Inferring Synaptic Plasticity Rule from Change of Population Activity

Jizheng Dong

Sep 10 2019

1 Background

Information about external stimuli is thought to be stored in cortical circuits through the change of synaptic connectivity. When a particular stimulus is repeatedly encountered, the modifications of network connectivity would lead to changes in neuronal activity. Here we ask what plasticity rules are consistent with the differences in the statistics of visual response to novel and familiar stimuli in the inferior temporal cortex, an area underlying the visual object recognition.

We refer the learning rule of synaptic plasticity using the statistic and mathematics methods. We want to infer the dependence of the presumptive learning rule on postsynaptic firing rate and the inferred learning rule is appropriate for the real situation.

2 Inferring synaptic plasticity rule

2.1 Introduction

Here we consider a faring rate model with a plasticity rule that modifies the strength of recurrent synapses as a function of the firing rate of pre- and postsynaptic neurons.

The spikes of neurons are caused by the activities of other neurons. We describe this progress as the firing rate of neuron r_i is determined by its input h_i from other neurons via a transfer function.

$$r_i = \Phi_i(h_i) \tag{1}$$

r_i: the firing rate of neuron i with $i = 1, \ldots, N$
N: the number of neurons in the network
h_i: the inputs of neuron i
Φ_i: the transfer function($f - I$ curve)

The input current h_i is the sum of the external input I_{iX} and the recurrent input, which is the sum of presynaptic firing rates r_j, weighted by the synaptic strength W_{ij}.

$$h_i = I_{iX} + \sum_{j=1}^{N} W_{ij} r_j \tag{2}$$

$$r_i + \Delta r_i = \Phi_i(h_i + \Delta h_i) \tag{3}$$

Δr can be attained by comparing the firing rate when a monkey faced the novel and familiar stimuli (the passive viewing task and the dimming-detection task, both tasks include novel and familiar stimuli)

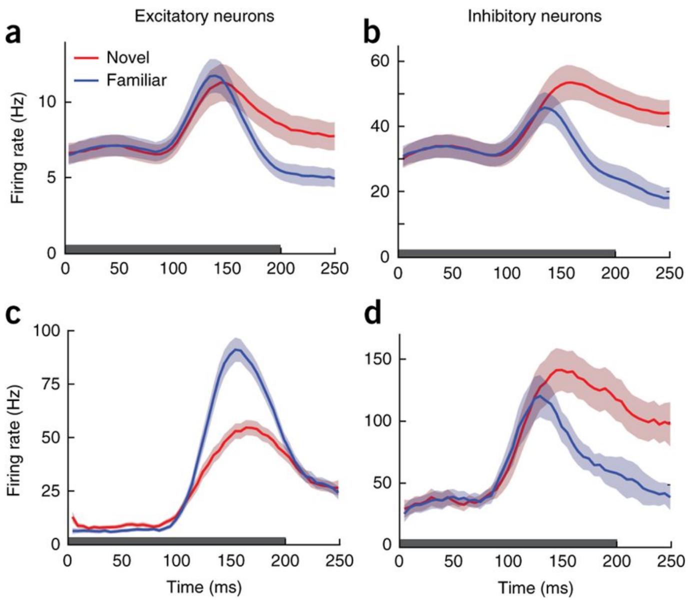


Figure 1: A visual response of inferior temporal cortical (ITC) neurons to novel and familiar stimuli.
a b : mean of firing rate.
c d : maximal of firing rate

Assumption 1 Here we assume that changes in network response are primarily due to changes in recurrent synapses. This assumption is justified by the observation that differences between responses to familiar and novel stimuli start to emerge a few tens of milliseconds after the activity onset(Figure 1).

