Least-Squares Estimation
Robert Stengel
Optimal Control and Estimation, MAE 546, Princeton University, 2018

- Estimating unknown constants from redundant measurements
  - Least-squares
  - Weighted least-squares
- Recursive weighted least-squares estimator

Copyright 2018 by Robert Stengel. All rights reserved. For educational use only.
http://www.princeton.edu/~stengel/MAE546.html
http://www.princeton.edu/~stengel/OptConEst.html

Perfect Measurement of a Constant Vector

- Given
  - Measurements, \( y \), of a constant vector, \( x \)
- Estimate \( x \)

- Assume that output, \( y \), is a perfect measurement and \( H \) is invertible

\[
y = H x
\]

- Estimate is based on inverse transformation

\[
\hat{x} = H^{-1} y
\]
Imperfect Measurement of a Constant Vector

• Given
  – “Noisy” measurements, \( z \), of a constant vector, \( x \)
• Effects of error can be reduced if measurement is redundant
• Noise-free output, \( y \)
  \[
  y = H x
  \]
• Measurement of output with error, \( z \)
  \[
  z = y + n = H x + n
  \]

Cost Function for Least-Squares Estimate

Measurement-error residual
  \[
  \varepsilon = z - H \hat{x} = z - \hat{y}
  \]
  dim(\( \varepsilon \)) = (r \times 1)

Squared measurement error = cost function, \( J \)

\[
J = \frac{1}{2} \varepsilon^T \varepsilon = \frac{1}{2} (z - H \hat{x})^T (z - H \hat{x})
\]

\[
= \frac{1}{2} (z^T z - \hat{x}^T H^T z - z^T H \hat{x} + \hat{x}^T H^T H \hat{x})
\]

What is the control parameter?

The estimate of \( x \) \( \hat{x} \)
  dim(\( \hat{x} \)) = (n \times 1)
Static Minimization Provides Least-Squares Estimate

Error cost function *(scalar)*

\[
J = \frac{1}{2} \left( z^T z - \hat{x}^T H^T z - z^T H \hat{x} + \hat{x}^T H^T H \hat{x} \right)
\]

Necessary condition for minimum

\[
\frac{\partial J}{\partial \hat{x}} = 0 = \frac{1}{2} \left[ 0 - (H^T z)^T - z^T H + (H^T \hat{x})^T + \hat{x}^T H^T H \right]
\]

\[
\hat{x}^T H^T H = z^T H
\]

Sufficient condition for minimum

\[
H^T H > 0
\]

Static Minimization Provides Least-Squares Estimate

Estimate employs left pseudo-inverse matrix

\[
\hat{x}^T (H^T H)(H^T H)^{-1} = \hat{x}^T = z^T H (H^T H)^{-1} \quad (row)
\]

*or*

\[
\hat{x} = (H^T H)^{-1} H^T z \quad (column)
\]
Example: Average Weight of a Pail of Jelly Beans

- Measurements are equally uncertain

\[ z_i = x + n_i , \quad i = 1 \text{ to } r \]

- Express measurements as

\[ z = Hx + n \]

- Output matrix

\[
H = \begin{bmatrix}
1 \\
1 \\
\vdots \\
1
\end{bmatrix}
\]

- Optimal estimate of \( x \) (scalar)

\[
\hat{x} = \left(H^T H\right)^{-1} H^T z
\]
Least-Squares Applications

• More generally, least-squares estimation is used for
  – Higher-degree curve-fitting
  – Multivariate estimation

Least-Squares Linear Fit to Noisy Data

Find trend line in noisy data

\[ y = a_0 + a_1x \]
\[ z = (a_0 + a_1x) + n \]

Error cost function

\[ J = \frac{1}{2} (z - H\hat{a})^T (z - H\hat{a}) \]

Least-squares output estimate

\[ \hat{y} = \hat{a}_0 + \hat{a}_1x \]
Measurements of Differing Quality

• Suppose some elements of the measurement, \( z \), are more uncertain than others

\[
z = Hx + n
\]

