Why doesn't the brain explode?

by

Tom Renner – Queens' College

Fourth-year undergraduate project

in Group F, 2013/2014

I hereby declare that, except where specifically indicated, the work submitted herein is my own original work.

Signed: Date:

Acknowledgements

I wish to thank Máté Lengyel for his clear explanations and support throughout the year, and Guillaume Hennnequin for his assistance with the stabilisation procedure, mathematics, and programming carried out.

Why doesn't the brain explode?

Tom Renner (Queens') — Fourth Year undergraduate project

Technical abstract

Neural network dynamics are difficult to model, due to the requirement for asymptotic stability coupled with the observation of high transient amplification of signals. In order to produce these high amplitude transients there must be strong synapses between excitatory neurons. However, a purely excitatory network is unstable, with positive feedback resulting in runaway amplification of signals. Neural networks therefore require a network of inhibitory neurons, which dampen activity, to stabilise the excitatory network dynamics.

It is not possible to analytically create inhibitory connections that stabilise a complex excitatory network. By using a linear approximation to neuronal dynamics, however, we can formulate the problem as a linear algebra matrix optimisation problem, with network operation characterised by the synaptic weight matrix, $\mathbf{W}$. This linear assumption, which is a poor approximation of measured neuronal dynamics, allows the stabilisation of the network using an iterative gradient-based method to force all of the eigenvalues of $\mathbf{W}$ to be below zero.

The largest real part of a matrix's eigenspectrum is called its spectral abscissa. A linear system will therefore be asymptotically stable if its spectral abscissa is below zero. Unfortunately, attempting to reduce the spectral abscissa of $\mathbf{W}$ directly is a non-smooth function of the elements of $\mathbf{W}$, so we use an upper bound on the spectral abscissa – the smoothed spectral abscissa. As its name suggests, this new metric is a smooth function, and therefore can be used in an iterative gradient-based method. It is calculated from the solutions to a primal-dual Lyapunov equation pair, which can be computed using specialised control libraries.

The stabilisation was carried out by optimising the inhibitory synapses of a network. By initialising the excitatory synapses to reflect biological observations and tuning the inhibitory network to stabilise the system dynamics we aimed to uncover statistics and patterns of the inhibitory connections.

We began by enforcing Dale's Law (the synapses from a neuron must be either all excitatory or all inhibitory) and a synaptic sparsity constraint. The effects of allowing synapses to be pruned from and added to the synaptic weight matrix were investigated and it was found that allowing this 'stabilisation plasticity' increased the correlation between inhibitory networks stabilising a given excitatory network. This was as expected, as allowing synapses to decay if they are not assisting the stabilisation, while keeping those important for stability, should result in some similarity between inhibitory networks. It

was also observed that all stabilising networks had a similar distribution of inhibitory synaptic weights; the majority of synapses having low strength, with a few with much greater weights.

Neural networks typically have more excitatory neurons than inhibitory. It was found that the Inhibitory→Excitatory synapses were significantly stronger than the Inhibitory→Inhibitory connections, which suggested that the reason for the imbalance in number of neurons is because little Inhibitory→Inhibitory inhibition is needed.

The effect of different strength excitatory networks was then investigated. It was found that the inhibitory network stabilised the excitatory network less well as the excitatory synapse strengths were increased. This was despite the average inhibitory strength increasing at a faster rate than the excitatory strength, indicating that it is the precise wiring of the inhibitory network, not its strength, that is important for stabilisation.

Higher than random synaptic reciprocity is defined as the increased probability of neuron j having an efferent connection to neuron i if neuron i has an efferent connection to neuron j. This effect has been observed in biological cortical networks, and was added to our model of the excitatory network. This initialisation scheme was found to produce an elliptical distribution of eigenvalues aligned along the real axis. This elliptical distribution of eigenvalues increases the spectral abscissa for a given strength of excitatory network, and hence these partially symmetric networks were harder to stabilise than those with random connectivity patterns.

Greater symmetry in the initial matrix was modelled, and it was noted that this increased the eccentricity of the elliptical eigenvalue distribution. The increased positive feedback caused by more reciprocal connections makes the system harder to stabilise, resulting in more oscillatory dynamics and a greater spectral abscissa of the stabilised network. However, despite being harder to stabilise, no class correlation is observed between the stabilising inhibitory networks. The variance was used to quantify the spread of the inhibitory synaptic weight values, and it was found that the variance is high, indicating a well tuned network, up to a high degree of symmetry.

Contents

1 Introduction

Investigating neuronal dynamics is difficult to do; the sheer complexity of the brain and cortex function makes detailed modelling of large networks impossible. However, with some simplifications to our neuronal and network models we can draw conclusions about network operation and structure. The aim of this project is to attempt to model the free dynamics of a neural network by applying control theory to stabilise a general network with structural similarities to a cortical network, to draw conclusions based on the resulting, stabilised structure.

Neurons are able to produce an electrical potential, called a spike. The frequency with which a neuron produces spikes is characterised by the firing rate. There is a base level firing rate at which neurons fire randomly, which is the rate of spike production for a neuron with no stimuli. Stimulating a neuron will cause a temporary elevation in the firing rate, during which period the neuron is said to be active. Neurons in the brain are divided into two categories – excitatory and inhibitory. When an excitatory neuron is active it will stimulate all the neurons to which it is connected to elevate their firing rates. This can cause a cascade of activity through the network, and will result in unstable behaviour if not carefully regulated. The regulatory force is provided by the network of inhibitory neurons. These have the opposite effect to excitatory neurons; when an inhibitory neuron is active it suppresses activity in neurons to which it is connected.

There is clearly a very delicate balance between the excitatory and inhibitory networks in brain cortices: too much inhibition and brain activity will be damped out too soon; too little inhibition and uncontrolled positive feedback can occur, leading to (theoretically, at least) the brain exploding!

A characteristic of neural network dynamics is is high transient amplification of signals, with asymptotic stability once these transients have died down. The approach used in this project is to initialise excitatory networks with different known structures, and then stabilise that network with an appropriately tuned inhibitory network. This can be achieved by using a metric called the smoothed spectral abscissa, which forms an upper bound on the largest real part of a matrix's eigenspectrum. This method, which stabilises general matrices, can be used to generate stable network models if a linear neuronal model is used.

The assumption of linearity is a strong one, as linearity is a poor estimation of measured neuronal dynamics. However, both this approximation and the stabilisation method used have been shown to produce dynamics that are a good model of measured cortical network dynamics[3]. It should be noted that the stabilisation method is not biologically plausible, but is useful for its ability to produce networks which have similar operation to neural measurements.

This project will use the smoothed spectral abscissa stabilisation method to stabilise

networks initialised with a variety of different synaptic structures. By examining the resulting stable network structures we aim to observe some trends that can be used to infer conclusions about the structure of cortical networks.

2 Mathematical theory

2.1 Neuronal dynamics

The firing rate of a neuron can be expressed in a general form as:

$$\tau_r \frac{dr_i}{dt} = -r_i(t) + \mathbf{F}\left(h_i + \sum_j w_{ij} r_j(t)\right) \tag{1}$$

where τ_r = time constant of the differential equation, $r_i(t)$ = firing rate of neuron i as a function of time, w_{ij} = strength of the efferent synapse from neuron i to neuron j, h_i = the input from sources external to our modelled network, and $\mathbf{F}(.)$ = our model of the neuron's response to stimuli, called the activation function. This gives the vectorised general dynamical model for a neuronal network:

$$\tau_r \frac{d\mathbf{r}}{dt} = -\mathbf{r}(t) + \mathbf{F}\left(\mathbf{h} + \mathbf{W}\mathbf{r}(t)\right) \tag{2}$$

It has been shown that using a linear model produces network dynamics which closely match those measured from experimental motor cortex neuron activity measurements[3], despite this being a poor approximation of the true activation function (Figure 1). This significantly reduces the computational complexity of the problem, and reduces the dynamics of the system to the linear case:

$$\tau_r \frac{d\mathbf{r}}{dt} = -\mathbf{r}(t) + \mathbf{h} + \mathbf{W}\mathbf{r}(t) \tag{3}$$

Recent experimental studies have examined movement generation in motor and premotor cortical areas[1]. A mechanism similar to a spring-loaded box has been suggested for

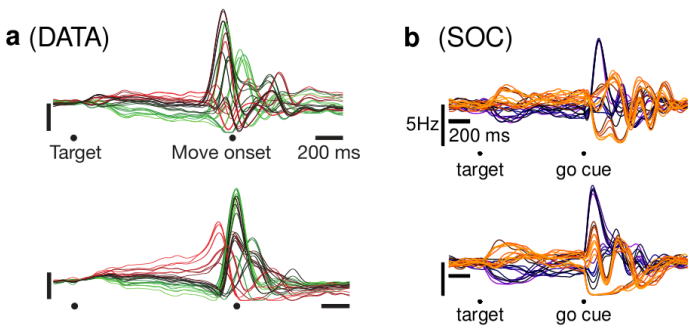


Figure 1: a) Activity traces from two sample neurons for the preparation and movement phase of 27 different actions.
b) Firing rates predicted by a Stability Optimised Circuit (SOC) constructed using smoothed spectral abscissa methods for two modelled neurons, initialised to 27 different preparatory states.
(Figure adapted from Hennequin *et al.* (2013).)

