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 $d\bar{a}^{\langle t \rangle}$ 和 $d\bar{c}^{\langle t \rangle}$,一部分来自于本单元损失函数的产生的误差 $d\hat{a}^{\langle t \rangle}$ 和 $d\hat{c}^{\langle t \rangle}$

因此有:

$$\mathrm{d}a^{\langle t
angle} = \mathrm{d}ar{a}^{\langle t
angle} + \mathrm{d}\hat{a}^{\langle t
angle} \ \mathrm{d}c^{\langle t
angle} = \mathrm{d}ar{a}^{\langle t
angle} + \mathrm{d}\hat{a}^{\langle t
angle} \ \mathrm{d}\hat{c}^{\langle t
angle} = \mathrm{d}a^{\langle t
angle} * \Gamma_o^{\langle t
angle} * (1 - anh \; (\; c^{\langle t
angle} \;)^2)$$

则有:

$$\mathrm{d}c^{\langle t
angle}=\mathrm{d}ar{c}^{\langle t
angle}+\mathrm{d}a^{\langle t
angle}*\Gamma_{o}^{\langle t
angle}*\left(1- anh\left(\left.c^{\langle t
angle}
ight)
ight.^{2}
ight)$$

$$egin{aligned} \mathrm{d}\Gamma_o^{\langle t
angle} &= \mathrm{d}a^{\langle t
angle} * anh(c^{\langle t
angle}) \ \mathrm{d} ilde{c}^{\langle t
angle} &= \mathrm{d}c^{\langle t
angle} * \Gamma_u^{\langle t
angle} \ \mathrm{d}\Gamma_f^{\langle t
angle} &= \mathrm{d}c^{\langle t
angle} * c^{\langle t-1
angle} \ \mathrm{d}\Gamma_u^{\langle t
angle} &= \mathrm{d}c^{\langle t
angle} * ilde{c}^{\langle t
angle} \end{aligned}$$

计算参数梯度

$$egin{align*} \mathrm{d}w_o &= \mathrm{d}\Gamma_o^{\langle t
angle} *\Gamma_o^{\langle t
angle} *(1-\Gamma_o^{\langle t
angle}) * \left(rac{a^{\langle t-1
angle}}{x^{\langle t
angle}}
ight)^T \ &= \mathrm{d}a^{\langle t
angle} * anh(c^{\langle t
angle}) *\Gamma_o^{\langle t
angle} *(1-\Gamma_o^{\langle t
angle}) * \left(rac{a^{\langle t-1
angle}}{x^{\langle t
angle}}
ight)^T \ &\mathrm{d}b_o = \sum_{aixs=1} \mathrm{d}\Gamma_o^{\langle t
angle} *\Gamma_o^{\langle t
angle} *(1-\Gamma_o^{\langle t
angle}) \ &= \sum_{aixs=1} \mathrm{d}a^{\langle t
angle} * anh(c^{\langle t
angle}) *\Gamma_o^{\langle t
angle} *(1-\Gamma_o^{\langle t
angle}) \end{aligned}$$

$$\begin{split} \mathrm{d}w_f &= \mathrm{d}\Gamma_f^{\langle t \rangle} * \Gamma_f^{\langle t \rangle} * (1 - \Gamma_f^{\langle t \rangle}) * \left(\frac{a^{\langle t-1 \rangle}}{x^{\langle t \rangle}}\right)^T \\ &= \mathrm{d}c^{\langle t \rangle} * c^{\langle t-1 \rangle} * \Gamma_f^{\langle t \rangle} * (1 - \Gamma_f^{\langle t \rangle}) * \left(\frac{a^{\langle t-1 \rangle}}{x^{\langle t \rangle}}\right)^T \\ \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}) \\ \mathrm{d}w_u &= \mathrm{d}\Gamma_u^{\langle t \rangle} * \Gamma_u^{\langle t \rangle} * (1 - \Gamma_u^{\langle t \rangle}) * \left(\frac{a^{\langle t-1 \rangle}}{x^{\langle t \rangle}}\right)^T \\ &= \mathrm{d}c^{\langle t \rangle} * \tilde{c}^{\langle t \rangle} * \Gamma_u^{\langle t \rangle} * (1 - \Gamma_u^{\langle t \rangle}) * \left(\frac{a^{\langle t-1 \rangle}}{x^{\langle t \rangle}}\right)^T \\ \mathrm{d}b_u &= \sum_{aixs=1} \mathrm{d}\Gamma_u^{\langle t \rangle} * \tilde{c}^{\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}) \\ \mathrm{d}w_c &= \mathrm{d}\tilde{c}^{\langle t \rangle} * (1 - (\tilde{c}^{\langle t \rangle})^2) * \left(\frac{a^{\langle t-1 \rangle}}{x^{\langle t \rangle}}\right)^T \\ &= \mathrm{d}c^{\langle t \rangle} * \Gamma_u^{\langle t \rangle} * (1 - (\tilde{c}^{\langle t \rangle})^2) * \left(\frac{a^{\langle t-1 \rangle}}{x^{\langle t \rangle}}\right)^T \\ \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{split}$$

计算关于隐藏状态、先前记忆状态和输入的导数

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

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

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