A general theory of learning and memory with Complex Synapses

based on work with Surya Ganguli

Subhaneil Lahiri

Stanford University, Applied Physics

June 6, 2013

A general theory of learning and memory with Complex Synapses

based on work with Surya Ganguli

Subhaneil Lahiri

Stanford University, Applied Physics

June 6, 2013

Introduction

We often model synaptic plasticity as the change of a single number (synaptic weight). In reality, there is a complex dynamical system inside a synapse.

Semi-realistic models of synaptic plasticity have terrible memory without synaptic complexity.

We will study the entire space of a broad class of models of complex synapses to find upper bounds on their performance.

Introduction

We often model synaptic plasticity as the change of a single number (synaptic weight). In reality, there is a complex dynamical system inside a synapse.

Semi-realistic models of synaptic plasticity have terrible memory without synaptic complexity.

We will study the entire space of a broad class of models of complex synapses to find upper bounds on their performance.

1. amplitude of psp.
2. finite number of values.

Concepts

Learning rule: how activity → potentiation/depression.
e.g. Hebb rule, STDP, Hopfield outer product, perceptron rule,. . .

Plasticity mechanism: how synapse responds to potentiation/depression.
e.g. Binary switch, Cascade model, Multistate model,. . .

Memory retrieval: how synaptic state affects neural activity.
e.g. Hopfield attractor,. . .

Concepts

Learning rule: how activity → potentiation/depression.
e.g. Hebb rule, STDP, Hopfield outer product, perceptron rule,. . .

Plasticity mechanism: how synapse responds to potentiation/depression.
e.g. Binary switch, Cascade model, Multistate model,. . .

Memory retrieval: how synaptic state affects neural activity.
e.g. Hopfield attractor,. . .

1. Avoid confusion: separate concepts
2. Talk exclusively about second one.
3. Abstract away other two.

Outline

1. Why complex synapses?
2. Modelling synaptic complexity
3. Upper bounds
4. Envelope memory curve

Outline
1. Why complex synapses?
2. Modelling synaptic complexity
3. Upper bounds
4. Envelope memory curve

1. review terrible properties of simple synapses.
2. mathematical formalism of model, quantify performance (memory decay over time)
3. upper bounds on single numbers that depend on whole memory curve (deca over time)
4. upper bounds at finite times

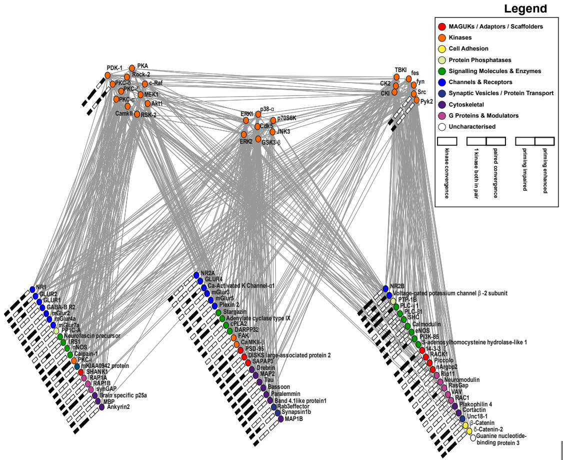
Section 1

Why complex synapses?

2013-06-06

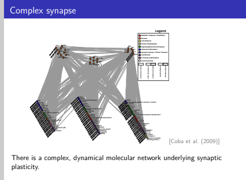
Section 1
Why complex synapses?

Complex synapse



[Coba et al. (2009)]

There is a complex, dynamical molecular network underlying synaptic plasticity.



1. Molecular network, post-synaptic density, from Seth Grant
2. Does this matter?
3. Could just be the machinery for changing synaptic weight
4. link back to questions on “There”

A classical perceptron (used as a recognition memory device) has a capacity $\propto N$, the number of synapses.

Requires synapses' dynamic range also $\propto N$.

If we restrict synaptic weight to a fixed, finite set of values,
$\implies$ tradeoff between learning and forgetting:
new memories overwriting old.

If we wish to store new memories rapidly, memory capacity $\sim \mathcal{O}(\log N)$.
[Amit and Fusi (1992), Amit and Fusi (1994)]

To circumvent this tradeoff, need to go beyond model of a synapse as a single number.

