Note Title 27-10-2014

# Least Squares Estimation (LS)

27/10

- Alternative to MMSE estimation
- Wide range of applications
- Originally developed by Gauss while studying planetary motions

Problem statement:

Let $\underline{x}_{N\times 1}$ vector $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$

Let $H$ be $N \times M$ matrix

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ \vdots & & & \\ h_{N1} & \cdots & & h_{NM} \end{bmatrix}$$

Given $\underline{x}$ & $H$ find $\underline{w}_{M\times 1}$ such that

$\|\underline{x} - H\underline{w}\|^2$ is minimum

Two examples of LS problem

1. Ergodic Averaging of MMSE

Recall MMSE problem

$$E\{|X - \underline{w}^* \underline{Y}|^2\}$$

X is a scalar random variable

$\underline{Y}$ is $M \times 1$ random vector

Suppose we have N realizations of X & $\underline{Y}$

$X(1), \ldots, X(N)$ $\quad$ $\underline{Y}(1), \underline{Y}(2), \ldots, \underline{Y}(N)$

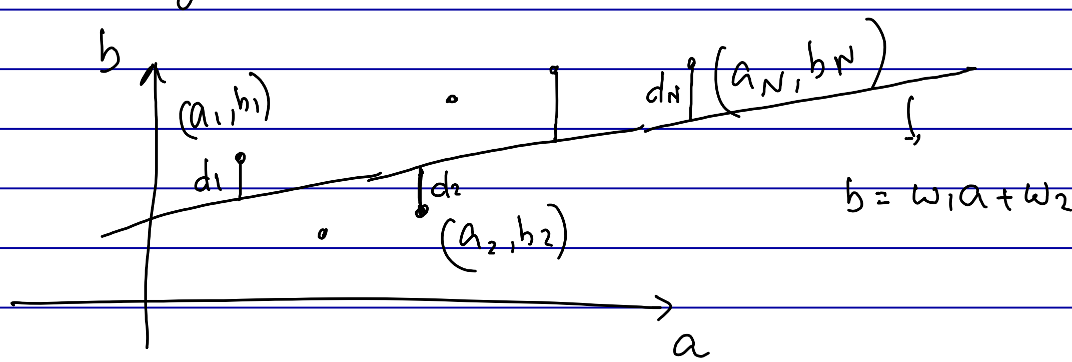
$$E|X - \underline{w}^* \underline{Y}|^2 \approx \frac{1}{N} \sum_{i=1}^{N} |X(i) - \underline{w}^* \underline{Y}(i)|^2$$

$$\sum_{i=1}^{N} |X(i) - \underline{w}^* \underline{Y}(i)|^2$$

$$= \underbrace{\| [X(1) \ X(2) \ \ldots \ X(N)] - \underline{w}^* [\underline{Y}(1) \ldots \underline{Y}(N)] \|^2}_{\text{If we take } (\cdot)^* \text{ then we have the LS problem.}}$$

2. Curve fitting problems.

We are given N points in 2D plane



$(a_1, b_1), (a_2, b_2), \ldots, (a_N, b_N)$ are given points

$$d_i = b_i - (w_1 a_i + w_2)$$

Find $w_1, w_2$ such that $\sum_{i=1}^{N} |d_i|^2$ is minimized

$$\underline{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}}_{\underline{b}} - \underline{w_1} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} - w_2 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$= \underline{b} - \underbrace{\begin{bmatrix} a_1 & 1 \\ \vdots & 1 \\ a_N & \vdots \\ & 1 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{\underline{w}}$$

$$\|\underline{d}\|^2 = \|\underline{b} - H\underline{w}\|^2$$

minimizing $\|\underline{d}\|^2$ wrt $w_1, w_2$ is LS problem

# Solution for LS Estimation problem

Three Approaches

1. Geometric Approach
2. Differentiation approach
3. Completion of Squares approach

## Geometric Approach.

Given $\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}_{N \times 1}$

Given $H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ \vdots & \vdots & & \vdots \\ h_{N1} & h_{N2} & & h_{NM} \end{bmatrix}_{N \times M}$

$$\downarrow \qquad \downarrow \qquad \downarrow$$

$$\underline{h}(1) \qquad \underline{h}(2) \qquad \underline{h}(M)$$

$\underset{N \times 1}{\underline{h}(i)}$ is $i^{th}$ column of $H$

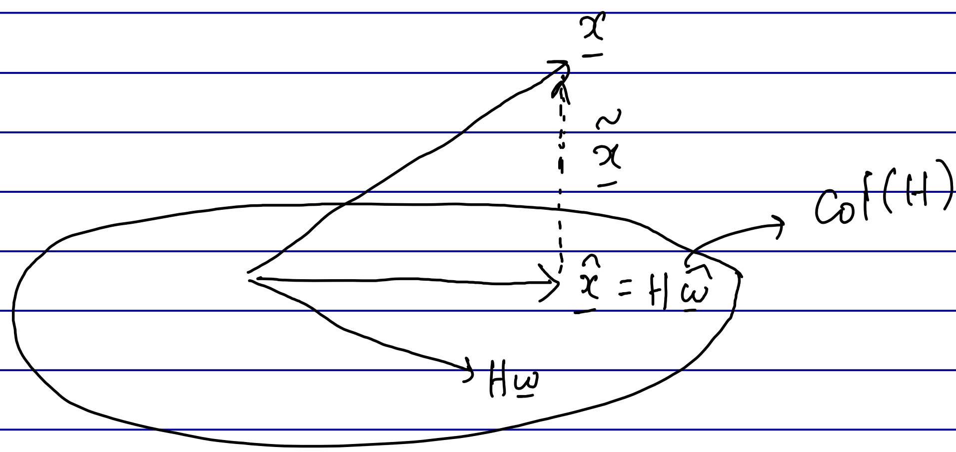
We want to find $\underline{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix}$

such that

$$\| \underline{x} - H \underline{w} \| \text{ is minimum}$$

$$\underline{x} - H\underline{w} = \underline{x} - \underbrace{\left(w_1 \underline{h}(1) + w_2 \underline{h}(2) + \dots + w_M \underline{h}(M)\right)}_{\in \text{ Column Space of } H}$$

$\text{Col}(H)$ is set of all possible linear combinations of columns of $H$.




