Systems Dynamics

Course ID: 267MI – Fall 2023

Thomas Parisini
Gianfranco Fenu

University of Trieste
Department of Engineering and Architecture

267MI –Fall 2023

Lecture 9
Least-Squares Estimation

Lecture 9: Table of Contents

Introduction to Least-Squares Estimation

Introduction to Least-Squares Estimation

Linear Regression

Least-Squares Estimation - Linear Regression

Linear regression

- This is the typical context suited to the use of the least-squares (LS) estimator
- We have $q+1$ variables $y(t), u_1(t), \ldots, u_q(t)$ over the time-window $t = 1, 2, \ldots, N$
- We want to compute (if possible) q parameters $\vartheta_1, \vartheta_2, \ldots, \vartheta_q$ such that

$$y(t) = \vartheta_1\, u_1(t) + \cdots \vartheta_q\, u_q(t)\,, \quad t = 1, \ldots, N \quad (\star)$$

- Relationship $(\star)$ is defined as the **linear regression** of the variable $y(t)$ on the variables $u_1(t), \ldots, u_q(t)$

Least-Squares Estimation - Linear Regression (cont.)

- The problem can be equivalently stated in vector form letting

$$\vartheta = \begin{bmatrix} \vartheta_1 \\ \vdots \\ \vartheta_q \end{bmatrix} \quad \varphi(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_q(t) \end{bmatrix}$$

  and hence getting

$$y(t) = \varphi(t)^\top \vartheta$$

- Clearly, in case of real data, an **error** $\varepsilon(t)$ **is always present**:

$$\varepsilon(t) = y(t) - \varphi(t)^\top \vartheta$$

Least-Squares Estimation - Linear Regression (cont.)

- The goal of the linear regression problem is to minimize the error $\varepsilon(t)$ by determining an optimal vector ϑ° such that such a **minimum** is achieved
- We introduce the **quadratic cost function**:

$$J(\vartheta) = \sum_{t=1}^{N} [\varepsilon(t)]^2 = \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta\right]^2$$

- Therefore, the **Least-Squares Estimator** is given by

$$\vartheta^\circ = \arg\min_{\vartheta} J(\vartheta)$$

Least-Squares Estimation - Linear Regression (cont.)

- Denoting by ϑ_i the i-th component of the vector ϑ, one has:

$$\frac{\partial J}{\partial \vartheta_i} = \frac{\partial}{\partial \vartheta_i} \left\{ \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta \right]^2 \right\}$$
$$= -2 \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta \right] u_i(t) \,, \quad i = 1, 2, \dots, q$$

and noticing that

$$\frac{\partial J}{\partial \vartheta} = \begin{bmatrix} \frac{\partial J}{\partial \vartheta_1} & \frac{\partial J}{\partial \vartheta_2} & \cdots & \frac{\partial J}{\partial \vartheta_q} \end{bmatrix}$$

it follows that

$$\frac{\partial J}{\partial \vartheta} = -2 \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta \right] \varphi(t)^\top$$

Least-Squares Estimation - Linear Regression (cont.)

- Imposing $\dfrac{\partial J}{\partial \vartheta} = [0\ 0\ \cdots\ 0]$ one gets:

$$-2 \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta\right] \varphi(t)^\top = [0\ 0\ \cdots\ 0] \quad \Longrightarrow \quad \sum_{t=1}^{N} y(t)\, \varphi(t)^\top = \sum_{t=1}^{N} \varphi(t)^\top \vartheta\, \varphi(t)^\top$$

and converting the equality between row-vectors into an equality between column-vectors:

$$\sum_{t=1}^{N} \varphi(t)\, y(t) = \left[\sum_{t=1}^{N} \varphi(t)\, \varphi(t)^\top\right] \vartheta$$

Least-Squares Normal Equations

$(q \text{ equations}, \ q \text{ unknowns})$

- If $\sum_{t=1}^{N} \varphi(t)\, \varphi(t)^\top$ is **non-singular**, it finally follows that:

$$\hat{\vartheta}_N = \left[\sum_{t=1}^{N} \varphi(t)\, \varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\, y(t)$$

