Lectures on Least Squares Methods

Part I: Linear Least Squares

Francesco Santoni

Department of Engineering
University of Perugia, Italy
email: francesco.santoni@unipg.it

February 2023

Table of Contents

Table of Contents

Prerequisites

1. Differential calculus and integrals with multiple variables
2. Linear algebra, from fundamentals to eigenvalues, eigenvectors, and spectral theorem
3. All the previous notions extended to the complex field
4. Fundamentals of probability theory: distributions, expected value, variance, covariance and their properties, Bayes theorem

Table of Contents

Basic concepts and notation

1. The least squares problem arises whenever one has a physical system described by a model in the form $\mathbf{b} = H\boldsymbol{\theta}$
   - H is the response function describing the system, in this case a linear function, i.e. a matrix, with $\boldsymbol{\theta}$ as its argument
   - $\boldsymbol{\theta}$ are the parameters or inputs of the system (independent variables)
   - $\mathbf{b}$ are the observations or outputs of the system (dependent variables)

Basic concepts and notation

1. The least squares problem arises whenever one has a physical system described by a model in the form $\mathbf{b} = H\boldsymbol{\theta}$
   - H is the response function describing the system, in this case a linear function, i.e. a matrix, with $\boldsymbol{\theta}$ as its argument
   - $\boldsymbol{\theta}$ are the parameters or inputs of the system (independent variables)
   - $\mathbf{b}$ are the observations or outputs of the system (dependent variables)
2. Experimentally, observations are affected by uncertainty due to system and measurement noise, and finite measurement resolution: $\mathbf{b} \neq H\boldsymbol{\theta} \Rightarrow \mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$
   - $\mathbf{b}$ is a column vector with N components, representing observations
   - $\boldsymbol{\theta}$ is a column vector with p parameters that are characteristic of the system, and that must be estimated
   - H is a known $N \times p$ matrix; N: number of equations, p number of parameters.
   - $\boldsymbol{\varepsilon}$ is the noise and generally it is assumed: $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, and $\mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I$ [a]

[a] Reminder: $\mathrm{cov}[\mathbf{X}] = \mathrm{E}\left[\mathbf{X}\mathbf{X}^T\right] - \mathrm{E}[\mathbf{X}]\,\mathrm{E}\left[\mathbf{X}^\dagger\right]$

Basic concepts and notation

3. Because of the noise, $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ is in general an inconsistent system of N equations
   - One then seeks the optimal solution that minimizes the cost function

$$\phi(\boldsymbol{\theta}) = ||\mathbf{b} - H\boldsymbol{\theta}||^2 = (\mathbf{b} - H\boldsymbol{\theta})^T (\mathbf{b} - H\boldsymbol{\theta})$$

   - Thus, the least squares estimator is $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} ||\mathbf{b} - H\boldsymbol{\theta}||^2$

Basic concepts and notation

3. Because of the noise, $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ is in general an inconsistent system of N equations
   - One then seeks the optimal solution that minimizes the cost function

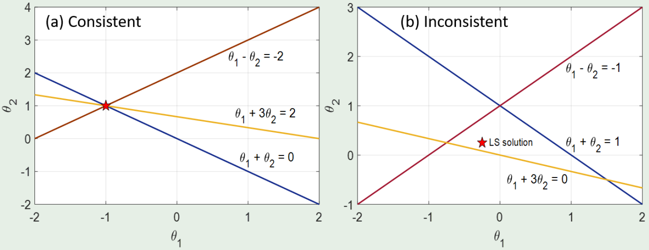
$$\phi(\boldsymbol{\theta}) = ||\mathbf{b} - H\boldsymbol{\theta}||^2 = (\mathbf{b} - H\boldsymbol{\theta})^T (\mathbf{b} - H\boldsymbol{\theta})$$

   - Thus, the least squares estimator is $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} ||\mathbf{b} - H\boldsymbol{\theta}||^2$

Example

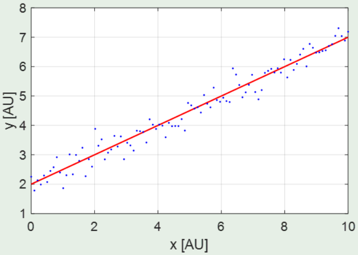
$$(a): \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \\ 2 \end{pmatrix}$$

$$(b): \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$



At this level, only **b** is affected by the uncertainty. When **b** is changed, lines are just translated, slopes are not changed. When also H is affected by the uncertainty, slopes change: this is the Total Least Squares method, discussed later on.

LS regression examples



Linear regression, N observations, $p = 2$ parameters:

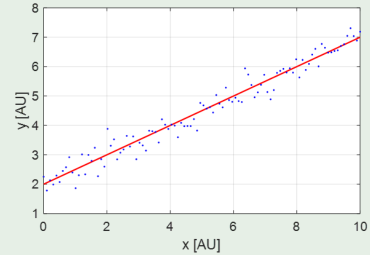
$$\mathbf{y} = m\mathbf{x} + q = \begin{pmatrix} \mathbf{x} & \mathbf{1} \end{pmatrix} \begin{pmatrix} m \\ q \end{pmatrix}$$

$$\mathbf{b} \equiv \mathbf{y}$$

$$H \equiv \begin{pmatrix} \mathbf{x} & \mathbf{1} \end{pmatrix}$$

$$\boldsymbol{\theta} \equiv \begin{pmatrix} m \\ q \end{pmatrix}$$

LS regression examples



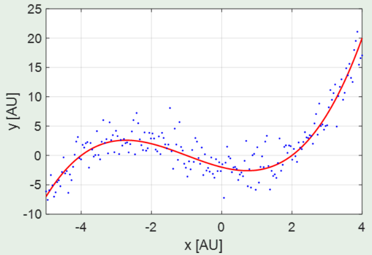
Linear regression, N observations, $p = 2$ parameters:

$$\mathbf{y} = m\mathbf{x} + q = \begin{pmatrix} \mathbf{x} & \mathbf{1} \end{pmatrix} \begin{pmatrix} m \\ q \end{pmatrix}$$

$$\mathbf{b} \equiv \mathbf{y}$$

$$H \equiv \begin{pmatrix} \mathbf{x} & \mathbf{1} \end{pmatrix}$$

$$\boldsymbol{\theta} \equiv \begin{pmatrix} m \\ q \end{pmatrix}$$



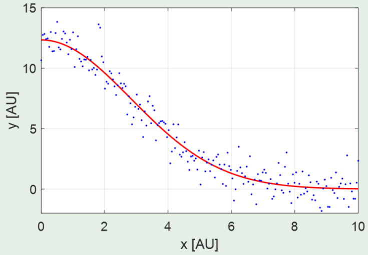
Polinomial regression, 3rd degree, N observations, $p = 4$ parameters:

$$\mathbf{y} = c_0 + c_1\mathbf{x} + c_2\mathbf{x}^2 + c_3\mathbf{x}^2 = \begin{pmatrix} \mathbf{x} & \mathbf{1} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

$$\mathbf{b} \equiv \mathbf{y} \qquad H \equiv \begin{pmatrix} \mathbf{1} & \mathbf{x} & \mathbf{x}^2 & \mathbf{x}^3 \end{pmatrix}$$

$$\boldsymbol{\theta} \equiv \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

LS regression examples



Exponential regression, N observations, $p = 2$ parameters:

$$\mathbf{y} = A e^{b\mathbf{x}^2}$$

A non-linear problem. It can be linearized by using logarithms

$$\log \mathbf{y} = \log A + b\mathbf{x}^2 =$$
$$= C + b\mathbf{x}^2 = \begin{pmatrix} 1 & \mathbf{x}^2 \end{pmatrix} \begin{pmatrix} C \\ b \end{pmatrix}$$

Warning: the uncertainty estimated for C will propagate non-linearly on A

Table of Contents

General terminology for estimators

1. A sample is a series of N observations $\mathbf{z} = (z_1 \cdots z_N)$ of a random variable $\mathbf{Z}$

General terminology for estimators

1. A sample is a series of N observations $\mathbf{z} = (z_1 \cdots z_N)$ of a random variable $\mathbf{Z}$
2. A statistic is any function of the observations $g(\mathbf{z}) = g(z_1 \cdots z_N)$ not dependent on unknown parameters

General terminology for estimators

1. A sample is a series of N observations $\mathbf{z} = (z_1 \cdots z_N)$ of a random variable $\mathbf{Z}$
2. A statistic is any function of the observations $g(\mathbf{z}) = g(z_1 \cdots z_N)$ not dependent on unknown parameters
3. Typically, formulating a hypothesis means assuming that observations are extracted from a probability density function p.d.f. $f(\mathbf{z} \,|\, \boldsymbol{\theta})$ dependent on some parameters $\boldsymbol{\theta} = (\theta_1 \cdots \theta_N)$ that must be determined

General terminology for estimators

1. An estimator is a statistic used to estimate the parameters of a p.d.f. The estimator of $\boldsymbol{\theta}$ is typically denoted by the symbol $\hat{\boldsymbol{\theta}}$

General terminology for estimators

1. An estimator is a statistic used to estimate the parameters of a p.d.f. The estimator of $\boldsymbol{\theta}$ is typically denoted by the symbol $\hat{\boldsymbol{\theta}}$
2. An estimate is the value of an estimator calculated for a given sample

General terminology for estimators

1. An estimator is a statistic used to estimate the parameters of a p.d.f. The estimator of $\boldsymbol{\theta}$ is typically denoted by the symbol $\hat{\boldsymbol{\theta}}$
2. An estimate is the value of an estimator calculated for a given sample
3. The procedure by which one comes to an estimate of the $\boldsymbol{\theta}$ parameters for a given sample is also called parameter fitting

General terminology for estimators

1. An estimator is a statistic used to estimate the parameters of a p.d.f. The estimator of $\boldsymbol{\theta}$ is typically denoted by the symbol $\hat{\boldsymbol{\theta}}$
2. An estimate is the value of an estimator calculated for a given sample
3. The procedure by which one comes to an estimate of the $\boldsymbol{\theta}$ parameters for a given sample is also called parameter fitting
4. The bias (or polarization) of an estimator is defined as the difference: $\mathbf{b} = \mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] - \boldsymbol{\theta}$

General terminology for estimators

1. An estimator is a statistic used to estimate the parameters of a p.d.f. The estimator of $\boldsymbol{\theta}$ is typically denoted by the symbol $\hat{\boldsymbol{\theta}}$
2. An estimate is the value of an estimator calculated for a given sample
3. The procedure by which one comes to an estimate of the $\boldsymbol{\theta}$ parameters for a given sample is also called parameter fitting
4. The bias (or polarization) of an estimator is defined as the difference: $\mathbf{b} = \mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] - \boldsymbol{\theta}$
5. An estimator is termed biased (or polarized) when $\mathbf{b} \neq 0$, otherwise it is termed unbiased (or non-polarized)

General terminology for estimators

1. An estimator is a statistic used to estimate the parameters of a p.d.f. The estimator of $\boldsymbol{\theta}$ is typically denoted by the symbol $\hat{\boldsymbol{\theta}}$
2. An estimate is the value of an estimator calculated for a given sample
3. The procedure by which one comes to an estimate of the $\boldsymbol{\theta}$ parameters for a given sample is also called parameter fitting
4. The bias (or polarization) of an estimator is defined as the difference: $\mathbf{b} = \mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] - \boldsymbol{\theta}$
5. An estimator is termed biased (or polarized) when $\mathbf{b} \neq 0$, otherwise it is termed unbiased (or non-polarized)
6. Tipically, observations are independent, hence the p.d.f. is $f_{sample} = f_1(z_1) f_2(z_2) \ldots f_N(z_N)$. If the sample consists of repeated observations of the same variable, then $f_1 = f_2 = \ldots = f_N = f$, and:

$$\mathrm{E}\left[\hat{\theta}(\mathbf{z})\right] = \int_D \hat{\theta}(\mathbf{z}) f_{sample}(\mathbf{z}|\theta)\, d\mathbf{z} = \int \ldots \int \hat{\theta}(\mathbf{z}) f_1(z_1) \ldots f_N(z_N)\, dz_1 \ldots dz_N$$

General terminology for estimators

Unbiased estimator example: the sample (or arithmetic) mean

The sample mean is an unbiased estimator of the expected value of a p.d.f. $f(z)$, given a sample of N observations z_i

$$\mu = \mathrm{E}[z] = \int z f(z)\, dz$$

$$\hat{\mu} = \bar{z} = \frac{1}{N} \sum_{i=1}^{N} z_i$$

$$\mathrm{E}[\hat{\mu}(\mathbf{z})] = \mathrm{E}\left[\frac{1}{N} \sum_{i=1}^{N} z_i\right] = \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}[z_i] = \frac{1}{N} \sum_{i=1}^{N} \mu = \frac{1}{N} N\mu = \mu$$

$$b = \mathrm{E}[\hat{\mu}(\mathbf{z})] - \mu = \mu - \mu = 0$$

Biased estimator example: the sample variance

The sample variance

$$s^2 = \frac{1}{N} \sum_{i=1}^{N} (z_i - \bar{z})^2$$

is a biased estimator of the variance σ^2, indeed, without performing all calculations

$$\mathrm{E}\left[s^2\right] = \frac{N-1}{N}\sigma^2$$

An unbiased estimator can be easily obtained:

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (z_i - \bar{z})^2 = \frac{N}{N-1} s^2$$

$$\mathrm{E}\left[S^2\right] = \frac{N}{N-1}\mathrm{E}\left[s^2\right] = \sigma^2$$

Table of Contents

Review of linear algebra

1. Linearly independent vectors: $\sum_i c_i \mathbf{v}_i = 0 \Leftrightarrow \forall i, \;\; c_i = 0$

Review of linear algebra

1. Linearly independent vectors: $\sum_i c_i \mathbf{v}_i = 0 \Leftrightarrow \forall i,\ c_i = 0$
2. The rank of a matrix $A \in \mathbb{C}^{m\times n}$ is the maximum number of linearly independent columns or rows: $\mathrm{rank}\,(A) \leq \min\,(m, n)$; $\mathrm{rank}\,(A) = \mathrm{rank}\,\left(A^{\dagger}\right)$.

Review of linear algebra

1. Linearly independent vectors: $\sum_i c_i \mathbf{v}_i = 0 \Leftrightarrow \forall i,\ \ c_i = 0$

2. The rank of a matrix $A \in \mathbb{C}^{m \times n}$ is the maximum number of linearly independent columns or rows: $\mathrm{rank}\,(A) \leq \min\,(m, n)$; $\mathrm{rank}\,(A) = \mathrm{rank}\,(A^{\dagger})$.

3. The rank of a matrix is the dimension of the space generated by its columns: $\mathrm{rank}\,(A) = \dim\,[\mathrm{Span}\,(\mathbf{a}_1, \ldots, \mathbf{a}_n)]$, $\mathrm{Span}\,(\mathbf{a}_1, \ldots, \mathbf{a}_n) \equiv \left\{ \mathbf{v} : \mathbf{v} = \sum_i c_i \mathbf{a}_i \right\}$

Review of linear algebra

1. Linearly independent vectors: $\sum_i c_i \mathbf{v}_i = 0 \Leftrightarrow \forall i,\ \ c_i = 0$
2. The rank of a matrix $A \in \mathbb{C}^{m \times n}$ is the maximum number of linearly independent columns or rows: $\mathrm{rank}\,(A) \leq \min\,(m, n)$; $\mathrm{rank}\,(A) = \mathrm{rank}\,\left(A^{\dagger}\right)$.
3. The rank of a matrix is the dimension of the space generated by its columns: $\mathrm{rank}\,(A) = \dim\,[\mathrm{Span}\,(\mathbf{a}_1, \ldots, \mathbf{a}_n)]$, $\mathrm{Span}\,(\mathbf{a}_1, \ldots, \mathbf{a}_n) \equiv \left\{ \mathbf{v} : \mathbf{v} = \sum_i c_i \mathbf{a}_i \right\}$
4. Kernel of A:
$\ker\,(A) \equiv \{\mathbf{v} : A\mathbf{v} = \mathbf{0}\}$, $\quad \forall A, (\mathbf{v} = \mathbf{0}) \in \ker\,(A)$, $\ \ker\,(A) \equiv \{\mathbf{0}\} \Rightarrow \dim\,[\ker\,(A)] = 0$
$\dim\,[\ker\,(A)]$ is called the nullity of A.

1. Linearly independent vectors: $\sum_i c_i \mathbf{v}_i = 0 \Leftrightarrow \forall i,\ c_i = 0$
2. The rank of a matrix $A \in \mathbb{C}^{m\times n}$ is the maximum number of linearly independent columns or rows: $\mathrm{rank}\,(A) \leq \min\,(m, n)$; $\mathrm{rank}\,(A) = \mathrm{rank}\,\left(A^\dagger\right)$.
3. The rank of a matrix is the dimension of the space generated by its columns: $\mathrm{rank}\,(A) = \dim\left[\mathrm{Span}\,(\mathbf{a}_1, \ldots, \mathbf{a}_n)\right]$, $\mathrm{Span}\,(\mathbf{a}_1, \ldots, \mathbf{a}_n) \equiv \left\{\mathbf{v} : \mathbf{v} = \sum_i c_i \mathbf{a}_i\right\}$
4. Kernel of A:
$\ker\,(A) \equiv \{\mathbf{v} : A\mathbf{v} = \mathbf{0}\}$, $\quad \forall A, (\mathbf{v} = \mathbf{0}) \in \ker\,(A)$, $\ker\,(A) \equiv \{\mathbf{0}\} \Rightarrow \dim\left[\ker\,(A)\right] = 0$
$\dim\left[\ker\,(A)\right]$ is called the nullity of A.

Rank-nullity theorem

$$\forall A \in \mathbb{C}^{m\times n},\ \mathrm{rank}\,(A) + \dim\left[\ker\,(A)\right] = n$$

A useful lemma

$$\forall A \in \mathbb{C}^{m \times n},\ \mathrm{rank}\,(A) = \mathrm{rank}\left(A^\dagger A\right)$$

Review of linear algebra

A useful lemma

$$\forall A \in \mathbb{C}^{m\times n},\ \mathrm{rank}\left(A\right) = \mathrm{rank}\left(A^{\dagger}A\right)$$

Proof.

- From the rank-nullity theorem, it follows that:

$$\mathrm{rank}\left(A\right) + \dim\left[\ker\left(A\right)\right] = n = \mathrm{rank}\left(A^{\dagger}A\right) + \dim\left[\ker\left(A^{\dagger}A\right)\right]$$

□

A useful lemma

$$\forall A \in \mathbb{C}^{m\times n},\ \mathrm{rank}\,(A) = \mathrm{rank}\left(A^{\dagger}A\right)$$

Proof.

- From the rank-nullity theorem, it follows that:

$$\mathrm{rank}\,(A) + \dim\left[\ker\left(A\right)\right] = n = \mathrm{rank}\left(A^{\dagger}A\right) + \dim\left[\ker\left(A^{\dagger}A\right)\right]$$

- Then, one can prove that ranks are equal by proving that kernels are the same, i.e. by showing that if $\mathbf{v} \in \ker(A)$, then $\mathbf{v} \in \ker\left(A^{\dagger}A\right)$, and *vice versa*:

$$\mathbf{v} \in \ker(A) \Rightarrow A\mathbf{v} = \mathbf{0} \Rightarrow A^{\dagger}A\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v} \in \ker\left(A^{\dagger}A\right)$$

$$\mathbf{v} \in \ker\left(A^{\dagger}A\right) \Rightarrow A^{\dagger}A\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v}^{\dagger}A^{\dagger}A\mathbf{v} = 0 \Rightarrow \|A\mathbf{v}\|^2 = 0 \Rightarrow A\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v} \in \ker(A)$$

□

Table of Contents

Ordinary Least Squares - OLS

OLS assumptions

- System $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ has more equations the parameters $(N \geq p)$
- H is a full-rank matrix: $\mathrm{rank}\,(H) = p$.

Ordinary Least Squares - OLS

OLS assumptions

- System $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ has more equations the parameters ($N \geq p$)
- H is a full-rank matrix: $\text{rank}\,(H) = p$.

Consistent system

- When $\boldsymbol{\varepsilon} = 0$ the system is:

$$\mathbf{b} = H\boldsymbol{\theta} = \begin{pmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_p \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} = \sum_{i=1}^{p} \theta_i \mathbf{h}_i$$

- The system has a solution when $\mathbf{b}$ is a linear combination of the columns of H:

$$\mathbf{b} \in \text{Span}\,(H) \Leftrightarrow rank\,(H) = \text{rank}\left[\begin{pmatrix} H & \mathbf{b} \end{pmatrix}\right]$$

When H is full-rank, the solution is unique.

Ordinary Least Squares - OLS

Inconsistent system

- In general $\boldsymbol{\varepsilon} \neq 0$ and the sistem is inconsistent: $rank\,(H) \neq \operatorname{rank}\left[\begin{pmatrix} H & \mathbf{b} \end{pmatrix}\right]$
- According to the lemma on the rank of $H^\dagger H$: $\operatorname{rank}(H) = p = \operatorname{rank}\left(H^\dagger H\right)$
- $H^\dagger H$ is a full-rank square $p \times p$ matrix, hence it is invertible

Ordinary Least Squares - OLS

Inconsistent system

- In general $\boldsymbol{\varepsilon} \neq 0$ and the sistem is inconsistent: $rank\,(H) \neq \mathrm{rank}\left[\begin{pmatrix} H & \mathbf{b} \end{pmatrix}\right]$
- According to the lemma on the rank of $H^{\dagger}H$: $\mathrm{rank}\,(H) = p = \mathrm{rank}\,(H^{\dagger}H)$
- $H^{\dagger}H$ is a full-rank square $p \times p$ matrix, hence it is invertible

Associated consistent system

- For the previous assumptions, the following system is consistent:

$$H^{\dagger}\mathbf{b} = H^{\dagger}H\boldsymbol{\theta} \Rightarrow \hat{\boldsymbol{\theta}} = \left(H^{\dagger}H\right)^{-1}H^{\dagger}\mathbf{b} = H^{+}\mathbf{b}$$

- The pseudo-inverse or Moore-Penrose matrix has been introduced:

$$H^{+} = \left(H^{\dagger}H\right)^{-1}H^{\dagger} \Rightarrow H^{+}H = I,\ HH^{+} \neq I$$

- H is a $N \times p$ matrix, and H^{+} is $p \times N$. When H is square ($N = p$), then $H^{+} = H^{-1}$

Ordinary Least Squares - OLS

Inconsistent system

- In general $\boldsymbol{\varepsilon} \neq 0$ and the sistem is inconsistent: $rank\,(H) \neq \text{rank}\left[\begin{pmatrix} H & \mathbf{b} \end{pmatrix}\right]$
- According to the lemma on the rank of $H^\dagger H$: $\text{rank}\,(H) = p = \text{rank}\left(H^\dagger H\right)$
- $H^\dagger H$ is a full-rank square $p \times p$ matrix, hence it is invertible

Associated consistent system

- For the previous assumptions, the following system is consistent:

$$H^\dagger \mathbf{b} = H^\dagger H \boldsymbol{\theta} \Rightarrow \hat{\boldsymbol{\theta}} = \left(H^\dagger H\right)^{-1} H^\dagger \mathbf{b} = H^+ \mathbf{b}$$

- The pseudo-inverse or Moore-Penrose matrix has been introduced:

$$H^+ = \left(H^\dagger H\right)^{-1} H^\dagger \Rightarrow H^+ H = I, \;\; H H^+ \neq I$$

- H is a $N \times p$ matrix, and H^+ is $p \times N$. When H is square ($N = p$), then $H^+ = H^{-1}$

What does the solution $\hat{\boldsymbol{\theta}} = \left(H^\dagger H\right)^{-1} H^\dagger \mathbf{b} = H^+ \mathbf{b}$ mean?

Ordinary Least Squares - OLS

OLS problem

- Full-rank (p) inconsistent system: $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} ||\mathbf{b} - H\boldsymbol{\theta}||^2$
- Associated consistent system: $H^\dagger \mathbf{b} = H^\dagger H\boldsymbol{\theta}$
- Cost function:

$$\phi(\boldsymbol{\theta}) = ||\mathbf{b} - H\boldsymbol{\theta}||^2 = (\mathbf{b} - H\boldsymbol{\theta})^\dagger (\mathbf{b} - H\boldsymbol{\theta}) =$$
$$= \boldsymbol{\theta}^\dagger H^\dagger H\boldsymbol{\theta} + \mathbf{b}^\dagger\mathbf{b} - \mathbf{b}^\dagger H\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger H^\dagger \mathbf{b} = \boldsymbol{\theta}^\dagger H^\dagger H\boldsymbol{\theta} + \mathbf{b}^\dagger\mathbf{b} - 2\mathrm{Re}\left(\boldsymbol{\theta}^\dagger H^\dagger \mathbf{b}\right)$$

Ordinary Least Squares - OLS

OLS problem

- Full-rank (p) inconsistent system: $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\mathbf{b} - H\boldsymbol{\theta}\|^2$
- Associated consistent system: $H^\dagger \mathbf{b} = H^\dagger H\boldsymbol{\theta}$
- Cost function:

$$\phi(\boldsymbol{\theta}) = \|\mathbf{b} - H\boldsymbol{\theta}\|^2 = (\mathbf{b} - H\boldsymbol{\theta})^\dagger (\mathbf{b} - H\boldsymbol{\theta}) =$$
$$= \boldsymbol{\theta}^\dagger H^\dagger H\boldsymbol{\theta} + \mathbf{b}^\dagger\mathbf{b} - \mathbf{b}^\dagger H\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger H^\dagger\mathbf{b} = \boldsymbol{\theta}^\dagger H^\dagger H\boldsymbol{\theta} + \mathbf{b}^\dagger\mathbf{b} - 2\mathrm{Re}\left(\boldsymbol{\theta}^\dagger H^\dagger\mathbf{b}\right)$$

OLS solution of the full-rank inconsistent system

The solution of the associated consistent system:

$$\hat{\boldsymbol{\theta}} = \left(H^\dagger H\right)^{-1} H^\dagger\mathbf{b} = H^+\mathbf{b}$$

is also the solution that minimizes the cost function

Ordinary Least Squares - OLS

OLS solution of the full-rank inconsistent system

The solution of the associated consistent system:

$$\hat{\boldsymbol{\theta}} = \left(H^{\dagger}H\right)^{-1}H^{\dagger}\mathbf{b} = H^{+}\mathbf{b}$$

is also the solution that minimizes the cost function $\phi(\boldsymbol{\theta}) = \|\mathbf{b} - H\boldsymbol{\theta}\|^2$

Ordinary Least Squares - OLS

OLS solution of the full-rank inconsistent system

The solution of the associated consistent system:

$$\hat{\boldsymbol{\theta}} = \left(H^{\dagger} H\right)^{-1} H^{\dagger} \mathbf{b} = H^{+} \mathbf{b}$$

is also the solution that minimizes the cost function $\phi(\boldsymbol{\theta}) = \|\mathbf{b} - H\boldsymbol{\theta}\|^2$

Proof.

