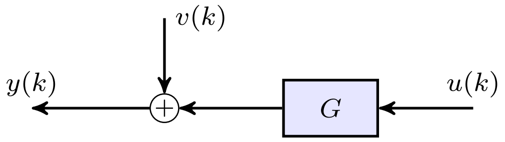
System Identification

Lecture 3: Least-squares estimation

Roy Smith

Least-Squares data-fitting

Framework



True system: $y \;=\; G_0\, u \;+\; v, \quad v \sim \mathcal{N}(0, \sigma_v^2), \quad v$ is i.i.d.

Data set: $\mathcal{Z} \;=\; \{\, (y_0, u_0, \ldots y_{N-1}, u_{N-1} \,\}\,.$

$$\underbrace{\begin{bmatrix} y_0 \\ \vdots \\ y_{N-1} \end{bmatrix}}_{y} \;=\; \underbrace{\begin{bmatrix} u_0 \\ \vdots \\ u_{N-1} \end{bmatrix}}_{\Phi} \theta \;+\; \underbrace{\begin{bmatrix} v_0 \\ \vdots \\ v_{N-1} \end{bmatrix}}_{\epsilon}$$

Least-Squares data-fitting

Objective: least-squares

"Closest fit" in the 2-norm sense:

$$\underset{\theta}{\text{minimise}} \quad \|y \,-\, \Phi\,\theta\|_2^2$$

Least-Squares

Normal equations

$$\left(\Phi^T\Phi\right)\,\theta \;=\; \Phi^T\,y$$

These are always consistent.

Least-Squares

Pseudo-inverse formulation

Notation: $X^\dagger$, Defining properties: $X = XX^\dagger X$, $X^\dagger = X^\dagger X X^\dagger$,
$XX^\dagger$ and $X^\dagger X$ are symmetric.

SVD: $U\Sigma V^T = \mathsf{svd}(X)$, $V\Sigma^\dagger U^T = \mathsf{svd}(X^\dagger)$, $\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}$, $\Sigma^\dagger = \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}$.

Least-Squares

QR formulation

Least-Squares

Geometric interpretation

Least-Squares

Geometric interpretation

Least-Squares

Linear regression problems

Least-Squares

Example

$$y \;=\; u\,G \;+\; v, \qquad y \in \mathcal{R}^N, u \in \mathcal{R}^N, v \in \mathcal{R}^N.$$

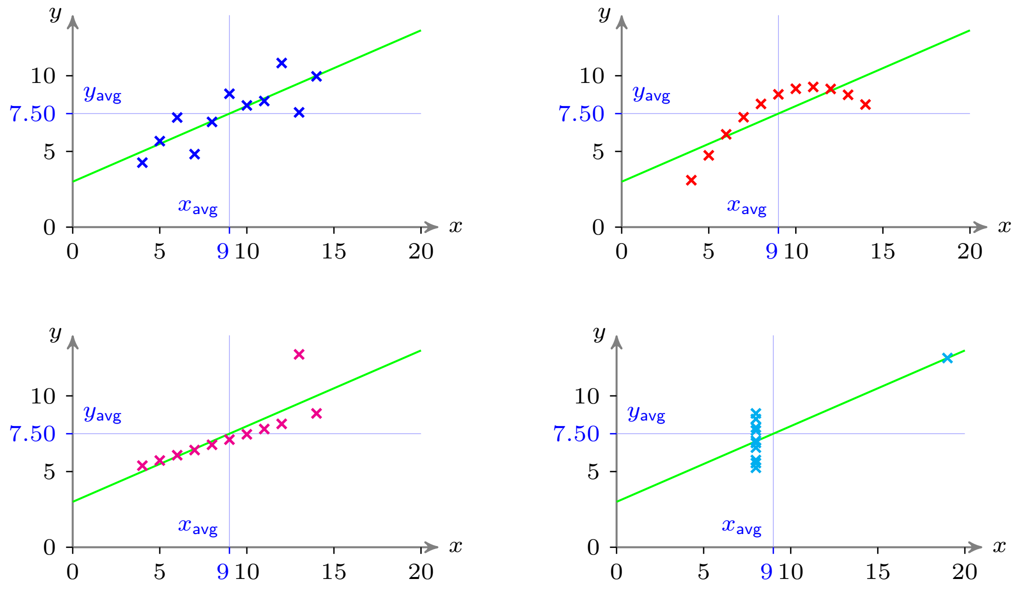
Normal equations: $(u^T u)G \;=\; u^T y,$

