A general theory of learning and memory with Complex Synapses

based on work with Surya Ganguli

Subhaneil Lahiri

Stanford University, Applied Physics

April 11, 2013

A general theory of learning and memory with Complex Synapses

based on work with Surya Ganguli

Subhaneil Lahiri

Stanford University, Applied Physics

April 11, 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 memory have terrible storage 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 memory have terrible storage 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.

Outline

Outline

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

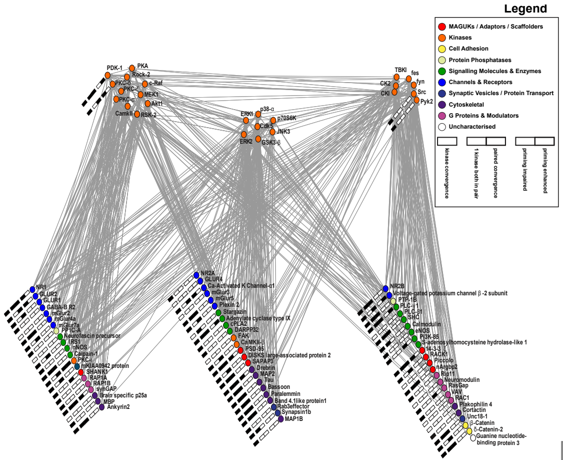
1. review terrible properties of simple synapses.
2. mathematical formalism of model, quantify performance (memory decay over time)
3. upper bounds on single numbers that describe performance at all times
4. upper bounds at finite times

Section 1

Why complex synapses?

Section 1
Why complex synapses?

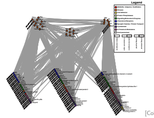
Complex synapse



[Coba et al. (2009)]

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

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”

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.

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.

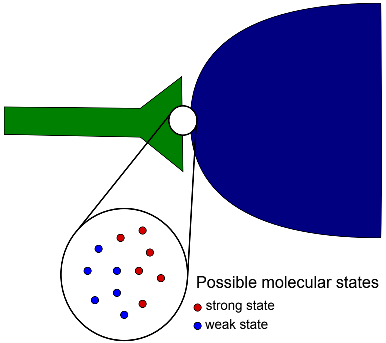
1. very plastic: learn easy, forget easy
2. little plasticity, remember better, learn harder
3. or sparse $\sim \log N/N$
4. one way around limit: complexity

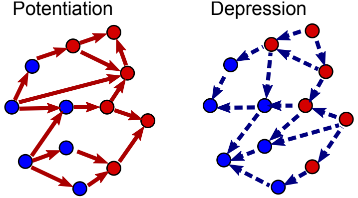
Section 2

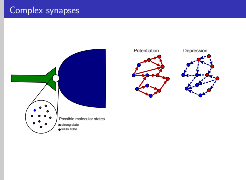
Modelling synaptic complexity

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

- There are N identical synapses with M internal functional states.
- States of different synapses are independent of each other.
- Which synapses eligible for plasticity chosen randomly.
- Potentiating/depressing plasticity events $\sim$ Poisson processes with rates $rf^{\text{pot/dep}}$, where $f^{\text{pot}} + f^{\text{dep}} = 1$.
- Potentiation and depression are described by Markov processes with transition probabilities $\mathbf{M}^{\text{pot/dep}}$.
- Synaptic weights of the internal states are given by vector $\mathbf{w}$. Can only take values ± 1.

Simplifying assumptions

- There are N identical synapses with M internal functional states.
- States of different synapses are independent of each other.
- Which synapses eligible for plasticity chosen randomly.
- Potentiating/depressing plasticity events $\sim$ Poisson processes with rates $rf^{\text{pot/dep}}$, where $f^{\text{pot}} + f^{\text{dep}} = 1$.
- Potentiation and depression are described by Markov processes with transition probabilities $\mathbf{M}^{\text{pot/dep}}$.
- Synaptic weights of the internal states are given by vector $\mathbf{w}$. Can only take values ± 1.

