

Laplacian envelope

Subhaneil Lahiri

May 29, 2015

Abstract

We try to find the continuous time Markov process that has the maximal Laplace transformed signal-to-noise curve.

Contents

1	Framework	1
1.1	Recognition memory	1
1.2	Signal-to-noise ratio	2
1.3	Markov models of synapses	3

1 Framework

1.1 Recognition memory

We will be trying to store patterns in a set of N synaptic weights, \vec{w} . Every time we try to store a pattern, these synapses are subjected to a plasticity event where each synapse is either potentiated or depressed, depending on the pattern. we will assume that these patterns are spatially and temporally independent.

At some time, suppose we wish to determine if a given pattern is one of those that we previously attempted to store. We wish to answer this question by looking at the synaptic weights directly (ideal observer). For that given pattern there will be an ideal set of synaptic weights, \vec{w}_{ideal} , where those synapses that were supposed to be potentiated are maximised and those that were supposed to be depressed are minimised. Suppose that the given pattern was actually seen at time 0 and we are observing the synapses at time t . The actual set of synaptic weights we see, $\vec{w}(t)$, will be a vector of random variables that differs from \vec{w}_{ideal} due to the stochasticity of the pattern encoding and all of the other (uncorrelated) pattern that are stored after it. As $t \rightarrow \infty$, the synaptic weights will become independent of the pattern stored at $t = 0$. Thus, the vector of random variables $\vec{w}(\infty)$ also describes the synaptic weights under the null hypothesis – if that given pattern had never been stored.

We can test if the pattern had been previously stored by computing $\vec{w}_{\text{ideal}} \cdot \vec{w}$ and comparing it to some threshold. For large N , this quantity will have a Gaussian distribution. There will be a ROC curve as a function of this threshold:

$$\begin{aligned} \text{TPR} &= \Phi_c \left(\frac{\Phi_c^{-1}(\text{FPR}) - \text{SNR}}{\text{NNR}} \right), \quad \text{where} \quad \Phi_c(x) = \int_x^\infty \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz, \\ \text{SNR} &= \frac{\langle \vec{w}_{\text{ideal}} \cdot \vec{w}(t) \rangle - \langle \vec{w}_{\text{ideal}} \cdot \vec{w}(\infty) \rangle}{\sqrt{\text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty))}}, \quad (1) \quad \boxed{\text{eq:ROC}} \\ \text{NNR} &= \sqrt{\frac{\text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(t))}{\text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty))}}, \end{aligned}$$

and TPR/FPR are the true/false positive rates. When the signal-to-noise ratio, SNR is larger than $\Phi_c^{-1}(\text{FPR})$, it is beneficial to decrease the noise-to-noise ratio, NNR. In the other case, it is beneficial to increase it. The expectations and variances are over the probability distribution of the synaptic states given the sequence of plasticity events, the probability distribution of the sequence of plasticity events themselves and the probability distribution of the pattern we are testing, \vec{w}_{ideal} .

The formulae above assumed that we know the time between storage and recognition. We should also compute the expectations and variances over probability distribution of the recall time, t , as well. If we only know average time, τ , a natural choice for this distribution is

$$P(t|\tau) = \frac{e^{-t/\tau}}{\tau}. \quad (2) \quad \boxed{\text{eq:recogtime}}$$

Different parts of the brain, that store memories for different timescales, could be characterised by different values of τ .

1.2 Signal-to-noise ratio

sec:snr

We will model the synapses as having two possible synaptic weights. As we are looking at the information contained in the synaptic weights, rather than modelling the readout via electrical activity of the neurons, what values the synaptic weights actually take is irrelevant. It will be convenient to call these two values ± 1 .

As discussed in §1.1, the signal-to-noise ratio is given by

$$\text{SNR}(t) = \frac{\langle \vec{w}_{\text{ideal}} \cdot \vec{w}(t) \rangle - \langle \vec{w}_{\text{ideal}} \cdot \vec{w}(\infty) \rangle}{\sqrt{\text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty))}}, \quad (3) \quad \boxed{\text{eq:SNRdef}}$$

Where the averages and variances are over the probability distribution of the synaptic states given the sequence of plasticity events, the probability distribution of the sequence of plasticity events themselves and the probability distribution of the pattern we are testing, \vec{w}_{ideal} . If we also average over the time between storage and recognition, we have

$$\overline{\text{SNR}}(\tau) = \int_0^\infty dt P(t|\tau) \text{SNR}(t) = \frac{1}{\tau} \int_0^\infty dt e^{-t/\tau} \text{SNR}(t), \quad (4) \quad \boxed{\text{eq:snrb}}$$

