1 Equation 3

$$A \leftarrow A - \eta_A \nabla_A L(W', D^{(val)}) \tag{1}$$

$$\nabla_A L(W', D^{(val)}) = \left(\frac{\partial E'}{\partial A} \frac{\partial W'}{\partial E'} + \frac{\partial F'}{\partial A} \frac{\partial W'}{\partial F'} \right) \nabla_{W'} L(W', D^{(val)}) \tag{2}$$

Consider,

$$\left(\frac{\partial E'}{\partial A} \frac{\partial W'}{\partial E'} \right) * \nabla_{W'} L(W', D^{(val)}) \tag{3}$$

$$\eta_e \eta_w \nabla^2_{AE} \sum_{i=1}^{n} a_i L(E, F, d_i^{(tr)}) * \nabla^2_{E'W} L\left(W, U, E', F'\right) * \nabla_{W'} L(W', D^{(val)}) \tag{4}$$

Evaluating from right to left,

$$\nabla^2_{E'W} L\left(W, U, E', F'\right) * \nabla_{W'} L(W', D^{(val)}) \tag{5}$$

Applying finite difference,

$$\frac{1}{2\alpha_{E'}} \nabla_{E'} L\left(W^+, E', F', U\right) - \nabla_{E'} L\left(W^-, E', F', U\right) \tag{6}$$

where,

$$W^{\pm} = W \pm \alpha_{E'} \nabla_{W'} L\left(W', E', F', D^{(val)}\right) \tag{7}$$

$$\alpha_{E'} = \frac{0.01}{\left\| \nabla_{W'} L\left(W', E', F', D^{(val)}\right) \right\|} \tag{8}$$

Consider $W' = G' + H'$

$$\alpha_{G'} = \frac{0.01}{\left\| \nabla_{G'} L\left(W', E', F', D^{(val)}\right) \right\|} \tag{9}$$

$$\alpha_{H'} = \frac{0.01}{\left\| \nabla_{H'} L\left(W', E', F', D^{(val)}\right) \right\|} \tag{10}$$

similarly,

$$G^{\pm} = G \pm \alpha_{G'} \nabla_{G'} L\left(W', E', F', D^{(val)}\right) \tag{11}$$

$$H^{'\pm} = H \pm \alpha_{H'} \nabla_{H'} L\left(W', E', F', D^{(val)}\right) \tag{12}$$

Hence, $W^+ = G'^+ + H'^+$ and $W^- = G'^- + H'^-$

$$\eta_e \eta_w \nabla^2_{AE} \sum_{i=1}^{n} a_i L\left(E, F, d_i^{(tr)}\right) \left(\frac{1}{2\alpha_{E'}} \nabla_{E'} L\left(W^+, E', F', U\right) - \nabla_{E'} L\left(W^-, E', F', U\right) \right) \tag{13}$$

Expanding (13) further,

$$\eta_e \eta_w \nabla^2_{AE} \sum_{i=1}^{n} a_i L\left(E, F, d_i^{(tr)}\right) * \nabla_{E'} L\left(W^+, E', F', U\right) \tag{14}$$

Applying finite difference approximation,

$$\left(\frac{1}{2\alpha_A} \nabla_A \sum_{i=1}^{n} a_i L\left(E'^+, F'^+, d_i^{tr}\right) - \nabla_A \sum_{i=1}^{n} a_i L\left(E'^-, F'^-, d_i^{tr}\right) \right) \tag{15}$$

where,

$$\alpha_{E'} = \frac{0.01}{\left\| \nabla_{E'} L\left(W^+, E', F', U\right) \right\|} \tag{16}$$

$$\alpha_{F'} = \frac{0.01}{\|\nabla_{F'} L\left(W^{+}, E', F', U\right)\|} \tag{17}$$

$$E^{\pm} = E \pm \alpha_{E'} \nabla_{E'} L\left(W^{+}, E', F', U\right) \tag{18}$$

$$F^{\pm} = F \pm \alpha_{F'} \nabla_{F'} L\left(W^{+}, E', F', U\right) \tag{19}$$

Expanding (13) further,

$$\nabla^2_{AE} \sum_{i=1}^{n} a_i L\left(E, F, d_i^{(tr)}\right) * \nabla_{E'} L\left(W^{-}, E', F', U\right) \tag{20}$$

Applying finite difference approximation,

$$\left(\frac{1}{2\alpha_A} \nabla_A \sum_{i=1}^{n} a_i L\left(E'^{+}, F'^{+}, d_i^{tr}\right) - \nabla_A \sum_{i=1}^{n} a_i L\left(E'^{-}, F'^{-}, d_i^{tr}\right)\right) \tag{21}$$

where,

$$\alpha_{E'} = \frac{0.01}{\|\nabla_{E'} L\left(W^{-}, E', F', U\right)\|} \tag{22}$$

$$\alpha_{F'} = \frac{0.01}{\|\nabla_{F'} L\left(W^{-}, E', F', U\right)\|} \tag{23}$$

$$E^{\pm} = E \pm \alpha_{E'} \nabla_{E'} L\left(W^{-}, E', F', U\right) \tag{24}$$

$$F^{\pm} = F \pm \alpha_{F'} \nabla_{F'} L\left(W^{-}, E', F', U\right) \tag{25}$$

Similarly for F'.