1. allows us to concentrate on synapse, not neuron/network
2. don't care if STDP...
3. r = total rate of plasticity events per synapse, $f^{\text{pot/dep}}$ = fraction of events that are potentiating/depressing.
4. matrix elements: transition prob from $i \to j$, given pot/dep
5. looks like binary synapse from outside. Inside...

Dynamics

At $t = 0$, the memory is created by $\mathbf{M}^{\text{pot/dep}}$ with probability $f^{\text{pot/dep}}$.

Forgetting caused by subsequent memories, evolving as

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

Eventually, this will settle into the equilibrium distribution:

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

Dynamics
At $t = 0$, the memory is created by $\mathbf{M}^{\text{pot/dep}}$ with probability $f^{\text{pot/dep}}$.
Forgetting caused by subsequent memories, evolving as
$$\frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = r\mathbf{p}(t)\mathbf{W}^{\mathrm{F}}, \qquad \mathbf{W}^{\mathrm{F}} = f^{\text{pot}}\mathbf{M}^{\text{pot}} + f^{\text{dep}}\mathbf{M}^{\text{dep}} - \mathbf{I},$$
Eventually, this will settle into the equilibrium distribution:
$$\mathbf{p}^{\infty}\mathbf{W}^{\mathrm{F}} = 0.$$

1. for this one, we keep track of pot/dep, look for inc/dec of $\mathbf{w}$
2. $\mathbf{W}^{\mathrm{F}}$ is forgetting matrix, $\mathbf{I}$ =identity, don't keep track of pot/dep
3. In equilibrium prior to memory creation

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

$$\begin{aligned}\text{Signal} &= \langle \vec{w}_{\text{ideal}} \cdot \vec{w}(t) - \vec{w}_{\text{ideal}} \cdot \vec{w}(\infty) \rangle \\ \text{Noise} &= \text{Var}\left(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty)\right)\end{aligned}$$

Related to reconstruction probability of single synapses.

$$\text{SNR}(t) \sim \sqrt{N}\, P(\text{strong/weak}, t|\text{pot/dep}, t=0) - \ldots (t=\infty).$$

Memory curve

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

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

Related to reconstruction probability of single synapses.

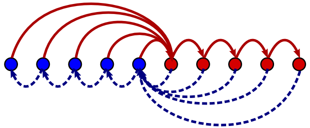
$\text{SNR}(t) \sim \sqrt{N}\, P(\text{strong/weak}, t|\text{pot/dep}, t=0) - \ldots (t=\infty)$.

1. of different synapses
2. ideal observer reads weights, not states
3. upper bound on electrical activity readout
4. ideal: pot→strong...
5. subtract baseline, some overlap even w/o encoding
6. if we ignore correlations...

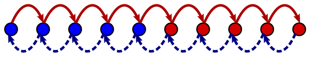
Example models

Two example models of complex synapses.

Cascade model

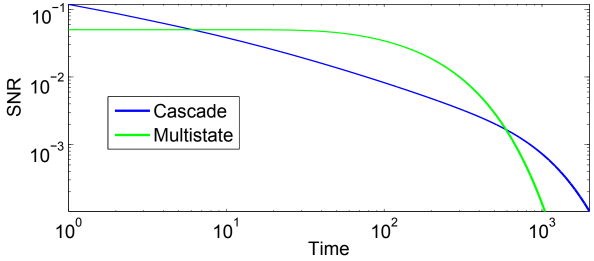


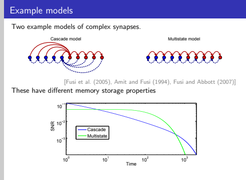
Multistate model



[Fusi et al. (2005), Amit and Fusi (1994), Fusi and Abbott (2007)]

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.