It is the I_{iX} in the equation that leads to neuronal firings in our recurrent network. The firing rate will change during a few tens of milliseconds after the activity onset.

$$h_i + \Delta h_i = I_{iX} + \sum_{j=1}^{N} \left(W(r_i, r_j) + \Delta W(r_i, r_j)\right) r_j \tag{4}$$

After the subtraction of equation (2) and (4), we can get

$$\Delta h_i = \sum_{j=1}^{N} \Delta W(r_i, r_j) r_j \tag{5}$$

2.2 Synaptic plasticity rule

Assumption 2 We assume that the learning rule is a separate function of pre- and postsynaptic rates

$$\Delta W_{ij} = \Delta W(r_i, r_j) = \alpha f(r_i) g(r_j) \tag{6}$$

f: refer to presynaptic neurons
g: refer to postsynaptic neurons
At first, we set $\alpha = 1$.

As we can see, many classical neural plasticity have adopted this assumption.

The Basic Hebb Rule

$$\tau \frac{dW_{ij}}{dt} = r_i \, r_j \tag{7}$$

The Covariance Rule

$$\tau \frac{dW_{ij}}{dt} = (r_i - \theta_{r_i}) \, r_j \tag{8}$$

$$\tau \frac{dW_{ij}}{dt} = r_i \, (r_j - \theta_{r_j}) \tag{9}$$

The BCM Rule

$$\tau \frac{dW_{ij}}{dt} = r_i \, r_j \, (r_j - \theta_{r_j}) \tag{10}$$

Those rules are all theory-driven methods. Can we find a data-driven method? The answer is yes. Sukbin Lim has done very beautiful work(S. Lim 2015).

But a new problem occurs. If we assume that the connection between neurons is all-to-all, the sum of postsynaptic neurons is constant. That is

$$\sum_{j=1}^{N} g(r_j) r_j = \mathsf{Const} \tag{11}$$

$$f(r_i) = \frac{\Delta h_i}{\sum_{j=1}^{N} g(r_j) r_j} \tag{12}$$

So that f_i will only be a deformed function of Δh_i. If we have a plot, we will find that they have a similar shape.

Besides, Δh_i primarily depends on Δr_i, so this rule is mainly determined by the Δr_i.

2.3 Add a random adjacent matrix

If we want to get a more realistic expression of f and g in the situation that neurons are connected randomly, we can create an adjacent matrix to show the connection of neurons, using 0 and 1 to refer to connection status. The connected probability is p.

Here is the adjacent matrix C

$$C = (c_{ij})_{i,j=1:N} \tag{13}$$

with

$$c_{ij} \in 0, 1 \tag{14}$$

$C_{ij} = 1$ means that neuron j is connected to neuron i, meanwhile, 0 means no connection. Therefore, original equations have changed to

$$\Delta h_i = f(r_i) \sum_{j=1}^{N} C_{ij} g(r_j) r_j \tag{15}$$

As a consequence, $\sum_{j=1}^{N} C_{ij} g(r_j) r_j$ is different for different neurons.

2.4 Add restriction to the problem

We can get the change of firing rate Δr_i when monkeys are performing two different tasks. But how can we get f and g or $f(r_i)$ and $g(r_i)$ for every r_i? It only has N equations, but has $2N$ unknowns.

N equations:

$$\begin{cases} \Delta h_1 & = f(r_1) \sum_{j=1}^{N} C_{1j} g(r_j) r_j \\ \vdots \\ \Delta h_i & = f(r_i) \sum_{j=1}^{N} C_{ij} g(r_j) r_j \\ \vdots \\ \Delta h_N & = f(r_N) \sum_{j=1}^{N} C_{Nj} g(r_j) r_j \end{cases} \tag{16}$$

$2N$ unknowns:

$$f = \begin{bmatrix} f(r_1) \\ \vdots \\ f(r_i) \\ \vdots \\ f(r_N) \end{bmatrix} \qquad g = \begin{bmatrix} g(r_1) \\ \vdots \\ g(r_i) \\ \vdots \\ g(r_N) \end{bmatrix} \tag{17}$$

It is impossible to solve this problem unless we have other restrictions. We must add more restrictions to get f and g.