Need to choose $\underline{w}$ such that euclidean distance between $\underline{x}$ and $H\underline{w}$ is minimized

Will use the following result from linear algebra

Theorem: Any vector $\underline{x}$ can be uniquely written as

$$\underline{x} = \underline{\hat{x}} + \underline{\tilde{x}}$$

where $\hat{\underline{x}} \in Col(H)$

and $\tilde{\underline{x}}$ is orthogonal to $Col(H)$

- $\hat{\underline{x}}$ is called projection of $\underline{x}$ onto $Col(H)$
- $\tilde{\underline{x}}$ is called projection of $\underline{x}$ onto orthogonal complement of $Col(H)$

**Theorem:**

Suppose $\hat{\underline{w}}$ is such that

$$H\hat{\underline{w}} = \hat{\underline{x}}.$$

Then $\hat{\underline{w}}$ is optimal LS solution

**Proof,**

Consider $\| \underline{x} - H\underline{w} \|^2$

$\hat{\underline{x}} \in col(H)$

$\tilde{\underline{x}}$ is orthogonal to $col(H)$

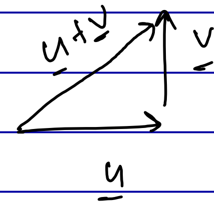
$$= \| \hat{\underline{x}} + \tilde{\underline{x}} - H\underline{w} \|^2$$

$$= \| H\hat{\underline{w}} + \tilde{\underline{x}} - H\underline{w} \|^2$$

$$= \| \tilde{\underline{x}} + (H\hat{\underline{w}} - H\underline{w}) \|^2$$

$\tilde{\underline{x}}$ is orthogonal to $H\hat{\underline{w}} - H\underline{w}$

Pythagoras Theorem $\rightarrow \; = \|\underline{\tilde{x}}\|^2 + \underbrace{\|H\underline{\hat{w}} - H\underline{w}\|^2}_{\geq 0}$



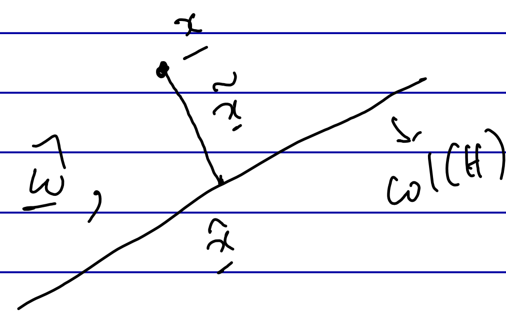

$$\geq \|\underline{\tilde{x}}\|^2$$

equality if and only if

$$H\underline{\hat{w}} = H\underline{w} = \underline{\hat{x}}$$

---

28/10 **Orthogonality Criterion**




For the optimal LS solution $\underline{\hat{w}}$,

$\underline{\tilde{x}} = \underline{x} - H\underline{\hat{w}}$ is orthogonal to the column space of H.

$\underline{x} - H\underline{\hat{w}}$ is orthogonal to every vector in col(H)

Let $\underline{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix}$

$$H\underline{p} = p_1 \underline{h}(1) + p_2 \underline{h}(2) + \dots + p_M \underline{h}(M)$$

$$H\underline{p} \in \mathrm{col}(H)$$

$$\underline{x} - H\hat{\underline{w}} \perp H\underline{p}$$

$$(H\underline{p})^* (\underline{x} - H\hat{\underline{w}}) = 0$$

for any $\underline{p}$, $\quad \underline{p}^* H^* (\underline{x} - H\hat{\underline{w}}) = 0$

$$\Rightarrow \quad H^* (\underline{x} - H\hat{\underline{w}}) = 0$$

$$\Rightarrow \quad \boxed{H^* H \hat{\underline{w}} = H^* \underline{x}}$$

Normal equations for LS

---

<u>Differentiation Approach to LS solution.</u>

Consider the cost function

$$J(\underline{w}) = \| \underline{x} - H\underline{w} \|^2$$

$$= (\underline{x} - H\underline{w})^* (\underline{x} - H\underline{w})$$

$$= \|\underline{x}\|^2 - \underline{x}^* H \underline{w} - \underline{w}^* H^* \underline{x} + \underline{w}^* H^* H \underline{w}$$

Take derivative wrt $\underline{w}$ and equate to zero

$$\nabla_{\underline{w}} J(\underline{w}) = 0$$

$$\Rightarrow \quad -\underline{x}^* H + \underline{w}^* H^* H = 0$$

$$\Rightarrow \quad \boxed{H^* H \underline{w} = H^* \underline{x}}$$

Optimal LS solution satisfies the normal equation

---

Properties of LS Solution.

Summary of LS

- LS criterion is minimize

$$\| \underline{x} - H\underline{w} \|^2 \quad \text{wrt } \underline{w}$$

given $\underline{x}$, $H$

Let $\hat{\underline{w}}$ be optimal solution

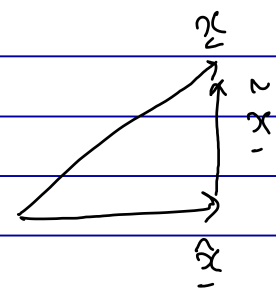
- $$H^* H \hat{\underline{w}} = H^* \underline{x}$$

Normal equations

- $H\hat{\underline{w}} = \hat{\underline{x}} \rightarrow$ projection of $\underline{x}$ onto col$(H)$

- $\tilde{\underline{x}} = \underline{x} - \hat{\underline{x}}$ is orthogonal to col$(H)$

- $\| \underline{x} - H\hat{\underline{w}} \|^2 = \| \tilde{\underline{x}} \|^2$




From Pythagoras theorem,

$$\| \underline{x} \|^2 = \| \tilde{\underline{x}} \|^2 + \| \hat{\underline{x}} \|^2$$

So $\| \tilde{\underline{x}} \|^2 = \| \underline{x} \|^2 - \| \hat{\underline{x}} \|^2$

The minimum LS cost

$$= \| \tilde{\underline{x}} \|^2 = \| \underline{x} \|^2 - \| \hat{\underline{x}} \|^2$$

$$= \| \underline{x} \|^2 - \| H\hat{\underline{w}} \|^2$$

x ———————————————— x

Properties of LS solution.

