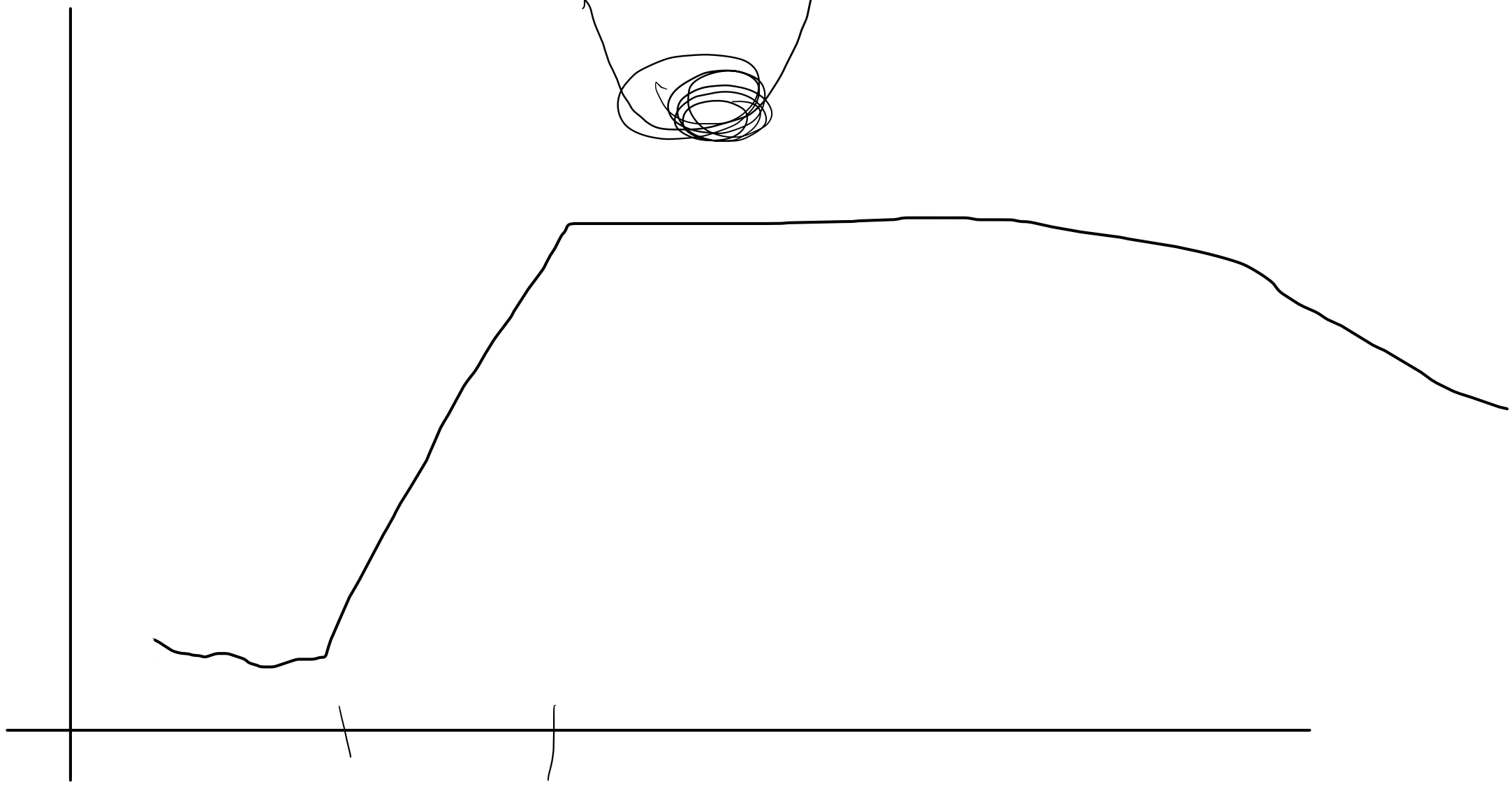
$$\theta_{nxt} = \theta - \eta \nabla_\theta MSE(\theta) \qquad (100)$$

Const $\qquad (\cdot 1) \qquad (10)$



Simulated annealing :->

Gradually reduce η (learning rate)