Estimate: $$\hat{G} \;=\; (u^T u)^{-1} u^T y \;=\; \frac{\sum_{n=0}^{N-1} u(n)y(n)}{\sum_{i=0}^{N-1} |u(i)|^2}$$

This is equivalent to:

$$\hat{G} \;=\; \sum_{n=0}^{N-1} \alpha_n \left(\frac{y(n)}{u(n)}\right), \quad \text{where} \quad \alpha_n \;=\; \frac{|u(n)|^2}{\sum_{i=0}^{N-1} \frac{1}{|u(i)|^2}}.$$

Least-Squares: Applicability



F.J. Anscombe, "Graphs in Statistical Analysis", *The American Statistician*, **27:1**, 17–21, 1973.

Example: Ceres orbit determination

1801, 1st January. Guissepe Piazzi finds and makes 22 observations of a "planet" over 44 days.

1801, February. Ceres is obscured by the sun. The observations correspond to 9 degrees of arc.

1801, June. Zach publishes the observational data. Many astronomers attempt to estimate when and where Ceres will reappear.

1801, September. Karl Fredrich Gauss estimates the orbit of Ceres and Zach publishes his empheris (along with many others). Those of Gauss differ significantly.

1801, 7th December. Zach locates Ceres within 1/2 a degree of Gauss's prediction. Gauss refuses to disclose his method and is accused by some of sorcery.

Dawn mission image of Ceres
(930 km. diameter)

Example: Ceres orbit determination

1802. Gauss also estimates the orbit of Pallas.

1805 Legendre publishes a paper on the method of least squares.

1807 Gauss becomes director of Göttingen observatory. Olbers discovers Vesta and Gauss takes only 10 hours to estimate its orbit. Ceres' orbit had taken him 100 hours.

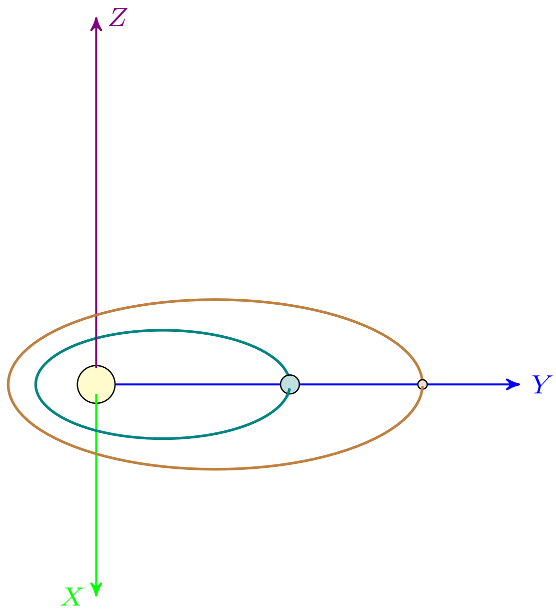
1809. Gauss finally publishes details of his method and claims priority over Legendre.

Gauss's approach: least squares

Basic assumptions: Kepler's laws of planetary motion.

- Planetary orbits are ellipses with the sun at one of the foci.
- A line drawn between the Sun and any planet sweeps out equal areas in equal time.
- The ratio of the square of the orbital periods of two planets is equal to the ratio of the cubes of their major semi-axes.

Parametrisation: Keplerian orbital elements



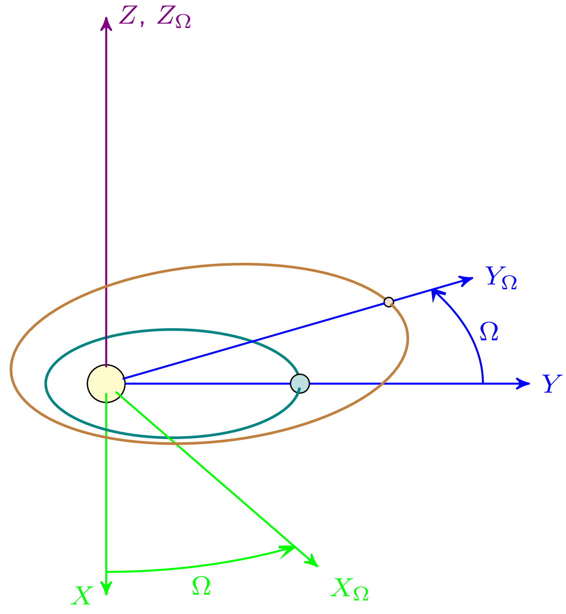
Five parameters define an orbit in space, and a sixth parameter gives the current position of the body on the orbit.