1. Solution for LS problem <u>always</u> exists

Given $\underline{x}$ and $H$ we can

always find $\hat{\underline{x}}$ which is

projection of $\underline{x}$ onto $\text{col}(H)$.

Since $\hat{\underline{x}} \in \text{col}(H)$, $\hat{\underline{x}}$ can

be written linear combination of

columns of $H$

$\exists \; \hat{w}_1, \hat{w}_2, \ldots, \hat{w}_M$ such that

$$\hat{\underline{x}} = \hat{w}_1 \underline{h}(1) + \hat{w}_2 \underline{h}(2) + \cdots + \hat{w}_M \underline{h}(M)$$

$$= H \hat{\underline{w}}$$

where $\hat{\underline{w}} = \begin{bmatrix} \hat{w}_1 \\ \hat{w}_2 \\ \vdots \\ \hat{w}_M \end{bmatrix}$

So LS solution $\hat{\underline{w}}$ always exist

$H^H H \hat{\underline{w}} = H^H \underline{x}$ will always have solution

## 2. Uniqueness of Solution

LS Solution $\hat{\underline{w}}$ is unique if and only if all the columns of $H$ are linearly independent, ie) $\text{rank}(H) = M$

$H_{N \times M}$

Proof: We need the following result

$$\text{rank}(H^* H) = \text{rank}(H)$$

$$\text{rank}(H) = M = \text{rank}(H^* H)$$

$H^* H \to M \times M$ matrix

$\to$ full rank

$\to$ it is invertible

Normal equation

$$H^* H \hat{\underline{w}} = H^* \underline{x}$$

$$\hat{\underline{w}} = (H^* H)^{-1} H^* \underline{x}$$

unique solution

3. If $\text{rank}(H) < M$ then

there are infinitely many LS solutions.

<u>Proof:</u> $H_{N \times M}$

$\text{rank}(H) < M$

$\Downarrow$

number of linearly independent rows or columns

The columns of $H$ are linearly dependent

since $\text{rank}(H) <$ number of columns

So $\exists\ p_1, p_2, \ldots, p_M$ (not all zeros)

such that

$$p_1 \underline{h}(1) + p_2 \underline{h}(2) + \cdots + p_M \underline{h}(M) = \underline{0}$$

$$H \underline{p} = \underline{0} \quad \text{where} \quad \underline{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix}$$

$\underline{p} \in$ Null space of $H$

$\Downarrow$

set of all $\underline{q}$ such that $H \underline{q} = \underline{0}$

$\alpha \underline{p} \in$ Null space of $H$ for any scalar $\alpha$

There are infinitely many vectors in Null space of $H$.

Now, let $\hat{\underline{w}}$ be a LS solution

then $\hat{\underline{w}} + \underline{q}$ for any $\underline{q} \in$ Null space of $H$

is also a LS solution

We can verify this easily

$$H^{\dagger} H \hat{\underline{w}} = H^{\dagger} \underline{x}$$

Now,
$$H^{\dagger} H (\hat{\underline{w}} + \underline{q}) = \underbrace{H^{\dagger} H \hat{\underline{w}}} + \underbrace{H^{\dagger} H \underline{q}}$$
$$= H^{\dagger} \underline{x} + \underline{0}$$

$\hat{\underline{w}} + \underline{q}$ also satisfies the normal equation.

$\hat{\underline{w}} + \underline{q}$ is also LS solution

So There are infinitely many LS solutions

---

Remarks.

1. Undetermined LS problem

$H_{N \times M}$ with $N < M$.

$$\text{rank}(H) \le \min(N, M)$$
$$\le N$$

$< \quad M$

$\Rightarrow$ There are infinite LS solutions

2. Over determined LS problem

$H_{N \times M}$ with $N \geq M$.

unique LS solution if $\text{rank}(H) = M$

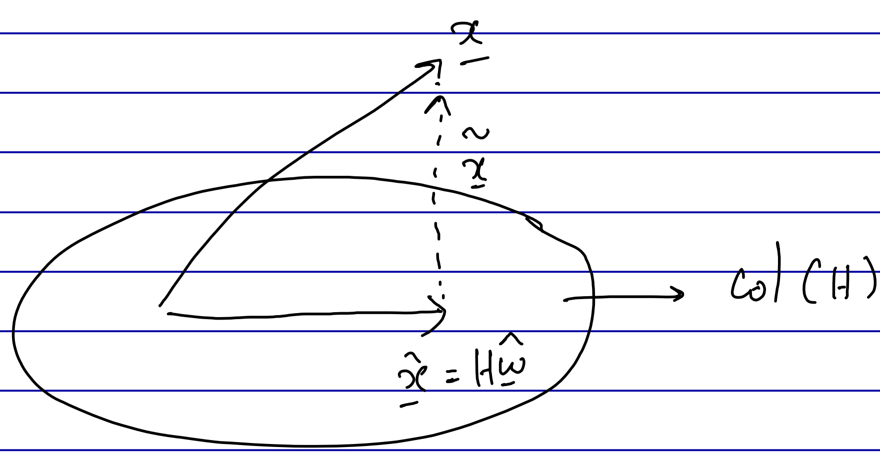
Infinitely many solutions if $\text{rank}(H) < M$.

---

<u>Projection Matrices.</u> $H_{N \times M}$ matrix

We assume $H$ is full column rank

ie $\text{rank}(H) = M$.




Given $\underline{x}$ and $H$, how to find the

projection of $\underline{x}$ onto $\text{Col}(H)$

Let $\underline{\hat{w}}$ be LS solution

$$H^* H \underline{\hat{w}} = H^* \underline{x}$$

$$\underline{\hat{w}} = (H^* H)^{-1} H^* \underline{x}$$ since $(H^* H)^{-1}$ exist.

