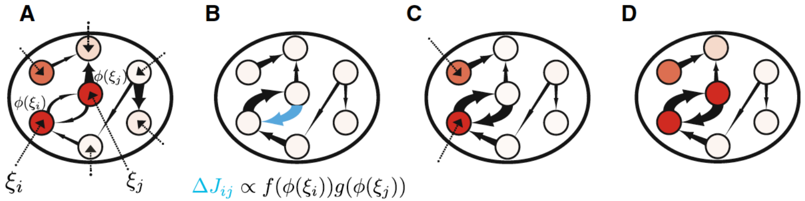
Information bounds and attractor dynamics of an associative memory trained via spike-timing dependent plasticity

Clayton Seitz

May 22, 2021

Introduction

RNNs trained with Hebbian learning rules



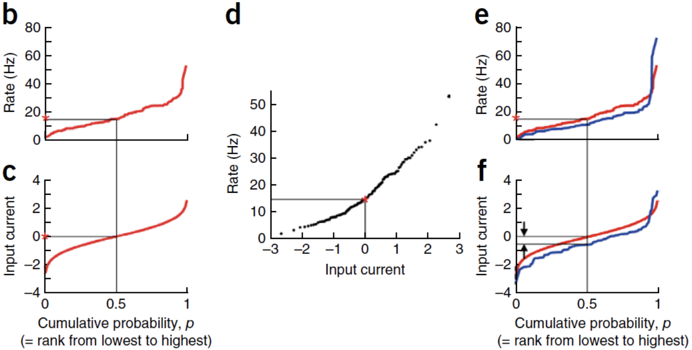
Let W_{ij} be a matrix of recurrent weights that evolves when stimulated by

$$\xi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp -\frac{1}{2}(\mathrm{r} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathrm{r} - \boldsymbol{\mu})$$

1

[1] [Peirera and Brunel, Neuron. 2018]

Inferring learning rules from firing rate distributions in ITC

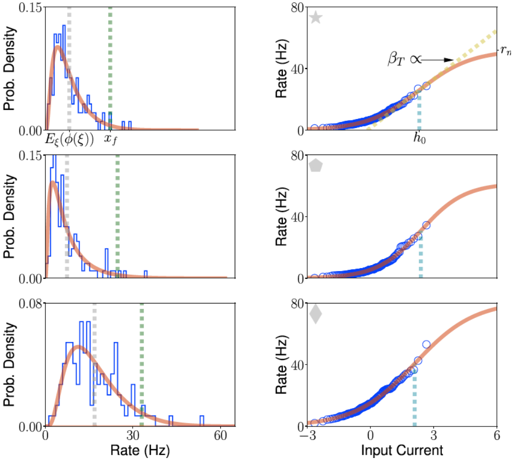


Inferring ΔW_{ij} from ITC neurons after presentation of novel and familiar images [2]

[2] [Lim et al., Nature Neuroscience. 2015]

Inferring the transfer function from ITC data

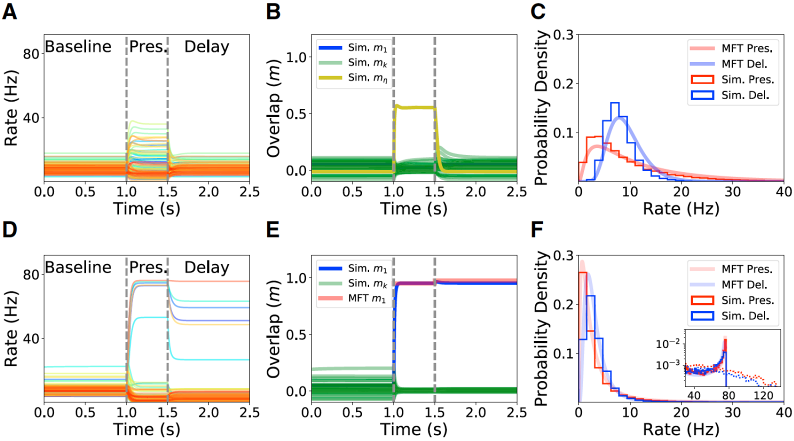
All you can really observe is the firing rate distribution. Assume the input currents are Gaussian