Orbital Elements:

- Orbit size and shape:
  - eccentricity (e)
  - semimajor axis (a)

[0] zero angle positions simplified in the figure

Parametrisation: Keplerian orbital elements



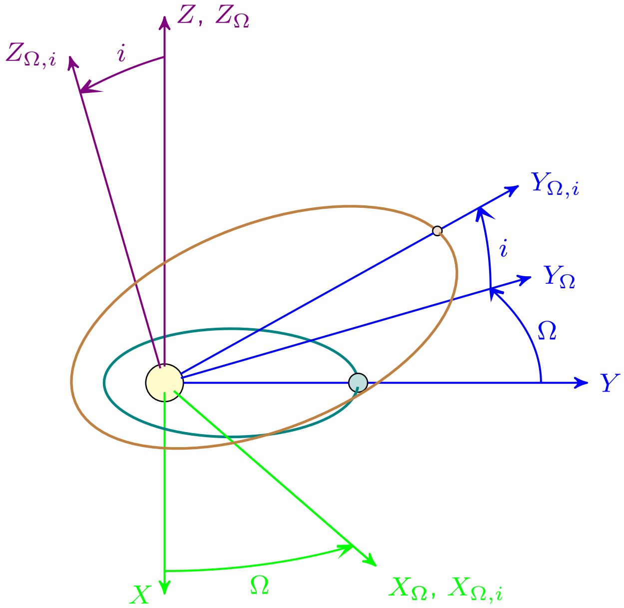
Five parameters define an orbit in space, and a sixth parameter gives the current position of the body on the orbit.

Orbital Elements:

- Orbit size and shape:
  - eccentricity (e)
  - semimajor axis (a)
- Longitude of ascending node (Ω)

[0]zero angle positions simplified in the figure

Parametrisation: Keplerian orbital elements



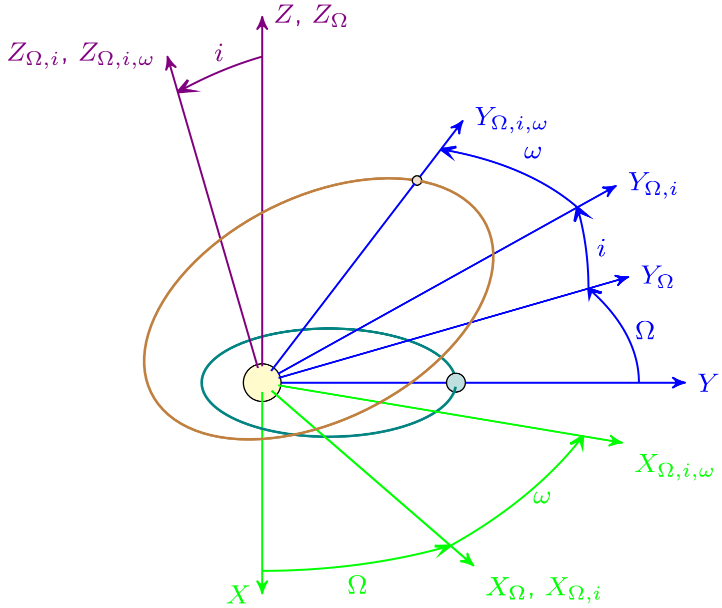
Five parameters define an orbit in space, and a sixth parameter gives the current position of the body on the orbit.

Orbital Elements:

- Orbit size and shape:
  - eccentricity (e)
  - semimajor axis (a)
- Longitude of ascending node (Ω)
- Inclination (i)