Storage capacity of synaptic memory
A classical perceptron (used as a recognition memory device) has a capacity $\propto N$, the number of synapses.
Requires synapses' dynamic range also $\propto N$.
If we restrict synaptic weight to a fixed, finite set of values,
$\implies$ tradeoff between learning and forgetting:
new memories overwriting old.
If we wish to store new memories rapidly, memory capacity $\sim \mathcal{O}(\log N)$.
[Amit and Fusi (1992), Amit and Fusi (1994)]
To circumvent this tradeoff, need to go beyond model of a synapse as a single number.

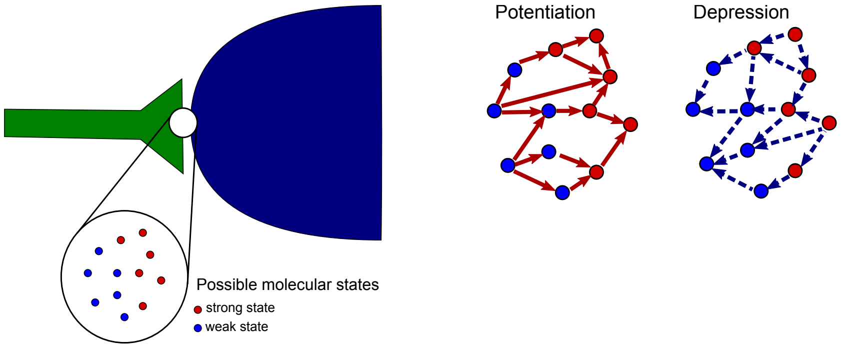
1. very plastic: learn easy, forget easy
2. little plasticity, remember better, learn harder
3. or sparse $\sim \log N/N$
4. one way around limit: complexity

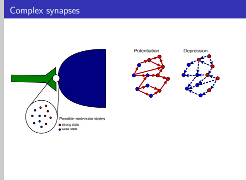
Section 2

Modelling synaptic complexity

Section 2
Modelling synaptic complexity

Complex synapses





1. functional states, not molecules
2. synaptic weight depends on state
3. many states can have same weight
4. stochastic transitions

Simplifying assumptions

- No spatial/temporal correlations in plasticity events.
- Potentiating/depressing plasticity events $\sim$ Poisson processes.
- Potentiation and depression are described by Markov processes.
- Synaptic weights can only take values ± 1.
- Ideal observer: read weights directly.

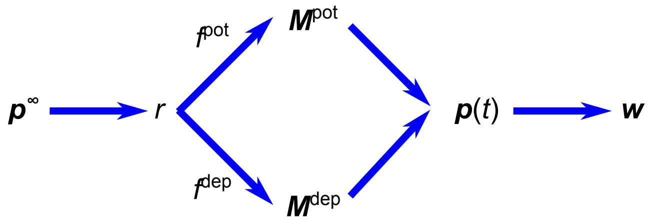
  [Fusi et al. (2005), Fusi and Abbott (2007), Barrett and van Rossum (2008)]

Simplifying assumptions
- No spatial/temporal correlations in plasticity events.
- Potentiating/depressing plasticity events $\sim$ Poisson processes.
- Potentiation and depression are described by Markov processes.
- Synaptic weights can only take values ± 1.
- Ideal observer: read weights directly.

  [Fusi et al. (2005), Fusi and Abbott (2007), Barrett and van Rossum (2008)]

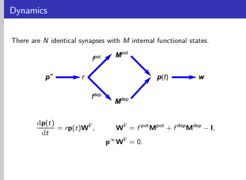
1. allows us to concentrate on synapse, not neuron/network
2. No filing system
3. don't care if STDP...
4. looks like binary synapse from outside. Inside...
5. ideal observer reads weights, not electrical activity: don't model neurons/network
6. upper bound on electrical activity readout

There are N identical synapses with M internal functional states.



$$\frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = r\mathbf{p}(t)\mathbf{W}^{\mathrm{F}}, \qquad \mathbf{W}^{\mathrm{F}} = f^{\text{pot}}\mathbf{M}^{\text{pot}} + f^{\text{dep}}\mathbf{M}^{\text{dep}} - \mathbf{I},$$

$$\mathbf{p}^\infty\mathbf{W}^{\mathrm{F}} = 0.$$



1. stoch process has steady state.
2. Prior activity puts it in this state. row vec.
3. plasticity events at rate r
4. fraction pot/dep
5. probs changed by Markov matrices, prob $i \to j$
6. Readout: synaptic weight vec when in each state.
7. Memory at $t = 0$, keep track of pot/dep
8. subsequent: average over pot/dep

