Order-Recursive Least Squares

<u>Question</u>: When is $\underline{\varepsilon}$ more than just $\underline{w}$?

True signal model of the form $a + bn + cn^2$ but we are fitting only $\hat{a} + \hat{b}n$.

In a given class of models, we can plot the error $J(\hat{\theta})$ vs. the number of parameters in the model to get an idea of the appropriate number of parameters.

<u>Note 1</u>: LS error will always decrease with noisy data even if we increase the # of parameters beyond the appropriate #.

Think of the geometric interpretation of LS. By adding a new column to H (basis vector) the signal model can always get closer to $\underline{x}$. In effect, what is happening is the model is starting to capture some of the noise. <u>This is known as overfitting</u>

<u>Note 2</u>: order-recursive LS is a method for incrementally updating $\hat{\theta}$ without having to perform $\hat{\theta}_p = (H_p^T H_p)^{-1} H_p^T x$ every time we add a column to H.

In fact, if H forms an orthonormal basis (as in Fourier analysis)

$$\hat{\theta}_p = H^T x = \begin{bmatrix} h_1^T \\ h_2^T \\ \vdots \\ h_p^T \end{bmatrix} \begin{bmatrix} x[0] \\ \vdots \\ x[p] \end{bmatrix}$$

so when we add h_{p+1} to H all we need to do is

$$\hat{\theta}_{p+1} = \begin{bmatrix} \hat{\theta}_p \\ \underline{h}_{p+1}^T \underline{x} \end{bmatrix}$$

This part doesn't change due to orthonormality of H.

SEQUENTIAL L.S.

Data set increasing with time. For $x[0], \ldots, x[N-1]$ we have $\hat{\theta}$, now we get $x[N]$. How does $\hat{\theta}$ change? Do we have to compute it from scratch?

Example: $x[n] = A + w[n]$ D.C. level

Define $\hat{A}[N]$ to be the L.S. estimator for A from the dataset $x[0], \ldots, x[N]$

We know that:

$$\hat{A}[N-1] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad ; \text{ and}$$

$$\hat{A}[N] = \frac{1}{N+1} \sum_{n=0}^{N} x[n]$$

We can rewrite $\hat{A}[N]$ as

$$\hat{A}[N] = \frac{1}{N+1} \left[\underbrace{\left(\sum_{n=0}^{N-1} x[n] \right)}_{N\hat{A}[N-1]} + x[N] \right]$$

$$(**) \qquad = \frac{N}{N+1} \hat{A}[N-1] + \frac{1}{N+1} x[N]$$

This is a weighted average. The weight for $x[N]$ goes to 0 as N becomes large, whereas the weight for $\hat{A}[N-1]$ goes to 1. Why?

Again rearranging the terms

$$(*) \quad \hat{A}[N] = \hat{A}[N-1] + \underbrace{\frac{1}{N+1}}_{\text{correction weight.}} \underbrace{\left(x[N] - \hat{A}[N-1] \right)}_{\text{Correction to estimate}}$$

How about $\underbrace{J_{min}}_{J(\hat{\theta})}$ as a function of N?

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

$$= \left(\sum_{n=0}^{N-1} \left(x[n] - \hat{A}[N]\right)^2\right) + \left(x[N] - \hat{A}[N]\right)^2$$

$$= \sum_{n=0}^{N-1} \Big(x[n] - \underbrace{\hat{A}[N-1] - \frac{1}{N+1}\left(x[N] - \hat{A}[N-1]\right)}_{\text{from } *}\Big)^2$$

$$+ \left(x[N] - \hat{A}[N]\right)^2$$

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

$$- \frac{2}{N+1}\cancel{\left[\sum_{n=0}^{N-1}\left(x[n] - \hat{A}[N-1]\right)\right]}_{0}\left(x[N] - \hat{A}[N-1]\right)$$

$$+ \frac{N}{(N+1)^2}\left(x[N] - \hat{A}[N-1]\right)^2$$

$$+ \left(x[N] - \hat{A}[N]\right)^2 \longrightarrow \left(x[N] - \frac{N}{N+1}\hat{A}[N-1] - \frac{1}{N+1}x[N]\right)^2$$

from $**$

$$= \left(\frac{N}{N+1}x[N] - \frac{N}{N+1}\hat{A}[N-1]\right)^2$$

$$= \frac{N^2}{(N+1)^2}\left(x[N] - \hat{A}[N-1]\right)^2$$

$$= J_{min}[N-1]$$

$$+ \frac{N}{(N+1)^2}(1+N)\left(x[N] - \hat{A}[N-1]\right)^2$$

$$= J_{min}[N-1]$$

$$+ \frac{N}{N+1}\left(x[N] - \hat{A}[N-1]\right)^2$$

error can not decrease with increasing N. Why not?

