L4

Q(8)

$$\frac{\partial \mathcal{L}}{\partial w_i} = \delta \cdot x_i$$

where x_i is input recvd ~~by~~ from a prev. layer that utilized activation fn.

$\Rightarrow x_i > 0 \Rightarrow \frac{\partial \mathcal{L}}{\partial w_i}$ depends on sign of δ.

$\Rightarrow$ Dirn of "step" $= \left(\frac{\partial \mathcal{L}}{\partial w_1}, \frac{\partial \mathcal{L}}{\partial w_2}, \text{---}, \frac{\partial \mathcal{L}}{\partial w_n}\right)$

$$= (\delta \cdot x_1, \delta \cdot x_2, \text{---}, \delta \cdot x_n)$$

$$= \delta(x_1, x_2, \text{---}, x_n)$$

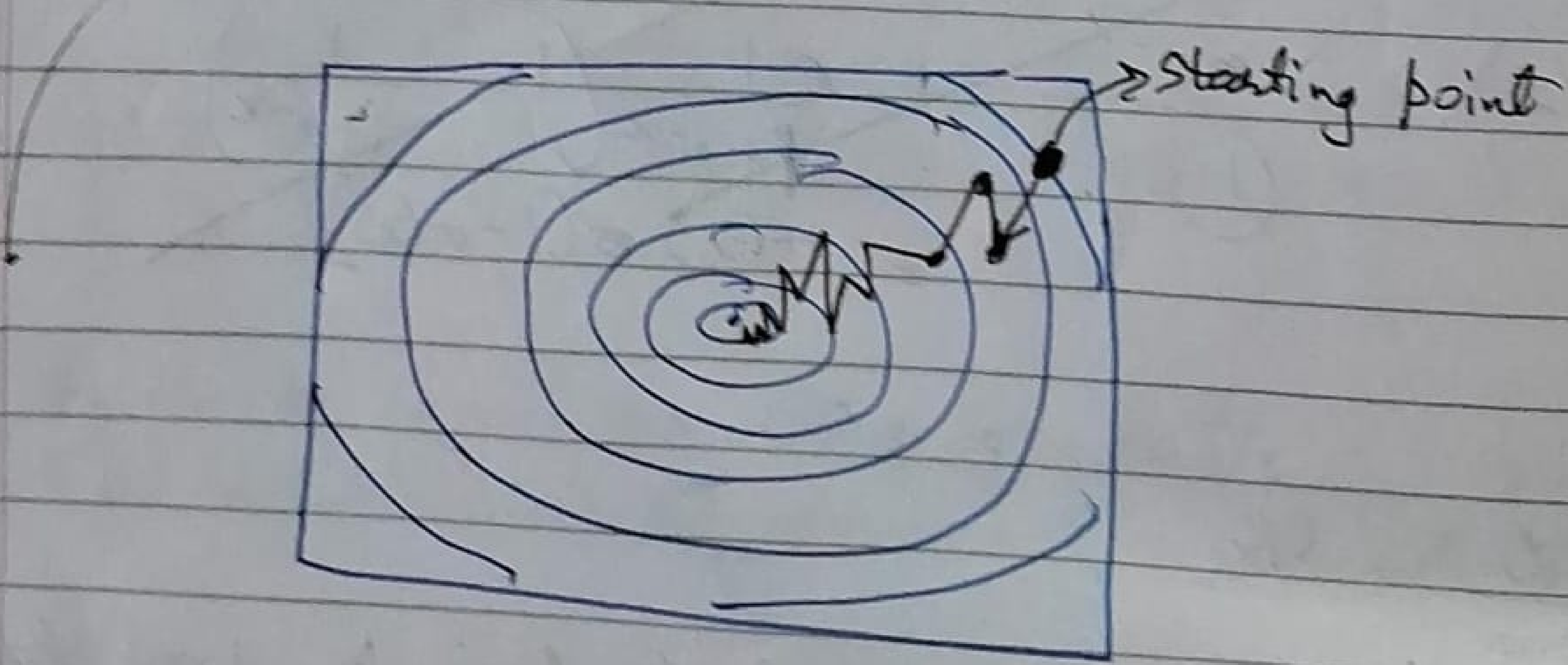
⇒ $w_i \leftarrow w_i - \alpha \frac{\partial \mathcal{L}}{\partial w_i}$

⇒ $w_i \leftarrow w_i - \alpha \delta x_i$

This means all w_i's ~~can~~ either increase or decrease simultaneously. The dirn of ~~motion~~ "step" is same and depends only on δ.

Problem with the above

Consider countour plot of $\mathcal{L}(W, b)$:



Suppose ~~two~~ in two iterations we get δs with opposite signs. ~~which is~~ This is possible because ~~the~~ all inputs increasing/decreasing simultaneously does not guarantee ~~actual~~ dirn of steepest descent

In such a case, we first "overshoot" (increase w_i's in one dirn) and then retract (move in another dirn). This causes "zig-zag" ~~conver~~ motion and chaotic convergence.