Least-squares Techniques

Dr. Kamesh Subbaro and Dr. Mangal Kothari

GIAN Course: IIT Kanpur

December 2017

Linear Least Squares Estimation

For a measurable quantity, x, the following two equations hold,

$$\begin{aligned} \text{measured value} &= \text{true value} + \text{measurement error} \\ \tilde{x} &= x + v \\ \text{measured value} &= \text{estimated value} + \text{residual error} \\ \tilde{x} &= \hat{x} + e \end{aligned}$$

Note,

- The actual measurement error (v) and the true value (x) is never known in practice.
- The errors in the process/mechanism that physically generate this error are usually approximated by some known process (Gaussian noise with known variance).
- The statistical properties are utilized to weight the relative importance of the various measurements used in the estimation scheme.
- The residual error is known exactly and is easily computed once an estimated value has been found.
- The residual error drives the estimator.

Linear Batch Estimation

Consider a batch of measurements obtained at discrete instants of time:

$$\{(\tilde{y}_1, t_1), (\tilde{y}_2, t_2), (\tilde{y}_3, t_3), \cdots, (\tilde{y}_m, t_m)\}$$

We wish to model these measurements via a mathematical model.
(**REMEMBER**: YOU ARE PROPOSING THE MODEL!)

$$y(t) = \sum_{i=1}^{n} x_i h_i(t), \quad m \geq n$$

where, $h_i(t) \in \{h_1(t), h_2(t), h_3(t), \cdots, h_n(t)\}$ are a set of *independent specified basis functions*. x_i are the constants whose values are to be determined. We seek to obtain optimum x-values based upon a measure of "how well" the model predicts the measurements. Errors in the prediction are usually due to,

- measurement errors
- incorrect x-values
- modelling errors, i.e. the proposed model was bad.

Linear Batch Estimation

Lets relate the measurements ($\tilde{y}_j$) and the estimated outputs ($\hat{y}_j$).

$$\tilde{y}_j \equiv \tilde{y}(t_j) = \sum_{i=1}^{n} x_i h_i(t_j) + v_j \quad j = 1, 2, \ldots, m$$

The estimated outputs are computed using the estimated values of x and the basis functions.

$$\hat{y}_j \equiv \hat{y}(t_j) = \sum_{i=1}^{n} \hat{x}_i h_i(t_j) \quad j = 1, 2, \ldots, m$$

What about v_j?

Clearly, $\tilde{y}_j = \sum_{i=1}^{n} \hat{x}_i h_i(t_j) + e_j$, where e_j is the residual error, $e_j = \tilde{y}_j - \hat{y}_j$. We can compactly represent the above by combining the measurements at all time instants and stacking them up as,

$$\tilde{\boldsymbol{y}} = \boldsymbol{H}\hat{\boldsymbol{x}} + \boldsymbol{e} \tag{1}$$

Linear Batch Estimation

where,

$$\begin{aligned}
\tilde{\boldsymbol{y}} &= [\tilde{y}_1\ \tilde{y}_2\ \cdots\ \tilde{y}_m]^T = \text{measured } y - \text{values} \\
\boldsymbol{e} &= [\tilde{e}_1\ \tilde{e}_2\ \cdots\ \tilde{e}_m]^T = \text{residual errors} \\
\hat{\boldsymbol{x}} &= [\hat{x}_1\ \hat{x}_2\ \cdots\ \hat{x}_n]^T = \text{estimated } x - \text{values}
\end{aligned}$$

$$\boldsymbol{H} = \begin{bmatrix} h_1(t_1) & h_2(t_1) & \cdots & h_n(t_1) \\ h_1(t_2) & h_2(t_2) & \cdots & h_n(t_2) \\ \vdots & \vdots & & \vdots \\ h_1(t_m) & h_2(t_m) & \cdots & h_n(t_m) \end{bmatrix}$$

Similarly one can also develop the following equations,

$$\tilde{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v} \tag{2}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{H}\hat{\boldsymbol{x}} \tag{3}$$

Equations (1) and (2) are commonly referred to as the “observation equations”.

