

Area maximisation in continuous time

Subhaneil Lahiri

July 10, 2012

Abstract

We try to find the continuous time Markov process that has the maximal area under the signal-to-noise curve.

Contents

1	Continuous time Markov processes	2
1.1	Notation	2
1.2	Fundamental matrix	3
1.3	First passage times	4
1.4	Mixing time (Kemeny's constant)	5
1.5	Sensitivity of equilibrium distribution	6
1.6	Subsets and flux	6
1.7	Lumpability	7
2	Signal-to-Noise ratio (SNR)	8
2.1	Framework	8
2.2	SNR curve	9
2.3	Degrees of freedom and constraints	10
2.4	Lumpability	10
2.5	Initial SNR and flux	11
2.6	Maximising initial SNR	11
3	Area maximisation	12
3.1	Area under signal-to-noise curve	12
3.2	Derivatives wrt. \mathbf{W}^\pm	14
3.2.1	Scaling mode	15
3.3	Kuhn-Tucker conditions	15
3.3.1	Triangularity	15
3.3.2	Increasing c_k	16
3.3.3	Shortcuts	17

3.3.4	Summary	17
3.4	Multi-state topology	18
A	Other quantities	19
A.1	Differences in c_k 's	20
A.2	Derivatives wrt. $\bar{\mathbf{T}}$	20
A.3	Initial SNR	21
A.4	Fixing mixing time (Kemeny's constant)	21
A.5	Scaling mode	22

1 Continuous time Markov processes

ContMarkov

In this section we'll provide a summary of all the relevant properties of ergodic Markov chains in continuous time. It is a straightforward generalisation of material that can be found in [1] with some ideas from [2].

1.1 Notation

sec:not

For any matrix \mathbf{A} , we define matrices \mathbf{A}^{dg} and $\bar{\mathbf{A}}$ as

$$\mathbf{A}_{ij}^{\text{dg}} \equiv \delta_{ij} \mathbf{A}_{ij}, \quad \bar{\mathbf{A}} \equiv \mathbf{A} - \mathbf{A}^{\text{dg}}. \quad (1) \quad \text{eq:dgdef}$$

We let \mathbf{e} denote a column-vector of ones and $\mathbf{E} = \mathbf{e}\mathbf{e}^T$ denote a matrix of ones.

A Markov process is described by a matrix of transitions rates, \mathbf{Q}_{ij} , from state i to j . The probabilities of being in each state at time t , the row-vector $\mathbf{p}(t)$, evolve according to

$$\frac{d\mathbf{p}(t)}{dt} = \mathbf{p}(t)\mathbf{Q}, \quad \mathbf{p}(t)\mathbf{e} = 1, \quad \mathbf{Q}\mathbf{e} = 0. \quad (2) \quad \text{eq:rowsum}$$

In Surya's notes, $\mathbf{Q} = r\mathbf{W}^F$.

The equilibrium probabilities, \mathbf{p}^∞ , satisfy

$$\mathbf{p}^\infty \mathbf{Q} = 0, \quad \mathbf{p}^\infty \mathbf{e} = 1. \quad (3) \quad \text{eq:equilibr}$$

As we assume an ergodic process, this eigenvalue is non-degenerate. If all other eigenvalues have strictly negative real parts, the process is regular (aperiodic).

We define additional matrices

$$\mathbf{\Lambda} \equiv (-\mathbf{Q}^{\text{dg}})^{-1}, \quad \mathbf{P} \equiv \mathbf{I} + \mathbf{\Lambda}\mathbf{Q}. \quad (4) \quad \text{eq:defDLP}$$

It can be shown that $\mathbf{\Lambda}_{ii}$ is the mean time it takes to leave state i and \mathbf{P}_{ij} is the probability the the next transition from state i goes to state j :

$$\mathbf{\Lambda}_{ii} = \frac{1}{\sum_{j \neq i} \mathbf{Q}_{ij}}, \quad \mathbf{P}_{ij} = \begin{cases} 0 & \text{if } i = j, \\ \frac{\mathbf{Q}_{ij}}{\sum_{k \neq j} \mathbf{Q}_{ik}} & \text{otherwise.} \end{cases} \quad (5) \quad \text{eq:LamdaPcmp}$$

Furthermore, we also define

$$\mathbf{D} \equiv \text{diag}(\mathbf{p}^\infty)^{-1}, \quad \implies \quad \mathbf{p}^\infty \mathbf{D} = \mathbf{e}^T. \quad (6) \quad \text{eq:pdotD}$$

1.2 Fundamental matrix

Definition 1: Fundamental matrix

$$\mathbf{Z} \equiv (-\mathbf{Q} + \mathbf{e}\boldsymbol{\pi})^{-1}, \quad (7) \quad \text{eq:funddef}$$

where $\boldsymbol{\pi}$ is any row-vector with $\boldsymbol{\pi}\mathbf{e} = 1/\tau \neq 0$.

Note that the canonical choice for the discrete time version, $\boldsymbol{\pi} = \mathbf{p}^\infty$, is not available here due to problems with units. It will be helpful to choose $\boldsymbol{\pi}$ to be independent of \mathbf{Q} , e.g. $\boldsymbol{\pi} = \mathbf{e}^T/(n\tau)$. All quantities that we calculate using \mathbf{Z} below will be independent of this choice.

Theorem 1:

The definition of \mathbf{Z} is valid, i.e. $(-\mathbf{Q} + \mathbf{e}\boldsymbol{\pi})$ is invertible.

Proof. Assume there exists an \mathbf{x} such that

$$(-\mathbf{Q} + \mathbf{e}\boldsymbol{\pi})\mathbf{x} = 0. \quad (8) \quad \text{eq:fundinvke}$$

Multiplying from the left with \mathbf{p}^∞ gives

$$\boldsymbol{\pi}\mathbf{x} = 0. \quad (9) \quad \text{eq:fundinvpi}$$

Substituting back into (8) gives

$$\mathbf{Q}\mathbf{x} = 0.$$

As we assume an ergodic process, the zero eigenvalue is non-degenerate. Therefore, $\mathbf{x} = \lambda\mathbf{e}$. Substituting this into (9) gives

$$\lambda\boldsymbol{\pi}\mathbf{e} = \frac{\lambda}{\tau} = 0.$$

As we defined $\boldsymbol{\pi}$ such that $1/\tau \neq 0$, this means $\lambda = 0 \implies \mathbf{x} = 0$. □

Corollary 2:

$$\boldsymbol{\pi}\mathbf{Z} = \mathbf{p}^\infty, \quad (10) \quad \text{eq:fundprob}$$

$$\mathbf{Z}\mathbf{e} = \tau\mathbf{e}, \quad (11) \quad \text{eq:fundrowsu}$$

$$\mathbf{I} + \mathbf{Q}\mathbf{Z} = \mathbf{e}\mathbf{p}^\infty, \quad (12) \quad \text{eq:fundQZ}$$

$$\mathbf{I} + \mathbf{Z}\mathbf{Q} = \tau\mathbf{e}\boldsymbol{\pi}. \quad (13) \quad \text{eq:fundZQ}$$

Proof. We can deduce (10) and (11) by pre/post-multiplying the following equations by \mathbf{Z} :

$$\begin{aligned} \mathbf{p}^\infty(-\mathbf{Q} + \mathbf{e}\boldsymbol{\pi}) &= \boldsymbol{\pi}, \\ (-\mathbf{Q} + \mathbf{e}\boldsymbol{\pi})\mathbf{e} &= \frac{\mathbf{e}}{\tau}. \end{aligned}$$

We can then deduce (12) and (13) by substituting these into

$$(-\mathbf{Q} + \mathbf{e}\boldsymbol{\pi})\mathbf{Z} = \mathbf{Z}(-\mathbf{Q} + \mathbf{e}\boldsymbol{\pi}) = \mathbf{I}.$$

□

1.3 First passage times

Definition 2: First passage time matrix

We define $\bar{\mathbf{T}}_{ij}$ as the mean time it takes the process to reach state j for the first time, starting from state i . We also define $\mathbf{T}_{ii}^{\text{dg}}$ as the mean time it takes the process to return to state i . As usual, $\mathbf{T} = \bar{\mathbf{T}} + \mathbf{T}^{\text{dg}}$.