Memory curve 2

Memory curve given by

$$\mathrm{SNR}(t) = \frac{\sqrt{N}(2f^{\text{pot}}f^{\text{dep}})}{\sqrt{4\mathbf{p}_+^\infty \mathbf{p}_-^\infty}}\, \mathbf{p}^\infty \left(\mathbf{M}^{\text{pot}} - \mathbf{M}^{\text{dep}}\right) \exp\left(rt\mathbf{W}^{\mathrm{F}}\right) \mathbf{w}.$$

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

Eigenmode decomposition:

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

Memory curve 2
Memory curve given by
$\mathrm{SNR}(t) = \frac{\sqrt{N}(2f^{\text{pot}}f^{\text{dep}})}{\sqrt{4\mathbf{p}_+^\infty \mathbf{p}_-^\infty}}\, \mathbf{p}^\infty \left(\mathbf{M}^{\text{pot}} - \mathbf{M}^{\text{dep}}\right) \exp\left(rt\mathbf{W}^{\mathrm{F}}\right) \mathbf{w}.$
Constraints: $\mathbf{M}_{ij}^{\text{pot/dep}} \in [0,1], \qquad \sum_j \mathbf{M}_{ij}^{\text{pot/dep}} = 1.$
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

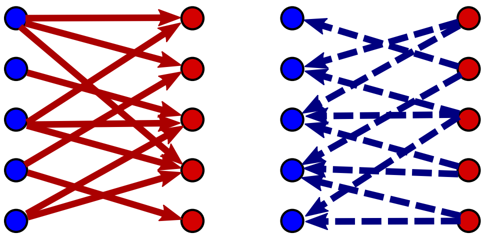
Upper bounds

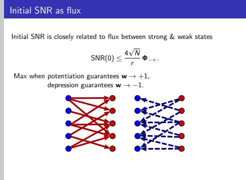
Initial SNR as flux

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

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

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

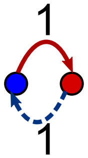




1. flux = eq prob × trans prob
2. usually saturated: pot never dec, dep never inc
3. transitions out of one node sum to 1
4. equivalent to two-state model: doesn't matter which strong/weak state, same prob of going to other set.

Two-state model

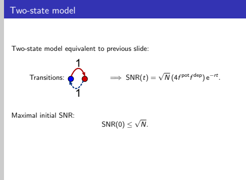
Two-state model equivalent to previous slide:

Transitions: 

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

Maximal initial SNR:

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

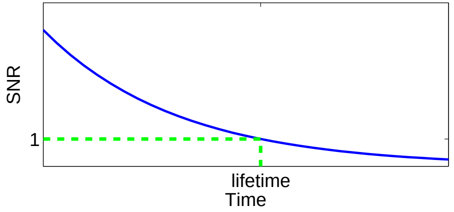


1. decays very quickly
2. $f^{\mathrm{pot}} = \frac{1}{2}$
3. Initial SNR not a good thing to optimise.

Area under memory curve

Memory lifetime bounded by area under SNR curve:

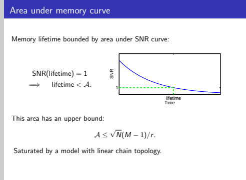
$$\begin{aligned} \mathrm{SNR}(\text{lifetime}) &= 1 \\ \implies \quad \text{lifetime} &< \mathcal{A}. \end{aligned}$$



This area has an upper bound:

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

Saturated by a model with linear chain topology.

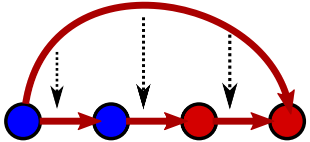

Area under memory curve
Memory lifetime bounded by area under SNR curve:
SNR(lifetime) = 1
⟹ lifetime < $\mathcal{A}$.
This area has an upper bound:
$\mathcal{A} \leq \sqrt{N}(M-1)/r$.
Saturated by a model with linear chain topology.