We give a simple proof for the real case. The complex case will be proved later in the more general context of singular value decomposition. When H is real:

$$\phi(\boldsymbol{\theta}) = \boldsymbol{\theta}^T H^T H \boldsymbol{\theta} + \mathbf{b}^T \mathbf{b} - 2\boldsymbol{\theta}^T H^T \mathbf{b} = \sum_{jkl} \theta_j H_{kj} H_{kl} \theta_l + \sum_j b_j^2 - 2 \sum_{jk} \theta_j H_{kj} b_k$$

The minimum is attained where the jacobian matrix (the gradient in this case) is zero:

$$\frac{\partial \phi}{\partial \theta_i} = \sum_{jkl} (\delta_{ij} H_{kj} H_{kl} \theta_l + \theta_j H_{kj} H_{kl} \delta_{il}) - 2 \sum_{jk} \delta_{ij} H_{kj} b_k = 2 \sum_{jk} H_{ji} H_{jk} \theta_k - 2 \sum_j H_{ji} b_j$$

OLS solution of the full-rank inconsistent system

The solution of the associated consistent system:

$$\hat{\boldsymbol{\theta}} = \left(H^{\dagger}H\right)^{-1} H^{\dagger}\mathbf{b} = H^{+}\mathbf{b}$$

is also the solution that minimizes the cost function $\phi\left(\boldsymbol{\theta}\right) = \|\mathbf{b} - H\boldsymbol{\theta}\|^2$

Proof.

The minimum is attained where the jacobian matrix (the gradient in this case) is zero:

$$\frac{\partial \phi}{\partial \theta_i} = 2\left(H^T H\boldsymbol{\theta}\right)_i - 2(H\mathbf{b})_i \Rightarrow \frac{\partial \phi}{\partial \boldsymbol{\theta}} = 2H^T H\boldsymbol{\theta} - 2H\mathbf{b} = 0 \Rightarrow H^T H\boldsymbol{\theta} = H\mathbf{b}$$

from which the solution follows. □

Table of Contents

We assumed: $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, and $\mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I$, $N \geq p$
Observations $\mathbf{b}$ are homoscedastic (from the greek *homo* "same" *skedasis* "dispersion", i.e. they all have the same variance) and uncorrelated

Expected value of the OLS estimator

The OLS estimator $\boldsymbol{\theta} = H^{+}\mathbf{b}$ is unbiased: $\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$

Properties of the OLS estimator

We assumed: $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, and $\mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I$, $N \geq p$
Observations $\mathbf{b}$ are homoscedastic (from the greek *homo* "same" *skedasis* "dispersion", i.e. they all have the same variance) and uncorrelated

Expected value of the OLS estimator

The OLS estimator $\boldsymbol{\theta} = H^{+}\mathbf{b}$ is unbiased: $\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$

Proof.

By a straightforward calculation:

$$\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \mathrm{E}\left[\left(H^{\dagger}H\right)^{-1}H^{\dagger}\mathbf{b}\right] = \mathrm{E}\left[\left(H^{\dagger}H\right)^{-1}H^{\dagger}(H\boldsymbol{\theta} + \boldsymbol{\varepsilon})\right] =$$

$$= \left(H^{\dagger}H\right)^{-1}H^{\dagger}H\mathrm{E}[\boldsymbol{\theta}] + \left(H^{\dagger}H\right)^{-1}H^{\dagger}\mathrm{E}[\boldsymbol{\varepsilon}] = \boldsymbol{\theta}$$

□

Properties of the OLS estimator

We assumed: $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, and $\mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I$, $N \geq p$
Observations **b** are homoscedastic (from the greek *homo* "same" *skedasis* "dispersion", i.e. they all have the same variance) and uncorrelated

Covariance of the OLS estimator

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2 \left(H^\dagger H\right)^{-1}$$

Properties of the OLS estimator

We assumed: $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, and $\mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I$, $N \geq p$
Observations $\mathbf{b}$ are homoscedastic (from the greek *homo* "same" *skedasis* "dispersion", i.e. they all have the same variance) and uncorrelated

Covariance of the OLS estimator

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2 \left(H^\dagger H\right)^{-1}$$

Proof.

By a straightforward calculation[a]:

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \mathrm{cov}\left[\left(H^\dagger H\right)^{-1} H^\dagger \mathbf{b}\right] = \mathrm{cov}\left[\boldsymbol{\theta} + \left(H^\dagger H\right)^{-1} H^\dagger \boldsymbol{\varepsilon}\right] =$$

$$= \left(H^\dagger H\right)^{-1} H^\dagger \mathrm{cov}\left[\boldsymbol{\varepsilon}\right] H \left(H^\dagger H\right)^{-1} = \left(H^\dagger H\right)^{-1} H^\dagger \sigma^2 I H \left(H^\dagger H\right)^{-1} = \sigma^2 \left(H^\dagger H\right)^{-1}$$

□

[a] Reminder: $\mathrm{cov}\left[A\mathbf{X}\right] = A\,\mathrm{cov}\left[\mathbf{X}\right] A^\dagger$

Properties of the OLS estimator

A reminder on positive semi-definite and definite matrices

- A Hermitian matrix $A = A^\dagger$ is positive semi-definite (respectively definite) iff $\mathbf{z}^\dagger A\mathbf{z} \geq 0$ (respectively $\mathbf{z}^\dagger A\mathbf{z} > 0$), $\forall \mathbf{z} \in \mathbb{C}^n$

Properties of the OLS estimator

A reminder on positive semi-definite and definite matrices

- A Hermitian matrix $A = A^\dagger$ is positive semi-definite (respectively definite) iff $\mathbf{z}^\dagger A\mathbf{z} \geq 0$ (respectively $\mathbf{z}^\dagger A\mathbf{z} > 0$), $\forall \mathbf{z} \in \mathbb{C}^n$
- The diagonal elements of a positive semi-definite (respectively definite) matrix A are always real positive semi-definite (respectively definite) values, indeed, by using the standard basis on $\mathbb{C}^n$, $\mathbf{z} \equiv \mathbf{e}_i$: $A_{ii} = \mathbf{e}_i^T A\mathbf{e}_i \geq 0$ (respectively $A_{ii} > 0$).

Properties of the OLS estimator

A reminder on positive semi-definite and definite matrices

- A Hermitian matrix $A = A^\dagger$ is positive semi-definite (respectively definite) iff $\mathbf{z}^\dagger A\mathbf{z} \geq 0$ (respectively $\mathbf{z}^\dagger A\mathbf{z} > 0$), $\forall \mathbf{z} \in \mathbb{C}^n$
- The diagonal elements of a positive semi-definite (respectively definite) matrix A are always real positive semi-definite (respectively definite) values, indeed, by using the standard basis on $\mathbb{C}^n$, $\mathbf{z} \equiv \mathbf{e}_i$: $A_{ii} = \mathbf{e}_i^T A\mathbf{e}_i \geq 0$ (respectively $A_{ii} > 0$).
- A matrix of the form $A^\dagger A$ is always positive semi-definite, indeed $\mathbf{z}^\dagger A^\dagger A\mathbf{z} = \|A\mathbf{z}\|^2 \geq 0$ by definition of norm.

Properties of the OLS estimator

Assumptions: $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, $\mathrm{E}\left[\boldsymbol{\varepsilon}\right] = 0$, and $\mathrm{cov}\left[\boldsymbol{\varepsilon}\right] = \sigma^2 I$, $N \geq p$

Gauss-Markov theorem

- The OLS estimator $\hat{\boldsymbol{\theta}}$ is the unbiased linear estimator with minimum variance, i.e., given any other unbiased linear estimator $\hat{\boldsymbol{\theta}}_L = C\mathbf{b}$, then

$$\mathrm{var}\left[\hat{\boldsymbol{\theta}}_L\right] \geq \mathrm{var}\left[\hat{\boldsymbol{\theta}}\right]$$

- The OLS estimator $\hat{\boldsymbol{\theta}}$ is the best linear unbiased estimator (BLUE), i.e., it has minimum squared error:

$$\mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}\right\|^2\right] \geq \mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|^2\right]$$

Properties of the OLS estimator

Proof.

- For the first point: first we need an unbiased $\hat{\boldsymbol{\theta}}_L$. C can always be written as $C = H^{+} + D$, for a suitable D:

$$\mathrm{E}\left[\hat{\boldsymbol{\theta}}_L\right] = \mathrm{E}\left[C\mathbf{b}\right] = \mathrm{E}\left[\left(\left(H^{\dagger}H\right)^{-1}H^{\dagger} + D\right)(H\boldsymbol{\theta} + \boldsymbol{\varepsilon})\right]$$

$$= \left(\left(H^{\dagger}H\right)^{-1}H^{\dagger} + D\right)H\boldsymbol{\theta} = (I + DH)\,\boldsymbol{\theta}$$

Hence $\hat{\boldsymbol{\theta}}_L$ is unbiased iff $DH = 0$. Then:

$$\mathrm{var}\left[\hat{\boldsymbol{\theta}}_L\right] = \mathrm{diag}\left(\mathrm{cov}\left[C\mathbf{b}\right]\right) = \mathrm{diag}\left(C\,\mathrm{cov}\left[\mathbf{b}\right]C^{\dagger}\right) = \mathrm{diag}\left(\sigma^2 CC^{\dagger}\right)$$

$$\sigma^2 CC^{\dagger} = \sigma^2\left(\left(H^{\dagger}H\right)^{-1}H^{\dagger} + D\right)\left(H\left(H^{\dagger}H\right)^{-1} + D^{\dagger}\right)$$

$$= \sigma^2\left(H^{\dagger}H\right)^{-1} + \sigma^2\left(H^{\dagger}H\right)^{-1}(DH)^{\dagger} + \sigma^2 DH\left(H^{\dagger}H\right)^{-1} + \sigma^2 DD^{\dagger}$$

$$= \mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] + \sigma^2 DD^{\dagger}$$

Since $DD^{\dagger}$ is positive semi-definite, then $\mathrm{var}\left[\hat{\boldsymbol{\theta}}_L\right] \geq \mathrm{var}\left[\hat{\boldsymbol{\theta}}\right]$

Properties of the OLS estimator

Proof.

- The second point follows from the first, and from the fact that $\hat{\boldsymbol{\theta}}_L$ and $\hat{\boldsymbol{\theta}}$ are unbiased.

$$\mathrm{var}\left[\hat{\boldsymbol{\theta}}_L\right] \geq \mathrm{var}\left[\hat{\boldsymbol{\theta}}\right]$$

$$\sum_i \mathrm{var}\left[\hat{\theta}_{L,i}\right] = \mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}}_L - \mathrm{E}\left[\hat{\boldsymbol{\theta}}_L\right]\right\|^2\right] \geq \mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}} - \mathrm{E}\left[\hat{\boldsymbol{\theta}}\right]\right\|^2\right] = \sum_i \mathrm{var}\left[\hat{\theta}_i\right]$$

$$\mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}\right\|^2\right] \geq \mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|^2\right] \qquad \square$$

Table of Contents

Weighted least squares

- Let us now consider the case: $\operatorname{cov}[\mathbf{b}] = \operatorname{cov}[\boldsymbol{\varepsilon}] = \boldsymbol{\Sigma} = \sigma_i^2\delta_{ij}$ (i.e. $\boldsymbol{\Sigma}$ is a diagonal matrix). When the variances σ_i^2 have different values, the random variable is called heteroscedastic.

Weighted least squares

- Let us now consider the case: $\text{cov}[\mathbf{b}] = \text{cov}[\boldsymbol{\varepsilon}] = \boldsymbol{\Sigma} = \sigma_i^2 \delta_{ij}$ (i.e. $\boldsymbol{\Sigma}$ is a diagonal matrix). When the variances σ_i^2 have different values, the random variable is called heteroscedastic.

- Without the homoscedasticity assumption, the Gauss-Markov theorem is not valid, but the heteroscedastic $\mathbf{b}$ can be suitably corrected in order to become homoscedastic.

Weighted least squares

- Let us now consider the case: $\text{cov}[\mathbf{b}] = \text{cov}[\boldsymbol{\varepsilon}] = \Sigma = \sigma_i^2\delta_{ij}$ (i.e. Σ is a diagonal matrix). When the variances σ_i^2 have different values, the random variable is called heteroscedastic.
- Without the homoscedasticity assumption, the Gauss-Markov theorem is not valid, but the heteroscedastic $\mathbf{b}$ can be suitably corrected in order to become homoscedastic.
- Let us define the weight matrix $W = \frac{1}{\sigma_i^2}\delta_{ij}$, and the weighted observations $\mathbf{b}_w = W^{\frac{1}{2}}\mathbf{b}$.

Weighted least squares

- Let us now consider the case: $\text{cov}\,[\mathbf{b}] = \text{cov}\,[\boldsymbol{\varepsilon}] = \boldsymbol{\Sigma} = \sigma_i^2\delta_{ij}$ (i.e. $\boldsymbol{\Sigma}$ is a diagonal matrix). When the variances σ_i^2 have different values, the random variable is called heteroscedastic.

- Without the homoscedasticity assumption, the Gauss-Markov theorem is not valid, but the heteroscedastic $\mathbf{b}$ can be suitably corrected in order to become homoscedastic.

- Let us define the weight matrix $W = \frac{1}{\sigma_i^2}\delta_{ij}$, and the weighted observations $\mathbf{b}_w = W^{\frac{1}{2}}\mathbf{b}$.

- Accordingly: $H_w = W^{\frac{1}{2}}H$, $\boldsymbol{\varepsilon}_w = W^{\frac{1}{2}}\boldsymbol{\varepsilon}$, and

$$\text{cov}\,[\mathbf{b}_w] = \text{cov}\,[\boldsymbol{\varepsilon}_w] = \text{cov}\left[W^{\frac{1}{2}}\boldsymbol{\varepsilon}\right] = W^{\frac{1}{2}}\boldsymbol{\Sigma}W^{\frac{1}{2}} = I,$$

i.e., $\mathbf{b}_w$ is homoscedastic.

Weighted least squares

- Let us now consider the case: $\text{cov}\,[\mathbf{b}] = \text{cov}\,[\boldsymbol{\varepsilon}] = \Sigma = \sigma_i^2 \delta_{ij}$ (i.e. Σ is a diagonal matrix). When the variances σ_i^2 have different values, the random variable is called heteroscedastic.

- Without the homoscedasticity assumption, the Gauss-Markov theorem is not valid, but the heteroscedastic $\mathbf{b}$ can be suitably corrected in order to become homoscedastic.

- Let us define the weight matrix $W = \frac{1}{\sigma_i^2}\delta_{ij}$, and the weighted observations $\mathbf{b}_w = W^{\frac{1}{2}}\mathbf{b}$.

- Accordingly: $H_w = W^{\frac{1}{2}}H$, $\boldsymbol{\varepsilon}_w = W^{\frac{1}{2}}\boldsymbol{\varepsilon}$, and

$$\text{cov}\,[\mathbf{b}_w] = \text{cov}\,[\boldsymbol{\varepsilon}_w] = \text{cov}\left[W^{\frac{1}{2}}\boldsymbol{\varepsilon}\right] = W^{\frac{1}{2}}\Sigma W^{\frac{1}{2}} = I,$$

  i.e., $\mathbf{b}_w$ is homoscedastic.

- Thus, the weighted LS estimator for the system $\mathbf{b}_w = H_w\boldsymbol{\theta} + \boldsymbol{\varepsilon}_w$ is BLUE:

$$\hat{\boldsymbol{\theta}} = \left(H_w^\dagger H_w\right)^{-1} H_w^\dagger \mathbf{b_w} = \left(H^\dagger W H\right)^{-1} H^\dagger W \mathbf{b} = H_w^+ \mathbf{b}$$

Weighted least squares

- Heteroscedastic observations $\mathbf{b}$ with non-diagonal covariance are called autocorrelated.

Weighted least squares

- Heteroscedastic observations $\mathbf{b}$ with non-diagonal covariance are called autocorrelated.
- The weighted LS estimator can be generalized to any positive definite covariance.

Weighted least squares

- Heteroscedastic observations **b** with non-diagonal covariance are called autocorrelated.
- The weighted LS estimator can be generalized to any positive definite covariance.

Lemma 1

A positive definite complex square matrix A is invertible. If A is positive semi-definite, but not positive definite, it is not invertible.

Proof.

If A is positive definite, it has only non-zero eigenvalues: $\forall \mathbf{z} \neq 0$, $A\mathbf{z} \neq 0$. Hence $\dim(\ker A) = 0$, and A is full-rank. Therefore, A is invertible. Otherwise, if A is positive semi-definite but not definite, it has a 0 eigenvalue and $\dim(\ker A) \neq 0 \Rightarrow A$ not invertible. □

Weighted least squares

- Heteroscedastic observations **b** with non-diagonal covariance are called autocorrelated.
- The weighted LS estimator can be generalized to any positive definite covariance.

Lemma 1

A positive definite complex square matrix A is invertible. If A is positive semi-definite, but not positive definite, it is not invertible.

Proof.

If A is positive definite, it has only non-zero eigenvalues: $\forall \mathbf{z} \neq 0$, $A\mathbf{z} \neq 0$. Hence $\dim(\ker A) = 0$, and A is full-rank. Therefore, A is invertible. Otherwise, if A is positive semi-definite but not definite, it has a 0 eigenvalue and $\dim(\ker A) \neq 0 \Rightarrow A$ not invertible. □

Lemma 2

The covariance matrix $\operatorname{cov}[\mathbf{b}]$ of a sample **b** is positive definite and invertible iff for any non-zero **z**, $\operatorname{var}\left[\mathbf{z}^\dagger \mathbf{b}\right] \neq 0$.

Proof.

Since the covariance is positive semi-definite by definition, it is invertible only if it is also positive definite. If $\operatorname{cov}[\mathbf{b}]$ is positive definite, then $\operatorname{var}\left[\mathbf{z}^\dagger \mathbf{b}\right] \neq 0$, indeed $0 \neq \mathbf{z}^\dagger \operatorname{cov}[\mathbf{b}]\, \mathbf{z}$ $= \operatorname{cov}\left[\mathbf{z}^\dagger \mathbf{b}\right] = \operatorname{var}\left[\mathbf{z}^\dagger \mathbf{b}\right]$, since $\mathbf{z}^\dagger \mathbf{b}$ is a scalar. Conversely, if for any non-zero **z**, $\operatorname{var}\left[\mathbf{z}^\dagger \mathbf{b}\right] \neq 0$, then $\operatorname{cov}[\mathbf{b}]$ is positive definite, hence invertible. □

Weighted least squares

- If $\Sigma = \mathrm{cov}\left[\mathbf{b}\right]$ is positive definite, its inverse can be factorized by Cholensky decomposition as $\Sigma^{-1} = \Omega\Omega^{\dagger}$, where Ω is an invertible lower-triangular matrix.

Weighted least squares

- If $\Sigma = \mathrm{cov}\,[\mathbf{b}]$ is positive definite, its inverse can be factorized by Cholensky decomposition as $\Sigma^{-1} = \Omega\Omega^{\dagger}$, where Ω is an invertible lower-triangular matrix.
- When the observations $\mathbf{b}$ are heteroscedastic but non-autocorrelated, then $\Omega = W^{\frac{1}{2}}$.

Weighted least squares

- If $\boldsymbol{\Sigma} = \mathrm{cov}\left[\mathbf{b}\right]$ is positive definite, its inverse can be factorized by Cholensky decomposition as $\boldsymbol{\Sigma}^{-1} = \Omega\Omega^{\dagger}$, where Ω is an invertible lower-triangular matrix.
- When the observations $\mathbf{b}$ are heteroscedastic but non-autocorrelated, then $\Omega = W^{\frac{1}{2}}$.
- As above, let us define weighted quantities $\mathbf{b}_{\Omega} = \Omega^{\dagger}\mathbf{b}$, $H_{\Omega} = \Omega^{\dagger}H$, $\boldsymbol{\varepsilon}_{\Omega} = \Omega^{\dagger}\boldsymbol{\varepsilon}$

Weighted least squares

- If $\Sigma = \mathrm{cov}\,[\mathbf{b}]$ is positive definite, its inverse can be factorized by Cholensky decomposition as $\Sigma^{-1} = \Omega\Omega^\dagger$, where Ω is an invertible lower-triangular matrix.
- When the observations $\mathbf{b}$ are heteroscedastic but non-autocorrelated, then $\Omega = W^{\frac{1}{2}}$.
- As above, let us define weighted quantities $\mathbf{b}_\Omega = \Omega^\dagger \mathbf{b}$, $H_\Omega = \Omega^\dagger H$, $\boldsymbol{\varepsilon}_\Omega = \Omega^\dagger \boldsymbol{\varepsilon}$

Generalized Weighted Least Squares

The weighted observations $\mathbf{b}_\Omega$ are homoscedastic and non-autocorrelated, therefore, the weighted LS estimator for the system $\mathbf{b}_\Omega = H_\Omega \boldsymbol{\theta} + \boldsymbol{\varepsilon}_\Omega$ is BLUE by Gauss-Markov theorem:

$$\hat{\boldsymbol{\theta}} = \left(H_\Omega^\dagger H_\Omega\right)^{-1} H_\Omega^\dagger \mathbf{b}_\Omega = \left(H^\dagger \Omega\Omega^\dagger H\right)^{-1} H^\dagger \Omega\Omega^\dagger \mathbf{b} = \left(H^\dagger \Sigma^{-1} H\right)^{-1} H^\dagger \Sigma^{-1} \mathbf{b} = H_\Omega^{+} \mathbf{b}$$

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \left(H_\Omega^\dagger H_\Omega\right)^{-1} = \left(H^\dagger \Omega\Omega^\dagger H\right)^{-1} = \left(H^\dagger \Sigma^{-1} H\right)^{-1}$$

Proof.

- $\mathrm{E}\left[\boldsymbol{\varepsilon}_\Omega\right] = \mathrm{E}\left[\Omega^\dagger \boldsymbol{\varepsilon}\right] = \Omega^\dagger \mathrm{E}\left[\boldsymbol{\varepsilon}\right] = 0$
- $\mathrm{cov}\left[\boldsymbol{\varepsilon}_\Omega\right] = \mathrm{cov}\left[\Omega^\dagger \boldsymbol{\varepsilon}\right] = \Omega^\dagger \mathrm{cov}\left[\boldsymbol{\varepsilon}\right] \Omega = \Omega^\dagger \Sigma \Omega = \Omega^\dagger \left(\Omega\Omega^\dagger\right)^{-1} \Omega = \Omega^\dagger \left(\Omega^\dagger\right)^{-1} \Omega^{-1}\Omega = I.$
- The assumptions of the Gauss-Markov theorem are therefore satisfied.

Table of Contents

Summary on the OLS estimator

- Given a system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, with N observations, p parameters, $\mathrm{rank} H = p$, $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, $\mathrm{cov}[\boldsymbol{\varepsilon}] = \Sigma$ positive definite, the OLS estimator is:

$$\hat{\boldsymbol{\theta}} = \left(H^\dagger \Sigma^{-1} H\right)^{-1} H^\dagger \Sigma^{-1} \mathbf{b} \qquad \left(= \left(H^\dagger H\right)^{-1} H^\dagger \mathbf{b} \text{ when } \Sigma = \sigma^2 I\right)$$

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \left(H^\dagger \Sigma^{-1} H\right)^{-1} \qquad \left(= \sigma^2 \left(H^\dagger H\right)^{-1} \text{ when } \Sigma = \sigma^2 I\right)$$

Summary on the OLS estimator

- Given a system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, with N observations, p parameters, $\mathrm{rank} H = p$, $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, $\mathrm{cov}[\boldsymbol{\varepsilon}] = \Sigma$ positive definite, the OLS estimator is:

$$\hat{\boldsymbol{\theta}} = \left(H^\dagger \Sigma^{-1} H\right)^{-1} H^\dagger \Sigma^{-1} \mathbf{b} \qquad \left(= \left(H^\dagger H\right)^{-1} H^\dagger \mathbf{b} \text{ when } \Sigma = \sigma^2 I\right)$$

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \left(H^\dagger \Sigma^{-1} H\right)^{-1} \qquad \left(= \sigma^2 \left(H^\dagger H\right)^{-1} \text{ when } \Sigma = \sigma^2 I\right)$$

- $\hat{\boldsymbol{\theta}}$ is unbiased, i.e., $\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$.

