Q3.

2.

$$\frac{\partial \mathrm{L}}{\partial \mathrm{x_t}} = \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial x_t}$$

$$= \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial Z}\frac{\partial Z}{\partial x_t}, \quad \text{where } Z = W_x x_t + W_h h_{t-1} + b$$

$$= \frac{\partial L}{\partial h_t} diag\left(\vec{1} - \tanh^2(W_x x_t + W_h h_{t-1} + b)\ \right) W_x$$

$$= \left(\frac{\partial L}{\partial h_t} \odot \left(\vec{1} - h_t^2\right)\right) W_x \qquad \text{note that } \odot \text{ gives a row vector}$$

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W_x}} = \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial W_x}$$

$$= \sum_i \frac{\partial L}{\partial h_{t,i}}\frac{\partial h_{t,i}}{\partial W_x}$$

$$= \left(\frac{\partial L}{\partial h_t} \odot \left(\vec{1} - h_t^2\right)\right)^T x_t^T$$

$$\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t-1}}} = \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial h_{t-1}}$$

$$= \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial Z}\frac{\partial Z}{\partial h_{t-1}}, \quad \text{where } Z = W_x x_t + W_h h_{t-1} + b$$

$$= \left(\frac{\partial L}{\partial h_t} \odot \left(\vec{1} - h_t^2\right)\right) W_h$$

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W_h}} = \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial W_h}$$

$$= \sum_i \frac{\partial L}{\partial h_{t,i}}\frac{\partial h_{t,i}}{\partial W_h}$$

$$= \left(\frac{\partial L}{\partial h_t} \odot \left(\vec{1} - h_t^2\right)\right)^T h_{t-1}^T$$

$$\frac{\partial \mathrm{L}}{\partial \mathrm{b}} = \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial b}$$

$$= \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial Z}\frac{\partial Z}{\partial b}, \quad \text{where } Z = W_x x_t + W_h h_{t-1} + b$$

$$= \frac{\partial L}{\partial h_t} \odot \left(\vec{1} - h_t^2\right)$$

4.

$\frac{\partial \mathrm{L}}{\partial \mathrm{x_t}}$ keeps the same:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{x_t}} = \left(\frac{\partial L}{\partial h_t} \odot \left(\vec{1} - h_t^2\right)\right) W_x$$

$\frac{\partial \mathrm{L}}{\partial \mathrm{W_x}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W_x}} = \sum_{\mathrm{t=1}}^{\mathrm{T}} \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial W_x}$$

$$= \sum_{\mathrm{t=1}}^{\mathrm{T}} \left(\frac{\partial L}{\partial h_t} \odot \left(\vec{1} - h_t^2\right)\right)^T x_t^T$$

$\frac{\partial \mathrm{L}}{\partial \mathrm{W_h}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W_h}} = \sum_{\mathrm{t=1}}^{\mathrm{T}} \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial W_h}$$

$$= \sum_{\mathrm{t=1}}^{\mathrm{T}} \left(\frac{\partial L}{\partial h_t} \odot \left(\vec{1} - h_t^2\right)\right)^T h_{t-1}^T$$

$\frac{\partial \mathrm{L}}{\partial \mathrm{b}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{b}} = \sum_{\mathrm{t=1}}^{\mathrm{T}} \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial b}$$

$$= \sum_{\mathrm{t=1}}^{\mathrm{T}} \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial Z} \frac{\partial Z}{\partial b}, \quad \text{where } Z = W_x x_t + W_h h_{t-1} + b$$

$$= \sum_{\mathrm{t=1}}^{\mathrm{T}} \frac{\partial L}{\partial h_t} \odot \left(\vec{1} - h_t^2\right)$$

$\frac{\partial \mathrm{L}}{\partial \mathrm{h_0}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{h_0}} = \frac{\partial L}{\partial h_1} \frac{\partial h_1}{\partial h_0}$$

$$= \frac{\partial L}{\partial h_1} \frac{\partial h_1}{\partial Z} \frac{\partial Z}{\partial h_0}, \quad \text{where } Z = W_x x_t + W_h h_{t-1} + b$$

$$= \frac{\partial L}{\partial h_1} \odot (\vec{1} - h_1^2) W_h$$

Q4.

