E-prop maths

Werner van der Veen

October 29, 2020

1 Proof: BPTT to E-prop

The main equation to be proved:

$$\frac{dE}{dW_{ji}} = \sum_t \frac{dE}{dz_j^t} \cdot \left[\frac{dz_j^t}{dW_{ji}}\right]_{\text{local}} \tag{1}$$

We start with the classical factorization of the loss gradients in an unrolled RNN:

$$\frac{dE}{dW_{ji}} = \frac{dE}{d\mathbf{h}_j^{t'}} \cdot \frac{\partial \mathbf{h}_j^{t'}}{\partial W_{ji}} \tag{2}$$

The summation indicates that weights are shared in an unrolled RNN.

We now decompose the first term into a series of learning signals $L_j^t = \frac{dE}{dz_j^t}$ and local factors $\frac{\partial \mathbf{h}_j^{t-t'}}{\partial \mathbf{h}_j^t}$ for t since the event horizon t':

$$\frac{dE}{d\mathbf{h}_j^{t'}} = \underbrace{\frac{dE}{dz_j^{t'}}}_{L_j^{t'}} \frac{\partial z_j^{t'}}{\partial \mathbf{h}_j^{t'}} + \frac{dE}{d\mathbf{h}_j^{t'+1}} \frac{\partial \mathbf{h}_j^{t'+1}}{\partial \mathbf{h}_j^{t'}} \tag{3}$$

Note that this equation is recursive. If we substitute the equation (3) into the classical factorization (2), we get:

$$\frac{dE}{dW_{ji}} = \sum_{t'} \left(L_j^{t'} \frac{\partial z_j^{t'}}{\partial \mathbf{h}_j^{t'}} + \frac{dE}{d\mathbf{h}_j^{t'+1}} \frac{\partial \mathbf{h}_j^{\prime t+1}}{\partial \mathbf{h}_j^{\prime t}} \right) \cdot \frac{\partial \mathbf{h}_j^{t'}}{\partial W_{ji}} \tag{4}$$

$$= \sum_{t'} \left(L_j^{t'} \frac{\partial z_j^{t'}}{\partial \mathbf{h}_j^{t'}} + \left(L_j^{t'+1} \frac{\partial z_j^{t'+1}}{\partial \mathbf{h}_j^{t'+1}} + (\cdots) \frac{\partial \mathbf{h}_j^{t'+2}}{\partial \mathbf{h}_j^{t'+1}} \right) \frac{\partial \mathbf{h}_j^{\prime t+1}}{\partial \mathbf{h}_j^{\prime t}} \right) \cdot \frac{\partial \mathbf{h}_j^{t'}}{\partial W_{ji}} \tag{5}$$

We write the term in parentheses into a second term indexed by t:

$$\frac{dE}{dW_{ji}} = \sum_{t'} \sum_{t \geq t'} L_j^t \frac{\partial z_j^t}{\partial \mathbf{h}_j^t} \frac{\partial \mathbf{h}_j^t}{\partial \mathbf{h}_j^{t-1}} \cdots \frac{\partial \mathbf{h}_j^{t+1}}{\partial \mathbf{h}_j^{t'}} \cdot \frac{\partial \mathbf{h}_j'^{t}}{\partial W_{ji}} \tag{6}$$

We then exchange the summation indices to pull out the learning signal L_j^t. This expresses the loss as a sum of learning signals multiplied by something we define as the eligibility trace. This eligibility trace consists of $\frac{\partial z_j^t}{\partial \mathbf{h}_j^t}$ and the eligibility vector ϵ_{ji}^t:

$$\frac{dE}{dW_{ji}} = \sum_t L_j^t \underbrace{\frac{\partial z_j^t}{\partial \mathbf{h}_j^t} \underbrace{\sum_{t \geq t'} \frac{\partial \mathbf{h}_j^t}{\partial \mathbf{h}_j^{t-1}} \cdots \frac{\partial \mathbf{h}_j^{t+1}}{\partial \mathbf{h}_j^{t'}} \cdot \frac{\partial \mathbf{h}_j'^{t}}{\partial W_{ji}}}_{\epsilon_{ji}^t}}_{e_{ji}^t} \tag{7}$$

This is the main e-prop equation.

2 Single-layer e-prop in pseudocode (LIF)

In LIF, $\{\mathbf{h}_j^t, \epsilon_{ji}^t\} \subset \mathbb{R}$.

for t in T **do**

$$z_j^t \leftarrow \begin{cases} 0, & \text{if } t - t_{z_j} < \delta t_{\text{ref}}. \\ H(v_j^t - v_{\text{th}}), & \text{otherwise.} \end{cases}$$

$$I_j^t \leftarrow \textstyle\sum_i W_{ji} z_i^t + \sum_u W_{ju} u(t)$$

$$v_j^{t+1} \leftarrow \alpha v_j^t + I_j^t - z_j^t \alpha v_j^t - z_j^{t-\delta t_{\text{ref}}} \alpha v_j^t$$

$$\epsilon_{ji}^{t+1} = \alpha (1 - z_j - z_j^{t-\delta t_{\text{ref}}}) \epsilon_{ji}^t + z_i^t$$

