Computational Neuroscience
Dr. Sharba Bandyopadhyay
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology Kharagpur
Week – 10
Lecture – 49

Lecture 49 : BCM Rule

Welcome. So, we continue our lectures on long term potentiation and long term depression and implementation of those in models of neurons. So, we as we discussed in the previous lecture, we are taking only rates as responses in this case and not really considering timing of spikes in general and so and with that we are building towards more realistic models if possible with biological basis. So, the first two rules that we have discussed are the correlation based rules and the covariance based rules. So, interestingly as you saw, so let us just remind ourselves we have this rate of change of W with the time constant. This is equal to vu and when we average over multiple input patterns, then this translates into Q into W where Q is the correlation matrix.

And similarly in the covariance rule, there are two of them $WdWdt$ is equal to $vminusthetav$ which we will replace with the average activity of v times u or r is equal to $vtimesuminusthetau$ where the $thetau$ again will be taken as the average of u. And these turn out to be when we average over multiple input patterns u, we get the rule that is the covariance rule which is CW. So, what are the problems with these two? Both of these are unstable that is the weights sort of diverge. Another problem with the correlation rule is that there is no LTD here because it is continuously increasing because v and u are both positive.

And so that is why the threshold is introduced in the covariance rule and it is made the covariance by making it the average of the input activity for the input threshold and the average of the output activity for the output threshold. And here we can get LTD. But now if you observe in this particular case where in the covariance rule, we have $tauWdWdt$ the original form where we have $vminusthetav$ which is the average of v times u. So, in this case if we set v to be 0 that is there is no output activity no matter what the input activity is. So then that becomes problem in the sense that you still get LTD although it is seemingly correct in the sense that the input activity is not driving any output activity and so the synapse will depress and die.

But evidence shows experimental evidence shows that you require some post-synaptic activity in order to get long term depression LTD. So here momentarily

v is 0 and we would still get depression or LTD which is not correct in the sense that it is not supported with experimental evidence. And similarly we need both pre and postsynaptic activity to observe LTD and so here it is in the other case we have $vtimesuminusaverageofu$. So again here if we set u to be 0 then again we see that there is still LTD without any presynaptic activity. Now of course if u is always 0 that is fine but when there are periods when u is 0 and you still keep on getting LTD due to postsynaptic spiking that is not supported by evidence.

So we require both pre and postsynaptic activity. So the problem of not LTD not being there is removed by introducing a threshold from the Hebb's correlation based rule. And by introducing a threshold the problem that we are getting is that there is LTD without presynaptic or postsynaptic activity depending on which more version of the model we are using. And so that led to probably the most realistic such rate based plasticity models which is the Bienenstock Cooper and Munro learning rule or the BCM rule. So in this case what we see is that we set the $dwdt$ with this $tauw$ as the time constant and that is equal to our vu times $vminusthetav$.

So here the rule is very similar to the previous one but if in this case if let us say our v is 0 then our weights are not updated because it is also proportional to v itself and so our w remains constant. So this way we remove the requirement of I mean we remove the limitation of the covariance rule where we are using v explicitly another v explicitly to control for that lack of postsynaptic activity leading to depression because in such scenario w becomes a constant whatever it was. So in terms of the stability if we look at this particular rule then if we consider again as we did for our norm w square and how it changes with time then let us also multiply that with $ttauw$ and so that becomes our twice $vtauw$ and in that case we multiply it with w and $dwdt$ and so now we replace $dwdt$ with the term above and what we have is twice $tauw$ times w times vu times $vminusthetav$. So for simplicity we will remove the $tauw$ because it is simply a scalar and so what we have on the right hand side that is the derivative of the norm times the $tauw$ is w times or rather the $tauwtauw$ is actually removed because with the $tauwdwdt$ is the term that we have sorry. So we have w times v times u times $vminusthetav$.

So now remember that we have our $vequalsw$ dot product with u which is if we take because v is a constant we take it out this is w times u and we simply get here twice v square and we have $vminusthetav$ and this $thetav$ we will again take it as average of v that is if the activity is above the overall average then we will have a potentiation and if the activity is below the overall average that would lead to depression. Remember this $tauw$ is extremely I mean the time constant are over very large time scales that is we require a number of such events to actually see changes in w. So w changes at a very very long time scale that is it is extremely slow. So here we have we put in this average of v and so now if we do the time average of multiple of multiple patterns of this case then we will get the average change of the norm of w and so if we look at this equation what we have is average of so basically the $dwsquaredt$ turns out to be average on average turns out to be average of twice v square $vminusv$. So in this case if we take one v inside then we have and the two can come out we have v times v square minus v times average of v and average of that.

