Complex Synapse Identification

Subhaneil Lahiri

August 4, 2020

Abstract

Notes on using HMM techniques to fit models of complex synapses

sec:intro

1 Introduction

ec:notation

2 Notation

Let t be the time of a plasticity event, assumed to be an integer. We denote a sequence of event times by $[0, T]$. Any sequence whose upper bound is smaller that its lower bound can be considered empty, and that variable is omitted. The same applies to sums and products.

We denote states of a synapse by subscripts i, j, k. The state at a particular time during a sequence is $i(t)$. An entire trajectory (sequence of states) is denoted by $i[0, T]$.

The synaptic weight is denoted by $\mathbf{w}$. Its value when the synapse is in state i is $\mathbf{w}_i$. Its observed value at time t during a sequence is $\mathbf{w}(t)$. An entire sequence of observations is denoted by $\mathbf{w}[0, T]$.

The different types of plasticity (e.g. potentiation and depression) are indicated by superscripts μ, ν. The type of plasticity used between times $(t, t+1)$ is $\mu(t)$. An entire sequence of plasticity types is denoted by $\mu[0, T-1]$.

When we are combining data from several sequences of plasticity events, the individual sequences will be denoted by superscripts a, b. The set of all sequences will be denoted by the superscript A.

The Markov matrices describing plasticity are $\mathbf{M}^{\mu}_{ij}$. The probabilities of the initial states are $\boldsymbol{\pi}_i$. The set of parameters defining a model (i.e. $\left\{\mathbf{M}^{\mu}_{ij}\right\}$ and $\boldsymbol{\pi}_i$) are denoted by $\mathbf{M}$ for convenience.

sec:em

3 Expectation-Maximisation (EM)

When we don't have any prior information about the model $\mathbf{M}$, we fit the model by maximising the likelihood function:

$$\begin{aligned}\mathbb{P}^A_{[0,T]}(\mathbf{w}\,|\,\mu,\mathbf{M}) &\equiv \mathbb{P}\big(\mathbf{w}^A[0,T]\,\big|\,\mu^A[0,T-1],\mathbf{M}\big)\\ &= \prod_a \mathbb{P}(\mathbf{w}^a[0,T]\,|\,\mu^a[0,T-1],\mathbf{M}) \equiv \prod_a \mathbb{P}^a_{[0,T]}(\mathbf{w}\,|\,\mu,\mathbf{M})\,.\end{aligned} \tag{1}$$

{eq:like}

To do this, we define an axillary function [1, 2]:

$$Q(\mathbf{M},\mathbf{M}') = \sum_{i^A[0,T]} \mathbb{P}^A_{[0,T]}(i,\mathbf{w}\,|\,\mu,\mathbf{M}) \log\left[\frac{\mathbb{P}^A_{[0,T]}(i,\mathbf{w}\,|\,\mu,\mathbf{M}')}{\mathbb{P}^A_{[0,T]}(i,\mathbf{w}\,|\,\mu,\mathbf{M})}\right]. \tag{2}$$

{eq:emq}

It can be shown that, if $\mathbf{M}'$ is chosen to maximise this quantity (which is easier than maximising the likelihood), the it will have a higher likelihood than $\mathbf{M}'$. In effect, $\mathbf{M}$ is used to estimate the hidden state trajectories, which are then used to re-estimate the model. This can be done iteratively to find a local maximum of the likelihood. If this is repeated several times with random initialisation, we can hope to find the global maximum.

