Linear Least-Squares Estimation

Waquar Ahmed
Prashant Tripathi
Y8 Batch

Objective

- Least Square Approach
- Least Square Estimate for Scalar and Vector Parameter
- Geometrical Interpretation

Least Square Approach

- All the previous methods required a Probabilistic model for the Data.
- Least Square is not statistically based

  (a) It does not need a *pdf* model

  (b) It need a Deterministic Signal Model

Least Square Approach

Minimize the LS Cost

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} \varepsilon^2[n] = \sum_{n=0}^{N-1} \left(x[n] - s[n;\boldsymbol{\theta}]\right)^2$$

- The Value Of ‘θ’ that minimizes the J(θ) is the LSE
- Method is equally valid for a Gaussian as well as non-Gaussian noise.
- It is applied when statistical characterization of data is unknown.

Small Example:

- Estimate of DC level

Assuming the given signal model as s[n] = A and observed signal x[n] for n = 0,1,…,N-1.

Now using LS approach we have

$$J(A) = \sum_{n=0}^{N-1} (x[n] - A)^2$$

$$set \frac{\partial J(A)}{\partial A} = 0 \Rightarrow \hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x}$$

Some more example:

Consider the signal model s[n] = $A\cos 2\pi f_0 n$

in which the frequency f_0 to be estimated.

Assume 'A' is known parameter.

LSE is found by minimizing

$$J(f_0) = \sum_{n=0}^{N-1} (x[n] - A\cos 2\pi f_0 n)^2$$

In this case LS error is highly non-linear in frequency. Hence minimization cannot be done in closed form.

Some inferences:

- Signal that is linear in unknown parameter is quadratic function in J.

- A Signal model that is linear in the unknown parameter is said to generate a linear least squares problem, else non-linear least square problem is said to be generated.

- Suppose in the previous example if ‘A’ is unknown parameter and f_0 is known then it can become a problem of linear least squares estimation.
- It may also be the possibility that both ‘A’ and f_0 are unknown and it become a case of vector parameter estimation.

Weighted Least Square Criterion

- Sometimes not all data samples are equally good:

  x[0], x[1],….,x[N-1]

  Say we know x[10] was poor in quality compared to other data. Thus we want to de-emphasize the x[10] in the sum of squares.

$$J(\theta) = \sum_{n=0}^{N-1} w_n (x[n] - S[n,\theta])^2$$

Linear Least Squares

- Take the case of scalar parameter,

  let $s[n] = \theta h[n]$

  The LS error criterion becomes

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - \theta h[n])^2$$

Minimizing this LSE gives

$$\hat{\theta} = \frac{\sum_{n=0}^{N-1} x[n]h[n]}{\sum_{n=0}^{N-1} h^2[n]}$$

LSE (Scalar parameter contd.)

- Substitute $\hat{\theta}$ in J(θ), we get

$$J_{\min} = \sum_{n=0}^{N-1} x^2[n] - \frac{\left(\sum_{n=0}^{N-1} x[n]h[n] \right)^2}{\sum_{n=0}^{N-1} h^2[n]}$$

LSE (Vector Parameter)

- Let us consider a vector parameter θ of dimension px1.
- For the signal s= { s[0],s[1],…s[N-1]} to be linear in the unknown parameters such that

  s = H θ , where H is the observation matrix N x p.

  LSE is found by minimizing

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - s[n])^2$$

$$= (X - H\theta)^T (X - H\theta)$$

LSE (Vector Parameters contd…)

$$J(\theta) = X^T X - X^T H\theta - \theta^T H^T X + \theta^T H^T H\theta$$
$$= X^T X - 2X^T H\theta + \theta^T H^T H\theta$$

Now differentiating J(θ) w.r.t θ

$$\frac{\partial J(\theta)}{\partial \theta} = -2H^T X + 2H^T H\theta$$

Setting the gradient equal to zero yields the LSE

$$\hat{\theta} = (H^T H)^{-1} H^T X$$

LSE (Vector Parameters contd…)

- The minimum LS error is found as

$$J_{\min} = J(\hat{\theta})$$
$$= (X - H\hat{\theta})^T (X - H\hat{\theta})$$
$$= X^T (I - H(H^T H)^{-1} H^T) X$$

- An extension of Linear LS problem is weighted LS

J(θ) $=(X - H\hat{\theta})^T W (X - H\hat{\theta})$

Geometrical Interpretation

- Linear LS approach using Geometrical Perspective.
- Let S=Hθ be the general signal model.
- If we denote columns of H by hi then we can write

$$S = [h1 \quad h2 \quad . \quad . \quad hp]\begin{bmatrix} \theta 1 \\ \theta 2 \\ . \\ . \\ \theta p \end{bmatrix}$$

