

2013-04-10

Complex synapses

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

Subhaneil Lahiri

Stanford University, Applied Physics

April 10, 2013

A general theory of learning and memory with Complex Synapses

based on work with Surya Ganguli

Subhaneil Lahiri

Stanford University, Applied Physics

April 10, 2013

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.

2013-04-10

Complex synapses

└ Introduction

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

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

2013-04-10

Complex synapses

└ Outline

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

Outline

1 Why complex synapses?

2 Modelling synaptic complexity

3 Upper bounds

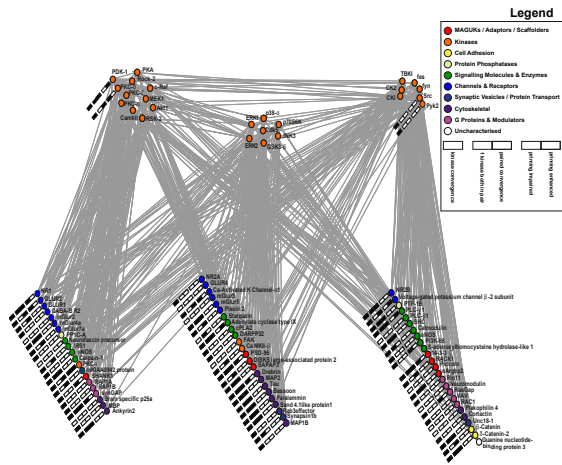
4 Envelope memory curve

Complex synapses
└ Why complex synapses?

Why complex synapses?

Why complex synapses?

Complex synapse



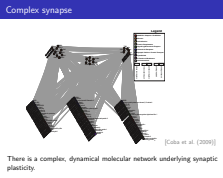
[Coba et al. (2009)]

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

2013-04-10

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"

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.

2013-04-10

Complex synapses

└ Why complex synapses?

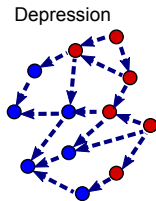
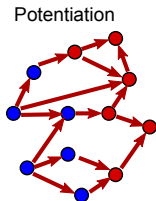
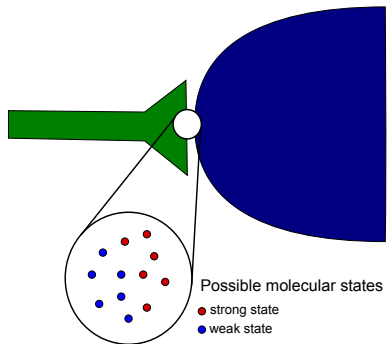
└ Storage capacity of synaptic memory

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

- Complex synapses
 - Modelling synaptic complexity

Modelling synaptic complexity

Modelling synaptic complexity

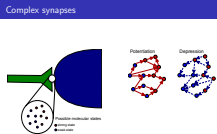


2013-04-10

Complex synapses

└ 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 .

Complex synapses

└ Modelling synaptic complexity

└ Simplifying assumptions

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 $i \rightarrow j$, given pot/dep
5. looks like binary synapse from outside. Inside...

- 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 .

2013-04-10

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

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

Eventually, this will settle into the equilibrium distribution:

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

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

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

Eventually, this will settle into the equilibrium distribution:

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

1. for this one, we keep track of pot/dep, look for inc/dec of \mathbf{w}
2. \mathbf{W}^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.

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

$$\text{Noise} = \text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty))$$

Related to reconstruction probability of single synapses.

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

2013-04-10

Complex synapses

└ Modelling synaptic complexity

└ 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. subtract baseline, some overlap even w/o encoding
6. if we ignore correlations...

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

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

$$\text{Noise} = \text{Var}(\vec{w}_{\text{ideal}} \cdot \vec{w}(\infty))$$

Related to reconstruction probability of single synapses.

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

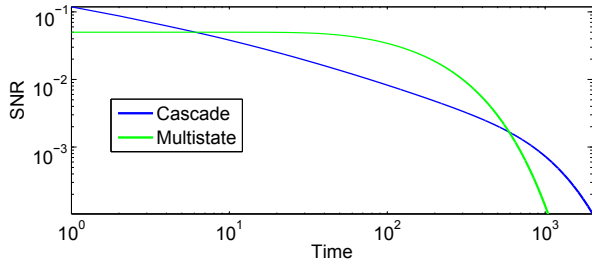
Example models

Two example models of complex synapses.



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

These have different memory storage properties

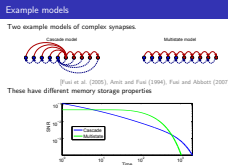


2013-04-10

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.

- 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?