Memory curve

$\vec{w}$ is the N-element vector of synaptic weights.

$$\text{Signal} = \left\langle \vec{w}_{\text{ideal}} \cdot \vec{w}(t) - \vec{w}_{\text{ideal}} \cdot \vec{w}(\infty) \right\rangle,$$
$$\text{Noise} = \sqrt{\text{Var}\left(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty)\right)} \sim \sqrt{N}.$$

We find:

$$\text{SNR}(t) = \sqrt{N}(2f^{\text{pot}}f^{\text{dep}})\,\mathbf{p}^{\infty}\left(\mathbf{M}^{\text{pot}} - \mathbf{M}^{\text{dep}}\right)\exp\left(rt\mathbf{W}^{\text{F}}\right)\mathbf{w}.$$

Memory curve
$\vec{w}$ is the N-element vector of synaptic weights.
$\text{Signal} = \left\langle \vec{w}_{\text{ideal}} \cdot \vec{w}(t) - \vec{w}_{\text{ideal}} \cdot \vec{w}(\infty) \right\rangle$,
$\text{Noise} = \sqrt{\text{Var}\left(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty)\right)} \sim \sqrt{N}$.
We find:
$\text{SNR}(t) = \sqrt{N}(2f^{\text{pot}}f^{\text{dep}})\,\mathbf{p}^{\infty}\left(\mathbf{M}^{\text{pot}} - \mathbf{M}^{\text{dep}}\right)\exp\left(rt\mathbf{W}^{\text{F}}\right)\mathbf{w}$.

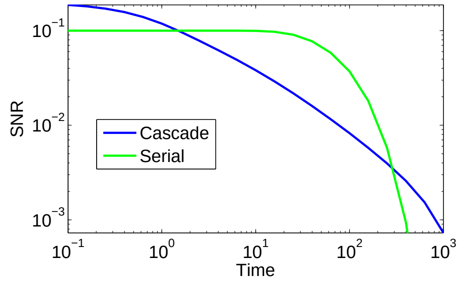
1. of different synapses
2. ideal observer reads weights, not states
3. upper bound on electrical activity readout
4. ideal: pot$\rightarrow$strong...
5. use $\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty)$ as null model
6. Noise: ignore correction when asymmetric. No effect.
7. Using $\mathbf{W}^{\text{F}}$ averages over pot/dep sequence (proof: expand)

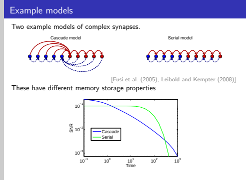
Example models

Two example models of complex synapses.

[Fusi et al. (2005), Leibold and Kempter (2008)]

These have different memory storage properties





1. previous work, also: Benna-Fusi
2. Multistate good at one time, bad at others,
3. Cascade, less well at that time, better over range of times.

Questions

- Can we understand the space of *all possible* synaptic models?
- How does structure (topology) of model $\rightarrow$ function (memory curve)?
- What are the limits on what can be achieved?
- Which transition topologies saturate these limits?

Questions

- Can we understand the space of *all possible* synaptic models?
- How does structure (topology) of model $\rightarrow$ function (memory curve)?
- What are the limits on what can be achieved?
- Which transition topologies saturate these limits?

1. not just individual models
2. understand net (link on topology)
3. avoid using word "optimal". depends on what want to do.

Constraints

Memory curve given by

$$\mathrm{SNR}(t) = \sqrt{N}(2f^{\mathrm{pot}}f^{\mathrm{dep}})\,\mathbf{p}^{\infty}\left(\mathbf{M}^{\mathrm{pot}} - \mathbf{M}^{\mathrm{dep}}\right)\exp\left(rt\mathbf{W}^{\mathrm{F}}\right)\mathbf{w}.$$

Constraints: $\mathbf{M}^{\mathrm{pot/dep}}_{ij} \in [0,1], \qquad \sum_j \mathbf{M}^{\mathrm{pot/dep}}_{ij} = 1.$

