

## Network Dynamics

### Event Rate

#### Hidden Layers

At time  $t$ , input  $\mathbf{h}^0(t-1)$  is received by the first hidden layer and used to compute its somatic voltages  $\mathbf{v}^1(t)$ :

$$\mathbf{v}^1(t) = \mathbf{W}^1 \mathbf{h}^0(t-1) + \mathbf{b}^1 \quad (1)$$

Then, the *event rates* of the first hidden layer units,  $\mathbf{h}^1(t)$ , are given by applying the softplus function  $\sigma$ :

$$\mathbf{h}^1(t) = \sigma(\mathbf{v}^1(t)) \quad (2)$$

Similarly, for all other hidden layers  $i$  of the network,

$$\mathbf{v}^i(t) = \mathbf{W}^i \mathbf{h}^{i-1}(t-1) + \mathbf{b}^i \quad (3)$$

$$\mathbf{h}^i(t) = \sigma(\mathbf{v}^i(t)) \quad (4)$$

#### Output Layer

The somatic voltages of output layer  $k$ ,  $\mathbf{v}^k(t)$ , are calculated as in equation (3). Unlike hidden layers, output layer units have a rectified linear activation function  $\psi$ :

$$\mathbf{h}^k(t) = \psi(\mathbf{v}^k(t)) \quad (5)$$

During training, a teaching signal  $\mathbf{t}$  is randomly presented at the output layer. When the teaching signal is present, the event rates of the final layer are equal to the teaching signal. That is, if a teaching signal arrives at time  $t^*$ , the event rates of the output layer  $k$  become:

$$\mathbf{r}^k(t^*) = \mathbf{t} \quad (6)$$

## Burst Probability & Burst Rate

#### Output Layer

At the output layer  $k$ , we set the *burst probabilities* to be a constant value  $\omega$ :

$$\mathbf{p}^k(t) = \omega \quad (7)$$

The *burst rates* of the output layer  $k$  at time  $t$  are given by:

$$\boldsymbol{\beta}^k(t) = \mathbf{p}^k(t) \mathbf{h}^k(t) \quad (8)$$

## Hidden Layers

At time  $t$ , for hidden layer  $i$ , the apical dendritic voltages  $\mathbf{u}^i(t)$  are:

$$\mathbf{u}^i(t) = \frac{\mathbf{Y}^i \boldsymbol{\beta}^{i+1}(t-1)}{\mathbf{Z}^i \boldsymbol{\beta}^i(t-1)} \quad (9)$$

where  $\boldsymbol{\beta}^i(t)$  are the *burst rates* of layer  $i$  at time  $t$  (see equation (8)). The *burst probabilities* of hidden layer  $i$ ,  $\mathbf{p}^i(t)$ , are given by applying the sigmoid function  $\sigma$ :

$$\mathbf{p}^i(t) = \sigma(\mathbf{u}^i(t)) \quad (10)$$

Finally, the *burst rates* of hidden layer  $i$  are as in equation (8).

## Recurrent Weight Updates

The recurrent weights  $\mathbf{Z}^i$  are updated continuously in order to descend the loss function  $L^{\mathbf{Z}^i}(t)$ :

$$L^{\mathbf{Z}^i}(t) = \|\mathbf{u}^i(t)\|_2^2 + \|\alpha - \mathbf{Z}^i\|_2^2 \quad (11)$$

where  $\alpha$  is a constant. The first term will update  $\mathbf{Z}^i$  in order to push the apical voltages  $\mathbf{u}^i$  toward 0, while the second term is a regularization term that will push  $\mathbf{Z}$  toward small values as long as  $\alpha$  is small. The parameter  $\alpha$  allows us to control how small  $\mathbf{Z}$  should be allowed to get. Thus, at every time step  $t$ ,  $\mathbf{Z}^i$  is updated according to:

$$\Delta \mathbf{Z}^i(t) = -\eta \frac{\partial L^{\mathbf{Z}^i}(t)}{\partial \mathbf{Z}^i} \quad (12)$$

$$= \eta \left( \left( \frac{(\mathbf{u}^i(t))^2}{\mathbf{Z}^i \boldsymbol{\beta}^i(t-1)} \right) \otimes \boldsymbol{\beta}^i(t-1) + (\alpha - \mathbf{Z}^i) \right) \quad (13)$$