Least-Squares Formula

Introduction to Least-Squares Estimation

Geometric Interpretation

Least-Squares Estimation - Geometric Interpretation

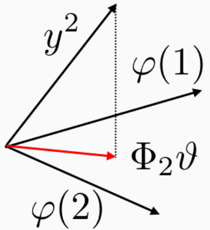
Let:

$$\varepsilon_\vartheta^N = \begin{bmatrix} \varepsilon_\vartheta(1) \\ \vdots \\ \varepsilon_\vartheta(N) \end{bmatrix} \quad y^N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \quad \Phi_N = \begin{bmatrix} \varphi(1)^\top \\ \vdots \\ \varphi(N)^\top \end{bmatrix}$$

Then, we write:

$$J(\vartheta) = \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta\right]^2 = \left\|y^N - \Phi_N\,\vartheta\right\|^2$$

Clearly $\left\|y^N - \Phi_N\,\vartheta\right\|$ is minimum when $y^N - \Phi_N\,\vartheta$ is orthogonal to $\Phi_N\,\vartheta$



Introduction to Least-Squares Estimation

Identifiability Condition

Least-Squares Estimation (cont.)

- Let's verify that $\hat{\vartheta}_N$ is a **minimum** by evaluating the definiteness of the symmetric matrix

$$\left[\frac{d^2 J}{d\vartheta^2}\right]_{i,j} = \frac{\partial^2 J}{\partial\vartheta_i \partial\vartheta_j}, \quad i,j = 1,\ldots,q$$

We have

$$\left(\frac{\partial J}{\partial \vartheta}\right)^\top = 2\left\{\left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]\vartheta - \sum_{t=1}^{N} \varphi(t)\,y(t)\right\}$$

and hence:

$$\frac{d^2 J}{d\vartheta^2} = 2\left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]$$

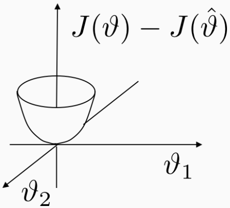
Clearly, this matrix is symmetric and positive semi-definite and thus $\hat{\vartheta}_N$ is a **local minimum** of $J(\vartheta)$.

Least-Squares Estimation (cont.)

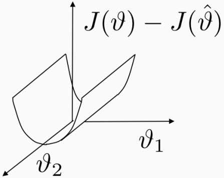
- Therefore, considering the quadratic form

$$J(\vartheta) - J(\hat{\vartheta}) = \frac{1}{2}(\vartheta - \hat{\vartheta})^\top \left.\frac{d^2 J}{d\vartheta^2}\right|_{\hat{\vartheta}} (\vartheta - \hat{\vartheta})$$

two possible scenarios may occur:



$$\det\left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right] \neq 0$$



$$\det\left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right] = 0$$

Least-Squares Estimation (cont.)

- Then:
  - If $\det\left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right] \neq 0 \implies$ $\hat{\vartheta}_N$ is the unique global minimum
  - If $\det\left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right] = 0 \implies$ $\hat{\vartheta}_N$ is one among the infinite global minima
- The condition

$$\det\left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right] \neq 0$$

is called **Identifiability Condition**

Probabilistic Properties of the Least-Squares Estimator

Probabilistic Properties of the Least-Squares Estimator

- Suppose that the identifiability condition is verified:

$$\det\left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right] \neq 0$$

  and then

$$\hat{\vartheta}_N = \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,y(t)$$

- **Assumption:** $y(t) = \varphi(t)^\top \vartheta^\circ + \xi(t)$ where the process is uncorrelated with $u(\cdot)$ and $E[\xi(t)] = 0$

  Therefore:

  We are assuming that the true relationship between $y(t)$ and $u_1(t), \ldots, u_q(t)$ is linear + uncorrelated zero-mean noise

Probabilistic Properties of the Least-Squares Estimator

Bias

Probabilistic Properties of the Least-Squares Estimator (cont.)