Summary on the OLS estimator

- Given a system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, with N observations, p parameters, $\operatorname{rank} H = p$, $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, $\operatorname{cov}[\boldsymbol{\varepsilon}] = \Sigma$ positive definite, the OLS estimator is:

$$\hat{\boldsymbol{\theta}} = \left(H^\dagger \Sigma^{-1} H\right)^{-1} H^\dagger \Sigma^{-1} \mathbf{b} \qquad \left(= \left(H^\dagger H\right)^{-1} H^\dagger \mathbf{b} \text{ when } \Sigma = \sigma^2 I\right)$$

$$\operatorname{cov}\left[\hat{\boldsymbol{\theta}}\right] = \left(H^\dagger \Sigma^{-1} H\right)^{-1} \qquad \left(= \sigma^2 \left(H^\dagger H\right)^{-1} \text{ when } \Sigma = \sigma^2 I\right)$$

- $\hat{\boldsymbol{\theta}}$ is unbiased, i.e., $\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$.

- The Gauss-Markov theorem states that $\hat{\boldsymbol{\theta}}$ is the minimum variance estimator and the best linear unbiased estimator (BLUE), i.e., if $\hat{\boldsymbol{\theta}}_L$ is any other linear unbiased estimator:

$$\operatorname{var}\left[\hat{\boldsymbol{\theta}}_L\right] \geq \operatorname{var}\left[\hat{\boldsymbol{\theta}}\right]$$

$$\mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}\right\|^2\right] \geq \mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|^2\right].$$

Summary on the OLS estimator

- Given a system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, with N observations, p parameters, $\mathrm{rank} H = p$, $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, $\mathrm{cov}[\boldsymbol{\varepsilon}] = \Sigma$ positive definite, the OLS estimator is:

$$\hat{\boldsymbol{\theta}} = \left(H^\dagger \Sigma^{-1} H\right)^{-1} H^\dagger \Sigma^{-1} \mathbf{b} \qquad \left(= \left(H^\dagger H\right)^{-1} H^\dagger \mathbf{b} \text{ when } \Sigma = \sigma^2 I\right)$$

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \left(H^\dagger \Sigma^{-1} H\right)^{-1} \qquad \left(= \sigma^2 \left(H^\dagger H\right)^{-1} \text{ when } \Sigma = \sigma^2 I\right)$$

- $\hat{\boldsymbol{\theta}}$ is unbiased, i.e., $\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$.

- The Gauss-Markov theorem states that $\hat{\boldsymbol{\theta}}$ is the minimum variance estimator and the best linear unbiased estimator (BLUE), i.e., if $\hat{\boldsymbol{\theta}}_L$ is any other linear unbiased estimator:

$$\mathrm{var}\left[\hat{\boldsymbol{\theta}}_L\right] \geq \mathrm{var}\left[\hat{\boldsymbol{\theta}}\right]$$

$$\mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}\right\|^2\right] \geq \mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|^2\right].$$

So far, so good! BUT when $\mathrm{rank} H < p$, $H^\dagger H$ is not invertible and $\hat{\boldsymbol{\theta}}$ is not defined.

Summary on the OLS estimator

- Given a system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, with N observations, p parameters, $\mathrm{rank} H = p$, $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, $\mathrm{cov}[\boldsymbol{\varepsilon}] = \Sigma$ positive definite, the OLS estimator is:

$$\hat{\boldsymbol{\theta}} = \left(H^\dagger \Sigma^{-1} H\right)^{-1} H^\dagger \Sigma^{-1} \mathbf{b} \qquad \left(= \left(H^\dagger H\right)^{-1} H^\dagger \mathbf{b} \text{ when } \Sigma = \sigma^2 I\right)$$

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \left(H^\dagger \Sigma^{-1} H\right)^{-1} \qquad \left(= \sigma^2 \left(H^\dagger H\right)^{-1} \text{ when } \Sigma = \sigma^2 I\right)$$

- $\hat{\boldsymbol{\theta}}$ is unbiased, i.e., $\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$.

- The Gauss-Markov theorem states that $\hat{\boldsymbol{\theta}}$ is the minimum variance estimator and the best linear unbiased estimator (BLUE), i.e., if $\hat{\boldsymbol{\theta}}_L$ is any other linear unbiased estimator:

$$\mathrm{var}\left[\hat{\boldsymbol{\theta}}_L\right] \geq \mathrm{var}\left[\hat{\boldsymbol{\theta}}\right]$$

$$\mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}\right\|^2\right] \geq \mathrm{E}\left[\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|^2\right].$$

So far, so good! BUT when $\mathrm{rank} H < p$, $H^\dagger H$ is not invertible and $\hat{\boldsymbol{\theta}}$ is not defined.

How to proceed then when $\mathrm{rank}(H) < p$?

Table of Contents

Under-determined linear system

- In this section we consider the case $N < p$ and $\mathrm{rank}\,(H) = N$, i.e. a system with less equations than parameters.
- The most general case $(\mathrm{rank}\,(H) \leq \min\,(N, p)$, $\forall N$ and $\forall p)$ will be treated later on.

Under-determined linear system

- In this section we consider the case $N < p$ and $\operatorname{rank}(H) = N$, i.e. a system with less equations than parameters.
- The most general case $(\operatorname{rank}(H) \leq \min(N, p), \forall N$ and $\forall p)$ will be treated later on.
- Since $\mathbf{b} \in \mathbb{C}^N$ and $\operatorname{rank}(H) = N$, then $\operatorname{rank}(H) = \operatorname{rank}\left[\begin{pmatrix} H & \mathbf{b} \end{pmatrix}\right]$, and the undetermined system $H\boldsymbol{\theta} = \mathbf{b}$ is consistent.

Under-determined linear system

- In this section we consider the case $N < p$ and $\mathrm{rank}\,(H) = N$, i.e. a system with less equations than parameters.
- The most general case $(\mathrm{rank}\,(H) \leq \min\,(N, p), \forall N$ and $\forall p)$ will be treated later on.
- Since $\mathbf{b} \in \mathbb{C}^N$ and $\mathrm{rank}\,(H) = N$, then $\mathrm{rank}\,(H) = \mathrm{rank}\left[\begin{pmatrix} H & \mathbf{b} \end{pmatrix}\right]$, and the undetermined system $H\boldsymbol{\theta} = \mathbf{b}$ is consistent.
- For the rank-nullity theorem $\dim\,(\ker H) = p - N$, therefore, there exist nonzero vectors $\mathbf{v} \in \ker H$, s.t. $H\mathbf{v} = 0 \Rightarrow H\,(\boldsymbol{\theta} + \mathbf{v}) = H\boldsymbol{\theta} = \mathbf{b}$, i.e., the system has infinite solutions.

Under-determined linear system

- In this section we consider the case $N < p$ and $\mathrm{rank}\,(H) = N$, i.e. a system with less equations than parameters.
- The most general case $(\mathrm{rank}\,(H) \leq \min\,(N, p)$, $\forall N$ and $\forall p)$ will be treated later on.
- Since $\mathbf{b} \in \mathbb{C}^N$ and $\mathrm{rank}\,(H) = N$, then $\mathrm{rank}\,(H) = \mathrm{rank}\left[\begin{pmatrix} H & \mathbf{b} \end{pmatrix}\right]$, and the undetermined system $H\boldsymbol{\theta} = \mathbf{b}$ is consistent.
- For the rank-nullity theorem $\dim\,(\ker H) = p - N$, therefore, there exist nonzero vectors $\mathbf{v} \in \ker H$, s.t. $H\mathbf{v} = 0 \Rightarrow H\,(\boldsymbol{\theta} + \mathbf{v}) = H\boldsymbol{\theta} = \mathbf{b}$, i.e., the system has infinite solutions.
- The solution can be made unique by requiring that $\|\boldsymbol{\theta}\|^2 = \boldsymbol{\theta}^\dagger\boldsymbol{\theta}$ is minimum.

Under-determined linear system

- In this section we consider the case $N < p$ and $\mathrm{rank}\,(H) = N$, i.e. a system with less equations than parameters.
- The most general case $(\mathrm{rank}\,(H) \leq \min\,(N, p)$, $\forall N$ and $\forall p)$ will be treated later on.
- Since $\mathbf{b} \in \mathbb{C}^N$ and $\mathrm{rank}\,(H) = N$, then $\mathrm{rank}\,(H) = \mathrm{rank}\left[\begin{pmatrix} H & \mathbf{b} \end{pmatrix}\right]$, and the undetermined system $H\boldsymbol{\theta} = \mathbf{b}$ is consistent.
- For the rank-nullity theorem $\dim\,(\ker H) = p - N$, therefore, there exist nonzero vectors $\mathbf{v} \in \ker H$, s.t. $H\mathbf{v} = 0 \Rightarrow H\,(\boldsymbol{\theta} + \mathbf{v}) = H\boldsymbol{\theta} = \mathbf{b}$, i.e., the system has infinite solutions.
- The solution can be made unique by requiring that $\|\boldsymbol{\theta}\|^2 = \boldsymbol{\theta}^\dagger\boldsymbol{\theta}$ is minimum.
- Hence we have the following constrained optimization problem:

$$\begin{cases} \hat{\boldsymbol{\theta}} = \arg\min\limits_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|^2 \\ \mathbf{g}\,(\boldsymbol{\theta}) = H\boldsymbol{\theta} - \mathbf{b} = 0 \end{cases}$$

Under-determined linear system

$$\begin{cases} \hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} ||\boldsymbol{\theta}||^2 \\ \mathbf{g}(\boldsymbol{\theta}) = H\boldsymbol{\theta} - \mathbf{b} = 0 \end{cases}$$

- The problem can be solved by using Lagrange multipliers. As for the the full-rank system, we now treat only the system in the real field. The complex case will be treated later on. Let us define the Lagrangian function with the Lagrange multiplier $\boldsymbol{\lambda} \in \mathbb{R}^N$:

$$L(\boldsymbol{\theta}, \lambda) = \boldsymbol{\theta}^T\boldsymbol{\theta} + \boldsymbol{\lambda}^T\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}^T\boldsymbol{\theta} + \boldsymbol{\lambda}^T(H\boldsymbol{\theta} - \mathbf{b})$$

Under-determined linear system

$$\begin{cases} \hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} ||\boldsymbol{\theta}||^2 \\ \mathbf{g}(\boldsymbol{\theta}) = H\boldsymbol{\theta} - \mathbf{b} = 0 \end{cases}$$

- The problem can be solved by using Lagrange multipliers. As for the the full-rank system, we now treat only the system in the real field. The complex case will be treated later on. Let us define the Lagrangian function with the Lagrange multiplier $\boldsymbol{\lambda} \in \mathbb{R}^N$:

$$L(\boldsymbol{\theta}, \lambda) = \boldsymbol{\theta}^T\boldsymbol{\theta} + \boldsymbol{\lambda}^T\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}^T\boldsymbol{\theta} + \boldsymbol{\lambda}^T(H\boldsymbol{\theta} - \mathbf{b})$$

- The constrained problem becomes an unconstrained problem. Imposing the gradient is zero, the constraint is directly included in the second equation:

$$\begin{cases} \frac{\partial L}{\partial \boldsymbol{\theta}} = 2\boldsymbol{\theta} + H^T\boldsymbol{\lambda} = 0 \\ \frac{\partial L}{\partial \boldsymbol{\lambda}} = H\boldsymbol{\theta} - \mathbf{b} = \mathbf{g}(\boldsymbol{\theta}) = 0 \end{cases}$$

Under-determined linear system

$$\begin{cases} \hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|^2 \\ \mathbf{g}(\boldsymbol{\theta}) = H\boldsymbol{\theta} - \mathbf{b} = 0 \end{cases}$$

- The problem can be solved by using Lagrange multipliers. As for the the full-rank system, we now treat only the system in the real field. The complex case will be treated later on. Let us define the Lagrangian function with the Lagrange multiplier $\boldsymbol{\lambda} \in \mathbb{R}^N$:

$$L(\boldsymbol{\theta}, \lambda) = \boldsymbol{\theta}^T\boldsymbol{\theta} + \boldsymbol{\lambda}^T \mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}^T\boldsymbol{\theta} + \boldsymbol{\lambda}^T (H\boldsymbol{\theta} - \mathbf{b})$$

- The constrained problem becomes an unconstrained problem. Imposing the gradient is zero, the constraint is directly included in the second equation:

$$\begin{cases} \frac{\partial L}{\partial \boldsymbol{\theta}} = 2\boldsymbol{\theta} + H^T\boldsymbol{\lambda} = 0 \\ \frac{\partial L}{\partial \boldsymbol{\lambda}} = H\boldsymbol{\theta} - \mathbf{b} = \mathbf{g}(\boldsymbol{\theta}) = 0 \end{cases}$$

- Therefore: $\boldsymbol{\theta} = -\frac{1}{2}H^T\boldsymbol{\lambda} \Rightarrow -\frac{1}{2}HH^T\boldsymbol{\lambda} = \mathbf{b} \Rightarrow \boldsymbol{\lambda} = -2\left(HH^T\right)^{-1}\mathbf{b}$, and finally:

$$\hat{\boldsymbol{\theta}} = H^\dagger \left(HH^\dagger\right)^{-1} \mathbf{b}$$

Transpose ^{T} has been substituted with conjugate transpose †, since the solution is correct also in the complex field, as will be proved later on. $HH^\dagger$ is invertible because it is an $N \times N$ matrix and $\mathrm{rank}(H) = N$.

Under-determined linear system

LS solution of the underdetermined linear system

The system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, with $N < p$, $\operatorname{rank} H = N$, $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$, $\operatorname{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I$, has the following LS solution:

$$\hat{\boldsymbol{\theta}} = H^\dagger \left(HH^\dagger\right)^{-1} \mathbf{b} \Rightarrow \operatorname{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2 H^\dagger \left(HH^\dagger\right)^{-2} H$$

Furthermore, the norm $\left\|\hat{\boldsymbol{\theta}}\right\|^2$ is minimum

Under-determined linear system

LS solution of the underdetermined linear system

The system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, with $N < p$, $\mathrm{rank} H = N$, $\mathrm{E}\left[\boldsymbol{\varepsilon}\right] = 0$, $\mathrm{cov}\left[\boldsymbol{\varepsilon}\right] = \sigma^2 I$, has the following LS solution:

$$\hat{\boldsymbol{\theta}} = H^{\dagger}\left(HH^{\dagger}\right)^{-1}\mathbf{b} \Rightarrow \mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2 H^{\dagger}\left(HH^{\dagger}\right)^{-2} H$$

Furthermore, the norm $\left\|\hat{\boldsymbol{\theta}}\right\|^2$ is minimum

Proof.

Theorem already proved, except for the covariance:

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \mathrm{cov}\left[H^{+}\mathbf{b}\right] = \mathrm{cov}\left[H^{\dagger}\left(HH^{\dagger}\right)^{-1}\mathbf{b}\right] = H^{\dagger}\left(HH^{\dagger}\right)^{-1}\mathrm{cov}\left[\boldsymbol{\varepsilon}\right]\left(HH^{\dagger}\right)^{-1} H =$$

$$= \sigma^2 H^{\dagger}\left(HH^{\dagger}\right)^{-1}\left(HH^{\dagger}\right)^{-1} H = \sigma^2 H^{\dagger}\left(HH^{\dagger}\right)^{-2} H$$

Under-determined linear system

- So far, only the case $N \geq p$, $\mathrm{rank}\,(H) = p$, and the case $N < p$, $\mathrm{rank}\,(H) = N$ have been treated.

Under-determined linear system

- So far, only the case $N \geq p$, $\mathrm{rank}\,(H) = p$, and the case $N < p$, $\mathrm{rank}\,(H) = N$ have been treated.
- If we define the pseudo-inverse for the undetermined linear system as $H^{+} = H^{\dagger}\left(HH^{\dagger}\right)^{-1}$, we see that $HH^{+} = I$ and $H^{+}H \neq I$.

Under-determined linear system

- So far, only the case $N \geq p$, $\mathrm{rank}\,(H) = p$, and the case $N < p$, $\mathrm{rank}\,(H) = N$ have been treated.
- If we define the pseudo-inverse for the undetermined linear system as $H^{+} = H^{\dagger}\left(HH^{\dagger}\right)^{-1}$, we see that $HH^{+} = I$ and $H^{+}H \neq I$.
- For OLS we saw $HH^{+} \neq I$ and $H^{+}H = I$.

Under-determined linear system

- So far, only the case $N \geq p$, $\operatorname{rank}(H) = p$, and the case $N < p$, $\operatorname{rank}(H) = N$ have been treated.
- If we define the pseudo-inverse for the undetermined linear system as $H^{+} = H^{\dagger}\left(HH^{\dagger}\right)^{-1}$, we see that $HH^{+} = I$ and $H^{+}H \neq I$.
- For OLS we saw $HH^{+} \neq I$ and $H^{+}H = I$.
- We will see that, in general, it might be $HH^{+} \neq I$ and $H^{+}H \neq I$, but $HH^{+}H = H$ is always true.

Under-determined linear system

- So far, only the case $N \geq p$, $\mathrm{rank}\,(H) = p$, and the case $N < p$, $\mathrm{rank}\,(H) = N$ have been treated.
- If we define the pseudo-inverse for the undetermined linear system as $H^{+} = H^{\dagger}\left(HH^{\dagger}\right)^{-1}$, we see that $HH^{+} = I$ and $H^{+}H \neq I$.
- For OLS we saw $HH^{+} \neq I$ and $H^{+}H = I$.
- We will see that, in general, it might be $HH^{+} \neq I$ and $H^{+}H \neq I$, but $HH^{+}H = H$ is always true.
- The most general case is $\mathrm{rank}\,(H) = r \leq \min\,(N, p)$, $\forall N$ and $\forall p$.

Under-determined linear system

- So far, only the case $N \geq p$, $\mathrm{rank}\,(H) = p$, and the case $N < p$, $\mathrm{rank}\,(H) = N$ have been treated.
- If we define the pseudo-inverse for the undetermined linear system as $H^{+} = H^{\dagger}\left(HH^{\dagger}\right)^{-1}$, we see that $HH^{+} = I$ and $H^{+}H \neq I$.
- For OLS we saw $HH^{+} \neq I$ and $H^{+}H = I$.
- We will see that, in general, it might be $HH^{+} \neq I$ and $H^{+}H \neq I$, but $HH^{+}H = H$ is always true.
- The most general case is $\mathrm{rank}\,(H) = r \leq \min\,(N, p)$, $\forall N$ and $\forall p$.
- The general case can be treated by means of a powerful technique: Singular Value Decomposition.

Under-determined linear system

- So far, only the case $N \geq p$, $\operatorname{rank}(H) = p$, and the case $N < p$, $\operatorname{rank}(H) = N$ have been treated.
- If we define the pseudo-inverse for the undetermined linear system as $H^+ = H^\dagger \left(HH^\dagger\right)^{-1}$, we see that $HH^+ = I$ and $H^+H \neq I$.
- For OLS we saw $HH^+ \neq I$ and $H^+H = I$.
- We will see that, in general, it might be $HH^+ \neq I$ and $H^+H \neq I$, but $HH^+H = H$ is always true.
- The most general case is $\operatorname{rank}(H) = r \leq \min(N, p)$, $\forall N$ and $\forall p$.
- The general case can be treated by means of a powerful technique: Singular Value Decomposition.
- A general solution will be found that reduces to those already obtained for the two special cases discussed so far.

Under-determined linear system

- So far, only the case $N \geq p$, $\operatorname{rank}(H) = p$, and the case $N < p$, $\operatorname{rank}(H) = N$ have been treated.
- If we define the pseudo-inverse for the undetermined linear system as $H^{+} = H^{\dagger}\left(HH^{\dagger}\right)^{-1}$, we see that $HH^{+} = I$ and $H^{+}H \neq I$.
- For OLS we saw $HH^{+} \neq I$ and $H^{+}H = I$.
- We will see that, in general, it might be $HH^{+} \neq I$ and $H^{+}H \neq I$, but $HH^{+}H = H$ is always true.
- The most general case is $\operatorname{rank}(H) = r \leq \min(N, p)$, $\forall N$ and $\forall p$.
- The general case can be treated by means of a powerful technique: Singular Value Decomposition.
- A general solution will be found that reduces to those already obtained for the two special cases discussed so far.
- In the next section, Singular Value Decomposition will be introduced and demonstrated.

Table of Contents

Review of linear algebra preliminary to SVD

- Let be given $A \in \mathbb{C}^{N \times p}$, $\mathrm{rank}\,(A) = r \leq \min\,(N, p)$.

Review of linear algebra preliminary to SVD

- Let be given $A \in \mathbb{C}^{N \times p}$, $\mathrm{rank}\,(A) = r \leq \min\,(N, p)$.
- Then, $A^{\dagger}A \in \mathbb{C}^{p \times p}$, $AA^{\dagger} \in \mathbb{C}^{N \times N}$ are semi-positive definite, and $\mathrm{rank}\left(A^{\dagger}A\right) = \mathrm{rank}\left(AA^{\dagger}\right) = r$.

Review of linear algebra preliminary to SVD

- Let be given $A \in \mathbb{C}^{N \times p}$, $\mathrm{rank}\,(A) = r \leq \min\,(N, p)$.
- Then, $A^{\dagger}A \in \mathbb{C}^{p \times p}$, $AA^{\dagger} \in \mathbb{C}^{N \times N}$ are semi-positive definite, and $\mathrm{rank}\left(A^{\dagger}A\right) = \mathrm{rank}\left(AA^{\dagger}\right) = r$.
- For the rank-nullity theorem:

$$\dim\left(\ker A^{\dagger}A\right) = p - r$$

$$\dim\left(\ker AA^{\dagger}\right) = N - r$$

Review of linear algebra preliminary to SVD

- Let be given $A \in \mathbb{C}^{N \times p}$, $\mathrm{rank}\,(A) = r \leq \min\,(N, p)$.
- Then, $A^{\dagger}A \in \mathbb{C}^{p \times p}$, $AA^{\dagger} \in \mathbb{C}^{N \times N}$ are semi-positive definite, and $\mathrm{rank}\,\left(A^{\dagger}A\right) = \mathrm{rank}\,\left(AA^{\dagger}\right) = r$.
- For the rank-nullity theorem:

$$\dim\left(\ker A^{\dagger}A\right) = p - r$$

$$\dim\left(\ker AA^{\dagger}\right) = N - r$$

- Then, $A^{\dagger}A$ has $p - r$ orthogonal eigenvectors associated with the eigenvalue 0, and $AA^{\dagger}$ has $N - r$ orthogonal eigenvectors associated with the eigenvalue 0.

Review of linear algebra preliminary to SVD

- Let be given $A \in \mathbb{C}^{N\times p}$, $\operatorname{rank}(A) = r \leq \min(N, p)$.
- Then, $A^\dagger A \in \mathbb{C}^{p\times p}$, $AA^\dagger \in \mathbb{C}^{N\times N}$ are semi-positive definite, and $\operatorname{rank}\left(A^\dagger A\right) = \operatorname{rank}\left(AA^\dagger\right) = r$.
- For the rank-nullity theorem:

$$\dim\left(\ker A^\dagger A\right) = p - r$$

$$\dim\left(\ker AA^\dagger\right) = N - r$$

- Then, $A^\dagger A$ has $p - r$ orthogonal eigenvectors associated with the eigenvalue 0, and $AA^\dagger$ has $N - r$ orthogonal eigenvectors associated with the eigenvalue 0.
- Since $A^\dagger A$ and $AA^\dagger$ are Hermitian, they have an orthonormal basis of eigenvectors. E.g.:

$$A^\dagger AV = A^\dagger A\begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_p \end{bmatrix} = \\ = V\begin{bmatrix} \sigma_1^2 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \sigma_r^2 & \vdots \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} \end{bmatrix}_{p\times p} = V\Sigma_p^2 \quad ; \qquad AA^\dagger U = AA^\dagger\begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_N \end{bmatrix} = \\ = U\begin{bmatrix} \sigma_1^2 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \sigma_r^2 & \vdots \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} \end{bmatrix}_{N\times N} = U\Sigma_N^2$$

- Same symbols σ_i^2 have been used for both Σ_N^2 and Σ_p^2, indeed, as it will be proved in the following, the eigenvalues of $A^\dagger A$ and $AA^\dagger$ are the same.