[0]zero angle positions simplified in the figure

Parametrisation: Keplerian orbital elements



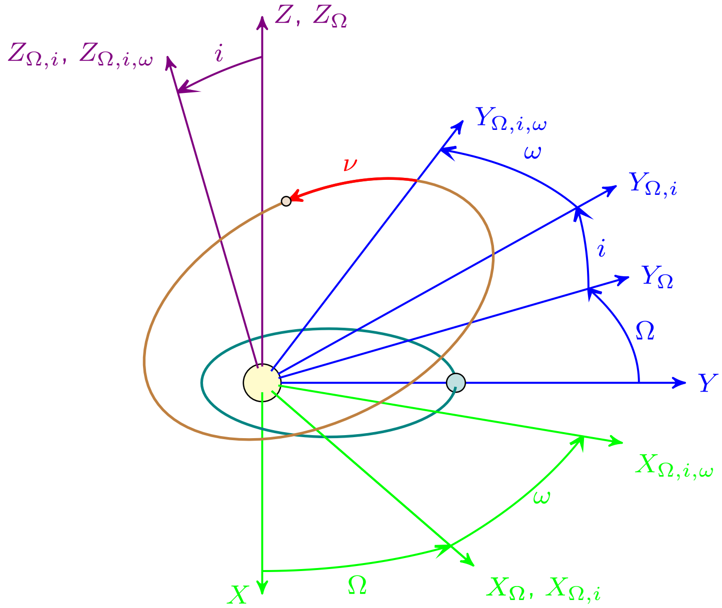
Five parameters define an orbit in space, and a sixth parameter gives the current position of the body on the orbit.

Orbital Elements:

- Orbit size and shape:
  - eccentricity (e)
  - semimajor axis (a)
- Longitude of ascending node (Ω)
- Inclination (i)
- Argument of periapsis (ω)

[0] zero angle positions simplified in the figure

Parametrisation: Keplerian orbital elements



Five parameters define an orbit in space, and a sixth parameter gives the current position of the body on the orbit.

Orbital Elements:

- Orbit size and shape:
  - eccentricity (e)
  - semimajor axis (a)
- Longitude of ascending node (Ω)
- Inclination (i)
- Argument of periapsis (ω)
- True anomaly (ν)

[0] zero angle positions simplified in the figure

Example: Ceres orbit determination

Gauss's approach: least squares

- Select three observations: 1 January, 21 January and 11 February.
- Calculate a nominal orbit matching the observations.
  - Calculate a Taylor series approximation to the nonlinear differential equations.
  - Iterative refinement of the nominal orbit.
- Adjust linearized versions of the orbit to minimise the sum of squares of the error in all 22 observations.

Problem formulation

$$y = \begin{bmatrix} \text{RA of obs. 1} \\ \text{Dec. of obs. 1} \\ \vdots \\ \text{RA of obs. 22} \\ \text{Dec. of obs. 22} \end{bmatrix} = H(\theta, t), \qquad \text{where } \theta = \text{ vector of orbital elements.}$$

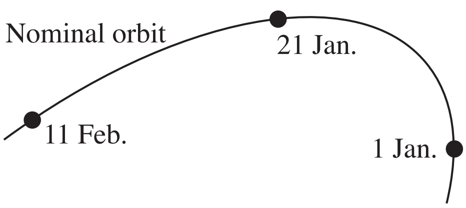
Example: Ceres orbit determination

Problem formulation

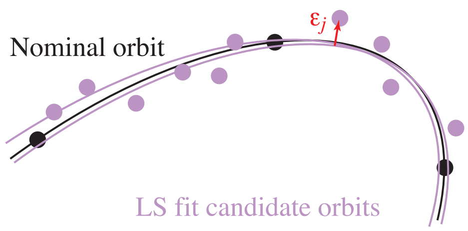
$\epsilon = y_{\text{obs}} - H(\theta, t)$, $\quad \epsilon$ is the vector of residuals.

Define a "cost function": $\quad J(\theta) = \dfrac{\epsilon^T R^{-1} \epsilon}{2}$, $\quad$ with weighting: $R > 0$.

Optimisation: $\quad \theta_* = \underset{\theta}{\operatorname{argmin}}\, J(\theta)$.



Nominal orbit determination
6 equations in 6 unknown parameters.



Estimated orbit errors
44 equations in 6 unknown parameters.

Example: Ceres orbit determination

Linearisation and iterative refinement

The orbit is specified by θ_i. Start with initial case ($i = 0$). Close to this,

$$H(\theta) \approx H(\theta_i) + H_\theta(\theta - \theta_i), \qquad \text{where } H_\theta = \left.\frac{\partial H}{\partial \theta}\right|_{\theta_i}.$$

Define $\delta\theta = \theta - \theta_i$ and $\delta y = y_{\text{obs}} - H(\theta_i)$. Then,

$$J_{\text{linear}}(\delta\theta) = \frac{(\delta y - H_\theta \delta\theta)^T R^{-1} (\delta y - H_\theta \delta\theta)^T}{2} \approx J(\theta_i + \delta\theta).$$

