A general theory of learning and memory with Complex Synapses

based on work with Surya Ganguli

Subhaneil Lahiri

Stanford University, Applied Physics

September 1, 2014

Complex synapses

A general theory of learning and memory with Complex Synapses based on work with Surya Ganguli Subbaneil Lahiri

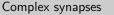
> Stanford University, Applied Pt September 1, 2014

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.



2014-09-01

—Introduction

introduction

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

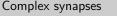
- 1. amplitude of psp.
- 2. finite number of values.

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.

ag. Hubb rule, STDP, Hopfield outer product, perceptron rule....

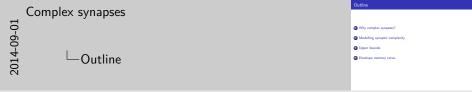
Plasticity mechanism: how synapse responds to potentiation/depression ag. Binary switch, Cascade model, Multitate model...

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

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

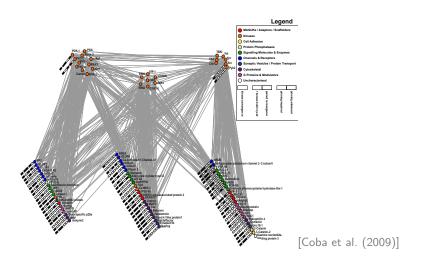


- 1. review terrible properties of simple synapses.
- mathematical formalism of model, quantify performance (memory decay over time)
- 3. upper bounds on single numbers that depend on whole memory curve (deca over time)
- 4. upper bounds at finite times

Section 1

Why complex synapses?

Complex synapse



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

Complex synapses?

Why complex synapses?

Complex synapse



- 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"

Subhaneil Lahiri (Stanford) Complex synapses September 1, 2014 6 / 32

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,

⇒ 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.



Complex synapses Why complex synapses? -Storage capacity of synaptic memory

new memories overwriting old. If we wish to store new memories rapidly, memory capacity $\sim O(\log N)$

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

torage capacity of synaptic memory

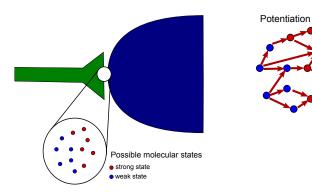
-> tradeoff between learning and forgetting:

- 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

Complex synapses





Complex synapses



Complex synapses

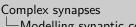
- 1. functional states, not molecules
- 2. synaptic weight depends on state
- 3. many states can have same weight
- 4. stochastic transitions

Depression

Simplifying assumptions

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

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



Modelling synaptic complexity

-Simplifying assumptions

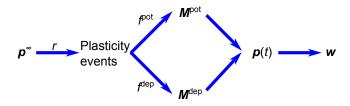
implifying assumptions

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

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

Dynamics

There are N identical synapses with M internal functional states.



$$rac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = r\mathbf{p}(t)\mathbf{W}^{\mathrm{F}}, \qquad \mathbf{W}^{\mathrm{F}} = f^{\mathsf{pot}}\mathbf{M}^{\mathsf{pot}} + f^{\mathsf{dep}}\mathbf{M}^{\mathsf{dep}} - \mathbf{I},$$
 $\mathbf{p}^{\infty}\mathbf{W}^{\mathrm{F}} = 0.$

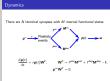
4日 → 4日 → 4 目 → 4 目 → 9 Q ○

Complex synapses

Modelling synaptic complexity

└─Dynamics

0 7 1 1



- 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 $i \rightarrow j$
- 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.

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

We find:

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

Complex synapses

-Modelling synaptic complexity

└─Memory curve

 \widetilde{w} is the N-shement vector of synaptic weights. Signal = $\langle \widetilde{w}_{clout} : \widetilde{w}(t) - \widetilde{w}_{clout} : \widetilde{w}(\infty) \rangle$, Notice = $\sqrt{Var}\langle \widetilde{w}_{clout} : \widetilde{w}(\infty) \rangle \sim \sqrt{N}$. We find: $SNS(t) = \sqrt{N}(2f^{act}f^{act}) v^{ac} \left(M^{act} - M^{acp}\right) von \left(f^{act}W^{c}\right) w$.

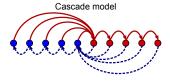
Memory curve

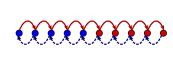
- 1. of different synapses
- 2. ideal observer reads weights, not states
- 3. upper bound on electrical activity readout
- 4. ideal: pot→strong...
- 5. use $\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty)$ as null model
- 6. Noise: ignore correction when asymmetric. No effect.
- 7. Using \mathbf{W}^{F}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

Example models

Subhaneil Lahiri (Stanford)

Two example models of complex synapses.