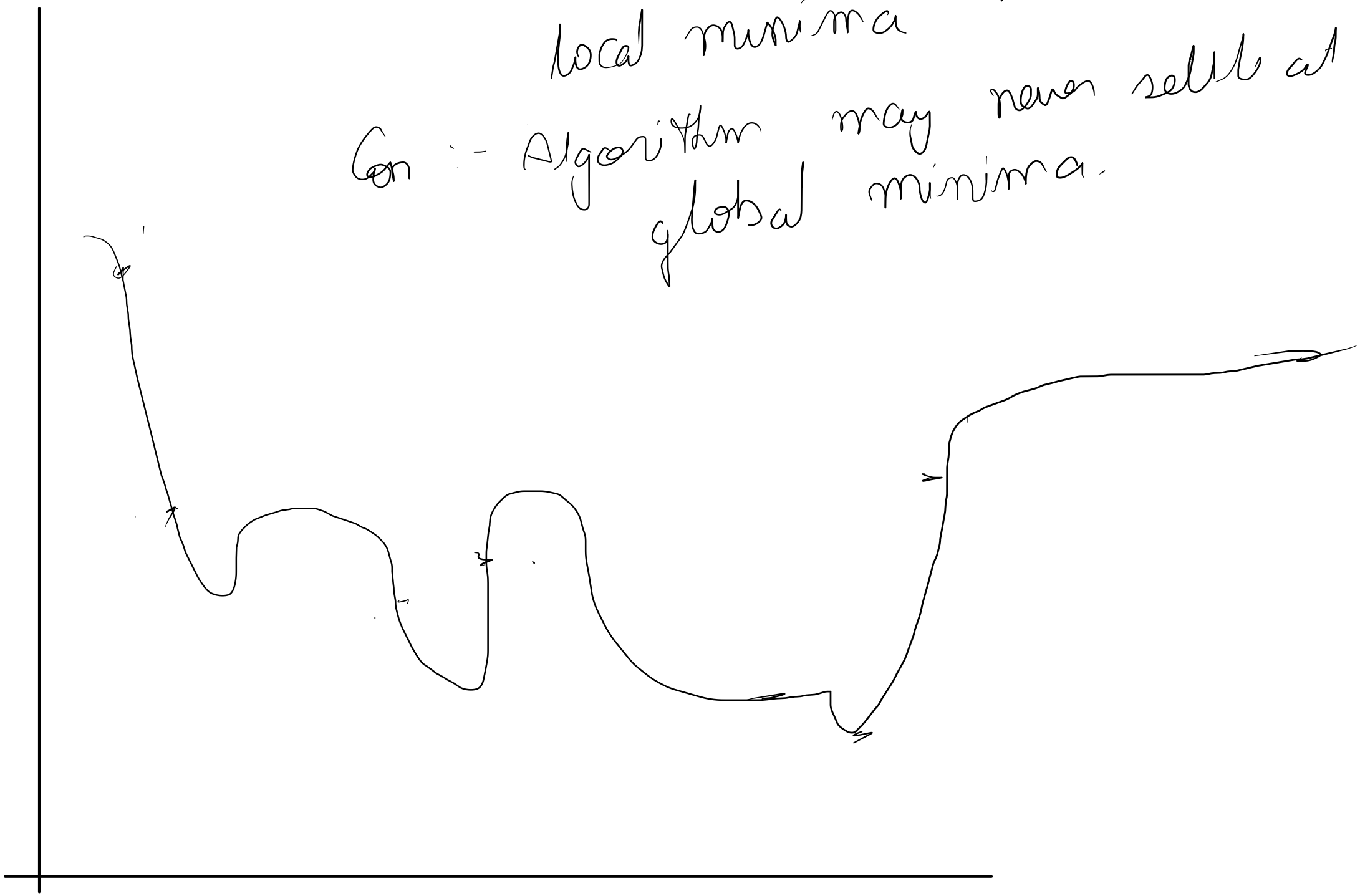
(+) :- Big initial steps & smaller steps at the end.

This will help algorithm to settle down at global minima.

Learning Schedule :- function determines the learning rate at each iteration.

Pro :- very less chance of getting stuck in local minima

Con :- Algorithm may never settle at global minima.



SoC