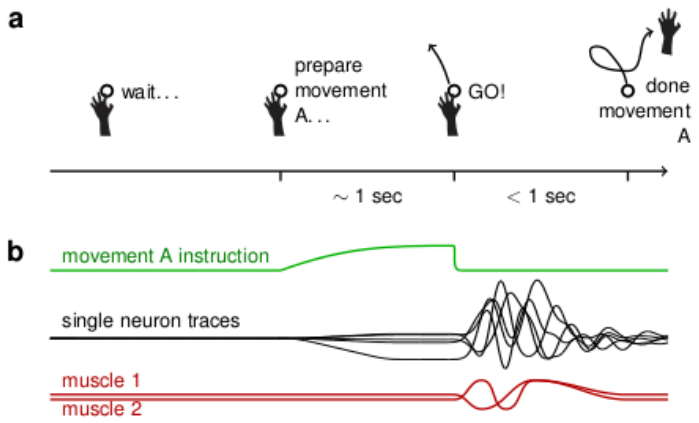


Figure 2: The spring-box model of cortical motor dynamics. Preparatory signals are built up to a steady state, and then free network dynamics is observed after the 'go' cue. (Figure adapted from Hennequin *et al.* (2013).)

the dynamics of these areas, where the system is driven into a specific state by preparatory stimuli, and then the network dynamics upon release of these stimuli orchestrate a sequence of motor commands resulting in the desired action. This is the model of network action used throughout the project, an illustration of which is shown in Figure 2.

These network dynamics exhibit specific, strong characteristics which we wish to model. Stability of the network is clearly a necessity, but large transient amplification of signals is also observed. Randomly connected, globally balanced networks are stable, but whether an integrate-and-fire or rate based model is used, these fail to capture the free dynamical transient amplification seen in experimental results[2]. A stronger coupled network can exhibit the complex dynamics desired, however, this is at the expense of inherently chaotic dynamics. These stronger coupled networks are therefore also not a suitable neural network model, as the chaotic behaviour will have high sensitivity to noise. This is clearly undesirable in the 'spring-box' dynamical system, as the system response is determined by the initial input conditions.

2.2 Defining The Smoothed Spectral Abscissa

Our goal is to be able to produce networks with strong excitatory connections but overall stability. Networks of this form have free dynamical responses that produce the high transient amplification we are attempting to model. However, it is not possible to directly create network structures that achieve this while respecting biological constraints. We therefore use an optimisation scheme that iteratively updates matrix values, subject to constraints, to force an unstable network into stability.

From Equation 3 it is clear that the matrix $\mathbf{W}$ dictates the dynamical response of the system to input, and the matrix $\mathbf{A} = \mathbf{W} - \mathbf{I}$ dictates the stability. We begin with a reference network, with some (known) initial structure of excitatory and inhibitory synapses, initialised with constant values for all synapses of the same type (initialisation discussed further in Section 3). This initial network is not generally stable. For the linear dynamical system we are assuming, the system will be stable if the largest real part of

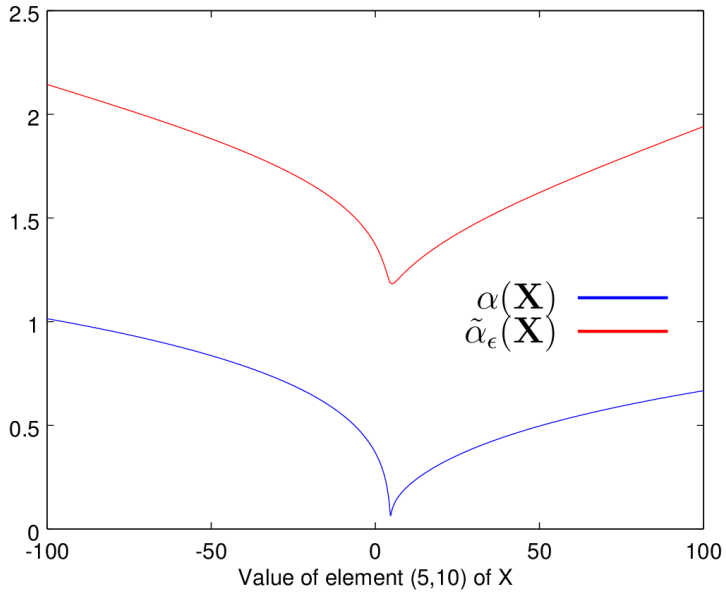


Figure 3: Variation of $\alpha(\mathbf{X})$ and $\tilde{\alpha}_\epsilon(\mathbf{X})$ with the value of an element of the example matrix, $\mathbf{X}$. The non-linearity in $\alpha(\mathbf{X})$ has been smoothed out such that $\tilde{\alpha}_\epsilon(\mathbf{X})$ is a smooth function of the element value.

the eigenspectrum of $\mathbf{A}$ is less than zero. This 'largest real part of the eigenspectrum' is defined as the spectral abscissa of $\mathbf{A}$, and is denoted $\alpha(\mathbf{A})$.

It would seem reasonable to attempt to stabilise our reference matrices by reducing the spectral abscissa of $\mathbf{A}$ to be less than zero (or, equivalently, the spectral abscissa of $\mathbf{W}$ below one). This approach would suggest optimising the network with respect to the spectral abscissa, and hence stabilising the dynamics. However, the problem of optimising $\alpha(\mathbf{A})$ with respect to $\mathbf{A}$ is a non-smooth function, and therefore not suitable for use with general gradient-based methods. We therefore use an upper bound on $\alpha(\mathbf{A})$, which is a smooth function of $\mathbf{A}$, such that gradient-based methods can be used. The upper bound used is called the smoothed spectral abscissa, $\tilde{\alpha}_\epsilon(\mathbf{A})$. Figure 3 shows the smoothing properties of $\tilde{\alpha}_\epsilon(\mathbf{A})$.

The smoothed spectral abscissa is defined as the solution of[7]:

$$f(\mathbf{A}, s) = \|\mathbf{H}_s\|^2_{\mathcal{H}_2}, \tag{4}$$

$$\tilde{\alpha}_\epsilon(\mathbf{A}) : s, \text{ such that } f(\mathbf{A}, s) = \epsilon^{-1} \tag{5}$$

$\|.\|_{\mathcal{H}_2}$ is the H_2-norm, and the transfer function $\mathbf{H}_s(z) = \mathbf{V}(z\mathbf{I}-(\mathbf{A}-s\mathbf{I}))^{-1}\mathbf{U}$. Vanbiervliet *et al.* (2009) state that "the matrices U and V are to be seen as respective input and output weighting matrices, with (A, U) controllable and (V, A) observable". It will be shown that we do not need to consider these matrices closely, as the smoothed spectral abscissa is calculated by an alternative method (see Section 2.3). This function $f(A, s)$ has several useful properties.

Firstly:

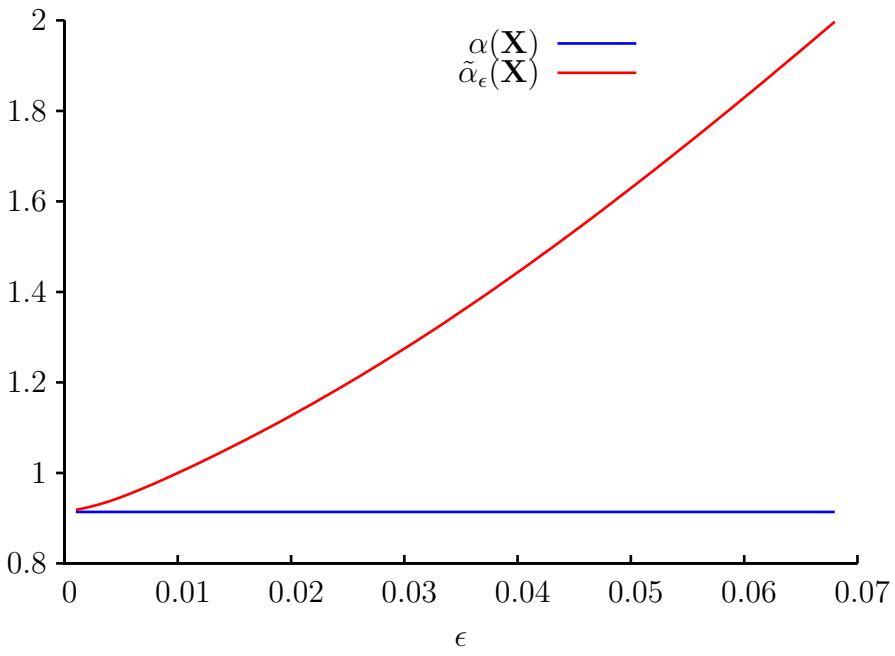


Figure 4: Using an example random matrix $\mathbf{X}$, with $\alpha(\mathbf{X}) = 0.914$. ϵ is used to characterise how tight the bound of $\tilde{\alpha}_\epsilon(\mathbf{X})$ is to $\alpha(\mathbf{X})$.

$$\frac{\partial f(\mathbf{A}, s)}{\partial s} < 0 \text{ , if } s > \alpha(\mathbf{A}) \tag{6}$$

meaning that the minimisation of $f(A, s)$ is a convex problem as long we ensure $s > \alpha(\mathbf{A})$. This is the property that transforms the non-linear, computationally intractible problem of reducing the spectral abscissa below zero to a form we can deal with computationally.

Secondly:

$$\tilde{\alpha}_\epsilon(\mathbf{A}) > \alpha(\mathbf{A}) \text{ , } \forall \epsilon > 0 \tag{7}$$

$$\lim_{\epsilon \to 0} \tilde{\alpha}_\epsilon(\mathbf{A}) = \alpha(\mathbf{A}) \tag{8}$$

which shows us a second important property of the smoothed spectral abscissa – that it tends to the spectral abscissa as ϵ tends to zero (shown in Figure 4). We, therefore, see that the parameter ϵ dictates how tight the upper bound of the smoothed spectral abscissa is to the spectral abscissa. A tight bound (given by smaller ϵ) is clearly desirable, as the smoothed spectral abscissa will therefore be a good approximation of the spectral abscissa, as initially desired. As the smoothed spectral abscissa is a valid upper bound on the spectral abscissa, and as is it also a smooth function of $\mathbf{A}$ (shown in Figure 3) it can be used to stabilise our linear network dynamics.

2.3 Computing the smoothed spectral abscissa and its derivatives

In the form defined in Equation 5 it is not obvious how to go about solving for $\tilde{\alpha}_\epsilon(\mathbf{A})$, as this representation does not allow an easily calculable equation for its derivatives.