This is similar to the average SNR up to time τ , but with the hard cut-off replaced with a smooth exponential. This can be expressed in terms of the Laplace transform of the SNR curve in (3):

$$A(s) = \int_0^\infty dt e^{-st} \text{SNR}(t). \quad \overline{\text{SNR}}(\tau) = \frac{A(1/\tau)}{\tau}, \quad (5) \quad \text{eq:laplace}$$

First, let's look at the variances, remembering that the states and plasticity events of each synapse are independent and identically distributed:

$$\begin{aligned} \text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(t)) &= \sum_{\alpha\beta} \left\langle \vec{w}_{\text{ideal}}^\alpha \vec{w}^\alpha(t) \vec{w}_{\text{ideal}}^\beta \vec{w}^\beta(t) \right\rangle - \left(\sum_{\alpha} \langle \vec{w}_{\text{ideal}}^\alpha \vec{w}^\alpha(t) \rangle \right)^2 \\ &= \sum_{\alpha} \langle (\vec{w}_{\text{ideal}}^\alpha)^2 (\vec{w}^\alpha(t))^2 \rangle + \sum_{\alpha \neq \beta} \langle \vec{w}_{\text{ideal}}^\alpha \vec{w}^\alpha(t) \rangle \langle \vec{w}_{\text{ideal}}^\beta \vec{w}^\beta(t) \rangle \\ &\quad - \left(\sum_{\alpha} \langle \vec{w}_{\text{ideal}}^\alpha \vec{w}^\alpha(t) \rangle \right)^2 \\ &= N \langle 1 \rangle + N(N-1) \langle \vec{w}_{\text{ideal}}^1 \vec{w}^1(t) \rangle^2 - N^2 \langle \vec{w}_{\text{ideal}}^1 \vec{w}^1(t) \rangle^2 \\ &= N(1 - \langle \vec{w}_{\text{ideal}}^1 \vec{w}^1(t) \rangle^2), \end{aligned} \quad (6) \quad \text{eq:noise}$$

where we used $\vec{w}^\alpha = \pm 1$. We can compute $\text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty))$ by taking $t \rightarrow \infty$.

For the numerator, we can write

$$\langle \vec{w}_{\text{ideal}} \cdot \vec{w}(t) \rangle = \sum_{\alpha} \langle \vec{w}_{\text{ideal}}^\alpha \vec{w}^\alpha(t) \rangle = N \langle \vec{w}_{\text{ideal}}^1 \vec{w}^1(t) \rangle, \quad (7) \quad \text{eq:overlap}$$

If the elements of \vec{w}_{ideal} take values ± 1 with probability $f^{\text{pot/dep}}$,

$$\langle \vec{w}_{\text{ideal}}^1 \vec{w}^1(t) \rangle = f^{\text{pot}} \langle \vec{w}^1(t) \rangle_{\text{pot}, t=0} - f^{\text{dep}} \langle \vec{w}^1(t) \rangle_{\text{dep}, t=0} \quad (8) \quad \text{eq:overlap}$$

To compute this quantity, we will need to discuss how we model individual synapses.

1.3 Markov models of synapses

We model a synapse as having M internal states. The synaptic weight depends deterministically on the state, given by the M -element column vector \mathbf{w} . We denote the probability distribution across these states by the row vector $\mathbf{p}(t)$. We denote a column vector of ones by \mathbf{e} , so the normalisation condition is $\mathbf{p}(t)\mathbf{e} = 1$.

When the synapse is subjected to a plasticity event the internal state will change stochastically, described by the matrices $\mathbf{M}^{\text{pot/dep}}$ whose i, j elements are the transition probabilities from state i to state j . In between these plasticity events the synapse will undergo stochastic transitions described by a matrix \mathbf{H} whose i, j element is the transition rate from state i to state j for $i \neq j$ and $\mathbf{H}_{ii} = -\sum_{j \neq i} \mathbf{H}_{ij}$. This allows us to describe homeostatic plasticity as well as plasticity processes that occur over a period of time –