We simplify the denote of f and g as follows:

$$\begin{bmatrix} f(r_1) \\ \vdots \\ f(r_i) \\ \vdots \\ f(r_N) \end{bmatrix} = \begin{bmatrix} f_1 \\ \vdots \\ f_i \\ \vdots \\ f_N \end{bmatrix} \qquad \begin{bmatrix} g(r_1) \\ \vdots \\ g(r_i) \\ \vdots \\ g(r_N) \end{bmatrix} = \begin{bmatrix} g_1 \\ \vdots \\ g_i \\ \vdots \\ g_N \end{bmatrix} \tag{18}$$

Here we want get more smooth f and g. That is to minimize

$$\sum_{i=1}^{N-1} \left(\frac{f_{i+1} - f_i}{r_{i+1} - r_i} \right)^2 + \sum_{i=1}^{N-1} \left(\frac{g_{i+1} - g_i}{r_{i+1} - r_i} \right)^2 \tag{19}$$

and we consider the upper term as a function of $g_1, \cdots, g_N$

$$H(g_1, \cdots, g_N) = \sum_{i=1}^{N-1} \left(\frac{f_{i+1} - f_i}{r_{i+1} - r_i} \right)^2 + \sum_{i=1}^{N-1} \left(\frac{g_{i+1} - g_i}{r_{i+1} - r_i} \right)^2 \tag{20}$$

with

$$f_i = \frac{\Delta h_i}{\sum_{j=1}^{N} C_{ij} g_j r_j} \tag{21}$$

that is

$$H(g_1, \cdots, g_N) = \sum_{i=1}^{N-1} \left(\frac{\frac{\Delta h_{i+1}}{\sum_{j=1}^{N} C_{i+1,j} g_j r_j} - \frac{\Delta h_i}{\sum_{j=1}^{N} C_{ij} g_j r_j}}{r_{i+1} - r_i} \right)^2 + \sum_{i=1}^{N-1} \left(\frac{g_{i+1} - g_i}{r_{i+1} - r_i} \right)^2 \tag{22}$$

To minimize $H(g_1, \cdots, g_N)$, we can make

$$\frac{\partial H(g_1, \ldots, g_N)}{\partial g_k} = 0 \tag{23}$$

for $k = 1, \ldots, N$.
Then we will have $2N$ equations and $2N$ unknowns.

2.5 Using mathematical method to solve the problem

Now we have $2N$ equations and $2N$ unknowns. But it is difficult to get the exact result. Before solving this problem, we can simplify those equations.

2.5.1 Using $Taylor$ expansion to simplify cost function $H(g)$

$Taylor$ expansion:

$$\begin{aligned}\frac{1}{x+a} &= \frac{1}{a} - \frac{x}{a^2} + \frac{x^2}{a^3} + \cdots \\ &= \frac{1}{a}\left(1 - \frac{x}{a} + \frac{x^2}{a^2}\right) + \cdots\end{aligned} \tag{24}$$

and

$$\begin{aligned}\frac{1}{x^2+2ax+a^2} &= \frac{1}{a^2} - \frac{2x}{a^3} + \frac{3x^2}{a^4} + \cdots \\ &= \frac{1}{a^2}\left(1 - \frac{2x}{a} + \frac{3x^2}{a^2}\right) + \cdots\end{aligned} \tag{25}$$

as $x << a$.