• Give the more uncertain measurements less weight in arriving at the minimum-cost estimate

• Let \( S \) = measure of uncertainty; then express error cost in terms of \( S^{-1} \)

\[
J = \frac{1}{2} \varepsilon^T S^{-1} \varepsilon
\]

Error Cost and Necessary Condition for a Minimum

Error cost function, \( J \)

\[
J = \frac{1}{2} \varepsilon^T S^{-1} \varepsilon = \frac{1}{2} (z - H \hat{x})^T S^{-1} (z - H \hat{x})
\]

\[
= \frac{1}{2} (z^T S^{-1} z - \hat{x}^T H^T S^{-1} z - z^T S^{-1} H \hat{x} + \hat{x}^T H^T S^{-1} H \hat{x})
\]

Necessary condition for a minimum

\[
\frac{\partial J}{\partial \hat{x}} = 0
\]

\[
= \frac{1}{2} \left[ 0 - (H^T S^{-1} z)^T - z^T S^{-1} H + (H^T S^{-1} H \hat{x})^T + \hat{x}^T H^T S^{-1} H \right]
\]

Sufficient condition for a minimum

\[
H^T S^{-1} H > 0
\]
Weighted Least-Squares Estimate of a Constant Vector

Necessary condition for a minimum

\[ \left[ \hat{x}^T H^T S^{-1} H - z^T S^{-1} H \right] = 0 \]

\[ \hat{x}^T H^T S^{-1} H = z^T S^{-1} H \]

Weighted left pseudo-inverse provides the solution

\[ \hat{x} = \left( H^T S^{-1} H \right)^{-1} H^T S^{-1} z \]

Return of the Jelly Beans

Error-weighting matrix

\[ S^{-1} \triangleq \Lambda = \begin{bmatrix} a_{i1} & 0 & \ldots & 0 \\ 0 & a_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{rr} \end{bmatrix} \]

Optimal estimate of average jelly bean weight

\[ \hat{x} = \hat{x} = \left( H^T S^{-1} H \right)^{-1} H^T S^{-1} z \]

Weighted Estimate of \( x \) (scalar)

\[ \hat{x} = \frac{\sum_{i=1}^{r} a_{ii} z_i}{\sum_{i=1}^{r} a_{ii}} \]
Weighted Least Squares ("Kriging") Estimates
(Wiener–Kolmogorov Interpolation between measurement points)

- Curve, \(y(x)\), between measurement points, \(x_i\), is the mean of a stationary process with covariance derived from the measurements, \(y(x)\) or other known source
- Curve, \(y(x)\), is a distance-weighted linear combination of all of the points

\[
y(x) = w^T(x)y(x) = \begin{bmatrix} w_1(x) & w_2(x) & \cdots & w_k(x) \end{bmatrix} \begin{bmatrix} y(x_1) \\ y(x_2) \\ \vdots \\ y(x_k) \end{bmatrix}
\]

\(w\): Minimum-least-squares weighting function

How to Chose the Error Weighting Matrix

a) Normalize the cost function according to expected measurement error, \(S_A\)

\[
J = \frac{1}{2} \varepsilon^T S_A^{-1} \varepsilon = \frac{1}{2} (z - y)^T S_A^{-1} (z - y) = \frac{1}{2} (z - H \hat{x})^T S_A^{-1} (z - H \hat{x})
\]

b) Normalize the cost function according to expected measurement residual, \(S_B\)

\[
J = \frac{1}{2} \varepsilon^T S_B^{-1} \varepsilon = \frac{1}{2} (z - H \hat{x})^T S_B^{-1} (z - H \hat{x})
\]

\[
\text{dim}(S_A) = \text{dim}(S_B) = (r \times r)
\]
Measurement Error Covariance, $S_A$