$\underline{\hat{x}} \rightarrow$ projection of $\underline{x}$ onto $col(H)$ is given by

$$\underline{\hat{x}} = H \underline{\hat{w}}$$

$$\boxed{\underline{\hat{x}} = H (H^* H)^{-1} H^* \underline{x}}$$

Define Projection matrix $P_H = H (H^* H)^{-1} H^*$

$$\underline{\hat{x}} = P_H \underline{x}$$

Now the projection onto orthogonal complement of $col(H)$

$\underline{\tilde{x}}$ is given by

$$\underline{\tilde{x}} = \underline{x} - \underline{\hat{x}}$$

$$= \underline{x} - P_H \underline{x}$$

$$\tilde{\underline{x}} = (\underline{I} - P_H)\,\underline{x}$$

$\times$ ———————— $\times$

Properties of projection matrix

(Conjugate)

① Symmetry $\quad P_H^* = P_H$

$$P_H^* = \left( H (H^* H)^{-1} H^* \right)^*$$

$$= (H^*)^* \left( (H^* H)^{-1} \right)^* (H)^*$$

$$(A^{-1})^* = (A^*)^{-1}$$

$$= H \left( (H^* H)^* \right)^{-1} H^*$$

$$= H (H^* H)^{-1} H^*$$

$$= P_H$$

② Idempotent $\quad P_H^2 = P_H$

$$P_H^2 = P_H \cdot P_H$$

$$= H (H^* H)^{-1} \underbrace{H^* \cdot H}\, \underbrace{(H^* H)}^{-1} H^*$$

$$= H (H^* H)^{-1} H^* = P_H$$

(a) This implies $P_H \hat{\underline{x}} = \hat{\underline{x}}$

$$P_H \cdot (P_H \underline{x}) = P_H^2 \underline{x}$$
$$= P_H \underline{x}$$
$$= \hat{\underline{x}}$$

Projection of a vector which is already in col(H) (subspace) onto the same subspace gives the same vector.

(b) $P_H \tilde{\underline{x}} = \underline{0}$

Projection of a vector orthogonal to the subspace gives $\underline{0}$

$$P_H \tilde{\underline{x}} = P_H (I - P_H) \underline{x}$$
$$= (P_H - P_H^2) \underline{x}$$
$$= (P_H - P_H) \underline{x}$$
$$= \underline{0}$$

x ——— x

ASP HW3

31/10

# Weighted Least Squares (WLS)

(Conjugate)

Let $A$ be symmetric and positive definite matrix.

LS : $\min_{\underline{w}} \|\underline{x} - H\underline{w}\|^2$

WLS : $\min_{\underline{w}} (\underline{x} - H\underline{w})^* A (\underline{x} - H\underline{w})$

$A \to$ flexibility to weigh each coordinate of $\underline{x} - H\underline{w}$ seperately

For instance, if $A = \begin{bmatrix} a_1 & & & 0 \\ & a_2 & & \\ & & \ddots & \\ 0 & & & a_N \end{bmatrix}$

then $i$th coordinate of $\underline{x} - H\underline{w}$ gets weighted by $a_i$

Want to solve WLS problem:

By using change of coordinates we can transform WLS problem to LS problem

Let eigen decomposition of

$$A = U \Delta U^*$$

$U \rightarrow$ orthonormal eigen vector matrix

$\Delta \rightarrow$ diagonal eigen value matrix

Let $\Delta = \begin{bmatrix} \delta_1 & & 0 \\ & \delta_2 & \\ & & \ddots \\ 0 & & \delta_N \end{bmatrix}$

Define $\Delta^{1/2} = \begin{bmatrix} \sqrt{\delta_1} & & 0 \\ & \sqrt{\delta_2} & \\ & & \ddots \\ 0 & & \sqrt{\delta_N} \end{bmatrix}$

Consider WLS cost.

$$(\underline{x} - H\underline{w})^* A (\underline{x} - H\underline{w})$$

$$= (\underline{x} - H\underline{w})^* U \Delta U^* (\underline{x} - H\underline{w})$$

Change of variables

$\underline{b} = \Delta^{1/2} U^* \underline{x}$

$B = \Delta^{1/2} U^* H$

$$= \underbrace{(\underline{x} - H\underline{w})^* U \Delta^{1/2}}_{(\underline{b} - B\underline{w})^*} \underbrace{\Delta^{1/2} U^* (\underline{x} - H\underline{w})}_{(\underline{b} - B\underline{w})}$$

$$= \| \underline{b} - B\underline{w} \|^2$$

$\searrow$ LS cost

By doing change of variables

WLS problem $\min_{\underline{w}} (\underline{x} - H\underline{w})^* A (\underline{x} - H\underline{w})$ is equivalent to LS problem $\min_{\underline{w}} \| \underline{b} - B\underline{w} \|^2$

We know how to solve LS problem

Normal equation is

$$B^* B \hat{\underline{w}} = B^* \underline{b}$$

Now,

$$B^* B = (\Delta^{1/2} U^* H)^* (\Delta^{1/2} U^* H)$$

$$= H^* \underbrace{U \Delta^{1/2} \Delta^{1/2} U^*}_{A} H$$

$$= H^* A H$$

Similarly

$$B^* \underline{b} = H^* A \underline{x}$$

Optimal WLS satisfies

$$H^* A H \hat{\underline{w}} = H^* A \underline{x}$$ → Normal eqn for WLS

Properties WLS solution

1. Solution $\hat{\underline{w}}$ will always exist

2. If $rank(H) = M$ (full column rank)

then WLS Solution $\hat{\underline{w}}$ is unique

3. If $rank(H) < M$ then

infinitely many WLS solutions

4 minimum WLS cost :

$$= \underline{x}^* A \underline{x} - (H\hat{\underline{w}})^* A (H\hat{\underline{w}})$$

Let us verify this

$$(\underline{x} - H\hat{\underline{w}})^* A (\underline{x} - H\hat{\underline{w}})$$

$$= \underline{x}^* A (\underline{x} - H\hat{\underline{w}}) - (H\hat{\underline{w}})^* A (\underline{x} - H\hat{\underline{w}})$$

$$= \underline{x}^* A \underline{x} - \underline{x}^* A H \hat{\underline{w}} - \hat{\underline{w}}^* \underbrace{H^* A (\underline{x} - H\hat{\underline{w}})}_{H^* A \underline{x} - H^* A H \hat{\underline{w}}}$$

||

$$x^* A H = (H^* A x)^*$$

$$= (H^* A H \hat{w})^*$$

$\parallel$, $\underline{0}$ from WLS normal equation

So min WLS cost is

$$\underline{x}^* A x - (H^* A H \hat{\underline{w}})^* \underline{\hat{w}}$$

$$= \underline{x}^* A x - (H \hat{\underline{w}})^* A (H \hat{w})$$

---

## Regularized Least Squares

Cost function includes a term on the norm of the solution.