$\mathbf{M}^{\text{pot/dep}}$ puts the synapse into some state where the process \mathbf{H} takes over and does the rest. The matrices $\mathbf{M}^{\text{pot/dep}}$ are transition matrices of discrete time Markov processes, whereas \mathbf{H} is the transition matrix of a continuous time Markov process. Their elements must satisfy the following constraints

$$\begin{aligned} \mathbf{M}_{ij}^{\text{pot/dep}} &\geq 0, & \sum_j \mathbf{M}_{ij}^{\text{pot/dep}} &= 1, \\ \mathbf{H}_{i \neq j} &\geq 0, & \sum_j \mathbf{H}_{ij} &= 0. \end{aligned} \tag{9} \quad \text{eq:constr}$$

The upper bounds $\mathbf{M}_{ij}^{\text{pot/dep}} \leq 1$ and $\mathbf{H}_{ii} \leq 0$ follow automatically from these. There is no upper bound on $\mathbf{H}_{i \neq j}$ (or lower bound on \mathbf{H}_{ii}) as these are rates rather than probabilities.

Given a sequence of plasticity events at times t_1, t_2, t_3, \dots , the probability distribution across the internal states is

$$\mathbf{p}(t_n) = \mathbf{p}(t_0) e^{\mathbf{H}(t_1-t_0)} \mathbf{M}^{\text{pot}} e^{\mathbf{H}(t_2-t_1)} \mathbf{M}^{\text{dep}} e^{\mathbf{H}(t_3-t_4)} \mathbf{M}^{\text{dep}} \dots \mathbf{M}^{\text{pot}} e^{\mathbf{H}(t_n-t_{n-1})}. \tag{10} \quad \text{eq:plastseq}$$

In computing the expectation in (8) the plasticity event that occurs at $t = 0$ is fixed, but all of the previous and subsequent events are marginalised over. If the plasticity events occur at Poisson rate r , with potentiation and depression occurring independently with probability $f^{\text{pot/dep}}$, averaging over all previous and subsequent events gives

$$P(\text{state} = i, t \mid \text{pot/dep}, 0) = \left[\mathbf{p}^\infty \mathbf{M}^{\text{pot/dep}} e^{r(f^{\text{pot}} \mathbf{M}^{\text{pot}} + f^{\text{dep}} \mathbf{M}^{\text{dep}} - \mathbf{I})t + \mathbf{H}t} \right]_i, \tag{11} \quad \text{eq:plastseq}$$

where \mathbf{p}^∞ is the steady state distribution of the continuous time transition matrix that appears in the exponential:

$$\mathbf{p}^\infty (r \mathbf{W}^F + \mathbf{H}) = 0, \quad \text{where} \quad \mathbf{W}^F = f^{\text{pot}} \mathbf{M}^{\text{pot}} + f^{\text{dep}} \mathbf{M}^{\text{dep}} - \mathbf{I}. \tag{12} \quad \text{eq:eq}$$

This results in

$$\begin{aligned} \langle \vec{w}_{\text{ideal}}^1 \vec{w}^1(t) \rangle &= \mathbf{p}^\infty (f^{\text{pot}} \mathbf{M}^{\text{pot}} - f^{\text{dep}} \mathbf{M}^{\text{dep}}) e^{(r \mathbf{W}^F + \mathbf{H})t} \mathbf{w}, \\ \langle \vec{w}_{\text{ideal}}^1 \vec{w}^1(\infty) \rangle &= \mathbf{p}^\infty (f^{\text{pot}} \mathbf{M}^{\text{pot}} - f^{\text{dep}} \mathbf{M}^{\text{dep}}) \mathbf{e} \mathbf{p}^\infty \mathbf{w} \\ &= \mathbf{p}^\infty (f^{\text{pot}} \mathbf{e} - f^{\text{dep}} \mathbf{e}) \mathbf{p}^\infty \mathbf{w} \\ &= (f^{\text{pot}} - f^{\text{dep}}) \mathbf{p}^\infty \mathbf{w} \\ &= (f^{\text{pot}} - f^{\text{dep}}) \mathbf{p}^\infty e^{(r \mathbf{W}^F + \mathbf{H})t} \mathbf{w}. \end{aligned} \tag{13} \quad \text{eq:overlapwt}$$

