Least Squares Optimization

eg. 1) $y(x,t) = x_1 + x_2 \exp(-x_3 t)$ N.L

$t \rightarrow$ meas. time

$x \rightarrow$ model params

$y(x,t) = x_1 + x_2 t + x_3 t^2$ L.

Problem:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left[\sum_{j=1}^{m} \underbrace{\left| y_j - y(x, t_j) \right|^2} \right]$$

$m \rightarrow$ total no of meas.

Least squares optimization.

$f(x)$ = Cost function (c.f) / Objective fn

$y_j - y(x,t_j) \rightarrow$ residual $\rightarrow r_j(x)$

residual vector: $r(x) = [r_1(x) \ \cdots \ r_m(x)]^T$

$$f(x) = \frac{1}{2} \sum_{j=1}^{m} [r_j(x)]^2 = \frac{1}{2} \| r(x) \|^2$$

$\boxed{m \geq n \rightarrow \text{realistic case}}$

$m < n \rightarrow \infty$ solutions

[Compressive Sensing.]

$r(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$

$f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$

$$f(x) = \frac{1}{2} \sum_{j=1}^{m} (r_j(x))^2$$

$$f \rightarrow \nabla f \rightarrow \begin{bmatrix} \partial f / \partial x_1 \\ \vdots \\ \partial f / \partial x_n \end{bmatrix}$$

Jacobian of $r(x)$?

$$J_{ji} = \left(\frac{\partial r_j}{\partial x_i} \right) \rightarrow \begin{bmatrix} (\nabla r_1)^T \longleftrightarrow \\ (\nabla r_2)^T \longleftrightarrow \\ \vdots \end{bmatrix} \rightarrow \left[\partial r_1 / \partial x_1 \ \ \partial r_1 / \partial x_2 \ \cdots \ \partial r_1 / \partial x_n \right]$$

$$\nabla f \rightarrow \frac{\partial f}{\partial x_i} = \sum_{j=1}^{m} r_j(x) \frac{\partial r_j(x)}{\partial x_i} = \sum_{j=1}^{m} \frac{\partial r_j}{\partial x_i} r_j = J^T(x)\, r(x)$$

Hessian $H_{ik} = \frac{\partial^2 f}{\partial x_i \partial x_k}$ defn.

$$= \sum_j \left[\frac{\partial r_j(x)}{\partial x_k} \frac{\partial r_j(x)}{\partial x_i} + r_j(x) \frac{\partial^2 r_j(x)}{\partial x_k \partial x_i} \right]$$

$$\nabla^2 f = J^T(x) J(x) + \boxed{\sum r_j(x) \nabla^2 r_j(x)}$$

When can we ignore

① r_j is small (e.g. when close to a soln)

② $\nabla^2 r_j$ is small (if r is approx affine)

When I ignore them: $\nabla^2 f = J^T J$ — Linear least squares.

$r_j(x) \rightarrow$ Affine function

$\Rightarrow \quad r(x) = Ax - b$

$$\Rightarrow f(x) = \frac{1}{2}\|r(x)\|^2 = \frac{1}{2}\|Ax - b\|^2$$

$$\nabla f = A^T(Ax - b) = A^T r(x) \quad \Rightarrow J = A$$

$$\nabla^2 f = J^T J = A^T A.$$ Hessian must be P.D.

P.D. $\quad z^T \nabla^2 f z > 0, \ \forall z \neq 0.$

$z^T A^T A z > 0$

$\|Az\|^2 > 0$

$$Az = \begin{bmatrix} | & | & | & | & | \\ | & | & | & | & | \end{bmatrix}\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$

A must have full col. rank for $\nabla^2 f$ to be P.D.

f is a convex fn. ∴ if we find x^*, s.t. $\nabla f(x^*) = 0$.

then x^* is the soln.

$$\nabla f = A^T(Ax - b) \rightarrow \boxed{A^T A x = A^T b}$$ Normal eqns.

Solving using the SVD.

$$A = U S V^T$$

A: $m \times n$, U: $m \times m$, S: $m \times n$, V^T: $n \times n$, $\sigma_i \geq 0$

$m > n$: $\begin{bmatrix} \ddots \\ 0 \end{bmatrix}$ $\quad$ $m < n$: $\begin{bmatrix} \ddots & 0 \end{bmatrix}$

$$A^T A = V S^T U^T U S V^T = V \underbrace{S^T S}_{n \times n} V^T$$

$$A^T A x = A^T b \rightarrow V S^T S V^T x = V S^T U^T b$$ L.M by V^T

$$S^T S V^T x = S^T U^T b$$

Invertible? $\quad \sigma_i^2, \ i = 1 \ldots n.$ $\quad$ if $m \geq n$ AND full col rank then $\sigma_n \neq 0$. in this case $S^T S$ invertible.

$$V^T x = (S^T S)^{-1} S^T U^T b$$

$[\diagdown\ 0]$

$$x = V\underbrace{(S^TS)^{-1}S^T}U^Tb.$$

$$x = V D^{-1} U_1^T b$$

U_1: first n cols of U

$$= \begin{array}{c} v_1 \quad \cdot \quad v_n \\ \begin{bmatrix} | & | & | & | \end{bmatrix} \end{array} \begin{bmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_n \end{bmatrix} \begin{bmatrix} u_1 \text{---} \\ \cdots \\ u_n \text{---} \end{bmatrix} \begin{bmatrix} \uparrow \\ b \\ \downarrow \end{bmatrix}$$

$$x = \sum_{i=1}^{n} \left(\frac{1}{\sigma_i} u_i^T b\right) \times v_i$$

Soln least squares problem.

A has full col rank

$$x^* = \sum_{i=1}^{n-1} \left(\frac{1}{\sigma_i} u_i^T b\right) v_i + \underbrace{\left(\frac{1}{\sigma_n} u_n^T b\right) v_n}$$

pseudo inverse

noise

$k = 10^4$, $\sigma_1 = 1$, $\Rightarrow \sigma_n = 10^{-4}$, $b = b_0 + \Delta$

Reduce error by truncating SVD

$$x^* \approx \underbrace{\sum_{i=1}^{r} \left(\frac{1}{\sigma_i} u_i^T b\right) v_i} \leftarrow v_n$$

$$= \frac{1}{\sigma_i} v_i (u_i^T b) \rightarrow \frac{1}{\sigma_i} v_i u_i^T b$$

$$x = \sum_{i=1}^{r} \frac{1}{\sigma_i} (v_i u_i^T)\, b$$