The gradient of $J_{\text{linear}}(\theta)$ is,

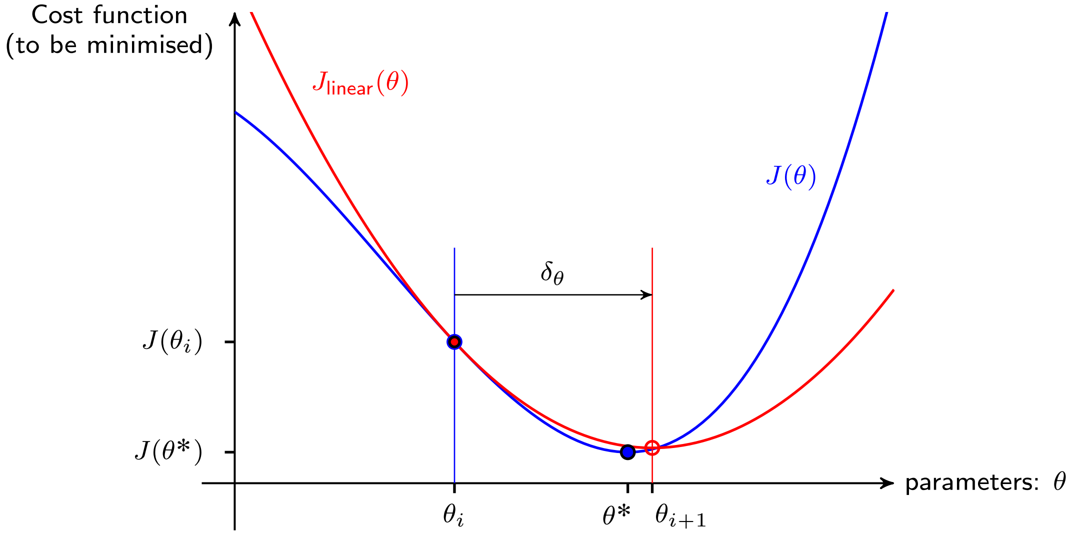
$$\frac{\partial J_{\text{linear}}(\theta)}{\partial \theta} = -H_\theta^T R^{-1}(\delta y - H_\theta \delta\theta).$$

Setting this to zero is equivalent to solving,

$$H_\theta^T R^{-1} H_\theta \delta\theta = H_\theta^T R^{-1} \delta y. \qquad \text{(linear least-squares problem)}$$

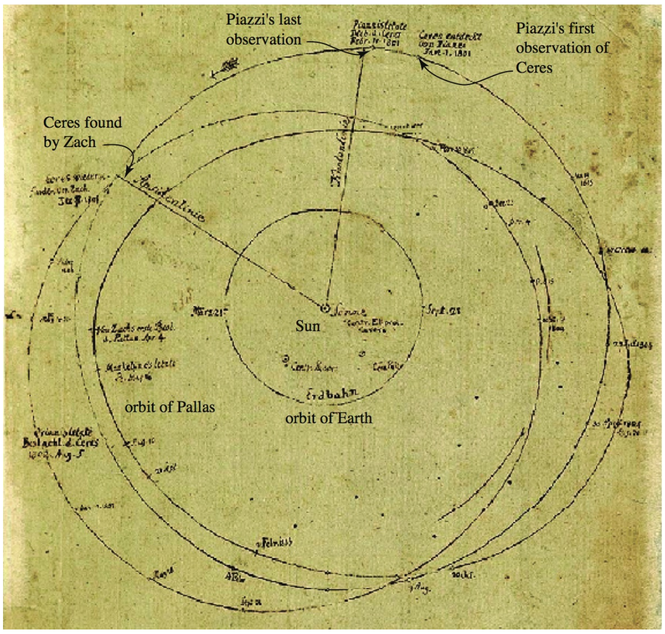
Solve for $\delta\theta$ and iteratively update: $\theta_{i+1} = \theta_i + \delta\theta$.

Linearization and iterative refinement



Example: Ceres orbit determination

Gauss's sketch



References

Skizze der Bahnen der Kleinplaneten Ceres, Pallas und Vesta. *"Astronomische Untersuchen und Rechnungen vornehmlich ber die Ceres Ferdinandea,"* 1802. SUB Göttingen: Cod. Ms. Gauß Handbuch 4, Bl. 1.