3

[3] [Peirera and Brunel, Neuron. 2018]

Presenting novel and familiar stimuli to the network



4

[4] [Peirera and Brunel, Neuron. 2018]

Time evolution of the RNN

The time evolution of the firing rate per neuron is given by

$$\tau_E \frac{dr_i}{dt} = -r_i + \Phi_E \left(\sum W_{ij}^0 r_j - \sum W_{ij}^1 r_j + \xi_i \right)$$

$$\tau_I \frac{dr_i}{dt} = -r_i + \Phi_I \left(\sum W_{ij}^2 r_j + \xi_i \right)$$

Let the stimulus current ξ be Gaussian

$$\xi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp -\frac{1}{2}(\mathrm{r} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathrm{r} - \boldsymbol{\mu})$$

[5]

[5] [J.J. Hopfield PNAS. 1982]

Time evolution of synaptic connectivity

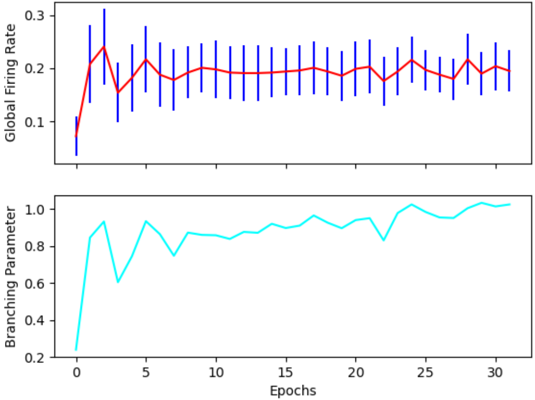
We neglect the fact that ΔW_{ij} is dependent on W_{ij}. In other words, we neglect the fact that r_i and r_j are not independent during learning

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

Assuming the functions are separable drastically simplifies the training procedure

Relating branching to firing rates

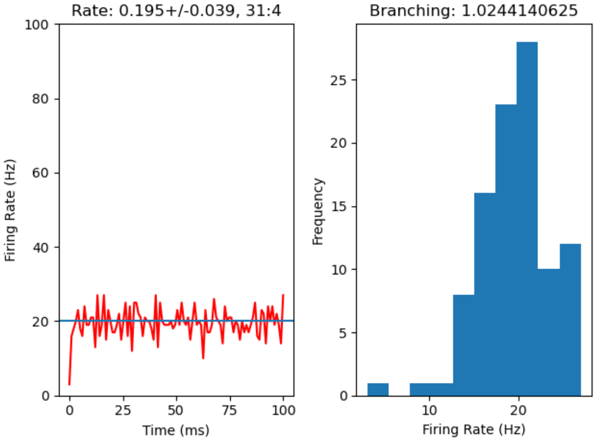
$$\mathcal{L} = \alpha \sum_t (r(t) - \hat{r})^2$$



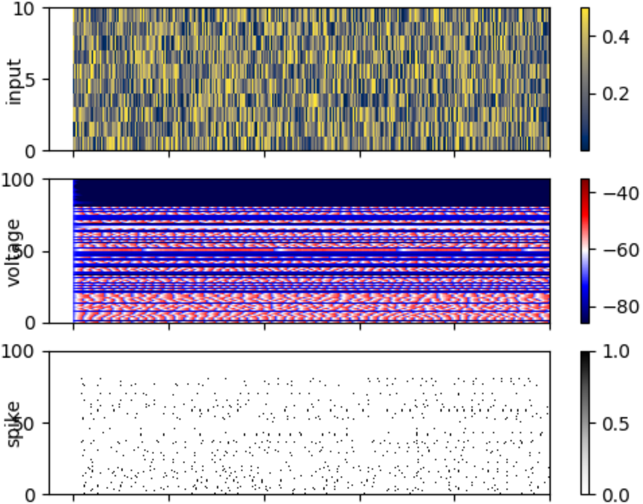
$$p_{ee} = 0.16 \; p_{ie} = 0.318 \; p_{ei} = 0.244 \; p_{ii} = 0.343$$

