More intuitions about Taylor Series

Can we get the answer from some basic mathematics?

1. The real aim is:
   a. $w \Rightarrow w + \eta\Delta w$
   b. $b \Rightarrow b + \eta\Delta b$
   c. $Loss(w) > Loss(w + \eta\Delta w)$
   d. $Loss(b) > Loss(b + \eta\Delta b)$
   e. $Loss(w, b) > Loss(w + \eta\Delta w, b + \eta\Delta b)$
   f. $Loss(\theta) > Loss(\theta + \eta\Delta\theta)$ (where $\theta = [w, b]$)

2. Vectorized Taylor Series: $L(\theta + \eta u) = L(\theta) + \eta * u^T \nabla_\theta L(\theta) + \frac{\eta^2}{2!} * u^T \nabla^2 L(\theta) u + \frac{\eta^3}{3!} * \ldots$
3. Where, $u = \Delta\theta$
4. Here, we know that in practice, η is very small ie (0.001) etc
5. So η^2 ,η^3… all end up being negligible, so remove those corresponding terms
6. New Vectorized Taylor Series: $L(\theta + \eta u) \approx L(\theta) + \eta * u^T \nabla_\theta L(\theta)$
   a. Here, ∇_θ refers to Gradient w.r.t θ and it consists of the partial derivatives of $L(\theta)$ w.r.t w and b, stacked up into a vector
   b. $L(\theta + \eta u) \in \mathbb{R}$
   c. $L(\theta) \in \mathbb{R}$
   d. $\eta \in \mathbb{R}$ $\frac{\partial Loss}{\partial w}$ $\frac{\partial Loss}{\partial b}$
   e. $u^T(\nabla_\theta L(\theta))$ = Dot product of $\Delta\theta$ transposed and the partial derivative vector ∇_θ ($\in \mathbb{R}$)

[Δw	Δb]

$\frac{\partial Loss}{\partial w}$
$\frac{\partial Loss}{\partial b}$