Neural Networks and Biological Modeling

Professor Wulfram Gerstner
Laboratory of Computational Neuroscience

CORRECTION QUESTION SET 12

Exercise 1: Synaptic Plasticity: the BCM rule

1.1   First consider a presynaptic neuron j_1 taken from the first population. The dynamics of the strength of the synapse with neuron i is such that

1. During the firing of the first population

$$\frac{d}{dt}w_{ij_1} = a_2^{\text{corr}}\Phi(\nu_i^{\text{post}} - 20\,\text{Hz})\nu_{j_1}^{\text{pre}}, \tag{1}$$

    with

$$\nu_i^{\text{post}} = \sum_{1^{\text{st}}\text{population}} w_{ij}\nu_j^{\text{pre}} = 30\,\text{Hz}. \tag{2}$$

    According to the graph in fig 1b we see that $\frac{d}{dt}w_{ij_1}$ is strictly positive. Since increased weights lead to increasing postsynaptic rate ν_i^{post} which in turn leads to increasing $\frac{d}{dt}w_{ij_1}$, the weights keep on increasing.

2. During the firing of the second population $\nu_{j_1}^{\text{pre}} = 0$ and hence $\frac{d}{dt}w_{ij_1} = 0$. The synaptic strengths stay fixed.

Therefore we conclude that the synapses of the first population of neurons become stronger with time.

Now consider a neuron j_2 belonging to the second population. During the firing of the first population, since $\nu_{j_2}^{\text{pre}} = 0$, we have

$$\frac{d}{dt}w_{ij_2} = 0 \tag{3}$$

During the firing of the second group, since initially we have $\nu_i^{\text{post}} = 10\,\text{Hz}$ which is below threshold, $\frac{d}{dt}w_{ij_2}$ is negative and will remain negative (see fig 1b) until the weights reach their minimum value (say zero – afterwards, weights are artificially kept null). The neuron i has become selective to one of the two groups, thus achieving group discrimination (which is good for memory).

1.2   The post-synaptic frequencies during the firing of the two groups are respectively $25\,\text{Hz}$ and $30\,\text{Hz}$, which is above ϑ. We are back to the situation described above, and the weights of both groups will grow indefinitely. No discrimination is made. One could argue that weights of group 1 will grow faster than those of group 2. This is true, but physically "growing without bound" actually means "growing until a limit is reached", and eventually the weights of both groups will end up being similar.

1.3   The idea is to place the threshold ϑ (which is the border between depression and potentiation) always at the right point to allow discrimination. If the postsynaptic rate increases (in average), then the threshold should follow quickly. To set things, let us assume that $\theta = \langle \nu_i \rangle$ (averaged post-synaptic rate, which can be implemented with a standard low pass filter with long time constant). In the situation given in 1.2, the threshold would slide until it reaches $\frac{25+30}{2} = 27.5$, in which case the situation becomes analog to 1.1, allowing group selection. In fact, this explanation is simplistic

in the sense that the weights also change with time: as the threshold slides towards the average rate, the average rate also increases because weights increase in both groups. It can be shown that the threshold must vary more rapidly than $\langle \nu_i \rangle$. In practice, one usually takes $\theta = \langle \nu_i^2 \rangle$ (average *squared* post-synaptic rate).

Exercise 2: Hopfield networks and Hebbian learning

2.1 We begin by calculating the change of weights Δw_{ij}^μ induced by presenting pattern μ to the network for $0.5s$:

$$\Delta w_{ij}^\mu = \int_0^{0.5} \frac{d}{dt} w_{ij}^\mu dt = \int_0^{0.5} a_2(\nu_i^\mu - 10)(\nu_j^\mu - 10)dt = 0.5\ a_2(\nu_i^\mu - 10)(\nu_j^\mu - 10) = 0.5\ a_2\ 10^2\ p_i^\mu p_j^\mu. \tag{4}$$

The last equality is easily explained: for all i and μ we see that if $p_i^\mu = 1$ then $\nu_i^\mu - 10 = 20 - 10 = 10 = 10p_i^\mu$, and similarly if $p_i^\mu = -1$ then $\nu_i^\mu - 10 = 0 - 10 = 10p_i^\mu$.

Thus, by choosing $a_2 = \frac{1}{50}$ and summing over all prototype presentations, we have $w_{ij}^{final} = \sum_\mu p_i^\mu p_j^\mu$, as we wanted. This exercise is intended to convince you that it is possible to learn memories in a fully interconnected network using a simple Hebbian learning rule.