When we have several sequences, we can write

$$\begin{aligned}Q(\mathbf{M},\mathbf{M}') &= \sum_{i^A[0,T]} \prod_b \mathbb{P}^b_{[0,T]}(i,\mathbf{w}\,|\,\mu,\mathbf{M}) \sum_a \log\left[\frac{\mathbb{P}^a_{[0,T]}(i,\mathbf{w}\,|\,\mu,\mathbf{M}')}{\mathbb{P}^a_{[0,T]}(i,\mathbf{w}\,|\,\mu,\mathbf{M})}\right]\\ &= \sum_a \prod_{b\neq a} \mathbb{P}^b_{[0,T]}(\mathbf{w}\,|\,\mu,\mathbf{M}) \sum_{i^a[0,T]} \mathbb{P}^a_{[0,T]}(i,\mathbf{w}\,|\,\mu,\mathbf{M}) \log\left[\frac{\mathbb{P}^a_{[0,T]}(i,\mathbf{w}\,|\,\mu,\mathbf{M}')}{\mathbb{P}^a_{[0,T]}(i,\mathbf{w}\,|\,\mu,\mathbf{M})}\right] \qquad (3)\\ &= \mathbb{P}^A_{[0,T]}(\mathbf{w}\,|\,\mu,\mathbf{M}) \sum_a \frac{Q^a(\mathbf{M},\mathbf{M}')}{\mathbb{P}^a_{[0,T]}(\mathbf{w}\,|\,\mu,\mathbf{M})}.\end{aligned}$$

{eq:emqsever}

sec:bw

3.1 Single sequence (Baum-Welch)

Suppose we have a single sequence of plasticity events. Then the joint likelihood of a trajectory and sequence observed synaptic weights is given by

$$\mathbb{P}(i[0,T],\mathbf{w}[0,T]\,|\,\mu[0,T-1],\mathbf{M}) = \boldsymbol{\pi}_{i(0)}\,\boldsymbol{\Pi}_{i(0)}(0) \prod_{t=0}^{T-1} \mathbf{M}^{\mu(t)}_{i(t)i(t+1)}\,\boldsymbol{\Pi}_{i(t+1)}(t+1), \tag{4}$$

$$\text{where}\quad \boldsymbol{\Pi}_i(t) = \delta_{\mathbf{w}_i\mathbf{w}(t)}$$

{eq:trajlike}

We wish to maximise (2) subject to normalisation constraints. Therefore we maximise the Lagrangian:

$$\mathcal{L} = Q(\mathbf{M},\mathbf{M}') + \lambda\left(1-\sum_i \boldsymbol{\pi}'_i\right) + \sum_{\mu i} \lambda^\mu_i \left(1-\sum_j \mathbf{M}'^{\mu}_{ij}\right). \tag{5}$$

{eq:lagrange

We find

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{M}'^{\mu}_{ij}} &= \frac{N^{\mu}_{ij}}{\mathbf{M}'^{\mu}_{ij}} - \lambda^{\mu}_{i}, \\ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\pi}'_{i}} &= \frac{\alpha_i(0)\,\beta_i(0)}{\boldsymbol{\pi}'_{i}} - \lambda,\end{aligned} \tag{6}$$

{eq:singledi

where the numerator N^{μ}_{ij} and the forward/backward variables α and β are defined as:

$$\begin{aligned}N^{\nu}_{jk} &= \sum_{t=0}^{T-1} \delta_{\mu(t)\nu}\, \mathbb{P}(i[t,t+1] = (j,k), \mathbf{w}[0,T] \,|\, \mu[0,T-1], \mathbf{M}) \\ &= \sum_{t=1}^{T} \delta_{\mu(t)\nu}\, \alpha_j(t)\, \beta_k(t+1)\, \mathbf{M}^{\nu}_{jk}\, \mathbf{\Pi}_k(t+1), \\ \alpha_j(t) &= \mathbb{P}(i(t) = j, \mathbf{w}[0,t] \,|\, \mu[0,t-1], \mathbf{M}) \\ &= \sum_{i[0,t]} \boldsymbol{\pi}_{i(0)}(0)\, \mathbf{\Pi}_{i(0)}(0) \prod_{s=0}^{t-1} \mathbf{M}^{\mu(s)}_{i(s)i(s+1)}\, \mathbf{\Pi}_{i(s+1)}(s+1)\, \delta_{i(t)j}, \\ \beta_j(t) &= \mathbb{P}(\mathbf{w}[t+1,T] \,|\, i(t) = j, \mu[t,T-1], \mathbf{M}) \\ &= \sum_{i[t,T]} \delta_{i(t)j} \prod_{s=t}^{T-1} \mathbf{M}^{\mu(s)}_{i(s)i(s+1)}\, \mathbf{\Pi}_{i(s+1)}(s+1).\end{aligned} \tag{7}$$