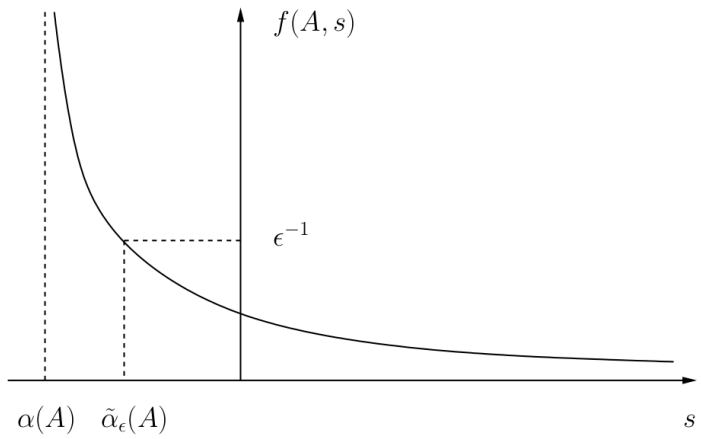


Figure 5: A typical form for $f(\mathbf{A}, s)$. A bisection method is used, evaluating $f(.)$ and $\frac{\partial f(\mathbf{A},s)}{\partial s}$ to find $\tilde{\alpha}_\epsilon(\mathbf{A})$.
(Figure adapted from Vanbiervliet *et al.* (2009).)

However, it can be shown that[7]:

$$f(\mathbf{A}, s) = \mathrm{Tr}(\mathbf{P}) = \mathrm{Tr}(\mathbf{Q}) \tag{9}$$

where the matrices $\mathbf{Q}$ and $\mathbf{P}$ are the solutions to the primal-dual Lyapunov pair:

$$0 = (\mathbf{A} - s\mathbf{I})\mathbf{P} + \mathbf{P}(\mathbf{A} - s\mathbf{I})^T + \mathbf{U}\mathbf{U}^T \tag{10}$$

$$0 = (\mathbf{A} - s\mathbf{I})^T\mathbf{Q} + \mathbf{Q}(\mathbf{A} - s\mathbf{I}) + \mathbf{V}^T\mathbf{V} \tag{11}$$

Using this alternative formulation of the problem, the derivatives of $f(\mathbf{A}, s)$ are simple to calculate.

$$\frac{\partial f(\mathbf{A}, s)}{\partial s} = -2\ \mathrm{Tr}(\mathbf{QP}) = -2\ \mathrm{Tr}(\mathbf{PQ}) \tag{12}$$

$$\frac{\partial f(\mathbf{A}, s)}{\partial \mathbf{A}} = 2\ \mathbf{Q}\ \mathbf{P} \tag{13}$$

We use a bisection method to calculate the value of $\tilde{\alpha}_\epsilon(\mathbf{A})$, using Equation 9 to evaluate $f(\mathbf{A}, s)$ and Equation 12 for $\frac{\partial f(\mathbf{A},s)}{\partial s}$ (see Figure 5). By evaluating these derivatives at $s = \tilde{\alpha}_\epsilon(\mathbf{A})$, and recalling that the minimisation of $f(\mathbf{A}, s)$ is a convex problem (Equation 6) it follows that a gradient descent based scheme to stabilise our network dynamics will be possible for our input reference network, as long as the primal-dual Lyapunov pair (equations 10 and 11) can be solved. Fortunately, this is a standard control problem, and so there exist libraries of functions to solve such pairs of equations in an efficient manner.

It is worth noting that the minimisation problem will become more non-linear as we reduce the bound parameter ϵ, and so will become harder computationally to solve. This translates to increased computation when gradient descent algorithms are used.

3 Biological considerations

Thus far we have discussed the mathematical methods involved in stabilising an initial reference network, but for these stabilised networks to have relevance to real cortical networks we need to take into account some experimental observations from biological networks both when initialising the reference network and during the stabilisation procedure.

3.1 Column constraints

Dale's Law states that any individual neuron can have either excitatory or inhibitory efferent synapses. When considering the system as characterised by the synaptic weight matrix, $\mathbf{W}$, this results in a columnwise constraint on the sign of elements w_{ij} of $\mathbf{W}$. To correctly model their action, excitatory connections (*i.e.* where the presynaptic neuron is excitatory) take positive values, and inhibitory connections take negative values. We define the first n columns of our $N \times N$ synaptic weight matrix to describe the strengths of excitatory connections, and the final $N - n$ columns the inhibitory connections. This gives $\mathbf{W}$ the structure shown in Figure 6.

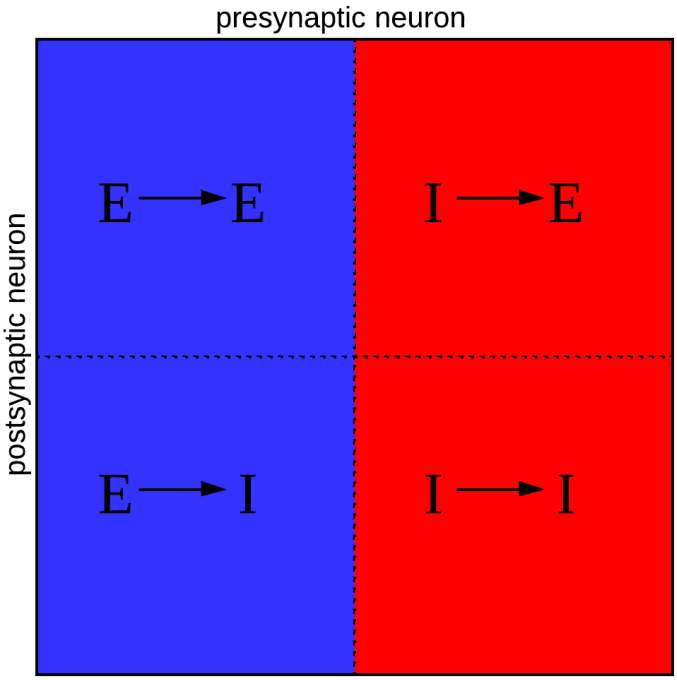


Figure 6: By organising the excitatory and inhibitory connections by column, and applying Dale's Law, we obtain weight matrices with this structure. Elements in the blue section are ≥ 0 and those in the red section are ≤ 0. (Note: Notation of the form E→E will be used throughout this report.)

3.2 Initialisation constants

The composition of neurons in cortical networks is typically assumed to be $n = fN$ excitatory and $N - n = (1 - f)N$ inhibitory, with $f \approx 0.8$. This necessarily changes the overall bias of the matrix $\mathbf{W}$ towards having more positive than negative elements. At initialisation of the synaptic weight matrix all excitatory (inhibitory) synapses are initialised to the same positive (negative) value. However, not every element of $\mathbf{W}$ is initialised to a non-zero value, as it is clearly not the case that the neural network is fully connected. If we assume $f > 0.5$, as is usual, an initialisation scheme with the same sparsity for excitatory and inhibitory sub-networks will result in an overall positive bias on $\mathbf{W}$, if both types of synapses are initialised to be the same strength.

As this project explores the effects of changing both the sparsity and the excitatory proportion parameter f, it is clearly necessary to control for the varying levels of positive (or even negative) bias that can result from these changes. This is achieved by setting the initial strengths of synapses in the excitatory and inhibitory networks (c_E and c_I respectively) to have different values. As the stabilisation procedure is an eigenvalue-based algorithm, it is therefore appropriate to use an eigenvalue metric to control for changing network bias. The metric used is the spectral radius, with c_E and c_I chosen to create reference matrices with controlled eigenspectrum radii.

The procedure used attempts to impose a radius, R, on the eigenspectrum of W. For a network with the same sparsity across the whole network, p, it can be shown that[5]:

$$\omega_o^2 = \frac{R^2}{p(1-p)} \tag{14}$$

$$\omega_E^2 = \frac{\omega_o^2(1-f)}{f} \tag{15}$$

$$\omega_I^2 = \frac{\omega_o^2 f}{1-f} \tag{16}$$

where $c_E = \frac{\omega_E}{\sqrt{N}}$, and $c_I = \frac{-\omega_I}{\sqrt{N}}$. These results follow from the global balance condition:

$$f\omega_E = (1-f)\omega_I \tag{17}$$

and greatly simplify the initialisation procedure.

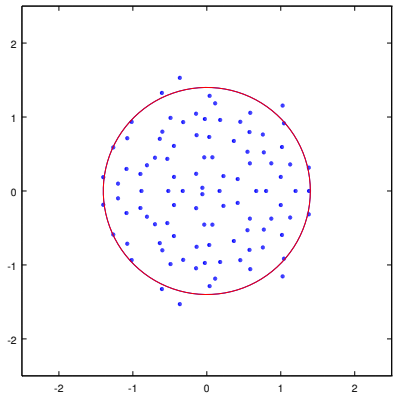


Figure 7: Eigenspectrum of a reference matrix. The excitatory and inhibitory weights ω_E and ω_I have been chosen to restrict the eigenspectrum to within the circle. The few eigenvalues located outside the circle are a result of the random sampling involved in the initialisation procedure (see Section 4.1.)

3.3 Non-random connectivity patterns

Experimental studies have found statistical properties of the cortex at a finer level of detail than simple sparsity constraints on the excitatory and inhibitory sub-networks[4]. The existence of these connectivity patterns is reassuring, as we would not expect complex and highly specific brain functions to arise from random connectivity patterns. The study by Song *et al.*[6] of rat visual cortex connectivity patterns revealed several non-random motifs, as well as distributions of synapse strength.

Connectivity patterns can be used to better initialise our **W** matrix, by imposing non-random structure on the reference network. The most significant effect found by Song *et al.* is the overrepresentation of reciprocal connections between neurons. That is, if a synapse exists propagating signals from neuron A to neuron B, it is approximately four times more likely than random that a symmetric synapse exists propagating signals from B to A (with confidence $p < 0.0001$), as shown in Figure 8. By introducing this connectivity pattern to the reference networks we hope to gain a more relevant insight into the structure of the stabilising inhibitory networks present in cortical networks.