Bias:

$$\begin{aligned}\hat{\vartheta}_N &= \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,y(t) \\ &= \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,\left[\varphi(t)^\top \vartheta^\circ + \xi(t)\right] \\ &= \vartheta^\circ + \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,\xi(t)\end{aligned}$$

Hence:

$$\hat{\vartheta}_N - \vartheta^\circ = \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,\xi(t)$$

$$\Longrightarrow E\left(\hat{\vartheta}_N - \vartheta^\circ\right) = \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,E[\xi(t)] = 0$$

$$\Longrightarrow E\left(\hat{\vartheta}_N\right) = \vartheta^\circ$$ **The LS estimator is unbiased**

Probabilistic Properties of the Least-Squares Estimator (cont.)

Important Remark:

- In the bias analysis of the LS estimator we have considered the regression vector $\varphi(t)$ as **known and set** (not random any more).
- On the other hand, carrying out the bias analysis considering $\varphi(s,t)$ as a random vector (hence a function of the result s of a random experiment), would lead to a **biased** LS estimator for any finite value of N.

Probabilistic Properties of the Least-Squares Estimator

Variance

Probabilistic Properties of the Least-Squares Estimator (cont.)

Variance:

Further Assumption: $\xi(t) \sim WN(0, \lambda^2)$

Let us introduce the symmetric matrix $S(N) = \sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top$

Hence:

$$\begin{aligned}
\operatorname{var}\left(\hat{\vartheta}_N\right) &= E\left[\left(\hat{\vartheta}_N - \vartheta^\circ\right)\left(\hat{\vartheta}_N - \vartheta^\circ\right)^\top\right] \\
&= E\left\{\left[S(N)^{-1}\sum_{t=1}^{N}\varphi(t)\,\xi(t)\right]\left[S(N)^{-1}\sum_{s=1}^{N}\varphi(s)\,\xi(s)\right]^\top\right\} \\
&= E\left\{\left[S(N)^{-1}\sum_{t=1}^{N}\varphi(t)\,\xi(t)\right]\left[\sum_{s=1}^{N}\xi(s)\,\varphi(s)^\top S(N)^{-1}\right]\right\} \\
&= S(N)^{-1}\,E\left[\sum_{t=1}^{N}\varphi(t)\,\xi(t)\sum_{s=1}^{N}\xi(s)\,\varphi(s)^\top\right]S(N)^{-1}
\end{aligned}$$

Probabilistic Properties of the Least-Squares Estimator (cont.)

In the product $\sum_{t=1}^{N} \varphi(t)\,\xi(t) \sum_{s=1}^{N} \xi(s)\,\varphi(s)^\top$ we have two kinds of terms:

- $\varphi(t)\,\xi(t)^2\,\varphi(t)^\top$ if $t = s$
- $\varphi(t)\,\xi(t)\,\xi(s)\,\varphi(s)^\top$ if $t \neq s$

But:

$$\xi(t) \sim WN(0, \lambda^2) \implies E\left[\xi(t)\xi(s)\right] = \begin{cases} \lambda^2 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$$

Hence:

$$E\left[\sum_{t=1}^{N} \varphi(t)\,\xi(t) \sum_{s=1}^{N} \xi(s)\,\varphi(s)^\top\right] = \sum_{t=1}^{N} \lambda^2\,\varphi(t)\,\varphi(t)^\top = \lambda^2\,S(N)$$

and thus

$$\text{var}\left(\hat{\vartheta}_N\right) = S(N)^{-1}\,\lambda^2\,S(N)\,S(N)^{-1} = \lambda^2\,S(N)^{-1}$$

Probabilistic Properties of the Least-Squares Estimator

Asymptotic Characteristics

Probabilistic Properties of the Least-Squares Estimator (cont.)

