Least squares Review

$\underline{y}$ = observational vector

$\langle n^2 \rangle$ = variance of ~~as~~ error/noise in the observations (e.g. the error of the instrument or any other thing you want to consider noise - see keeling curve example)

$\underline{\underline{E}}$ model or hypothesis to check. (Matrix)

$\underline{x}$ vector of model parameters

STEP 1:

$$\underset{\text{known}}{\underline{y}} = \underset{\text{hypothesis}}{\underline{\underline{E}}}\ \overset{\text{unknowns}}{\underline{x}} + \underline{n}$$

unknown but we assume we know its statistics

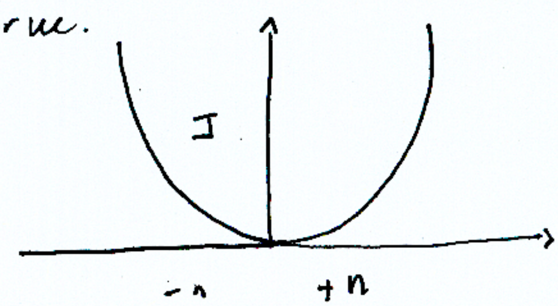
$\underline{\underline{R}}_{nn} = \langle \underline{n}\,\underline{n}^T \rangle$

$\langle \underline{n} \rangle = \emptyset$

STEP 2: define a cost function

$$J = \underline{n}^T \underline{n} = (\underline{y} - \underline{\underline{E}}\,\underline{x})^T (\underline{y} - \underline{\underline{E}}\,\underline{x})$$

this is a quadratic form. Finding the extremum of J is equivalent to finding the MINIMUM



$$\frac{\partial J}{\partial \underline{x}} = 0 \implies \underline{\underline{E}}^T \underline{\underline{E}}\,\underline{x} - \underline{\underline{E}}^T \underline{y} = \emptyset$$

"Normal Equation"

$$\hat{\underline{x}} = (\underline{\underline{E}}^T \underline{\underline{E}})^{-1} \underline{\underline{E}}^T \underline{y}$$

STEP 3: compute the uncertainty in our estimate of $\underline{x}$

$$\langle (\underline{x} - \hat{\underline{x}})(\underline{x} - \hat{\underline{x}})^T \rangle = \underline{\underline{P}}$$

Model parameter uncertainty.

If ~~the~~ we set $\underline{y}_0 = \underline{\underline{E}}\,\underline{x}$ true observed value with no noise (hypothetical)

$$\underline{\underline{P}} = \langle [(\underline{\underline{E}}^T \underline{\underline{E}})^{-1} \underline{\underline{E}}^T (\underbrace{\underline{y}_0 - \underline{y}}_{\underline{n}})][\qquad]^T \rangle$$

$$\underline{\underline{P}} = (\underline{\underline{E}}^T \underline{\underline{E}})^{-1} \underline{\underline{E}}^T \underbrace{\langle \underline{n}\,\underline{n}^T \rangle}_{\underline{\underline{R}}_{nn}} \underline{\underline{E}} (\underline{\underline{E}}^T \underline{\underline{E}})^{-1}$$

STEP 4: Inspection of the estimated error

$$\hat{\underline{n}} = \underline{y} - \underline{\underline{E}}\,\hat{\underline{x}}$$

Formal Testing of model error through "chi-square" test.

$$J = \sum_{m=1}^{M} n_m^2$$

if we assume P_n is gaussian and that each $n_m n_n \sim \delta_{mn}$ independent that

P_J = "chi-square" distribution

This implies that

$$\langle J \rangle = \int_{-\infty}^{\infty} J \, P_J(J) \, dJ = \underbrace{M - N}$$

degrees of freedom

M = number of equation (rows of $\underline{\underline{E}}$)

N = number of model parameters (length of $\underline{x}$)

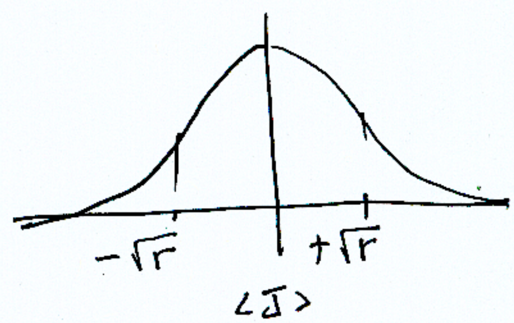
$$\langle J^2 \rangle = 2(M - N)$$

so we can check if our value of J

$$\langle J \rangle - \sqrt{\langle J^2 \rangle} > J > \langle J \rangle + \sqrt{\langle J^2 \rangle}$$

in other words if $r = M - N$

$$J = r \pm \sqrt{r}$$



NOTE if $\underline{\underline{R}}_{nn} = \sigma_n^2 \underline{\underline{I}}$ and $\sigma_n^2 \neq 1$

than the evaluation of J for the chi-square

$$J = \underline{n}^T \underline{\underline{R}}_{nn}^{-1} \underline{n}^*$$

Alternatively if $\underline{\underline{R}}_{nn}$ is diagonal with all elements $= \sigma_n^2$ you can post multiply $J \sigma_n^2$ and verify against chi-square.

other test exist to check if the residuals $\hat{\underline{n}}$ are random and have no additional structure.