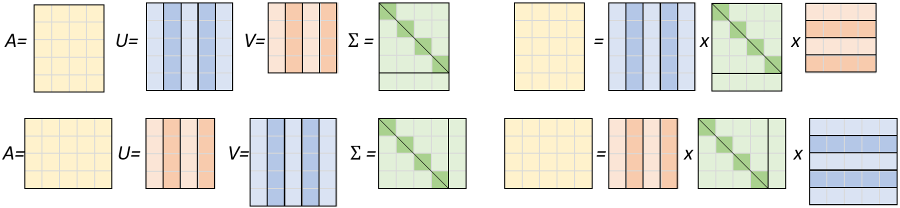
Table of Contents

Singular Value Decomposition

Singular Value Decomposition

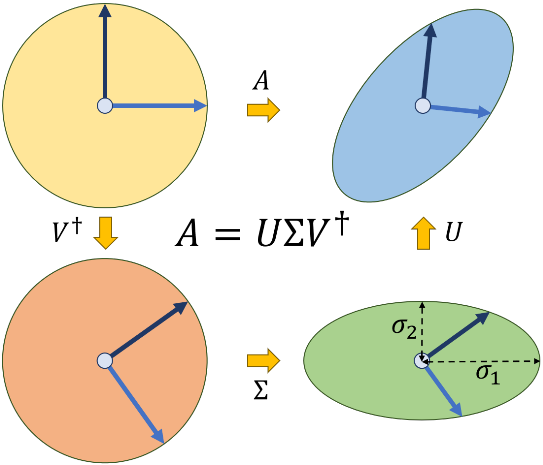
Any matrix $A \in \mathbb{C}^{N \times p}$, of any rank $r \leq \min(N, p)$, can be factorized in the form $A = U \Sigma V^{\dagger}$,

- $\Sigma \in \mathbb{R}^{N \times p}$ is a diagonal matrix with r positive elements that can always be ordered as $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$; σ_i are the so called singular values
- $U \in \mathbb{C}^{N \times N}$ and $V \in \mathbb{C}^{p \times p}$ are unitary matrices
- U, V and Σ can be found by solving the eigenvalue problems $A^{\dagger} A V = V \Sigma_p^2$ and $A A^{\dagger} U = U \Sigma_N^2$, where $\Sigma_p^2 = \Sigma^{\dagger} \Sigma$ and $\Sigma_N^2 = \Sigma \Sigma^{\dagger}$.



Singular Value Decomposition

- Geometrical interpretation: rotation, scaling and rotation



Singular Value Decomposition

Proof.

- Let us first consider the case $N \geq p$. Any matrix $A \in \mathbb{C}^{N \times p}$ is a linear application that is completely defined by the values it takes on a given basis $\mathbf{v}_{1\ldots p}$ of the domain $\mathbb{C}^p$:

$$\begin{cases} A\mathbf{v}_1 = \sigma_1 \mathbf{u}_1 \\ \vdots \\ A\mathbf{v}_p = \sigma_p \mathbf{u}_p \end{cases}$$

- $\mathbf{u}_i \in \mathbb{C}^N$ are unit vectors, $\sigma_i \geq 0$, and it is always possible to reorder the basis so that the σ_i are in descending order.

Singular Value Decomposition

Proof.

- Let us first consider the case $N \geq p$. Any matrix $A \in \mathbb{C}^{N \times p}$ is a linear application that is completely defined by the values it takes on a given basis $\mathbf{v}_{1 \ldots p}$ of the domain $\mathbb{C}^p$:

$$\begin{cases} A\mathbf{v}_1 = \sigma_1 \mathbf{u}_1 \\ \vdots \\ A\mathbf{v}_p = \sigma_p \mathbf{u}_p \end{cases}$$

- $\mathbf{u}_i \in \mathbb{C}^N$ are unit vectors, $\sigma_i \geq 0$, and it is always possible to reorder the basis so that the σ_i are in descending order.
- A convenient choice of the basis is an orthonormal set of eigenvectors: $A^\dagger A V = V\Lambda$, where $\Lambda = \lambda_i \delta_{ij}$, and $V = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_p]$ is unitary.

Singular Value Decomposition

Proof.

- Let us first consider the case $N \geq p$. Any matrix $A \in \mathbb{C}^{N\times p}$ is a linear application that is completely defined by the values it takes on a given basis $\mathbf{v}_{1\dots p}$ of the domain $\mathbb{C}^p$:

$$\begin{cases} A\mathbf{v}_1 = \sigma_1 \mathbf{u}_1 \\ \vdots \\ A\mathbf{v}_p = \sigma_p \mathbf{u}_p \end{cases}$$

- $\mathbf{u}_i \in \mathbb{C}^N$ are unit vectors, $\sigma_i \geq 0$, and it is always possible to reorder the basis so that the σ_i are in descending order.
- A convenient choice of the basis is an orthonormal set of eigenvectors: $A^\dagger A V = V\Lambda$, where $\Lambda = \lambda_i \delta_{ij}$, and $V = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_p]$ is unitary.
- This choice implies that also $\tilde{U} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_p]$ are orthonormal. Indeed, if $\sigma_{i,j} \neq 0$:

$$\mathbf{u}_i^\dagger \mathbf{u}_j = \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^\dagger A^\dagger A \mathbf{v}_j = \frac{\lambda_j}{\sigma_i \sigma_j} \mathbf{v}_i^\dagger \mathbf{v}_j = \frac{\lambda_j}{\sigma_i \sigma_j} \delta_{ij} \Rightarrow \begin{cases} i \neq j \Rightarrow \mathbf{u}_i^\dagger \mathbf{u}_j = 0 \\ i = j \Rightarrow \mathbf{u}_i^\dagger \mathbf{u}_j = \|\mathbf{u_i}\|^2 = 1 \end{cases}$$

$\lambda_i = \sigma_i^2$ because each $\mathbf{u}_i$ is a unit vector by construction.

Singular Value Decomposition

Proof.

- Let us first consider the case $N \geq p$. Any matrix $A \in \mathbb{C}^{N\times p}$ is a linear application that is completely defined by the values it takes on a given basis $\mathbf{v}_{1\ldots p}$ of the domain $\mathbb{C}^p$:

$$\begin{cases} A\mathbf{v}_1 = \sigma_1 \mathbf{u}_1 \\ \vdots \\ A\mathbf{v}_p = \sigma_p \mathbf{u}_p \end{cases}$$

- $\mathbf{u}_i \in \mathbb{C}^N$ are unit vectors, $\sigma_i \geq 0$, and it is always possible to reorder the basis so that the σ_i are in descending order.
- A convenient choice of the basis is an orthonormal set of eigenvectors: $A^\dagger AV = V\Lambda$, where $\Lambda = \lambda_i \delta_{ij}$, and $V = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_p]$ is unitary.
- This choice implies that also $\tilde{U} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_p]$ are orthonormal. Indeed, if $\sigma_{i,j} \neq 0$:

$$\mathbf{u}_i^\dagger \mathbf{u}_j = \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^\dagger A^\dagger A \mathbf{v}_j = \frac{\lambda_j}{\sigma_i \sigma_j} \mathbf{v}_i^\dagger \mathbf{v}_j = \frac{\lambda_j}{\sigma_i \sigma_j} \delta_{ij} \Rightarrow \begin{cases} i \neq j \Rightarrow \mathbf{u}_i^\dagger \mathbf{u}_j = 0 \\ i = j \Rightarrow \mathbf{u}_i^\dagger \mathbf{u}_j = \|\mathbf{u_i}\|^2 = 1 \end{cases}$$

  $\lambda_i = \sigma_i^2$ because each $\mathbf{u}_i$ is a unit vector by construction.
- When $i > r$, $\sigma_i = 0$, and it is always possible to complete $[\mathbf{u}_1 \ \cdots \ \mathbf{u}_r]$ to $[\mathbf{u}_1 \ \cdots \ \mathbf{u}_p]$ by adding $p - r$ orthonormal vectors however chosen (e.g., Gram-Schmidt).

Singular Value Decomposition

Proof.

- We came up to: $AV = A\left[\mathbf{v}_1 \;\; \cdots \;\; \mathbf{v}_p\right] = \left[\mathbf{u}_1 \;\; \cdots \;\; \mathbf{u}_p\right] \operatorname{diag}\left(\sigma_1 \;\; \cdots \;\; \sigma_r \;\; 0_{r+1} \;\; \cdots \;\; 0_p\right) = \tilde{U}\tilde{\Sigma}$.

Singular Value Decomposition

Proof.

- We came up to: $AV = A\,[\mathbf{v}_1 \;\; \cdots \;\; \mathbf{v}_p] = [\mathbf{u}_1 \;\; \cdots \;\; \mathbf{u}_p]\,\mathrm{diag}\,(\sigma_1 \;\; \cdots \;\; \sigma_r \;\; 0_{r+1} \;\; \cdots \;\; 0_p) = \tilde{U}\tilde{\Sigma}$.
- $\tilde{U}$ can be completed to a basis of $\mathbb{C}^N$:

$$AV = \begin{bmatrix} \tilde{U} & \mathbf{u}_{p+1} \cdots \mathbf{u}_N \end{bmatrix} \begin{bmatrix} \tilde{\Sigma} \\ \mathbf{0}_{(N-p)\times p} \end{bmatrix} = U\Sigma$$

Singular Value Decomposition

Proof.

- We came up to: $AV = A\left[\mathbf{v}_1 \ \cdots \ \mathbf{v}_p\right] = \left[\mathbf{u}_1 \ \cdots \ \mathbf{u}_p\right] \operatorname{diag}\left(\sigma_1 \ \cdots \ \sigma_r \ 0_{r+1} \ \cdots \ 0_p\right) = \tilde{U}\tilde{\Sigma}$.
- $\tilde{U}$ can be completed to a basis of $\mathbb{C}^N$:

$$AV = \left[\begin{array}{cc} \tilde{U} & \mathbf{u}_{p+1} \cdots \mathbf{u}_N \end{array}\right] \left[\begin{array}{c} \tilde{\Sigma} \\ \mathbf{0}_{(N-p)\times p} \end{array}\right] = U\Sigma$$

- $AV = U\Sigma \Rightarrow AVV^\dagger = U\Sigma V^\dagger \Rightarrow A = U\Sigma V^\dagger$

Singular Value Decomposition

Proof.

- We came up to: $AV = A\,[\mathbf{v}_1 \ \cdots \ \mathbf{v}_p] = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_p]\,\mathrm{diag}\,(\sigma_1 \ \cdots \ \sigma_r \ 0_{r+1} \ \cdots \ 0_p) = \tilde{U}\tilde{\Sigma}$.
- $\tilde{U}$ can be completed to a basis of $\mathbb{C}^N$:

$$AV = \begin{bmatrix} \tilde{U} & \mathbf{u}_{p+1} \cdots \mathbf{u}_N \end{bmatrix} \begin{bmatrix} \tilde{\Sigma} \\ \mathbf{0}_{(N-p)\times p} \end{bmatrix} = U\Sigma$$

- $AV = U\Sigma \Rightarrow AVV^\dagger = U\Sigma V^\dagger \Rightarrow A = U\Sigma V^\dagger$
- The eigenvalue problems for U, V and Σ can be derived as follows:

$$V^\dagger A^\dagger AV = \Sigma^\dagger U^\dagger U\Sigma = \Sigma^\dagger\Sigma = \Sigma_p^2 \Rightarrow A^\dagger AV = V\Sigma_p^2$$

$$AA^\dagger = U\Sigma V^\dagger V\Sigma^\dagger U^\dagger = U\Sigma\Sigma^\dagger U^\dagger = U\Sigma_N^2 U^\dagger \Rightarrow AA^\dagger = U\Sigma_N^2$$

Singular Value Decomposition

Proof.

- We came up to: $AV = A\left[\mathbf{v}_1 \ \cdots \ \mathbf{v}_p\right] = \left[\mathbf{u}_1 \ \cdots \ \mathbf{u}_p\right] \operatorname{diag}\left(\sigma_1 \ \cdots \ \sigma_r \ 0_{r+1} \ \cdots \ 0_p\right) = \tilde{U}\tilde{\Sigma}$.
- $\tilde{U}$ can be completed to a basis of $\mathbb{C}^N$:

$$AV = \left[\begin{array}{cc} \tilde{U} & \mathbf{u}_{p+1} \cdots \mathbf{u}_N \end{array}\right] \left[\begin{array}{c} \tilde{\Sigma} \\ \mathbf{0}_{(N-p)\times p} \end{array}\right] = U\Sigma$$

- $AV = U\Sigma \Rightarrow AVV^\dagger = U\Sigma V^\dagger \Rightarrow A = U\Sigma V^\dagger$
- The eigenvalue problems for U, V and Σ can be derived as follows:

$$V^\dagger A^\dagger AV = \Sigma^\dagger U^\dagger U\Sigma = \Sigma^\dagger\Sigma = \Sigma_p^2 \Rightarrow A^\dagger AV = V\Sigma_p^2$$

$$AA^\dagger = U\Sigma V^\dagger V\Sigma^\dagger U^\dagger = U\Sigma\Sigma^\dagger U^\dagger = U\Sigma_N^2 U^\dagger \Rightarrow AA^\dagger = U\Sigma_N^2$$

- For the case $N < p$, let us define $\bar{N} = p$ and $\bar{p} = N$, and $\bar{A} = A^\dagger \in \mathbb{C}^{\bar{N}\times\bar{p}}$, $\bar{N} > \bar{p}$:

$$\begin{array}{ccc} \bar{A} = \bar{U}\bar{\Sigma}\bar{V}^\dagger & \bar{A}^\dagger\bar{A}\bar{V} = \bar{V}\bar{\Sigma}_{\bar{p}}^2 & \bar{A}\bar{A}^\dagger\bar{U}\bar{\Sigma}_{\bar{N}}^2 \\ & AA^\dagger\bar{V} = \bar{V}\bar{\Sigma}_{\bar{p}}^2 & A^\dagger A\bar{U}\bar{\Sigma}_{\bar{N}}^2 \\ \bar{V} = U & \bar{U} = V & \bar{\Sigma} = \Sigma^\dagger \end{array}$$

$$\bar{A} = A^\dagger = V\Sigma^\dagger U^\dagger \Rightarrow A = U\Sigma V^\dagger \quad \square$$

Singular Value Decomposition

$N \geq p \Rightarrow A = U\Sigma V^\dagger =$

- $$\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_N \end{pmatrix} \begin{pmatrix} \sigma_1 & \cdots & \cdots & \cdots \\ \vdots & \ddots & & \vdots \\ \vdots & & \sigma_r & \vdots \\ \cdots & \cdots & \cdots & 0_{(p-r)\times(p-r)} \\ & & 0_{(N-p)\times p} & \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^\dagger \\ \vdots \\ \mathbf{v}_p^\dagger \end{pmatrix}$$

Singular Value Decomposition

- $N \geq p \Rightarrow A = U\Sigma V^\dagger =$

$$\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_N \end{pmatrix} \begin{pmatrix} \sigma_1 & \cdots & \cdots & \cdots \\ \vdots & \ddots & & \vdots \\ \vdots & & \sigma_r & \vdots \\ \cdots & \cdots & \cdots & 0_{(p-r)\times(p-r)} \\ & & 0_{(N-p)\times p} & \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^\dagger \\ \vdots \\ \mathbf{v}_p^\dagger \end{pmatrix}$$

- $N < p \Rightarrow A = U\Sigma V^\dagger =$

$$\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_N \end{pmatrix} \begin{pmatrix} \sigma_1 & \cdots & \cdots & \cdots & \\ \vdots & \ddots & & \vdots & \\ \vdots & & \sigma_r & \vdots & 0_{N\times(p-N)} \\ \cdots & \cdots & \cdots & 0_{(N-r)\times(N-r)} & \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^\dagger \\ \vdots \\ \mathbf{v}_p^\dagger \end{pmatrix}$$

Singular Value Decomposition

- $N \geq p \Rightarrow A = U\Sigma V^\dagger =$
$$\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_N \end{pmatrix} \begin{pmatrix} \sigma_1 & \cdots & \cdots & \cdots \\ \vdots & \ddots & & \vdots \\ \vdots & & \sigma_r & \vdots \\ \cdots & \cdots & \cdots & 0_{(p-r)\times(p-r)} \\ & & 0_{(N-p)\times p} & \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^\dagger \\ \vdots \\ \mathbf{v}_p^\dagger \end{pmatrix}$$
- $N < p \Rightarrow A = U\Sigma V^\dagger =$
$$\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_N \end{pmatrix} \begin{pmatrix} \sigma_1 & \cdots & \cdots & \cdots & \\ \vdots & \ddots & & \vdots & \\ \vdots & & \sigma_r & \vdots & 0_{N\times(p-N)} \\ \cdots & \cdots & \cdots & 0_{(N-r)\times(N-r)} & \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^\dagger \\ \vdots \\ \mathbf{v}_p^\dagger \end{pmatrix}$$

Alternative expression of the SVD

$$A = U\Sigma V^\dagger = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\dagger \quad \forall A \in \mathbb{C}^{N\times p}, \forall N, \forall p$$

Table of Contents

Introduction to the general LS solution

- For $N \geq p$ and $\mathrm{rank}\,(H) = r = p$, the OLS solution of the inconsistent system $\mathbf{b} = H\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \left(H^{\dagger}H\right)^{-1}H^{\dagger}\mathbf{b} = H^{+}\mathbf{b}$. A corrected observation vector $\hat{\mathbf{b}} = H\hat{\boldsymbol{\theta}}$ is defined, s.t. the cost function $\phi = \left\|\mathbf{b} - \hat{\mathbf{b}}\right\|^2 = \left\|\mathbf{b} - H\hat{\boldsymbol{\theta}}\right\|^2$ is minimum.

Introduction to the general LS solution

- For $N \geq p$ and $\mathrm{rank}\,(H) = r = p$, the OLS solution of the inconsistent system $\mathbf{b} = H\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \left(H^{\dagger}H\right)^{-1}H^{\dagger}\mathbf{b} = H^{+}\mathbf{b}$. A corrected observation vector $\hat{\mathbf{b}} = H\hat{\boldsymbol{\theta}}$ is defined, s.t. the cost function $\phi = \left\|\mathbf{b} - \hat{\mathbf{b}}\right\|^2 = \left\|\mathbf{b} - H\hat{\boldsymbol{\theta}}\right\|^2$ is minimum.
- When $r < \min\,(N, p)$, the rank-nullity theorem implies $\dim\,(\ker H) = p - r > 0$.

Introduction to the general LS solution

- For $N \geq p$ and $\mathrm{rank}\,(H) = r = p$, the OLS solution of the inconsistent system $\mathbf{b} = H\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \left(H^{\dagger}H\right)^{-1}H^{\dagger}\mathbf{b} = H^{+}\mathbf{b}$. A corrected observation vector $\hat{\mathbf{b}} = H\hat{\boldsymbol{\theta}}$ is defined, s.t. the cost function $\phi = \left\|\mathbf{b} - \hat{\mathbf{b}}\right\|^2 = \left\|\mathbf{b} - H\hat{\boldsymbol{\theta}}\right\|^2$ is minimum.
- When $r < \min\,(N, p)$, the rank-nullity theorem implies $\dim\,(\ker H) = p - r > 0$.
- Hence, $\exists \mathbf{v}_0 \neq \mathbf{0} : \; H\mathbf{v}_0 = \mathbf{0}$.

Introduction to the general LS solution

- For $N \geq p$ and $\operatorname{rank}(H) = r = p$, the OLS solution of the inconsistent system $\mathbf{b} = H\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \left(H^{\dagger}H\right)^{-1}H^{\dagger}\mathbf{b} = H^{+}\mathbf{b}$. A corrected observation vector $\hat{\mathbf{b}} = H\hat{\boldsymbol{\theta}}$ is defined, s.t. the cost function $\phi = \left\|\mathbf{b} - \hat{\mathbf{b}}\right\|^2 = \left\|\mathbf{b} - H\hat{\boldsymbol{\theta}}\right\|^2$ is minimum.
- When $r < \min(N, p)$, the rank-nullity theorem implies $\dim(\ker H) = p - r > 0$.
- Hence, $\exists \mathbf{v}_0 \neq \mathbf{0} : \; H\mathbf{v}_0 = \mathbf{0}$.
- Therefore, $\phi = \left\|\mathbf{b} - H\hat{\boldsymbol{\theta}}\right\|^2 = \left\|\mathbf{b} - H\left(\hat{\boldsymbol{\theta}} + \mathbf{v}_0\right)\right\|^2$ and the LS problem has an infinite number of solutions.

Introduction to the general LS solution

- For $N \geq p$ and $\mathrm{rank}(H) = r = p$, the OLS solution of the inconsistent system $\mathbf{b} = H\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \left(H^\dagger H\right)^{-1} H^\dagger \mathbf{b} = H^+ \mathbf{b}$. A corrected observation vector $\hat{\mathbf{b}} = H\hat{\boldsymbol{\theta}}$ is defined, s.t. the cost function $\phi = \left\|\mathbf{b} - \hat{\mathbf{b}}\right\|^2 = \left\|\mathbf{b} - H\hat{\boldsymbol{\theta}}\right\|^2$ is minimum.
- When $r < \min(N, p)$, the rank-nullity theorem implies $\dim(\ker H) = p - r > 0$.
- Hence, $\exists \mathbf{v}_0 \neq \mathbf{0} : \; H\mathbf{v}_0 = \mathbf{0}$.
- Therefore, $\phi = \left\|\mathbf{b} - H\hat{\boldsymbol{\theta}}\right\|^2 = \left\|\mathbf{b} - H\left(\hat{\boldsymbol{\theta}} + \mathbf{v}_0\right)\right\|^2$ and the LS problem has an infinite number of solutions.
- The solution can be made unique and it will be shown that:

General SVD pseudo-inverse

- The general form of the pseudo-inverse of $H = U\Sigma V^\dagger$ is $H^+ = V\Sigma^+ U^\dagger$.
- The unique LS solution $\hat{\boldsymbol{\theta}} = H^+\mathbf{b}$ is s.t. both $\left\|\mathbf{b} - H\hat{\boldsymbol{\theta}}\right\|^2$ and $\left\|\hat{\boldsymbol{\theta}}\right\|^2$ are minimum.
- $HH^+H = H$ is always true, but $H^+H = I$ or $HH^+ = I$ do not hold in general.
- $OLS : r = p \leq N \Rightarrow H^+ = \left(H^\dagger H\right)^{-1} H^\dagger$, $r = N < p \Rightarrow H^+ = H^\dagger \left(HH^\dagger\right)^{-1}$.
- $HH^+ = (HH^+)^\dagger$, $H^+H = (H^+H)^\dagger$

Table of Contents

General LS solution

General SVD pseudo-inverse

- The general form of the pseudo-inverse of $H = U\Sigma V^\dagger$ is $H^+ = V\Sigma^+ U^\dagger$.
- The unique LS solution $\hat{\boldsymbol{\theta}} = H^+\mathbf{b}$ is s.t. both $\left\|\mathbf{b} - H\hat{\boldsymbol{\theta}}\right\|^2$ and $\left\|\hat{\boldsymbol{\theta}}\right\|^2$ are minimum.
- $HH^+H = H$ is always true, but $H^+H = I$ or $HH^+ = I$ do not hold in general.
- $OLS : r = p \leq N \Rightarrow H^+ = \left(H^\dagger H\right)^{-1} H^\dagger$, $r = N < p \Rightarrow H^+ = H^\dagger \left(HH^\dagger\right)^{-1}$.
- $HH^+ = (HH^+)^\dagger$, $H^+H = (H^+H)^\dagger$

- Remark: as it will be shown, the pseudo-inverse of Σ, Σ^+ is obtained by transposing Σ and by replacing the elements of the diagonal with the reciprocals of their respective nonzero elements of Σ. E.g.:

$$\Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \Sigma^+ = \begin{pmatrix} 1/3 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- $N \geq p,\ \ r = p \Rightarrow \Sigma^+\Sigma = I$;
- $N \leq p,\ \ r = N \Rightarrow \Sigma\Sigma^+ = I$;
- $r < \min(N, p) \Rightarrow \Sigma^+\Sigma \neq I$, and $\Sigma\Sigma^+ \neq I$, but $\Sigma\Sigma^+\Sigma = \Sigma$ is always true.
- $\Sigma\Sigma^+ = (\Sigma\Sigma^+)^T = {\Sigma^+}^T\Sigma^T$; $\Sigma^+\Sigma = (\Sigma^+\Sigma)^T = \Sigma^T{\Sigma^+}^T$

General LS solution

An explanatory example on Σ and Σ^+

- If Σ is $N \times p$, then Σ^+ is $p \times N$, $\Sigma^+\Sigma$ is $p \times p$ and $\Sigma\Sigma^+$ is $N \times N$. E.g.:

$$\Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}; \quad \Sigma^+ = \begin{pmatrix} 1/3 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad \Sigma\Sigma^+ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- Matrices in the same form as $\Sigma\Sigma^+$, with only 0 and 1, can be called selection matrices of rank r, and denoted by the symbol I_n^r, where the superscript denotes rank, while the subscript denotes dimensions. Hence, $\Sigma\Sigma^+ = I_N^r$ and $\Sigma^+\Sigma = I_p^r$; obviously, $\mathrm{tr}\,(\Sigma\Sigma^+) = \mathrm{tr}\,(\Sigma^+\Sigma) = r$. In this example $\Sigma\Sigma^+ = I_4^2$.

General LS solution

Proof.

- $H = U\Sigma V^\dagger = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\dagger$
- Unitarity: $\mathbf{v}_i^\dagger \mathbf{v}_j = \mathbf{u}_i^\dagger \mathbf{u}_j = \delta_{ij}$

General LS solution

Proof.

- $H = U\Sigma V^\dagger = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\dagger$
- Unitarity: $\mathbf{v}_i^\dagger \mathbf{v}_j = \mathbf{u}_i^\dagger \mathbf{u}_j = \delta_{ij}$
- $j > r \Rightarrow H\mathbf{v}_j = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\dagger \mathbf{v}_j = \mathbf{0} \Rightarrow \mathbf{v}_j$ are an orthonormal basis of $\ker H$.

General LS solution

Proof.