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

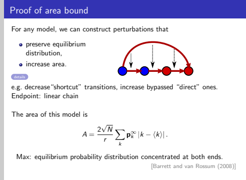
e.g. decrease "shortcut" transitions, increase bypassed "direct" ones.
Endpoint: linear chain

The area of this model is

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

Max: equilibrium probability distribution concentrated at both ends.

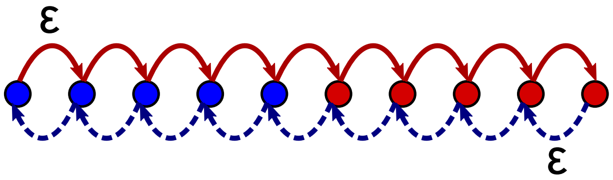
[Barrett and van Rossum (2008)]



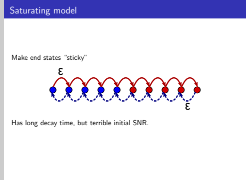
1. relies on order & technical condition
2. max given $\mathbf{p}^\infty$
3. now max wrt. $\mathbf{p}^\infty$
4. keep c.o.m. in middle
5. similar result, slightly different conditions: linear weights, mutual info

Saturating model

Make end states “sticky”



Has long decay time, but terrible initial SNR.



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:

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

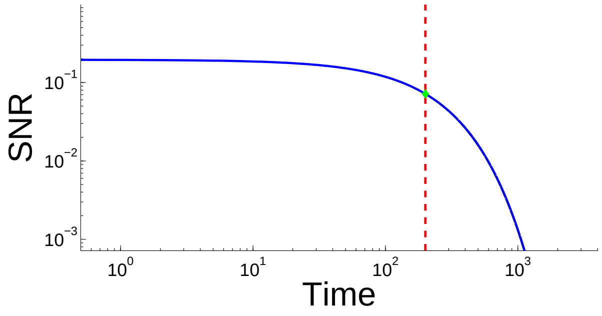
We can maximise wrt. $\mathcal{I}_a, \tau_a$.

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. $\mathcal{I}_a, \tau_a$.

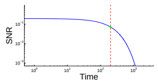
1. from eigenmode decomposition
2. from initial, area bounds