( similar to leaky LMS)

Reg. LS criterion is

$$\min_{\underline{w}} \; \underline{w}^* \Pi \underline{w} + \| \underline{x} - H \underline{w} \|^2$$

$\Pi$ is a symmetric positive definite (weighting) matrix

$\underline{x}$, $H$ are given data

Here also we can do change of variables and reduce it to a LS problem.

Define $\underline{b} = \begin{bmatrix} \underline{0}_{M\times 1} \\ \underline{x} \end{bmatrix}$

$(M+N) \times 1$ vector.

Let eigen decomposition of

$$\Pi = U \Delta U^*$$

Define matrix $B = \begin{bmatrix} \Delta^{1/2} U^* \\ \hdashline H \end{bmatrix}$

$(M+N) \times M$ matrix

Consider $\| \underline{b} - B \underline{w} \|^2$

$$= \left\| \begin{bmatrix} \underline{0} \\ \underline{x} \end{bmatrix} - \begin{bmatrix} \Delta^{1/2} U^* \\ H \end{bmatrix} \underline{w} \right\|^2$$

$$= \left\| \begin{bmatrix} 0 - \Delta^{1/2} U^* \underline{w} \\ \underline{x} - H \underline{w} \end{bmatrix} \right\|^2$$

easy to verify $\rightsquigarrow$ $= \|0 - \Delta^{1/2} U^* \underline{w}\|^2 + \|\underline{x} - H\underline{w}\|^2$

$$= \underline{w}^* U \Delta^{1/2} \Delta^{1/2} \underline{w} + \|\underline{x} - H\underline{w}\|^2$$

$$= \underline{w}^* \Pi \underline{w} + \|\underline{x} - H\underline{w}\|^2$$

We have reduced reg. LS problem to (standard) LS problem

$$\min_{\underline{w}} \underline{w}^* \Pi \underline{w} + \|\underline{x} - H\underline{w}\|^2 \qquad \min_{\underline{w}} \|\underline{b} - B\underline{w}\|^2$$

Optimal Reg. LS solution $\hat{\underline{w}}$ satisfies normal equation

$$B^* B \hat{\underline{w}} = B^* \underline{b}$$

$\Downarrow$

we can write this in terms of original variables

$$(\Pi + H^* H) \hat{\underline{w}} = H^* \underline{x}$$

$\Pi + H^* H$ is always full rank

So reg. LS has unique solution

$$\hat{\underline{w}} = \left(\Pi + H^* H\right)^{-1} H^* \underline{x}$$

3/11

Minimum Reg. LS cost:

$$\xi = (\underline{b} - B\hat{\underline{w}})^* (\underline{b} - B\hat{\underline{w}})$$

using the equivalence with standard LS

$$= \underline{b}^* (\underline{b} - B\hat{\underline{w}}) - \hat{\underline{w}}^* B^* \underbrace{(\underline{b} - B\hat{\underline{w}})}_{=\underline{0}}$$

Normal equation

$$= \underline{b}^* \underline{b} - \underline{b}^* B \hat{\underline{w}}$$

$$\underline{b} = \begin{bmatrix} \underline{0}_{M\times 1} \\ \underline{x} \end{bmatrix}$$

$$= \underline{x}^* \underline{x} - \underline{x}^* H \hat{\underline{w}}$$

$$B = \begin{bmatrix} \Delta^{1/2} U \\ \hline H \end{bmatrix}$$

Simplify further

$$= \underline{x}^* (\underline{x} - H\hat{\underline{w}})$$

$$= \underline{x}^* \left(\underline{x} - H\left(\Pi + H^* H\right)^{-1} H^* \underline{x}\right)$$

$$= \underline{x}^* \left(I - H\left(\Pi + H^* H\right)^{-1} H^*\right) \underline{x}$$

Matrix inversion lemma

$$= \underline{x}^* \left(I + H \Pi^{-1} H^*\right)^{-1} \underline{x}$$

# Weighted Regularized Least Squares

Criterion: $\min_{\underline{w}} \quad \underline{w}^* \Pi \underline{w} + (\underline{x} - H\underline{w})^* A (\underline{x} - H\underline{w})$

Given data $\underline{x}$, $H$

Your choice of $\Pi$, $A$ (Symmetric & positive definite)

Again, using similar change of variables ideas, we can solve this problem.

Normal Equation: $(\Pi + H^* A H)\, \hat{\underline{w}} = H^* A \underline{x}$

Unique Solution will always exist

$$\hat{\underline{w}} = (\Pi + H^* A H)^{-1} H^* A \underline{x}$$

Corresponding minimum cost

$$\xi = \underline{x}^* A (\underline{x} - H\hat{\underline{w}})$$

$$= \underline{x}^* \left[ A^{-1} + H \Pi^{-1} H^* \right]^{-1} \underline{x}$$

# Recursive Least Squares (RLS)

Given $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$ ($N \times 1$) $\quad H = \begin{bmatrix} h_{11} & \cdots & h_{1M} \\ \vdots & & \\ h_{N1} & \cdots & h_{NM} \end{bmatrix}$

$N \times M$ matrix

& Choice of $\Pi$ (symmetric & positive definite)

Regularized LS $\quad \min_{\underline{w}} \; \underline{w}^* \Pi \underline{w} + \| \underline{x} - H\underline{w} \|^2$

We know the optimal solution is

$$\hat{\underline{w}} = \left( \Pi + H^* H \right)^{-1} H^* \underline{x}$$

Since we have $N$ dimensional data

( $\underline{x}$ has $N$ entries

$H$ has $N$ rows ) we

denote this dimension explicitly & write

$$\hat{\underline{w}}_N = \left( \Pi + H_N^* H_N \right)^{-1} H_N^* \underline{x}_N$$

Say we have solved the LS problem with N dimensional data.