Linear Batch Estimation

Gauss' least squares principle selects the optimum $\hat{x}$ by minimizing the sum square of the residual errors, given by

$$J = \frac{1}{2}\boldsymbol{e}^T\boldsymbol{e} = \frac{1}{2}\left(\tilde{\boldsymbol{y}} - \boldsymbol{H}\hat{\boldsymbol{x}}\right)^T\left(\tilde{\boldsymbol{y}} - \boldsymbol{H}\hat{\boldsymbol{x}}\right)$$

or

$$J = J(\hat{\boldsymbol{x}}) = \frac{1}{2}\left(\tilde{\boldsymbol{y}}^T\tilde{\boldsymbol{y}} - 2\tilde{\boldsymbol{y}}^T\boldsymbol{H}\hat{\boldsymbol{x}} + \hat{\boldsymbol{x}}^T\boldsymbol{H}^T\boldsymbol{H}\hat{\boldsymbol{x}}\right)$$

The multiplier $1/2$ has a statistical significance (will be discussed later). Now, use matrix calculus differentiation rules to obtain the necessary condition for the minimum.

Necessary condition: $\nabla_{\hat{\boldsymbol{x}}} J = \boldsymbol{H}^T\boldsymbol{H}\hat{\boldsymbol{x}} - \boldsymbol{H}^T\tilde{\boldsymbol{y}} = 0$, i.e., $\hat{\boldsymbol{x}} = \left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\tilde{\boldsymbol{y}}$

Sufficient condition: $\nabla^2_{\hat{\boldsymbol{x}}} J \equiv \dfrac{\partial^2 J}{\partial\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}^T} = \boldsymbol{H}^T\boldsymbol{H} > 0$ ($\boldsymbol{H}^T\boldsymbol{H}$ is positive definite)

The inverse of $\boldsymbol{H}^T\boldsymbol{H}$ is required. A good choice of the basis functions is important.

Weighted Least Squares Estimation

If each measurement is made with different precisions, it is better if this aspect is captured by weighting the measurements accordingly. The choice of the weights is non-unique but it turns out that the inverse of the measurement error variance is an intuitive choice.

$$J = \frac{1}{2}\boldsymbol{e}^T \boldsymbol{W}\boldsymbol{e}$$

Necessary condition: $\nabla_{\hat{\boldsymbol{x}}} J = \boldsymbol{H}^T \boldsymbol{W}\boldsymbol{H}\hat{\boldsymbol{x}} - \boldsymbol{H}^T \boldsymbol{W}\tilde{\boldsymbol{y}} = 0$, i.e.,
$\hat{\boldsymbol{x}} = \left(\boldsymbol{H}^T \boldsymbol{W}\boldsymbol{H}\right)^{-1} \boldsymbol{H}^T \boldsymbol{W}\tilde{\boldsymbol{y}}$

Sufficient condition: $\nabla^2_{\hat{\boldsymbol{x}}} J \equiv \dfrac{\partial^2 J}{\partial \hat{\boldsymbol{x}} \hat{\boldsymbol{x}}^T} = \boldsymbol{H}^T \boldsymbol{W}\boldsymbol{H} > 0$ ($\boldsymbol{H}^T \boldsymbol{W}\boldsymbol{H}$ is positive definite)

$\boldsymbol{W}$ is typically chosen to be a diagonal matrix. A subset of the w_{ii} are chosen much larger than the others to reflect the preciseness of that specific subset of measurements.

Nonlinear Least Squares Estimation

- Most real world estimation problems are nonlinear
- Linear versions of the estimation problem and associated developments apply only to a subset of problems encountered in practice.
- Most nonlinear estimation problems can be accurately solved by a judiciously chosen successive approximation procedure.

Most widely used Successive Approximation Procedure $\Rightarrow$ Nonlinear Least Squares - also known as **Gaussian Least Squares Differential Correction**

(Early application by Gauss in 1800s to determine planetary orbits from telescope measurements of the "line of sight angles" to the planets)

Nonlinear Least Squares Estimation

Measurement Model: $\tilde{\boldsymbol{y}} = f(\boldsymbol{x}) + \boldsymbol{v}$

$$\begin{aligned}
\tilde{\boldsymbol{y}} = [\tilde{y}_1\ \tilde{y}_2\ \ldots\ \tilde{y}_m]^T &= \text{measured } y\text{—values} \\
\boldsymbol{f}(x) = [f_1\ f_2\ \ldots\ f_m]^T &= \text{independent functions} \\
\boldsymbol{x} = [x_1\ x_2\ \ldots\ x_n]^T &= \text{true } x\text{—values} \\
\boldsymbol{v} = [v_1\ v_2\ \ldots\ v_m]^T &= \text{measurement errors}
\end{aligned}$$

