Compute and update hidden state gradient: $\boldsymbol{\delta_h} = \left(\frac{\partial \mathbf{h}^{(t+1)}}{\partial \mathbf{h}^{(t)}}\right)^{\mathsf{T}} \boldsymbol{\delta_h} +$
(

Part I

$\text{ialo}^{(t)} \overline{\partial \mathbf{h}^{(t)\mathsf{T}} \nabla_{\mathbf{o}^{(t)}} \mathcal{L}}$

Acumulate parameter gradients: $\boldsymbol{\delta_c} + = \left(\frac{\partial \mathbf{o}^{(t)}}{\partial \mathbf{c}}\right)^{\mathsf{T}} \nabla_{\mathbf{o}^{(t)}} \mathcal{L}$ $\boldsymbol{\delta_b} + = \left(\frac{\partial \mathbf{h}^{(t)}}{\partial \mathbf{b}}\right)^{\mathsf{T}} \nabla_{\mathbf{h}^{(t)}} \mathcal{L}$ $\boldsymbol{\delta_V} + = \sum_i \left(\frac{\partial \mathcal{L}}{\partial o_i^{(t)}}\right) \nabla_{\mathbf{V}^{(t)}} o_i^{(t)}$ $\boldsymbol{\delta_W} + = \sum_i (\delta_{\mathbf{h}})_i \nabla_{\mathbf{W}^{(t)}} h_i^{(t)}$ $\boldsymbol{\delta_U} + = \sum_i (\delta_{\mathbf{h}})_i \nabla_{\mathbf{U}^{(t)}} h_i^{(t)}$ **Weight and bias update:** $\{W, U, V, \mathbf{b}, \mathbf{c}\} - = \alpha \cdot \{\boldsymbol{\delta_W}, \boldsymbol{\delta_U}, \boldsymbol{\delta_V}, \boldsymbol{ta_b}, \boldsymbol{\delta_c}\}$