Consider the first time the process leaves state i . On average, this will take time Λ_{ii} . With probability \mathbf{P}_{ij} , it will go directly to j , so the conditional mean time would be Λ_{ii} . On the other hand, if it goes to some other state, k , with probability \mathbf{P}_{ik} , the conditional mean time would be $\bar{\mathbf{T}}_{kj} + \Lambda_{ii}$. Combining these, we get the recursion relation

$$\begin{aligned} \mathbf{T}_{ij} &= \sum_{k \neq j} \mathbf{P}_{ik} (\bar{\mathbf{T}}_{kj} + \Lambda_{ii}) + \mathbf{P}_{ij} \Lambda_{ii} \\ &= \sum_{k \neq j} \mathbf{P}_{ik} \bar{\mathbf{T}}_{kj} + \sum_k \mathbf{P}_{ik} \Lambda_{ii} \\ &= \sum_{k \neq j} \mathbf{P}_{ik} \bar{\mathbf{T}}_{kj} + \Lambda_{ii}, \\ \mathbf{T} &= \mathbf{P} \bar{\mathbf{T}} + \Lambda \mathbf{E}, \end{aligned} \tag{14}$$

or, using (4),

$$\mathbf{T}^{\text{dg}} = \Lambda \mathbf{Q} \bar{\mathbf{T}} + \Lambda \mathbf{E}. \tag{15}$$

Theorem 3:

The recurrence times are given by

$$\mathbf{T}^{\text{dg}} = \Lambda \mathbf{D}. \tag{16}$$

Proof. Pre-multiply (15) by $\mathbf{p}^\infty \Lambda^{-1}$:

$$\mathbf{p}^\infty \Lambda^{-1} \mathbf{T}^{\text{dg}} = \mathbf{e}^T,$$

or in component form

$$\mathbf{p}_i^\infty \Lambda_{ii}^{-1} \mathbf{T}_{ii}^{\text{dg}} = 1.$$

This is the same as (16) in component form: $\mathbf{T}_{ii}^{\text{dg}} = \Lambda_{ii} / \mathbf{p}_i^\infty$. \square

I think the extra factor of Λ_{ii} , compared to the discrete case [1, Th.4.4.5], occurs because in this case we are demanding that the process leaves the initial state once before returning, whereas in the discrete case we only measure the time it takes to go to the initial state after the first time-step.

Theorem 4:

The solution of (15) is unique.

Proof. We have already seen that the diagonal part is determined by (15). Suppose we have two solutions for the off-diagonal part, $\bar{\mathbf{T}}$ and $\bar{\mathbf{T}}'$. Subtracting (15) for these two solutions gives

$$\Lambda \mathbf{Q} (\bar{\mathbf{T}} - \bar{\mathbf{T}}') = 0.$$

As we are assuming an ergodic process, the zero eigenvalue is non-degenerate. Therefore

$$\bar{\mathbf{T}} - \bar{\mathbf{T}}' = \mathbf{e}\mathbf{y},$$

for some row-vector \mathbf{y} . Looking at the diagonal components

$$\mathbf{y}_i = \bar{\mathbf{T}}_{ii} - \bar{\mathbf{T}}'_{ii} = 0,$$

and therefore $\bar{\mathbf{T}} - \bar{\mathbf{T}}' = 0$. □

Theorem 5:

The off-diagonal mean first passage times are given by

$$\bar{\mathbf{T}} = (\mathbf{E}\mathbf{Z}^{\text{dg}} - \mathbf{Z})\mathbf{D}. \quad (17) \quad \text{eq:fptfund}$$

Proof. We just need to show that this satisfies (15) with (16), and that the diagonal elements vanish. Premultiply (17) by $\mathbf{\Lambda}\mathbf{Q}$ and subtract from (15):

$$\begin{aligned} \mathbf{T}^{\text{dg}} - \mathbf{\Lambda}\mathbf{Q}\bar{\mathbf{T}} &= \mathbf{\Lambda}(\mathbf{I} + \mathbf{Q}\mathbf{Z})\mathbf{D} \\ &= \mathbf{\Lambda}\mathbf{e}\mathbf{p}^\infty\mathbf{D} \\ &= \mathbf{\Lambda}\mathbf{E}, \end{aligned}$$

where we used (12) and (6). The vanishing of the diagonal elements follows trivially from the component form of (17):

$$\bar{\mathbf{T}}_{ij} = \frac{\mathbf{Z}_{jj} - \mathbf{Z}_{ij}}{\mathbf{p}_j^\infty}. \quad (18) \quad \text{eq:fptfundcm}$$

□

1.4 Mixing time (Kemeny's constant)

ec:mixtime
kemenyconst

Theorem 6: The quantity

$$\eta \equiv \sum_j \bar{\mathbf{T}}_{ij} \mathbf{p}_j^\infty \quad (19) \quad \text{eq:mixdef}$$

is independent of i .

Proof. We use (17), (11) and the transpose of (6):

$$\begin{aligned} \bar{\mathbf{T}}(\mathbf{p}^\infty)^\text{T} &= (\mathbf{E}\mathbf{Z}^{\text{dg}} - \mathbf{Z})\mathbf{D}(\mathbf{p}^\infty)^\text{T} \\ &= (\mathbf{e}\mathbf{e}^\text{T}\mathbf{Z}^{\text{dg}} - \mathbf{Z})\mathbf{e} \\ &= (\mathbf{e}^\text{T}\mathbf{Z}^{\text{dg}}\mathbf{e})\mathbf{e} - \mathbf{Z}\mathbf{e} \\ &= (\text{tr } \mathbf{Z} - \tau)\mathbf{e}. \end{aligned}$$

which proves (19) with $\eta = \text{tr } \mathbf{Z} - \tau$. □

Note that it is essential that we use $\bar{\mathbf{T}}$ and not \mathbf{T} here, as that would lead to $\eta_i = \eta + \mathbf{\Lambda}_{ii}$, unlike the discrete time version [3] where this would only shift η by 1.

1.5 Sensitivity of equilibrium distribution

Suppose that the Markov process, defined by \mathbf{Q} , depends on some parameter α . Differentiating (7) gives

$$\frac{d\mathbf{Z}}{d\alpha} = \mathbf{Z} \frac{d\mathbf{Q}}{d\alpha} \mathbf{Z}. \quad (20) \quad \text{eq:diffZ}$$

We can substitute this into the derivative of (10):

$$\frac{d\mathbf{p}^\infty}{d\alpha} = \boldsymbol{\pi} \mathbf{Z} \frac{d\mathbf{Q}}{d\alpha} \mathbf{Z} = \mathbf{p}^\infty \frac{d\mathbf{Q}}{d\alpha} \mathbf{Z}. \quad (21) \quad \text{eq:diffp}$$

We can rewrite this in component form and use the fact that $\mathbf{Q}_{ii} = -\sum_{i \neq j} \mathbf{Q}_{ij}$:

$$\begin{aligned} \frac{d\mathbf{p}_k^\infty}{d\alpha} &= \sum_{i,j} \mathbf{p}_i^\infty \frac{d\mathbf{Q}_{ij}}{d\alpha} \mathbf{Z}_{jk} \\ &= \sum_{i \neq j} \mathbf{p}_i^\infty \frac{d\mathbf{Q}_{ij}}{d\alpha} \mathbf{Z}_{jk} + \sum_i \mathbf{p}_i^\infty \frac{d\mathbf{Q}_{ii}}{d\alpha} \mathbf{Z}_{ik} \\ &= \sum_{i \neq j} \mathbf{p}_i^\infty \frac{d\mathbf{Q}_{ij}}{d\alpha} (\mathbf{Z}_{jk} - \mathbf{Z}_{ik}) \\ &= \sum_{i \neq j} \frac{d\mathbf{Q}_{ij}}{d\alpha} \mathbf{p}_i^\infty \mathbf{p}_k^\infty (\bar{\mathbf{T}}_{ik} - \bar{\mathbf{T}}_{jk}), \end{aligned} \quad (22) \quad \text{eq:diffpT}$$

which is the result of [4] that we need. Note that the summand vanishes for $i = j$, so we can drop the restriction $i \neq j$ from the range of the sum.

1.6 Subsets and flux

Let us denote the set of states by \mathcal{S} . Consider a subset $\mathcal{A} \subset \mathcal{S}$. We can define a projection operator onto this subset:

$$(\mathbf{I}^{\mathcal{A}})_{ij} = \begin{cases} 1 & \text{if } i = j \in \mathcal{A}, \\ 0 & \text{otherwise.} \end{cases} \quad (23) \quad \text{eq:proj}$$

We will use superscripts/subscripts to denote projection onto/summation over a subset:

$$\begin{aligned} \boldsymbol{\pi}^{\mathcal{A}} &= \boldsymbol{\pi} \mathbf{I}^{\mathcal{A}}, \quad \mathbf{M}^{\cdot \mathcal{A}} = \mathbf{M} \mathbf{I}^{\mathcal{A}}, \quad \mathbf{M}^{\mathcal{A} \cdot} = \mathbf{I}^{\mathcal{A}} \mathbf{M}, \quad \mathbf{x}^{\mathcal{A}} = \mathbf{I}^{\mathcal{A}} \mathbf{x}, \\ \boldsymbol{\pi}_{\mathcal{A}} &= \boldsymbol{\pi} \mathbf{e}^{\mathcal{A}}, \quad \mathbf{M}_{\cdot \mathcal{A}} = \mathbf{M} \mathbf{e}^{\mathcal{A}}, \quad \mathbf{M}_{\mathcal{A} \cdot} = (\mathbf{e}^{\mathcal{A}})^{\mathbf{T}} \mathbf{M}, \quad \mathbf{x}_{\mathcal{A}} = (\mathbf{e}^{\mathcal{A}})^{\mathbf{T}} \mathbf{x}, \end{aligned} \quad (24) \quad \text{eq:projsum}$$

where $\boldsymbol{\pi}$ is a row vector, \mathcal{M} is a matrix and \mathbf{x} is a column vector.

We can define a flux matrix, a.k.a. ergodic flow:

$$\boldsymbol{\Phi} = \mathbf{D}^{-1} \mathbf{Q}, \quad \Phi_{ij} = \mathbf{p}_i^\infty \mathbf{Q}_{ij}. \quad (25) \quad \text{eq:flux}$$

This measures the flow of probability between states in the equilibrium distribution. Detailed balance, a.k.a. reversibility, is equivalent to $\boldsymbol{\Phi} = \boldsymbol{\Phi}^{\mathbf{T}}$.