Eigenmode decomposition:

$$\mathrm{SNR}(t) = \sqrt{N}\sum_a \mathcal{I}_a\,\mathrm{e}^{-rt/\tau_a}.$$

Constraints
Memory curve given by
$\mathrm{SNR}(t) = \sqrt{N}(2f^{\mathrm{pot}}f^{\mathrm{dep}})\,\mathbf{p}^{\infty}\left(\mathbf{M}^{\mathrm{pot}} - \mathbf{M}^{\mathrm{dep}}\right)\exp\left(rt\mathbf{W}^{\mathrm{F}}\right)\mathbf{w}.$
Constraints: $\mathbf{M}^{\mathrm{pot/dep}}_{ij} \in [0,1], \qquad \sum_j \mathbf{M}^{\mathrm{pot/dep}}_{ij} = 1.$
Eigenmode decomposition:
$\mathrm{SNR}(t) = \sqrt{N}\sum_a \mathcal{I}_a\,\mathrm{e}^{-rt/\tau_a}.$

1. prefactors don't do anything, ignore
2. prior state, encoding, forgetting, readout
3. difficult to to apply
4. what are constraints on these?

Section 3

Upper bounds

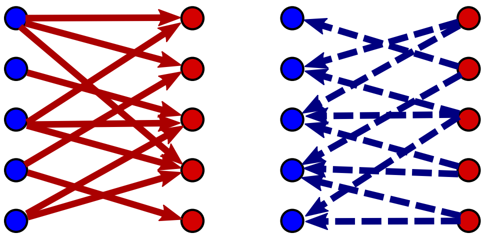
Section 3
Upper bounds

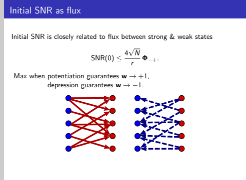
Initial SNR as flux

Initial SNR is closely related to flux between strong & weak states

$$\mathrm{SNR}(0) \leq \frac{4\sqrt{N}}{r}\,\mathbf{\Phi}_{-+}.$$

Max when potentiation guarantees $\mathbf{w} \to +1$,
depression guarantees $\mathbf{w} \to -1$.

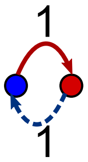




1. flux = eq prob × trans prob
2. usually saturated: pot never dec, dep never inc
3. transitions out of one node sum to 1
4. equivalent to two-state model: doesn't matter which strong/weak state, same prob of going to other set.

Two-state model

Two-state model equivalent to previous slide:

Transitions: 

$$\implies \mathrm{SNR}(t) = \sqrt{N}\,(4 f^{\text{pot}} f^{\text{dep}})\,\mathrm{e}^{-rt}.$$

Maximal initial SNR:

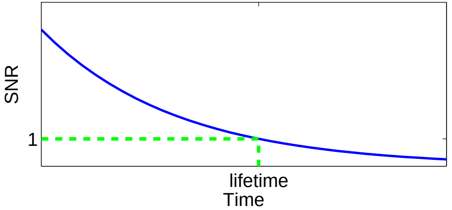
$$\mathrm{SNR}(0) \le \sqrt{N}.$$

Two-state model
Two-state model equivalent to previous slide:
Transitions: $\implies \mathrm{SNR}(t) = \sqrt{N}\,(4 f^{\text{pot}} f^{\text{dep}})\,\mathrm{e}^{-rt}.$
Maximal initial SNR: $\mathrm{SNR}(0) \le \sqrt{N}.$

1. decays very quickly
2. $f^{\text{pot}} = \frac{1}{2}$
3. Initial SNR not a good thing to optimise.

Memory lifetime bounded by area under SNR curve:

$$\begin{aligned} \mathrm{SNR}(\text{lifetime}) &= 1 \\ \Longrightarrow \quad \text{lifetime} &< \mathcal{A}. \end{aligned}$$

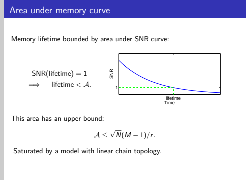


This area has an upper bound:

$$\mathcal{A} \leq \sqrt{N}(M-1)/r.$$

Saturated by a model with linear chain topology.

2013-06-06

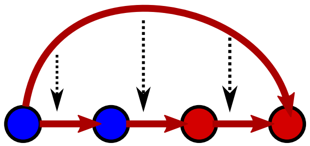


1. lifetime = area under green ¡ area under blue
2. capacity $\sim r$ lifetime, #new memories before we forget original.
3. reminder: N =#synapses, M =#states
4. proof next slide

Proof of area bound

For any model, we can construct perturbations that

- preserve equilibrium distribution,
- increase area.



details

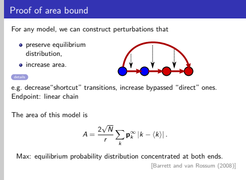
e.g. decrease“shortcut” transitions, increase bypassed “direct” ones.
Endpoint: linear chain

The area of this model is

$$A = \frac{2\sqrt{N}}{r} \sum_k \mathbf{p}^\infty_k \left| k - \langle k \rangle \right| .$$

Max: equilibrium probability distribution concentrated at both ends.