Using those two approximate equations, with

$$x = C_{ik} g_k r_k \tag{26}$$

or

$$x = C_{i+1,k} g_k r_k \tag{27}$$

and

$$a = \sum_{j \neq k} C_{ij} g_j r_j \tag{28}$$

$$a_1 = \sum_{j \neq k} C_{i+1,j} g_j r_j \tag{29}$$

we can get

$$\begin{aligned}H_k(g_1, \cdots, g_N) &= \sum_{i=1}^{n-1} \frac{1}{(r_{i+1}-r_i)^2}\left(\frac{\Delta h_{i+1}^2}{a_1^2}(1 - \frac{-2C_{i+1,k} g_k r_k}{a_1})\right) \\ &+ \sum_{i=1}^{n-1} \frac{1}{(r_{i+1}-r_i)^2}\left(\frac{\Delta h_i^2}{a^2}(1 - \frac{-2C_{ik} g_k r_k}{a})\right) \\ &- \sum_{i=1}^{n-1} \frac{2}{(r_{i+1}-r_i)^2}\frac{\Delta h_i}{a}\frac{\Delta h_{i+1}}{a_1}\left(1 - \frac{C_{i+1,k} g_k r_k}{a_1} - \frac{C_{ik} g_k r_k}{a}\right) \\ &+ \frac{(g_{k+1}-g_k)^2}{(r_{k+1}-r_k)^2} + \frac{(g_k - g_{k-1})^2}{(r_k - r_{k-1})^2} + \cdots\end{aligned} \tag{30}$$

2.5.2 Taking the derivative of $H(g)$ of g_k

Then we take the derivative of $H_k(g_1, \ldots, g_n)$ of g_k and make the derivative equal to 0

$$\frac{\partial H_k(g_1, \ldots, g_n)}{\partial g_k} = 0 \tag{31}$$

$$\begin{aligned}
&\sum_{i=1}^{n-1} \frac{1}{(r_{i+1} - r_i)^2} \left(\frac{\Delta h_{i+1}^2}{a_1^2} \frac{-2C_{i+1,k} r_k}{a_1} \right) \\
&+ \sum_{i=1}^{n-1} \frac{1}{(r_{i+1} - r_i)^2} \left(\frac{\Delta h_i^2}{a^2} \frac{-2C_{ik} r_k}{a} \right) \\
&- \sum_{i=1}^{n-1} \frac{2}{(r_{i+1} - r_i)^2} \frac{\Delta h_i}{a} \frac{\Delta h_{i+1}}{a_1} \left(-\frac{C_{i+1,k} r_k}{a_1} - \frac{C_{ik} r_k}{a} \right) \\
&+ \frac{g_k - g_{k+1}}{(r_{k+1} - r_k)^2} + \frac{g_k - g_{k-1}}{(r_k - r_{k-1})^2} = 0
\end{aligned} \tag{32}$$

2.5.3 Using the iteration method to solve the nonlinear equations

We will get nonlinear equations of $g_1, \ldots, g_n$. Then we will use the iterative method to solve the nonlinear equations.

The iterative method is giving $g_1, \ldots, g_n$ an initial value

$$g_i^{(0)} = c \tag{33}$$

and then using the iterative formula to calculate and update the value

$$g_i^{(m+1)} = G_i^{(m)}(g_1^{(m)}, \ldots, g_n^{(m)}) \tag{34}$$

After every calculation, the value $g_1, \ldots, g_n$ will change.

For example, we use the iterative formula as following.
The initial value

$$\Delta W_{ij} = \alpha \, (r_i - \bar{r}) \, (\sum_{j=1}^{n} C_{ij} (r_i - \bar{r}) r_j) \tag{35}$$

$$\alpha = 1 \tag{36}$$

$$f_i = r_i - \bar{r} \tag{37}$$

$$g_j = r_j - \bar{r} \tag{38}$$

$$\bar{r} = mean(r) \tag{39}$$

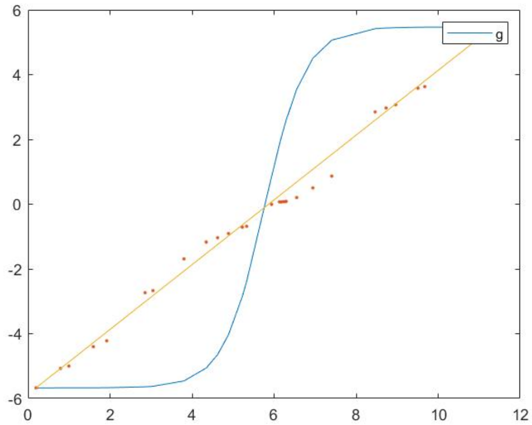
and $g^{(0)}$ is a sigmoid function

$$g_i^{(0)} = g(r_i)^{(0)} = \frac{\max(r) - \min(r)}{1 + \exp(\frac{\max(r) + \min(r)}{2} - r_i)} + \min(r) - mean(r) \tag{40}$$