Constructing the envelope



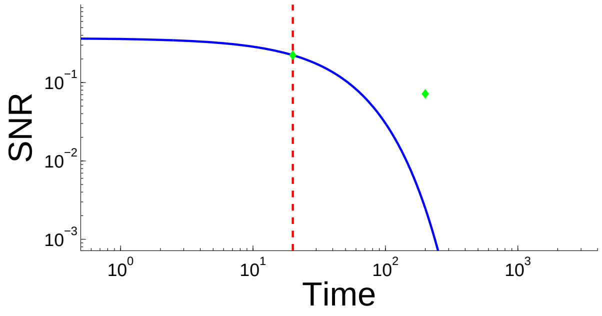
Maximise SNR at one time

Constructing the envelope

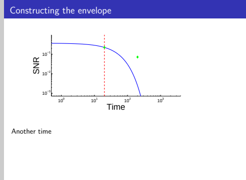


Maximise SNR at one time

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

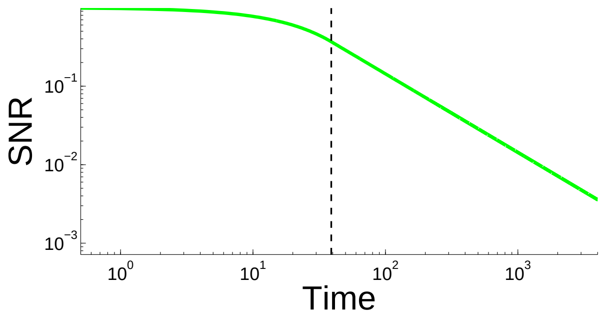


Another time

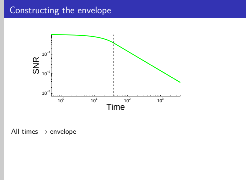


1. One exp. only constrains SNR at that time, not others
2. get another bound

Constructing the envelope

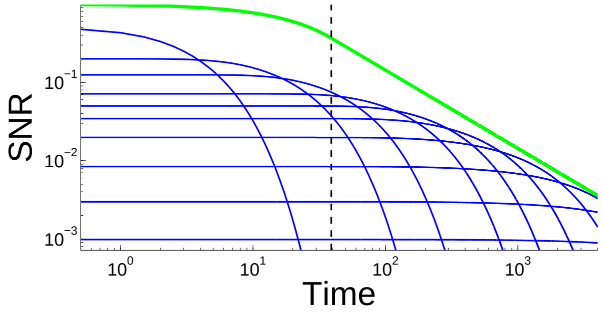


All times $\rightarrow$ envelope



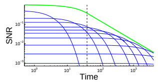
1. One exp. only constrains SNR at that time, not others
2. get another bound
3. vary time of max. no curve can cross this.
4. Regions: init(1); area(1,2)
5. is it tight? can any constrained set of exps be acheived?

Constructing the envelope



Memory curves of example models.

Constructing the envelope

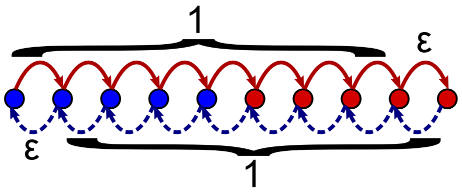


Memory curves of example models.

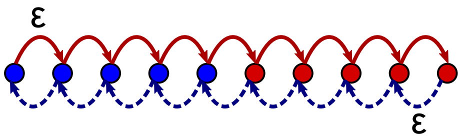
1. One exp. only constrains SNR at that time, not others
2. get another bound
3. vary time of max. no curve can cross this.
4. Regions: init(1); area(1,2)
5. is it tight? can any constrained set of exps be acheived?
6. no
7. One exp. discuss models later

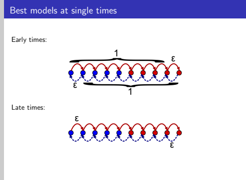
Best models at single times

Early times:



Late times:





1. shorten length of chain, keeping deterministic
2. Area maximising.
3. two mechs for slowing forgetting: time (lower trans prob) and space (diffusion length)

Additional constraint

Conjecture: additional constraint

$$\mathcal{I}_a \sqrt{\tau_a} \leq \mathcal{O}(1).$$

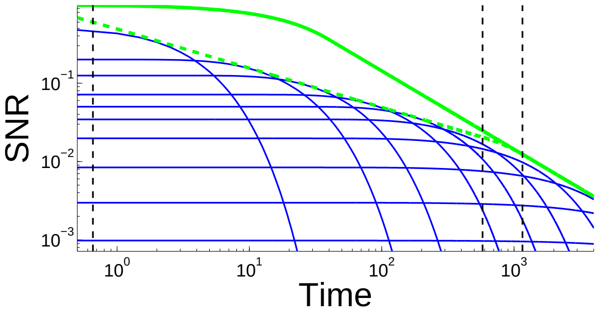
Saturated by a diffusive chain:

$$\text{SNR}(0) \sim \frac{1}{n}, \qquad \text{time-scale} \sim n^2.$$

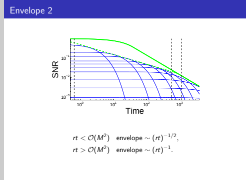
Additional constraint
Conjecture: additional constraint
$\mathcal{I}_a \sqrt{\tau_a} \leq \mathcal{O}(1).$
Saturated by a diffusive chain:
$\text{SNR}(0) \sim \frac{1}{n}, \qquad \text{time-scale} \sim n^2.$

1. Tested experimentally. Discuss later

Envelope 2



$$rt < \mathcal{O}(M^2) \quad \text{envelope} \sim (rt)^{-1/2},$$
$$rt > \mathcal{O}(M^2) \quad \text{envelope} \sim (rt)^{-1}.$$



