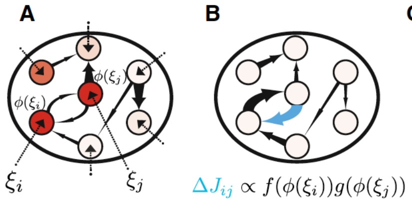
Information bounds and attractor dynamics of a Hebbian associative memory

Clayton Seitz

May 24, 2021

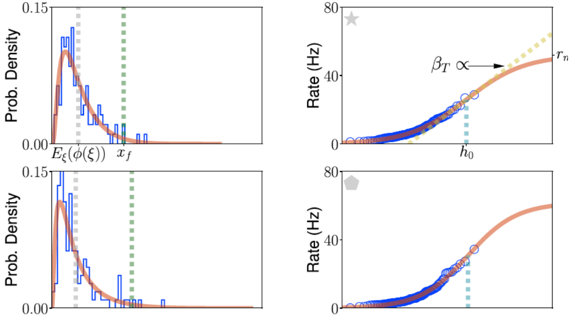
RNNs trained with a Hebbian learning rule



Measure Δr to infer the learning rule ΔW_{ij}.[1]

[1] [?]

Inferring the transfer function from ITC data



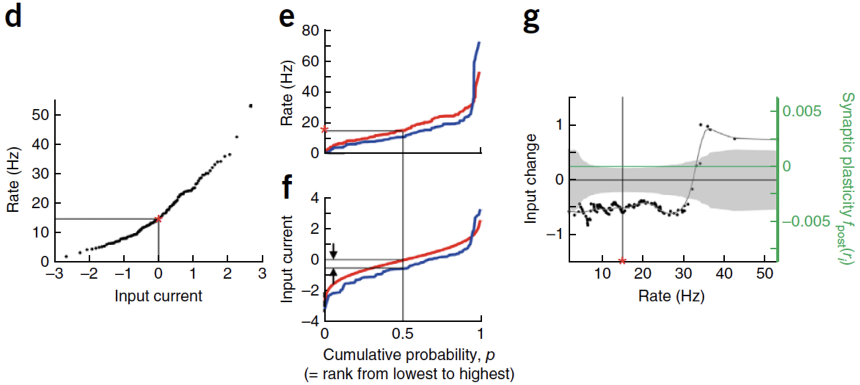
Measuring the *static* transfer function from novel images assuming that input currents are Gaussian variables

$$\phi(\xi) = \frac{r_{max}}{1 + \exp \beta(\xi - \xi_0)}$$

2

[2] **[?]**

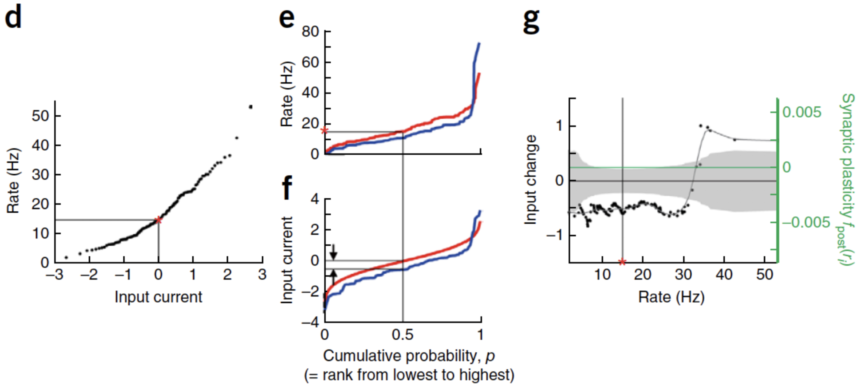
Inferring the learning rule from ITC data



Inferring the change in input current ξ_{in} from the change in firing rate in novel relative to familiar stimuli [3]

[3] [?]

Inferring the learning rule from ITC data



The change in input current to a neuron can then be read from the firing rate of that neuron when presented a novel stimulus

$$\Delta\xi_i(r) \propto (2q + 1 - \tanh(\beta(r - x)))$$

4

[4] [?]

A Hebbian update for synaptic weights

Assuming that $\Delta W_{ij} \propto f(r_i)g(r_j)$, the change in input current is related to synaptic plasticity by

$$\Delta\xi_i \propto f(r_i) \sum_j g(r_j) r_j$$

which we have fit from the data as

$$\Delta\xi_i(r) \propto (2q + 1 - \tanh(\beta(r - x)))$$

so we can write

$$f(r_i) = \frac{(2q + 1 - \tanh(\beta(r - x)))}{\sum_j g(r_j) r_j}$$

A Hebbian update for synaptic weights

During training, we stimulate the network with

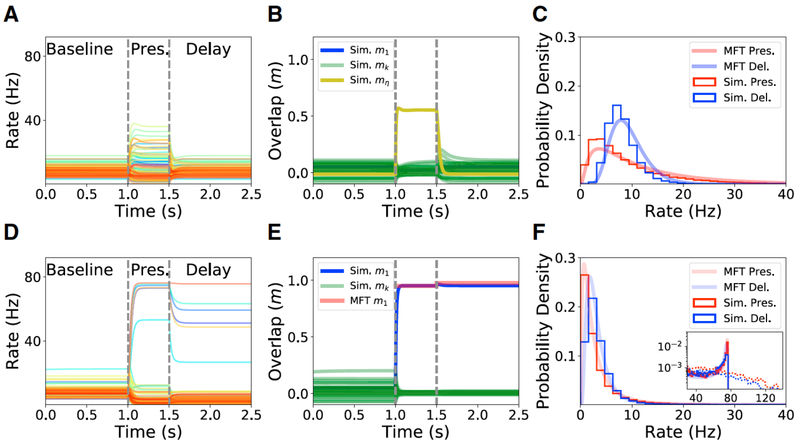
$$\xi_{in}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp -\frac{1}{2}(\boldsymbol{r} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{r} - \boldsymbol{\mu})$$

5

---

5 [**?**]

Presenting novel and familiar stimuli to the network

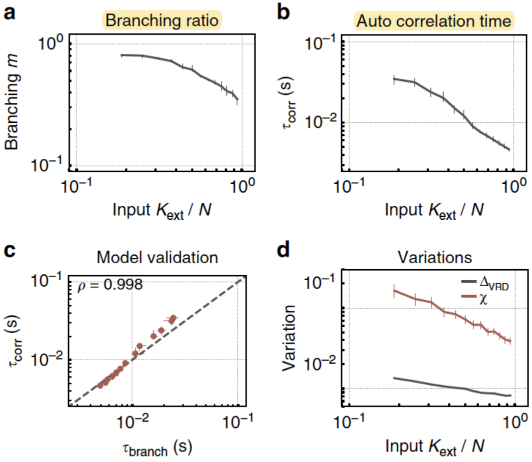


6

6 [?]

Do these networks optimize information transmission?

Are these networks functioning at a critical point? What about the balance between input and recurrence? (Cramer et al. 2020)



A coding theory perspective

How much information does the response R carry about the input pattern S i.e. $I(R;S)$ on novel and familiar stimuli?

What is the fundamental coding capacity of these networks?