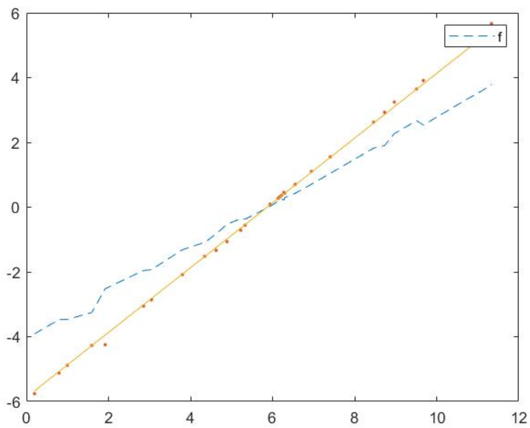
The iterative formula

$$
\begin{aligned}
g_k^{(m+1)} &= \frac{(r_{k+1}-r_k)^2(r_k-r_{k-1})^2}{(r_{k+1}-r_k)^2+(r_k-r_{k-1})^2}\left(\frac{g_{k+1}^{(m)}}{(r_{k+1}-r_k)^2}+\frac{g_{k-1}^{(m)}}{(r_k-r_{k-1})^2}\right)\\
&+ \frac{(r_{k+1}-r_k)^2(r_k-r_{k-1})^2}{(r_{k+1}-r_k)^2+(r_k-r_{k-1})^2}\left(\sum_{i=1}^{N-1}\frac{1}{(r_{i+1}-r_i)^2}\frac{\Delta h_{i+1}^2}{(a_1^{(m)})^2}\frac{2C_{i+1,k}r_k}{a_1^{(m)}}\right.\\
&+ \sum_{i=1}^{N-1}\frac{1}{(r_{i+1}-r_i)^2}\frac{\Delta h_i^2}{(a^{(m)})^2}\frac{2C_{ik}r_k}{a^{(m)}}\\
&\left.- \sum_{i=1}^{N-1}\frac{2}{(r_{i+1}-r_i)^2}\frac{\Delta h_i}{a^{(m)}}\frac{\Delta h_{i+1}}{a_1^{(m)}}\left(\frac{C_{i+1,k}r_k}{a_1^{(m)}}+\frac{C_{ik}r_k}{a^{(m)}}\right)\right)
\end{aligned}
\tag{41}
$$

Here is one of the results



(a) g



(b) f

The yellow solid line is the desired result

$$\Delta W_{ij} = f_i g_j$$

$$g(r_i) = r_i - mean(r)$$

The blue dotted line is calculated data using initially set the parameter as

$$\Delta W_{ij} = f(r_i)g(r_j)$$

$$g(r_j)^{(0)} = \frac{max(r) - min(r)}{1 + \exp(\frac{max(r)+min(r)}{2} - r_j)} + min(r) - mean(r)$$

The red point is the result after iteration.

We can see that the iterative result of g is very similar to the desired result. It is a linear function. And the iterative result of f is also linear in this situation.

2.6 Additional benefit

Before any stimulus is given, the input of neuron i is

$$I^{(1)} = \Phi(Wr^{(1)}) \tag{42}$$

If we give a random stimulus, it will lead to the first change of synapse

$$\Delta W_{ij}^{(1)} \tag{43}$$

If we give anther totally different stimulus, what will happen?
It will lead to another neural firing and the firing rate subjects to a distribution

$$\begin{aligned}
\Delta W^{(1)} r^{(2)} &= f(r_j^{(1)}) \sum_{j=1}^{N} g(r_j^{(1)}) r_j^{(2)} \\
&= f(r_j^{(1)}) \sum_{j=1}^{N} \left(r_j^{(1)} - mean(r^{(1)}) \right) r_j^{(2)} \\
&= f(r_j^{(1)}) \sum_{j=1}^{N} \left(r_j^{(1)} r_j^{(2)} - mean(r^{(1)}) r_j^{(2)} \right) \\
&= f(r_j^{(1)}) \left(\sum_{j=1}^{N} r_j^{(1)} r_j^{(2)} - mean(r^{(1)}) \sum_{j=1}^{n} r_j^{(2)} \right)
\end{aligned} \tag{44}$$