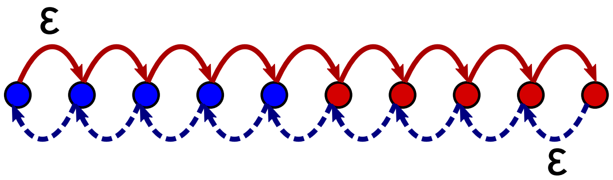
[Barrett and van Rossum (2008)]



1. relies on order & technical condition
2. max given $\mathbf{p}^\infty$
3. now max wrt. $\mathbf{p}^\infty$
4. keep c.o.m. in middle
5. similar result, slightly different conditions: linear weights, mutual info

Saturating model

Make end states “sticky”



Has long decay time, but terrible initial SNR.

$$\lim_{\varepsilon \to 0} A = \sqrt{N}(M-1)/r.$$

Saturating model
Make end states “sticky”
Has long decay time, but terrible initial SNR.
$\lim_{\varepsilon \to 0} A = \sqrt{N}(M-1)/r.$

1. Difficult to get out of end state.
2. Area not a good thing to optimise

Section 4

Envelope memory curve

Section 4
Envelope memory curve

Bounding finite time SNR

SNR curve:

$$\mathrm{SNR}(t) = \sqrt{N} \sum_a \mathcal{I}_a \, \mathrm{e}^{-rt/\tau_a}.$$

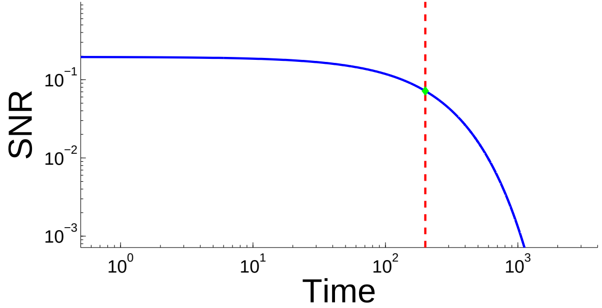
subject to constraints:

$$\sum_a \mathcal{I}_a \leq 1, \qquad \sum_a \mathcal{I}_a \tau_a \leq M - 1.$$

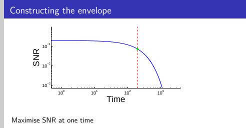
We can maximise wrt. $\mathcal{I}_a, \tau_a$.

Bounding finite time SNR
SNR curve: $\mathrm{SNR}(t) = \sqrt{N} \sum_a \mathcal{I}_a \, \mathrm{e}^{-rt/\tau_a}$.
subject to constraints: $\sum_a \mathcal{I}_a \leq 1, \quad \sum_a \mathcal{I}_a \tau_a \leq M - 1.$
We can maximise wrt. $\mathcal{I}_a, \tau_a$.

1. from eigenmode decomposition
2. from initial, area bounds

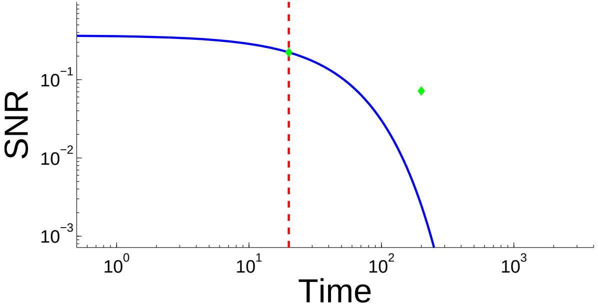


Maximise SNR at one time

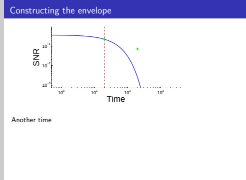


1. One exp. only constrains SNR at that time, not others

Constructing the envelope

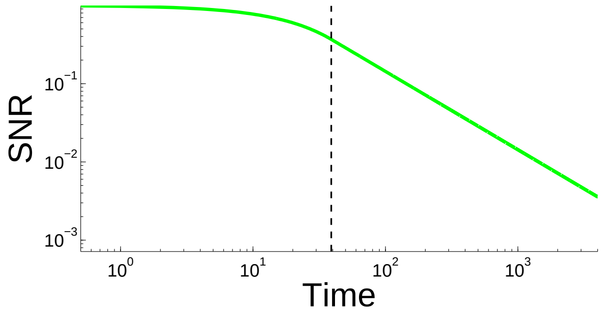


Another time



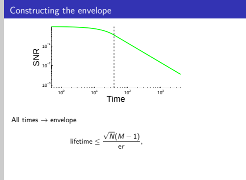
1. One exp. only constrains SNR at that time, not others
2. get another bound