Expected value of outer product of measurement error vector

$$S_A = E\left[ (z - y)(z - y)^T \right] = E\left[ (z - Hx)(z - Hx)^T \right] = E\left[ nn^T \right] \triangleq R$$

Measurement Residual Covariance, $S_B$

Expected value of outer product of measurement residual vector

$$S_B = E\left[ \varepsilon \varepsilon^T \right] = E\left[ (z - H\hat{x})(z - H\hat{x})^T \right] = E\left[ (H\varepsilon + n)(H\varepsilon + n)^T \right]$$

$$S_B = HE\left[ \varepsilon \varepsilon^T \right]H^T + HE\left[ \varepsilon n^T \right] + E\left[ nn^T \right]$$

$$\triangleq HPH^T + HM + M^T H^T + R$$

Requires iteration (“adaptation”) of the estimate to find $S_B$

where

$$P = E\left[ (x - \hat{x})(x - \hat{x})^T \right]$$

$$M = E\left[ (x - \hat{x})n^T \right]$$

$$R = E\left[ nn^T \right]$$
Recursive Least-Squares Estimation

- Prior unweighted and weighted least-squares estimators use “batch-processing” approach
  - All information is gathered prior to processing
  - All information is processed at once

- Recursive approach
  - Optimal estimate has been made from prior measurement set
  - New measurement set is obtained
  - Optimal estimate is improved by incremental change (or correction) to the prior optimal estimate

Prior Optimal Estimate

Initial measurement set and state estimate, with $S = S_A = R$

\[
\begin{align*}
  z_1 &= H_1 x + n_1 \\
  \hat{x}_1 &= \left( H_1^T R_1^{-1} H_1 \right)^{-1} H_1^T R_1^{-1} z_1
\end{align*}
\]

State estimate minimizes

\[
J_1 = \frac{1}{2} \varepsilon_1^T R_1^{-1} \varepsilon_1 = \frac{1}{2} (z_1 - H_1 \hat{x}_1)^T R_1^{-1} (z_1 - H_1 \hat{x}_1)
\]
New Measurement Set

New measurement

\[ z_2 = H_2 x + n_2 \]

\( R_2 \): Second measurement error covariance

Concatenation of old and new measurements

\[ z \triangleq \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \]

\[ \dim(z_2) = \dim(n_2) = r_2 \times 1 \]
\[ \dim(H_2) = r_2 \times n \]
\[ \dim(R_2) = r_2 \times r_2 \]

Cost of Estimation Based on Both Measurement Sets

Cost function incorporates estimate made after incorporating \( z_2 \)

\[ J_2 = \begin{pmatrix} (z_1 - H_1 \hat{x}_2)^T \\ (z_2 - H_2 \hat{x}_2)^T \end{pmatrix} \begin{pmatrix} R_1^{-1} & 0 \\ 0 & R_2^{-1} \end{pmatrix} \begin{pmatrix} (z_1 - H_1 \hat{x}_2) \\ (z_2 - H_2 \hat{x}_2) \end{pmatrix} \]

\[ = (z_1 - H_1 \hat{x}_2)^T R_1^{-1} (z_1 - H_1 \hat{x}_2) + (z_2 - H_2 \hat{x}_2)^T R_2^{-1} (z_2 - H_2 \hat{x}_2) \]

Both residuals refer to \( \hat{x}_2 \)
Optimal Estimate Based on Both Measurement Sets

Simplification occurs because weighting matrix is block diagonal

\[
\hat{x}_2 = \left( H_1^T R_1^{-1} H_1 + H_2^T R_2^{-1} H_2 \right)^{-1} \left( H_1^T R_1^{-1} z_1 + H_2^T R_2^{-1} z_2 \right)
\]