In general, the expected value operator is not multiplicative, i.e. $E[r^{(1)}r^{(2)}]$ is not necessarily equal to $E[r^{(1)}] \cdot E[r^{(2)}]$. However, if $r^{(1)}$ and $r^{(2)}$ are independent, then

$$E(r^{(1)}r^{(2)}) = E(r^{(1)})E(r^{(2)}) \tag{45}$$

$$\Delta W^{(1)} r^{(2)} = 0$$

If we have two totally different stimuli, which will lead to the different reaction of the recurrent network. If the firing rate is subjected to independent distribution, the first change of synapse will have no influence on the second first rate.

2.7 The improvement for the convergence of iterative method

We find that

$$\frac{1}{(r_{i+1} - r_i)^2} \tag{46}$$

in the iterative formula will cause instability, so the result will not be convergent to the desired result.

$$
\begin{aligned}
g_k^{(m+1)} &= \frac{(r_{k+1}-r_k)^2(r_k-r_{k-1})^2}{(r_{k+1}-r_k)^2+(r_k-r_{k-1})^2}\left(\frac{g_{k+1}^{(m)}}{(r_{k+1}-r_k)^2}+\frac{g_{k-1}^{(m)}}{(r_k-r_{k-1})^2}\right) \\
&+\frac{(r_{k+1}-r_k)^2(r_k-r_{k-1})^2}{(r_{k+1}-r_k)^2+(r_k-r_{k-1})^2}\Big(\sum_{i=1}^{N-1}\frac{1}{(r_{i+1}-r_i)^2}\frac{\Delta h_{i+1}^2}{(a_1^{(m)})^2}\frac{2C_{i+1,k}r_k}{a_1^{(m)}} \\
&+\sum_{i=1}^{N-1}\frac{1}{(r_{i+1}-r_i)^2}\frac{\Delta h_i^2}{(a^{(m)})^2}\frac{2C_{ik}r_k}{a^{(m)}} \\
&-\sum_{i=1}^{N-1}\frac{2}{(r_{i+1}-r_i)^2}\frac{\Delta h_i}{a^{(m)}}\frac{\Delta h_{i+1}}{a_1^{(m)}}\Big(\frac{C_{i+1,k}r_k}{a_1^{(m)}}+\frac{C_{ik}r_k}{a^{(m)}}\Big)\Big)
\end{aligned} \tag{47}
$$

Therefore, we set a threshold *eps* for the $(r_{i+1}-r_i)^2$. If $(r_{i+1}-r_i)^2 < eps$, we make it equal to a fixed value.

Besides, the α in the

$$\Delta W(r_i, r_j) = \alpha f(r_i)g(r_j) \tag{48}$$

will have an influence on the result through changing the Δh_i

$$\Delta h_i = \alpha f(r_i)\Big(\sum_{j=1}^{n} C_{ij}g(r_j)r_j\Big) \tag{49}$$

After many trials, we find it will induce a reasonable result when

$$\alpha \approx \frac{1}{N} \tag{50}$$

The probability of convergence will apparently increase after the change of α.

3 Acknowledgments

I would first like to thank my advisor Prof. Lim for steering me in the right direction whenever she thought I needed it. She was always patient and affable when we talked.

I would also like to thank Jintao and Xikang. They inspired me many times during our discussion about the research.

Finally, I would also like to acknowledge my friends Yiqing, Zhenni, Mace, Siri, Liliu, Jingjie for their companionship in ECNU.