Relating branching to firing rates

$$\mathcal{L} = \alpha \sum_t (r(t) - \hat{r})^2$$



But optimization of $\mathcal{L}$ shows a sparse response

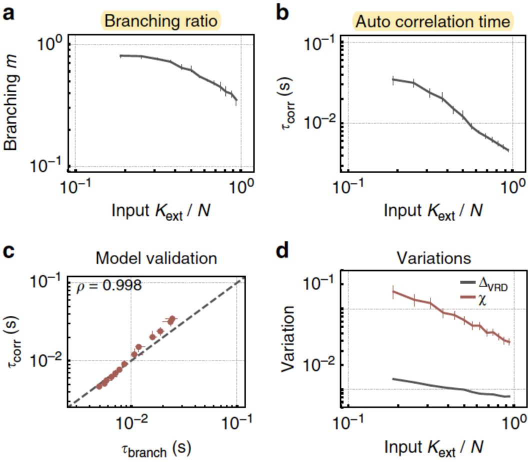


Balancing excitation and inhibition

The above analysis says nothing about excitatory and inhibitory subpopulations. How can you get critical dynamics in this case?

Balancing internal and recurrent inputs

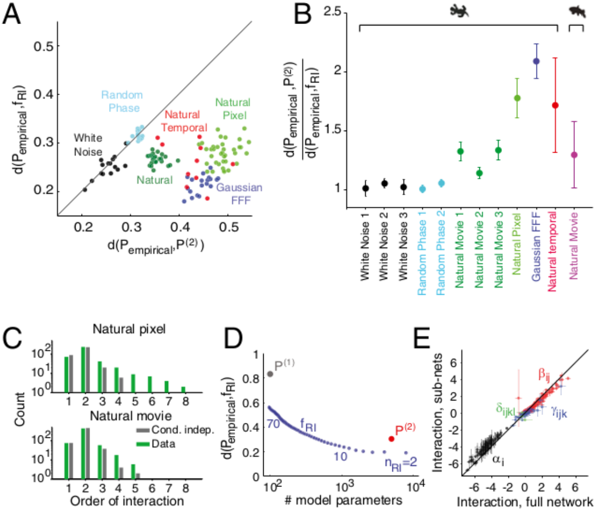
We know something about the balance of excitation and inhibition that gives critical dynamics. What about the balance between input and recurrence? (Cramer et al. 2020)



Balancing internal and recurrent inputs

Higher order correlations

Does the correlation structure of the network depend on the correlation structure of the stimulus?



RNN Gradients

Say we have a model $\Phi = (W^0, W^1)$ and want to use gradient descent to train a network to have a target rate or a target branching parameter. The rate and its associated loss for a single unit is

$$r(t) = \frac{1}{\Delta t} \int_t^{t+\Delta t} d\tau \langle \rho(\tau) \rangle \qquad \mathcal{L} = \alpha (r - r_0)^2$$

We would like the standard update

$$\Delta W_{ij} = -\eta \frac{\partial \mathcal{L}}{\partial W_{ij}}$$

But it is intractable to compute $\frac{\partial \mathcal{L}}{\partial W_{ij}}$ since $\rho(t)$ depends on other neurons through space and time.

Factorizing loss gradients for BPTT

BPTT involves unrolling an RNN into a large feedforward network where each layer is a time step.

$$\frac{\partial \mathcal{L}}{\partial W_{ij}^t} = \frac{\partial \mathcal{L}}{\partial h_j^t} \frac{\partial h_j^t}{\partial W_{ij}^t}$$

and the total gradient is a sum over the layers (time)

$$\frac{\partial \mathcal{L}}{\partial W_{ij}^t} = \sum_t \frac{\partial \mathcal{L}}{\partial h_j^t} \frac{\partial h_j^t}{\partial W_{ij}^t}$$