Combining these allows us to write the numerator of (3) as

$$\begin{aligned} \langle \vec{w}_{\text{ideal}} \cdot \vec{w}(t) \rangle - \langle \vec{w}_{\text{ideal}} \cdot \vec{w}(\infty) \rangle &= N \mathbf{p}^\infty \mathbf{K} e^{(r \mathbf{W}^F + \mathbf{H})t} \mathbf{w}, \\ \text{where} \quad \mathbf{K} &= f^{\text{pot}} (\mathbf{M}^{\text{pot}} - \mathbf{I}) - f^{\text{dep}} (\mathbf{M}^{\text{dep}} - \mathbf{I}) \end{aligned} \tag{14} \quad \text{eq:signal}$$

Combining with (6) gives

$$\text{SNR}(t) = \frac{\sqrt{N} \mathbf{p}^\infty \mathbf{K} e^{(r \mathbf{W}^F + \mathbf{H})t} \mathbf{w}}{\sqrt{1 - (f^{\text{pot}} - f^{\text{dep}})^2 (\mathbf{p}_+^\infty - \mathbf{p}_-^\infty)^2}}. \tag{15} \quad \text{eq:SNRcurveF}$$

We can absorb \mathbf{H} into $\mathbf{Q}^{\text{pot/dep}}$ by defining

$$\begin{aligned}\widetilde{\mathbf{M}}^{\text{pot/dep}} &= \mathbf{M}^{\text{pot/dep}} + \frac{\mathbf{H}}{2r f^{\text{pot/dep}}}, & \widetilde{\mathbf{W}}^{\text{F}} &= f^{\text{pot}} \widetilde{\mathbf{M}}^{\text{pot}} + f^{\text{dep}} \widetilde{\mathbf{M}}^{\text{dep}} - \mathbf{I}, \\ & & \widetilde{\mathbf{K}} &= f^{\text{pot}} (\widetilde{\mathbf{M}}^{\text{pot}} - \mathbf{I}) - f^{\text{dep}} (\widetilde{\mathbf{M}}^{\text{dep}} - \mathbf{I}).\end{aligned}\tag{16}$$

eq:absorbhom

Then the SNR curve is

$$\text{SNR}(t) = \frac{\sqrt{N} \mathbf{p}^\infty \widetilde{\mathbf{K}} e^{rt} \widetilde{\mathbf{W}}^{\text{F}} \mathbf{w}}{\sqrt{1 - (f^{\text{pot}} - f^{\text{dep}})^2 (\mathbf{p}_+^\infty - \mathbf{p}_-^\infty)^2}}.\tag{17}$$

eq:SNRcurveh

The denominator will not play any role in what follows, as the models that maximize the various measures of memory performance all have some sort of balance between potentiation and depression, either with $f^{\text{pot}} = f^{\text{dep}}$ or $\mathbf{p}_+^\infty = \mathbf{p}_-^\infty$. We can perform the maximisation in two steps. First maximise the numerator at fixed $\mathbf{p}^\infty \mathbf{w}$. Then maximise the ratio wrt. $\mathbf{p}^\infty \mathbf{w}$. This will allow us to ignore the denominator until the very end.

This formula takes the same form as the one in our NIPS paper [1], where the process \mathbf{H} was not included. However, the process \mathbf{H} still has an impact due to the constraints.

The constraints on $\widetilde{\mathbf{M}}^{\text{pot/dep}}$ are less stringent than the constraints on $\mathbf{M}^{\text{pot/dep}}$, (9). Clearly

$$\widetilde{\mathbf{M}}_{i \neq j}^{\text{pot/dep}} \geq 0, \quad \sum_j \widetilde{\mathbf{M}}_{ij}^{\text{pot/dep}} = 1.\tag{18}$$

eq:constrhof

The “off-diagonal” constraints are more complicated. There can be many $\mathbf{M}^{\text{pot/dep}}, \mathbf{H}$ that result in the same $\widetilde{\mathbf{M}}^{\text{pot/dep}}$. However, the fact that $\mathbf{M}_{ij}^{\text{pot/dep}} \in [0, 1]$ and $\mathbf{H}_{i \neq j} \geq 0$ imply that $\forall(i \neq j)$:

$$2r \max \left\{ f^{\text{pot}} (\widetilde{\mathbf{M}}_{ij}^{\text{pot}} - 1), f^{\text{dep}} (\widetilde{\mathbf{M}}_{ij}^{\text{dep}} - 1) \right\} \leq \mathbf{H}_{ij} \leq 2r \min \left\{ f^{\text{pot}} \widetilde{\mathbf{M}}_{ij}^{\text{pot}}, f^{\text{dep}} \widetilde{\mathbf{M}}_{ij}^{\text{dep}} \right\}\tag{19}$$

