Error comes from two parts, one from the next unit's back propagation $\mathrm{d}\bar{a}^{\langle t\rangle}$ and $\mathrm{d}\bar{c}^{\langle t\rangle}$, and another is from the loss function of present unit $\mathrm{d}\hat{a}^{\langle t\rangle}$ and $\mathrm{d}\hat{c}^{\langle t\rangle}$.

Thus:

$$\mathrm{d}a^{\langle t\rangle} = \mathrm{d}\bar{a}^{\langle t\rangle} + \mathrm{d}\hat{a}^{\langle t\rangle}$$

$\mathrm{d}c^{\langle t\rangle} = \mathrm{d}\bar{c}^{\langle t\rangle} + \mathrm{d}\hat{c}^{\langle t\rangle}$

$\mathrm{d}\hat{c}^{\langle t\rangle} = \mathrm{d}a^{\langle t\rangle} * \Gamma_o^{\langle t\rangle} * (1 - (\tanh(c^{\langle t\rangle})^2)$

Thus:

$$\mathrm{d}c^{\langle t\rangle} = \mathrm{d}\bar{c}^{\langle t\rangle} + \mathrm{d}a^{\langle t\rangle} * \Gamma_o^{\langle t\rangle} * (1 - \tanh(c^{\langle t\rangle})^2)$$

$\mathrm{d}\Gamma_o^{\langle t\rangle} = \mathrm{d}a^{\langle t\rangle} * \tanh(c^{\langle t\rangle})$

$\mathrm{d}\tilde{c}^{\langle t\rangle} = \mathrm{d}c^{\langle t\rangle} * \Gamma_u^{\langle t\rangle}$

$\mathrm{d}\Gamma_f^{\langle t\rangle} = \mathrm{d}c^{\langle t\rangle} * c^{\langle t-1\rangle}$

$\mathrm{d}\Gamma_u^{\langle t\rangle} = \mathrm{d}c^{\langle t\rangle} * \tilde{c}^{\langle t\rangle}$

Computing the gradients of parameters:

$$\begin{aligned}\mathrm{d}w_o &= \mathrm{d}\Gamma_o^{\langle t\rangle} * \Gamma_o^{\langle t\rangle} * (1 - \Gamma_o^{\langle t\rangle}) \cdot \begin{pmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{pmatrix}^T \\ &= \mathrm{d}a^{\langle t\rangle} * \tanh(c^{\langle t\rangle}) * \Gamma_o^{\langle t\rangle} * (1 - \Gamma_o^{\langle t\rangle}) \cdot \begin{pmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{pmatrix}^T\end{aligned}$$

$$\begin{aligned}\mathrm{d}b_o &= \sum_{aixs=1} \mathrm{d}\Gamma_o^{\langle t\rangle} * \Gamma_o^{\langle t\rangle} * (1 - \Gamma_o^{\langle t\rangle}) \\ &= \sum_{aixs=1} \mathrm{d}a^{\langle t\rangle} * \tanh(c^{\langle t\rangle}) * \Gamma_o^{\langle t\rangle} * (1 - \Gamma_o^{\langle t\rangle})\end{aligned}$$

$$\begin{aligned}\mathrm{d}w_f &= \mathrm{d}\Gamma_f^{\langle t\rangle} * \Gamma_f^{\langle t\rangle} * (1 - \Gamma_f^{\langle t\rangle}) \cdot \begin{pmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{pmatrix}^T \\ &= \mathrm{d}c^{\langle t\rangle} * c^{\langle t-1\rangle} * \Gamma_f^{\langle t\rangle} * (1 - \Gamma_f^{\langle t\rangle}) \cdot \begin{pmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{pmatrix}^T\end{aligned}$$

$$\begin{aligned}\mathrm{d}b_f &= \sum_{aixs=1} \mathrm{d}\Gamma_f^{\langle t\rangle} * \Gamma_f^{\langle t\rangle} * (1 - \Gamma_f^{\langle t\rangle}) \\ &= \sum_{aixs=1} \mathrm{d}c^{\langle t\rangle} * c^{\langle t-1\rangle} * \Gamma_f^{\langle t\rangle} * (1 - \Gamma_f^{\langle t\rangle})\end{aligned}$$