The flux between two subsets is a particularly useful quantity:

$$\Phi_{AB} = \mathbf{p}^{\infty A} \mathbf{Q} \mathbf{e}^B. \quad (26) \quad \text{eq:subflux}$$

One can show that

$$\Phi_{AA^c} = \Phi_{A^cA} = -\Phi_{AA} = -\Phi_{A^cA^c} \quad (27) \quad \text{eq:compflux}$$

using $(\mathbf{p}^{\infty A} + \mathbf{p}^{\infty A^c}) \mathbf{Q} = 0$ and $\mathbf{Q}(\mathbf{e}^A + \mathbf{e}^{A^c}) = 0$.

1.7 Lumpability

Suppose we have partitioned the states into disjoint subsets, $\{\mathcal{A}_\alpha\}$:

$$\bigcup_{\alpha} \mathcal{A}_\alpha = \mathcal{S}, \quad \mathcal{A}_\alpha \cap \mathcal{A}_\beta = \delta_{\alpha\beta} \mathcal{A}_\alpha. \quad (28) \quad \text{eq:partition}$$

We will use α instead of \mathcal{A}_α in superscripts and subscripts in the following. The fact that these subsets are disjoint and exhaustive allows us to define the function

$$\sigma(i) = \alpha \quad \Longleftrightarrow \quad i \in \mathcal{A}_\alpha. \quad (29) \quad \text{eq:whichsey}$$

We can use this partition to define a new stochastic process associated with the original Markov chain. At time t , if the state of the original process is i , the state of the new process is $\sigma(i)$.

One may ask if this new process is a Markov chain. The answer is yes, if the original Markov chain has a property called lumpability wrt. the partition (see [1, §6.3] for the discrete time version and [5, 6] for continuous time):

$$\sigma(i) = \sigma(j) \quad \Longrightarrow \quad \mathbf{Q}_{i\alpha} = \mathbf{Q}_{j\alpha} \equiv \sum_{k \in \mathcal{A}_\alpha} \mathbf{Q}_{jk}, \quad (30) \quad \text{eq:lump}$$

i.e. the total transition rate from some state to some subset is the same for all starting states within the same subset. This common value is the transition rate for the new lumped Markov chain.

This can be rewritten with the aid of two matrices

$$U_{\alpha i} = \frac{\delta_{\alpha\sigma(i)}}{|\mathcal{A}_\alpha|}, \quad V_{i\alpha} = \delta_{\sigma(i)\alpha}. \quad (31) \quad \text{eq:lumpmats}$$

Left multiplication by U averages over subsets, right multiplication by V sums over subsets. For U , we chose the uniform measure in each subset. Any measure would work equally well, e.g. one proportional to the equilibrium distribution:

$$U_{\alpha i} = \frac{\mathbf{p}_i^{\infty \alpha}}{\mathbf{p}_\alpha^{\infty}}. \quad (32) \quad \text{eq:altlumpma}$$

One can show that $(UV)_{\alpha\beta} = \delta_{\alpha\beta}$. The matrix VU is also interesting. It has a block diagonal structure, with each block corresponding to a subset. Each block is a discrete-time ergodic Markov matrix (it is an independent trials process with probabilities given by

the measure chosen for U). This means that the right eigenvectors with eigenvalue 1 will be those that are constant in each subset:

$$VU\mathbf{x} = \mathbf{x} \quad \Longleftrightarrow \quad \mathbf{x} = \sum_{\alpha} x_{\alpha} \mathbf{e}^{\alpha}. \quad (33) \quad \text{eq:setconst}$$

This allows us to write the lumpability condition (30), and the transition matrix for the lumped process compactly:

$$VU\mathbf{Q}V = \mathbf{Q}V, \quad \hat{\mathbf{Q}} = U\mathbf{Q}V. \quad (34) \quad \text{eq:lumpcompa}$$

By induction, one can show that similar relations hold for all powers:

$$VU\mathbf{Q}^nV = \mathbf{Q}^nV, \quad \hat{\mathbf{Q}}^n = U\mathbf{Q}^nV, \quad (35) \quad \text{eq:lumppower}$$

and, via the Taylor series, for the exponential as well:

$$VUe^{t\mathbf{Q}}V = e^{t\mathbf{Q}}V, \quad e^{t\hat{\mathbf{Q}}} = Ue^{t\mathbf{Q}}V. \quad (36) \quad \text{eq:lumpexp}$$

The equilibrium distribution of the lumped process is given by

$$\hat{\mathbf{p}}^{\infty} = \mathbf{p}^{\infty}V. \quad (37) \quad \text{eq:lumpeq}$$

2 Signal-to-Noise ratio (SNR)

sec:SNR

In this section we will look at the signal-to-noise curve, and put an upper bound on its initial value. We will only consider ergodic Markov chains. Transient states would be unoccupied in equilibrium and would not be accessed by the signal creation process, therefore they could be removed from the analysis. Absorbing chains are degenerate cases: they have zero initial signal but infinite decay times, so they can only be approached as the limit of a sequence of ergodic chains.

2.1 Framework

:framework

The individual potentiation/depression events will be described by *discrete*-time Markov chains:

$$\mathbf{M}^{\pm} \equiv \mathbf{I} + \mathbf{W}^{\pm}, \quad \mathbf{M}^{\pm} \mathbf{e} = \mathbf{e}, \quad \mathbf{M}_{ij}^{\pm} \in [0, 1]. \quad (38) \quad \text{eq:MWdef}$$

The initial signal creation event occurs at time $t = 0$, but all subsequent potentiation/depression events occur at random times according to Poisson processes with rates rf^{\pm} , where $f^{+} + f^{-} = 1$ are the fraction of plasticity events that are potentiating/depressing respectively. This means that the “forgetting” process will be described by the *continuous*-time Markov chain:

$$\mathbf{Q} = r\mathbf{W}^F \equiv r(f^{+}\mathbf{W}^{+} + f^{-}\mathbf{W}^{-}). \quad (39) \quad \text{eq:forgettin}$$

We only require that this Markov chain is ergodic. The Markov chains described by \mathbf{M}^\pm need not be.

We assume that the probability distribution starts in the equilibrium distribution (3). During the initial signal creation, a fraction f^+ will change to $\mathbf{p}^\infty \mathbf{M}^+$ and a fraction f^- will change to $\mathbf{p}^\infty \mathbf{M}^-$. After this, probabilities will evolve according to (2).

2.2 SNR curve

As we ignore correlations between different synapses, the only quantity of interest will be the number of synapses with each strength. The signal can be defined as the increase in synaptic strength due to potentiation plus the decrease in synaptic strength due to depression. If there are N synapses and the strength of the synapse in each state is given by the column vector, \mathbf{w} , this is given by

$$S(t) = N f^+ [\mathbf{p}(t|+) - \mathbf{p}^\infty] \mathbf{w} - N f^- [\mathbf{p}(t|-) - \mathbf{p}^\infty] \mathbf{w}, \quad (40) \quad \text{eq:signal}$$

where $\mathbf{p}(t|\pm)$ is the probability of being in each state conditional on potentiation/depression at $t = 0$:

$$\mathbf{p}(t|\pm) = \mathbf{p}^\infty \mathbf{M}^\pm e^{rt \mathbf{W}^F}. \quad (41) \quad \text{eq:initprob}$$

We can rewrite the signal as

$$S(t) = N \mathbf{p}^\infty (f^+ \mathbf{W}^+ - f^- \mathbf{W}^-) e^{rt \mathbf{W}^F} \mathbf{w} - N (2f^+ f^-) \mathbf{p}^\infty (\mathbf{W}^+ - \mathbf{W}^-) e^{rt \mathbf{W}^F} \mathbf{w}. \quad (42) \quad \text{eq:signalr}$$

For the noise, we can use the standard deviation of the synaptic strength:

$$\text{Noise} = \sqrt{N \left(\sum_i \mathbf{p}_i^\infty \mathbf{w}_i^2 - (\mathbf{p}^\infty \mathbf{w})^2 \right)}. \quad (43) \quad \text{eq:noise}$$

From now on we shall assume that there are only two possible synaptic strengths, strong and weak, which we will describe by $\mathbf{w}_i = \pm 1$. This splits the set of states into two subsets:

$$+ = \{k | \mathbf{w}_k = +1\}, \quad - = \{k | \mathbf{w}_k = -1\}. \quad (44) \quad \text{eq:plusminus}$$

This simplifies the noise:

$$\text{Noise} = \sqrt{N (4\mathbf{p}_+^\infty \mathbf{p}_-^\infty)}. \quad (45) \quad \text{eq:noisepm}$$

The factors of $\mathbf{p}_+^\infty \mathbf{p}_-^\infty$ will have little effect on anything that follows, so we will drop them.¹ This leaves the signal-to-noise ratio as

$$\text{SNR}(t) = \sqrt{N} (2f^+ f^-) \mathbf{p}^\infty (\mathbf{W}^+ - \mathbf{W}^-) e^{rt \mathbf{W}^F} \mathbf{w}. \quad (46) \quad \text{eq:SNRcurve}$$

¹They will not affect the optimal transition rates for most quantities, and the optima will correspond to the symmetric case when $\mathbf{p}_+^\infty = \mathbf{p}_-^\infty = \frac{1}{2}$.