- $H = U\Sigma V^\dagger = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\dagger$
- Unitarity: $\mathbf{v}_i^\dagger \mathbf{v}_j = \mathbf{u}_i^\dagger \mathbf{u}_j = \delta_{ij}$
- $j > r \Rightarrow H\mathbf{v}_j = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\dagger \mathbf{v}_j = \mathbf{0} \Rightarrow \mathbf{v}_j$ are an orthonormal basis of $\ker H$.
- Cost function with SVD: $\|\mathbf{b} - H\boldsymbol{\theta}\|^2 = \left\|\mathbf{b} - U\Sigma V^\dagger \boldsymbol{\theta}\right\|^2 = \left\|U^\dagger \mathbf{b} - \Sigma V^\dagger \boldsymbol{\theta}\right\|^2$

General LS solution

Proof.

- $H = U\Sigma V^\dagger = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\dagger$
- Unitarity: $\mathbf{v}_i^\dagger \mathbf{v}_j = \mathbf{u}_i^\dagger \mathbf{u}_j = \delta_{ij}$
- $j > r \Rightarrow H\mathbf{v}_j = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\dagger \mathbf{v}_j = \mathbf{0} \Rightarrow \mathbf{v}_j$ are an orthonormal basis of $\ker H$.
- Cost function with SVD: $\|\mathbf{b} - H\boldsymbol{\theta}\|^2 = \left\|\mathbf{b} - U\Sigma V^\dagger\boldsymbol{\theta}\right\|^2 = \left\|U^\dagger\mathbf{b} - \Sigma V^\dagger\boldsymbol{\theta}\right\|^2$
- By defining $\mathbf{y} \equiv V^\dagger\boldsymbol{\theta}$ and $\mathbf{c} \equiv U^\dagger\mathbf{b}$:
  $\|\mathbf{b} - H\boldsymbol{\theta}\|^2 = \|\mathbf{c} - \Sigma\mathbf{y}\|^2 = \sum_{i=1}^{r} |c_i - \sigma_i y_i|^2 + \sum_{i=r+1}^{p} |c_i|^2.$

General LS solution

Proof.

- $H = U\Sigma V^\dagger = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\dagger$
- Unitarity: $\mathbf{v}_i^\dagger \mathbf{v}_j = \mathbf{u}_i^\dagger \mathbf{u}_j = \delta_{ij}$
- $j > r \Rightarrow H\mathbf{v}_j = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\dagger \mathbf{v}_j = \mathbf{0} \Rightarrow \mathbf{v}_j$ are an orthonormal basis of $\ker H$.
- Cost function with SVD: $\|\mathbf{b} - H\boldsymbol{\theta}\|^2 = \left\|\mathbf{b} - U\Sigma V^\dagger \boldsymbol{\theta}\right\|^2 = \left\|U^\dagger \mathbf{b} - \Sigma V^\dagger \boldsymbol{\theta}\right\|^2$
- By defining $\mathbf{y} \equiv V^\dagger \boldsymbol{\theta}$ and $\mathbf{c} \equiv U^\dagger \mathbf{b}$:
  $\|\mathbf{b} - H\boldsymbol{\theta}\|^2 = \|\mathbf{c} - \Sigma\mathbf{y}\|^2 = \sum_{i=1}^{r} |c_i - \sigma_i y_i|^2 + \sum_{i=r+1}^{p} |c_i|^2.$
- The cost function is minimum for $y_i = c_i/\sigma_i$, $i = 1, \ldots, r$:

$$\mathbf{y} = \begin{pmatrix} c_1/\sigma_1 \\ \vdots \\ c_r/\sigma_r \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix} = U^\dagger \mathbf{b} = \begin{pmatrix} \mathbf{u}_1^\dagger \\ \vdots \\ \mathbf{u}_N^\dagger \end{pmatrix} \mathbf{b}$$

General LS solution

Proof.

- The cost function is minimum for $y_i = c_i/\sigma_i$, $i = 1, \ldots, r$:

$$V^\dagger \boldsymbol{\theta} \equiv \mathbf{y} = \begin{pmatrix} \mathbf{u}_1^\dagger \Big/ \sigma_1 \\ \vdots \\ \mathbf{u}_r^\dagger \Big/ \sigma_r \\ \mathbf{0} \end{pmatrix} \mathbf{b} = \Sigma^+ U^\dagger \mathbf{b}$$

General LS solution

Proof.

- The cost function is minimum for $y_i = c_i/\sigma_i$, $i = 1, \ldots, r$:

$$V^\dagger \boldsymbol{\theta} \equiv \mathbf{y} = \begin{pmatrix} \mathbf{u}_1^\dagger \big/ \sigma_1 \\ \vdots \\ \mathbf{u}_r^\dagger \big/ \sigma_r \\ \mathbf{0} \end{pmatrix} \mathbf{b} = \Sigma^+ U^\dagger \mathbf{b}$$

- $\hat{\boldsymbol{\theta}} = V\mathbf{y} = V\Sigma^+ U^\dagger \mathbf{b} = \sum_{i=1}^{r} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^\dagger \mathbf{b}$.

General LS solution

Proof.

- The cost function is minimum for $y_i = c_i/\sigma_i$, $i = 1, \ldots, r$:

$$V^\dagger \boldsymbol{\theta} \equiv \mathbf{y} = \begin{pmatrix} \mathbf{u}_1^\dagger \big/ \sigma_1 \\ \vdots \\ \mathbf{u}_r^\dagger \big/ \sigma_r \\ \mathbf{0} \end{pmatrix} \mathbf{b} = \Sigma^+ U^\dagger \mathbf{b}$$

- $\hat{\boldsymbol{\theta}} = V\mathbf{y} = V\Sigma^+ U^\dagger \mathbf{b} = \sum_{i=1}^{r} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^\dagger \mathbf{b}$.
- Any other solution can be written in the form: $\hat{\boldsymbol{\theta}} + \mathbf{v}_{\text{ker}} = \sum_{i=1}^{r} \frac{\mathbf{u}_i^\dagger \mathbf{b}}{\sigma_i} \mathbf{v}_i + \sum_{i=r+1}^{p} a_i \mathbf{v}_i$.

General LS solution

Proof.

- The cost function is minimum for $y_i = c_i/\sigma_i$, $i = 1, \ldots, r$:

$$V^\dagger \boldsymbol{\theta} \equiv \mathbf{y} = \begin{pmatrix} \mathbf{u}_1^\dagger \big/ \sigma_1 \\ \vdots \\ \mathbf{u}_r^\dagger \big/ \sigma_r \\ \mathbf{0} \end{pmatrix} \mathbf{b} = \boldsymbol{\Sigma}^+ U^\dagger \mathbf{b}$$

- $\hat{\boldsymbol{\theta}} = V\mathbf{y} = V\boldsymbol{\Sigma}^+ U^\dagger \mathbf{b} = \sum_{i=1}^{r} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^\dagger \mathbf{b}$.
- Any other solution can be written in the form: $\hat{\boldsymbol{\theta}} + \mathbf{v}_{\text{ker}} = \sum_{i=1}^{r} \frac{\mathbf{u}_i^\dagger \mathbf{b}}{\sigma_i} \mathbf{v}_i + \sum_{i=r+1}^{p} a_i \mathbf{v}_i$.
- Since $i \le r, j > r \Rightarrow \mathbf{v}_i^\dagger \mathbf{v}_j = 0$, then $\hat{\boldsymbol{\theta}} \perp \mathbf{v}_{\text{ker}} \Rightarrow \left\| \hat{\boldsymbol{\theta}} + \mathbf{v}_{\text{ker}} \right\|^2 = \left\| \hat{\boldsymbol{\theta}} \right\|^2 + \left\| \mathbf{v}_{\text{ker}} \right\|^2 \ge \left\| \hat{\boldsymbol{\theta}} \right\|^2$

General LS solution

Remark 1

- In general the pseudoinverse A^+ of a matrix A is exactly the inverse A^{-1} when A is invertible, i.e. when A is a full-rank square matrix.

General LS solution

Remark 1

- In general the pseudoinverse A^+ of a matrix A is exactly the inverse A^{-1} when A is invertible, i.e. when A is a full-rank square matrix.
- Indeed, be $A = U\Sigma V^\dagger$ and $A^+ = V\Sigma^+ U^\dagger$; since A is square, both Σ and Σ^+ are square; since A is full-rank, all the diagonal elements of both Σ and Σ^+ are non-zero, hence $\Sigma\Sigma^+ = \Sigma^+\Sigma = I$.

General LS solution

Remark 1

- In general the pseudoinverse A^+ of a matrix A is exactly the inverse A^{-1} when A is invertible, i.e. when A is a full-rank square matrix.

- Indeed, be $A = U\Sigma V^\dagger$ and $A^+ = V\Sigma^+ U^\dagger$; since A is square, both Σ and Σ^+ are square; since A is full-rank, all the diagonal elements of both Σ and Σ^+ are non-zero, hence $\Sigma\Sigma^+ = \Sigma^+\Sigma = I$.

- Thus:

$$AA^+ = U\Sigma V^\dagger V\Sigma^+ U^\dagger = U\Sigma\Sigma^+ U^\dagger = UIU^\dagger = I$$
$$A^+A = V\Sigma^+ U^\dagger U\Sigma V^\dagger = V\Sigma^+\Sigma V^\dagger = VIV^\dagger = I$$

General LS solution

Remark 2

- In general $(AB)^+ \neq B^+A^+$, but for some special cases the equality holds true.

General LS solution

Remark 2

- In general $(AB)^+ \neq B^+A^+$, but for some special cases the equality holds true.
- Later on, the following inverses should be expressed as a function of U, $\mathbf{\Sigma}$ and V:

$$\left(H^\dagger H\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p},\ \text{rank}H = p \leq N;$$

$$\left(HH^\dagger\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p},\ \text{rank}H = N \leq p;$$

General LS solution

Remark 2

- In general $(AB)^+ \neq B^+A^+$, but for some special cases the equality holds true.
- Later on, the following inverses should be expressed as a function of U, Σ and V:

$$\left(H^\dagger H\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p},\ \text{rank} H = p \leq N;$$

$$\left(HH^\dagger\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p},\ \text{rank} H = N \leq p;$$

- $\left(H^\dagger H\right)^{-1} = \left(H^\dagger H\right)^+ = H^+H^{+\dagger}$ and $\left(HH^\dagger\right)^{-1} = \left(HH^\dagger\right)^+ = H^{+\dagger}H^+$ are valid.

General LS solution

Remark 2

- In general $(AB)^+ \neq B^+A^+$, but for some special cases the equality holds true.
- Later on, the following inverses should be expressed as a function of U, Σ and V:

$$\left(H^\dagger H\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p},\ \text{rank}H = p \leq N;$$
$$\left(HH^\dagger\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p},\ \text{rank}H = N \leq p;$$

- $(H^\dagger H)^{-1} = (H^\dagger H)^+ = H^+H^{+\dagger}$ and $(HH^\dagger)^{-1} = (HH^\dagger)^+ = H^{+\dagger}H^+$ are valid.
- Indeed, in the first case, $H^\dagger H = \left(V\Sigma^T U^\dagger\right)\left(U\Sigma V^\dagger\right) = V\Sigma^T\Sigma V^\dagger$, and $H^+H^{+\dagger} = \left(V\Sigma^+U^\dagger\right)\left(U\Sigma^{+T}V^\dagger\right) = V\Sigma^+\Sigma^{+T}V^\dagger$

General LS solution

Remark 2

- In general $(AB)^+ \neq B^+A^+$, but for some special cases the equality holds true.
- Later on, the following inverses should be expressed as a function of U, Σ and V:
$$\left(H^\dagger H\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p}, \ \text{rank}H = p \leq N;$$
$$\left(HH^\dagger\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p}, \ \text{rank}H = N \leq p;$$
- $\left(H^\dagger H\right)^{-1} = \left(H^\dagger H\right)^{+} = H^+H^{+\dagger}$ and $\left(HH^\dagger\right)^{-1} = \left(HH^\dagger\right)^{+} = H^{+\dagger}H^+$ are valid.
- Indeed, in the first case, $H^\dagger H = \left(V\Sigma^T U^\dagger\right)\left(U\Sigma V^\dagger\right) = V\Sigma^T\Sigma V^\dagger$, and $H^+H^{+\dagger} = \left(V\Sigma^+U^\dagger\right)\left(U\Sigma^{+T}V^\dagger\right) = V\Sigma^+\Sigma^{+T}V^\dagger$
- Since Σ is $N \times p$, and all the elements on the main diagonal are non-zero, then $\Sigma^T\Sigma = \Sigma_p^2 = \text{diag}\left(\sigma_1^2 \cdots \sigma_p^2\right)$. Similarly, $\Sigma^+\Sigma^{+T} = \Sigma_p^{+2} = \text{diag}\left(1/\sigma_1^2 \cdots 1/\sigma_p^2\right)$. Hence $\Sigma_p^2\Sigma_p^{+2} = \Sigma_p^{+2}\Sigma_p^2 = I$.

General LS solution

Remark 2

- In general $(AB)^+ \neq B^+A^+$, but for some special cases the equality holds true.
- Later on, the following inverses should be expressed as a function of U, Σ and V:

$$\left(H^\dagger H\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p},\ \text{rank}H = p \leq N;$$

$$\left(HH^\dagger\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p},\ \text{rank}H = N \leq p;$$

- $\left(H^\dagger H\right)^{-1} = \left(H^\dagger H\right)^+ = H^+H^{+\dagger}$ and $\left(HH^\dagger\right)^{-1} = \left(HH^\dagger\right)^+ = H^{+\dagger}H^+$ are valid.
- Indeed, in the first case, $H^\dagger H = \left(V\Sigma^T U^\dagger\right)\left(U\Sigma V^\dagger\right) = V\Sigma^T\Sigma V^\dagger$, and $H^+H^{+\dagger} = \left(V\Sigma^+U^\dagger\right)\left(U\Sigma^{+T}V^\dagger\right) = V\Sigma^+\Sigma^{+T}V^\dagger$
- Since Σ is $N \times p$, and all the elements on the main diagonal are non-zero, then $\Sigma^T\Sigma = \Sigma_p^2 = \text{diag}\left(\sigma_1^2 \cdots \sigma_p^2\right)$. Similarly, $\Sigma^+\Sigma^{+T} = \Sigma_p^{+2} = \text{diag}\left(1/\sigma_1^2 \cdots 1/\sigma_p^2\right)$. Hence $\Sigma_p^2\Sigma_p^{+2} = \Sigma_p^{+2}\Sigma_p^2 = I$.
- With these expressions it is easy to verify that

$$H^+H^{+\dagger} = V\Sigma^+\Sigma^{+T}V^\dagger = \left(H^\dagger H\right)^{-1} = \left(H^\dagger H\right)^+$$

General LS solution

Remark 2

- In general $(AB)^{+} \neq B^{+}A^{+}$, but for some special cases the equality holds true.
- Later on, the following inverses should be expressed as a function of U, Σ and V:

$$\left(H^{\dagger}H\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p}, \ \mathrm{rank}H = p \leq N;$$

$$\left(HH^{\dagger}\right)^{-1} \text{ for } H \in \mathbb{C}^{N\times p}, \ \mathrm{rank}H = N \leq p;$$

- $\left(H^{\dagger}H\right)^{-1} = \left(H^{\dagger}H\right)^{+} = H^{+}H^{+\dagger}$ and $\left(HH^{\dagger}\right)^{-1} = \left(HH^{\dagger}\right)^{+} = H^{+\dagger}H^{+}$ are valid.
- Indeed, in the first case, $H^{\dagger}H = \left(V\Sigma^{T}U^{\dagger}\right)\left(U\Sigma V^{\dagger}\right) = V\Sigma^{T}\Sigma V^{\dagger}$, and $H^{+}H^{+\dagger} = \left(V\Sigma^{+}U^{\dagger}\right)\left(U\Sigma^{+T}V^{\dagger}\right) = V\Sigma^{+}\Sigma^{+T}V^{\dagger}$
- Since Σ is $N \times p$, and all the elements on the main diagonal are non-zero, then $\Sigma^{T}\Sigma = \Sigma_p^2 = \mathrm{diag}\left(\sigma_1^2 \cdots \sigma_p^2\right)$. Similarly, $\Sigma^{+}\Sigma^{+T} = \Sigma_p^{+2} = \mathrm{diag}\left(1/\sigma_1^2 \cdots 1/\sigma_p^2\right)$. Hence $\Sigma_p^2\Sigma_p^{+2} = \Sigma_p^{+2}\Sigma_p^2 = I$.
- With these expressions it is easy to verify that

$$H^{+}H^{+\dagger} = V\Sigma^{+}\Sigma^{+T}V^{\dagger} = \left(H^{\dagger}H\right)^{-1} = \left(H^{\dagger}H\right)^{+}$$

- Similarly, it can be proved that:

$$H^{+\dagger}H^{+} = U\Sigma^{+T}\Sigma^{+}U^{\dagger} = \left(HH^{\dagger}\right)^{-1} = \left(HH^{\dagger}\right)^{+}$$

General LS solution

Proof.

- $HH^+H = U\Sigma V^\dagger V\Sigma^+ U^\dagger U\Sigma V^\dagger = U\Sigma\Sigma^+\Sigma V^\dagger = U\Sigma V^\dagger = H$

General LS solution

Proof.

- $HH^{+}H = U\Sigma V^{\dagger} V \Sigma^{+} U^{\dagger} U \Sigma V^{\dagger} = U\Sigma\Sigma^{+}\Sigma V^{\dagger} = U\Sigma V^{\dagger} = H$
- $N \geq p,\ \ r = p \Rightarrow \Sigma^{+}\Sigma = I \Rightarrow H^{+}H = V\Sigma^{+}U^{\dagger}U\Sigma V^{\dagger} = V\Sigma^{+}\Sigma V^{\dagger} = VV^{\dagger} = I;$
- $N \geq p,\ \ r = p \Rightarrow \left(H^{\dagger}H\right)^{-1}H^{\dagger} = \left(H^{\dagger}H\right)^{+}H^{\dagger} = V\Sigma^{+}\Sigma^{+T}V^{\dagger}V\Sigma^{T}U^{\dagger} = V\Sigma^{+}U^{\dagger}$

General LS solution

Proof.

- $HH^+H = U\Sigma V^\dagger V\Sigma^+ U^\dagger U\Sigma V^\dagger = U\Sigma\Sigma^+\Sigma V^\dagger = U\Sigma V^\dagger = H$
- $N \geq p,\ \ r = p \Rightarrow \Sigma^+\Sigma = I \Rightarrow H^+H = V\Sigma^+ U^\dagger U\Sigma V^\dagger = V\Sigma^+\Sigma V^\dagger = VV^\dagger = I;$
- $N \geq p,\ \ r = p \Rightarrow \left(H^\dagger H\right)^{-1} H^\dagger = \left(H^\dagger H\right)^+ H^\dagger = V\Sigma^+\Sigma^{+T} V^\dagger V\Sigma^T U^\dagger = V\Sigma^+ U^\dagger$
- $N \leq p,\ \ r = N \Rightarrow \Sigma\Sigma^+ = I \Rightarrow HH^+ = U\Sigma V^\dagger V\Sigma^+ U^\dagger = U\Sigma\Sigma^+ U^\dagger = UU^\dagger = I;$
- $N \leq p,\ \ r = N \Rightarrow H^\dagger\left(HH^\dagger\right)^{-1} = H^\dagger\left(HH^\dagger\right)^+ = V\Sigma^T U^\dagger U\Sigma^{+T}\Sigma^+ U^\dagger = V\Sigma^+ U^\dagger$

General LS solution

Proof.

- $HH^+H = U\Sigma V^\dagger V\Sigma^+ U^\dagger U\Sigma V^\dagger = U\Sigma\Sigma^+\Sigma V^\dagger = U\Sigma V^\dagger = H$
- $N \geq p,\ r = p \Rightarrow \Sigma^+\Sigma = I \Rightarrow H^+H = V\Sigma^+ U^\dagger U\Sigma V^\dagger = V\Sigma^+\Sigma V^\dagger = VV^\dagger = I;$
- $N \geq p,\ r = p \Rightarrow \left(H^\dagger H\right)^{-1} H^\dagger = \left(H^\dagger H\right)^{+} H^\dagger = V\Sigma^+\Sigma^{+T} V^\dagger V\Sigma^T U^\dagger = V\Sigma^+ U^\dagger$
- $N \leq p,\ r = N \Rightarrow \Sigma\Sigma^+ = I \Rightarrow HH^+ = U\Sigma V^\dagger V\Sigma^+ U^\dagger = U\Sigma\Sigma^+ U^\dagger = UU^\dagger = I;$
- $N \leq p,\ r = N \Rightarrow H^\dagger\left(HH^\dagger\right)^{-1} = H^\dagger\left(HH^\dagger\right)^{+} = V\Sigma^T U^\dagger U\Sigma^{+T}\Sigma^+ U^\dagger = V\Sigma^+ U^\dagger$
- Proof that $HH^+ = (HH^+)^\dagger$, $H^+H = (H^+H)^\dagger$ is now obvious □

General SVD pseudo-inverse

- The general form of the pseudo-inverse of $H = U\Sigma V^\dagger$ is $H^+ = V\Sigma^+ U^\dagger$.
- The unique LS solution $\hat{\boldsymbol{\theta}} = H^+\mathbf{b}$ is s.t. both $\left\|\mathbf{b} - H\hat{\boldsymbol{\theta}}\right\|^2$ and $\left\|\hat{\boldsymbol{\theta}}\right\|^2$ are minimum.
- $HH^+H = H$ is always true, but $H^+H = I$ or $HH^+ = I$ do not hold in general.
- $OLS : r = p \leq N \Rightarrow H^+ = \left(H^\dagger H\right)^{-1} H^\dagger,\ r = N < p \Rightarrow H^+ = H^\dagger\left(HH^\dagger\right)^{-1}.$
- $HH^+ = (HH^+)^\dagger,\ H^+H = (H^+H)^\dagger$

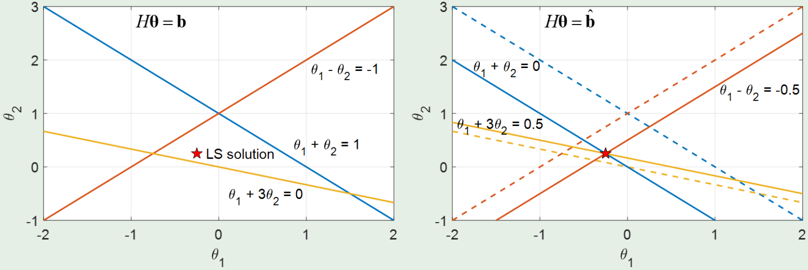
General LS solution

Example

- With Matlab, the SVD can be obtained by using the command `[U, S, V] = svd(H)`

$$H\boldsymbol{\theta} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \mathbf{b} \qquad \hat{\boldsymbol{\theta}} = \begin{pmatrix} -0.25 \\ 0.25 \end{pmatrix} \qquad \hat{\mathbf{b}} = H\hat{\boldsymbol{\theta}} = \begin{pmatrix} 0 \\ -0.5 \\ 0.5 \end{pmatrix} \qquad H\boldsymbol{\theta} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ -0.5 \\ 0.5 \end{pmatrix} = \hat{\mathbf{b}}$$

$$U = \begin{pmatrix} 0.3651 & 0.4472 & -0.8165 \\ -0.1826 & 0.8944 & 0.4082 \\ 0.9129 & 0 & 0.4082 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} 3.4641 & 0 \\ 0 & 1.4142 \\ 0 & 0 \end{pmatrix} \qquad V = \begin{pmatrix} 0.3162 & 0.9487 \\ 0.9487 & -0.3162 \end{pmatrix}$$



General LS solution

An explanatory example on V

- In the following we will have to deal with product of the form $V^\dagger V_r$ or $V_r^\dagger V$, where V_r is the matrix formed by taking the first r columns of V, hence it is useful to visualize these products. If V is $p \times p$:

$$V^\dagger V_r = \begin{bmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix}_{r\times r} \\ \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{pmatrix}_{(p-r)\times r} \end{bmatrix} = \begin{bmatrix} I_r \\ 0_{(p-r)\times r} \end{bmatrix}$$

$$V_r^\dagger V = \begin{bmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix}_{r\times r} & \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}_{r\times(p-r)} \end{bmatrix} = \begin{bmatrix} I_r & 0_{r\times(p-r)} \end{bmatrix}$$

- They can be called expansion or selection matrices and denoted by the symbol $I_{p\times r}$ or $I_{r\times p}$. Obviously, entirely similar results apply to U.