$$= \sum_{i=1}^{p} \theta_i h_i$$

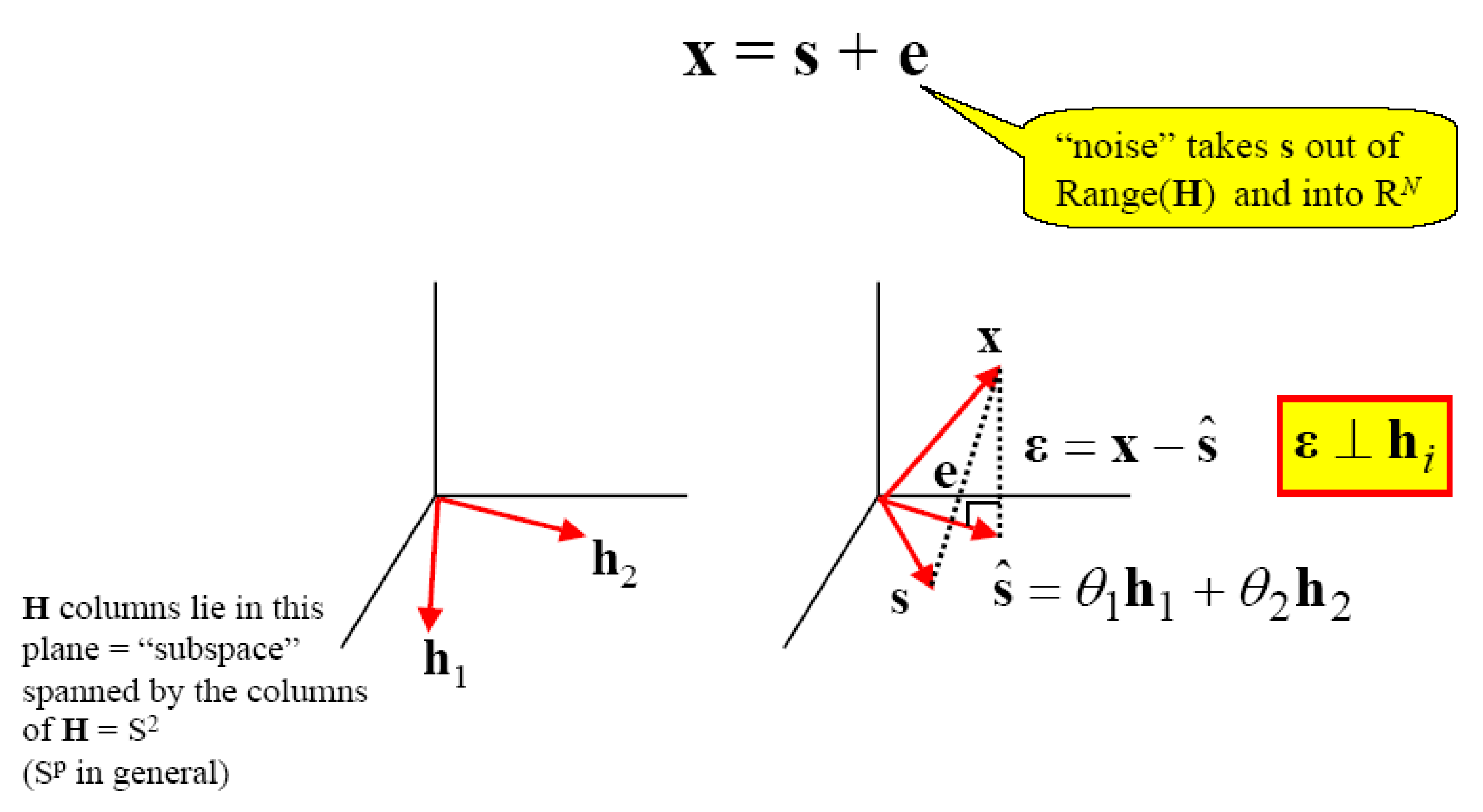
- Signal model can be seen to be linear combination of the “signal vectors” {h1,h2,….hp}

Geometrical Interpretation(contd…)

- Linear LS approach attempts to minimize the square of the distance from the data vector x to a signal vector.
- Data vector can lie anywhere in an N dimensional space.
- Signal vector must lie in p-dimensional subspace of N-dimensional space.
- Full rank of H assures that columns are linearly independent

Geometrical Interpretation(contd…)

- LS example with N=3, p=2



Geometrical Interpretation(contd…)

- Error vector is orthogonal to signal subspace

$$\left(X - \hat{S}\right) \perp S^2$$

$$\Rightarrow \left(X - \hat{S}\right) \perp h1$$

$$\left(X - \hat{S}\right) \perp h2$$

Using the definition of orthogonality, we have

$$\left(X - \hat{S}\right)^T h1 = 0$$

$$\left(X - \hat{S}\right)^T h2 = 0$$

Geometrical Interpretation(contd…)

- Using $S = \theta_1 h_1 + \theta_2 h_2$, we can write

$$(X - H\theta)^T [h1\ h2] = 0^T$$

$$(X - H\theta)^T H = 0^T$$

$$\Rightarrow \theta = (H^T H)^{-1} H^T X$$

- So finally we have got the same solution.