We can express this in terms of the one parameter family of transition matrices:

$$\begin{aligned}\mathbf{W}(\alpha) &= \alpha \mathbf{W}^+ + (1 - \alpha) \mathbf{W}^-, & \implies & \mathbf{W}^F = \mathbf{W}(f^+), \\ & & & \mathbf{W}^+ - \mathbf{W}^- = \frac{d\mathbf{W}}{d\alpha}, \\ & & & \mathbf{p}^\infty \frac{d\mathbf{W}}{d\alpha} = -\frac{d\mathbf{p}^\infty}{d\alpha} \mathbf{W}^F.\end{aligned}\tag{47} \quad \text{eq:Walpha}$$

Then (46) becomes

$$\text{SNR}(t) = \sqrt{N}(2f^+f^-) \frac{d\mathbf{p}^\infty}{d\alpha} (-\mathbf{W}^F) e^{rt\mathbf{W}^F} \mathbf{w}.\tag{48} \quad \text{eq:SNRalpha}$$

2.3 Degrees of freedom and constraints

We would like to use the components of \mathbf{W}^\pm as the independent degrees of freedom when we try to maximise quantities associated with the SNR curve (46). However, (38) imposes some restrictions on the space of allowed values. We can solve some of these by setting

$$\mathbf{W}_{ii}^\pm = -\sum_{j \neq i} \mathbf{W}_{ij}^\pm.\tag{49} \quad \text{eq:Wdiag}$$

Then we treat the off-diagonal components as the independent degrees of freedom. These must satisfy the following inequalities:

$$\mathbf{W}_{ij}^\pm \geq 0 \quad \text{for } i \neq j,\tag{50} \quad \text{eq:Wmin}$$

$$\sum_{j \neq i} \mathbf{W}_{ij}^\pm \leq 1.\tag{51} \quad \text{eq:Wmax}$$

The inequalities $\mathbf{W}_{ij}^\pm \leq 1$, $-1 \leq \mathbf{W}_{ii}^\pm \leq 0$ follow automatically from these.

In §3.2.1, we will see that the area under the SNR curve is invariant under the scaling $\mathbf{W}^\pm \rightarrow \lambda \mathbf{W}^\pm$. This means that we can ignore the constraint (51) when maximising the area: after the area has been maximised, we can use this degree of freedom to enforce (51) without changing the area. However, this scaling does change the SNR curve, in particular its initial value, so we can't ignore (51) when maximising that.

2.4 Lumpability

Suppose that we have a partition such that \mathbf{W}^+ and \mathbf{W}^- are simultaneously lumpable, and that all the states in each subset have the same synaptic strength (see §1.7):

$$VU\mathbf{W}^\pm V = \mathbf{W}^\pm V, \quad VU\mathbf{w} = \mathbf{w}.\tag{52} \quad \text{eq:lumpables}$$

We can define a new synapse with

$$\widehat{\mathbf{W}}^\pm = U\mathbf{W}^\pm V, \quad \widehat{\mathbf{w}} = U\mathbf{w}, \quad \widehat{\mathbf{p}}^\infty = \mathbf{p}^\infty V.\tag{53} \quad \text{eq:lumpedsyn}$$

This synapse has an SNR curve:

$$\begin{aligned}
\frac{\text{SNR}(t)}{\sqrt{N}(2f^+f^-)} &= \hat{\mathbf{p}}^\infty(\widehat{\mathbf{W}}^+ - \widehat{\mathbf{W}}^-)e^{rt\widehat{\mathbf{W}}^F}\widehat{\mathbf{w}}. \\
&= \mathbf{p}^\infty VU(\mathbf{W}^+ - \mathbf{W}^-)VUe^{rt\mathbf{W}^F}VU\mathbf{w}. \\
&= \mathbf{p}^\infty(\mathbf{W}^+ - \mathbf{W}^-)VUe^{rt\mathbf{W}^F}VU\mathbf{w}. \\
&= \mathbf{p}^\infty(\mathbf{W}^+ - \mathbf{W}^-)e^{rt\mathbf{W}^F}VU\mathbf{w}. \\
&= \mathbf{p}^\infty(\mathbf{W}^+ - \mathbf{W}^-)e^{rt\mathbf{W}^F}\mathbf{w}.
\end{aligned} \tag{54} \quad \text{eq:lumpedSNR}$$

i.e. the lumped process has exactly the same SNR as the original one.

2.5 Initial SNR and flux

Using the first line of (42), we can write the initial SNR as

$$\frac{\text{SNR}(0)}{\sqrt{N}} = I = (\mathbf{p}^{\infty+} + \mathbf{p}^{\infty-})(f^+\mathbf{W}^+ - f^-\mathbf{W}^-)(\mathbf{e}^+ - \mathbf{e}^-). \tag{55} \quad \text{eq:init}$$

Using $\mathbf{W}^\pm(\mathbf{e}^+ + \mathbf{e}^-) = 0$ and (27):

$$r\mathbf{p}^{\infty-}(f^+\mathbf{W}^+ + f^-\mathbf{W}^-)\mathbf{e}^+ = \Phi_{-+} = \Phi_{+-} = r\mathbf{p}^{\infty+}(f^+\mathbf{W}^+ + f^-\mathbf{W}^-)\mathbf{e}^-,$$

we can rewrite (55) as

$$I = \frac{4\Phi_{-+}}{r} - 4\mathbf{p}^{\infty+}f^+\mathbf{W}^+\mathbf{e}^- - 4\mathbf{p}^{\infty-}f^-\mathbf{W}^-\mathbf{e}^+. \tag{56} \quad \text{eq:initflux}$$

The last two terms are guaranteed to be negative, as the diagonal parts of \mathbf{W}^\pm cannot contribute. Therefore

$$\text{SNR}(0) \leq \frac{4\sqrt{N}\Phi_{-+}}{r}. \tag{57} \quad \text{eq:initfluxi}$$

This inequality is saturated if potentiation never takes it from a + state to a - state and depression never takes it from a - state to a + state.

2.6 Maximising initial SNR

The flux will be maximised if potentiation is guaranteed to take it to a + state and depression is guaranteed to take it to a - state:

$$\sum_{j \in +} \mathbf{W}_{ij}^+ = 1 \quad \forall i \in -, \quad \sum_{j \in -} \mathbf{W}_{ij}^- = 1 \quad \forall i \in +. \tag{58} \quad \text{eq:guarantee}$$

This implies that the process is lumpable to a two state system (see §1.7 and §2.4):

$$\begin{aligned}\mathbf{W}^+ &= \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{W}^- = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \\ \implies \mathbf{W}^F &= \begin{pmatrix} -f^+ & f^+ \\ f^- & -f^- \end{pmatrix}, \quad \mathbf{p}^\infty = (f^- \quad f^+), \quad (59) \quad \text{eq:binarylum}\end{aligned}$$

which leads to the initial SNR ²

$$\text{SNR}(0) \leq \sqrt{N}(4f^+f^-). \quad (60) \quad \text{eq:binarySNR}$$

This is maximised at $f^+ = f^- = \frac{1}{2}$:

$$\text{SNR}(0) \leq \sqrt{N}. \quad (61) \quad \text{eq:initmax}$$

3 Area maximisation

In this section we will find an upper bound on the area under the signal-to-noise curve. As in §2, we will only consider ergodic Markov chains. We will see in §3.4 that the optimal chain is absorbing, so it lies on the boundary of the (open) set of ergodic chains, but it still puts an upper bound on the area.

3.1 Area under signal-to-noise curve

The signal-to-noise curve is given by (48). The area is computed by integrating this

$$\begin{aligned}A &= \frac{\sqrt{N}(2f^+f^-)}{r} \frac{d\mathbf{p}^\infty}{d\alpha} \left[-e^{rt\mathbf{W}^F} \right]_0^\infty \mathbf{w} \\ &= \frac{\sqrt{N}(2f^+f^-)}{r} \frac{d\mathbf{p}^\infty}{d\alpha} (\mathbf{I} - e\mathbf{p}^\infty) \mathbf{w} \\ &= \frac{\sqrt{N}(2f^+f^-)}{r} \frac{d\mathbf{p}^\infty}{d\alpha} \mathbf{w}. \quad (62) \quad \text{eq:area}\end{aligned}$$

We can rewrite this using (22), with $A = \sqrt{N}(2f^+f^-)\hat{A}$ and $q_{ij} \equiv \frac{d\mathbf{W}_{ij}^F}{d\alpha} = \mathbf{W}_{ij}^+ - \mathbf{W}_{ij}^-$

$$\hat{A} = \sum_{i,j,k} \mathbf{p}_i^\infty q_{ij} (\bar{\mathbf{T}}_{ik} - \bar{\mathbf{T}}_{jk}) \mathbf{p}_k^\infty \mathbf{w}_k. \quad (63) \quad \text{eq:areaT}$$

²Note that including the dropped factors in (45) would only change this to $\text{SNR}(0) \leq \sqrt{N(4f^+f^-)}$, which would have no effect on what follows.