Suppose we add one more data point.

ie) Add one more entry $x_{N+1}$ to $\underline{x}$

$$\underline{x}_{N+1} = \begin{bmatrix} \left.\begin{matrix} x_1 \\ \vdots \\ x_N \end{matrix}\right\} \to \text{previous data} \\ x_{N+1} \to \text{New data point.} \end{bmatrix}$$

Add one more row to H

$$H_{N+1} = \begin{bmatrix} \left.\begin{matrix} h_{11} & \cdots & h_{1M} \\ \vdots & & \\ h_{N1} & \cdots & h_{NM} \end{matrix}\right\} \to \text{previous data} \\ h_{(N+1)1} \quad \cdots \quad h_{(N+1)M} \to \text{new data} \end{bmatrix}$$

$\underline{h}^*_{N+1}$

Now, we want to solve the problem

$$\min_{\underline{w}} \quad \underline{w}^* \Pi \underline{w} + \| \underline{x}_{N+1} - H_{N+1} \underline{w} \|^2$$

Note that size of $\underline{w}$ has not changed

We know $\hat{\underline{w}}_{N+1} = \underbrace{\left( \Pi + H^*_{N+1} H_{N+1} \right)^{-1}}_{M\times M \text{ inversion}} H^*_{N+1} \underline{x}_{N+1}$

Can we compute $\hat{\underline{w}}_{N+1}$ efficiently (without $M \times M$ matrix inversion)

using the fact that we have already computed $\hat{\underline{w}}_N$.

Yes, Recursive Least Squares (RLS) gives a method for this (using Matrix inversion lemma)

<u>Derivation of RLS.</u>

Define $$P_N = \left(\Pi + H_N^* H_N\right)^{-1}$$

$$P_{N+1} = \left(\Pi + H_{N+1}^* H_{N+1}\right)^{-1}$$

Now, $$\hat{\underline{w}}_N = P_N H_N^* \underline{x}_N$$

$$\hat{\underline{w}}_{N+1} = P_{N+1} H_{N+1}^* \underline{x}_{N+1}$$

$$H_{N+1} = \begin{bmatrix} H_N \\ \hline \underline{h}_{N+1}^* \end{bmatrix}$$

$\underline{h}_{N+1}^*$ is the last row of $H_{N+1}$

$$P_{N+1}^{-1} = \Pi + H_{N+1}^* H_{N+1}$$

$$= \Pi + \begin{bmatrix} H_N^* & \underline{h}_{N+1} \end{bmatrix} \begin{bmatrix} H_N \\ \hline \underline{h}_{N+1}^* \end{bmatrix}$$

$$= \underbrace{\Pi + H_N^* H_N}_{P_N^{-1}} + \underline{h}_{N+1} \cdot \underline{h}_{N+1}^*$$

$$P_{N+1}^{-1} = P_N^{-1} + \underline{h}_{N+1} \underline{h}_{N+1}^*$$

We need to compute $P_{N+1}$

$$P_{N+1} = \left(P_{N+1}^{-1}\right)^{-1} = \left(P_N^{-1} + \underline{h}_{N+1} \underline{h}_{N+1}^*\right)^{-1}$$

We can use matrix inversin lemma

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}$$

Take $A = P_N^{-1}$ $B = \underline{h}_{N+1}$ $C = 1$ $D = \underline{h}_{N+1}^*$

recursim for $P_N$ $\rightarrow$

$$P_{N+1} = P_N - \frac{P_N \underline{h}_{N+1} \underline{h}_{N+1}^* P_N}{\left(1 + \underline{h}_{N+1}^* P_N \underline{h}_{N+1}\right)}$$

5/11 The new LS estimate $\hat{\underline{w}}_{N+1}$ is obtained as

$$\hat{\underline{w}}_{N+1} = P_{N+1} H^*_{N+1} \underline{x}_{N+1}$$

$$= \left[ P_N - \frac{P_N \underline{h}_{N+1} \underline{h}^*_{N+1} P_N}{1 + \underline{h}^*_{N+1} P_N \underline{h}_{N+1}} \right] \left[ H_N^* \ \underline{h}_{N+1} \right] \times \begin{bmatrix} \underline{x}_N \\ x_{N+1} \end{bmatrix}$$

$$= \underbrace{P_N H_N^* \underline{x}_N}_{\hat{\underline{w}}_N} - \frac{P_N \underline{h}_{N+1}}{1 + \underline{h}^*_{N+1} P_N \underline{h}_{N+1}} \underline{h}^*_{N+1} \underbrace{P_N H_N^* \underline{x}_N}_{\hat{\underline{w}}_N}$$

$$- P_N \underline{h}_{N+1} \left( 1 - \frac{\underline{h}^*_{N+1} P_N \underline{h}_{N+1}}{1 + \underline{h}^*_{N+1} P_N \underline{h}_{N+1}} \right) x_{N+1}$$

$$\hat{\underline{w}}_{N+1} = \hat{\underline{w}}_N + \frac{P_N \underline{h}_{N+1}}{1 + \underline{h}^*_{N+1} P_N \underline{h}_{N+1}} \left( x_{N+1} - \underline{h}^*_{N+1} \hat{\underline{w}}_N \right)$$

$\hookrightarrow$ recursion for LS solution

For convenience, we define two terms

Conversion factor $\gamma(N+1) = \dfrac{1}{1 + \underline{h}^*_{N+1} P_N \underline{h}_{N+1}}$

Gain factor $\underline{g}(N+1) = P_N \underline{h}_{N+1} \gamma(N+1)$

We can verify that

$$\gamma(N+1) = 1 - \underline{h}^*_{N+1} P_N \underline{h}_{N+1}$$

and $\underline{g}(N+1) = P_N \underline{h}_{N+1}$

x ——— x

## Summary of RLS

Consider regularized LS problem

$$\min_{\underline{w}} \quad \underline{w}^* \Pi \underline{w} + \| \underline{x} - H \underline{w} \|^2$$

The optimal solution can be recursively computed as follows.