2013-04-10

Complex synapses

└─Modelling synaptic complexity

└─Questions

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 given by

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

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

Eigenmode decomposition:

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

2013-04-10

Complex synapses

└ Modelling synaptic complexity

└ Memory curve 2

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

Memory curve given by

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

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

Eigenmode decomposition:

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

Section 3

Upper bounds

2013-04-10

Complex synapses
└ Upper bounds

Section 3

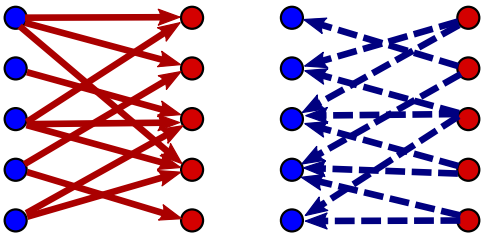
Upper bounds

Initial SNR as flux

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

$$\text{SNR}(0) \leq \frac{4\sqrt{N}}{r} \Phi_{-+}.$$

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



2013-04-10

Complex synapses

└ Upper bounds

└ Initial SNR

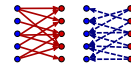
└ Initial SNR as flux

Initial SNR as flux

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

$$\text{SNR}(0) \leq \frac{4\sqrt{N}}{r} \Phi_{-+}$$

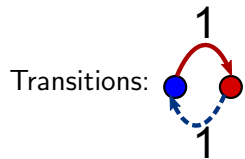
Max when potentiation guarantees $\mathbf{w} \rightarrow +1$,
depression guarantees $\mathbf{w} \rightarrow -1$.



1. usually saturated: pot never dec, dep never inc
2. transitions out of one node sum to 1
3. 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:



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

Maximal initial SNR:

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

2013-04-10

Complex synapses
└ Upper bounds
└ Initial SNR
└ Two-state model

Two-state model

Two-state model equivalent to previous slide:

Transitions:

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

Maximal initial SNR:

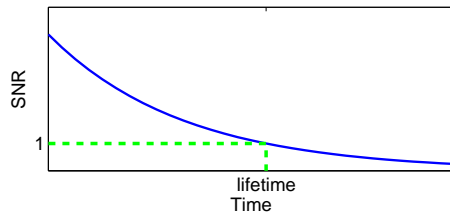
$$\text{SNR}(0) \leq \sqrt{N}.$$

1. decays very quickly
2. $f^{\text{pot}} = \frac{1}{2}$

Area under memory curve

Memory lifetime bounded by area under SNR curve:

$$\text{SNR}(\text{lifetime}) = 1 \\ \Rightarrow \text{lifetime} < \mathcal{A}.$$



This area has an upper bound:

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

Saturated by a model with linear chain topology.

2013-04-10

Complex synapses

└ Upper bounds

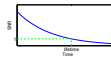
└ Area under memory curve

└ Area under memory curve

Area under memory curve

Memory lifetime bounded by area under SNR curve:

$$\text{SNR}(\text{lifetime}) = 1 \\ \Rightarrow \text{lifetime} < \mathcal{A}.$$



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: $N = \text{\#synapses}$, $M = \text{\#states}$
4. proof next slide

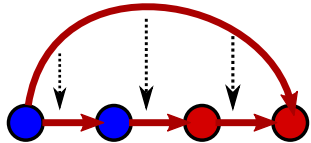
Proof of area bound

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

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

Max: equilibrium probability distribution concentrated at both ends.

[Barrett and van Rossum (2008)]



2013-04-10

Complex synapses

└ Upper bounds

└ Area under memory curve

└ Proof of area bound

Proof of area bound

For any model, we can construct perturbations that

- preserve equilibrium distribution,
- increase area.

e.g.

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)]



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

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

subject to constraints:

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

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

SNR curve:

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

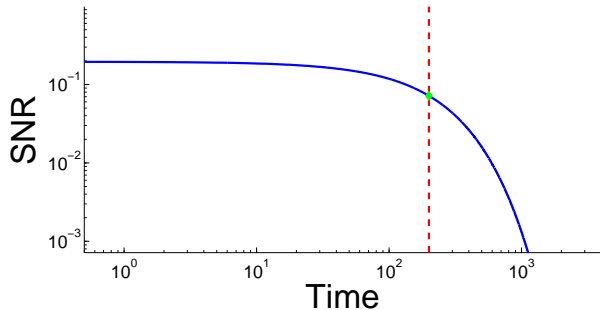
subject to constraints:

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

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

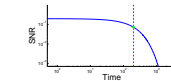
2013-04-10

Complex synapses

└ Envelope memory curve

└ Constructing the envelope

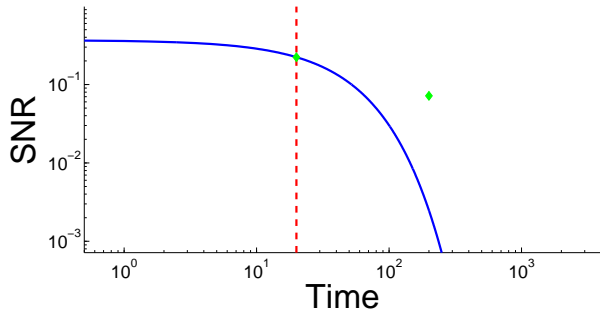
Constructing the envelope



Maximise SNR at one time

1. only constrains SNR at that time, not others
2. get another bound
3. vary time of max. no curve can cross this. is it tight? can any constrained set of exps be acheived?
4. no