{eq:alphabet

Note that, whilst disjoint sequences $\mathbf{w}[s,s']$ and $\mathbf{w}[t,t']$ are not independent, they are conditionally independent given the state at any time between the two intervals. Therefore:

$$\alpha_j(t)\,\beta_j(t) = \mathbb{P}(i(t) = j, \mathbf{w}[0,T] \,|\, \mathbf{M})\,. \tag{8}$$

{eq:statepro

These quantities usually suffer from underflow. This can be avoided by using normalised forward/backward variables [3]:

$$\begin{aligned}\tilde{\alpha}_i(t) &= \alpha_i(t) \prod_{s=0}^{t} \eta(s), \\ \tilde{\beta}_i(t) &= \beta_i(t) \prod_{s=t+1}^{T} \eta(s),\end{aligned} \tag{9}$$

{eq:normalph

where $\eta[0,T]$ is chosen so that all of the $\tilde{\alpha}[0,T]$ are normalised: $\sum_i \tilde{\alpha}_i(t) = 1$. These

quantities have the interpretations:

$$\begin{aligned}
\tilde{\alpha}_j(t) &= \mathbb{P}(i(t) = j \,|\, \mathbf{w}[0,t], \mu[0,t-1], \mathbf{M})\,, \\
\eta(t)^{-1} &= \mathbb{P}(\mathbf{w}(t) \,|\, \mathbf{w}[0,t-1], \mu[0,t-1], \mathbf{M})\,, \\
\left[\prod_{s=0}^{t} \eta(s)\right]^{-1} &= \mathbb{P}(\mathbf{w}[0,t] \,|\, \mu[0,t-1], \mathbf{M})\,, \\
\left[\prod_{s=t_0}^{t_1} \eta(s)\right]^{-1} &= \mathbb{P}(\mathbf{w}[t_0,t_1] \,|\, \mathbf{w}[0,t_0-1], \mu[0,t_1-1], \mathbf{M})\,.
\end{aligned} \tag{10}$$

{eq:norminte

Note that the new backward variables $\tilde{\beta}$ are not normalised and have no particularly interesting interpretation, beyond the following:

$$\tilde{\alpha}_j(t)\,\tilde{\beta}_j(t) = \mathbb{P}(i(t) = j \,|\, \mathbf{w}[0,T], \mu[0,T-1], \mathbf{M})\,. \tag{11}$$

{eq:normstat

Then we can write

$$\begin{aligned}
\widetilde{N}^{\nu}_{jk} &= \sum_{t=0}^{T-1} \delta_{\mu(t)\nu}\, \mathbb{P}(i[t,t+1] = (jk) \,|\, \mathbf{w}[0,T], \mu[0,T-1], \mathbf{M}) \\
&= \sum_{t=0}^{T-1} \delta_{\mu(t)\nu}\, \tilde{\alpha}_j(t)\, \tilde{\beta}_k(t+1)\, \eta(t+1)\, \mathbf{M}^{\nu}_{jk}\, \mathbf{\Pi}_k(t+1), \\
\frac{\partial \mathcal{L}}{\partial \mathbf{M}'^{\mu}_{ij}} &= \mathbb{P}_{[0,T]}(\mathbf{w} \,|\, \mu, \mathbf{M})\; \frac{\widetilde{N}^{\mu}_{ij}}{\mathbf{M}'^{\mu}_{ij}} - \lambda^{\mu}_i, \\
\frac{\partial \mathcal{L}}{\partial \boldsymbol{\pi}'_i} &= \mathbb{P}_{[0,T]}(\mathbf{w} \,|\, \mu, \mathbf{M})\; \frac{\tilde{\alpha}_i(0)\, \tilde{\beta}_i(0)}{\boldsymbol{\pi}'_i} - \lambda,
\end{aligned} \tag{12}$$

{eq:singledi

The factors of $\mathbb{P}_{[0,T]}(\mathbf{w} \,|\, \mu, \mathbf{M})$ can be absorbed by a redefinition of the Lagrange multipliers.