Definition 3: Partial mixing times

We define the \pm mixing times as

$$\begin{aligned}\eta_i^\pm &\equiv \sum_k \bar{\mathbf{T}}_{ik} \mathbf{p}_k^\infty \left(\frac{1 \pm \mathbf{w}_k}{2} \right) = \sum_{k \in \pm} \bar{\mathbf{T}}_{ik} \mathbf{p}_k^\infty \\ &= \sum_k (\mathbf{Z}_{kk} - \mathbf{Z}_{ik}) \left(\frac{1 \pm \mathbf{w}_k}{2} \right) = \sum_{k \in \pm} (\mathbf{Z}_{kk} - \mathbf{Z}_{ik}).\end{aligned}\tag{64} \quad \text{eq:mixingpm}$$

We can think of η_i^+ as a measure of the “distance” to the $\mathbf{w}_k = +1$ states and η_i^- as the “distance” to the $\mathbf{w}_k = -1$ states.

Using (19), we can write:

$$\begin{aligned}\eta_i^+ + \eta_i^- &= \eta, \\ 2(\eta_i^+ - \eta_j^+) &= \sum_k (\bar{\mathbf{T}}_{ik} - \bar{\mathbf{T}}_{jk}) \mathbf{p}_k^\infty \mathbf{w}_k = \sum_k (\mathbf{Z}_{jk} - \mathbf{Z}_{ik}) \mathbf{w}_k.\end{aligned}\tag{65} \quad \text{eq:mixingrel}$$

We could arrange the states in order of decreasing η^+ , which is the same as the order of increasing η^- .

We can rewrite (63) as

$$\begin{aligned}\hat{A} &= 2 \sum_{i,j} q_{ij} \mathbf{p}_i^\infty (\eta_i^+ - \eta_j^+) = -2 \sum_{i,j} q_{ij} \mathbf{p}_i^\infty \eta_j^+ \\ &= 2 \sum_{i,j} q_{ij} \mathbf{p}_i^\infty (\eta_j^- - \eta_i^-) = 2 \sum_{i,j} q_{ij} \mathbf{p}_i^\infty \eta_j^-.\end{aligned}\tag{66} \quad \text{eq:areaEta}$$

We can also express it in terms of the fundamental matrix (7) as

$$\hat{A} = \sum_{i,j,k,l} q_{ij} \boldsymbol{\pi}_l \mathbf{Z}_{li} (\mathbf{Z}_{jk} - \mathbf{Z}_{ik}) \mathbf{w}_k = \boldsymbol{\pi} \mathbf{Z} q \mathbf{Z} \mathbf{w}.\tag{67} \quad \text{eq:areaZ}$$

It is also helpful to define the following quantities:

$$\begin{aligned}c_k &= \frac{d \ln \mathbf{p}_k^\infty}{d\alpha} = \sum_{ij} \mathbf{p}_i^\infty q_{ij} (\bar{\mathbf{T}}_{ik} - \bar{\mathbf{T}}_{jk}) = -(\mathbf{p}^\infty q \bar{\mathbf{T}})_k = \frac{(\mathbf{p}^\infty q \mathbf{Z})_k}{\mathbf{p}_k^\infty}, \\ a_i &= \sum_j q_{ij} \mathbf{p}_i^\infty (\eta_i^+ - \eta_j^+), \\ \implies \hat{A} &= \sum_k c_k \mathbf{p}_k^\infty \mathbf{w}_k = 2 \sum_i a_i.\end{aligned}\tag{68} \quad \text{eq:areacoeff}$$

Note that the optimal choice of \mathbf{w} is $\mathbf{w}_k = \text{sgn}(c_k)$.

sec:deriv

3.2 Derivatives wrt. \mathbf{W}^\pm

As discussed in §2.3, we will regard the off-diagonal elements of \mathbf{W}_{ij}^\pm to be the independent variables, with $\mathbf{W}_{ii}^\pm = -\sum_{j \neq i} \mathbf{W}_{ij}^\pm$ imposed by hand. Thus,

$$\frac{\partial \mathbf{W}_{ij}^F}{\partial \mathbf{W}_{gh}^\pm} = f^\pm \delta_{gi} (\delta_{hj} - \delta_{ij}), \quad \frac{\partial q_{ij}}{\partial \mathbf{W}_{gh}^\pm} = \pm \delta_{gi} (\delta_{hj} - \delta_{ij}). \quad (69)$$

eq:basicderi

The implicit $g \neq h$ that comes with all derivatives is unnecessary, as the derivatives above vanish when $g = h$.

In particular, differentiating (7),

$$\frac{\partial \mathbf{Z}_{ij}}{\partial \mathbf{W}_{gh}^\pm} = r f^\pm \mathbf{Z}_{ig} (\mathbf{Z}_{hj} - \mathbf{Z}_{gj}). \quad (70)$$

eq:derivZ

We can then differentiate expression (67) to get

$$\begin{aligned} \frac{\partial \hat{A}}{\partial \mathbf{W}_{gh}^\pm} &= r f^\pm \sum_{ijkl} q_{ij} \mathbf{w}_k \pi_l [\mathbf{Z}_{lg} (\mathbf{Z}_{hi} - \mathbf{Z}_{gi}) (\mathbf{Z}_{jk} - \mathbf{Z}_{ik}) + \mathbf{Z}_{li} (\mathbf{Z}_{jg} - \mathbf{Z}_{ig}) (\mathbf{Z}_{hk} - \mathbf{Z}_{gk})] \\ &\quad \pm \sum_{kl} \mathbf{w}_k \pi_l \mathbf{Z}_{lg} (\mathbf{Z}_{hk} - \mathbf{Z}_{gk}) \\ &= 2r f^\pm \sum_{ij} q_{ij} \mathbf{p}_i^\infty \mathbf{p}_g^\infty [(\bar{\mathbf{T}}_{gi} - \bar{\mathbf{T}}_{hi}) (\eta_i^+ - \eta_j^+) + (\bar{\mathbf{T}}_{ig} - \bar{\mathbf{T}}_{jg}) (\eta_g^+ - \eta_h^+)] \\ &\quad \pm 2\mathbf{p}_g^\infty (\eta_g^+ - \eta_h^+) \\ &= 2r f^\pm \sum_{ij} q_{ij} \mathbf{p}_i^\infty \mathbf{p}_g^\infty (\bar{\mathbf{T}}_{gi} - \bar{\mathbf{T}}_{hi}) (\eta_i^+ - \eta_j^+) \pm 2\mathbf{p}_g^\infty (\eta_g^+ - \eta_h^+) \left[1 + \frac{\partial \ln(\mathbf{p}_g^\infty)}{\partial \ln(f^\pm)} \right] \\ &= 2r f^\pm \mathbf{p}_g^\infty \left[\sum_i a_i (\bar{\mathbf{T}}_{gi} - \bar{\mathbf{T}}_{hi}) + c_g (\eta_g^+ - \eta_h^+) \right] \pm 2\mathbf{p}_g^\infty (\eta_g^+ - \eta_h^+). \end{aligned} \quad (71)$$

eq:derivA

where a_i and c_k were defined in (68).

It is sometimes useful to consider the following derivatives:

$$\frac{\partial}{\partial \mathbf{W}_{gh}^F} \equiv \frac{\partial}{\partial \mathbf{W}_{gh}^+} + \frac{\partial}{\partial \mathbf{W}_{gh}^-}, \quad \frac{\partial}{\partial q_{gh}} \equiv f^- \frac{\partial}{\partial \mathbf{W}_{gh}^+} - f^+ \frac{\partial}{\partial \mathbf{W}_{gh}^-}. \quad (72)$$

eq:derivqw

Each of these derivatives behaves as their names suggest:

$$\frac{\partial \mathbf{W}_{ij}^F}{\partial \mathbf{W}_{gh}^F} = \frac{\partial q_{ij}}{\partial q_{gh}} = \delta_{gi} (\delta_{hj} - \delta_{ij}), \quad \frac{\partial q_{ij}}{\partial \mathbf{W}_{gh}^F} = \frac{\partial \mathbf{W}_{ij}^F}{\partial q_{gh}} = 0. \quad (73)$$

eq:derivqwef

This is because we could treat \mathbf{W}^F and q as the independent variables. However, the boundaries of the allowed region are more easily expressed in terms of \mathbf{W}^\pm .

ec:rescale

3.2.1 Scaling mode

Consider the following differential operator:

$$\Delta \equiv \sum_{g,h} \mathbf{W}_{gh}^+ \frac{\partial}{\partial \mathbf{W}_{gh}^+} + \mathbf{W}_{gh}^- \frac{\partial}{\partial \mathbf{W}_{gh}^-}. \quad (74) \quad \text{eq:scaleop}$$

This corresponds to the scaling, $\mathbf{W}^\pm \rightarrow (1+\epsilon)\mathbf{W}^\pm$. Intuitively, this has two effects: it scales up the initial potentiation/depression and it scales down all timescales. This intuition is confirmed by the following results:

$$\begin{aligned} \Delta \mathbf{Z} &= \tau \mathbf{e} \mathbf{p}^\infty - \mathbf{Z}, & \Delta \mathbf{p}^\infty &= 0, & \Delta \mathbf{T} &= -\mathbf{T}, \\ \Delta \eta_i^\pm &= -\eta_i^\pm, & \Delta q_{ij} &= q_{ij}, & \Delta \hat{A} &= 0, \end{aligned} \quad (75) \quad \text{eq:scaleeffe}$$

The anomalous bit in the scaling of \mathbf{Z} is due to the lack of dependence of $\boldsymbol{\pi}$ and τ on \mathbf{W}^\pm .