Table of Contents

Geometrical interpretation of LS

- Let us define the residual: $\mathbf{r} = \mathbf{b} - H\hat{\boldsymbol{\theta}} = \mathbf{b} - HH^{+}\mathbf{b} = \left(I - HH^{+}\right)\mathbf{b} = P_{H\perp}\mathbf{b}$

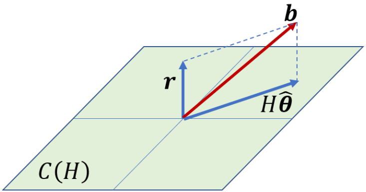
Geometrical interpretation of LS

- Let us define the residual: $\mathbf{r} = \mathbf{b} - H\hat{\boldsymbol{\theta}} = \mathbf{b} - HH^{+}\mathbf{b} = \left(I - HH^{+}\right)\mathbf{b} = P_{H\perp}\mathbf{b}$
- Let us also define $P_{H\|} = (I - P_{H\perp}) = HH^{+}$

Geometrical interpretation of LS

- Let us define the residual: $\mathbf{r} = \mathbf{b} - H\hat{\boldsymbol{\theta}} = \mathbf{b} - HH^{+}\mathbf{b} = \left(I - HH^{+}\right)\mathbf{b} = P_{H\perp}\mathbf{b}$
- Let us also define $P_{H\|} = (I - P_{H\perp}) = HH^{+}$

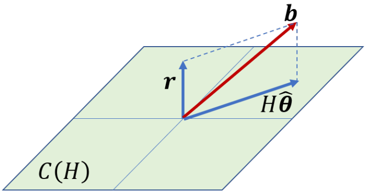
- $P_{H\perp}$ and $P_{H\|}$ are orthogonal projections



Geometrical interpretation of LS

- Let us define the residual: $\mathbf{r} = \mathbf{b} - H\hat{\boldsymbol{\theta}} = \mathbf{b} - HH^{+}\mathbf{b} = \left(I - HH^{+}\right)\mathbf{b} = P_{H\perp}\mathbf{b}$
- Let us also define $P_{H\|} = (I - P_{H\perp}) = HH^{+}$

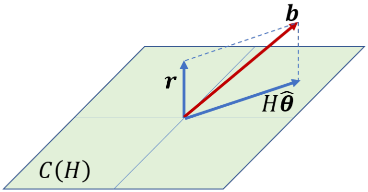
- $P_{H\perp}$ and $P_{H\|}$ are orthogonal projections
- It is straightforward to prove they are idempotent and symmetric
- $P_{H\perp}P_{H\perp} = P_{H\perp}$, $P_{H\|}P_{H\|} = P_{H\|}$ idempotency
- $P_{H\perp}{}^{\dagger} = P_{H\perp}$, $P_{H\|}{}^{\dagger} = P_{H\|}$ symmetry



Geometrical interpretation of LS

- Let us define the residual: $\mathbf{r} = \mathbf{b} - H\hat{\boldsymbol{\theta}} = \mathbf{b} - HH^{+}\mathbf{b} = \left(I - HH^{+}\right)\mathbf{b} = P_{H\perp}\mathbf{b}$
- Let us also define $P_{H\|} = (I - P_{H\perp}) = HH^{+}$

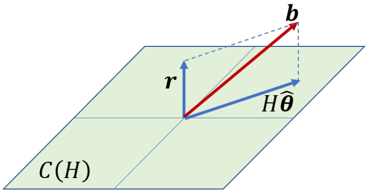
- $P_{H\perp}$ and $P_{H\|}$ are orthogonal projections
- It is straightforward to prove they are idempotent and symmetric
- $P_{H\perp}P_{H\perp} = P_{H\perp}$, $P_{H\|}P_{H\|} = P_{H\|}$ idempotency
- $P_{H\perp}{}^{\dagger} = P_{H\perp}$, $P_{H\|}{}^{\dagger} = P_{H\|}$ symmetry
- Also $P_{H\perp}P_{H\|} = 0$



Geometrical interpretation of LS

- Let us define the residual: $\mathbf{r} = \mathbf{b} - H\hat{\boldsymbol{\theta}} = \mathbf{b} - HH^{+}\mathbf{b} = \left(I - HH^{+}\right)\mathbf{b} = P_{H\perp}\mathbf{b}$
- Let us also define $P_{H\|} = (I - P_{H\perp}) = HH^{+}$

- $P_{H\perp}$ and $P_{H\|}$ are orthogonal projections
- It is straightforward to prove they are idempotent and symmetric
- $P_{H\perp}P_{H\perp} = P_{H\perp}$, $P_{H\|}P_{H\|} = P_{H\|}$ idempotency
- ${P_{H\perp}}^{\dagger} = P_{H\perp}$, ${P_{H\|}}^{\dagger} = P_{H\|}$ symmetry
- Also $P_{H\perp}P_{H\|} = 0$

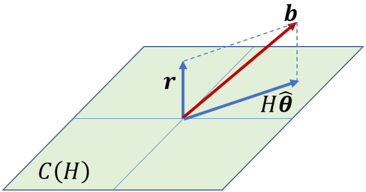


- Since $P_{H\|}\mathbf{b} = H\hat{\boldsymbol{\theta}}$, $P_{H\|}$ projects $\mathbf{b}$ onto column space $C(H)$ of H

Geometrical interpretation of LS

- Let us define the residual: $\mathbf{r} = \mathbf{b} - H\hat{\boldsymbol{\theta}} = \mathbf{b} - HH^{+}\mathbf{b} = \left(I - HH^{+}\right)\mathbf{b} = P_{H\perp}\mathbf{b}$
- Let us also define $P_{H\|} = (I - P_{H\perp}) = HH^{+}$

- $P_{H\perp}$ and $P_{H\|}$ are orthogonal projections
- It is straightforward to prove they are idempotent and symmetric
- $P_{H\perp}P_{H\perp} = P_{H\perp}$, $P_{H\|}P_{H\|} = P_{H\|}$ idempotency
- ${P_{H\perp}}^{\dagger} = P_{H\perp}$, ${P_{H\|}}^{\dagger} = P_{H\|}$ symmetry
- Also $P_{H\perp}P_{H\|} = 0$

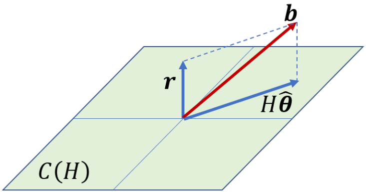


- Since $P_{H\|}\mathbf{b} = H\hat{\boldsymbol{\theta}}$, $P_{H\|}$ projects $\mathbf{b}$ onto column space $C(H)$ of H
- $P_{H\perp}$ projects $\mathbf{b}$ onto space $C_{\perp}(H)$ orthogonal to $C(H)$

Geometrical interpretation of LS

- Let us define the residual: $\mathbf{r} = \mathbf{b} - H\hat{\boldsymbol{\theta}} = \mathbf{b} - HH^{+}\mathbf{b} = \left(I - HH^{+}\right)\mathbf{b} = P_{H\perp}\mathbf{b}$
- Let us also define $P_{H\parallel} = (I - P_{H\perp}) = HH^{+}$

- $P_{H\perp}$ and $P_{H\parallel}$ are orthogonal projections
- It is straightforward to prove they are idempotent and symmetric
- $P_{H\perp}P_{H\perp} = P_{H\perp}$, $P_{H\parallel}P_{H\parallel} = P_{H\parallel}$ idempotency
- ${P_{H\perp}}^{\dagger} = P_{H\perp}$, ${P_{H\parallel}}^{\dagger} = P_{H\parallel}$ symmetry
- Also $P_{H\perp}P_{H\parallel} = 0$



- Since $P_{H\parallel}\mathbf{b} = H\hat{\boldsymbol{\theta}}$, $P_{H\parallel}$ projects $\mathbf{b}$ onto column space $C(H)$ of H
- $P_{H\perp}$ projects $\mathbf{b}$ onto space $C_{\perp}(H)$ orthogonal to $C(H)$
- The residual $\mathbf{r}$ accounts for the observed component of $\mathbf{b}$ that are not accounted for by the model $H\hat{\boldsymbol{\theta}}$

Geometrical interpretation of LS

- Since $\mathbf{r} = P_{H\perp}\mathbf{b}$, $P_{H\perp}$ is also called residual maker matrix.

Geometrical interpretation of LS

- Since $\mathbf{r} = P_{H\perp}\mathbf{b}$, $P_{H\perp}$ is also called residual maker matrix.
- Also $P_{H\perp}H = \left(I - HH^{+}\right)H = H - HH^{+}H = H - H = 0$.

Geometrical interpretation of LS

- Since $\mathbf{r} = P_{H\perp}\mathbf{b}$, $P_{H\perp}$ is also called residual maker matrix.
- Also $P_{H\perp}H = \left(I - HH^{+}\right) H = H - HH^{+}H = H - H = 0$.
- Hence, $\mathbf{r} = P_{H\perp}\mathbf{b} = P_{H\perp}\left(H\boldsymbol{\theta} + \boldsymbol{\varepsilon}\right) = P_{H\perp}\boldsymbol{\varepsilon}$

Geometrical interpretation of LS

- Since $\mathbf{r} = P_{H\perp}\mathbf{b}$, $P_{H\perp}$ is also called residual maker matrix.
- Also $P_{H\perp}H = \left(I - HH^{+}\right)H = H - HH^{+}H = H - H = 0$.
- Hence, $\mathbf{r} = P_{H\perp}\mathbf{b} = P_{H\perp}\left(H\boldsymbol{\theta} + \boldsymbol{\varepsilon}\right) = P_{H\perp}\boldsymbol{\varepsilon}$
- Thus, the cost function is $\phi = \|\mathbf{r}\|^2 = \mathbf{r}^{\dagger}\mathbf{r} = \boldsymbol{\varepsilon}^{\dagger}P_{H\perp}^{\dagger}P_{H\perp}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\dagger}P_{H\perp}\boldsymbol{\varepsilon}$

Geometrical interpretation of LS

- Since $\mathbf{r} = P_{H\perp}\mathbf{b}$, $P_{H\perp}$ is also called residual maker matrix.
- Also $P_{H\perp}H = \left(I - HH^{+}\right)H = H - HH^{+}H = H - H = 0$.
- Hence, $\mathbf{r} = P_{H\perp}\mathbf{b} = P_{H\perp}\left(H\boldsymbol{\theta} + \boldsymbol{\varepsilon}\right) = P_{H\perp}\boldsymbol{\varepsilon}$
- Thus, the cost function is $\phi = \|\mathbf{r}\|^2 = \mathbf{r}^\dagger\mathbf{r} = \boldsymbol{\varepsilon}^\dagger P_{H\perp}^\dagger P_{H\perp}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^\dagger P_{H\perp}\boldsymbol{\varepsilon}$
- The expected value can be computed easily:

$$
\begin{aligned}
\mathrm{E}\left[\phi\left(\hat{\boldsymbol{\theta}}\right)\right] &= \mathrm{E}\left[\boldsymbol{\varepsilon}^\dagger P_{H\perp}\boldsymbol{\varepsilon}\right] = \mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{\varepsilon}^\dagger P_{H\perp}\boldsymbol{\varepsilon}\right)\right] = \mathrm{E}\left[\mathrm{tr}\left(P_{H\perp}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\dagger\right)\right] = \\
&= \mathrm{tr}\left(P_{H\perp}\mathrm{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\dagger\right]\right) = \mathrm{tr}\left(P_{H\perp}\mathrm{cov}\left[\boldsymbol{\varepsilon}\right]\right) = \mathrm{tr}\left(P_{H\perp}\sigma^2 I\right) = \sigma^2\mathrm{tr}P_{H\perp} \\
\mathrm{tr}P_{H\perp} &= \mathrm{tr}\left(I_N - HH^{+}\right) = \mathrm{tr}\left(I_N - U\Sigma\Sigma^{+}U^\dagger\right) = N - \mathrm{tr}\left(\Sigma^{+}U^\dagger U\Sigma\right) = \\
&= N - \mathrm{tr}\left(\Sigma^{+}\Sigma\right) = N - \mathrm{tr}I_p^r = N - r
\end{aligned}
$$

Geometrical interpretation of LS

- Since $\mathbf{r} = P_{H\perp}\mathbf{b}$, $P_{H\perp}$ is also called residual maker matrix.
- Also $P_{H\perp}H = \left(I - HH^{+}\right)H = H - HH^{+}H = H - H = 0$.
- Hence, $\mathbf{r} = P_{H\perp}\mathbf{b} = P_{H\perp}\left(H\boldsymbol{\theta} + \boldsymbol{\varepsilon}\right) = P_{H\perp}\boldsymbol{\varepsilon}$
- Thus, the cost function is $\phi = \|\mathbf{r}\|^2 = \mathbf{r}^\dagger\mathbf{r} = \boldsymbol{\varepsilon}^\dagger P_{H\perp}^\dagger P_{H\perp}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^\dagger P_{H\perp}\boldsymbol{\varepsilon}$
- The expected value can be computed easily:

$$\mathrm{E}\left[\phi\left(\hat{\boldsymbol{\theta}}\right)\right] = \mathrm{E}\left[\boldsymbol{\varepsilon}^\dagger P_{H\perp}\boldsymbol{\varepsilon}\right] = \mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{\varepsilon}^\dagger P_{H\perp}\boldsymbol{\varepsilon}\right)\right] = \mathrm{E}\left[\mathrm{tr}\left(P_{H\perp}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\dagger\right)\right] =$$

$$= \mathrm{tr}\left(P_{H\perp}\mathrm{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\dagger\right]\right) = \mathrm{tr}\left(P_{H\perp}\mathrm{cov}\left[\boldsymbol{\varepsilon}\right]\right) = \mathrm{tr}\left(P_{H\perp}\sigma^2 I\right) = \sigma^2\mathrm{tr}P_{H\perp}$$

$$\mathrm{tr}P_{H\perp} = \mathrm{tr}\left(I_N - HH^{+}\right) = \mathrm{tr}\left(I_N - U\Sigma\Sigma^{+}U^\dagger\right) = N - \mathrm{tr}\left(\Sigma^{+}U^\dagger U\Sigma\right) =$$

$$= N - \mathrm{tr}\left(\Sigma^{+}\Sigma\right) = N - \mathrm{tr}I_p^r = N - r$$

Estimator of σ^2

If σ^2 is not known a priori, an unbiased estimator can be obtained from the residual:

$$\hat{\sigma}^2 = \frac{\phi\left(\hat{\boldsymbol{\theta}}\right)}{N-r} = \frac{\left\|\mathbf{r}\left(\hat{\boldsymbol{\theta}}\right)\right\|^2}{N-r} \Rightarrow \mathrm{E}\left[\hat{\sigma}^2\right] = \frac{\mathrm{E}\left[\phi\left(\hat{\boldsymbol{\theta}}\right)\right]}{N-r} = \frac{\sigma^2\left(N-r\right)}{N-r} = \sigma^2$$

Table of Contents

Properties of the general LS estimator

Covariance of the general LS estimator

1. For the general LS estimator, when $\mathrm{cov}\left[\boldsymbol{\varepsilon}\right] = \sigma^2 I$:

$$\hat{\boldsymbol{\theta}} = V\boldsymbol{\Sigma}^{+}U^{\dagger}\mathbf{b} \Rightarrow \mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2 V\boldsymbol{\Sigma}^{+}\boldsymbol{\Sigma}^{+T}V^{\dagger}$$

2. When $N \geq p$, and $\mathrm{rank}H = p$ (OLS):

$$\hat{\boldsymbol{\theta}} = \left(H^{\dagger}H\right)^{-1}H^{\dagger}\mathbf{b} \Rightarrow \mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2\left(H^{\dagger}H\right)^{-1}$$

3. When $N < p$, and $\mathrm{rank}H = N$:

$$\hat{\boldsymbol{\theta}} = H^{\dagger}\left(HH^{\dagger}\right)^{-1}\mathbf{b} \Rightarrow \mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2 H^{\dagger}\left(HH^{\dagger}\right)^{-2}H$$

The general covariance expression 1 yields the same values as the particular expressions 2 and 3, valid under the specified assumptions.

Properties of the general LS estimator

Proof.

1. $\text{cov}\,[\mathbf{b}] = \text{cov}\,[\varepsilon] = \sigma^2 I \Rightarrow$

$$\text{cov}\left[\hat{\boldsymbol{\theta}}\right] = \text{cov}\left[H^{+}\mathbf{b}\right] = \text{cov}\left[V\Sigma^{+}U^{\dagger}\mathbf{b}\right] = V\Sigma^{+}U^{\dagger}\text{cov}\,[\varepsilon]\,U\Sigma^{+T}V^{\dagger} = \sigma^2 V\Sigma^{+}\Sigma^{+T}V^{\dagger}$$

Properties of the general LS estimator

Proof.

1. $\text{cov}\left[\mathbf{b}\right] = \text{cov}\left[\boldsymbol{\varepsilon}\right] = \sigma^2 I \Rightarrow$

$$\text{cov}\left[\hat{\boldsymbol{\theta}}\right] = \text{cov}\left[H^+\mathbf{b}\right] = \text{cov}\left[V\Sigma^+ U^\dagger \mathbf{b}\right] = V\Sigma^+ U^\dagger \text{cov}\left[\boldsymbol{\varepsilon}\right] U\Sigma^{+T} V^\dagger = \sigma^2 V\Sigma^+\Sigma^{+T} V^\dagger$$

2. $\text{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2 \left(H^\dagger H\right)^{-1} = \sigma^2 \left(H^\dagger H\right)^{+} = \sigma^2 V\Sigma^+ U^\dagger U\Sigma^{+T} V^\dagger = \sigma^2 V\Sigma^+\Sigma^{+T} V^\dagger$

Properties of the general LS estimator

Proof.

1. $\text{cov}\,[\mathbf{b}] = \text{cov}\,[\varepsilon] = \sigma^2 I \Rightarrow$

$$\text{cov}\left[\hat{\boldsymbol{\theta}}\right] = \text{cov}\left[H^+\mathbf{b}\right] = \text{cov}\left[V\Sigma^+U^\dagger\mathbf{b}\right] = V\Sigma^+U^\dagger\text{cov}\,[\varepsilon]\,U\Sigma^{+T}V^\dagger = \sigma^2 V\Sigma^+\Sigma^{+T}V^\dagger$$

2. $\text{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2\left(H^\dagger H\right)^{-1} = \sigma^2\left(H^\dagger H\right)^{+} = \sigma^2 V\Sigma^+U^\dagger U\Sigma^{+T}V^\dagger = \sigma^2 V\Sigma^+\Sigma^{+T}V^\dagger$

3. $\text{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2 H^\dagger\left(HH^\dagger\right)^{-1}\left(HH^\dagger\right)^{-1}H =$

$$\sigma^2 V\Sigma^T\Sigma^{+T}\Sigma^+\Sigma^{+T}\Sigma^+\Sigma V^\dagger = \sigma^2 V\Sigma^+\Sigma\Sigma^+\Sigma^{+T}\Sigma^+\Sigma V^\dagger$$

Since $\Sigma\Sigma^+ = I$ when $\text{rank}H = N$, we get

$$\sigma^2 V\Sigma^+\Sigma^{+T}\Sigma^+\Sigma V^\dagger = \sigma^2 V\Sigma^+\Sigma^{+T}\Sigma^T\Sigma^{+T}V^\dagger = \sigma^2 V\Sigma^+\Sigma^{+T}\Sigma^T\Sigma^{+T}V^\dagger =$$

$$= \sigma^2 V\Sigma^+\left(\Sigma\Sigma^+\right)^T\Sigma^{+T}V^\dagger = \sigma^2 V\Sigma^+\Sigma^{+T}V^\dagger$$

Properties of the general LS estimator

- Is the Gauss-Markov theorem valid for the general LS estimator $\hat{\boldsymbol{\theta}} = V\boldsymbol{\Sigma}^{+}U^{\dagger}\mathbf{b}$?

Properties of the general LS estimator

- Is the Gauss-Markov theorem valid for the general LS estimator $\hat{\boldsymbol{\theta}} = V\boldsymbol{\Sigma}^{+}U^{\dagger}\mathbf{b}$?
- Gauss-Markov theorem assumptions:
  - Homoscedasticity: OK (always attainable by using weigths)

Properties of the general LS estimator

- Is the Gauss-Markov theorem valid for the general LS estimator $\hat{\boldsymbol{\theta}} = V\boldsymbol{\Sigma}^{+}U^{\dagger}\mathbf{b}$?

- Gauss-Markov theorem assumptions:
  - Homoscedasticity: OK (always attainable by using weigths)
  - LS estimator is unbiased: let's check...

Properties of the general LS estimator

- Is the Gauss-Markov theorem valid for the general LS estimator $\hat{\boldsymbol{\theta}} = V\Sigma^{+}U^{\dagger}\mathbf{b}$?
- Gauss-Markov theorem assumptions:
  - Homoscedasticity: OK (always attainable by using weigths)
  - LS estimator is unbiased: let's check...

$$\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \mathrm{E}\left[V\Sigma^{+}U^{\dagger}\mathbf{b}\right] = \mathrm{E}\left[V\Sigma^{+}U^{\dagger}\left(H\boldsymbol{\theta} + \boldsymbol{\varepsilon}\right)\right] = \mathrm{E}\left[V\Sigma^{+}U^{\dagger}\left(U\Sigma V^{\dagger}\boldsymbol{\theta} + \boldsymbol{\varepsilon}\right)\right] =$$

$$= V\Sigma^{+}\Sigma V^{\dagger}\boldsymbol{\theta} + V\Sigma^{+}U^{\dagger}\mathrm{E}\left[\boldsymbol{\varepsilon}\right] = V\Sigma^{+}\Sigma V^{\dagger}\boldsymbol{\theta}$$

Properties of the general LS estimator

- Is the Gauss-Markov theorem valid for the general LS estimator $\hat{\boldsymbol{\theta}} = V\Sigma^{+}U^{\dagger}\mathbf{b}$?
- Gauss-Markov theorem assumptions:
  - Homoscedasticity: OK (always attainable by using weigths)
  - LS estimator is unbiased: let's check...

$$\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \mathrm{E}\left[V\Sigma^{+}U^{\dagger}\mathbf{b}\right] = \mathrm{E}\left[V\Sigma^{+}U^{\dagger}\left(H\boldsymbol{\theta}+\boldsymbol{\varepsilon}\right)\right] = \mathrm{E}\left[V\Sigma^{+}U^{\dagger}\left(U\Sigma V^{\dagger}\boldsymbol{\theta}+\boldsymbol{\varepsilon}\right)\right] =$$
$$= V\Sigma^{+}\Sigma V^{\dagger}\boldsymbol{\theta} + V\Sigma^{+}U^{\dagger}\mathrm{E}\left[\boldsymbol{\varepsilon}\right] = V\Sigma^{+}\Sigma V^{\dagger}\boldsymbol{\theta}$$

- If $r = \mathrm{rank} H = p$ then $\Sigma^{+}\Sigma = I \Rightarrow \mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$ but in general, for any r, $\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] \neq \boldsymbol{\theta}$

Properties of the general LS estimator

- Is the Gauss-Markov theorem valid for the general LS estimator $\hat{\boldsymbol{\theta}} = V\Sigma^{+}U^{\dagger}\mathbf{b}$?

- Gauss-Markov theorem assumptions:
  - Homoscedasticity: OK (always attainable by using weigths)
  - LS estimator is unbiased: let's check...

$$\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \mathrm{E}\left[V\Sigma^{+}U^{\dagger}\mathbf{b}\right] = \mathrm{E}\left[V\Sigma^{+}U^{\dagger}\left(H\boldsymbol{\theta}+\boldsymbol{\varepsilon}\right)\right] = \mathrm{E}\left[V\Sigma^{+}U^{\dagger}\left(U\Sigma V^{\dagger}\boldsymbol{\theta}+\boldsymbol{\varepsilon}\right)\right] =$$
$$= V\Sigma^{+}\Sigma V^{\dagger}\boldsymbol{\theta} + V\Sigma^{+}U^{\dagger}\mathrm{E}\left[\boldsymbol{\varepsilon}\right] = V\Sigma^{+}\Sigma V^{\dagger}\boldsymbol{\theta}$$

- If $r = \mathrm{rank} H = p$ then $\Sigma^{+}\Sigma = I \Rightarrow \mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$ but in general, for any r, $\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] \neq \boldsymbol{\theta}$

- Gauss-Markov is not valid for the general LS estimator, hence in general $\hat{\boldsymbol{\theta}}$ is not BLUE.

Properties of the general LS estimator

- Is the Gauss-Markov theorem valid for the general LS estimator $\hat{\boldsymbol{\theta}} = V\Sigma^+ U^\dagger \mathbf{b}$?
- Gauss-Markov theorem assumptions:
  - Homoscedasticity: OK (always attainable by using weigths)
  - LS estimator is unbiased: let's check...

$$\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \mathrm{E}\left[V\Sigma^+ U^\dagger \mathbf{b}\right] = \mathrm{E}\left[V\Sigma^+ U^\dagger (H\boldsymbol{\theta} + \boldsymbol{\varepsilon})\right] = \mathrm{E}\left[V\Sigma^+ U^\dagger \left(U\Sigma V^\dagger \boldsymbol{\theta} + \boldsymbol{\varepsilon}\right)\right] =$$
$$= V\Sigma^+\Sigma V^\dagger \boldsymbol{\theta} + V\Sigma^+ U^\dagger \mathrm{E}\left[\boldsymbol{\varepsilon}\right] = V\Sigma^+\Sigma V^\dagger \boldsymbol{\theta}$$

- If $r = \mathrm{rank} H = p$ then $\Sigma^+\Sigma = I \Rightarrow \mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] = \boldsymbol{\theta}$ but in general, for any r, $\mathrm{E}\left[\hat{\boldsymbol{\theta}}\right] \neq \boldsymbol{\theta}$
- Gauss-Markov is not valid for the general LS estimator, hence in general $\hat{\boldsymbol{\theta}}$ is not BLUE.
- We will see how, for any rank r, it is always possible to extract r independent BLUE estimators from $\hat{\boldsymbol{\theta}}$.

Table of Contents

Generalized Gauss-Markov theorem

- $C(H)$ and $R(H)$ are respectively the column space and the row space of the matrix H.