2.2 Just expanding the given learning rule, we have $\frac{d}{dt} w_{ij} = a_2\vartheta^2 - a_2\vartheta\nu_i^{\text{post}} - a_2\vartheta\nu_i^{\text{pre}} + a_2\nu_i^{\text{post}}\nu_i^{\text{pre}}$. To map this into the general formulation we choose: $a_0 = a_2\vartheta^2$, $a_1^{\text{pre}} = -a_2\vartheta$, $a_1^{\text{post}} = -a_2\vartheta$, $a_2^{\text{corr}} = a_2$.

2.3 The learning is unsupervised, since the network learns implicit associations present in the input without any additional teaching signal, e.g. when the prototype being presented changed. We should note however, that learning should be limited to the first presentation of each pattern. During retrieval, we do not want the weights to change. Such a distinction of "learn a new pattern" and "retrieve a known pattern" could be triggered by a novelty related neuromodulator. In that sense, learning is not purely unsupervised.

Exercise 3: Spike-time dependent plasticity by local variables

From the learning rule

$$\frac{d}{dt} w_{ij} = a_+ x_j^{\text{pre}}\delta(t - t_i^{\text{post}}) - a_- y^{\text{post}}\delta(t - t_j^{\text{pre}}) \tag{5}$$

one can see that weight changes occur only when one of the neurons fire (pre or post).

Let us consider an isolated pair of spikes, "pre before post". The presynaptic neuron fires at time t_j^{pre} and the postsynaptic neuron fires at time t_i^{post}, with $\Delta t = t_j^{\text{pre}} - t_i^{\text{post}} < 0$.

- When the presynaptic spike occurs, the weight is depressed by an amount $a_- \cdot y_i(t_j^{\text{pre}})$. However, it has been a long time since the postsynaptic neuron spiked ("isolated pair"). Therefore the trace y_i has already decayed back to 0: no weight change occurs.

- When the postsynaptic neuron fires, the weight is potentiated by an amount $a_+ \cdot x_j(t_i^{\text{post}})$. When the presynaptic spike occured previously ($|\Delta t|$ ms ago), the trace x_j had jumped from 0 to $\frac{1}{\tau_+}$, and has been decaying until now (now $= t_i^{\text{post}}$). We know that the decay is exponential with time constant τ_+, so that $x_j(t_i^{\text{post}}) = \frac{1}{\tau_+} e^{-|\Delta t|/\tau_+}$.

In conclusion, the total weight change resulting from a "pre-before-post" pair depends on the time interval Δt according to

$$\Delta w(\Delta t < 0) = \frac{a_+}{\tau_+}\ e^{-|\Delta t|/\tau_+}.$$

Let us now consider an isolated pair of spikes in reverse order, "post before pre". The same reasoning holds: no potentiation occurs at the postsynaptic spike ($x_j^{\text{post}} = 0$), and the weight is depressed by $\frac{a_-}{\tau_-}e^{-|\Delta t|/\tau_-}$ when the presynaptic neuron fires. In conclusion, the total weight change resulting from a "post-before-pre" pair depends on the time interval Δt according to

$$\Delta w(\Delta t > 0) = -\frac{a_-}{\tau_-}\ e^{-|\Delta t|/\tau_-}$$

We have derived an analytical expression for both parts of the learning window. It is plotted in figure 1.
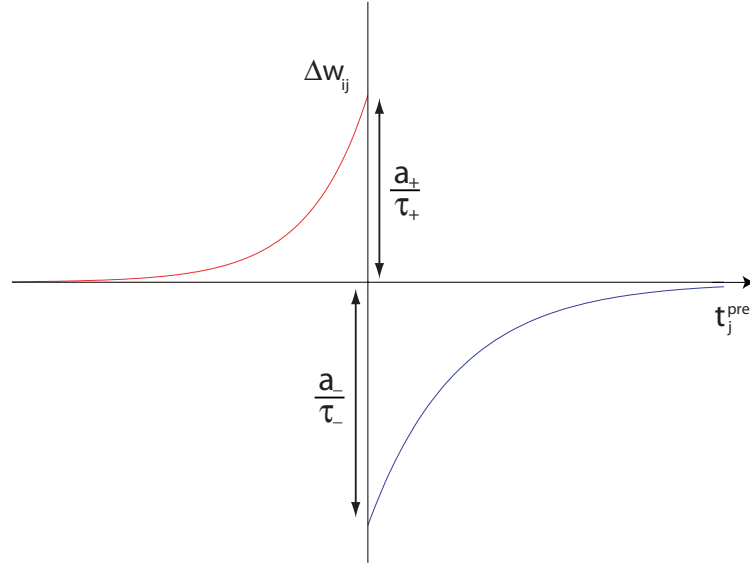

Figure 1:   The STDP window

Exercise 4: From spike-time dependent plasticity to rate models

4.1   The weight change Δw_{ij} is given by: $\Delta w_{ij} = \sum_{f,f'} W(t_i^f - t_j^{f'})$, where the sum runs over all spikes occurring within T. We can write this quantity as an integral over spike trains:

$$\Delta w_{ij} = \int_0^T dt \int_{-\infty}^{+\infty} W(s)S_i(t)S_j(t-s)ds \tag{6}$$