## Feedback Weight Updates

The feedback weights  $\mathbf{Y}^i$  are updated continuously in order to descend the loss function  $L^{\mathbf{Y}^i}(t)$ :

$$L^{\mathbf{Y}^i}(t) = \|\gamma - \mathbf{u}_{\max}^i\|_2^2 \quad (14)$$

where  $\gamma$  is a constant, and  $\mathbf{u}_{\max}^i$  are the maximum possible values for the apical voltages that can be driven by feedback, if recurrent activity is kept fixed:

$$\mathbf{u}_{\max}^i := \frac{|\mathbf{Y}^i| \mathbf{1}}{\mathbf{Z}^i \boldsymbol{\beta}^i(t-1)} \quad (15)$$

where  $\mathbf{1}$  is a vector of ones of the same size as layer  $i + 1$ . Descending this loss function will update  $\mathbf{Y}^i$  in order to produce an upper bound and lower bound of  $\mathbf{u}_{\max}^i$  and  $-\mathbf{u}_{\max}^i$ , respectively, for the apical voltages  $\mathbf{u}^i$ . The parameter  $\gamma$  allows us to control what value the upper and lower bounds should have. Thus, at every time step  $t$ ,  $\mathbf{Y}^i$  is updated according to:

$$\Delta \mathbf{Y}^i(t) = -\eta \frac{\partial L^{\mathbf{Y}^i}(t)}{\partial \mathbf{Y}^i} \quad (16)$$

$$= \eta \left( \frac{(\gamma - \mathbf{u}_{\max}^i)}{\mathbf{Z}^i \boldsymbol{\beta}^i(t-1)} |\mathbf{Y}^i| \right) \quad (17)$$

## Feedforward Weight Updates

For a network with output layer  $k$ , the feedforward weights  $\mathbf{W}^i$  are updated continuously in order to descend the global loss function  $L(t)$ :

$$L(t) = \|\boldsymbol{\beta}^k(t) - \boldsymbol{\beta}^k(t-1)\|_2^2 \quad (18)$$

First, we will show the feedforward weight updates that would be prescribed by backpropagation of error. For the output layer,

$$\Delta \mathbf{W}^k(t) = -\eta \frac{\partial L(t)}{\partial \mathbf{W}^k} \quad (19)$$

$$= -\eta \frac{\partial L(t)}{\partial \boldsymbol{\beta}^k(t-1)} \frac{\partial \boldsymbol{\beta}^k(t-1)}{\partial \mathbf{r}^k(t-1)} \frac{\partial \mathbf{r}^k(t-1)}{\partial \mathbf{v}^k(t-1)} \frac{\partial \mathbf{v}^k(t-1)}{\partial \mathbf{W}^k} \quad (20)$$

$$= (\boldsymbol{\beta}^k(t) - \boldsymbol{\beta}^k(t-1)) \omega H(\mathbf{v}^k(t-1)) \otimes \mathbf{h}^{k-1}(t-2) \quad (21)$$

where  $H$  is the Heaviside step function. Let us denote  $\boldsymbol{\delta}^k$  as:

$$\boldsymbol{\delta}^k := \frac{\partial L(t)}{\partial \boldsymbol{\beta}^k(t-1)} \frac{\partial \boldsymbol{\beta}^k(t-1)}{\partial \mathbf{r}^k(t-1)} \frac{\partial \mathbf{r}^k(t-1)}{\partial \mathbf{v}^k(t-1)} \quad (22)$$

For the last hidden layer  $k-1$ ,

$$\Delta \mathbf{W}^{k-1}(t) = -\eta \frac{\partial L(t)}{\partial \mathbf{W}^{k-1}} \quad (23)$$

$$= -\eta \frac{\partial L(t)}{\partial \boldsymbol{\beta}^k(t-1)} \frac{\partial \boldsymbol{\beta}^k(t-1)}{\partial \mathbf{r}^k(t-1)} \frac{\partial \mathbf{r}^k(t-1)}{\partial \mathbf{v}^k(t-1)} \frac{\partial \mathbf{v}^k(t-1)}{\partial \mathbf{W}^k} \quad (24)$$

$$= (\boldsymbol{\beta}^k(t) - \boldsymbol{\beta}^k(t-1)) \omega H(\mathbf{v}^k(t-1)) \otimes \mathbf{h}^{k-1}(t-2) \quad (25)$$