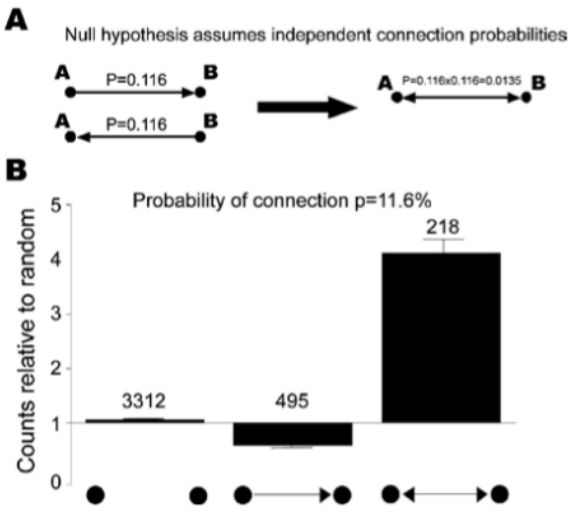


Figure 8: Analysis of cortical networks in rats showing the overrepresentation of bidirectional connectivity patterns.
A The calculation of the null hypothesis probability of bidirectional connections, calculated assuming independence of connections.
B Measured numbers of bidirectional connections are four times higher than predicted by a random network.
(Figure adapted from Song *et al.* (2005).)

4 Implementation

Having examined the mathematical and biological theory underpinning our methods, we now turn to the details of implementing these methods in practice.

4.1 Generating reference networks

The generation of the reference network for an optimisation is the first step in the method. The number of inhibitory columns and the desired size of $\mathbf{W}$ is passed to the function, and the sparsity and desired spectral radius of the resulting network is defined. From these values the initialisation constants are found as described in Section 3.2. The non-random connectivity patterns discussed in Section 3.3 are used optionally, controlled by a Boolean variable passed to the function.

For the simpler case, where these connectivity patterns is not included, the initialisation of the network is achieved by generating a random number between 0 and 1 for each element of the network. If that number is less than the desired sparsity, then the element is initialised to the relevant initialisation constant, otherwise it is set to zero. However, this will clearly not model the elevated probabilities of bidirectional connections. In order to include this added observation, we divide $\mathbf{W}$ along the diagonal. The elements in the upper right of the matrix are initialised randomly from a sparsity constraint, as before. Then, for the elements in the bottom left of $\mathbf{W}$:

$$p_{ij} = \begin{cases} p+c_{ij}(1-p) \text{ if } w_{ji}\neq 0 \\ p(1-c_{ij}) \text{ if } w_{ji}=0 \end{cases} \tag{18}$$

where p_{ij} = the probability that element $w_{ij} \neq 0$ (which is equal to the sparsity, p, when $\mathbf{W}$ is initialised without the bidirectional connectivity patterns), and c_{ij} is the normalised covariance between elements w_{ij} and w_{ji}. From this understanding of this covariance, we define $c_{max} = 1$ and $c_{min} = \frac{-p}{1-p}$. We then let:

$$c_{ij} = \kappa c_{max} \text{ , for connections between neurons of the same type} \tag{19}$$

$$c_{ij} = \kappa c_{min} \text{ , for connections between neurons of different types} \tag{20}$$

$$0 \leq \kappa \leq 1 \tag{21}$$

where κ is a parameter quantifying the strength of the symmetrical structure in the resulting $\mathbf{W}$ matrix. This is illustrated in Figure 9. The diagonal elements of $\mathbf{W}$ are also set to zero, as we do not desire any neuron to be connected to itself.

An example of the structure of reference matrix that results from this generation method (using $\kappa = 1$) is shown in Figure 10.

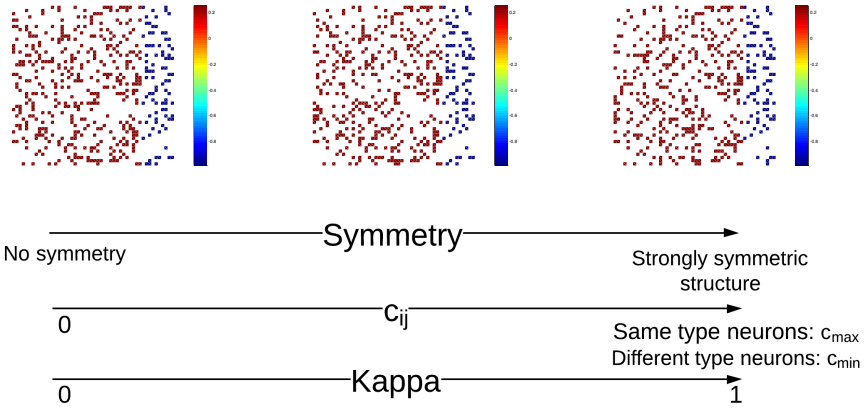


Figure 9: As we increase κ the strength of the symmetrical structure in the generated $\mathbf{W}$ matrix increases, as the correlation coefficient c_{ij} deviates from zero. The heat maps of example matrices generated in this way vary from random connectivity ($\kappa = 0$) to very strong symmetry in the diagonal ($\kappa = 1$).

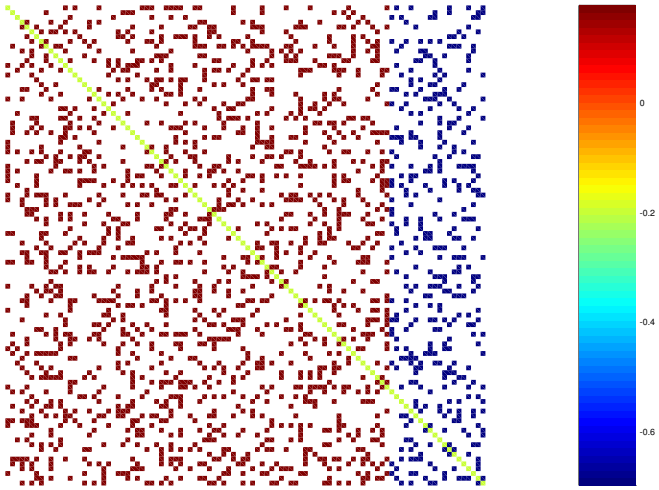


Figure 10: A typical reference matrix structure resulting from including the bidirectional connectivity pattern in the generation model. The mirror-like feature clearly visible across the diagonal (highlighted for clarity) of the matrix shows the symmetric bias encoded for by κ.

4.2 Solving for smoothed spectral abscissa and its derivatives

To stabilise our generated matrices we have to solve the primal-dual Lyapunov equation pair detailed in Equations 10 and 11. This is a standard control problem, and as such optimised software libraries are used in our algorithm to solve it. This is achieved by interfacing to the library functions in FORTRAN software libraries from the rest of the program (written in C++).

Having found $\frac{\partial f(\mathbf{A},s)}{\partial s}$, we use a Newton-Raphson bisection method to solve for the smoothed spectral abscissa, using Equation 5 and a chosen value of ϵ. Using $\epsilon = 0.01$ was found to be an acceptable compromise between the tightness of the bound on $\alpha(\mathbf{W})$ and computation time for the smoothed spectral abscissa. As $f(\mathbf{A}, s)$ is only defined for $s > \alpha(\mathbf{W})$, we must ensure that the Newton-Raphson root finding method doesn't attempt to evaluate $f(\mathbf{A}, s)$ outside this range. If an evaluation is attempted for $s < \alpha(\mathbf{W})$, we set $s = \alpha(\mathbf{W}) + 0.0001$ and restart. This ensures that $f(A, s)$ and its derivatives exist for all evaluations carried out in finding $\tilde{\alpha}_\epsilon(\mathbf{W})$.

The spring-box dynamics desired from the stabilised $\mathbf{W}$ resulting from this approach are achieved by keeping the excitatory network from the initialised network constant while tuning only the inhibitory network to enforce stability[3]. Furthermore, we only optimise the non-zero elements of the inhibitory network. Without this constraint the optimisation would tune all elements in the inhibitory columns of $\mathbf{W}$, including those initialised to zero, resulting in a fully populated inhibitory network. This is clearly not biologically plausible, and hence is prevented.

4.3 Gradient descent

As we can calculate values for $\tilde{\alpha}_\epsilon$ and its derivatives, a gradient descent method is used to stabilise $\mathbf{W}$. It is important to tune the step-size of this method, η, for two reasons. Firstly, having η too large will cause the algorithm to become unstable. Secondly, small values of η values will cause the algorithm to converge very slowly. We therefore use an adaptation to the standard gradient descent to allow a varying η. We initialise η to a high value found by observation of test trials ($\eta = 10$) and reduce it if unstable behaviour is detected. We define unstable behaviour to have occurred if any consecutive values of $\tilde{\alpha}_\epsilon$ during the stabilisation procedure are increasing, as the problem is convex.

Convergence of the algorithm is detected by a difference measure between values of $\tilde{\alpha}_\epsilon$. A record of previous values of $\tilde{\alpha}_\epsilon$ is kept, and if after a separation of one hundred iterations of the gradient descent algorithm the change in $\tilde{\alpha}_\epsilon$ is less than a threshold (set to 0.001), the algorithm is said to have converged. This form of convergence criterion is necessary as there is no analytical convergence criterion for gradient descent methods.

4.4 Reparameterisation

The reference network is forced to obey the desired biological constraints. However, the stabilisation procedure as described thus far does not consider them. Specifically, the sign of the synaptic weights w_{ij} are not constrained to remain the same as at initialisation, which is required by Dale's Law. We therefore reparameterise the stabilisation problem to address this.