Initialize $\hat{\underline{w}}_{-1} = \underline{0}$ $\quad P_{-1} = \Pi^{-1}$

For $i \geq 0$, compute

$$\gamma(i) = \frac{1}{\underline{h}_i^* P_{i-1} \underline{h}_i}$$

where $\underline{h}_i^*$ is $i^{th}$ row of $H$ matrix

and $\underline{g}(i) = P_{i-1} \underline{h}_i \gamma(i)$

Optimal reg. LS solution
(with $i$ data samples)

$$\underline{\hat{\omega}}_i = \underline{\hat{\omega}}_{i-1} + \underline{g}_i \left[ x_i - \underline{h}_i^* \underline{\hat{\omega}}_{i-1} \right]$$

$x_i \rightarrow i^{th}$ coordinate of $\underline{x}$

$$P_i = P_{i-1} - \frac{\underline{g}(i)\, \underline{g}^*(i)}{\gamma(i)}$$

$P_i$ has interpretation $P_i = \left( \Pi + H_i^* H_i \right)^{-1}$

$H_i \rightarrow$ matrix with first $i$ rows of $H$.

x ———————— x

## Remarks:

**Regularization** The presence of of regularization term (using positive definite $\Pi$), we are guaranteed to have unique solution.

$$\left(\Pi + H_N^* H_N\right)^{-1} \text{ exists for any } N \text{ rows of } H$$

Without regularization; the matrix $H_N^* H_N$ may not be invertible

1) If $N < M$ then $H_N^* H_N$ is not invertible.

2) If $N \geq M$ then $H_N^* H_N$ may or may not be invertible

**Conversion factor**

Define two error quantities

a priori estimation error

$$e_a(N+1) = x_{N+1} - \underline{h}_{N+1}^* \underline{\hat{w}}_N$$

last entry in $\underline{x}_{N+1} - H_{N+1} \underline{\hat{w}}_N$

$\underline{\hat{w}}_N \rightarrow$ LS estimate using N data samples

aposteriori estimation error

$$e_p(N+1) = x_{N+1} - \underline{h}^*_{N+1} \underline{\hat{w}}_{N+1}$$

$\downarrow$

error after updating LS solution using new data sample

We have

$$e_p(N+1) = x_{N+1} - \underline{h}^*_{N+1} \underline{\hat{w}}_{N+1}$$

$$= x_{N+1} - \underline{h}^*_{N+1} \left( \underline{\hat{w}}_N + \underline{g}(N+1) \underbrace{\left[ x_{N+1} - \underline{h}^*_{N+1} \underline{\hat{w}}_N \right]}_{e_a(N+1)} \right)$$

$$= \underbrace{x_{N+1} - \underline{h}^*_{N+1} \underline{\hat{w}}_N}_{e_a(N+1)} - \underline{h}^*_{N+1} \underline{g}(N+1)\, e_a(N+1)$$

$$= \left[ 1 - \underline{h}^*_{N+1} \underline{g}(N+1) \right] e_a(N+1)$$

$$\boxed{e_p(N+1) = \gamma(N+1)\, e_a(N+1)}$$

From the definition, $\gamma(N+1) = \dfrac{1}{1 + \underbrace{\underline{h}^*_{N+1} P_N \underline{h}_{N+1}}_{\geq 0}}$

$$0 < \gamma(N+1) \le 1 .$$

We always have

$$|e_p(N)| \le |e_a(N)| \quad \text{for any } N.$$

x ———— x

Recursion for the minimum cost

$$\xi = \underline{x}^* \underline{x} - \underline{x}^* H \hat{\underline{w}}$$

(from before)

$$= \underline{x}^* (\underline{x} - H \hat{\underline{w}})$$

With N observations

$$\xi(N) = \underline{x}_N^* (\underline{x}_N - H_N \hat{\underline{w}}_N)$$

With the additional $(N+1)^{th}$ observation

$$\xi(N+1) = \underline{x}_{N+1}^* (\underline{x}_{N+1} - H_{N+1} \hat{\underline{w}}_{N+1})$$

$$\hat{\underline{w}}_{N+1} = \hat{\underline{w}}_N + \underline{g}(N+1) \underbrace{[x_{N+1} - \underline{h}_{N+1}^* \hat{\underline{w}}_N]}_{e_a(N+1)}$$

$$\xi(N+1) = \begin{bmatrix} \underline{x}_N \\ x_{N+1} \end{bmatrix}^{*} \left( \begin{bmatrix} \underline{x}_N \\ x_{N+1} \end{bmatrix} - \begin{bmatrix} H_N \\ \underline{h}^{*}_{N+1} \end{bmatrix} \times \left( \underline{\hat{\omega}}_N + \underline{g}(N+1)\, e_a(N+1) \right) \right)$$

$$= \begin{bmatrix} \underline{x}^{*}_N & x^{*}_{N+1} \end{bmatrix} \begin{bmatrix} \underline{x}_N - H_N \left( \underline{\hat{\omega}}_N + \underline{g}(N+1)\, e_a(N+1) \right) \\ x_{N+1} - \underline{h}^{*}_{N+1} \left( \underline{\hat{\omega}}_N + \underline{g}(N+1)\, e_a(N+1) \right) \end{bmatrix}$$

$$= \underbrace{\underline{x}^{*}_N \left( \underline{x}_N - H_N \underline{\hat{\omega}}_N \right)}_{\xi(N)} - \underline{x}_N H_N \underline{g}(N+1)\, e_a(N+1)$$