As the area is invariant under this scaling, we can consider the \mathbf{W}^\pm to be projective coordinates. Therefore we don't need to enforce the constraint (51) while looking for the maximum area, as we can use this null-mode to enforce it afterwards without changing the area. We also don't have to worry about the boundary at infinity, as sending some of them to infinity is equivalent to sending the rest to zero.

kuhntucker

3.3 Kuhn-Tucker conditions

Consider the Lagrangian

$$\mathcal{L} = \hat{A} + \sum_{\pm} \sum_{i \neq j} \mu_{ij}^\pm \mathbf{W}_{ij}^\pm + \lambda \mathbf{p}^\infty \mathbf{w}. \quad (76) \quad \text{eq:lagrangia}$$

The last term ensures that we hold \mathbf{p}_\pm^∞ fixed during this extremisation, so that we can safely ignore the factors that we dropped from the noise in (45). Necessary conditions for an extremum are

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_{gh}^\pm} = 0, \quad \mu_{gh}^\pm \geq 0, \quad \mathbf{W}_{gh}^\pm \geq 0, \quad \mu_{gh}^\pm \mathbf{W}_{gh}^\pm = 0. \quad (77) \quad \text{eq:extremum}$$

with $g \neq h$. This enforces the constraints (50), but not (51). As discussed in §2.3 and §3.2.1, that can be enforced after finding the maximum with the null scaling degree of freedom.

triangular

3.3.1 Triangularity

Consider

$$\frac{\partial \mathcal{L}}{\partial q_{gh}} = f^- \frac{\partial \mathcal{L}}{\partial \mathbf{W}_{gh}^+} - f^+ \frac{\partial \mathcal{L}}{\partial \mathbf{W}_{gh}^-} = (f^- \mu_{gh}^+ - f^+ \mu_{gh}^-) + 2\mathbf{p}_g^\infty (\eta_g^+ - \eta_h^+) = 0. \quad (78) \quad \text{eq:shiftqder}$$

This corresponds to the shift

$$\mathbf{W}_{ij}^+ \rightarrow \mathbf{W}_{ij}^+ + f^- \epsilon_{ij}, \quad \mathbf{W}_{ij}^- \rightarrow \mathbf{W}_{ij}^- - f^+ \epsilon_{ij}, \quad \sum_j \epsilon_{ij} = 0, \quad (79) \quad \text{eq:shiftq}$$

which leaves \mathbf{W}^F unchanged, and therefore \mathbf{p}^∞ , \mathbf{T} and η^\pm as well.

Assume $\eta_g^+ > \eta_h^+$. Then

$$f^- \mu_{gh}^+ - f^+ \mu_{gh}^- < 0 \quad \implies \quad \mu_{gh}^- > 0 \quad \implies \quad \mathbf{W}_{gh}^- = 0. \quad (80) \quad \text{eq:lowertria}$$

Similarly, if $\eta_g^+ < \eta_h^+$, then

$$f^- \mu_{gh}^+ - f^+ \mu_{gh}^- > 0 \quad \implies \quad \mu_{gh}^+ > 0 \quad \implies \quad \mathbf{W}_{gh}^+ = 0. \quad (81) \quad \text{eq:uppertria}$$

Thus, if we arrange the states in order of decreasing η^+ , \mathbf{W}^+ is upper-triangular and \mathbf{W}^- is lower triangular.

We have ignored the possibility that $\mathbf{p}_g^\infty = 0$, as this would imply that $\mathbf{T}_{ig} = \infty$, which would in turn imply that the Markov process is not ergodic.

3.3.2 Increasing c_k

Consider the following combinations of derivatives:

$$\Delta_{gh} \equiv \frac{1}{\mathbf{p}_g^\infty} \left(\frac{\partial}{\partial \mathbf{W}_{gh}^F} \right) + \frac{1}{\mathbf{p}_h^\infty} \left(\frac{\partial}{\partial \mathbf{W}_{hg}^F} \right), \quad (82) \quad \text{eq:areacoeff}$$

$$(83)$$

Note that they are only well defined if all the states have non-zero equilibrium probabilities (see the comment in §3.3.1 about this being satisfied for ergodic chains).

One can show that the equilibrium probabilities, \mathbf{p}^∞ , are invariant under these operators:

$$\Delta_{gh} \mathbf{p}_i^\infty = 0, \quad (84) \quad \text{eq:sareacoeff}$$

which makes it possible to integrate the perturbation:

$$\mathbf{W}^\pm \rightarrow \mathbf{W}^\pm + \mathbf{D}\boldsymbol{\epsilon}, \quad \begin{aligned} \boldsymbol{\epsilon} &= \boldsymbol{\epsilon}^T, \\ \boldsymbol{\epsilon}\boldsymbol{\epsilon} &= 0. \end{aligned} \quad (85) \quad \text{eq:areacoeff}$$

But more interestingly:

$$\Delta_{gh} \mathcal{L} = \frac{\mu_{gh}^+ + \mu_{gh}^-}{\mathbf{p}_g^\infty} + \frac{\mu_{hg}^+ + \mu_{hg}^-}{\mathbf{p}_h^\infty} + 2r(c_g - c_h)(\eta_g^+ - \eta_h^+), \quad (86) \quad \text{eq:areacoeff}$$

$$(87)$$

where c_k were defined in (68).

Using the non-negativity of the Kuhn-Tucker multipliers, μ_{ij}^\pm , (86) tells us that if we arrange the states in order of decreasing η_i^+ , the optimal process will have non-decreasing c_k (if any of the η_k^+ are degenerate, we can choose their order to ensure this).

Note that, according to §3.3.1, either \mathbf{W}_{gh}^+ or \mathbf{W}_{gh}^- will be zero at the maximum, therefore we can expect one of $\mu_{gh}^+ + \mu_{gh}^-$ to be non-zero. This would rule out degeneracy of the c_k or η_k^+ . Looking at (78) closely, the only way $\mu_{gh}^+ + \mu_{gh}^-$ could be zero is if $\eta_g^+ = \eta_h^+$ or $\mathbf{p}_g^\infty = 0$.

3.3.3 Shortcuts

Now consider the following combinations of derivatives for $m > 1$:

$$\tilde{\Delta}_{g,m}^\pm \equiv \left[\sum_{k=0}^{m-1} \frac{1}{\mathbf{p}_{g\pm k}^\infty} \left(\frac{\partial}{\partial \mathbf{W}_{g\pm k, g\pm(k+1)}^\pm} \right) \right] - \frac{1}{\mathbf{p}_g^\infty} \left(\frac{\partial}{\partial \mathbf{W}_{g, g\pm m}^\pm} \right). \quad (88)$$

Once again, they are only well defined if all the states have non-zero equilibrium probabilities (see the comment in §3.3.1 about this being satisfied for ergodic chains).

One can show that the equilibrium probabilities, \mathbf{p}^∞ , are invariant under these operators:

$$\tilde{\Delta}_{g,m}^\pm \mathbf{p}_i^\infty = 0, \quad (89)$$

which makes it possible to integrate the perturbation:

$$\begin{aligned} \mathbf{W}^\pm \rightarrow \mathbf{W}^\pm + \mathbf{D}\boldsymbol{\epsilon}^{\pm(g,m)}, \quad & \begin{aligned} (\boldsymbol{\epsilon}^{\pm(g,m)})_{g, g\pm m} &= -\epsilon, \\ (\boldsymbol{\epsilon}^{\pm(g,m)})_{g\pm k, g\pm(k+1)} &= \epsilon \quad \forall k \in [0, m-1], \\ (\boldsymbol{\epsilon}^{\pm(g,m)})_{g\pm k, g\pm k} &= -\epsilon \quad \forall k \in [1, m-1]. \end{aligned} \end{aligned} \quad (90)$$

But more interestingly for our purposes:

$$\tilde{\Delta}_{g,m}^\pm \mathcal{L} = \left[\sum_{k=0}^{m-1} \frac{\mu_{g\pm k, g\pm(k+1)}^\pm}{\mathbf{p}_{g\pm k}^\infty} - \frac{\mu_{g, g\pm m}^\pm}{\mathbf{p}_g^\infty} \right] + 2r f^\pm \sum_{k=0}^{m-1} \left(\eta_{g\pm k}^+ - \eta_{g\pm(k+1)}^+ \right) (c_{g\pm k} - c_g), \quad (91)$$

If we put the states in order of decreasing η_k^+ , the results of the §3.3.2 tell us that the c_k are non-decreasing. This implies that the last term of the final expression in (91) is non-negative. If it is non-zero (there would need to be a lot of degeneracy for it to be zero), this would imply that $\mu_{g, g\pm m}^\pm > 0$, which in turn implies that $\mathbf{W}_{g, g\pm m}^\pm = 0$. This would tell us that the process with the maximal area has to have a multi-state topology.