Constructing the envelope



Another time

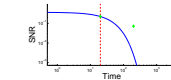
2013-04-10

Complex synapses

└ Envelope memory curve

└ Constructing the envelope

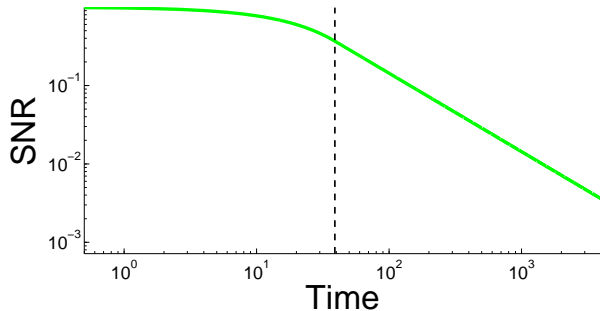
Constructing the envelope



Another time

1. only constrains SNR at that time, not others
2. get another bound
3. vary time of max. no curve can cross this. is it tight? can any constrained set of exps be acheived?
4. no

Constructing the envelope



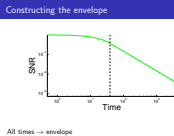
All times \rightarrow envelope

2013-04-10

Complex synapses

└ Envelope memory curve

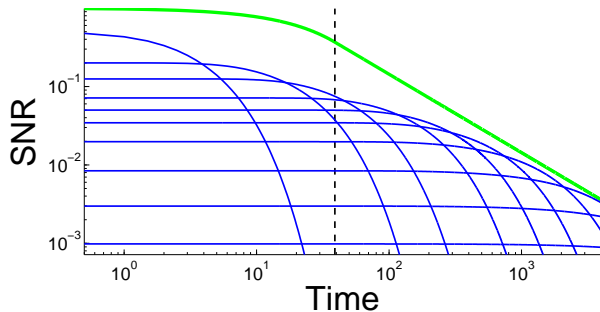
└ Constructing the envelope



All times \rightarrow envelope

1. only constrains SNR at that time, not others
2. get another bound
3. vary time of max. no curve can cross this. is it tight? can any constrained set of expts be achieved?
4. no

Constructing the envelope



Memory curves of example models.

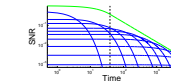
2013-04-10

Complex synapses

└ Envelope memory curve

└ Constructing the envelope

Constructing the envelope



Memory curves of example models.

1. only constrains SNR at that time, not others
2. get another bound
3. vary time of max. no curve can cross this. is it tight? can any constrained set of expts be acheived?
4. no

Thanks to:

- Surya Ganguli
- Stefano Fusi
- Marcus Benna

2013-04-10

Complex synapses

└ Envelope memory curve

└ Acknowledgements

Thanks to:
• Surya Ganguli
• Stefano Fusi
• Marcus Benna

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, PubMed:19401593.

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.

13

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡

2013-04-10

Complex synapses

Envelope memory curve

References

- 14** M. P. Coles, A. J. Pickensinger, J. M. Collins, M. V. Kozanits, R. T. Umes, S. Swamy, M. D. Conning, J. S. Choudhary, and S. G. Grant. "Neurotransmitters drive combinatorial multistate postsynaptic density networks". *Sci Signal*, 2(68):ra19, 2009. PubMed: 19401593.
- 15** D. J. Amit and S. Fusi. "Constraints on learning in dynamic synapses". *Network: Computation in Neural Systems*, 3(4):443–464, 1992.
- 16** D. J. Amit and S. Fusi. "Learning in neural networks with material synapses". *Neural Computation*, 6(5):957–982, 1994.



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

2013-04-10

Complex synapses

└ Envelope memory curve

└ References

- 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.



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

31

2013-04-10

Complex synapses
└ Envelope memory curve
└ References

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

Technical 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](#)

2013-04-10

Complex synapses

Envelope memory curve

- └ Technical detail: ordering states

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

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

$$\eta = \sum_j \mathbf{T}_j \mathbf{p}_j^{\text{SC}},$$

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

[Kameny and Snell (1980)]

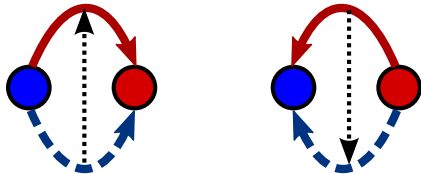
We define:

$$\eta_0^+ = \sum_{j \in \text{strong}} \mathbf{T}_0 \mathbf{p}_j^\infty, \quad \eta_0^- = \sum_{j \in \text{weak}} \mathbf{T}_0 \mathbf{p}_j^\infty.$$

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

Technical detail: upper/lower triangular

With states in order:



Endpoint: potentiation goes right, depression goes left.

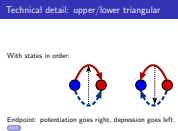
[back](#)

2013-04-10

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!