By letting:

$$w_{ij} = b_{ij} e^{v_{ij}} \tag{22}$$

$$b_{ij} = \begin{cases} +1 \text{ for excitatory pre-synaptic neurons} \\ -1 \text{ for inhibitory pre-synaptic neurons} \\ 0 \text{ for no synapse present} \end{cases} \tag{23}$$

we observe that:

$$\frac{\partial f(A,s)}{\partial v_{ij}} = w_{ij} \frac{\partial f(A,s)}{\partial w_{ij}} \tag{24}$$

Reparameterising the problem in this way has introduced the new variable v_{ij}. As we cannot stabilise in a biologically plausible way using w_{ij} directly, we carry out gradient descent on the elements v_{ij} using Equation 24. From Equation 22 we see that, due to the exponential, we can let the sign of v_{ij} vary freely without affecting the sign of w_{ij}. This is because the sign of our synaptic weights w_{ij} is encoded for entirely by b_{ij}. We therefore use our gradient descent algorithm as described above to optimise our new variables v_{ij}, and from this calculate the resulting synaptic weight matrix $\mathbf{W}$ using Equation 22. As the signs of our weights w_{ij} remain constant throughout the stabilisation procedure, hence our stabilised $\mathbf{W}$ will obey Dale's Law. We also note that Equation 24 shows that the desire to only tune the non-zero elements in the inhibitory network is automatically satisfied by optimising over our new parameters.

4.5 Stochastic plasticity

The optimisation procedure aims to find an inhibitory network that is tuned to match the initialised excitatory network. However, the probabilistic initialisation of $\mathbf{W}$ makes no attempt to match the inhibitory synapses present to the structure of the excitatory network. The reparameterisation of the the gradient descent optimisation procedure necessarily prevents any synapses from decaying to zero, and no $w_{ij} = 0$ are changed by the stabilisation procedure to retain the desired network sparsity. Together these effects mean that there is no mechanism for changes in the structure of the inhibitory network in the procedure as described thus far.

In order to introduce optimisation of the structure as well as the synaptic weights of the inhibitory network, we must allow synapses to decay to zero. The reparameterisation

does not allow this, and so we introduce a threshold value for v_{ij}, such that if the synaptic strength w_{ij} becomes very small the synapse is said to have decayed. A new synapse is then created by initialising an element of the inhibitory network that was previously zero to a non-zero synaptic strength. The new synapse created connects the same pre-synaptic neuron as for the decayed synapse to a new post-synaptic neuron. This constrains w_{ik}, the synapse to be created (where $w_{ik} = 0$) to be from the same column of $\mathbf{W}$ as the decayed synapse, w_{ij}.

In this way we allow the inhibitory network to vary throughout the stability optimisation procedure. By allowing synapses to decay and reform in the inhibitory network we allow it to depart from the structure of the reference network. It should be noted that while this is constrained to produce plausible network structures, it is not a synaptic plasticity rule. It is possible that nature achieves similarly stabilised networks through forms of inhibitory plasticity, but the network analysis used here is clearly not a biological process. The method of using the smoothed spectral abscissa to stabilise the network is a shortcut to the end result, used because we do not have enough information about the biological processes at play to simulate the growth of a neural network of substantial size that results in the complex dynamics shown here.

4.6 Network dynamics

This stabilisation procedure is used as it produces networks that display high amplitude transient oscillations, similar to those seen in experimental motor cortex readings[3]. However, in order to evoke these transient dynamics the state, $\mathbf{x}(t)$, must be tuned to a 'preferred' initial state of the network, $\mathbf{a}$, at the point of release (*i.e.* $\mathbf{x}(t = 0) = \mathbf{a}$). In order to find the initial state that will maximise the transient amplification of the network's response ($\mathbf{a}_1$), we define an energy measure $\mathcal{E}(\mathbf{a})$:

$$\mathcal{E}(\mathbf{a}) = \frac{2}{\tau_r} \int_0^\infty \|\mathbf{x}(t)\|^2 dt \tag{25}$$

which will be finite in value for any stable network, as $\mathbf{x}(t)$ will decay exponentially for large t. As we are using a linear neuronal model, this can be solved analytically.

$$\mathcal{E}(\mathbf{a}) = \mathbf{a}^T \left[2 \int_0^\infty e^{t(\mathbf{W}-\mathbf{I})^T} e^{t(\mathbf{W}-\mathbf{I})} dt \right] \mathbf{a} := \mathbf{a}^T \mathbf{Q} \mathbf{a} \tag{26}$$

This $\mathbf{Q}$ is the solution to the Lyapunov equation:

$$(\mathbf{W} - \mathbf{I})^T \mathbf{Q} + \mathbf{Q}(\mathbf{W} - \mathbf{I}) = -2\mathbf{I} \tag{27}$$

and hence is found trivially. As $\mathbf{Q}$ is a positive symmetric matrix its principle eigenvector is the initial condition $\mathbf{a}_1$ which maximises the energy measure $\mathcal{E}$, and the evoked energy for this initial condition is the corresponding eigenvalue of $\mathbf{Q}$. In order to illustrate the

transient dynamics for the stabilised matrices, we choose an input, $\mathbf{x}_{ip}$ such that the system is in state $\mathbf{a}_1$ when the system is released. From the definition of the linear system dynamics (Equation 3), and noting that when the system is released the dynamics should be in steady state, we see that:

$$0 = (\mathbf{W} - \mathbf{I})\mathbf{a}_1 + \mathbf{x}_{ip} \tag{28}$$

$$\mathbf{x}_{ip} = (\mathbf{I} - \mathbf{W})\mathbf{a}_1 \tag{29}$$

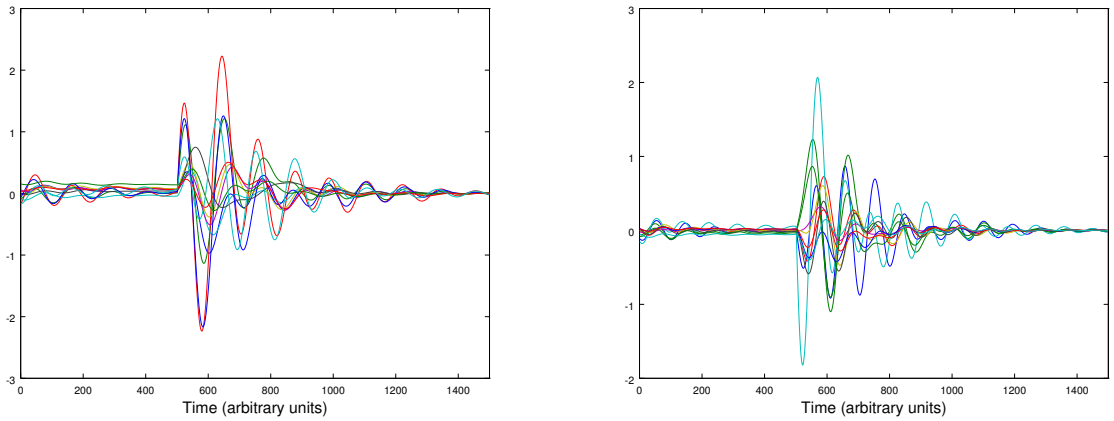
This input, $\mathbf{x}_{ip}$, is then applied in order to visualise the transient amplification response of the stabilised networks.

5 Results

5.1 Stabilisation plasticity

Hennequin *et al.* (2013) have shown that unstable networks initialised with $f = 0.5$ and obeying Dale's Law, stabilised using the smoothed spectral abscissa method, exhibit dynamics that are comparable to motor cortex dynamics. This result was replicated, and an example dynamical response from this procedure can be seen in Figure 11b. For an initial investigation into the importance of the inhibitory network structure, the inhibitory networks of matrices stabilised in this way were compared with those for which synapse decay and reformation was not allowed during the stabilisation procedure. Without this 'stabilisation plasticity' allowed, the stabilisation procedure can vary the strengths of the synapses created at initialisation of the synaptic weight matrix, $\mathbf{W}$, but their locations, and hence the inhibitory network structure, remains constant.

In order to isolate the effects of plasticity on the network structure the generated networks had constant sparsity with no reciprocal connectivity statistics implemented. The dynamics with and without stabilisation plasticity look qualitatively similar, and both exhibit the desired strong transient amplification but asymptotic stability required (Figures 11a and 11b).



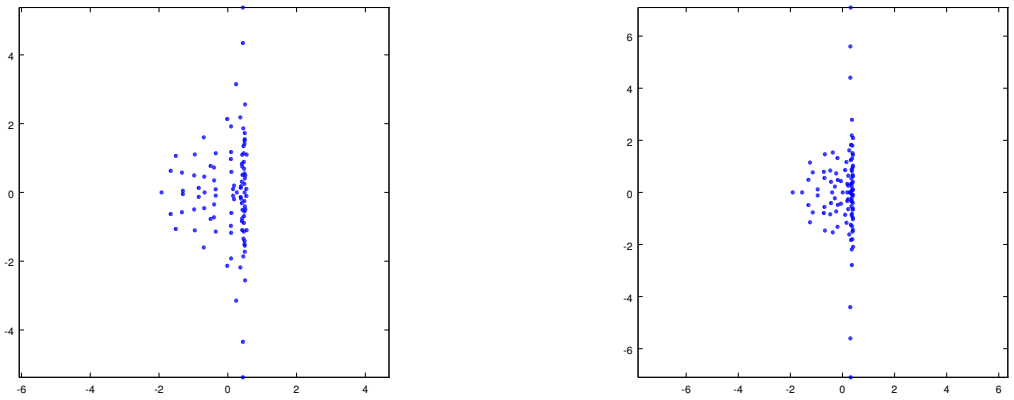
(a) No stabilisation plasticity

(b) With stabilisation plasticity

Figure 11: Example of transient amplification dynamics for stabilised $f = 0.5$ networks. Networks were generated using only a sparsity constraint ($p = 0.2$) and obeying Dale's Law. The system was subject to a preparatory input, held until time = 500, and then released. Arbitrary units are used as the form of the response will be the same at all scales.