Estimated y-values:

$$\hat{\boldsymbol{y}} = f(\hat{\boldsymbol{x}})$$

$$\boldsymbol{e} = \tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}} \equiv \Delta \boldsymbol{y}$$

$$\begin{aligned}
\hat{\boldsymbol{y}} = [\hat{y}_1\ \hat{y}_2\ \ldots\ \hat{y}_m]^T &= \text{estimated } y\text{—values} \\
\hat{\boldsymbol{x}} = [\hat{x}_1\ \hat{x}_2\ \ldots\ \hat{x}_n]^T &= \text{estimated } x\text{—values} \\
\boldsymbol{e} = [e_1\ e_2\ \ldots\ e_m]^T &= \text{residual errors}
\end{aligned}$$

Rewrite Measurement Model:

$$\tilde{\boldsymbol{y}} = f(\hat{\boldsymbol{x}}) + \boldsymbol{e}$$

As before, we seek an estimate ($\hat{\boldsymbol{x}}$) for $\boldsymbol{x}$ that minimizes

$$J = \frac{1}{2}\boldsymbol{e}^T W \boldsymbol{e} = \frac{1}{2}[\tilde{\boldsymbol{y}} - \boldsymbol{f}(\hat{\boldsymbol{x}})]^T W [\tilde{\boldsymbol{y}} - \boldsymbol{f}(\hat{\boldsymbol{x}})]$$

W is an $m \times m$ weighting matrix used to weight the relative importance of each measurement.

Gauss' Procedure:

Assume *current* estimates of the unknown $\boldsymbol{x}$-values are available

$$\boldsymbol{x}_c = [x_{1c}\ x_{2c}\ \ldots\ x_{nc}]^T$$

Whatever the unknown objective $\boldsymbol{x}$-values, $\hat{\boldsymbol{x}}$ are, we assume they are related to the respective current estimates, by an also unknown set of corrections $\Delta\boldsymbol{x}$.

$$\hat{\boldsymbol{x}} = \boldsymbol{x}_c + \Delta\boldsymbol{x}$$

Linearize $\boldsymbol{f}(\hat{\boldsymbol{x}})$ about $\boldsymbol{x}_c$.

$$\boldsymbol{f}(\hat{\boldsymbol{x}}) \approx \boldsymbol{f}(\boldsymbol{x}_c) + H\Delta\boldsymbol{x}$$

where,

$$H = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}|_{\boldsymbol{x}_c}$$

The Gradient Matrix H is known as the *Jacobian* matrix. The measurement residual after the correction can now be linearly approximated as

$$\Delta\boldsymbol{y} \equiv \tilde{\boldsymbol{y}} - \boldsymbol{f}(\hat{\boldsymbol{x}}) \approx \tilde{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{x}_c) - H\Delta\boldsymbol{x} = \Delta\boldsymbol{y}_c - H\Delta\boldsymbol{x}$$

where, the residual "before the correction" is

$$\Delta\boldsymbol{y}_c \equiv \tilde{\boldsymbol{y}} - \boldsymbol{f}(\boldsymbol{x}_c)$$

Objective: Minimize weighted sum squares J

Strategy: To determine approximate corrections in $\Delta \boldsymbol{x}$, select particular corrections that lead to *minimum sum of squares of the linearly predicted residuals*, J_p:

$$J = \frac{1}{2}\Delta \boldsymbol{y}^T W \Delta \boldsymbol{y} \approx J_p \equiv \frac{1}{2}(\Delta \boldsymbol{y}_c - H\Delta \boldsymbol{x})^T W (\Delta \boldsymbol{y}_c - H\Delta \boldsymbol{x})$$

Following the minimization procedure as before, one obtains

$$\Delta \boldsymbol{x} = (H^T W H)^{-1} H^T W \Delta \boldsymbol{y}_c$$

An initial guess $\Delta \boldsymbol{x}_c$ is essential to begin the procedure. A stopping condition with an accuracy dependent tolerance is given by

$$\delta J \equiv \frac{|J_i - J_{i-1}|}{J_i} < \frac{\varepsilon}{\|W\|}$$

where, ε is some prescribed small value. If the condition is not satisfied then the update procedure is iterated with the new estimate.