Regarding the computation of $\frac{\partial r(i,\mathbf{y}_i,\Theta)}{\partial \mathbf{E}^{(l)}}$, the key idea is to write out a recursive function with respect to $\frac{\partial r(i,\mathbf{y}_i,\Theta)}{\partial \mathbf{E}^{(l)}}$ based on the chain rule. More specifically, we first consider the element-wise computation and apply the chain rule as follows.

$$\frac{\partial r(i,\mathbf{y}_i,\Theta)}{\partial \mathbf{E}^{(l)}[c,d]} = \sum_{e=1}^{n}\sum_{f=1}^{h_{l+1}} \frac{\partial r(i,\mathbf{y}_i,\Theta)}{\partial \mathbf{E}^{(l+1)}[e,f]} \frac{\partial \mathbf{E}^{(l+1)}[e,f]}{\partial \mathbf{E}^{(l)}[c,d]} \tag{22}$$

We take derivative on both sides of $\mathbf{E}^{(l)} = \sigma(\hat{\mathbf{A}}\mathbf{E}^{(l-1)}\mathbf{W}^{(l)})$ and get

$$\frac{\partial \mathbf{E}^{(l+1)}[e,f]}{\partial \mathbf{E}^{(l)}[c,d]} = \sigma'\big(\hat{\mathbf{A}}\mathbf{E}^{(l)}\mathbf{W}^{(l+1)}\big)[e,f]\hat{\mathbf{A}}[e,c]\mathbf{W}^{(l+1)}[d,f] \tag{23}$$

Combining Eqs. (22) and (23) together, we get the following element-wise first-order derivative.

$$\frac{\partial r(i,\mathbf{y}_i,\Theta)}{\partial \mathbf{E}^{(l)}[c,d]} = \sum_{e=1}^{n}\sum_{f=1}^{h_{l+1}} \hat{\mathbf{A}}^T[c,e]\cdot\left(\frac{\partial r(i,\mathbf{y}_i,\Theta)}{\partial \mathbf{E}^{(l+1)}}\sigma'\big(\hat{\mathbf{A}}\mathbf{E}^{(l)}\mathbf{W}^{(l+1)}\big)\right)[e,f] \cdot \big(\mathbf{W}^{(l+1)}\big)^T[f,d] \tag{24}$$

Written Eq. (24) into matrix form, we complete the proof.

PROOF OF THEOREM 1

We prove case by case.

Case 1. When $i = l$, since the activation function σ is the ReLU function, it is trivial that the subgradient of its second-order derivative is always 0 since the first-order derivative is the indicator function. Thus, $\mathfrak{H}_{l,l} = 0$.

Case 2. When $i = l-1$, to get Eq. (11), it is trivial to prove by taking derivative on both sides of Eq. (9). See the proof of Proposition 2 for the proof of Eq. (12).

Case 3. When $i < l-1$, we first take derivative on both sides of Eq. (9) in i-th hidden layer. Then we have

$$\mathfrak{H}_{l,i}[:,:,c,d] = \left(\hat{\mathbf{A}}\frac{\partial \mathbf{E}^{(i-1)}}{\partial \mathbf{W}^{(i)}[c,d]}\right)^T\left(\frac{\partial R}{\partial \mathbf{E}^{(i)}}\circ\sigma'_l\right) \tag{25}$$

To compute $\frac{\partial \mathbf{E}^{(l-1)}}{\partial \mathbf{W}^{(i)}[c,d]}$, we first consider an arbitrary (a,b)-th element in $\frac{\partial \mathbf{E}^{(l-1)}}{\partial \mathbf{W}^{(i)}[c,d]}$, i.e., $\frac{\partial \mathbf{E}^{(l-1)}[a,b]}{\partial \mathbf{W}^{(i)}[c,d]}$. Then we take the derivative on both sides of $\frac{\partial \mathbf{E}^{(l-1)}[a,b]}{\partial \mathbf{W}^{(i)}[c,d]}$, which gives us

$$\begin{aligned}\frac{\partial \mathbf{E}^{(l-1)}[a,b]}{\partial \mathbf{W}^{(i)}[c,d]} &= \sigma'_{l-1}[a,b]\cdot\sum_{e=1}^{n}\sum_{f=1}^{h_i}\hat{\mathbf{A}}[a,e]\frac{\partial \mathbf{E}^{(l-2)}[e,f]}{\partial \mathbf{W}^{(i)}[c,d]}\mathbf{W}^{(l-1)}[f,b]\\ &= \left(\sigma'_{l-1}\circ\left(\hat{\mathbf{A}}\frac{\partial \mathbf{E}^{(l-2)}}{\partial \mathbf{W}^{(i)}[c,d]}\mathbf{W}^{(l-1)}\right)\right)[a,b]\end{aligned} \tag{26}$$

It is trivial to get Eq. (13) by writing out the matrix form of Eq. (26).

Case 4. When $i = l+1$, to get Eq. (14), it is trivial to prove by taking derivative on both sides of Eq. (9). Regarding the computation of $\frac{\partial^2 R}{\partial \mathbf{E}^{(l)}[a,b]\partial \mathbf{W}^{(l+1)}[c,d]}$, by Eq. (10), we have

$$\frac{\partial R}{\partial \mathbf{E}^{(l)}[a,b]} = \left(\hat{\mathbf{A}}^T\left(\frac{\partial R}{\partial \mathbf{E}^{(l+1)}}\circ\sigma'\big(\hat{\mathbf{A}}\mathbf{E}^{(l)}\mathbf{W}^{(l+1)}\big)\right)\big(\mathbf{W}^{(l+1)}\big)^T\right)[a,b] \tag{27}$$

Then we take derivative on both sides and get

$$\begin{aligned}\frac{\partial^2 R}{\partial \mathbf{E}^{(l)}[a,b]\partial \mathbf{W}^{(l+1)}[c,d]} &= \sum_{e=1}^{n}\sum_{f=1}^{h_{l+1}}\hat{\mathbf{A}}^T[a,e]\cdot\left(\frac{\partial R}{\partial \mathbf{E}^{(l+1)}}\sigma'_{l+1}\right)[e,f] \cdot\frac{\partial \mathbf{W}^{(l+1)}[b,f]}{\partial \mathbf{W}^{(l+1)}[c,d]}\\ &= \mathbf{I}[b,c]\sum_{e=1}^{n}\hat{\mathbf{A}}^T[a,e]\cdot\left(\frac{\partial R}{\partial \mathbf{E}^{(l+1)}}\sigma'_{l+1}\right)[e,d]\\ &= \mathbf{I}[b,c]\left(\hat{\mathbf{A}}^T\left(\frac{\partial R}{\partial \mathbf{E}^{(l+1)}}\circ\sigma'_{l+1}\right)\right)[a,d]\end{aligned} \tag{28}$$

Case 5. When $i > l+1$, we get Eq. (15) by taking derivative on both sides of Eq. (10).

Putting everything (Cases 1 – 5) together, we complete the proof.