Demanding a maximum and setting the derivatives to zero leaves us with the update rule [1]:

$$\begin{aligned}
\mathbf{M}'^{\mu}_{ij} &= \frac{\widetilde{N}^{\mu}_{ij}}{\sum_k \widetilde{N}^{\mu}_{ik}}, \\
\boldsymbol{\pi}'_i &= \frac{\tilde{\alpha}_i(0)\, \tilde{\beta}_i(0)}{\sum_j \tilde{\alpha}_j(0)\, \tilde{\beta}_j(0)}.
\end{aligned} \tag{13}$$

{eq:BWupdate

This can be iterated until a maximum is found.

sec:rj

3.2 Multiple sequences (Rabiner-Juang)

When we have multiple sequences of plasticity events, we will still want to maximise the Lagrangian (5), but now we will use the auxiliary function (3). This results in

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{M}'^{\mu}_{ij}} &= \mathbb{P}^{A}_{[0,T]}(\mathbf{w} \,|\, \mu, \mathbf{M}) \sum_a \frac{1}{\mathbb{P}^{a}_{[0,T]}(\mathbf{w} \,|\, \mu, \mathbf{M})} \frac{N^{a\mu}_{ij}}{\mathbf{M}'^{\mu}_{ij}} - \lambda^{\mu}_i, \\
&= \mathbb{P}^{A}_{[0,T]}(\mathbf{w} \,|\, \mu, \mathbf{M}) \frac{\sum_a \widetilde{N}^{a\mu}_{ij}}{\mathbf{M}'^{\mu}_{ij}} - \lambda^{\mu}_i, \\
\frac{\partial \mathcal{L}}{\partial \boldsymbol{\pi}'_i} &= \mathbb{P}^{A}_{[0,T]}(\mathbf{w} \,|\, \mu, \mathbf{M}) \sum_a \frac{1}{\mathbb{P}^{a}_{[0,T]}(\mathbf{w} \,|\, \mu, \mathbf{M})} \frac{\alpha^a_i(0)\, \beta^a_i(0)}{\boldsymbol{\pi}'_i} - \lambda, \\
&= \mathbb{P}^{A}_{[0,T]}(\mathbf{w} \,|\, \mu, \mathbf{M}) \frac{\sum_a \tilde{\alpha}^a_i(0)\, \tilde{\beta}^a_i(0)}{\boldsymbol{\pi}'_i} - \lambda.
\end{aligned} \tag{14}$$

{eq:multidif

Once again, we can absorb the factors of $\mathbb{P}^{A}_{[0,T]}(\mathbf{w} \,|\, \mu, \mathbf{M})$ by a redefinition of the Lagrange multipliers.

Defining $\widetilde{N}^{A\mu}_{ij} = \sum_a \widetilde{N}^{a\mu}_{ij}$ and setting the derivatives to zero leaves us with the update rule [4]:

$$\begin{aligned}
\mathbf{M}'^{\mu}_{ij} &= \frac{\widetilde{N}^{A\mu}_{ij}}{\sum_k \widetilde{N}^{A\mu}_{ik}}, \\
\boldsymbol{\pi}'_i &= \frac{\sum_a \tilde{\alpha}^a_i(0)\, \tilde{\beta}^a_i(0)}{\sum_{bj} \tilde{\alpha}^b_j(0)\, \tilde{\beta}^b_j(0)}.
\end{aligned} \tag{15}$$

{eq:RJupdate

This can be iterated until a maximum is found.