A generalization of the previous result:

Let data point n have variance σ_n^2, then a weighted least squares estimator is

$$\hat{A}[N-1] = \frac{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} x[n]}{\sum_{n=0}^{N-1} 1/\sigma_n^2}$$

$$① \quad \hat{A}[N] = \hat{A}[N-1] + \underbrace{\frac{1/\sigma_N^2}{\sum_{n=0}^{N} 1/\sigma_n^2}}_{K[N]} \left(x[n] - \hat{A}[N-1] \right)$$

$K[N]$: gain factor for N'th correction

$0 \le K[N] \le 1$
$\hookrightarrow \sigma_N^2 \to 0$

if all $\sigma_n^2 = \sigma^2$ then $K[N] = 1/N+1$ as before

*$K[N]$ can be rewritten as

$$② \quad K[N] = \frac{1/\sigma_N^2}{1/\sigma_N^2 + 1/\mathrm{var}(\hat{A}[N-1])}$$

$$= \frac{\mathrm{var}(\hat{A}[N-1])}{\mathrm{var}(\hat{A}[N-1]) + \sigma_N^2}$$

from variance of sum of random vars

Equations ② & ① define a recursion. Start with $\hat{A}[0] = x[0]$ and $\mathrm{var}(\hat{A}[0]) = \sigma_0^2$

Vector case

C : data covariance . Diagonal with $\sigma_0^2, \sigma_1^2, \dots \sigma_n^2$

Let $H[n] = \begin{bmatrix} H[n-1] \\ h^T[n] \end{bmatrix}$ Adding a row for data point n.

$$(*) \quad \hat{\theta}[n] = \left(H^T[n] C^{-1}[n] H[n]\right)^{-1} H^T[n] C^{-1}[n] \underline{x}[n]$$

where $\underline{x}[n] = \begin{bmatrix} x[0] \\ \vdots \\ x[n] \end{bmatrix}$

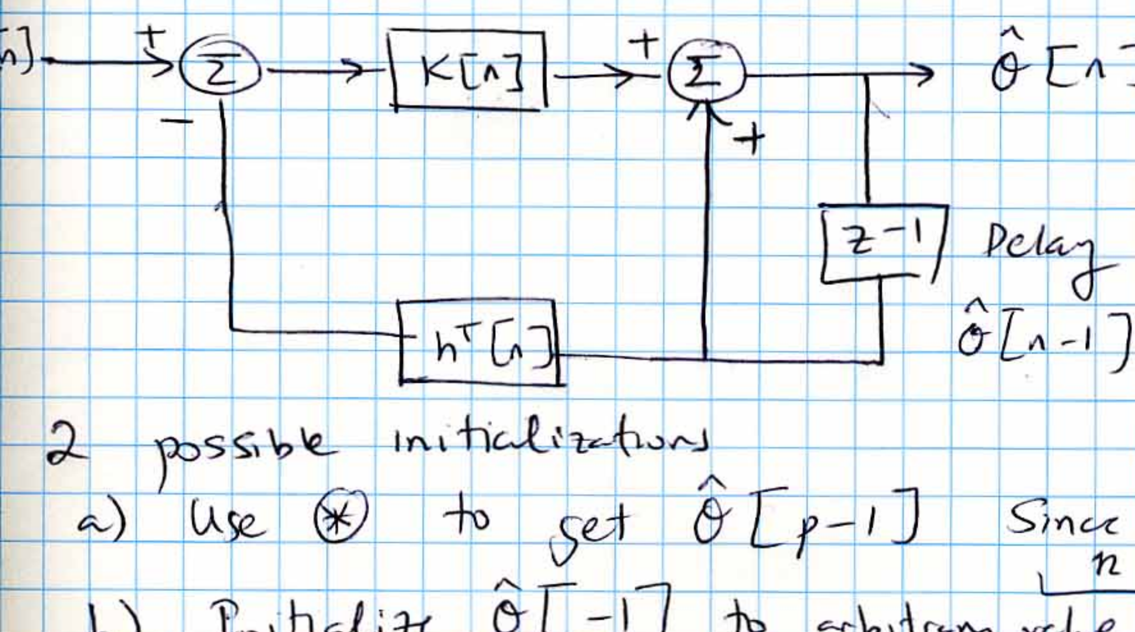
This can be written in recursive way as well

$$(1) \quad \hat{\theta}[n] = \hat{\theta}[n-1] + K[n]\left(x[n] - h^T[n]\hat{\theta}[n-1]\right)$$

$\hookrightarrow$ The data point $x[n]$. Not $\underline{x}[n]$

$$\text{where } K[n] = \frac{\Sigma[n-1]\, h[n]}{\sigma_n^2 + h^T[n] \underbrace{\Sigma[n-1]}_{\substack{\text{covariance} \\ \text{matrix for } \hat{\theta}[n-1]}} h[n]}$$

$$(2) \quad \Sigma[n] = \left(I - K[n] h^T[n]\right)\Sigma[n-1]$$



2 possible initializations

a) Use $(*)$ to get $\hat{\theta}[p-1]$ Since we need $\underbrace{n \geq p}$ for inversion

b) Initialize $\hat{\theta}[-1]$ to arbitrary value (0)

$\Sigma[-1] = \alpha I$ where α is large (low confidence)