$$+ \; \underline{x}^{*}_{N+1} \left( e_a(N+1) \right) \left( 1 - \underline{h}^{*}_{N+1} \underline{g}(N+1) \right)$$

after some algebra $\;= \; \xi(N) + |e_a(N+1)|^2 \, \gamma(N+1)$

$$\xi(N+1) = \xi(N) + \gamma(N+1)\, |e_a(N+1)|^2$$

We had $e_p(N+1) = x_{N+1} - \underline{\vec{h}}_{N+1} \underline{\hat{\omega}}_{N+1}$

$$e_p(N+1) = e_a(N+1)\,\gamma(N+1)$$

So $|e_a(N+1)|^2 \gamma(N+1) = e_a^*(N+1)\, e_p(N+1)$

$$= \frac{|e_p(N+1)|^2}{\gamma(N+1)}$$

x ———————— x

## Exponentially weighted RLS

Choose a weighting factor $0 << \lambda < 1$
($\lambda$ is close to 1)

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \quad H = \begin{bmatrix} h_{11} & \cdots & h_{1M} \\ \vdots & & \\ h_{N1} & \cdots & h_{NM} \end{bmatrix} \quad \underline{\omega} = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_M \end{bmatrix}$$

$$\underline{x} - H\underline{\omega} = \underline{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

Exponentially weighted cost

$$\min_{\underline{w}} \quad \lambda^N \underline{w}^* \Pi \underline{w} + \sum_{i=1}^{N} \lambda^{(N+1-i)} |d_i|^2$$

Equivalently

$$\min_{\underline{w}} \quad \lambda^N \underline{w}^* \Pi \underline{w} + (\underline{x} - H\underline{w})^* \overbrace{\begin{bmatrix} \lambda^N & & 0 \\ & \ddots & \\ 0 & & \lambda^2 \\ & & & \lambda \end{bmatrix}}^{\Lambda_N} (\underline{x} - H\underline{w})$$

$$\min_{\underline{w}} \quad \lambda^N \underline{w}^* \Pi \underline{w} + (\underline{x} - H\underline{w})^* \Lambda_N (\underline{x} - H\underline{w})$$

Optimal solution from normal equations

$$\left( \lambda^N \Pi + H_N^* \Lambda_N H_N \right) \hat{\underline{w}}_N = H_N^* \Lambda_N \underline{x}_N$$

Let $P_N = \left( \lambda^N \Pi + H_N^* \Lambda_N H_N \right)^{-1}$

Recursive way of updating $\hat{\underline{w}}_N$

Initialization $\underline{w}_0 = \underline{0}$ $\quad P_0 = \Pi^{-1}$

$\mathcal{E}(0) = 0$

For $i \geq 1$

$$\gamma(i) = \frac{1}{\left(1 + \underline{h}_i^* P_{i-1} \underline{h}_i\right)}$$

$\underline{h}_i^* \rightarrow$ ith row of H

$$\underline{g}(i) = \lambda^{-1} P_{i-1} \underline{h}_i \gamma(i)$$

$$e_a(i) = x_i - \underline{h}_i^* \hat{\underline{w}}_{i-1}$$

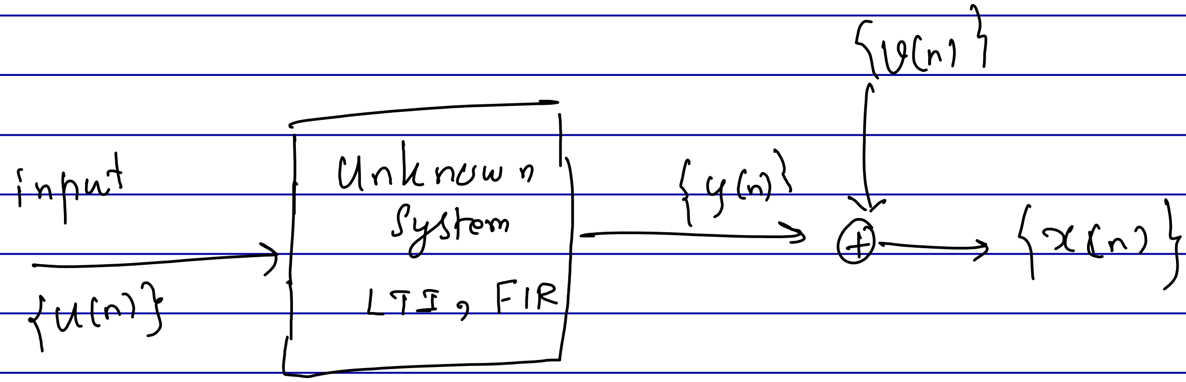
$$\longrightarrow \hat{\underline{w}}_i = \hat{\underline{w}}_{i-1} + \underline{g}_i e_a(i)$$

Optimal weighted RLS solution with i data samples

$$P_i = \lambda^{-1} P_{i-1} - \frac{\underline{g}_i \underline{g}_i^*}{\gamma(i)}$$

$$\xi(i) = \xi(i-1) + \gamma(i) |e_a(i)|^2$$

## Least Squares in System identification




$$y(n) = \sum_{\ell=0}^{L-1} \alpha_\ell \, u(n-\ell)$$

$\alpha_0, \alpha_1, \ldots, \alpha_{L-1}$ are unknown

Collect the measurements (N+1 samples)

$$\underbrace{\begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N) \end{bmatrix}}_{\underline{x}} = \underbrace{\begin{bmatrix} u(0) & u(-1) & \cdots & u(-L+1) \\ & & & \\ u(N) & u(N-1) & \cdots & u(N-L+1) \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_L \end{bmatrix}}_{\underline{\alpha}} + \underbrace{\begin{bmatrix} v(0) \\ \vdots \\ v(N) \end{bmatrix}}_{\underline{v}}$$

$$\underline{x} = H \underline{\alpha} + \underline{v}$$

Find $\min_{\underline{\alpha}} \|\underline{x} - H\underline{\alpha}\|^2$

$$\hat{\underline{\alpha}}_{LS} = (H^* H)^{-1} H^* \underline{x}$$

Assuming $H$ is full column rank.