Taking the expectation, we have:

$$\langle\,\Delta w_{ij}\,\rangle = \int_0^T dt \int_{-\infty}^{+\infty} W(s)\,\langle\,S_i(t)S_j(t-s)\,\rangle\,ds \tag{7}$$

$$= \int_0^T dt \int_{-\infty}^{+\infty} W(s)\,\langle\,S_i(t)\,\rangle\,\langle\,S_j(t-s)\,\rangle\,ds \tag{8}$$

$$= \int_0^T dt \int_{-\infty}^{+\infty} W(s)\underbrace{\langle\,S_i(t)\,\rangle}_{=\nu_i}\underbrace{\langle\,S_j(t-s)\,\rangle}_{=\nu_j}\,ds \tag{9}$$

$$= \int_0^T dt \int_{-\infty}^{+\infty} W(s)\nu_i\nu_j ds \tag{10}$$

$$= T\nu_i\nu_j \int_{-\infty}^{+\infty} W(s)ds. \tag{11}$$

The second equality comes from the fact that the two Poisson processes are independent.
In words, Eq. 11 tells us that the expected weight change is equal to the area of the STDP window multiplied by the firing rates and the time interval, as we might have intuitively thought.

4.2 We have:

$$\Delta w_{ij} = \int_0^T dt \int_{-\infty}^{+\infty} W(s)S_i(t)S_j(t-s)ds \tag{12}$$

Taking the expectation, we have:

$$\langle\,\Delta w_{ij}\,\rangle = \int_0^T dt \int_{-\infty}^{+\infty} W(s)\,\langle\,S_i(t)S_j(t-s)\,\rangle\,ds \tag{13}$$

$$= \int_0^T dt \int_{-\infty}^{+\infty} W(s)\left\langle\,\sum_k w_{ik}\int_0^\infty \epsilon(s')S_k(t-s')ds'\,S_j(t-s)\,\right\rangle ds \tag{14}$$

$$= \int_0^T dt \int_{-\infty}^{+\infty} W(s)\sum_k w_{ik}\int_0^\infty \epsilon(s')\,\langle\,S_k(t-s')S_j(t-s)\,\rangle\,ds\,ds'. \tag{15}$$

We can split now the input spikes into those coming from synapse j and all the rest coming from other synapses $k \neq j$.

$$\langle\,\Delta w_{ij}\,\rangle = \int_0^T dt \int_{-\infty}^{+\infty} W(s)\int_0^\infty \epsilon(s')\Big[\sum_{k\neq j} w_{ik}\,\langle\,S_k(t-s')S_j(t-s)\,\rangle + w_{ij}\,\langle\,S_j(t-s')S_j(t-s)\,\rangle\Big]ds\,ds'$$

$$\tag{16}$$

Due to the independence of the presynaptic spike trains we have:
$\langle\,S_k(t-s')S_j(t-s)\,\rangle = \langle\,S_k(t-s')\,\rangle\,\langle\,S_j(t-s)\,\rangle = \nu_k\nu_j$.

As we found in exercise 3 of exercise set 11, the autocorrelation of a Poisson process is:
$\langle\,S_j(t-s')S_j(t-s)\,\rangle = \nu_j^2 + \nu_j\delta(s-s')$.

Thus we obtain:

$$\frac{\langle \Delta w_{ij} \rangle}{T} = \frac{1}{T} \int_0^T dt \int_{-\infty}^{+\infty} W(s) \int_0^\infty \epsilon(s') \Big[\sum_{k \neq j} w_{ik} \nu_k \nu_j + w_{ij} \nu_j^2 + w_{ij} \nu_j \delta(s - s') \Big] ds \; ds' \quad (17)$$

$$= \frac{1}{T} T \int_{-\infty}^{+\infty} W(s) \int_0^\infty \epsilon(s') \Big[\sum_k w_{ik} \nu_k \nu_j + w_{ij} \nu_j \delta(s - s') \Big] ds \; ds' \quad (18)$$

$$= \int_{-\infty}^{+\infty} W(s) \int_0^\infty \epsilon(s') \Big[\sum_k w_{ik} \nu_k \nu_j + w_{ij} \nu_j \delta(s - s') \Big] ds \; ds' \quad (19)$$

$$= \sum_k w_{ik} \nu_k \nu_j \int_{-\infty}^{+\infty} \int_0^\infty W(s) \epsilon(s') ds \; ds' + w_{ij} \nu_j \int_0^\infty W(s') \epsilon(s') ds'. \quad (20)$$