The eigenspectra of these matrices show the stabilisation procedure has successfully forced all of the eigenvalues of $\mathbf{W}$ to below 1 both with and without plasticity (Figure 12),

as expected from the stable dynamics. The heat maps of example inhibitory networks show further similarities between the two, with the networks characterised by a few strong synapses amidst a large number of relatively weak connections (Figure 13). This implies that the synaptic strengths in the inhibitory network are matched to the excitatory network, with the synapses most important for stability strengthened significantly above the majority.



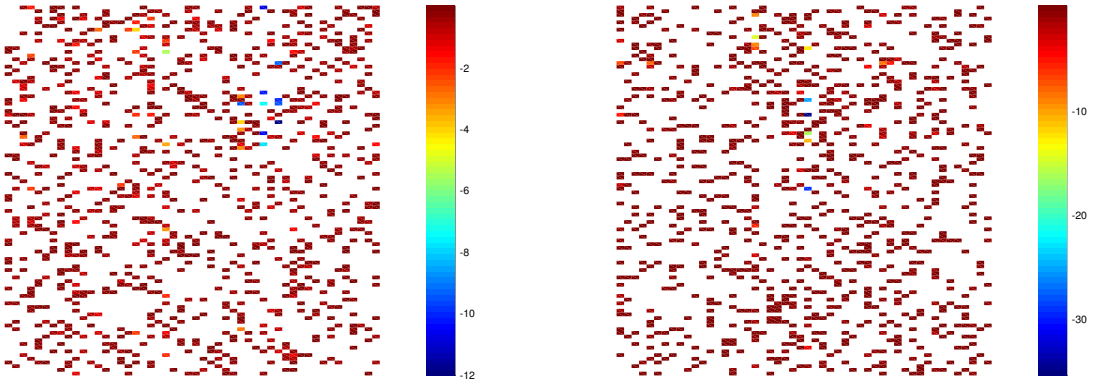
(a) No stabilisation plasticity

(b) With stabilisation plasticity

Figure 12: Example of eigenvalue distribution for stabilised $f = 0.5$ networks. The previously spherical distribution has been compressed into an approximate disk and all eigenvalues have been forced below 1.

In order to make stronger claims about the structure of networks stabilised in this way, 25 stabilisation trials were carried out with stabilisation plasticity and 25 without. For each set of 25 trials the excitatory network was kept constant, and the inhibitory network was re-initialised. The results from above were shown to be generally true, with all stabilising networks exhibiting the 'few strong synapses' trend.

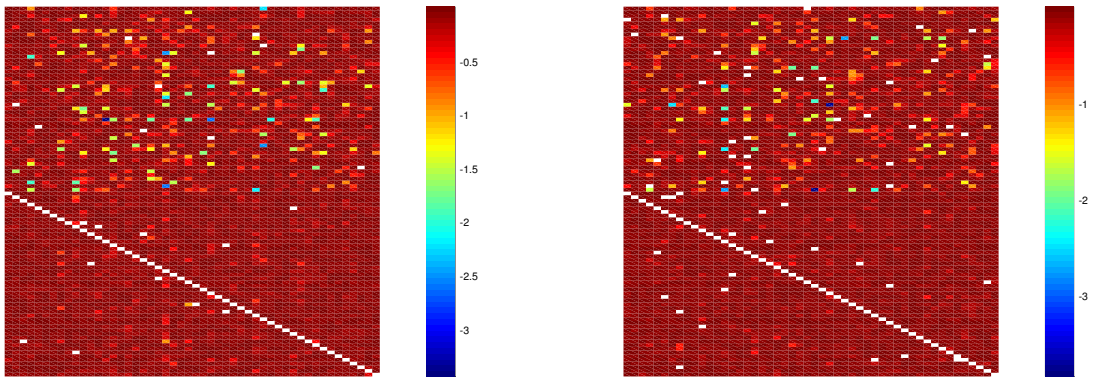
Figures 14a and 14b show the result of averaging all of the stabilising networks. As can be seen by the low sparsity of these average network plots, there is a lot of variation in the structure of the stabilising inhibitory networks for these trials. There is also no obvious visible difference in the structure of the average networks for the trials with and without stabilising plasticity. However, it can be seen that a majority of the weight of the inhibitory matrices lies in the top half of this network, that is, in the 'inhibitory connected to excitatory' section (denoted I→E hereafter). This is as expected, as it indicates that the majority of the stabilising power of the inhibitory networks is used to dampen activity in the excitatory neurons.



(a) Without synaptic plasticity

(b) With synaptic plasticity

Figure 13: The heat maps of the stabilising inhibitory networks show similar structures. A few strong connections tuned to the particular excitatory network amongst other, weaker inhibitory connections well stabilises the excitatory network. White indicates that there is no synapse present at at that location.



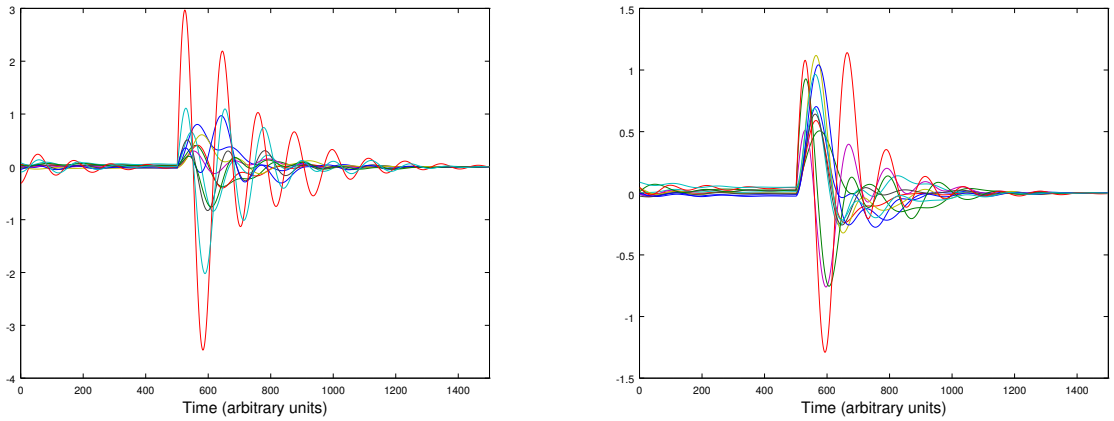
(a) No stabilisation plasticity

(b) With stabilisation plasticity

Figure 14: The trial averaged inhibitory networks. Both have stronger synapses in the top half of the network, corresponding to synapses with an excitatory post-synaptic neuron.

Other than this effect, however, the average inhibitory networks appear to be homogeneous. This would suggest that there are many possible solutions that stabilise the given excitatory network. During simulation it was observed that stabilisation took very different amounts of program time across trials while stabilising to a similar final value of smoothed spectral abscissa, despite the networks being initialised with the same spectral radius. This indicates that, while there are several possible inhibitory networks that will stabilise the overall network, some structures are in some sense worse (*i.e.* harder to force into stability) than others.

It is clear that a large amount of stabilising power is being unused, as the inhibitory network has few and weak internal connections. A more efficient balance would therefore be to have a higher proportion of excitatory neurons. This is the result we expect from biological observations, as $f = 0.8$ is a more usual value for cortex networks. The same trials as for $f = 0.5$ networks were carried out on networks using this more realistic value of f, resulting in the dynamics seen in Figures 15a and 15b. These show that changing the proportion of excitatory neurons has not adversely affected the transient amplification characteristic we are trying to model; in fact the richest dynamics are observed when $f = 0.8$ and synaptic plasticity is used (Figure 15b).



(a) No stabilisation plasticity

(b) With stabilisation plasticity

Figure 15: Transient amplification dynamics for networks with $f = 0.8$.

To quantify the similarity between the stabilising inhibitory networks for different trials we use a class correlation measure. For this we take the normalised correlation (r_{pq}) between each pair of networks, defined as:

$$r_{pq} = \frac{\sum_k (\mathbf{w}_{p,k} - \bar{\mathbf{w}}_p)(\mathbf{w}_{q,k} - \bar{\mathbf{w}}_q)}{\sigma_p \sigma_q} \tag{30}$$

Here we have defined $\mathbf{w}_q$ to be a vectorised version of the inhibitory network of trial q, $\bar{\mathbf{w}}_q$ to be the mean of that vector, and σ_q its standard deviation. $\mathbf{w}_{p,k}$ is element k of the

	Without stabilising plasticity		With stabilising plasticity	
f	0.5	0.8	0.5	0.8
Class correlation	0.034	0.040	0.045	0.039
Structural correlation	0.001	0.003	0.011	0.006

Table 1: Class correlation values for two values of f with and without stabilising plasticity. Class correlation was calculated both for the full inhibitory network and also for the network structure. This was achieved by normalising all non-zero elements to remove strength correlation effects.

---

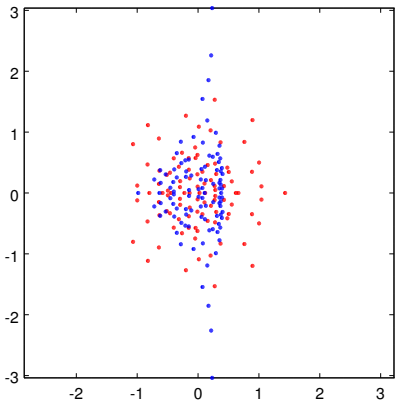
vector $\mathbf{w}_q$. The class correlation is then the mean of the correlations between all possible pairs in the class. Calculating this for the networks discussed so far yields the results in the first row of Table 1.

The class correlation values match our observation that the inhibitory networks have similar forms whether stabilising plasticity is used or not. The structural correlation quantity is calculated in the same way, but removes any dependence on the synaptic strengths by normalising any synapses present to 1. Without stabilising plasticity this structural correlation is very low due to the randomisation in the reference network initialisation. The higher values of the structural correlation for the networks using stabilising plasticity indicates that these networks have slightly higher than random structural correlation. This shows that there is at least some convergence of the inhibitory networks towards similar solutions, with synapses that do not affect the stability of the network being allowed to decay.