2.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{x_t}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{x_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{f_t}} \frac{\partial f_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial \mathrm{i_t}} \frac{\partial i_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial \tilde{\mathrm{c}}_\mathrm{t}} \frac{\partial \tilde{c}_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial \mathrm{o_t}} \frac{\partial o_t}{\partial x_t}$$

$$= \frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} \frac{\partial c_t}{\partial f_t} \frac{\partial f_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial c_t} \frac{\partial c_t}{\partial i_t} \frac{\partial i_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial c_t} \frac{\partial c_t}{\partial \tilde{\mathrm{c}}_\mathrm{t}} \frac{\partial \tilde{c}_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \frac{\partial h_t}{\partial o_t} \frac{\partial o_t}{\partial x_t}$$

$$= \left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} \odot c_{t-1} \odot f_t \odot (1 - f_t)\right) W_x^f + \left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} \odot \tilde{c}_t \odot i_t \odot (1 - i_t)\right) W_x^i + \left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} \odot i_t \odot (1 - \tilde{c}_t^2)\right) W_x^c + \left(\frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot \tanh(c_t) \odot o_t \odot (1 - o_t)\right) W_x^o$$

- $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t-1}}}$: (similar with $\frac{\partial \mathrm{L}}{\partial \mathrm{x_t}}$)

$$\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t-1}}} = \left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} \odot c_{t-1} \odot f_t \odot (1 - f_t)\right) W_h^f + \left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} \odot \tilde{c}_t \odot i_t \odot (1 - i_t)\right) W_h^i + \left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} \odot i_t \odot (1 - \tilde{c}_t^2)\right) W_h^c + \left(\frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot \tanh(c_t) \odot o_t \odot (1 - o_t)\right) W_h^o$$

- $\frac{\partial \mathrm{L}}{\partial \mathrm{c_{t-1}}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{c_{t-1}}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot f_t$$

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W_x^{\mathit{f}}}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W_x^{\mathit{f}}}} = \frac{\partial L}{\partial c_t} \frac{\partial c_t}{\partial f_t} \frac{\partial f_t}{\partial W_x^f} = \left(\frac{\partial L}{\partial c_t} \odot \mathrm{c_{t-1}} \odot f_t \odot (1 - \mathrm{f_t})\right)^T x_t^T \ \in \mathbb{R}^{\mathrm{mxd}}$$

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W_h^{\mathit{f}}}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W_h^{\mathit{f}}}} = \left(\frac{\partial L}{\partial c_t} \odot \mathrm{c_{t-1}} \odot f_t \odot (1 - \mathrm{f_t})\right)^T h_{t-1}^T \ \in \mathbb{R}^{\mathrm{mxm}}$$

- $\frac{\partial \mathrm{L}}{\partial b^f}$:

$$\frac{\partial \mathrm{L}}{\partial f} = \frac{\partial L}{\partial c_t} \odot c_{\mathrm{t}-1} \odot f_t \odot (1 - \mathrm{f_t})$$

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{i}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{i}} = \left(\frac{\partial L}{\partial c_t} \odot \tilde{c}_{\mathrm{t}} \odot i_t \odot (1 - \mathrm{i_t})\right)^T x_t^T \ \in \mathbb{R}^{\mathrm{mxd}}$$

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{i}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{i}} = \left(\frac{\partial L}{\partial c_t} \odot \tilde{c}_{\mathrm{t}} \odot i_t \odot (1 - \mathrm{i_t})\right)^T h_{t-1}^T \ \in \mathbb{R}^{\mathrm{mxm}}$$

- $\frac{\partial \mathrm{L}}{\partial b^i}$:

$$\frac{\partial \mathrm{L}}{\partial b^i} = \frac{\partial L}{\partial c_t} \odot \tilde{c}_{\mathrm{t}} \odot i_t \odot (1 - \mathrm{i_t})$$

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{c}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{c}} = \left(\frac{\partial L}{\partial c_t} \odot i_{\mathrm{t}} \odot (1 - \tilde{\mathrm{c}}_t^2)\right)^T x_t^T \ \in \mathbb{R}^{\mathrm{mxd}}$$