Generalized Gauss-Markov theorem

- $C(H)$ and $R(H)$ are respectively the column space and the row space of the matrix H.
- $\dim C = \dim R = \mathrm{rank} H = r$

Generalized Gauss-Markov theorem

- $C(H)$ and $R(H)$ are respectively the column space and the row space of the matrix H.
- $\dim C = \dim R = \mathrm{rank} H = r$
- Any vector $\boldsymbol{\lambda}^{\dagger} \in R(H)$ can be written as $\boldsymbol{\lambda}^{\dagger} = \mathbf{a}^{\dagger} H \Leftrightarrow \boldsymbol{\lambda} = H^{\dagger}\mathbf{a}$ for some $\mathbf{a}$, i.e., $\boldsymbol{\lambda}^{\dagger} \in R(H) \Leftrightarrow \boldsymbol{\lambda} \in C(H^{\dagger})$

Generalized Gauss-Markov theorem

- $C(H)$ and $R(H)$ are respectively the column space and the row space of the matrix H.
- $\dim C = \dim R = \text{rank} H = r$
- Any vector $\boldsymbol{\lambda}^\dagger \in R(H)$ can be written as $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H \Leftrightarrow \boldsymbol{\lambda} = H^\dagger \mathbf{a}$ for some $\mathbf{a}$, i.e., $\boldsymbol{\lambda}^\dagger \in R(H) \Leftrightarrow \boldsymbol{\lambda} \in C(H^\dagger)$
- Statement of the theorem (proof will require an additional theoretical framework):

Generalized Gauss-Markov theorem

- $C(H)$ and $R(H)$ are respectively the column space and the row space of the matrix H.
- $\dim C = \dim R = \text{rank} H = r$
- Any vector $\boldsymbol{\lambda}^\dagger \in R(H)$ can be written as $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H \Leftrightarrow \boldsymbol{\lambda} = H^\dagger \mathbf{a}$ for some $\mathbf{a}$, i.e., $\boldsymbol{\lambda}^\dagger \in R(H) \Leftrightarrow \boldsymbol{\lambda} \in C(H^\dagger)$
- Statement of the theorem (proof will require an additional theoretical framework):

Generalized Gauss-Markov theorem

- Given any system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ with N equations and p unknown parameters, s.t. $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$ and $\mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I_N$.
- Be $r = \text{rank} H \leq \min(N, p)$ and $\hat{\boldsymbol{\theta}} = H^+ \mathbf{b} = V\boldsymbol{\Sigma}^+ U^\dagger \mathbf{b}$ the generalized LS estimator.
- Be $\boldsymbol{\lambda}_i^\dagger$, $i = 1 \cdots r$, any set of linearly independent vectors $\in R(H)$.
- Then, $\boldsymbol{\lambda}_i^\dagger \hat{\boldsymbol{\theta}}$ are unbiased minimum variance estimators of $\boldsymbol{\lambda}_i^\dagger \boldsymbol{\theta}$ and are BLUE.

Generalized Gauss-Markov theorem

- $C(H)$ and $R(H)$ are respectively the column space and the row space of the matrix H.
- $\dim C = \dim R = \mathrm{rank} H = r$
- Any vector $\boldsymbol{\lambda}^\dagger \in R(H)$ can be written as $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H \Leftrightarrow \boldsymbol{\lambda} = H^\dagger \mathbf{a}$ for some $\mathbf{a}$, i.e., $\boldsymbol{\lambda}^\dagger \in R(H) \Leftrightarrow \boldsymbol{\lambda} \in C(H^\dagger)$
- Statement of the theorem (proof will require an additional theoretical framework):

Generalized Gauss-Markov theorem

- Given any system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ with N equations and p unknown parameters, s.t. $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$ and $\mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I_N$.
- Be $r = \mathrm{rank} H \leq \min(N, p)$ and $\hat{\boldsymbol{\theta}} = H^+\mathbf{b} = V\boldsymbol{\Sigma}^+ U^\dagger \mathbf{b}$ the generalized LS estimator.
- Be $\boldsymbol{\lambda}_i^\dagger$, $i = 1 \cdots r$, any set of linearly independent vectors $\in R(H)$.
- Then, $\boldsymbol{\lambda}_i^\dagger \hat{\boldsymbol{\theta}}$ are unbiased minimum variance estimators of $\boldsymbol{\lambda}_i^\dagger \boldsymbol{\theta}$ and are BLUE.

- The theorem states that it is always possible to find at most r linear combinations of the components of $\hat{\boldsymbol{\theta}}$, which are BLUE estimators.

Generalized Gauss-Markov theorem

- $C(H)$ and $R(H)$ are respectively the column space and the row space of the matrix H.
- $\dim C = \dim R = \mathrm{rank} H = r$
- Any vector $\boldsymbol{\lambda}^\dagger \in R(H)$ can be written as $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H \Leftrightarrow \boldsymbol{\lambda} = H^\dagger \mathbf{a}$ for some $\mathbf{a}$, i.e., $\boldsymbol{\lambda}^\dagger \in R(H) \Leftrightarrow \boldsymbol{\lambda} \in C(H^\dagger)$
- Statement of the theorem (proof will require an additional theoretical framework):

Generalized Gauss-Markov theorem

- Given any system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ with N equations and p unknown parameters, s.t. $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$ and $\mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I_N$.
- Be $r = \mathrm{rank} H \leq \min(N, p)$ and $\hat{\boldsymbol{\theta}} = H^+\mathbf{b} = V\boldsymbol{\Sigma}^+ U^\dagger \mathbf{b}$ the generalized LS estimator.
- Be $\boldsymbol{\lambda}_i^\dagger$, $i = 1 \cdots r$, any set of linearly independent vectors $\in R(H)$.
- Then, $\boldsymbol{\lambda}_i^\dagger \hat{\boldsymbol{\theta}}$ are unbiased minimum variance estimators of $\boldsymbol{\lambda}_i^\dagger \boldsymbol{\theta}$ and are BLUE.

- The theorem states that it is always possible to find at most r linear combinations of the components of $\hat{\boldsymbol{\theta}}$, which are BLUE estimators.
- There are infinite possible choices of $\boldsymbol{\lambda}_i^\dagger$.

Table of Contents

Estimable linear functions

Definition of estimable linear function

A linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ of the unknown parameter $\boldsymbol{\theta}$ is estimable if, given observations $\mathbf{b}$ s.t. $\mathrm{E}[\mathbf{b}] = \mathrm{E}[H\boldsymbol{\theta} + \boldsymbol{\varepsilon}] = H\boldsymbol{\theta}$, there exists an unbiased linear estimator $\mathbf{a}^\dagger \mathbf{b}$ for some $\mathbf{a}$, s.t. $\mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$.

Estimable linear functions

Definition of estimable linear function

A linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ of the unknown parameter $\boldsymbol{\theta}$ is estimable if, given observations $\mathbf{b}$ s.t. $\mathrm{E}[\mathbf{b}] = \mathrm{E}[H\boldsymbol{\theta} + \boldsymbol{\varepsilon}] = H\boldsymbol{\theta}$, there exists an unbiased linear estimator $\mathbf{a}^\dagger \mathbf{b}$ for some $\mathbf{a}$, s.t. $\mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$.

Lemma on the estimability of linear functions

A linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable iff $\boldsymbol{\lambda}^\dagger \in R(H)$, i.e. iff $\exists \mathbf{a}$ s.t. $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H$.

Estimable linear functions

Definition of estimable linear function

A linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$ of the unknown parameter $\boldsymbol{\theta}$ is estimable if, given observations $\mathbf{b}$ s.t. $\mathrm{E}[\mathbf{b}] = \mathrm{E}[H\boldsymbol{\theta} + \boldsymbol{\varepsilon}] = H\boldsymbol{\theta}$, there exists an unbiased linear estimator $\mathbf{a}^{\dagger}\mathbf{b}$ for some $\mathbf{a}$, s.t. $\mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$.

Lemma on the estimability of linear functions

A linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$ is estimable iff $\boldsymbol{\lambda}^{\dagger} \in R(H)$, i.e. iff $\exists\mathbf{a}$ s.t. $\boldsymbol{\lambda}^{\dagger} = \mathbf{a}^{\dagger}H$.

Proof.

- If $\boldsymbol{\lambda}^{\dagger} = \mathbf{a}^{\dagger}H$, then:

$$\mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \mathbf{a}^{\dagger}\mathrm{E}[\mathbf{b}] = \mathbf{a}^{\dagger}H\boldsymbol{\theta} = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$$

Estimable linear functions

Definition of estimable linear function

A linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$ of the unknown parameter $\boldsymbol{\theta}$ is estimable if, given observations $\mathbf{b}$ s.t. $\mathrm{E}[\mathbf{b}] = \mathrm{E}[H\boldsymbol{\theta} + \boldsymbol{\varepsilon}] = H\boldsymbol{\theta}$, there exists an unbiased linear estimator $\mathbf{a}^{\dagger}\mathbf{b}$ for some $\mathbf{a}$, s.t. $\mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$.

Lemma on the estimability of linear functions

A linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$ is estimable iff $\boldsymbol{\lambda}^{\dagger} \in R(H)$, i.e. iff $\exists\mathbf{a}$ s.t. $\boldsymbol{\lambda}^{\dagger} = \mathbf{a}^{\dagger}H$.

Proof.

- If $\boldsymbol{\lambda}^{\dagger} = \mathbf{a}^{\dagger}H$, then:

$$\mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \mathbf{a}^{\dagger}\mathrm{E}[\mathbf{b}] = \mathbf{a}^{\dagger}H\boldsymbol{\theta} = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$$

- If $\mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$, then:

$$\mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \mathbf{a}^{\dagger}H\boldsymbol{\theta} = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta},\ \forall\boldsymbol{\theta} \Rightarrow \mathbf{a}^{\dagger}H = \boldsymbol{\lambda}^{\dagger} \quad \square$$

Estimable linear functions

Lemma: uniqueness of the unbiased estimator

If a linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$ is estimable, there exists a unique unbiased estimator $\mathbf{a}_{||}^{\dagger}\mathbf{b}$, s.t. $\mathbf{a}_{||} \in C(H)$, and $\mathrm{E}\left[\mathbf{a}_{||}^{\dagger}\mathbf{b}\right] = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$

Estimable linear functions

Lemma: uniqueness of the unbiased estimator

If a linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, there exists a unique unbiased estimator $\mathbf{a}_{||}^\dagger \mathbf{b}$, s.t. $\mathbf{a}_{||} \in C(H)$, and $\mathrm{E}\left[\mathbf{a}_{||}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$

Proof.

Existence:

- Since $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, $\exists \mathbf{a} \in \mathbb{C}^N$, s.t. $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H$, and $\mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$.

Estimable linear functions

Lemma: uniqueness of the unbiased estimator

If a linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, there exists a unique unbiased estimator $\mathbf{a}_{||}^\dagger \mathbf{b}$, s.t. $\mathbf{a}_{||} \in C(H)$, and $\mathrm{E}\left[\mathbf{a}_{||}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$

Proof.

Existence:

- Since $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, $\exists \mathbf{a} \in \mathbb{C}^N$, s.t. $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H$, and $\mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$.
- $\mathbf{a} = P_{H||}\mathbf{a} + P_{H\perp}\mathbf{a} = \mathbf{a}_{||} + \mathbf{a}_{\perp}$, where $\mathbf{a}_{||} \in C(H)$ and $\mathbf{a}_{\perp} \in C_{\perp}(H)$.

Estimable linear functions

Lemma: uniqueness of the unbiased estimator

If a linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, there exists a unique unbiased estimator $\mathbf{a}_{||}^\dagger \mathbf{b}$, s.t. $\mathbf{a}_{||} \in C(H)$, and $\mathrm{E}\left[\mathbf{a}_{||}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$

Proof.

Existence:

- Since $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, $\exists \mathbf{a} \in \mathbb{C}^N$, s.t. $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H$, and $\mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$.
- $\mathbf{a} = P_{H||}\mathbf{a} + P_{H\perp}\mathbf{a} = \mathbf{a}_{||} + \mathbf{a}_\perp$, where $\mathbf{a}_{||} \in C(H)$ and $\mathbf{a}_\perp \in C_\perp(H)$.
- $\mathrm{E}\left[\mathbf{a}_\perp^\dagger \mathbf{b}\right] = \mathbf{a}_\perp^\dagger H\boldsymbol{\theta} = \mathbf{a}^\dagger P_{H\perp}^\dagger H = \mathbf{a}^\dagger P_{H\perp} H = 0$ ($\mathbf{a}_\perp$ is orthogonal to the columns of H).

Estimable linear functions

Lemma: uniqueness of the unbiased estimator

If a linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, there exists a unique unbiased estimator $\mathbf{a}_{||}^\dagger \mathbf{b}$, s.t. $\mathbf{a}_{||} \in C(H)$, and $\mathrm{E}\left[\mathbf{a}_{||}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$

Proof.

Existence:

- Since $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, $\exists \mathbf{a} \in \mathbb{C}^N$, s.t. $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H$, and $\mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$.
- $\mathbf{a} = P_{H||}\mathbf{a} + P_{H\perp}\mathbf{a} = \mathbf{a}_{||} + \mathbf{a}_\perp$, where $\mathbf{a}_{||} \in C(H)$ and $\mathbf{a}_\perp \in C_\perp(H)$.
- $\mathrm{E}\left[\mathbf{a}_\perp^\dagger \mathbf{b}\right] = \mathbf{a}_\perp^\dagger H \boldsymbol{\theta} = \mathbf{a}^\dagger P_{H\perp}^\dagger H = \mathbf{a}^\dagger P_{H\perp} H = 0$ ($\mathbf{a}_\perp$ is orthogonal to the columns of H).
- $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta} = \mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \mathrm{E}\left[\mathbf{a}_{||}^\dagger \mathbf{b}\right] + \mathrm{E}\left[\mathbf{a}_\perp^\dagger \mathbf{b}\right] = \mathrm{E}\left[\mathbf{a}_{||}^\dagger \mathbf{b}\right]$.

Estimable linear functions

Lemma: uniqueness of the unbiased estimator

If a linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, there exists a unique unbiased estimator $\mathbf{a}_{||}^\dagger \mathbf{b}$, s.t. $\mathbf{a}_{||} \in C(H)$, and $\mathrm{E}\left[\mathbf{a}_{||}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$

Proof.

Existence:

- Since $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, $\exists \mathbf{a} \in \mathbb{C}^N$, s.t. $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H$, and $\mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$.
- $\mathbf{a} = P_{H||}\mathbf{a} + P_{H\perp}\mathbf{a} = \mathbf{a}_{||} + \mathbf{a}_\perp$, where $\mathbf{a}_{||} \in C(H)$ and $\mathbf{a}_\perp \in C_\perp(H)$.
- $\mathrm{E}\left[\mathbf{a}_\perp^\dagger \mathbf{b}\right] = \mathbf{a}_\perp^\dagger H \boldsymbol{\theta} = \mathbf{a}^\dagger P_{H\perp}^\dagger H = \mathbf{a}^\dagger P_{H\perp} H = 0$ ($\mathbf{a}_\perp$ is orthogonal to the columns of H).
- $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta} = \mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \mathrm{E}\left[\mathbf{a}_{||}^\dagger \mathbf{b}\right] + \mathrm{E}\left[\mathbf{a}_\perp^\dagger \mathbf{b}\right] = \mathrm{E}\left[\mathbf{a}_{||}^\dagger \mathbf{b}\right]$.

Uniqueness:

- If $\exists \mathbf{c}_{||} \in C(H)$, s.t. $\mathrm{E}\left[\mathbf{c}_{||}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$, then $0 = \mathrm{E}\left[\mathbf{a}_{||}^\dagger \mathbf{b}\right] - \mathrm{E}\left[\mathbf{c}_{||}^\dagger \mathbf{b}\right] = \left(\mathbf{a}_{||} - \mathbf{c}_{||}\right)^\dagger H \boldsymbol{\theta}$, $\forall \boldsymbol{\theta}$.

Estimable linear functions

Lemma: uniqueness of the unbiased estimator

If a linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, there exists a unique unbiased estimator $\mathbf{a}_{\|}^\dagger \mathbf{b}$, s.t. $\mathbf{a}_{\|} \in C(H)$, and $\mathrm{E}\left[\mathbf{a}_{\|}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$

Proof.

Existence:

- Since $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, $\exists \mathbf{a} \in \mathbb{C}^N$, s.t. $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H$, and $\mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$.
- $\mathbf{a} = P_{H\|}\mathbf{a} + P_{H\perp}\mathbf{a} = \mathbf{a}_{\|} + \mathbf{a}_{\perp}$, where $\mathbf{a}_{\|} \in C(H)$ and $\mathbf{a}_{\perp} \in C_{\perp}(H)$.
- $\mathrm{E}\left[\mathbf{a}_{\perp}^\dagger \mathbf{b}\right] = \mathbf{a}_{\perp}^\dagger H\boldsymbol{\theta} = \mathbf{a}^\dagger P_{H\perp}^\dagger H = \mathbf{a}^\dagger P_{H\perp} H = 0$ ($\mathbf{a}_{\perp}$ is orthogonal to the columns of H).
- $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta} = \mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \mathrm{E}\left[\mathbf{a}_{\|}^\dagger \mathbf{b}\right] + \mathrm{E}\left[\mathbf{a}_{\perp}^\dagger \mathbf{b}\right] = \mathrm{E}\left[\mathbf{a}_{\|}^\dagger \mathbf{b}\right]$.

Uniqueness:

- If $\exists \mathbf{c}_{\|} \in C(H)$, s.t. $\mathrm{E}\left[\mathbf{c}_{\|}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$, then $0 = \mathrm{E}\left[\mathbf{a}_{\|}^\dagger \mathbf{b}\right] - \mathrm{E}\left[\mathbf{c}_{\|}^\dagger \mathbf{b}\right] = \left(\mathbf{a}_{\|} - \mathbf{c}_{\|}\right)^\dagger H\boldsymbol{\theta}$, $\forall \boldsymbol{\theta}$.
- $\left(\mathbf{a}_{\|} - \mathbf{c}_{\|}\right)^\dagger H = 0 \Rightarrow \left(\mathbf{a}_{\|} - \mathbf{c}_{\|}\right) \in C_{\perp}(H)$; but, by assumption: $\left(\mathbf{a}_{\|} - \mathbf{c}_{\|}\right) \in C(H)$

Estimable linear functions

Lemma: uniqueness of the unbiased estimator

If a linear function $\lambda(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, there exists a unique unbiased estimator $\mathbf{a}_{\|}^\dagger \mathbf{b}$, s.t. $\mathbf{a}_{\|} \in C(H)$, and $\mathrm{E}\left[\mathbf{a}_{\|}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$

Proof.

Existence:

- Since $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$ is estimable, $\exists \mathbf{a} \in \mathbb{C}^N$, s.t. $\boldsymbol{\lambda}^\dagger = \mathbf{a}^\dagger H$, and $\mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$.
- $\mathbf{a} = P_{H\|}\mathbf{a} + P_{H\perp}\mathbf{a} = \mathbf{a}_{\|} + \mathbf{a}_{\perp}$, where $\mathbf{a}_{\|} \in C(H)$ and $\mathbf{a}_{\perp} \in C_{\perp}(H)$.
- $\mathrm{E}\left[\mathbf{a}_{\perp}^\dagger \mathbf{b}\right] = \mathbf{a}_{\perp}^\dagger H \boldsymbol{\theta} = \mathbf{a}^\dagger P_{H\perp}^\dagger H = \mathbf{a}^\dagger P_{H\perp} H = 0$ ($\mathbf{a}_{\perp}$ is orthogonal to the columns of H).
- $\boldsymbol{\lambda}^\dagger \boldsymbol{\theta} = \mathrm{E}\left[\mathbf{a}^\dagger \mathbf{b}\right] = \mathrm{E}\left[\mathbf{a}_{\|}^\dagger \mathbf{b}\right] + \mathrm{E}\left[\mathbf{a}_{\perp}^\dagger \mathbf{b}\right] = \mathrm{E}\left[\mathbf{a}_{\|}^\dagger \mathbf{b}\right]$.

Uniqueness:

- If $\exists \mathbf{c}_{\|} \in C(H)$, s.t. $\mathrm{E}\left[\mathbf{c}_{\|}^\dagger \mathbf{b}\right] = \boldsymbol{\lambda}^\dagger \boldsymbol{\theta}$, then $0 = \mathrm{E}\left[\mathbf{a}_{\|}^\dagger \mathbf{b}\right] - \mathrm{E}\left[\mathbf{c}_{\|}^\dagger \mathbf{b}\right] = \left(\mathbf{a}_{\|} - \mathbf{c}_{\|}\right)^\dagger H \boldsymbol{\theta}$, $\forall \boldsymbol{\theta}$.
- $\left(\mathbf{a}_{\|} - \mathbf{c}_{\|}\right)^\dagger H = 0 \Rightarrow \left(\mathbf{a}_{\|} - \mathbf{c}_{\|}\right) \in C_{\perp}(H)$; but, by assumption: $\left(\mathbf{a}_{\|} - \mathbf{c}_{\|}\right) \in C(H)$
- The only vector that is in both $C(H)$ and $C_{\perp}(H)$ is $\left(\mathbf{a}_{\|} - \mathbf{c}_{\|}\right) = 0$, then $\mathbf{a}_{\|} = \mathbf{c}_{\|}$. □

Estimable linear functions

Lemma: estimator of minimum variance

- The unique unbiased estimator $\mathbf{a}_{\|}^{\dagger}\mathbf{b}$ has minimum variance, i.e., for any other unbiased estimator $\mathbf{a}^{\dagger}\mathbf{b}$ s.t. $\mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$, then $\mathrm{var}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] \geq \mathrm{var}\left[\mathbf{a}_{\|}^{\dagger}\mathbf{b}\right]$.
- The unique unbiased estimator $\mathbf{a}_{\|}^{\dagger}\mathbf{b}$ is BLUE, i.e. $\mathrm{E}\left[\left|\mathbf{a}^{\dagger}\mathbf{b} - \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}\right|^{2}\right] \geq \mathrm{E}\left[\left|\mathbf{a}_{\|}^{\dagger}\mathbf{b} - \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}\right|^{2}\right]$.

Estimable linear functions

Lemma: estimator of minimum variance

- The unique unbiased estimator $\mathbf{a}_{\|}^{\dagger}\mathbf{b}$ has minimum variance, i.e., for any other unbiased estimator $\mathbf{a}^{\dagger}\mathbf{b}$ s.t. $\mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$, then $\mathrm{var}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] \geq \mathrm{var}\left[\mathbf{a}_{\|}^{\dagger}\mathbf{b}\right]$.
- The unique unbiased estimator $\mathbf{a}_{\|}^{\dagger}\mathbf{b}$ is BLUE, i.e. $\mathrm{E}\left[\left|\mathbf{a}^{\dagger}\mathbf{b} - \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}\right|^2\right] \geq \mathrm{E}\left[\left|\mathbf{a}_{\|}^{\dagger}\mathbf{b} - \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}\right|^2\right]$.

Proof.

- Each vector $\mathbf{a}$ defining an unbiased estimator for $\boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$ can be written as $\mathbf{a} = \mathbf{a}_{\|} + \mathbf{a}_{\perp}$, where $\mathbf{a}_{\|}$ is unique by the previous lemma.

Estimable linear functions

Lemma: estimator of minimum variance

- The unique unbiased estimator $\mathbf{a}_{||}^{\dagger}\mathbf{b}$ has minimum variance, i.e., for any other unbiased estimator $\mathbf{a}^{\dagger}\mathbf{b}$ s.t. $\mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$, then $\mathrm{var}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] \geq \mathrm{var}\left[\mathbf{a}_{||}^{\dagger}\mathbf{b}\right]$.
- The unique unbiased estimator $\mathbf{a}_{||}^{\dagger}\mathbf{b}$ is BLUE, i.e. $\mathrm{E}\left[\left|\mathbf{a}^{\dagger}\mathbf{b} - \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}\right|^{2}\right] \geq \mathrm{E}\left[\left|\mathbf{a}_{||}^{\dagger}\mathbf{b} - \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}\right|^{2}\right]$.

Proof.

- Each vector $\mathbf{a}$ defining an unbiased estimator for $\boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$ can be written as $\mathbf{a} = \mathbf{a}_{||} + \mathbf{a}_{\perp}$, where $\mathbf{a}_{||}$ is unique by the previous lemma.
- $$\mathrm{var}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \mathbf{a}^{\dagger}\mathrm{cov}\,[\mathbf{b}]\,\mathbf{a} = \sigma^{2}\left\|\mathbf{a}\right\|^{2} = \sigma^{2}\left(\left\|\mathbf{a}_{||}\right\|^{2} + \left\|\mathbf{a}_{\perp}\right\|^{2}\right) =$$
$$= \mathrm{var}\left[\mathbf{a}_{||}^{\dagger}\mathbf{b}\right] + \sigma^{2}\left\|\mathbf{a}_{\perp}\right\|^{2} \geq \mathrm{var}\left[\mathbf{a}_{||}^{\dagger}\mathbf{b}\right].$$

Estimable linear functions

Lemma: estimator of minimum variance

- The unique unbiased estimator $\mathbf{a}_{\|}^{\dagger}\mathbf{b}$ has minimum variance, i.e., for any other unbiased estimator $\mathbf{a}^{\dagger}\mathbf{b}$ s.t. $\mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$, then $\mathrm{var}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] \geq \mathrm{var}\left[\mathbf{a}_{\|}^{\dagger}\mathbf{b}\right]$.
- The unique unbiased estimator $\mathbf{a}_{\|}^{\dagger}\mathbf{b}$ is BLUE, i.e. $\mathrm{E}\left[\left|\mathbf{a}^{\dagger}\mathbf{b} - \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}\right|^2\right] \geq \mathrm{E}\left[\left|\mathbf{a}_{\|}^{\dagger}\mathbf{b} - \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}\right|^2\right]$.