$$\begin{aligned}\mathrm{d}w_u &= \mathrm{d}\Gamma_u^{\langle t\rangle} * \Gamma_u^{\langle t\rangle} * (1 - \Gamma_u^{\langle t\rangle}) \cdot \begin{pmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{pmatrix}^T \\ &= \mathrm{d}c^{\langle t\rangle} * \tilde{c}^{\langle t\rangle} * \Gamma_u^{\langle t\rangle} * (1 - \Gamma_u^{\langle t\rangle}) \cdot \begin{pmatrix} a^{\langle t-1\rangle} \\ x^{\langle t\rangle} \end{pmatrix}^T\end{aligned}$$

$$
\begin{aligned}
\mathrm{d}b_u &= \sum_{aixs=1} \mathrm{d}\Gamma_u^{\langle t \rangle} * \Gamma_u^{\langle t \rangle} * (1 - \Gamma_u^{\langle t \rangle}) \\
&= \sum_{aixs=1} \mathrm{d}c^{\langle t \rangle} * \tilde{c}^{\langle t \rangle} * \Gamma_u^{\langle t \rangle} * (1 - \Gamma_u^{\langle t \rangle})
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{d}w_c &= \mathrm{d}\tilde{c}^{\langle t \rangle} * (1 - (\tilde{c}^{\langle t \rangle})^2) \cdot \begin{pmatrix} a^{\langle t-1 \rangle} \\ x^{\langle t \rangle} \end{pmatrix}^T \\
&= \mathrm{d}c^{\langle t \rangle} * \Gamma_u^{\langle t \rangle} * (1 - (\tilde{c}^{\langle t \rangle})^2) \cdot \begin{pmatrix} a^{\langle t-1 \rangle} \\ x^{\langle t \rangle} \end{pmatrix}^T
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{d}b_c &= \sum_{aixs=1} \mathrm{d}\tilde{c}^{\langle t \rangle} * (1 - (\tilde{c}^{\langle t \rangle})^2) \\
&= \sum_{aixs=1} \mathrm{d}c^{\langle t \rangle} * \Gamma_u^{\langle t \rangle} * (1 - (\tilde{c}^{\langle t \rangle})^2)
\end{aligned}
$$

Computing the gradients of hidden state, previous memory state and input:

$$
\mathrm{d}c^{\langle t-1 \rangle} = \mathrm{d}c^{\langle t \rangle} * \Gamma_f^{\langle t \rangle}
$$

$$
\begin{aligned}
\mathrm{d}a^{\langle t-1 \rangle} = \;& \mathrm{d}\Gamma_o^{\langle t \rangle} * \Gamma_o^{\langle t \rangle} * (1 - \Gamma_o^{\langle t \rangle}) * \hat{w}_o^T + \mathrm{d}\Gamma_f^{\langle t \rangle} * \Gamma_f^{\langle t \rangle} * (1 - \Gamma_f^{\langle t \rangle}) * \hat{w}_f^T \\
&+ \mathrm{d}\Gamma_u^{\langle t \rangle} * \Gamma_u^{\langle t \rangle} * (1 - \Gamma_u^{\langle t \rangle}) * \hat{w}_u^T + \mathrm{d}\tilde{c}^{\langle t \rangle} * (1 - (\tilde{c}^{\langle t \rangle})^2) * \hat{w}_c^T
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{d}x^{\langle t \rangle} = \;& \mathrm{d}\Gamma_o^{\langle t \rangle} * \Gamma_o^{\langle t \rangle} * (1 - \Gamma_o^{\langle t \rangle}) * \tilde{w}_o^T + \mathrm{d}\Gamma_f^{\langle t \rangle} * \Gamma_f^{\langle t \rangle} * (1 - \Gamma_f^{\langle t \rangle}) * \tilde{w}_f^T \\
&+ \mathrm{d}\Gamma_u^{\langle t \rangle} * \Gamma_u^{\langle t \rangle} * (1 - \Gamma_u^{\langle t \rangle}) * \tilde{w}_u^T + \mathrm{d}\tilde{c}^{\langle t \rangle} * (1 - (\tilde{c}^{\langle t \rangle})^2) * \tilde{w}_c^T
\end{aligned}
$$

Where $\hat{w}^T$ denote $w[:, : n_a]$, $\tilde{w}^T$ denote $w[:, n_a : n_a + n_x]$