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{c}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{c}} = \left(\frac{\partial L}{\partial c_t} \odot i_{\mathrm{t}} \odot (1 - \tilde{\mathrm{c}}_t^2)\right)^T h_{t-1}^T \ \in \mathbb{R}^{\mathrm{mxm}}$$

- $\frac{\partial \mathrm{L}}{\partial b^c}$:

$$\frac{\partial \mathrm{L}}{\partial b^c} = \frac{\partial L}{\partial c_t} \odot i_{\mathrm{t}} \odot (1 - \tilde{\mathrm{c}}_t^2)$$

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{o}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{o}} = \left(\frac{\partial L}{\partial c_t} \odot \tanh(\mathrm{c_t}) \odot \tilde{c}_{\mathrm{t}} \odot (1 - \tilde{c}_{\mathrm{t}})\right)^T x_t^T \ \in \mathbb{R}^{\mathrm{mxd}}$$

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{o}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{o}} = \left(\frac{\partial L}{\partial c_t} \odot \tanh(\mathrm{c_t}) \odot \tilde{c}_{\mathrm{t}} \odot (1 - \tilde{c}_{\mathrm{t}})\right)^T h_{t-1}^T \ \in \mathbb{R}^{\mathrm{mxm}}$$

- $\frac{\partial \mathrm{L}}{\partial b^o}$:

$$\frac{\partial \mathrm{L}}{\partial b^o} = \frac{\partial L}{\partial c_t} \odot \tanh(\mathrm{c_t}) \odot \tilde{c}_{\mathrm{t}} \odot (1 - \tilde{c}_{\mathrm{t}})$$

4.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{x_t}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{x_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{f_t}}\frac{\partial f_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial \mathrm{i_t}}\frac{\partial i_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial \tilde{\mathrm{c}}_{\mathrm{t}}}\frac{\partial \tilde{c}_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial \mathrm{o_t}}\frac{\partial o_t}{\partial x_t}$$

$$= \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}}\frac{\partial h_t}{\partial c_t}\frac{\partial c_t}{\partial f_t}\frac{\partial f_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}}\frac{\partial h_t}{\partial c_t}\frac{\partial c_t}{\partial i_t}\frac{\partial i_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}}\frac{\partial h_t}{\partial c_t}\frac{\partial c_t}{\partial \tilde{\mathrm{c}}_{\mathrm{t}}}\frac{\partial \tilde{c}_t}{\partial x_t} + \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}}\frac{\partial h_t}{\partial o_t}\frac{\partial o_t}{\partial x_t}$$

$$= \left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} \odot c_{t-1} \odot f_t \odot (1 - f_t)\right) W_x^f + \left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} \odot \tilde{c}_t \odot i_t \odot (1 - i_t)\right) W_x^i +$$

$$\left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} \odot i_t \odot (1 - \tilde{c}_t^2)\right) W_x^c + \left(\frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot \tanh(c_t) \odot o_t \odot (1 - o_t)\right) W_x^o$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{h_0}}$: (similar with $\frac{\partial \mathrm{L}}{\partial \mathrm{x_t}}$)

$$\frac{\partial \mathrm{L}}{\partial \mathrm{h_0}} = \left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_1}} \odot c_0 \odot f_1 \odot (1 - f_1)\right) W_h^f + \left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_1}} \odot \tilde{c}_1 \odot i_1 \odot (1 - i_1)\right) W_h^i +$$

$$\left(\frac{\partial \mathrm{L}}{\partial \mathrm{c_1}} \odot i_1 \odot (1 - \tilde{c}_1^2)\right) W_h^c + \left(\frac{\partial \mathrm{L}}{\partial \mathrm{h_1}} \odot \tanh(c_1) \odot o_1 \odot (1 - o_1)\right) W_h^o$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_1}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_1}} \odot o_1 \odot (1 - \tanh^2(c_1)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_2}} \odot \mathrm{f}_2$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_2}} = 0$ if 2>T.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W_x^f}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W_x^f}} = \sum\nolimits_{\mathrm{t}} \left(\frac{\partial L}{\partial c_t} \odot c_{\mathrm{t-1}} \odot f_t \odot (1 - \mathrm{f_t})\right)^T x_t^T \in \mathbb{R}^{\mathrm{mxd}}$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W_h^f}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{f}} = \sum\nolimits_{\mathrm{t}} \left(\frac{\partial L}{\partial c_t} \odot c_{\mathrm{t-1}} \odot f_t \odot (1 - \mathrm{f}_{\mathrm{t}}) \right)^T h_{t-1}^T \ \in \mathbb{R}^{\mathrm{mxm}}$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial b^f}$:

$$\frac{\partial \mathrm{L}}{\partial f} = \sum\nolimits_{\mathrm{t}} \frac{\partial L}{\partial c_t} \odot c_{\mathrm{t-1}} \odot f_t \odot (1 - \mathrm{f}_{\mathrm{t}})$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{i}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{i}} = \sum\nolimits_{\mathrm{t}} \left(\frac{\partial L}{\partial c_t} \odot \tilde{c}_{\mathrm{t}} \odot i_t \odot (1 - \mathrm{i}_{\mathrm{t}}) \right)^T x_t^T \ \in \mathbb{R}^{\mathrm{mxd}}$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{i}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{i}} = \sum\nolimits_{\mathrm{t}} \left(\frac{\partial L}{\partial c_t} \odot \tilde{c}_{\mathrm{t}} \odot i_t \odot (1 - \mathrm{i}_{\mathrm{t}}) \right)^T h_{t-1}^T \ \in \mathbb{R}^{\mathrm{mxm}}$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial b^i}$:

$$\frac{\partial \mathrm{L}}{\partial b^i} = \sum\nolimits_{t} \frac{\partial L}{\partial c_t} \odot \tilde{c}_{\mathrm{t}} \odot i_t \odot (1 - \mathrm{i}_{\mathrm{t}})$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{c}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{c}} = \sum\nolimits_{t} \left(\frac{\partial L}{\partial c_t} \odot i_{\mathrm{t}} \odot (1 - \tilde{\mathrm{c}}_t^2) \right)^T x_t^T \ \in \mathbb{R}^{\mathrm{mxd}}$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{c}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{c}} = \sum_t \left(\frac{\partial L}{\partial c_t} \odot i_{\mathrm{t}} \odot (1 - \tilde{\mathrm{c}}_t^2) \right)^T h_{t-1}^T \ \in \mathbb{R}^{\mathrm{mxm}}$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial b^c}$:

$$\frac{\partial \mathrm{L}}{\partial b^c} = \sum_t \frac{\partial L}{\partial c_t} \odot i_{\mathrm{t}} \odot (1 - \tilde{\mathrm{c}}_t^2)$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{o}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{x}}^{o}} = \sum_t \left(\frac{\partial L}{\partial c_t} \odot \tanh(\mathrm{c_t}) \odot \tilde{c}_{\mathrm{t}} \odot (1 - \tilde{c}_{\mathrm{t}}) \right)^T x_t^T \ \in \mathbb{R}^{\mathrm{mxd}}$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{o}}$:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{W}_{\mathrm{h}}^{o}} = \sum_t \left(\frac{\partial L}{\partial c_t} \odot \tanh(\mathrm{c_t}) \odot \tilde{c}_{\mathrm{t}} \odot (1 - \tilde{c}_{\mathrm{t}}) \right)^T h_{t-1}^T \ \in \mathbb{R}^{\mathrm{mxm}}$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.

- $\frac{\partial \mathrm{L}}{\partial b^o}$:

$$\frac{\partial \mathrm{L}}{\partial b^o} = \sum_t \frac{\partial L}{\partial c_t} \odot \tanh(\mathrm{c_t}) \odot \tilde{c}_{\mathrm{t}} \odot (1 - \tilde{c}_{\mathrm{t}})$$

where $\frac{\partial \mathrm{L}}{\partial \mathrm{c_t}} = \frac{\partial \mathrm{L}}{\partial \mathrm{h_t}} \odot o_t \odot (1 - \tanh^2(c_t)) + \frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} \odot \mathrm{f}_{t+1}$, and $\frac{\partial \mathrm{L}}{\partial \mathrm{h_{t+1}}} = 0$ if t+1>T.