eq:homrange

These two bounds are consistent when

$$-f^{\text{dep}} \leq f^{\text{pot}} \widetilde{\mathbf{M}}_{ij}^{\text{pot}} - f^{\text{dep}} \widetilde{\mathbf{M}}_{ij}^{\text{dep}} \leq f^{\text{pot}} \quad \forall i \neq j.\tag{20}$$

eq:homconsis

To ensure that $\mathbf{M}_{ii}^{\text{pot/dep}} \geq 0$, we also need to ensure that $\sum_{j \neq i} \mathbf{M}_{ij}^{\text{pot/dep}} \leq 1$.

$$\begin{aligned}\sum_{j \neq i} \mathbf{M}_{ij}^{\text{pot}} &= \sum_{j \neq i} \left(\widetilde{\mathbf{M}}_{ij}^{\text{pot}} - \frac{\mathbf{H}_{ij}}{2r f^{\text{pot}}} \right) \\ &\geq \sum_{j \neq i} \left(\widetilde{\mathbf{M}}_{ij}^{\text{pot}} - \frac{1}{f^{\text{pot}}} \min \left\{ f^{\text{pot}} \widetilde{\mathbf{M}}_{ij}^{\text{pot}}, f^{\text{dep}} \widetilde{\mathbf{M}}_{ij}^{\text{dep}} \right\} \right) \\ &= \frac{1}{f^{\text{pot}}} \sum_{j \neq i} \left[f^{\text{pot}} \widetilde{\mathbf{M}}_{ij}^{\text{pot}} - f^{\text{dep}} \widetilde{\mathbf{M}}_{ij}^{\text{dep}} \right]_+, \\ \sum_{j \neq i} \mathbf{M}_{ij}^{\text{dep}} &= \frac{1}{f^{\text{dep}}} \sum_{j \neq i} \left[f^{\text{dep}} \widetilde{\mathbf{M}}_{ij}^{\text{dep}} - f^{\text{pot}} \widetilde{\mathbf{M}}_{ij}^{\text{pot}} \right]_+, \end{aligned}\tag{21}$$

eq:homdiag

where rectification function is $[x]_{\pm} = x\Theta(\pm x)$, in terms of the Heaviside step function. Demanding that these quantities are at most 1 is sufficient to ensure that (20) is satisfied.

Therefore we find the set of constraints

$$\begin{aligned} \widetilde{\mathbf{M}}_{i \neq j}^{\text{pot/dep}} &\geq 0, & \sum_{j \neq i} \left[f^{\text{pot}} \widetilde{\mathbf{M}}_{ij}^{\text{pot}} - f^{\text{dep}} \widetilde{\mathbf{M}}_{ij}^{\text{dep}} \right]_+ &\leq f^{\text{pot}}, \\ \sum_j \widetilde{\mathbf{M}}_{ij}^{\text{pot/dep}} &= 1, & \sum_{j \neq i} \left[f^{\text{dep}} \widetilde{\mathbf{M}}_{ij}^{\text{dep}} - f^{\text{pot}} \widetilde{\mathbf{M}}_{ij}^{\text{pot}} \right]_+ &\leq f^{\text{dep}}. \end{aligned} \tag{22} \quad \boxed{\text{eq:constrh}}$$

We showed that these are necessary conditions for the existence of $\mathbf{M}^{\text{pot/dep}}, \mathbf{H}$ satisfying (9). We can see that they are sufficient by choosing $\mathbf{H}_{ij} = 2r \min \left\{ f^{\text{pot}} \widetilde{\mathbf{M}}_{ij}^{\text{pot}}, f^{\text{dep}} \widetilde{\mathbf{M}}_{ij}^{\text{dep}} \right\}$ for $i \neq j$ and $\mathbf{H}_{ii} = -\sum_{j \neq i} \mathbf{H}_{ij}$.

We can then treat the off-diagonal elements of $\widetilde{\mathbf{M}}^{\text{pot/dep}}$ as the independent degrees of freedom, subject to these constraints, when maximising the SNR.

References

2013synapse

- [1] S. Lahiri and S. Ganguli, “A memory frontier for complex synapses,” in *Advances in Neural Information Processing Systems 26*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, eds., pp. 1034–1042. NIPS, 2013.