误差由两部分组成，一部分来自于后一个单元反向传播记为$\mathrm{d}\bar{a}^{\langle t\rangle}$和$\mathrm{d}\bar{c}^{\langle t\rangle}$，一部分来自于本单元损失函数的产生的误差$\mathrm{d}\hat{a}^{\langle t\rangle}$和$\mathrm{d}\hat{c}^{\langle t\rangle}$

因此有：

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

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

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

则有：

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

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

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

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

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

计算参数梯度

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

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

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

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

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

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

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

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

计算关于隐藏状态、先前记忆状态和输入的导数

$$
dc^{\langle t-1 \rangle} = dc^{\langle t \rangle} * \Gamma_f^{\langle t \rangle}
$$

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

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

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