3.3.4 Summary

Using the Kuhn-Tucker formalism, we have shown that, with the states arranged in order of non-increasing η_i^+ :

- There can be no ergodic maximum for which \mathbf{W}^+ contains backwards transitions or \mathbf{W}^- contains forwards transitions.
- There can be no ergodic maximum with the c_k decreasing.
- The c_k may only be degenerate at an ergodic maximum if the corresponding η_k^+ are also degenerate.
- If the c_k increase and the η_i^+ decrease, there can be no ergodic maximum with shortcuts.

These were shown by finding allowed perturbations that increase the area.

This leaves two possibilities for the maximum area Markov chain. Either there is no degeneracy and no shortcuts, which implies the Multi-state topology that we'll discuss in §3.4, or there is some degeneracy, which would allow shortcuts provided that they do not bypass an entire degenerate set (see (91)).

Degeneracy tends to be very delicate. It is usually hard to arrange without some symmetry relating degenerate states. Such a symmetry would imply lumpability (see §1.7). The lumped chain would not have any shortcuts, as an entire degenerate set cannot be bypassed. As this lumped chain has the same area (see §2.4), we would need only consider the multi-state topology.

3.4 Multi-state topology

The multi-state topology is defined by (see [7, 8]):

$$\mathbf{W}_{ij}^+ = q_i^+ \delta_{i+1,j}, \quad \mathbf{W}_{ij}^- = q_j^- \delta_{i,j+1}. \quad (92)$$

It saturates various inequalities:

$$\begin{aligned} \bar{\mathbf{T}}_{ik} - \bar{\mathbf{T}}_{jk} &= \begin{cases} \bar{\mathbf{T}}_{ij}, & \text{if } i \leq j \leq k \text{ or } i \geq j \geq k, \\ -\bar{\mathbf{T}}_{ji}, & \text{if } j \leq i \leq k \text{ or } j \geq i \geq k, \end{cases} \\ r\mathbf{p}_i^\infty \mathbf{W}_{ij}^F (\bar{\mathbf{T}}_{ij} + \bar{\mathbf{T}}_{ji}) &= 1 \quad \text{if } i = j \pm 1, \end{aligned} \quad (93)$$

and it satisfies detailed balance (a.k.a. reversibility a.k.a. $\mathcal{L}_{\mathbf{p}^\infty}^2$ self-adjointness):

$$f^+ q_i^+ \mathbf{p}_i^\infty = f^- q_i^- \mathbf{p}_{i+1}^\infty, \quad (94)$$

which means we can always choose the transition rates, q_i^\pm , to give any desired equilibrium probabilities, \mathbf{p}_i^∞ .

This allows us to calculate the c_k 's:

$$\begin{aligned}
c_k &= \sum_{i < k} \mathbf{T}_{i,i+1} (\mathbf{p}_i^\infty q_i^+ + \mathbf{p}_{i+1}^\infty q_i^-) - \sum_{i \geq k} \mathbf{T}_{i+1,i} (\mathbf{p}_i^\infty q_i^+ + \mathbf{p}_{i+1}^\infty q_i^-), \\
c_{k+1} - c_k &= (\mathbf{T}_{k,k+1} + \mathbf{T}_{k+1,k}) \left(\frac{\mathbf{p}_k^\infty \mathbf{W}_{k,k+1}^F}{f^+} + \frac{\mathbf{p}_{k+1}^\infty \mathbf{W}_{k+1,k}^F}{f^-} \right) = \frac{1}{r f^+ f^-}, \\
\sum_k c_k \mathbf{p}_k^\infty &= \sum_{ij} \mathbf{p}_i^\infty q_{ij} (\eta - \eta) = 0, \\
\Rightarrow c_k &= \frac{k - \sum_j j \mathbf{p}_j^\infty}{r f^+ f^-},
\end{aligned} \tag{95} \quad \text{eq:areacoeff}$$

where we used (93) to derive the first two equations respectively and Th.6 to derive the third. This allows us to write the area as

$$A = \frac{2\sqrt{N}}{r} \sum_k \left[k - \sum_j j \mathbf{p}_j^\infty \right] \mathbf{p}_k^\infty \mathbf{w}_k = \frac{2\sqrt{N}}{r} \sum_k \left| k - \sum_j j \mathbf{p}_j^\infty \right| \mathbf{p}_k^\infty, \tag{96} \quad \text{eq:multistat}$$

where we used $\mathbf{w}_k = \text{sgn}(c_k)$, as discussed after (68).

First let us maximise (96) at fixed $\mathbf{p}_\pm^\infty = \sum_k \mathbf{p}_k^\infty \left(\frac{1 \pm \mathbf{w}_k}{2} \right)$. Clearly this will happen when we put all of the probability at the ends: $\mathbf{p}_1^\infty = \mathbf{p}_-^\infty$ and $\mathbf{p}_n^\infty = \mathbf{p}_+^\infty$ are the only non-zero \mathbf{p}_k^∞ . This gives an area of ³

$$A \leq \frac{\sqrt{N}}{r} (n-1) (4\mathbf{p}_+^\infty \mathbf{p}_-^\infty). \tag{97} \quad \text{eq:multistat}$$

This is maximised at $\mathbf{p}_+^\infty = \mathbf{p}_-^\infty = \frac{1}{2}$:

$$A \leq \frac{\sqrt{N}}{r} (n-1). \tag{98} \quad \text{eq:maxarea}$$

Note that the chain that achieves this is not ergodic, the two states at each end are absorbing. This is similar to the results found numerically in [9] in a slightly different situation.

A Other quantities

Here we collect useless facts.

³Note that including the dropped factors in (45) would only change this to $A \leq \sqrt{N(4\mathbf{p}_+^\infty \mathbf{p}_-^\infty)}(n-1)/r$, which would have no effect on what follows.

A.1 Differences in c_k 's

In the general case, making full use of [1, Cor.6.2.7], we can show that:

$$\begin{aligned}
 c_{k+1} - c_k &= \frac{N_{kk}^{k+1}}{\mathbf{p}_k^\infty} \left[\sum_{i < j} \mathbf{p}_i^\infty \frac{\mathbf{W}_{ij}^F}{f^+} (H_{i,k}^{k+1} + H_{j,k+1}^k - 1) + \sum_{i > j} \mathbf{p}_i^\infty \frac{\mathbf{W}_{ij}^F}{f^+} (H_{j,k}^{k+1} + H_{i,k+1}^k - 1) \right] \\
 &= \frac{N_{kk}^{k+1}}{\mathbf{p}_k^\infty} \left[\sum_{i < j} \mathbf{p}_i^\infty \frac{\mathbf{W}_{ij}^F}{f^+} (1 - H_{j,k}^{k+1} + H_{i,k+1}^k) + \sum_{i > j} \mathbf{p}_i^\infty \frac{\mathbf{W}_{ij}^F}{f^+} (1 - H_{i,k}^{k+1} + H_{j,k+1}^k) \right], \\
 \eta_k^+ - \eta_{k+1}^+ &= \mathbf{T}_{k,k+1} \mathbf{p}_+^\infty - \sum_{\{i | \mathbf{w}_i > 0\}} N_{ii}^{k+1} (1 - H_{k,k+1}^i) \\
 &= \sum_{\{i | \mathbf{w}_i > 0\}} N_{ii}^k (1 - H_{k+1,k}^i) - \mathbf{T}_{k+1,k} \mathbf{p}_+^\infty \\
 = \eta_{k+1}^- - \eta_k^- &= \mathbf{T}_{k+1,k} \mathbf{p}_-^\infty - \sum_{\{i | \mathbf{w}_i < 0\}} N_{ii}^k (1 - H_{k+1,k}^i) \\
 &= \sum_{\{i | \mathbf{w}_i < 0\}} N_{ii}^{k+1} (1 - H_{k,k+1}^i) - \mathbf{T}_{k,k+1} \mathbf{p}_-^\infty.
 \end{aligned} \tag{99}$$

eq:areacoeff

We need to show that the positivity of one ($\forall k$) implies that the positivity of the other.

A.2 Derivatives wrt. $\bar{\mathbf{T}}$

The Markov chain, \mathbf{W}^F , is completely determined by the off-diagonal mean first passage times, $\bar{\mathbf{T}}$, [1, Th.4.4.12]:

$$\begin{aligned}
 \mathbf{p}^\infty &= \eta \left(\bar{\mathbf{T}}^{-1} \mathbf{e} \right)^\top, \\
 r \mathbf{W}^F &= \mathbf{Q} = (\mathbf{D} - \mathbf{E}) \bar{\mathbf{T}}^{-1}.
 \end{aligned} \tag{100}$$

eq:fptdet

where η can be determined by normalising \mathbf{p}^∞ and \mathbf{D} is determined by \mathbf{p}^∞ , (6). Clearly q will be undetermined.

We can then define derivatives wrt. $\bar{\mathbf{T}}$:

$$\frac{\partial}{\partial \bar{\mathbf{T}}_{gh}} = \sum_{ij} \left(\frac{\partial \mathbf{W}_{ij}^F}{\partial \bar{\mathbf{T}}_{gh}} \right) \frac{\partial}{\partial \mathbf{W}_{ij}^F}, \tag{101}$$

eq:derivTdef

where

$$\frac{\partial \mathbf{W}_{ij}^F}{\partial \bar{\mathbf{T}}_{gh}} = r \mathbf{p}_h^\infty \mathbf{W}_{ig}^F (\mathbf{W}_{ij}^F - \mathbf{W}_{hj}^F), \quad \frac{\partial \mathbf{p}_i^\infty}{\partial \bar{\mathbf{T}}_{gh}} = -r \mathbf{p}_h^\infty \mathbf{p}_i^\infty \mathbf{W}_{ig}^F. \tag{102}$$

eq:derivTWp

These derivatives of q will vanish.