—— summary —— and alternative formulations of LSQ

① $J = \underline{n}^T \underline{n}$; $\hat{x} = \underbrace{(\underline{\underline{E}}^T \underline{\underline{E}})^{-1} \underline{\underline{E}}^T}_{\underline{\underline{G}}} \underline{y}$ $\underline{\underline{C}}_{xx} = \underline{\underline{G}} \, \underline{\underline{R}}_{nn} \underline{\underline{G}}^T$

② $J = \underline{n}^T \underline{\underline{W}} \, \underline{n}$; $\hat{x} = \underbrace{(\underline{\underline{E}}^T \underline{\underline{W}} \, \underline{\underline{E}})^{-1} \underline{\underline{E}}^T \underline{\underline{W}}}_{\underline{\underline{G}}_2} \underline{y}$ $\underline{\underline{C}}_{xx} = \underline{\underline{G}}_2 \underline{\underline{R}}_{nn} \underline{\underline{G}}^T$

③ $J = \underline{n}^T \underline{\underline{W}} \, \underline{n} + \underline{x}^T \underline{\underline{S}} \, \underline{x}$; $\hat{x} = \underbrace{(\underline{\underline{E}}^T \underline{\underline{W}} \, \underline{\underline{E}} + \underline{\underline{S}})^{-1} \underline{\underline{E}}^T \underline{\underline{W}}}_{\underline{\underline{G}}_3} \underline{y}$

$\underline{\underline{C}}_{xx} = \underline{\underline{G}}_3 \underline{\underline{R}}_{nn} \underline{\underline{G}}^T$

CASE $\underline{\underline{1}}$ is the most simple one.

CASE $\underline{\underline{2}}$ the weight $\underline{\underline{W}}$ can be used to give different "importance" to the errors. Assume you wanted to disregard datum y_{10} in your fit

$$\underline{\underline{W}} = \begin{bmatrix} 1_1 & & & & & & & \\ & 1 & & & & & 0 & \\ & & 1 & & & & & \\ & & & 1 & & & & \\ & & & & 0.0001 & & & \\ & 0 & & & & 1 & & \\ & & & & & & 1 & \\ & & & & & & & 1 \end{bmatrix}$$

$W_{10,10}$

The means that

$$J = n_1 w_{11} n_1 + \dots \; n_{10} w_{10,10} n_{10} \dots \; n_M w_{MM} n_M)$$

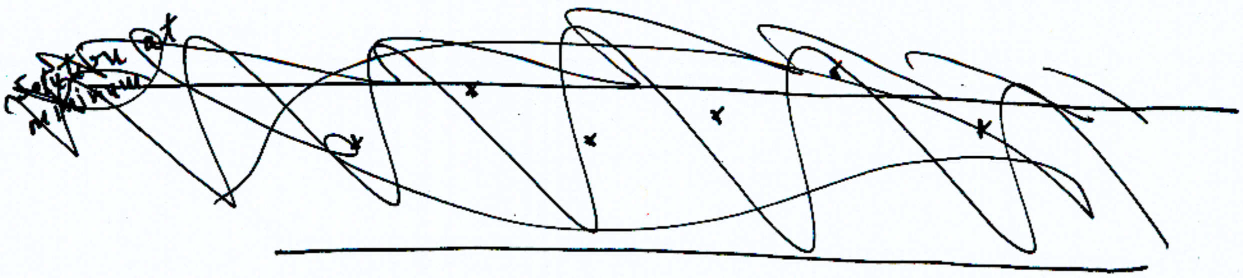
this term is so small $y_{10} = \underline{\underline{E}} \, \underline{x} + n_{10}$

no matter what $\hat{n}_{10} = \cancel{Ex}\, y_{10} - E x$ is

$\hat{n}_{10} W_{10,10} \hat{n}_{10}$ is small

J

$\hat{n}_{10}$ is already very small so the fitting does not care.



CASE 3

If you ~~know~~ have some apriori information on C_{xx} you can impose this statistical constraint

by $S = C_{xx}^{-1}$

$$J = n W n^T + x C_{xx}^{-1} x$$

Let us look at CO_2 and SST examples. These show how this machinery can be used to fit functions to data, since it is obvious that neither CO_2 and SST can be understood by an algebraic model

in other words the dynamics cannot be fully captured by ...