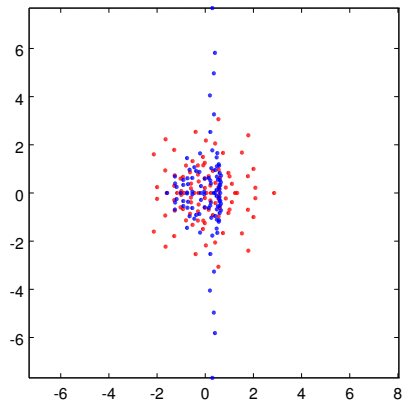
Across all networks the class correlation is significantly higher than the structural correlation, which reflects the fact that the tuning of the synaptic weights to the excitatory network is the dominant effect in the stabilisation.

5.2 The effect of spectral radius

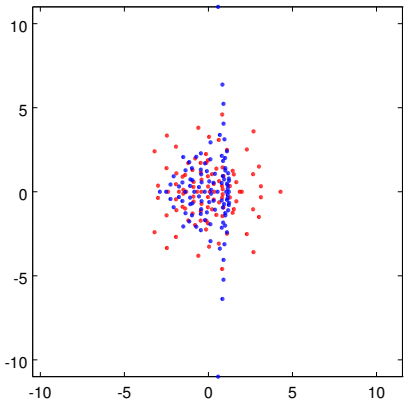
Previously the spectral radius was kept constant in order to observe the effects of stabilisation plasticity. In order to investigate how spectral radius affects the structure of the stabilising inhibitory network, we allow stabilisation plasticity during the gradient descent algorithm, but initialise the synaptic weight matrices to have different spectral radii. In order to isolate the effects of the difference in spectral radius we generate only one synaptic weight matrix and attempt to stabilise a copy of this matrix scaled by a constant β in each trial. We therefore have, for each of 25 trials at each value of β, $\mathbf{W}_{unstable} = \beta \mathbf{W}_{generated}$, where $\mathbf{W}_{unstable}$ is the initial matrix that is then stabilised. The $\mathbf{W}_{generated}$ used has spectral radius $= 1.4$, $f = 0.8$, and we analyse $\beta = 1, 2, 3$. Further values of β are not analysed as the excitatory network becomes too strong to be stabilised by our method at this scale. The form of the eigenvalue distributions of the $\mathbf{W}$ matrices before and after the stabilisation procedure has been carried out is shown in Figure 16.



(a) $\beta = 1$



(b) $\beta = 2$



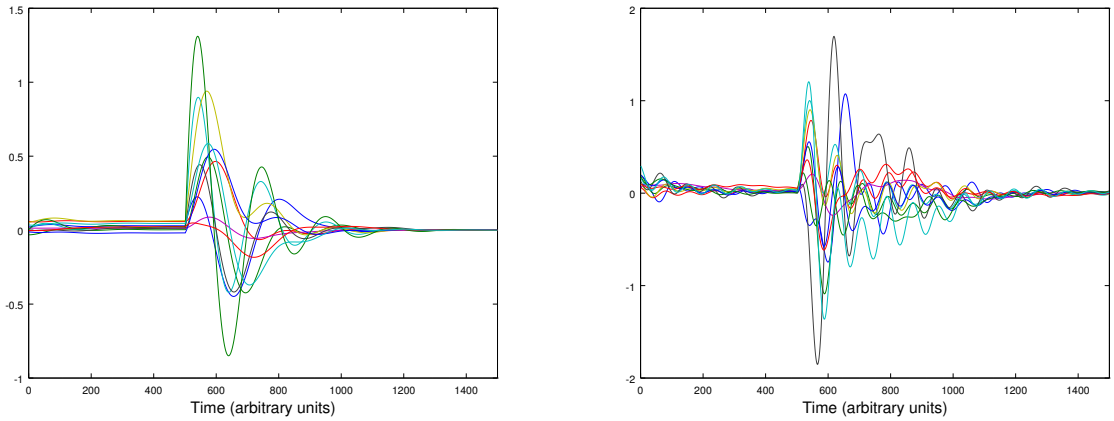
(c) $\beta = 3$

Figure 16: Eigenspectra for the initial and final $\mathbf{W}$ matrices for varying levels of β. The red points are the initial eigenvalues, distributed within a circle of radius 1.4β, and the stabilised spectra are in blue.

We can see that the method has compressed the circularly distributed eigenvalues of the initial weight matrix into a vertical 'disk' structure as before. The height of the disk scales approximately proportionally with β.

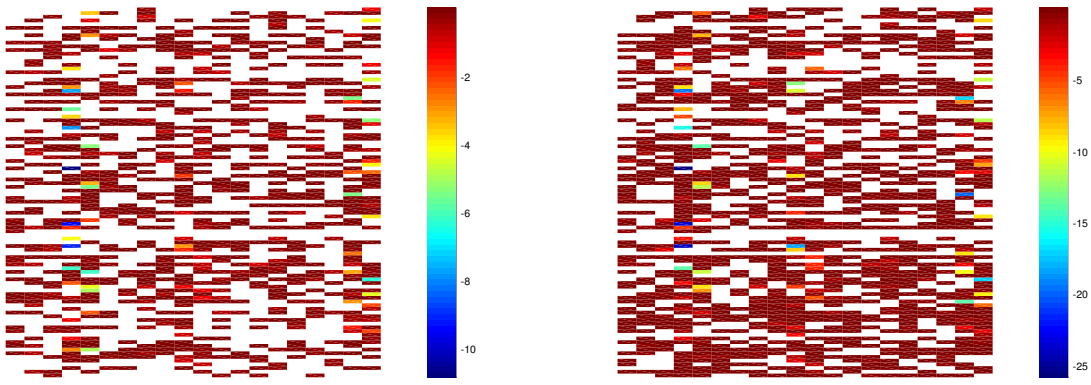
The larger spectral radius results in the system being harder to stabilise. This is observed as the mean spectral abscissa values increase with β (see Table 2). This shows that for $\beta = 3$ we have failed to produce a stable network, despite the method converging and the eigenspectrum having the expected shape. As would be expected, the dynamics for the converged networks reflect these spectral abscissa values (Figure 17). We can see that a larger spectral abscissa results in a more oscillatory network with larger transient amplification.

The average inhibitory networks for the stable values of β are shown in Figure 18. These exhibit some similar properties to those discussed in Section 5.1; there are a few



(a) $\beta = 1$ (b) $\beta = 2$

Figure 17: The dynamical responses for networks initialised with β values that can be stabilised. Larger β makes the network harder to stabilise, and hence results in larger transient amplification and greater oscillation in the dynamics.



(a) $\beta = 1$ (b) $\beta = 2$

Figure 18: The average of the inhibitory networks for the stable values of β analysed. The networks are more sparse than previously (Figure 14) due to the use of the same $\mathbf{W}_{generated}$ for all trials.

synapses that are much stronger than the average for the network, and the average synaptic weight in the I→E section is higher than that in the I→I section. However, as we used a constant structure for our $\mathbf{W}_{unstable}$ across all trials, there is much more structure seen in the average inhibitory network. This structure is quantified by the class correlation and structural correlation values shown in Table 2.

Table 2 shows that for $\beta = 1$ the class correlation is significantly higher than the structural correlation. This implies that the synaptic strengths are well correlated, on top of the structural correlation. The fact that this effect is not observed for $\beta = 2$ networks implies that the strength, not just the structure, of the excitatory network influences the form of the stabilising inhibitory network. This conclusion is reinforced by the fact that the mean synaptic strength for the inhibitory networks increases in magnitude faster than the corresponding increase in β. The $\beta = 3$ networks have a mean inhibitory synapse strength of approximately four times that of the $\beta = 1$ networks, and yet the final networks for the $\beta = 3$ trials are not stable, whereas for $\beta = 1$ the spectral abscissa is just 0.382. The mean inhibitory synaptic weight has therefore increased at a greater rate than the initial spectral radius, but the stabilisation performance has become significantly worse.

β	1	2	3
Class correlation	0.968	0.658	—
Structural correlation	0.797	0.670	—
Mean synaptic strength	-0.716	-1.731	-2.8368
Mean spectral abscissa	0.382	0.704	1.119

Table 2: Class correlation values for the two stable values of β, and the mean strength of non-zero synaptic weights in the inhibitory networks after the stabilisation algorithm (note that for $\beta = 3$ the resulting network is still unstable)

5.3 Introducing symmetry

So far the networks stabilised have modelled the proportion of the network which is excitatory and the synaptic sparsity. However, subject to these constraints, each element in a network has been chosen randomly, resulting in random connectivity patterns in the initial network. As discussed in Section 3.3, this random connectivity model is not consistent with experimental observations. We therefore introduce positive correlation to the reference network, with the strength of the correlation given by the parameter κ (see Section 4.1). An example of a reference network initialised with strong positive correlation is given in Figure 19.

The partial symmetry imposed on the initial network by setting $\kappa > 0$ has several effects. To investigate these we initially use $\kappa = 0.4$ and contrast with our earlier findings for networks initialised with random connectivity structure. The first notable difference with non-zero κ is in the initial eigenspectrum (Figure 20). The positive correlation has flattened the previously spherical distribution of eigenvalues, and stretched it parallel to the real axis. If negative correlation were to be implemented it would similarly stretch the eigenspectrum, but parallel to the imaginary axis. The stretching of the eigenspectrum from the positive correlation used increases the spectral abscissa. Positively correlated initial networks are therefore more difficult to stabilise than randomly connected networks for the same excitatory synaptic strengths.

The stabilised eigenspectrum is as expected, forming a vertical disc with a short tail of negative eigenvalues close to the real axis (Figure 21a). Comparing with Figure 12 we see that the stabilisation algorithm has resulted in a very similar stable eigenspectrum, despite the larger spectral abscissa in this case. Figure 21b shows that the stable network also exhibits similar dynamics to the uncorrelated case. These results are encouraging, as the motivation for this analysis was based on the observation of the network dynamics. It is therefore pleasing to see these dynamics retained when further biological considerations are included.