This means that the derivatives of the area are:

$$\frac{\partial \hat{A}}{\partial \bar{\mathbf{T}}_{gh}} = -\mathbf{p}_h^\infty \left[\sum_i \mathbf{p}_i^\infty q_{ig} \mathbf{w}_h + r \sum_{ijk} (\mathbf{W}_{ig}^F + \mathbf{W}_{kg}^F) \mathbf{p}_i^\infty q_{ij} (\bar{\mathbf{T}}_{ik} - \bar{\mathbf{T}}_{jk}) \mathbf{p}_k^\infty \mathbf{w}_k \right]. \tag{103}$$

derivTarea

sec:initSNR

A.3 Initial SNR

Let us write $\text{SNR}(0) = \sqrt{N}(2f^+f^-)\hat{I}$. Using (46), we can write

$$\begin{aligned}\hat{I} &= \frac{d\mathbf{p}^\infty}{d\alpha}(-\mathbf{W}^F)\mathbf{w} &= -\pi\mathbf{Z}\frac{d\mathbf{Q}}{d\alpha}\mathbf{Z}\mathbf{W}^F\mathbf{w} &= -\pi\mathbf{Z}\frac{d\mathbf{W}^F}{d\alpha}\mathbf{Z}\mathbf{Q}\mathbf{w} \\ &= \pi\mathbf{Z}\frac{d\mathbf{W}^F}{d\alpha}(\mathbf{I} - \tau\mathbf{e}\pi)\mathbf{w} &= \pi\mathbf{Z}\frac{d\mathbf{W}^F}{d\alpha}\mathbf{w} &= \mathbf{p}^\infty\frac{d\mathbf{W}^F}{d\alpha}\mathbf{w}.\end{aligned}\tag{104}$$

eq:initSNR

We can also write this in component form

$$\begin{aligned}\hat{I} &= \sum_{i,j} \mathbf{p}_i^\infty q_{ij} \mathbf{w}_j \\ &= \sum_{i \neq j} \mathbf{p}_i^\infty q_{ij} \mathbf{w}_j + \sum_i \mathbf{p}_i^\infty q_{ii} \mathbf{w}_i \\ &= \sum_{i \neq j} q_{ij} \mathbf{p}_i^\infty (\mathbf{w}_j - \mathbf{w}_i).\end{aligned}\tag{105}$$

eq:initSNRcm

We can differentiate the expression in terms of \mathbf{Z} to get

$$\begin{aligned}\frac{\partial \hat{I}}{\partial \mathbf{W}_{gh}^\pm} &= rf^\pm \sum_{ijk} \pi_k \mathbf{Z}_{kg} (\mathbf{Z}_{hi} - \mathbf{Z}_{gi}) q_{ij} \mathbf{w}_j \pm \sum_k \pi_k \mathbf{Z}_{kg} (\mathbf{w}_h - \mathbf{w}_g) \\ &= rf^\pm \sum_{ij} \mathbf{p}_g^\infty q_{ij} (\bar{\mathbf{T}}_{gi} - \bar{\mathbf{T}}_{hi}) \mathbf{p}_i^\infty \mathbf{w}_j \pm \mathbf{p}_g^\infty (\mathbf{w}_h - \mathbf{w}_g).\end{aligned}\tag{106}$$

eq:initSNRdif

If we wished to maximise the area with fixed initial SNR, $\hat{I} = I_0$, we would add $\beta(\hat{I} - I_0)$ to the Lagrangian (76) and extremise wrt β .

A.4 Fixing mixing time (Kemeny's constant)

sec:fixmix

The mixing time, η (19), can be thought of as the mean time to pass between two states drawn from the equilibrium distribution [3, 10]. It should be a measure of the time it takes to reach equilibrium, and therefore related to the time it takes the SNR to decay. It could be interesting to hold it fixed while maximising the area, as this could help prevent degenerate solutions with very low initial SNR but long decay time. This can be done by adding $\gamma(\eta - \eta_0)$ to the Lagrangian (76) and extremising wrt γ .

It would help to calculate its derivatives.

$$\begin{aligned}\frac{\partial \eta}{\partial \mathbf{W}_{gh}^\pm} &= \frac{\partial}{\partial \mathbf{W}_{gh}^\pm} \left[\sum_i \mathbf{Z}_{ii} - \tau \right] \\ &= rf^\pm \sum_i \mathbf{Z}_{ig} (\mathbf{Z}_{hi} - \mathbf{Z}_{gi}) \\ &= rf^\pm (\mathbf{Z}_{hg}^2 - \mathbf{Z}_{gg}^2).\end{aligned}\tag{107}$$

eq:mixingdif

A.5 Scaling mode

Consider the following differential operator:

$$\Delta \equiv \sum_{g,h} \mathbf{W}_{gh}^+ \frac{\partial}{\partial \mathbf{W}_{gh}^+} + \mathbf{W}_{gh}^- \frac{\partial}{\partial \mathbf{W}_{gh}^-}. \quad (108) \quad \text{eq:scaleopap}$$

This corresponds to the scaling, $\mathbf{W}^\pm \rightarrow (1+\epsilon)\mathbf{W}^\pm$. Intuitively, this has two effects: it scales up the initial potentiation/depression and it scales down all timescales. This intuition is confirmed by the following results:

$$\begin{aligned} \Delta \mathbf{Z} &= \tau \mathbf{e} \mathbf{p}^\infty - \mathbf{Z}, \quad \Delta \mathbf{p}^\infty = 0, \quad \Delta \mathbf{T} = -\mathbf{T}, \quad \Delta \eta_i^\pm = -\eta_i^\pm, \quad \Delta q_{ij} = q_{ij}, \\ \Delta \hat{A} &= 0, \quad \Delta \hat{I} = \hat{I}, \quad \Delta \eta = -\eta. \end{aligned} \quad (109) \quad \text{eq:scaleeffe}$$

The anomalous bit in the scaling of \mathbf{Z} is due to the lack of dependence of $\boldsymbol{\pi}$ and τ on \mathbf{W}^\pm .

Suppose we don't hold anything fixed when maximising the area. Then we'd have

$$\Delta \mathcal{L} = 0,$$

i.e. this scaling is a null direction. Area maximisation can only determine the ratios of transition rates, not their absolute values.

Now suppose we only fix \hat{I} when maximising the area. Then we'd have

$$\Delta \mathcal{L} = \beta \hat{I} = \beta I_0 = 0.$$

As the Lagrange multiplier is zero, the constraint doesn't change the derivatives of the Lagrangian wrt \mathbf{W}^\pm . The sole effect of the constraint would be to fix the scaling mode. It wouldn't change the ratios of transition rates, topology, etc.

Exactly the same conclusions would follow if we only fixed η when maximising the area. This would also only fix the scaling mode. However, if we held them both fixed, it would have more of an effect:

$$\Delta \mathcal{L} = \beta \hat{I} - \gamma \eta = \beta I_0 - \gamma \eta_0 = 0. \quad (110) \quad \text{eq:scalecons}$$

The Lagrange multipliers are non-zero, so this could affect the topology of the optimal Markov chain.

References

- [1] J. Kemeny and J. Snell, *Finite markov chains*. Springer, 1960.

- [2] J. Hunter, "A survey of generalized inverses and their use in stochastic modelling," *Research Letters in the Information and Mathematical Sciences* **1** (2000) no. 1, 25–36. [link].

- [3] J. Hunter, "Mixing times with applications to perturbed Markov chains," *Linear algebra and its applications* **417** (2006) no. 1, 108–123.

2000markov	[4] G. Cho and C. Meyer, “Markov chain sensitivity measured by mean first passage times,” <i>Linear Algebra and its Applications</i> 316 (2000) no. 1-3, 21–28. [link].
8markovian	[5] C. Burke and M. Rosenblatt, “A Markovian function of a Markov chain,” <i>The Annals of Mathematical Statistics</i> 29 (1958) no. 4, 1112–1122.
umpability	[6] F. Ball and G. F. Yeo, “Lumpability and Marginalisability for Continuous-Time Markov Chains,” <i>Journal of Applied Probability</i> 30 (1993) no. 3, 518–528.
94learning	[7] D. J. Amit and S. Fusi, “Learning in neural networks with material synapses,” <i>Neural Computation</i> 6 (1994) no. 5, 957–982.
multistate	[8] S. Fusi and L. F. Abbott, “Limits on the memory storage capacity of bounded synapses,” <i>Nat. Neurosci.</i> 10 (Apr, 2007) 485–493, PubMed:17351638.
08discrete	[9] A. B. Barrett and M. C. van Rossum, “Optimal learning rules for discrete synapses,” <i>PLoS Comput. Biol.</i> 4 (Nov, 2008) e1000230, PubMed:19043540.
2002kemeny	[10] M. Levene and G. Loizou, “Kemeny’s constant and the random surfer,” <i>The American mathematical monthly</i> 109 (2002) no. 8, 741–745.