Now we know that v is greater than 0. So let us say v has a minimum value let us some particular $vnaught$ and so we can replace this v by this $vnaught$ and can say that this is going to be larger than that since v is positive and so ultimately our the norm square derivative depends on the average of v square minus v and average of v. So this average if we take the average in is essentially v square minus v times v and what we have here is v square minus v square which is simply the variance of v. So since the variance is positive variance of v that is of the output activity v square this is always greater than equal to 0 unless our v is constant then what we are seeing is again that the norm of the weight that is the length in this case let us say Euclidean norms and length of the weight keeps on increasing. So this being greater than 0 this is increasing and hence even the BCM learning rule in this form is unstable.

So this how do we get rid of this instability and again we have to remember that we have to base our changes and equations and whatever we derive based on realistic biological systems based on observations that we see. So what they actually did I mean there are multiple versions of this rule the BCM learning rule. So here we have again vu times $vminusthetav$ and what the rule adds on to this is that let there be a sliding threshold or a changing threshold. So in other words we introduced the $thetavdt$ with a time constant $theta$ which is v square minus $thetav$. So essentially what are we doing here so because having the $thetav$ fixed leads to instability.

Instability meaning that our length of the weight keeps on increasing unbounded or all of them I mean all the elements of the w reach some saturation value in terms of the plasticity even though there is LTD in there I mean LTD and LTP both are at play still the norm keeps on increasing. So what this does is by introducing this $theta\ v$ dynamics or the sliding threshold you meet the threshold change dynamically such that as the output activity increases that is gradually as the output activity increases our $theta\ v$ is made to go higher and higher. So this time constant of $tautheta$ is such that it is faster than the time constant of this

$tauw$ in the sense that over a period of time in which the w is changing that is we are playing multiple patterns and w is changing very slowly within this period $thetav$ is changing much faster and when w is getting finally updated or increased our $thetav$ has gone over multiple values depending on what the output activity has been in the in the over time. So as we see so if our if we have sort of a fixed output activity going on then $theta\ v$ would approach this v square value which is very large. So what are we then trying to do is basically so if let us say we have number of synapses or inputs and so we have $w1$, $w2$ up to wnb if we say that nb is the number of synapses that we have or the number of inputs that we have on to the postsynaptic neuron then if let us say some of the weights increase and make it make the output activity large.

Let us say so you are starting out with a flat set of weights and based on the input activity some of the weights are increased and the v became larger the output activity became larger and so the threshold is set to a larger value. With the larger value of the threshold what happens is that it gets more and more difficult for the system to get potentiation and in fact it leads to depression because of the $vminusthetav$ term here. And so it gets more difficult for every synapse in fact the weaker especially the weaker synapses to increase themselves and so this leads to a competition between the synapses. So that is those outputs that those inputs that are causing stronger outputs are weighted more and the other sort of synaptic weights are dying down gradually. So here we do not have the problem of unbounded growth of w because inherently we are putting in a threshold or rather the inherent properties of these equations is putting in a threshold based on the output activity that is stopping from the system from growing out of hand or because for the waves to diverge.

So because our $thetav$ will be going towards v square the more activity we get more input activity we get it is more likely that our $vminusthetav$ is negative and because our v square I mean v is positive v square is very large and so it will take a while for the system to get potentiation based on the inputs. And it is likely that the largest inputs the largest the weights that have increased the largest they might be able to drive it further forward. So this competition that is introduced between the synapses actually is important here and so which was missing in all the earlier models. So here ultimately in the BCM rule what we have is that we have both LTP and LTD based on both input and output activity. We remove the problem of zero activity based depression so this is removed.

We have stability in growth of synaptic weight which is an important factor which was not there in the earlier models. And on top of this we are introducing competition among the synapses. So this was I was saying that BCM rule it was the closest possible probably in terms of biological learning in with these kind of rate models in neurons because that in fact the more large output activity turns out to be the more difficult it gets for the neuron to increase its responses or increase the synaptic inputs. That is an observation which on the long time scale is also called homeostatic plasticity but the mechanisms of that are slightly different than what the way this has been presented in the BCM rule. So there are ways in which that or there biological mechanisms by which actually there is scaling of the synapses.

So here we are introducing a constraint on the increase in the weights based on the output activity. So there are learning rules that are based on direct constraints on the weights and they are more similar to the the homeostatic plasticity kind of rules where where we observe some maintenance of activity at an average level and that requires direct constraints on the weights. So we will be discussing those direct constraint based models or implementation of synaptic plasticity in our subsequent lectures. Thank you.