example

Vector Models

- Scalar observations, several parameters $\theta_1, \theta_2, \dots, \theta_K$:

$$
\begin{aligned}
y_n &= c_1\theta_1 + c_2\theta_2 + \dots + c_K\theta_K + r_n \\
&= \boldsymbol{c}\boldsymbol{\theta} + r_n
\end{aligned}
$$

- Vector observations, several parameters:

$$
\begin{aligned}
\boldsymbol{y}_n &= \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1K} \\ c_{21} & c_{22} & \dots & c_{1K} \\ \vdots & \vdots & \ddots & \vdots \\ c_{d_y1} & c_{d_y2} & \dots & c_{d_yK} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \end{bmatrix} + \boldsymbol{r}_n \\
&= \boldsymbol{C}_n\boldsymbol{\theta} + \boldsymbol{r}_n,
\end{aligned}
$$

General Linear Model: Definition

- Batch notation:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \\ \vdots \\ \boldsymbol{C}_N \end{bmatrix} \boldsymbol{\theta} + \begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \vdots \\ \boldsymbol{r}_N \end{bmatrix}$$

- Compact notation:

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{\theta} + \boldsymbol{r},$$

  with $\mathrm{Cov}\{\boldsymbol{r}\} = \boldsymbol{R}$.
- This is the general linear model, both the scalar and vector cases can be expressed in this way.

General Linear Model: Least Squares

Task

Derive the least squares estimator for the general linear model

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{\theta} + \boldsymbol{r}.$$

What are the estimator's mean and covariance?

Hints

Some vector calculus identities[1]:

$$\frac{\partial \boldsymbol{x}^\mathsf{T}\boldsymbol{a}}{\partial \boldsymbol{x}} = \left(\frac{\partial \boldsymbol{a}^\mathsf{T}\boldsymbol{x}}{\partial \boldsymbol{x}}\right)^\mathsf{T} = \boldsymbol{a}$$

$$\frac{\partial \boldsymbol{x}^\mathsf{T}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$$

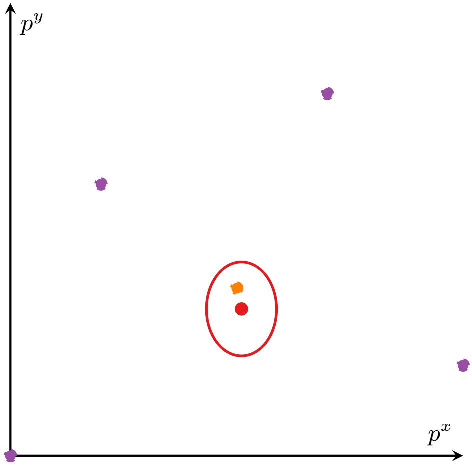
[1] See K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Technical University of Denmark, Tech. Rep., November 2012

General Linear Model: Least Squares

- The least squares estimator for the general linear model is
$$\hat{\boldsymbol{\theta}}_{\mathsf{LS}} = (\boldsymbol{G}^{\mathsf{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathsf{T}}\boldsymbol{y}$$
- Its statistical properties are

$$\begin{aligned}
\mathrm{E}\{\hat{\boldsymbol{\theta}}_{\mathsf{LS}}\} &= \boldsymbol{\theta} \\
\mathrm{Cov}\{\hat{\boldsymbol{\theta}}_{\mathsf{LS}}\} &= (\boldsymbol{G}^{\mathsf{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathsf{T}}\boldsymbol{R}\boldsymbol{G}((\boldsymbol{G}^{\mathsf{T}}\boldsymbol{G})^{-1})^{\mathsf{T}}.
\end{aligned}$$

Example: Localizing a Target (1)



Weighted Linear Least Squares (1/2)

- Recall the general linear model:

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{\theta} + \boldsymbol{r},$$

  with $\mathrm{Cov}\{\boldsymbol{r}\} = \boldsymbol{R}$.
- Weighted least squares cost function:

$$\begin{aligned} J_{\mathsf{WLS}}(\boldsymbol{\theta}) &= (\boldsymbol{y} - g(\boldsymbol{\theta}))^{\mathsf{T}} \boldsymbol{R}^{-1} (\boldsymbol{y} - g(\boldsymbol{\theta})) \\ &= (\boldsymbol{y} - \boldsymbol{G}\boldsymbol{\theta})^{\mathsf{T}} \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{G}\boldsymbol{\theta}) \end{aligned}$$

- Weighted linear least squares estimator:

$$\hat{\boldsymbol{\theta}}_{\mathsf{WLS}} = (\boldsymbol{G}^{\mathsf{T}} \boldsymbol{R}^{-1} \boldsymbol{G})^{-1} \boldsymbol{G}^{\mathsf{T}} \boldsymbol{R}^{-1} \boldsymbol{y}.$$

Weighted Linear Least Squares (2/2)

- Weighted linear least squares estimator:

$$\hat{\boldsymbol{\theta}}_{\mathsf{WLS}} = (\boldsymbol{G}^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{y}.$$

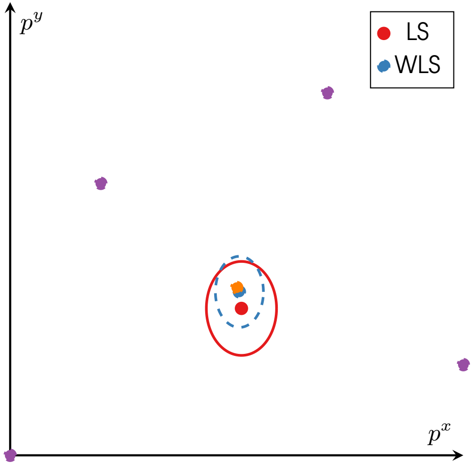
- Properties:

$$\begin{aligned} \mathrm{E}\{\hat{\boldsymbol{\theta}}_{\mathsf{WLS}}\} &= \boldsymbol{\theta} \\ \mathrm{Cov}\{\hat{\boldsymbol{\theta}}_{\mathsf{WLS}}\} &= (\boldsymbol{G}^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{G})^{-1} \end{aligned}$$

- It can be shown that $\boldsymbol{W} = \boldsymbol{R}^{-1}$ *minimizes* $\mathrm{Cov}\{\hat{\boldsymbol{\theta}}_{\mathsf{WLS}}\}$ over all choices for $\boldsymbol{W}$ and in this case

$$\mathrm{Cov}\{\hat{\boldsymbol{\theta}}_{\mathsf{WLS}}\} \leq \mathrm{Cov}\{\hat{\boldsymbol{\theta}}_{\mathsf{LS}}\}$$

Example: Localizing a Target (2)



Summary

- The general linear model is given by

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{\theta} + \boldsymbol{r},\ \mathrm{E}\{\boldsymbol{r}\} = 0,\ \mathrm{Cov}\{\boldsymbol{r}\} = \boldsymbol{R}$$

- The weighted linear least squares estimator is

$$\hat{\boldsymbol{\theta}}_{WLS} = (\boldsymbol{G}^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G})^{-1}\boldsymbol{G}^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{y}$$

  with $\mathrm{E}\{\hat{\boldsymbol{\theta}}_{WLS}\} = \boldsymbol{\theta}$ and $\mathrm{Cov}\{\hat{\boldsymbol{\theta}}_{WLS}\} = (\boldsymbol{G}^\mathsf{T}\boldsymbol{R}^{-1}\boldsymbol{G})^{-1}$.