$$h_j^{t+1} \leftarrow \begin{cases} -\gamma, & \text{if } t - t_{z_j} < \delta t_{\text{ref}}. \\ \gamma \max\left(0, 1 - \left| \frac{v_j^{t+1} - v_{\text{th}}}{v_{\text{th}}} \right| \right), & \text{otherwise.} \end{cases}$$

$$e_{ji}^{t+1} \leftarrow h_j^{t+1} \epsilon_{ji}^{t+1}$$

$$y_k^t = \kappa y_k^{t-1} + \textstyle\sum_j W_{kj}^{\text{out}} z_j^t + b_k^{\text{out}}$$

$$W \leftarrow W - \eta \textstyle\sum_t \left(\sum_k B_{jk} \left(y_k^t - y_k^{*,t} \right) \right) e_{ji}^t$$

end for

3 ALIF steps

$$I_j^t = \sum_{t \neq j} W_{ji}^{\text{rec}} z_i^t + \sum_i W_{ji}^{\text{in}} x_i^{t+1} \tag{8}$$

$$A_j^t = v_{\text{th}} + \beta a_j^t \tag{9}$$

$$z_j^t = H\left(v_j^t - A_j^t\right) \tag{10}$$

$$\psi_j^t = \frac{1}{v_{\text{th}}} 0.3 \max\left(0, 1 - \left|\frac{v_j^t - A_j^t}{v_{\text{th}}}\right|\right) \tag{11}$$

$$y_k^t = \kappa y_k^{t-1} + \sum_j W_{kj}^{\text{out}} z_j^t + b_k^{\text{out}} \tag{12}$$

$$e_{ji}^t = \psi_j^t \left(\epsilon_{ji,v}^t - \beta \epsilon_{ji,a}^t\right) \tag{13}$$

$$\bar{e}_{ji}^t = \kappa \bar{e}_{ji}^{t-1} + e_{ji}^t \tag{14}$$

$$v_j^{t+1} = \alpha v_j^t + I_j^t - z_j v_{\text{th}} \tag{15}$$

$$a_j^{t+1} = \rho a_j^t + z_j^t \tag{16}$$

$$\epsilon_{ji,v}^{t+1} = \alpha \epsilon_{ji,v}^t + z_i^t \tag{17}$$

$$\epsilon_{ji,a}^{t+1} = \psi_j^t \epsilon_{ji,v}^t + \left(\rho - \psi_j^t \beta\right) \epsilon_{ji,a}^t \tag{18}$$

$$\Delta W_{ji}^{\text{rec}} = -\eta \sum_t \left(\sum_k B_{jk} \left(y_k^t - y_k^{*,t}\right)\right) \bar{e}_{ji}^t \tag{19}$$

$$\tag{20}$$

Given at time t, with given observable state z_i^t (simplified):

$$v_j^{t+1} = \alpha v_j^t + \sum_{t \neq j} W_{ji}^{\text{rec},t} z_i^t + \sum_i W_{ji}^{\text{in}} x_i^{t+1} - H\left(v_j^t - v_{\text{th}} - \beta a_j^t\right) v_{\text{th}} \tag{21}$$

$$a_j^{t+1} = \rho a_j^t + H\left(v_j^t - v_{\text{th}} - \beta a_j^t\right) \tag{22}$$

$$\epsilon_{ji,v}^{t+1} = \alpha \epsilon_{ji,v}^t + z_i^t \tag{23}$$

$$\epsilon_{ji,a}^{t+1} = \frac{1}{v_{\text{th}}} 0.3 \max\left(0, 1 - \left|\frac{v_j^t - v_{\text{th}} - \beta a_j^t}{v_{\text{th}}}\right|\right) \epsilon_{ji,v}^t \tag{24}$$

$$+ \left(\rho - \frac{1}{v_{\text{th}}} 0.3 \max\left(0, 1 - \left|\frac{v_j^t - v_{\text{th}} - \beta a_j^t}{v_{\text{th}}}\right|\right) \beta\right) \epsilon_{ji,a}^t \tag{25}$$

$$\bar{e}_{ji}^t = \kappa \bar{e}_{ji}^{t-1} + \frac{1}{v_{\text{th}}} 0.3 \max\left(0, 1 - \left|\frac{v_j^t - v_{\text{th}} - \beta a_j^t}{v_{\text{th}}}\right|\right) \left(\epsilon_{ji,v}^t - \beta \epsilon_{ji,a}^t\right) \tag{26}$$

$$y_k^t = \kappa y_k^{t-1} + \sum_j W_{kj}^{\text{out}} z_j^t + b_k^{\text{out}} \tag{27}$$

$$W_{ji}^{\text{rec},t+1} = W_{ji}^{\text{rec},t} - \eta \sum_t \left(\sum_k B_{jk}\left(y_k^t - y_k^{*,t}\right)\right) \bar{e}_{ji}^t \tag{28}$$

$$\tag{29}$$

Effects of $\mathbf{h}$ on W:

$$\frac{\partial v_j^t}{\partial W_{ji}} = \epsilon_{ji,v}^{t-1} \tag{30}$$

$$\frac{\partial u_j^t}{\partial W_{ji}} = 0 \tag{31}$$