sec:priors

4 Sparseness promoting priors

When we have a prior distribution for the model, $\mathbb{P}(\mathbf{M})$, instead of maximising the likelihood (1), we can maximise the posterior:

$$\mathbb{P}(\mathbf{M} \,|\, \mathbf{w}[0,T]) = \frac{\mathbb{P}(\mathbf{w}[0,T], \mathbf{M})}{\mathbb{P}(\mathbf{w}[0,T])} = \frac{\mathbb{P}(\mathbf{w}[0,T] \,|\, \mathbf{M})\, \mathbb{P}(\mathbf{M})}{\mathbb{P}(\mathbf{w}[0,T])}. \tag{16}$$

{eq:posterio

As the denominator is independent of the model, this is equivalent to maximising the numerator, i.e. the joint distribution.

Using the same trick that was used for (2), we can define an auxilliary function

$$\begin{aligned}
\widetilde{Q}(\mathbf{M}, \mathbf{M}') &= \sum_{i^A[0,T]} \mathbb{P}\big(i^A[0,T], \mathbf{w}^A[0,T], \mathbf{M}\big) \log\left[\frac{\mathbb{P}\big(i^A[0,T], \mathbf{w}^A[0,T], \mathbf{M}'\big)}{\mathbb{P}(i^A[0,T], \mathbf{w}^A[0,T], \mathbf{M})}\right] \\
&= \mathbb{P}(\mathbf{M})\, Q(\mathbf{M}, \mathbf{M}') + \mathbb{P}\big(\mathbf{w}^A[0,T], \mathbf{M}\big) \log\left[\frac{\mathbb{P}(\mathbf{M}')}{\mathbb{P}(\mathbf{M})}\right].
\end{aligned} \tag{17}$$

{eq:emqprior

We will consider priors of the form

$$\mathbb{P}(\mathbf{M}) = \frac{\exp\left(-\beta \sum_{\mu ij} E^{\mu}_{ij}(\mathbf{M}^{\mu}_{ij})\right)}{Z(\beta)}. \tag{18}$$ {eq:priorsum

We will not put any prior on $\boldsymbol{\pi}$, so its update rule will be unaffected. The normalisation constant $Z(\beta)$ will not play any role either. Now the derivatives read:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{M}'^{\mu}_{ij}} = \mathbb{P}(\mathbf{w}[0,T],\mathbf{M}) \left(\frac{\widetilde{N}^{A\mu}_{ij}}{\mathbf{M}'^{\mu}_{ij}} - \beta \frac{\partial E^{\mu}_{ij}}{\partial \mathbf{M}'^{\mu}_{ij}} \right) - \lambda^{\mu}_{i}. \tag{19}$$ {eq:priordif

Once again, we can absorb the factors of $\mathbb{P}(\mathbf{w}[0,T],\mathbf{M})$ by a redefinition of the Lagrange multipliers.

:offdiagL1

4.1 Off-diagonal $\mathcal{L}^1$ penalty

Here we use the off-diagonal $\mathcal{L}^1$ norm as a penalty:

$$E^{\mu}_{ij} = (1-\delta_{ij})\mathbf{M}^{\mu}_{ij}. \tag{20}$$ {eq:L1penalt

Then the condition for a maximum (19) reads

$$\begin{aligned} \frac{\widetilde{N}^{A\mu}_{ij}}{\mathbf{M}'^{\mu}_{ij}} - (1-\delta_{ij})\beta - \lambda^{\mu}_{i} = 0, \\ \mathbf{M}'^{\mu}_{ij} = \frac{\widetilde{N}^{A\mu}_{ij}}{\lambda^{\mu}_{i} + (1-\delta_{ij})\beta}. \end{aligned} \tag{21}$$ {eq:L1priorm