Constructing the envelope



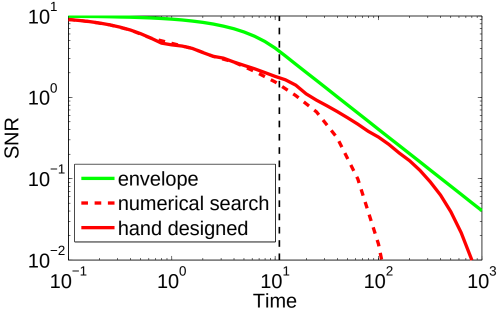
All times $\rightarrow$ envelope

$$\text{lifetime} \leq \frac{\sqrt{N}(M-1)}{er},$$

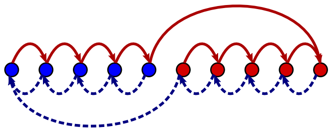


1. One exp. only constrains SNR at that time, not others
2. get another bound
3. vary time of max. no curve can cross this.
4. Regions: init(1); area(1,2)
5. Early: exp; Late: power-law, $\sim t^{-1}$
6. snr(lifetime)=1. Bounded by envelope=1.
7. is it tight? can any constrained set of exps be acheived?

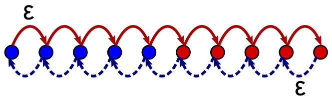
Achievable envelope

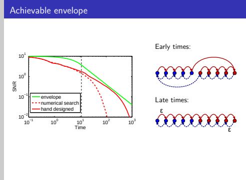


Early times:



Late times:





1. Best we've found, by numerical opt and hand chosen models.
2. vary length, keeping deterministic
3. Area maximising.
4. Gap $\sim \mathcal{O}(\sqrt{N})$.
5. Area bound active at early times $\implies$ need more constraints.

Summary

- We have formulated a general theory of learning and memory with complex synapses.
- The area under the memory curve of any model $<$ linear chain with same equilibrium distribution.
- We find a memory envelope: a single curve that cannot be exceeded by the memory curve of *any* synaptic model.
- Synaptic complexity (M internal states) raises the memory envelope linearly in M for times $> \mathcal{O}(M)$.
- Gap between envelope and what we can achieve at early times?
- Trade-off between SNR at different times?

Summary

- We have formulated a general theory of learning and memory with complex synapses.
- The area under the memory curve of any model $<$ linear chain with same equilibrium distribution.
- We find a memory envelope: a single curve that cannot be exceeded by the memory curve of *any* synaptic model.
- Synaptic complexity (M internal states) raises the memory envelope linearly in M for times $> \mathcal{O}(M)$.
- Gap between envelope and what we can achieve at early times?
- Trade-off between SNR at different times?

Acknowledgements

Thanks to:

- Surya Ganguli
- Stefano Fusi
- Marcus Benna
- David Sussillo
- Jascha Sohl-Dickstein

Acknowledgements

Thanks to:

- Surya Ganguli
- Stefano Fusi
- Marcus Benna
- David Sussillo
- Jascha Sohl-Dickstein

1. Last slide!

References I

M. P. Coba, A. J. Pocklington, M. O. Collins, M. V. Kopanitsa, R. T. Uren, S. Swamy, M. D. Croning, J. S. Choudhary, and S. G. Grant.
"Neurotransmitters drive combinatorial multistate postsynaptic density networks".
Sci Signal, 2(68):ra19, 2009.
7

D. J. Amit and S. Fusi.
"Constraints on learning in dynamic synapses".
Network: Computation in Neural Systems, 3(4):443–464, 1992.
8

D. J. Amit and S. Fusi.
"Learning in neural networks with material synapses".
Neural Computation, 6(5):957–982, 1994.
8

2013-06-06

References I

M. P. Coba, A. J. Pocklington, M. O. Collins, M. V. Kopanitsa, R. T. Uren, S. Swamy, M. D. Croning, J. S. Choudhary, and S. G. Grant.
"Neurotransmitters drive combinatorial multistate postsynaptic density networks".
Sci Signal, 2(68):ra19, 2009.