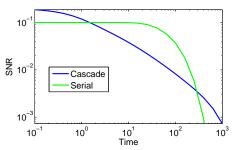




Serial model

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

These have different memory storage properties



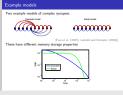
Complex synapses

September 1, 2014

Complex synapses

-Modelling synaptic complexity

Example models



- 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 → 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

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

Constraints:

$$oldsymbol{\mathsf{M}}^{\mathsf{pot}/\mathsf{dep}}_{ij} \in [0,1], \qquad \sum_j oldsymbol{\mathsf{M}}^{\mathsf{pot}/\mathsf{dep}}_{ij} = 1.$$

$$\sum_{j} {f M}_{ij}^{{\sf pot/dep}} = 1$$
 .

Eigenmode decomposition:

$$\mathsf{SNR}(t) = \sqrt{N} \sum_{a} \mathcal{I}_a \, \mathsf{e}^{-\mathsf{r} t/\tau_a}.$$

Complex synapses Modelling synaptic complexity

-Constraints



- 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

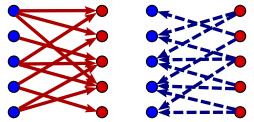


Initial SNR as flux

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

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

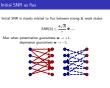
Max when potentiation guarantees $\mathbf{w} \to +1$, depression guarantees $\mathbf{w} \to -1$.





Complex synapses

Upper bounds
Initial SNR
Initial SNR as flux



- 1. flux = eq prob \times 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.

Subhaneil Lahiri (Stanford)

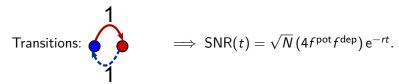
Complex synapses

September 1, 2014

17 / 3

Two-state model

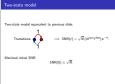
Two-state model equivalent to previous slide:



Maximal initial SNR:

$$SNR(0) \leq \sqrt{N}$$
.

Complex synapses
Upper bounds
Initial SNR
Two-state model

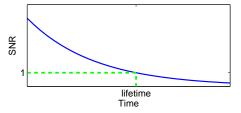


- 1. decays very quickly
- 2. $f^{pot} = \frac{1}{2}$
- 3. Initial SNR not a good thing to optimise.

Area under memory curve

Memory lifetime bounded by area under SNR curve:

$$\mathsf{SNR}(\mathsf{lifetime}) = 1$$
 \Longrightarrow $\mathsf{lifetime} < \mathcal{A}.$



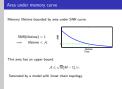
This area has an upper bound:

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

Saturated by a model with linear chain topology.

Complex synapses

Upper bounds
Area under memory curve
Area under memory curve



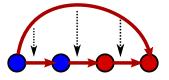
- 1. lifetime = area under green ; area under blue
- 2. capacity $\sim r$ lifetime, #new memories before we forget original.
- 3. reminder: N = #synapses, M = #states
- 4. proof next slide

2014-09-01

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

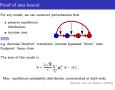
$$A = \frac{2\sqrt{N}}{r} \sum_{k} \mathbf{p}_{k}^{\infty} |k - \langle k \rangle|.$$

Max: equilibrium probability distribution concentrated at both ends.

[Barrett and van Rossum (2008)]

Complex synapses

Upper bounds
Area under memory curve
Proof of area bound



- 1. relies on order & technical condition
- 2. max given \mathbf{p}^{∞}

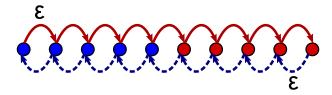
2014-09-01

- 3. now max wrt. \mathbf{p}^{∞}
- 4. keep c.o.m. in middle
- 5. similar result, slightly different conditions: linear weights, mutual info

Subhaneil Lahiri (Stanford) Complex synapses September 1, 2014 20 / 3

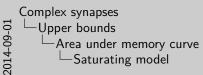
Saturating model

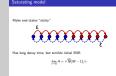
Make end states "sticky"



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

Complex synapses
Envelope memory curve

Section 4

Envelope memory curve

Bounding finite time SNR

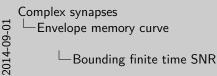
SNR curve:

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

subject to constraints:

$$\sum_{a} \mathcal{I}_{a} \leq 1, \qquad \sum_{a} \mathcal{I}_{a} \tau_{a} \leq M - 1.$$

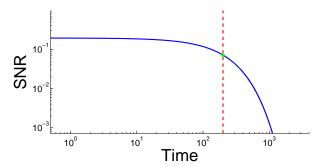
We can maximise wrt. \mathcal{I}_a, τ_a .