R. Bannister, *Am. J. Phys.*, **71**(12), pp. 1268–1275, Dec. 2003.

L. Weiss, "Gauss and Ceres," (`www.math.rutgers.edu/~cherlin/History/Papers1999/weiss.html`).

Regularised Least-Squares

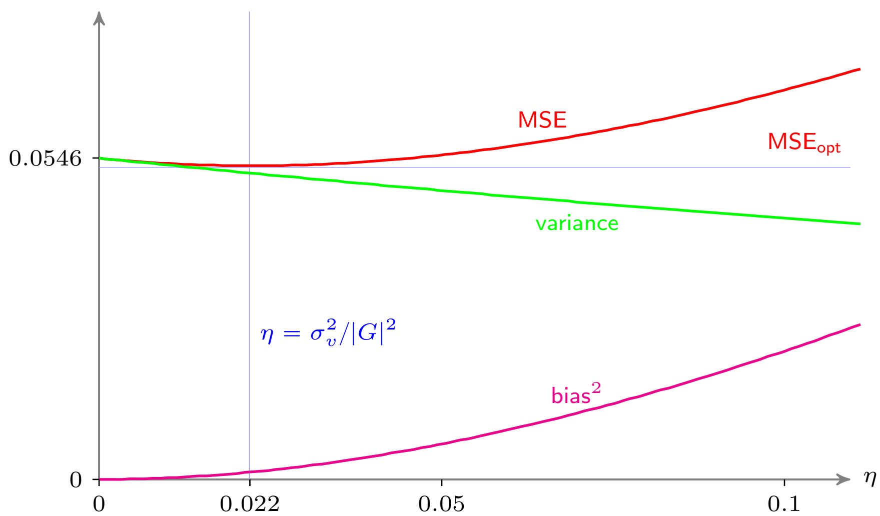
Tikhonov regularisation

Regularised Least-Squares

Example

Example

Bias-Variance trade-off



Statistical properties of the LS estimates

Model

$$Y = \Phi\theta + \epsilon, \qquad \epsilon = \begin{bmatrix} \epsilon(0) \\ \vdots \\ \epsilon(N-1) \end{bmatrix}$$

Error assumptions,

$$E\{\epsilon\} = 0, \qquad E\{\epsilon\epsilon^T\} = \sigma^2 I.$$

LS estimator properties

Assume that Φ is fixed, but consider ϵ as a stochastic variable.

$$E\{\hat{\theta}\} = \theta_0 \qquad \text{(unbiased estimator)}$$

$$\text{cov}\{\hat{\theta}\} = E\left\{(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^T\right\} = \sigma^2 \left(\Phi^T\Phi\right)^{-1}.$$

Statistical properties of the LS estimates

Bias and variance of the estimates

Least square estimation statistics

Model with correlated noise

$$Y = \Phi\theta + \epsilon, \qquad \text{with} \quad E\{\epsilon\epsilon^T\} = R.$$

LS estimator properties

Assume that Φ is fixed, but consider ϵ as a stochastic variable.

$$E\{\hat{\theta}\} = \theta_0 \qquad \text{(unbiased estimator)}$$

$$\operatorname{cov}\{\hat{\theta}\} = E\left\{(\hat{\theta}-\theta_0)(\hat{\theta}-\theta_0)^T\right\} = \left(\Phi^T\Phi\right)^{-1}\Phi^T R\Phi\left(\Phi^T\Phi\right)^{-1}$$

Note the simplification to the white noise case if $R = \sigma^2 I$.

Best linear unbiased estimator (BLUE or Markov estimate)

Model with correlated noise

$$Y = \Phi\theta + \epsilon, \qquad \text{with} \quad E\{\epsilon\epsilon^T\} = R.$$

Best linear estimator

Linear estimators: $\hat{\theta} = Z^T Y$

The linear estimator: $Z = R^{-1}\Phi\left(\Phi^T R^{-1}\Phi\right)^{-1}$

satisfies:

$$E\{\hat{\theta}_Z\} = \theta_0, \qquad \text{(unbiased)}$$

$$\operatorname{cov}\{\hat{\theta}_Z\} = \left(\Phi^T R^{-1}\Phi\right)^{-1} \leqslant \operatorname{cov}\{\hat{\theta}\} \text{ for any unbiased estimate.}$$

Note that the BLUE requires knowledge of the error covariance, R.

Bibliography

LS estimation statistics

Lennart Ljung, *System Identification;Theory for the User*, 2nd Ed., Prentice-Hall, 1999, [section 10.1 and appendix II].

F.J. Anscombe, "Graphs in Statistical Analysis", *The American Statistician*, **27:1**, 17–21, 1973.