Inverse can be put in more useful form using the Matrix Inversion Lemma

Development of Matrix Inversion Lemma

A & B are square, non-singular matrices

\[
B \triangleq A^{-1}; \text{ then } BA = AB = I
\]

\[
A = \begin{bmatrix}
    A_1 (m \times m) & A_2 (m \times n) \\
    A_3 (n \times m) & A_4 (n \times n)
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
    B_1 (m \times m) & B_2 (m \times n) \\
    B_3 (n \times m) & B_4 (n \times n)
\end{bmatrix} \triangleq \begin{bmatrix}
    A_1 & A_2 \\
    A_3 & A_4
\end{bmatrix}^{-1}
\]
### Development of Matrix Inversion Lemma

**AB product is Identity Matrix**

\[
AB = \begin{bmatrix}
A_1 & A_2 \\
A_3 & A_4
\end{bmatrix}
\begin{bmatrix}
B_1 & B_2 \\
B_3 & B_4
\end{bmatrix}
= \begin{bmatrix}
(A_1B_1 + A_2B_3) & (A_1B_2 + A_2B_4) \\
(A_3B_1 + A_4B_3) & (A_3B_2 + A_4B_4)
\end{bmatrix}
= \begin{bmatrix}
I_m & 0 \\
0 & I_n
\end{bmatrix}
\]

**BA product is Identity Matrix**

\[
BA = \begin{bmatrix}
B_1 & B_2 \\
B_3 & B_4
\end{bmatrix}
\begin{bmatrix}
A_1 & A_2 \\
A_3 & A_4
\end{bmatrix}
= \begin{bmatrix}
(B_1A_1 + B_2A_3) & (B_1A_2 + B_2A_4) \\
(B_3A_1 + B_4A_3) & (B_3A_2 + B_4A_4)
\end{bmatrix}
= \begin{bmatrix}
I_m & 0 \\
0 & I_n
\end{bmatrix}
\]

### Submatrices of B

**From first column of AB**

\[
B_3 = -A_4^{-1}A_3B_1
\]

\[
A_1B_1 - A_2A_4^{-1}A_3B_1 = I_m
\]

\[
\left( A_1 - A_2A_4^{-1}A_3 \right)B_1 = I_m
\]

\[
B_1 = \left( A_1 - A_2A_4^{-1}A_3 \right)^{-1}
\]

\[
B_3 = -A_4^{-1}A_3 \left( A_1 - A_2A_4^{-1}A_3 \right)^{-1}
\]
Solutions for B

Similar solutions from second column of $AB$

$$B = \begin{bmatrix} (A_1 - A_2 A_4^{-1} A_3)^{-1} & -A_1^{-1} A_2 (A_4 - A_3 A_1^{-1} A_2)^{-1} \\ -A_4^{-1} A_3 (A_1 - A_2 A_4^{-1} A_3)^{-1} & (A_4 - A_3 A_1^{-1} A_2)^{-1} \end{bmatrix}$$

Alternative solutions from $BA$

$$B = \begin{bmatrix} (A_1 - A_2 A_4^{-1} A_3)^{-1} & -(A_1 - A_2 A_4^{-1} A_3)^{-1} A_2^{-1} A_4 \\ -(A_4 - A_3 A_1^{-1} A_2)^{-1} A_3^{-1} A_1 & (A_4 - A_3 A_1^{-1} A_2)^{-1} \end{bmatrix}$$

Matrix Inversion Lemma

For symmetric $A$

$A_1, A_4, B_1,$ and $B_4$ are symmetric

$$A_3 = A_2^T$$ and $$B_3 = B_2^T$$

Substitution and elimination in $BA$

$$B_4 = \left[I_n - B_3 A_2\right] A_4^{-1}$$

Matrix inversion lemma for symmetric $A$

$$\begin{bmatrix} A_4 - A_2^T A_1^{-1} A_2 \end{bmatrix} = A_4^{-1} - A_4^{-1} A_2^T \left[A_2 A_4^{-1} A_2^T - A_1\right]^{-1} A_2 A_4^{-1}$$
Apply Matrix Inversion Lemma in Optimal Estimate

First, define

\[ P_1^{-1} \triangleq H_1^T R_1^{-1} H_1 \quad \text{dim}(P_1) = (n \times n) \]