1. dashed: conjecture. tight.
2. earlier: diffusion limited. later: stochastic limited.
3. regions: init(1); sqrt(2,3); area(3,4)
4. Benna-Fusi hugs envelope? cascade $\sim t^{-3/4}$

Lifetime bound

Lifetime of a memory bounded by where envelope crosses 1

$$N > \mathcal{O}(M^2) \quad \Longrightarrow \text{lifetime} \leq \frac{\sqrt{N}(M-1)}{er},$$

$$N > \mathcal{O}(M^2) \quad \Longrightarrow \text{lifetime} \leq \frac{\gamma^2 N}{2er}.$$

Lifetime bound

Lifetime of a memory bounded by where envelope crosses 1

$N > \mathcal{O}(M^2) \Longrightarrow \text{lifetime} \leq \frac{\sqrt{N}(M-1)}{er}$,
$N > \mathcal{O}(M^2) \Longrightarrow \text{lifetime} \leq \frac{\gamma^2 N}{2er}$.

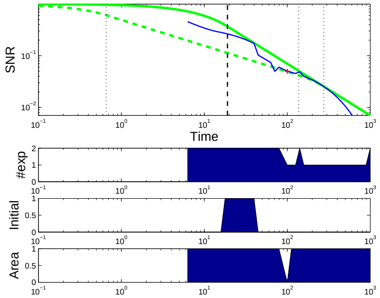
1. $\gamma \sim \mathcal{O}(1)$ constant in additional constaint
2. First t^{-1} assumes M low. Second $t^{-1/2}$ applies to Benna-Fusi.
3. Independent synapses?

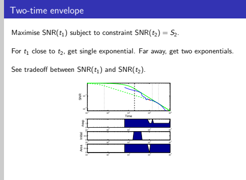
Two-time envelope

Maximise SNR(t_1) subject to constraint SNR(t_2) $= S_2$.

For t_1 close to t_2, get single exponential. Far away, get two exponentials.

See tradeoff between SNR(t_1) and SNR(t_2).





1. Max at multiple times, → multiple timescales? cascade? Benna-Fusi?
2. numerics not working. 2 exp solution need to solve 2 transcendental equations.

Additional constraint: other forms?

Involving eigenmodes:

$$\mathcal{I}_a\sqrt{\tau_a}, \qquad \sum_a \mathcal{I}_a\sqrt{\tau_a}, \qquad \sum_a \mathcal{I}_a^2\tau_a.$$

Not involving eigenmodes

$$\mathcal{A}\times \mathsf{SNR}(0), \qquad \int\!\mathrm{d}t\; \mathsf{SNR}(t)^2.$$

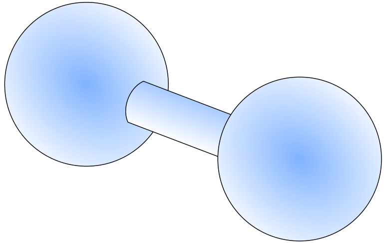
Additional constraint: other forms?
Involving eigenmodes:
$\mathcal{I}_a\sqrt{\tau_a}, \quad \sum_a \mathcal{I}_a\sqrt{\tau_a}, \quad \sum_a \mathcal{I}_a^2\tau_a.$
Not involving eigenmodes
$\mathcal{A}\times \mathsf{SNR}(0), \quad \int\!\mathrm{d}t\; \mathsf{SNR}(t)^2.$

1. as one-time max only involved one exp, would also work
2. right units
3. easier to work with?
4. L2 doesn't have nice expression in terms of matrices

Cheeger inequality

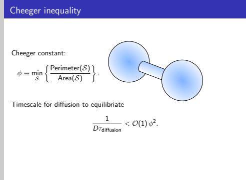
Cheeger constant:

$$\phi \equiv \min_{\mathcal{S}} \left\{ \frac{\text{Perimeter}(\mathcal{S})}{\text{Area}(\mathcal{S})} \right\}.$$