Demanding normalisation gives

$$\begin{aligned} \sum_{j\neq i} \frac{\widetilde{N}^{A\mu}_{ij}}{\lambda^{\mu}_{i}+\beta} + \frac{\widetilde{N}^{A\mu}_{ii}}{\lambda^{\mu}_{i}} = 1, \\ \implies \quad \lambda^{\mu}_{i} = \frac{1}{2}\left[\left(\sum_{j} \widetilde{N}^{A\mu}_{ij} - \beta \right) + \sqrt{\left(\sum_{j} \widetilde{N}^{A\mu}_{ij} - \beta \right)^2 + 4\beta \widetilde{N}^{A\mu}_{ii}} \right] \end{aligned} \tag{22}$$ {eq:L1norm}

Which produces the update rule

$$\begin{aligned} \mathbf{M}'^{\mu}_{ij} &= \frac{2\widetilde{N}^{A\mu}_{ij}}{\left(\sum_k \widetilde{N}^{A\mu}_{ik} + \beta\right) + \sqrt{\left(\sum_k \widetilde{N}^{A\mu}_{ik} - \beta\right)^2 + 4\beta \widetilde{N}^{A\mu}_{ii}}} \qquad \text{for } i \neq j, \\ \mathbf{M}'^{\mu}_{ii} &= \frac{2\widetilde{N}^{A\mu}_{ii}}{\left(\sum_k \widetilde{N}^{A\mu}_{ik} - \beta\right) + \sqrt{\left(\sum_k \widetilde{N}^{A\mu}_{ik} - \beta\right)^2 + 4\beta \widetilde{N}^{A\mu}_{ii}}}. \end{aligned} \tag{23}$$ {eq:L1update

sec:Lhalf

4.2 $\mathcal{L}^{\frac{1}{2}}$ penalty

Here we use the $\mathcal{L}^{\frac{1}{2}}$ norm as a penalty:

$$E_{ij}^{\mu} = 2\sqrt{\mathbf{M}_{ij}^{\mu}}. \tag{24}$$

{eq:Lhalfpen

Then the condition for a maximum (19) reads

$$\frac{\widetilde{N}_{ij}^{A\mu}}{\mathbf{M}_{ij}'^{\mu}} - \frac{\beta}{\sqrt{\mathbf{M}_{ij}'^{\mu}}} - \lambda_i^{\mu} = 0. \tag{25}$$

{eq:Lhalfpri

It is helpful to define

$$\begin{aligned} \widetilde{N}_{ij}^{A\mu} &= \beta^2 L_{ij}^{\mu}, \\ \lambda_i^{\mu} &= \left(\gamma_i^{\mu}\right)^{-2}. \end{aligned} \tag{26}$$

{eq:Lhalfvar

Then we have the update rule

$$\mathbf{M}_{ij}'^{\mu} = \frac{\left(\gamma_i^{\mu}\right)^2 \beta^2}{4} \left(\sqrt{\left(\gamma_i^{\mu}\right)^2 + 4L_{ij}^{\mu}} - \gamma_i^{\mu} \right)^2, \tag{27}$$

{eq:Lhalfupd

with γ_i^{μ} determined by:

$$\frac{\left(\gamma_i^{\mu}\right)^2 \beta^2}{4} \sum_j \left(\sqrt{\left(\gamma_i^{\mu}\right)^2 + 4L_{ij}^{\mu}} - \gamma_i^{\mu} \right)^2 = 1. \tag{28}$$

{eq:Lhalfnor

References

70baumwelch [1] L. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *The annals of mathematical statistics* **41** (1970) no. 1, 164–171.

oster2007EM [2] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)* (Oct., 2007).

2003HMMnorm [3] C. Zhai, "A Brief Note on the Hidden Markov Models (HMMs)." Lecture notes, 2003. `http://sifaka.cs.uiuc.edu/course/498cxz05f/hmm.pdf`.

93speechrec [4] L. R. Rabiner and B.-H. Juang, *Fundamentals of speech recognition*, vol. 14 of *Signal Processing*. Prentice Hall, Inc., Upper Saddle River, NJ, USA, 1993.