Interpretation:

Assume that ϑ° is scalar and hence also $\varphi(t)$ is scalar as well. Then:

$$y(t) = \varphi(t)\,\vartheta^\circ + \xi(t) = u(t)\,\vartheta^\circ + \xi(t)$$

and hence:

$$\hat{\vartheta}_N = \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,y(t) = \frac{\frac{1}{N}\sum_{t=1}^{N} u(t)\,y(t)}{\frac{1}{N}\sum_{t=1}^{N} u(t)^2}$$

But:

- $\frac{1}{N}\sum_{t=1}^{N} u(t)\,y(t)$ is the sample estimate of the cross-correlation $E\left[u(t)y(t)\right]$
- $\frac{1}{N}\sum_{t=1}^{N} u(t)^2$ is the sample estimate of $E\left[u(t)^2\right]$ (variance if $E\,(u) = 0$).

Probabilistic Properties of the Least-Squares Estimator (cont.)

Moreover:

$$\operatorname{var}\left(\hat{\vartheta}_N\right) = \lambda^2 \, S(N)^{-1} = \frac{1}{N} \frac{\lambda^2}{\dfrac{1}{N}\displaystyle\sum_{t=1}^{N} u(t)^2}$$

Therefore:

- $\operatorname{var}\left(\hat{\vartheta}_N\right)$ grows with λ^2. Hence, estimate's uncertainty grows with data uncertainty
- For given N and λ^2, $\operatorname{var}\left(\hat{\vartheta}_N\right)$ decreases when the sample variance of u increases and this is consistent with intuition: the noise influence on the signal containing the useful information decreases

Probabilistic Properties of the Least-Squares Estimator (cont.)

- $\dfrac{\lambda^2}{\dfrac{1}{N}\displaystyle\sum_{t=1}^{N} u(t)^2}$ is kind of a noise/signal ratio
- If the variance of u is bounded then

$$\lim_{N\to\infty} \operatorname{var}\left(\hat{\vartheta}_N\right) = 0$$

and, owing to the fact that the estimator is unbiased one has:

$$\lim_{N\to\infty} E\left(\left\|\hat{\vartheta}_N - \vartheta^\circ\right\|^2\right) = 0$$

that is, **the LS estimator converges in quadratic mean**

Probabilistic Properties of the Least-Squares Estimator (cont.)

Moreover, we can write

$$\hat{\vartheta}_N = \frac{1}{\sum_{t=1}^{N} u(t)^2} \sum_{t=1}^{N} u(t)\left[u(t)\,\vartheta^\circ + \xi(t)\right] = \vartheta^\circ + \frac{\frac{1}{N}\sum_{t=1}^{N} u(t)\,\xi(t)}{\frac{1}{N}\sum_{t=1}^{N} u(t)^2} \quad \longrightarrow \quad \vartheta^\circ + \frac{E\left[u(t)\,\xi(t)\right]}{E\left[u(t)^2\right]}$$

- If u is deterministic, one has:

$$\vartheta^\circ + \frac{E\left[u(t)\,\xi(t)\right]}{E\left[u(t)^2\right]} = \vartheta^\circ + u(t)\,\frac{E\left[\xi(t)\right]}{E\left[u(t)^2\right]} = \vartheta^\circ$$

- If u is stochastic but uncorrelated with ξ, one has:

$$\vartheta^\circ + \frac{E\left[u(t)\,\xi(t)\right]}{E\left[u(t)^2\right]} = \vartheta^\circ + \frac{E\left[u(t)\right]E\left[\xi(t)\right]}{E\left[u(t)^2\right]} = \vartheta^\circ$$

Least-Squares Estimator: Application Examples and Properties

Matlab live script

Given a realization of a white noise stationary stochastic process, the autocorrelation function and the spectrum can be estimated from the data. A **Matlab live script** illustrate how to perform the estimation.

Steps to retrieve the live script:

- Download as a ZIP archive the whole contents of the folder named "**Lecture9**," available in the "**Class Materials**" file area of the MS Teams course team, and uncompress it in a preferred folder.
- Add the chosen folder and subfolders to the Matlab path.
- Open the live script using the Matlab command:

```
open('L9_LSregressionEX_levelSensors.mlx');
```

- Explore the live script and run it.

267MI –Fall 2023

Lecture 9

Least-Squares Estimation

END