Timescale for diffusion to equilibriate

$$\frac{1}{D\tau_{\text{diffusion}}} < \mathcal{O}(1)\,\phi^2.$$



1. split into two pieces. pick smaller. higher dim.
2. bottleneck
3. if we want fast diffusion, need lhs large $\to$ no bottlenecks.
4. purely geometric
5. also inequality in other direction: want slow diffusion $\to$ need bottleneck. Not useful for us

Cheeger inequality: Markov chains

Cheeger constant:

$$\phi \equiv \min_{\mathcal{S}} \left\{ \frac{\mathbf{\Phi}_{\mathcal{S}\mathcal{S}^c}}{\mathbf{p}^\infty(\mathcal{S})} \right\}.$$

Timescale to equilibriate:

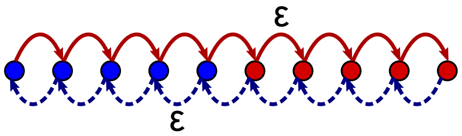
$$\frac{1}{\max_a \tau_a} < \mathcal{O}(1)\, \phi^2.$$

Simple proof assuming detailed balance. [Sinclair and Jerrum (1989)]
More complicated proof for general case. [Lawler and Sokal (1988)]

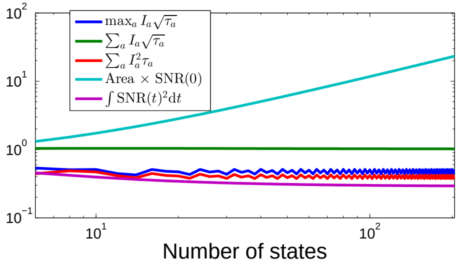
1. split states into two subsets. pick smaller.
2. again bottleneck
3. denominator varies
4. $\mathcal{O}(1)$ bit differs
5. bottleneck need not be between strong & weak.

Counter examples?

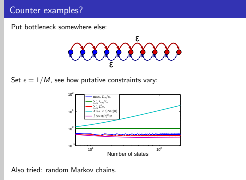
Put bottleneck somewhere else:



Set $\epsilon = 1/M$, see how putative constraints vary:



Also tried: random Markov chains.



1. eq prob concentrated near middle. bottleneck at ϵ.
2. high inital snr and long timescale in different modes.
3. only eigenmode dependent constraints survive (and L2 - but difficult to wor
   with).

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 $> \mathcal{O}(M^2)$.
- For times $< \mathcal{O}(M^2)$: conjecture that the model that reaches the envelope uses deterministic transitions $\rightarrow$ diffusive forgetting.

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 $> \mathcal{O}(M^2)$.
- For times $< \mathcal{O}(M^2)$: conjecture that the model that reaches the envelope uses deterministic transitions $\rightarrow$ diffusive forgetting.

Acknowledgements

Thanks to:

- Surya Ganguli
- Stefano Fusi
- Marcus Benna

Acknowledgements
Thanks to:
- Surya Ganguli
- Stefano Fusi
- Marcus Benna

1. Last slide!

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, PubMed:19401593.
6

D. J. Amit and S. Fusi.
"Constraints on learning in dynamic synapses".
Network: Computation in Neural Systems, 3(4):443–464, 1992.
7

D. J. Amit and S. Fusi.
"Learning in neural networks with material synapses".
Neural Computation, 6(5):957–982, 1994.
7 13

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, PubMed:19401593.
6
D. J. Amit and S. Fusi.
"Constraints on learning in dynamic synapses".
Network: Computation in Neural Systems, 3(4):443–464, 1992.
7
D. J. Amit and S. Fusi.
"Learning in neural networks with material synapses".
Neural Computation, 6(5):957–982, 1994.
7 13

