

## Original

$$\bar{c}(t) = m(t) \cdot \bar{d}$$

$$\theta = (\bar{c}/c_0)^p \bar{c}$$

$$\dot{m}_j(t) = \phi(c(t))d_j(t) - \epsilon m_j(t)$$

## Law and Cooper

$$y = \sigma\left(\sum_i w_i x_i\right)$$

$$\theta = E[y^2]$$

$$\frac{dw_i}{dt} = \frac{y(y - \theta)x_i}{\theta}$$

The Law and Cooper form has all of the same fixed points as the Intrator and Cooper form, but speed of synaptic modification increases when the threshold is small, and decreases as the threshold increases. The practical result is that the simulation can be run with artificially high learning rates, and wild oscillations are reduced. This form has been used primarily when running simulations of networks, where the run-time of the simulation can be prohibitive. ascholarped

## Squadrani et al version

$$z = ReLU\left(\sum_i w_i x_i\right)$$

$$\theta_t = \gamma\theta_{t-1} + (1 - \gamma)\langle z^2 \rangle_{b_t}$$

$$\frac{dw_i}{dt} = \frac{z(z - \theta)x_i}{\theta}$$

in the code with einstein notation:

$$\Delta W = \frac{\left(\frac{z(z-\theta)}{\theta}\right)_j^i x_k^j}{batchsize}$$

$$W \leftarrow \epsilon * \Delta W$$

for trace:

$$\langle \theta \rangle = \frac{1}{B} \sum_{t=1}^B \theta_t$$

update theta = use a moving average of previous batch-averaged quadratic postsynaptic activities where

$\gamma$  : decay-memory factor

$\langle \cdot \rangle_{b_t}$  : average over the batch of training patterns considered at the time step t