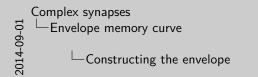


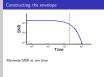
- 1. from eigenmode decomposition
- 2. from initial, area bounds

Constructing the envelope



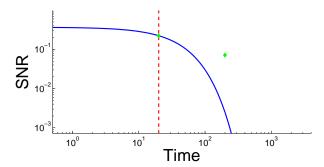
Maximise SNR at one time





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

Constructing the envelope



Another time



Complex synapses

—Envelope memory curve

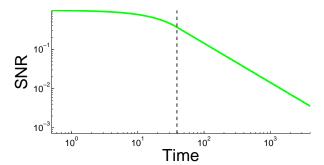


Constructing the envelope

Constructing the envelope

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

Constructing the envelope



All times \rightarrow envelope

ifetime
$$\leq \frac{\sqrt{N}(M-1)}{er}$$



Complex synapses

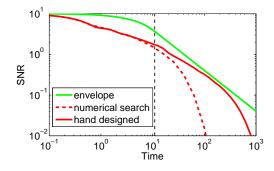
Envelope memory curve



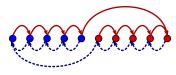


- 1. One exp. only constrains SNR at that time, not others
- 2. get another bound
- 3. vary time of max. no curve can cross this.
- 4. Regions: init(1); area(1,2)
- 5. Early: exp; Late: power-law, $\sim t^{-1}$
- 6. snr(lifetime)=1. Bounded by envelope=1.
- 7. is it tight? can any constrained set of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

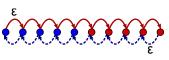
Achievable envelope



Early times:

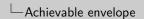


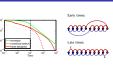
Late times:



Complex synapses

Envelope memory curve





Achievable envelope

- 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?

Complex synapses Envelope memory curve

-Summary

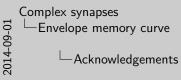
. We have formulated a general theory of learning and memory with

- 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 > O(M).
- · Gap between envelope and what we can achieve at early times? a Trade-off between SNR at different times?

Acknowledgements

Thanks to:

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



Acknowledgements

Thanks to:

Sinya Gangul

Stafan Fai

Marcia Bena

Danid Soullo

Jaicha Sold-Didutini

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).





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).





Complex synapses Envelope memory curve -References

M. P. Coba, A. J. Pocklington, M. O. Collins, M. V. Kocanitsa. R. T. Urer S. Swarry, M. D. Croning, J. S. Choudhary, and S. G. Grant. D. J. Amit and S. Fusi.

"Constraints on learning in dynamic synapses". D. J. Amit and S. Fusi. "Learning in neural networks with material synapses"

Subhaneil Lahiri (Stanford)

Complex synapses

September 1, 2014

References II



S. Fusi, P. J. Drew, and L. F. Abbott.

"Cascade models of synaptically stored memories".

Neuron, 45(4):599-611, (Feb. 2005).







S. Fusi and L. F. Abbott.

"Limits on the memory storage capacity of bounded synapses".

Nat. Neurosci., 10(4):485–493, (Apr. 2007).





A. B. Barrett and M. C. van Rossum.

"Optimal learning rules for discrete synapses".

PLoS Comput. Biol., 4(11):e1000230, (Nov, 2008).





Complex synapses Envelope memory curve

-References

S. Fusi, P. J. Drew, and L. F. Abbott "Cascade models of synaptically stored memories"

S. Funi and L. F. Abbott. "Limits on the memory storage capacity of bounded synapses

A. B. Barrett and M. C. van Rossum. "Optimal learning rules for discrete synapses".

References III

Complex synapses Envelope memory curve

References

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

Christian Leibold and Richard Kempter.

"Sparseness Constrains the Prolongation of Memory Lifetime via Synaptic Metaplasticity".

Cerebral Cortex, 18(1):67-77, (2008).



Techinical detail: ordering states

Let T_{ii} = mean first passage time from state i to state j. Then:

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

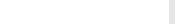
is independent of the initial state *i* (Kemeney's constant).

[?]

We define:

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

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



Techinical detail: ordering states Complex synapses Envelope memory curve ☐ Techinical detail: ordering states

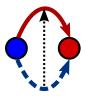
- 1. Measure "distance" to the strong/weak states.
- 2. sum to constant, \implies two orders same

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

 $\eta_i^+ = \sum_{} \mathbf{T}_{ij} \mathbf{p}_j^{\infty}, \quad \eta_i^- = \sum_{} \mathbf{T}_{ij} \mathbf{p}_j^{\infty}.$

Technical detail: upper/lower triangular

With states in order:





Endpoint: potentiation goes right, depression goes left.

back

ロ ト 4 周 ト 4 章 ト 4 章 ト 章 4 9 Q (P)

Complex synapses

Envelope memory curve

☐ Technical detail: upper/lower triangular



- 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!