Deriving e-prop from BPTT

Consider the first term above. The hidden state is computed by some function $h_j^t = F(z_j^t, h_j^{t-1}, W)$. Backpropagating through time is then

$$\frac{\partial \mathcal{L}}{\partial h_j^t} = \frac{\partial \mathcal{L}}{\partial z_j^t}\frac{\partial z_j^t}{\partial h_j^t} + \frac{\partial \mathcal{L}}{\partial h_j^{t+1}}\frac{\partial h_j^{t+1}}{\partial h_j^t}$$

which must be expressed recursively

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial h_j^t} &= \frac{\partial \mathcal{L}}{\partial z_j^t}\frac{\partial z_j^t}{\partial h_j^t} + \left(\frac{\partial \mathcal{L}}{\partial z_j^{t+1}}\frac{\partial z_j^{t+1}}{\partial h_j^{t+1}} + (\ldots)\frac{\partial h_j^{t+2}}{\partial h_j^{t+1}}\right)\frac{\partial h_j^{t+1}}{\partial h_j^t} \\
&= L_j^t\frac{\partial z_j^t}{\partial h_j^t} + \left(L_j^{t+1}\frac{\partial z_j^{t+1}}{\partial h_j^{t+1}} + (\ldots)\frac{\partial h_j^{t+2}}{\partial h_j^{t+1}}\right)\frac{\partial h_j^{t+1}}{\partial h_j^t} \\
&= L_j^t\frac{\partial z_j^t}{\partial h_j^t} + \left(L_j^{t+1}\frac{\partial z_j^{t+1}}{\partial h_j^{t+1}} + (\ldots)\frac{\partial h_j^{t+2}}{\partial h_j^{t+1}}\right)\frac{\partial h_j^{t+1}}{\partial h_j^t}
\end{aligned}$$

Deriving e-prop from BPTT

Plugging into the original factorization gives

$$\frac{\partial \mathcal{L}}{\partial W_{ij}} = \left(\sum_t L_j^t \frac{\partial z_j^t}{\partial h_j^t} + \left(L_j^{t+1} \frac{\partial z_j^{t+1}}{\partial h_j^{t+1}} + (\dots) \frac{\partial h_j^{t+2}}{\partial h_j^{t+1}} \right) \frac{\partial h_j^{t+1}}{\partial h_j^t} \right) \frac{\partial h_j^{t'}}{\partial W_{ij}}$$

You can then collect terms that are multiplied L_j^t

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial W_{ij}} &= \sum_t L_j^t \frac{\partial z_j^t}{\partial h_j^t} \left(\sum_{t' \leq t} \left(\prod_{t'} \frac{\partial h_j^{t'+1}}{\partial h_j^{t'}} \right) \frac{\partial h_j^{t'}}{\partial W_{ij}} \right) \\ &= \sum_t L_j^t \frac{\partial z_j^t}{\partial h_j^t} \epsilon_{\mathrm{ij}}^{\mathrm{t}} = \sum_t L_j^t e_{ij}^t \end{aligned}$$

Constraining the global firing rate distribution

We can define a constraint on the variance of the global firing rate (which simultaneously constrains the mean)

$$\mathcal{L} = \beta(\sigma - \sigma_r)^2 \qquad \sigma = \frac{1}{T}\sum_t (r - \mu_r)^2$$

where we constrain branching by constraining the variance s of the global firing rate where branching $\rightarrow 1$ as $s \rightarrow 0$.

$$L_j^t = \frac{\partial \mathcal{L}}{\partial z_j^t} = \frac{\partial \mathcal{L}}{\partial \sigma}\frac{\partial \sigma}{\partial n}\frac{\partial n}{\partial z_j^t} = \pm\beta(\sigma - \sigma_r)\cdot(r - \mu_r)$$

Think push-pull. Some variation is necessary for refractoriness.

Relating branching to firing rates and information

We have an ensemble of neurons with firing rates r $= (r_1, r_2..., r_K)$

$$\langle N \rangle = \sum_k r_k^t \Delta t = K \Delta t \langle r \rangle$$

and we would like $\langle N \rangle$ to be constant. We draw a rate-vector from the joint distribution r $\sim R(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$H(R) = H(P(r_1, r_2, ...r_k)) \leq H\left(\prod_k P(r_k)\right)$$

the upper bound maximizes mutual information at low noise

$$I(X; R) = H(R) - H(R|X)$$