Proof.

- Each vector $\mathbf{a}$ defining an unbiased estimator for $\boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$ can be written as $\mathbf{a} = \mathbf{a}_{\|} + \mathbf{a}_{\perp}$, where $\mathbf{a}_{\|}$ is unique by the previous lemma.
- $$\mathrm{var}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \mathbf{a}^{\dagger}\mathrm{cov}\left[\mathbf{b}\right]\mathbf{a} = \sigma^2 \|\mathbf{a}\|^2 = \sigma^2\left(\|\mathbf{a}_{\|}\|^2 + \|\mathbf{a}_{\perp}\|^2\right) =$$
$$= \mathrm{var}\left[\mathbf{a}_{\|}^{\dagger}\mathbf{b}\right] + \sigma^2\|\mathbf{a}_{\perp}\|^2 \geq \mathrm{var}\left[\mathbf{a}_{\|}^{\dagger}\mathbf{b}\right].$$
- $\mathrm{var}\left[\mathbf{a}^{\dagger}\mathbf{b}\right] = \mathrm{E}\left[\left|\mathbf{a}^{\dagger}\mathbf{b} - \mathrm{E}\left[\mathbf{a}^{\dagger}\mathbf{b}\right]\right|^2\right] = \mathrm{E}\left[\left|\mathbf{a}^{\dagger}\mathbf{b} - \boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}\right|^2\right]$, and BLUEness follows from the first part of the lemma. □

Estimable linear functions

Lemma: definition of the unbiased estimator

- The unique unbiased estimator $\mathbf{a}_{||}^{\dagger}\mathbf{b}$ for $\boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$, where $\boldsymbol{\lambda}^{\dagger} = \mathbf{a}_{||}^{\dagger}H \in R(H)$ is defined as $\mathbf{a}_{||}^{\dagger}\mathbf{b} = \boldsymbol{\lambda}^{\dagger}\hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}}$ is the general LS estimator $\hat{\boldsymbol{\theta}} = H^{+}\mathbf{b} = V\boldsymbol{\Sigma}^{+}U^{\dagger}\mathbf{b}$.

Estimable linear functions

Lemma: definition of the unbiased estimator

- The unique unbiased estimator $\mathbf{a}_{||}^{\dagger}\mathbf{b}$ for $\boldsymbol{\lambda}^{\dagger}\boldsymbol{\theta}$, where $\boldsymbol{\lambda}^{\dagger} = \mathbf{a}_{||}^{\dagger}H \in R(H)$ is defined as $\mathbf{a}_{||}^{\dagger}\mathbf{b} = \boldsymbol{\lambda}^{\dagger}\hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}}$ is the general LS estimator $\hat{\boldsymbol{\theta}} = H^{+}\mathbf{b} = V\boldsymbol{\Sigma}^{+}U^{\dagger}\mathbf{b}$.

Proof.

- Since $\mathbf{a}_{||} \in C(H)$ and $\mathbf{a}_{||} = P_{H||}\mathbf{a}_{||}$:

$$\mathbf{a}_{||}^{\dagger}\mathbf{b} = \mathbf{a}_{||}^{\dagger}P_{H||}^{\dagger}\mathbf{b} = \mathbf{a}_{||}^{\dagger}P_{H||}\mathbf{b} = \mathbf{a}_{||}^{\dagger}H\hat{\boldsymbol{\theta}} = \boldsymbol{\lambda}^{\dagger}\hat{\boldsymbol{\theta}}. \square$$

Table of Contents

Generalized Gauss-Markov theorem

We can now easily prove the:

Generalized Gauss-Markov theorem

- Given any system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ with N equations and p unknown parameters, s.t. $\mathrm{E}\left[\boldsymbol{\varepsilon}\right] = 0$ and $\mathrm{cov}\left[\boldsymbol{\varepsilon}\right] = \sigma^2 I_N$.
- Be $r = \mathrm{rank} H \leq \min\left(N, p\right)$ and $\hat{\boldsymbol{\theta}} = H^+\mathbf{b} = V\boldsymbol{\Sigma}^+U^\dagger\mathbf{b}$ the generalized LS estimator.
- Be $\boldsymbol{\lambda}_i^\dagger, \;\; i = 1\cdots r$, any set of linearly independent vectors $\in R\left(H\right)$.
- Then, $\boldsymbol{\lambda}_i^\dagger\hat{\boldsymbol{\theta}}$ are unbiased minimum variance estimators of $\boldsymbol{\lambda}_i^\dagger\boldsymbol{\theta}$ and are BLUE.

Generalized Gauss-Markov theorem

We can now easily prove the:

Generalized Gauss-Markov theorem

- Given any system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ with N equations and p unknown parameters, s.t. $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$ and $\mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I_N$.
- Be $r = \mathrm{rank}H \leq \min(N, p)$ and $\hat{\boldsymbol{\theta}} = H^{+}\mathbf{b} = V\boldsymbol{\Sigma}^{+}U^{\dagger}\mathbf{b}$ the generalized LS estimator.
- Be $\boldsymbol{\lambda}_i^{\dagger}$, $i = 1 \cdots r$, any set of linearly independent vectors $\in R(H)$.
- Then, $\boldsymbol{\lambda}_i^{\dagger}\hat{\boldsymbol{\theta}}$ are unbiased minimum variance estimators of $\boldsymbol{\lambda}_i^{\dagger}\boldsymbol{\theta}$ and are BLUE.

Proof.

- Since $\dim R(H) = \mathrm{rank}H = r$, it is possible to arbitrarily choose at most r linearly independent vectors $\boldsymbol{\lambda}_i^{\dagger} \in R(H)$, $i = 1 \cdots r$.

Generalized Gauss-Markov theorem

We can now easily prove the:

Generalized Gauss-Markov theorem

- Given any system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ with N equations and p unknown parameters, s.t. $\mathrm{E}[\boldsymbol{\varepsilon}] = 0$ and $\mathrm{cov}[\boldsymbol{\varepsilon}] = \sigma^2 I_N$.
- Be $r = \mathrm{rank} H \leq \min(N, p)$ and $\hat{\boldsymbol{\theta}} = H^+\mathbf{b} = V\boldsymbol{\Sigma}^+U^\dagger\mathbf{b}$ the generalized LS estimator.
- Be $\boldsymbol{\lambda}_i^\dagger$, $i = 1 \cdots r$, any set of linearly independent vectors $\in R(H)$.
- Then, $\boldsymbol{\lambda}_i^\dagger\hat{\boldsymbol{\theta}}$ are unbiased minimum variance estimators of $\boldsymbol{\lambda}_i^\dagger\boldsymbol{\theta}$ and are BLUE.

Proof.

- Since $\dim R(H) = \mathrm{rank} H = r$, it is possible to arbitrarily choose at most r linearly independent vectors $\boldsymbol{\lambda}_i^\dagger \in R(H)$, $i = 1 \cdots r$.
- For each $\boldsymbol{\lambda}_i^\dagger$, the estimable linear function $\lambda_i(\boldsymbol{\theta}) \equiv \boldsymbol{\lambda}_i^\dagger\boldsymbol{\theta}$ can be defined.

Generalized Gauss-Markov theorem

We can now easily prove the:

Generalized Gauss-Markov theorem

- Given any system $\mathbf{b} = H\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ with N equations and p unknown parameters, s.t. $\mathrm{E}\left[\boldsymbol{\varepsilon}\right] = 0$ and $\mathrm{cov}\left[\boldsymbol{\varepsilon}\right] = \sigma^2 I_N$.
- Be $r = \mathrm{rank} H \leq \min\left(N, p\right)$ and $\hat{\boldsymbol{\theta}} = H^+\mathbf{b} = V\boldsymbol{\Sigma}^+ U^\dagger\mathbf{b}$ the generalized LS estimator.
- Be $\boldsymbol{\lambda}_i^\dagger$, $i = 1 \cdots r$, any set of linearly independent vectors $\in R\left(H\right)$.
- Then, $\boldsymbol{\lambda}_i^\dagger\hat{\boldsymbol{\theta}}$ are unbiased minimum variance estimators of $\boldsymbol{\lambda}_i^\dagger\boldsymbol{\theta}$ and are BLUE.

Proof.

- Since $\dim R\left(H\right) = \mathrm{rank} H = r$, it is possible to arbitrarily choose at most r linearly independent vectors $\boldsymbol{\lambda}_i^\dagger \in R\left(H\right)$, $i = 1 \cdots r$.
- For each $\boldsymbol{\lambda}_i^\dagger$, the estimable linear function $\lambda_i\left(\boldsymbol{\theta}\right) \equiv \boldsymbol{\lambda}_i^\dagger\boldsymbol{\theta}$ can be defined.
- By all the previous lemmas, $\boldsymbol{\lambda}_i^\dagger\hat{\boldsymbol{\theta}}$ is the unbiased, minimum variance, and BLUE estimator of $\boldsymbol{\lambda}_i^\dagger\boldsymbol{\theta}$. $\square$

Covariance of the generalized Gauss-Markov estimator

- Let us define $\Lambda = [\boldsymbol{\lambda}_1 \ \cdots \ \boldsymbol{\lambda}_r]$. Hence, the generalized Gauss-Markov estimators can be collected in the single expression $\Lambda^\dagger \hat{\boldsymbol{\theta}}$.

Generalized Gauss-Markov theorem

Covariance of the generalized Gauss-Markov estimator

- Let us define $\Lambda = [\boldsymbol{\lambda}_1 \ \cdots \ \boldsymbol{\lambda}_r]$. Hence, the generalized Gauss-Markov estimators can be collected in the single expression $\Lambda^\dagger \hat{\boldsymbol{\theta}}$.
- In general, $\operatorname{cov}\left[\Lambda^\dagger \hat{\boldsymbol{\theta}}\right] = \Lambda^\dagger \operatorname{cov}\left[\hat{\boldsymbol{\theta}}\right] \Lambda = \sigma^2 \Lambda^\dagger V \Sigma^+ \Sigma^{+T} V^\dagger \Lambda$.

Generalized Gauss-Markov theorem

Covariance of the generalized Gauss-Markov estimator

- Let us define $\Lambda = [\boldsymbol{\lambda}_1 \ \cdots \ \boldsymbol{\lambda}_r]$. Hence, the generalized Gauss-Markov estimators can be collected in the single expression $\Lambda^\dagger \hat{\boldsymbol{\theta}}$.
- In general, $\mathrm{cov}\left[\Lambda^\dagger \hat{\boldsymbol{\theta}}\right] = \Lambda^\dagger \mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] \Lambda = \sigma^2 \Lambda^\dagger V \Sigma^+ \Sigma^{+T} V^\dagger \Lambda$.
- The expression of covariance depends on the arbitrary choice of Λ. Some choices of Λ yield particularly simple expressions of covariance.

Generalized Gauss-Markov theorem

Covariance of the generalized Gauss-Markov estimator

- Let us define $\Lambda = [\boldsymbol{\lambda}_1 \ \cdots \ \boldsymbol{\lambda}_r]$. Hence, the generalized Gauss-Markov estimators can be collected in the single expression $\Lambda^\dagger \hat{\boldsymbol{\theta}}$.
- In general, $\mathrm{cov}\left[\Lambda^\dagger \hat{\boldsymbol{\theta}}\right] = \Lambda^\dagger \mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] \Lambda = \sigma^2 \Lambda^\dagger V \Sigma^+ \Sigma^{+T} V^\dagger \Lambda$.
- The expression of covariance depends on the arbitrary choice of Λ. Some choices of Λ yield particularly simple expressions of covariance.
- Since $\boldsymbol{\theta} \in \mathbb{C}^p$, if $r = p$, then $\dim R\left(H\right) = p$, and it is possible to choice the standard basis $\boldsymbol{\lambda}_i = \mathbf{e}_i \Rightarrow \Lambda^\dagger \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$, whose covariance was already obtained: $\sigma^2 \left(H^\dagger H\right)^{-1}$.

Covariance of the generalized Gauss-Markov estimator

- Let us define $\Lambda = [\boldsymbol{\lambda}_1 \ \cdots \ \boldsymbol{\lambda}_r]$. Hence, the generalized Gauss-Markov estimators can be collected in the single expression $\Lambda^\dagger \hat{\boldsymbol{\theta}}$.
- In general, $\mathrm{cov}\left[\Lambda^\dagger \hat{\boldsymbol{\theta}}\right] = \Lambda^\dagger \mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] \Lambda = \sigma^2 \Lambda^\dagger V \Sigma^+ \Sigma^{+T} V^\dagger \Lambda$.
- The expression of covariance depends on the arbitrary choice of Λ. Some choices of Λ yield particularly simple expressions of covariance.
- Since $\boldsymbol{\theta} \in \mathbb{C}^p$, if $r = p$, then $\dim R(H) = p$, and it is possible to choice the standard basis $\boldsymbol{\lambda}_i = \mathbf{e}_i \Rightarrow \Lambda^\dagger \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$, whose covariance was already obtained: $\sigma^2 \left(H^\dagger H\right)^{-1}$.
- By noticing that $R(H) = R\left(H^\dagger H\right)$, it is possible to choose $\boldsymbol{\lambda}_i^\dagger = \mathbf{a}_i^\dagger H^\dagger H$, $\mathbf{a}_i \in C(H)$. Let us define $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_r] \Rightarrow \Lambda^\dagger = A^\dagger H^\dagger H$. The covariance is then

$$\mathrm{cov}\left[\Lambda^\dagger \hat{\boldsymbol{\theta}}\right] = \mathrm{cov}\left[A^\dagger H^\dagger H H^+ \mathbf{b}\right] = \sigma^2 A^\dagger H^\dagger H H^+ \left(H^\dagger H H^+\right)^\dagger A =$$
$$= \sigma^2 A^\dagger H^\dagger H H^+ H H^+ H A = \sigma^2 A^\dagger H^\dagger H A$$

Generalized Gauss-Markov theorem

Covariance of the generalized Gauss-Markov estimator

- Let us define $\Lambda = [\boldsymbol{\lambda}_1 \ \cdots \ \boldsymbol{\lambda}_r]$. Hence, the generalized Gauss-Markov estimators can be collected in the single expression $\Lambda^\dagger \hat{\boldsymbol{\theta}}$.
- In general, $\mathrm{cov}\left[\Lambda^\dagger \hat{\boldsymbol{\theta}}\right] = \Lambda^\dagger \mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] \Lambda = \sigma^2 \Lambda^\dagger V \Sigma^+ \Sigma^{+T} V^\dagger \Lambda$.
- The expression of covariance depends on the arbitrary choice of Λ. Some choices of Λ yield particularly simple expressions of covariance.
- Since $\boldsymbol{\theta} \in \mathbb{C}^p$, if $r = p$, then $\dim R(H) = p$, and it is possible to choice the standard basis $\boldsymbol{\lambda}_i = \mathbf{e}_i \Rightarrow \Lambda^\dagger \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}$, whose covariance was already obtained: $\sigma^2 \left(H^\dagger H\right)^{-1}$.
- By noticing that $R(H) = R\left(H^\dagger H\right)$, it is possible to choose $\boldsymbol{\lambda}_i^\dagger = \mathbf{a}_i^\dagger H^\dagger H$, $\mathbf{a}_i \in C(H)$. Let us define $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_r] \Rightarrow \Lambda^\dagger = A^\dagger H^\dagger H$. The covariance is then

$$\mathrm{cov}\left[\Lambda^\dagger \hat{\boldsymbol{\theta}}\right] = \mathrm{cov}\left[A^\dagger H^\dagger H H^+ \mathbf{b}\right] = \sigma^2 A^\dagger H^\dagger H H^+ \left(H^\dagger H H^+\right)^\dagger A = $$
$$= \sigma^2 A^\dagger H^\dagger H H^+ H H^+ H A = \sigma^2 A^\dagger H^\dagger H A$$

- If $\Lambda = V_r$, where V_r are the first r columns of V, the covariance is diagonal, and $V_r^\dagger \hat{\boldsymbol{\theta}}$ are the principal components of $\hat{\boldsymbol{\theta}}$:

$$\mathrm{cov}\left[\Lambda^\dagger \hat{\boldsymbol{\theta}}\right] = \sigma^2 V_r^\dagger V \Sigma^+ \Sigma^{+T} V^\dagger V_r = \sigma^2 I_{r\times p} \Sigma^+ \Sigma^{+T} I_{p\times r} = \sigma^2 \mathrm{diag}\left(1/\sigma_1^2 \ \cdots \ 1/\sigma_r^2\right)$$

Generalized Gauss-Markov theorem

Remark on V_r

- $\boldsymbol{\lambda} \in R(H)$. Are we confident that the columns $\mathbf{v}$ of $\Lambda = V_r$ are in $R(H)$?

Remark on V_r

- $\boldsymbol{\lambda} \in R(H)$. Are we confident that the columns $\mathbf{v}$ of $\Lambda = V_r$ are in $R(H)$?
- $\mathbf{v}$ are eigenvectors of $H^\dagger H$, i.e.

$$H^\dagger H\mathbf{v} = \sigma\mathbf{v} = H^\dagger (H\mathbf{v})$$

Remark on V_r

- $\boldsymbol{\lambda} \in R(H)$. Are we confident that the columns $\mathbf{v}$ of $\Lambda = V_r$ are in $R(H)$?
- $\mathbf{v}$ are eigenvectors of $H^\dagger H$, i.e.

$$H^\dagger H\mathbf{v} = \sigma\mathbf{v} = H^\dagger (H\mathbf{v})$$

- The last equality makes it clear that $\mathbf{v}$ is a linear combination of the colums of $H^\dagger$, where the coefficients of the combination are the components of the vector $H\mathbf{v}$. Hence $\mathbf{v} \in C\left(H^\dagger\right)$

Remark on V_r

- $\boldsymbol{\lambda} \in R(H)$. Are we confident that the columns $\mathbf{v}$ of $\Lambda = V_r$ are in $R(H)$?
- $\mathbf{v}$ are eigenvectors of $H^{\dagger}H$, i.e.

$$H^{\dagger}H\mathbf{v} = \sigma\mathbf{v} = H^{\dagger}(H\mathbf{v})$$

- The last equality makes it clear that $\mathbf{v}$ is a linear combination of the colums of $H^{\dagger}$, where the coefficients of the combination are the components of the vector $H\mathbf{v}$. Hence $\mathbf{v} \in C\left(H^{\dagger}\right)$
- Since $C\left(H^{\dagger}\right) \equiv R(H)$, it is proved that $\mathbf{v} \in R(H)$.

Generalized Gauss-Markov theorem

Example

- Let us consider the following system:

$$H\boldsymbol{\theta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

Generalized Gauss-Markov theorem

Example

- Let us consider the following system:

$$H\boldsymbol{\theta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

- Clearly, it is $r = \mathrm{rank} H = 2$. SVD yields the following matrices:

$$U = \begin{bmatrix} -0.5 & -0.5 & -0.5 & -0.5 \\ -0.5 & -0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & 0.5 & -0.5 \\ -0.5 & 0.5 & -0.5 & 0.5 \end{bmatrix} \quad V = \begin{bmatrix} -8.165 & 0 & -0.5774 \\ -0.4082 & -0.7071 & 0.5774 \\ -0.4082 & 0.7071 & 0.5774 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 2.4495 & 0 & 0 \\ 0 & 1.4142 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \Sigma^+ = \begin{bmatrix} 0.4082 & 0 & 0 & 0 \\ 0 & 0.7071 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Generalized Gauss-Markov theorem

Example

- Let us consider the following system:

$$H\boldsymbol{\theta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

- Let us assume $\boldsymbol{\theta} = [1\ 1\ 1]^\dagger$, hence $\mathbf{b}_0 = H\boldsymbol{\theta} = [2\ 2\ 2\ 2]^\dagger$.

Generalized Gauss-Markov theorem

Example

- Let us consider the following system:

$$H\boldsymbol{\theta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

- Let us assume $\boldsymbol{\theta} = [1\ 1\ 1]^\dagger$, hence $\mathbf{b}_0 = H\boldsymbol{\theta} = [2\ 2\ 2\ 2]^\dagger$.

- The LS estimator yields $\hat{\boldsymbol{\theta}} = V\boldsymbol{\Sigma}^+U^\dagger = \left[1.\bar{3}\ 0.\bar{6}\ 0.\bar{6}\right]^\dagger$, and $\hat{\mathbf{b}}_0 = [2\ 2\ 2\ 2]^\dagger$.

Generalized Gauss-Markov theorem

Example

- Let us consider the following system:

$$H\boldsymbol{\theta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

- Let us assume $\boldsymbol{\theta} = [1\ 1\ 1]^\dagger$, hence $\mathbf{b}_0 = H\boldsymbol{\theta} = [2\ 2\ 2\ 2]^\dagger$.

- The LS estimator yields $\hat{\boldsymbol{\theta}} = V\boldsymbol{\Sigma}^+U^\dagger = \left[1.\bar{3}\ 0.\bar{6}\ 0.\bar{6}\right]^\dagger$, and $\hat{\mathbf{b}}_0 = [2\ 2\ 2\ 2]^\dagger$.

- Thus, $\left\|\mathbf{b}_0 - \hat{\mathbf{b}}_0\right\|^2 = 0$ is effectively minimized, but even without noise, it is not possible to estimate parameters correctly, since the system is under-determined.

Generalized Gauss-Markov theorem

Example

- Let us consider the following system:

$$H\boldsymbol{\theta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

- Let us assume $\boldsymbol{\theta} = [1\ 1\ 1]^\dagger$, hence $\mathbf{b}_0 = H\boldsymbol{\theta} = [2\ 2\ 2\ 2]^\dagger$.
- The LS estimator yields $\hat{\boldsymbol{\theta}} = V\boldsymbol{\Sigma}^+ U^\dagger = \left[1.\bar{3}\ 0.\bar{6}\ 0.\bar{6}\right]^\dagger$, and $\hat{\mathbf{b}}_0 = [2\ 2\ 2\ 2]^\dagger$.
- Thus, $\left\|\mathbf{b}_0 - \hat{\mathbf{b}}_0\right\|^2 = 0$ is effectively minimized, but even without noise, it is not possible to estimate parameters correctly, since the system is under-determined.
- But if check principal components: $V_r^\dagger \boldsymbol{\theta} = [-1.633\ 0]^\dagger$ and $V_r^\dagger \hat{\boldsymbol{\theta}} = [-1.633\ 0]^\dagger$, perfectly matching.

Generalized Gauss-Markov theorem

Example

- Let us consider the following system:

$$H\boldsymbol{\theta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

- We add Gaussian noise with $\sigma = 0.1$: $\mathbf{b} = \mathbf{b}_0 + \boldsymbol{\varepsilon} = [1.9196\ 2.0697\ 2.0835\ 1.9756]^\dagger$.

Generalized Gauss-Markov theorem

Example

- Let us consider the following system:

$$H\boldsymbol{\theta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

- We add Gaussian noise with $\sigma = 0.1$: $\mathbf{b} = \mathbf{b}_0 + \boldsymbol{\varepsilon} = [1.9196\ 2.0697\ 2.0835\ 1.9756]^\dagger$.
- The LS estimator yields $\hat{\boldsymbol{\theta}} = V\Sigma^+U^\dagger = [1.3414\ 0.6532\ 0.6882]^\dagger$, with covariance:

$$\operatorname{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2 V \boldsymbol{\Sigma}^+ \boldsymbol{\Sigma}^{+T} V^\dagger = \begin{bmatrix} 0.0011 & 0.0006 & 0.0006 \\ 0.0006 & 0.0028 & -0.0022 \\ 0.0006 & -0.0022 & 0.0028 \end{bmatrix}$$

Generalized Gauss-Markov theorem

Example

- Let us consider the following system:

$$H\boldsymbol{\theta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

- We add Gaussian noise with $\sigma = 0.1$: $\mathbf{b} = \mathbf{b}_0 + \boldsymbol{\varepsilon} = [1.9196\ 2.0697\ 2.0835\ 1.9756]^\dagger$.

- The LS estimator yields $\hat{\boldsymbol{\theta}} = V\boldsymbol{\Sigma}^+ U^\dagger = [1.3414\ 0.6532\ 0.6882]^\dagger$, with covariance:

$$\mathrm{cov}\left[\hat{\boldsymbol{\theta}}\right] = \sigma^2 V\boldsymbol{\Sigma}^+\boldsymbol{\Sigma}^{+T}V^\dagger = \begin{bmatrix} 0.0011 & 0.0006 & 0.0006 \\ 0.0006 & 0.0028 & -0.0022 \\ 0.0006 & -0.0022 & 0.0028 \end{bmatrix}$$

- Principal components: $V_r^\dagger\boldsymbol{\theta} = [-1.633\ 0]^\dagger$ and $V_r^\dagger\hat{\boldsymbol{\theta}} = [-1.6429\ 0.0247]^\dagger$, with:

$$\mathrm{cov}\left[V_r^\dagger\hat{\boldsymbol{\theta}}\right] = \sigma^2 V_r^\dagger V\boldsymbol{\Sigma}^+\boldsymbol{\Sigma}^{+T}V^\dagger V_r = \sigma^2 \mathrm{diag}\left(1/\sigma_1^2\ \cdots\ 1/\sigma_r^2\right) = \begin{bmatrix} 0.0017 & 0 \\ 0 & 0.0050 \end{bmatrix}$$