D. J. Amit and S. Fusi.
"Constraints on learning in dynamic synapses".
Network: Computation in Neural Systems, 3(4):443–464, 1992.

D. J. Amit and S. Fusi.
"Learning in neural networks with material synapses".
Neural Computation, 6(5):957–982, 1994.

References II

S. Fusi, P. J. Drew, and L. F. Abbott.
"Cascade models of synaptically stored memories".
Neuron, 45(4):599–611, Feb 2005.
11 14

S. Fusi and L. F. Abbott.
"Limits on the memory storage capacity of bounded synapses".
Nat. Neurosci., 10(4):485–493, Apr 2007.
11

A. B. Barrett and M. C. van Rossum.
"Optimal learning rules for discrete synapses".
PLoS Comput. Biol., 4(11):e1000230, Nov 2008.
11 21

References II

S. Fusi, P. J. Drew, and L. F. Abbott.
"Cascade models of synaptically stored memories".
Neuron, 45(4):599–611, Feb 2005.
11 14

S. Fusi and L. F. Abbott.
"Limits on the memory storage capacity of bounded synapses".
Nat. Neurosci., 10(4):485–493, Apr 2007.
11

A. B. Barrett and M. C. van Rossum.
"Optimal learning rules for discrete synapses".
PLoS Comput. Biol., 4(11):e1000230, Nov 2008.
11 21

References III

Christian Leibold and Richard Kempter.
"Sparseness Constrains the Prolongation of Memory Lifetime via Synaptic Metaplasticity".
Cerebral Cortex, 18(1):67–77, 2008.
14

J.G. Kemeny and J.L. Snell.
Finite markov chains.
Springer, 1960.
34

References III
Christian Leibold and Richard Kempter.
"Sparseness Constrains the Prolongation of Memory Lifetime via Synaptic Metaplasticity".
Cerebral Cortex, 18(1):67–77, 2008.
J.G. Kemeny and J.L. Snell.
Finite markov chains.
Springer, 1960.

Techinical detail: ordering states

Let $\mathbf{T}_{ij}$ = mean first passage time from state i to state j. Then:

$$\eta = \sum_j \mathbf{T}_{ij} \mathbf{p}_j^\infty,$$

is independent of the initial state i (Kemeney's constant).

[Kemeny and Snell (1960)]

We define:

$$\eta_i^+ = \sum_{j \in \text{strong}} \mathbf{T}_{ij} \mathbf{p}_j^\infty, \qquad \eta_i^- = \sum_{j \in \text{weak}} \mathbf{T}_{ij} \mathbf{p}_j^\infty.$$

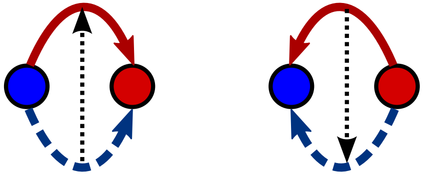
They can be used to arrange the states in an order (increasing η^- or decreasing η^+). back

Techinical detail: ordering states
Let $\mathbf{T}_{ij}$ = mean first passage time from state i to state j. Then:
$\eta = \sum_j \mathbf{T}_{ij} \mathbf{p}_j^\infty,$
is independent of the initial state i (Kemeney's constant).
[Kemeny and Snell (1960)]
We define:
$\eta_i^+ = \sum_{j \in \text{strong}} \mathbf{T}_{ij} \mathbf{p}_j^\infty, \quad \eta_i^- = \sum_{j \in \text{weak}} \mathbf{T}_{ij} \mathbf{p}_j^\infty.$
They can be used to arrange the states in an order (increasing η^- or decreasing η^+). back

1. Measure "distance" to the strong/weak states.
2. sum to constant, $\implies$ two orders same

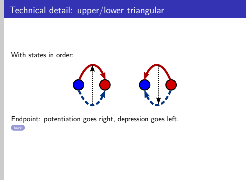
Technical detail: upper/lower triangular

With states in order:



Endpoint: potentiation goes right, depression goes left.

back



1. pot & dep with same initial & final state
2. pot/dep matrices are upper/lower triangular.
3. one other pert. too technical, even for bonus slide!