Question 2

$\mathcal{E}(\omega) \Rightarrow$ Error f^n

$$E(\omega) \approx E(\tilde{\omega}) + (\omega - \tilde{\omega})^T \nabla E \Big|_{\omega = \tilde{\omega}} + \frac{1}{2}(\omega - \tilde{\omega})^T H (\omega - \tilde{\omega})$$

$H \Rightarrow$ Hessian

To show # of independent element in eq equadratic ~~eror~~ f^n is $\omega(\omega+3)/2$

Suppose : objective f^n $f: R^n \to R$

gradient vector $\nabla f(x) = \left[\frac{\partial}{\partial x} f(x)\right]^T \in R^n \to$ vector

Hessian

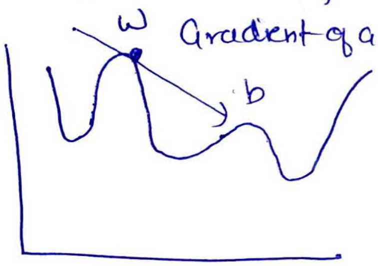
this works only if Hessian is Positive definite is convex

$\Rightarrow$ Matrix of 2^n derivative

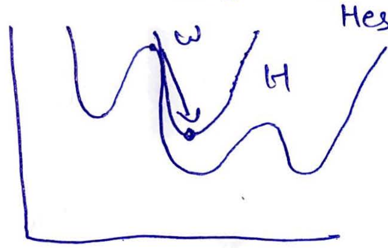
$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f x & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x_1) \\ \vdots & & \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x) & \cdots & \frac{\partial 2}{\partial x_n \partial x_n} f(x) \end{bmatrix} \in R^n$$

~~Since~~ In this ~~is~~ ~~a~~ vanilla version of Newton method there is no learning rate

First order optimizatio



using lr it minimize approx



this use gradient & Hessian to for quadratic approx. & minima to approx

& it has full step jump directly to min of local squared approx.

In Hessian matrix we have $(W = N)$ parameter so size of Hessian is $H_{ww} \cdot (H_{nn})$.

$b =$ Gradient vector of $\mathcal{E}$ at $\hat{\omega}$ $\quad b \equiv \nabla E \Big|_{\omega = \tilde{\omega}}$

$H = \nabla \nabla E$

$$H_{ij} = \frac{\partial \epsilon}{\partial \omega_i \partial \omega_j}\Big|_{\omega = \hat{\omega}}$$

local approx ~~of~~ ~~gra~~ of Error f^n

$$\nabla E_{(w-\tilde{w})^T} \simeq b + H(w-\hat{w})$$

as w is close to $\tilde{w}$, approx for error & gradient

Total independent parameter

$b + H(w-\tilde{w})$, $b \to$ vector of size W &

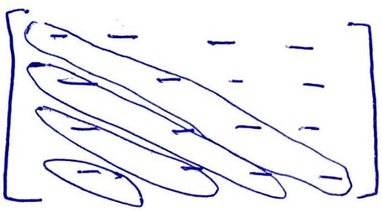
H is a symmetric Hessian matrix of size W, W.

total independent parameter in $b = |b| = W$

" " " " H $= W + \boxed{\frac{W(W-1)}{2}}$

$\downarrow$ diagonal

lower/upper Δ matrix as it is symmetric



Total element $= W + W + \frac{W(W-1)}{2}$

$$= 2W + \frac{W^2}{2} - \frac{W}{2}$$

$$= \frac{4W - W}{2} + \frac{W^2}{2}$$

$$= \frac{3W}{2} + \frac{W^2}{2} = \frac{W(W+3)}{2}$$