S. Fusi, P. J. Drew, and L. F. Abbott.
"Cascade models of synaptically stored memories".
Neuron, 45(4):599–611, Feb 2005, PubMed:15721245.
13

S. Fusi and L. F. Abbott.
"Limits on the memory storage capacity of bounded synapses".
Nat. Neurosci., 10(4):485–493, Apr 2007, PubMed:17351638.
13

A. B. Barrett and M. C. van Rossum.
"Optimal learning rules for discrete synapses".
PLoS Comput. Biol., 4(11):e1000230, Nov 2008, PubMed:19043540.
20

2013-04-11

References II

S. Fusi, P. J. Drew, and L. F. Abbott.
"Cascade models of synaptically stored memories".
Neuron, 45(4):599–611, Feb 2005, PubMed:15721245.

S. Fusi and L. F. Abbott.
"Limits on the memory storage capacity of bounded synapses".
Nat. Neurosci., 10(4):485–493, Apr 2007, PubMed:17351638.

A. B. Barrett and M. C. van Rossum.
"Optimal learning rules for discrete synapses".
PLoS Comput. Biol., 4(11):e1000230, Nov 2008, PubMed:19043540.

Alistair Sinclair and Mark Jerrum.

"Approximate counting, uniform generation and rapidly mixing Markov chains".

Information and Computation, 82(1):93 – 133, 1989.

ISSN 0890-5401.

35

Gregory F. Lawler and Alan D. Sokal.

"Bounds on the L^2 Spectrum for Markov Chains and Markov Processes: A Generalization of Cheeger's Inequality".

Transactions of the American Mathematical Society, 309(2):557–580, 1988.

ISSN 00029947.

35

2013-04-11

References III

Alistair Sinclair and Mark Jerrum.
"Approximate counting, uniform generation and rapidly mixing Markov chains".
Information and Computation, 82(1):93 – 133, 1989.
ISSN 0890-5401.
35

Gregory F. Lawler and Alan D. Sokal.
"Bounds on the L^2 Spectrum for Markov Chains and Markov Processes: A Generalization of Cheeger's Inequality".
Transactions of the American Mathematical Society, 309(2):557–580, 1988.
ISSN 00029947.
35

References IV

J.G. Kemeny and J.L. Snell.
Finite markov chains.
Springer, 1960.
43

References IV
J.G. Kemeny and J.L. Snell.
Finite markov chains.
Springer, 1960.
43

Techinical detail: ordering states

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

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

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

[Kemeny and Snell (1960)]

We define:

$$\eta^+_i = \sum_{j \in \text{strong}} \mathbf{T}_{ij} \mathbf{p}^\infty_j, \qquad \eta^-_i = \sum_{j \in \text{weak}} \mathbf{T}_{ij} \mathbf{p}^\infty_j.$$

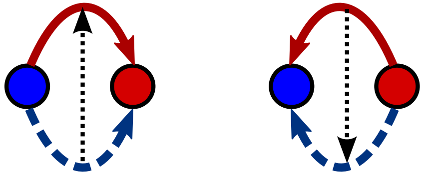
They can be used to arrange the states in an order (increasing η^- or decreasing η^+). back

Techinical detail: ordering states
Let $\mathbf{T}_{ij}$ = mean first passage time from state i to state j. Then:
$\eta = \sum_j \mathbf{T}_{ij} \mathbf{p}^\infty_j,$
is independent of the initial state i (Kemeney's constant).
[Kemeny and Snell (1960)]
We define:
$\eta^+_i = \sum_{j \in \text{strong}} \mathbf{T}_{ij} \mathbf{p}^\infty_j, \quad \eta^-_i = \sum_{j \in \text{weak}} \mathbf{T}_{ij} \mathbf{p}^\infty_j.$
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, $\Longrightarrow$ two orders same

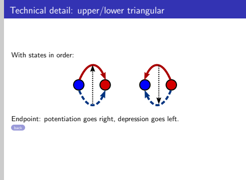
Technical detail: upper/lower triangular

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!