From matrix inversion lemma

\[
\left( H_1^T R_1^{-1} H_1 + H_2^T R_2^{-1} H_2 \right)^{-1} = \left( P_1^{-1} + H_2^T R_2^{-1} H_2 \right)^{-1} = P_1 - P_1 H_2^T \left( H_2 P_1 H_2^T + R_2 \right)^{-1} H_2 P_1
\]

Improved Estimate Incorporating New Measurement Set

\[ \hat{x}_1 = P_1 H_1^T R_1^{-1} z_1 \]

New estimate is a correction to the old

\[ \hat{x}_2 = \hat{x}_1 - P_1 H_2^T \left( H_2 P_1 H_2^T + R_2 \right)^{-1} H_2 \hat{x}_1 \]

\[ + P_1 H_2^T \left[ I_n - \left( H_2 P_1 H_2^T + R_2 \right)^{-1} H_2 P_1 H_2^T \right] R_2^{-1} z_2 \]

\[ = \left[ I_n - \left( H_2 P_1 H_2^T + R_2 \right)^{-1} H_2 P_1 H_2^T \right] \hat{x}_1 \]

\[ + P_1 H_2^T \left[ I_n - \left( H_2 P_1 H_2^T + R_2 \right)^{-1} H_2 P_1 H_2^T \right] R_2^{-1} z_2 \]
Simplify Optimal Estimate
Incorporating New Measurement Set

\[
I = A^{-1}A = AA^{-1}, \quad \text{with} \quad A \triangleq H_2 P_1 H_2^T + R_2
\]

\[
\hat{x}_2 = \hat{x}_1 - P_1 H_2^T \left( H_2 P_1 H_2^T + R_2 \right)^{-1} (z_2 - H_2 \hat{x}_1)
\]

\[
\triangleq \hat{x}_1 - K (z_2 - H_2 \hat{x}_1)
\]

**Estimator gain matrix**

\[
K = P_1 H_2^T \left( H_2 P_1 H_2^T + R_2 \right)^{-1}
\]

Recursive Optimal Estimate

- Prior estimate may be based on prior incremental estimate, and so on
- Generalize to a recursive form, with sequential index \( i \)

\[
\hat{x}_i = \hat{x}_{i-1} - P_{i-1} H_i^T \left( H_i P_{i-1} H_i^T + R_i \right)^{-1} (z_i - H_i \hat{x}_{i-1})
\]

\[
\triangleq \hat{x}_{i-1} - K_i (z_i - H_i \hat{x}_{i-1})
\]

\[
\text{dim}(x) = n \times 1; \quad \text{dim}(P) = n \times n
\]
\[
\text{dim}(z) = r \times 1; \quad \text{dim}(R) = r \times r
\]
\[
\text{dim}(H) = r \times n; \quad \text{dim}(K) = n \times r
\]

with

\[
P_i = \left( P_{i-1}^{-1} + H_i^T R_i^{-1} H_i \right)^{-1}
\]
Example of Recursive Optimal Estimate

\[
\begin{align*}
\hat{z} &= x + n \\
\hat{x}_i &= \hat{x}_{i-1} + p_{i-1} \left( p_{i-1} + 1 \right)^{-1} (z_i - \hat{x}_{i-1}) \\
\hat{x}_i &= \hat{x}_{i-1} + k_i (z_i - \hat{x}_{i-1})
\end{align*}
\]

With constant estimation error matrix, \( R \),
- Error covariance decreases at each step
- Estimator gain matrix, \( K \), invariably goes to zero as number of samples increases

Why?
- Each new sample has smaller effect on the average than the sample before
Next Time:
Propagation of Uncertainty in Dynamic Systems

Supplemental Material
Covariance Matrix is Expected Value of the Outer Product

- Transcript correlation for normal and abnormal samples
  - Identifies possible components of biological circuits

- Sample tissue correlation for RNA expression
  - Confirms or questions the classification of samples