Analysing our stabilised networks with some symmetry included shows that there is less correlation between the stabilising inhibitory networks than for randomly connected networks (Table 3). This is not as expected, the introduction of more structure into the excitatory network was expected to induce stronger structure in the inhibitory network. The heat map of an example inhibitory network with $\kappa = 0.4$ is shown in Figure 22a, from which we can see that stabilisation is again achieved by a few strong synapses within a network of weaker connections. These two results imply that while the form of any individual stabilising inhibitory network is similar whether or not correlation considerations are used at initialisation, imposing some symmetry on the reference network has removed any structural similarity between these inhibitory networks.

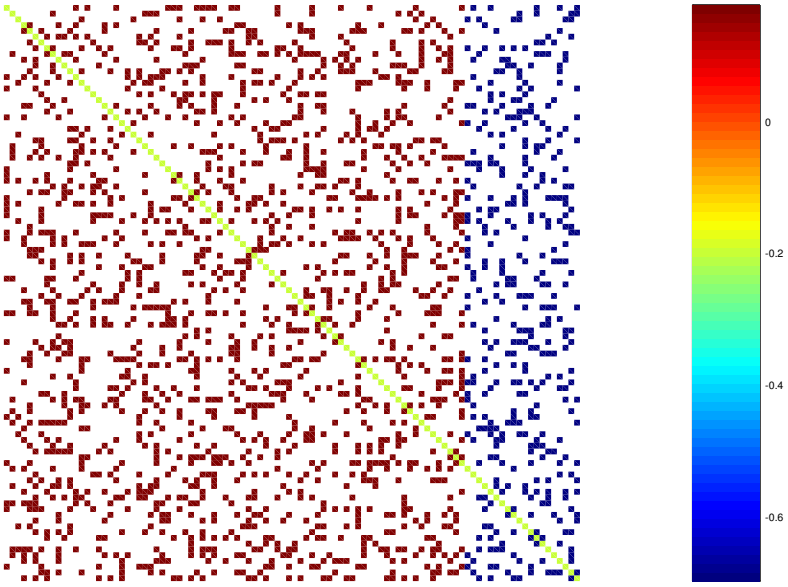


Figure 19: The heat map of a reference network initialised with $\kappa = 1$ and the diagonal highlighted in green such that the symmetrical properties are clearly seen. For this value of κ the E→E and I→I networks are fully symmetric in the diagonal, and if synapse ij exists in the I→E network, then synapse ji (in the E→I network) will not exist. Note that the E→I and I→E networks are not fully anti-symmetric due to the sparsity constraint.

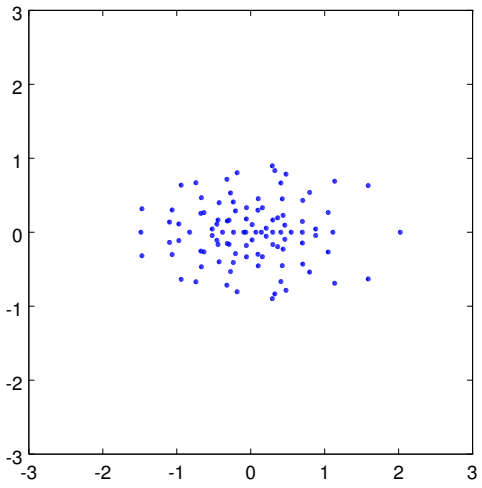
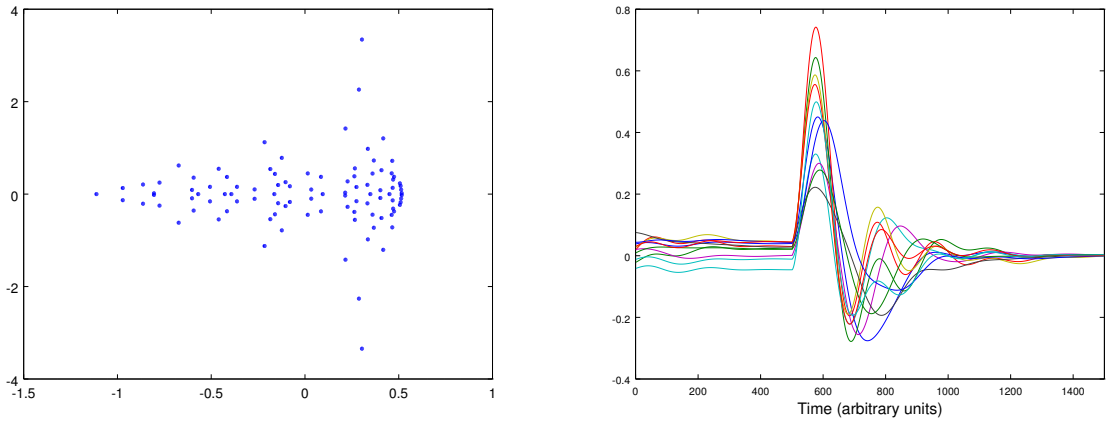


Figure 20: Elliptical eigenspectrum resulting from positive correlation in the generation of the reference matrix. The major axis of the ellipse lies parallel to the real axis for positive correlation, and parallel to the imaginary axis for negative correlation.

	$\kappa = 0.4$ (With some symmetry)	$\kappa = 0$ (No symmetry)
Class correlation	0.003	0.039
Structural correlation	0.002	0.006

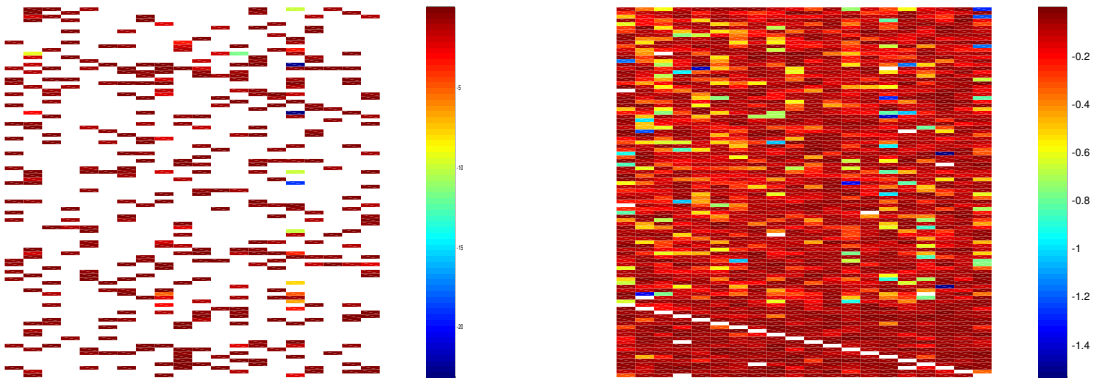
Table 3: Correlation values of stabilising inhibitory networks where the reference network has been initialised with and without correlation considerations. In both cases the E→E network is kept constant. The correlation constraint requires the E→I network to be reinitialised as well as the inhibitory network, whereas this is not the case when no correlation is implemented.



(a) Eigenspectrum of stabilised correlated network

(b) Stable dynamics again showing high transient amplification

Figure 21: The stabilised correlated network has similar properties to the stabilised uncorrelated network, with the same form of eigenspectrum and retaining the desired dynamical response characteristics.



(a) An example network

(b) Average over all networks

Figure 22: Heat maps of the stabilising inhibitory networks which stabilise correlated reference networks. An example network shows that the form of the stabilising networks is similar to that without the correlation on initialisation. The average of all networks, however, shows that the correlation between stabilising networks that was seen before is no longer present.

5.4 Varying levels of symmetry

Song *et al.* (2005) demonstrate that positive correlation is present in rat visual cortices and quantifies this correlative effect, finding that reciprocal connections in this area are four times more likely than predicted by randomly connected networks. However, the process of measuring network connectivity experimentally is very difficult, and as a result there is only limited information about the amount of synaptic symmetry present in neural networks across multiple brain areas. We therefore investigate the effects of different levels of symmetry in the initial weight matrix we carried out a comparative analysis on sets of networks initialised with values of κ between 0 and 1.

We found in Section 5.3 that imposing positive correlation on the generation of the synaptic weight matrix results in an elliptical eigenvalue structure lying along the real axis. Figure 23 shows that this ellipse becomes more eccentric as κ increases, with the eigenvalues having smaller imaginary components. This increased eccentricity also causes an increase in spectral abscissa of the initial matrices. Higher correlation in the synaptic weight matrix therefore makes it harder to stabilise the network.

The eigenspectrum of networks initialised using positive synapse correlation after stabilisation has the general form seen in Figure 21a; a disk-like structure with a small 'tail' of eigenvalues with negative real parts and small imaginary components. Changing the strength of the symmetry in $\mathbf{W}$ by altering κ varies both the height of the disk and the length of the tail (shown in Figure 24). The length of the tail (measured from the origin along the negative real axis) increases with κ, whereas the height of the disk decreases.

The increased difficulty in stabilising the network results in a larger spectral abscissa at convergence of the stabilisation algorithm (see Table 4). This has implications for the dynamics; Figure 25 shows the dynamic responses for networks with $\kappa = 0.2$ and $\kappa = 0.8$ (Figures 25a and 25b respectively). The greater difficulty in stabilising the $\kappa = 0.8$ network is evident from the greater oscillation present.

Since greater correlation increases the difficulty of stabilising the network, we might assume that this would result in greater correlation between the stabilising inhibitory networks. However, we can see from Table 4 that this is not the case, with negligible correlation and structural correlation between stabilising networks observed for all non-zero values of κ.

Introducing correlation into the generation model for the networks has nullified one of the three main conclusions about the structure of uncorrelated initial networks; the synapse structures are no longer correlated for constant excitatory networks. A second result was that stabilising networks were made up of a few strong connections amongst a larger network of weak connections. This structure is clear when a histogram of the inhibitory synaptic weights is viewed: Figure 26a, a histogram of the non-zero synaptic weights in the stabilising inhibitory network, shows a vast majority of the synaptic weights

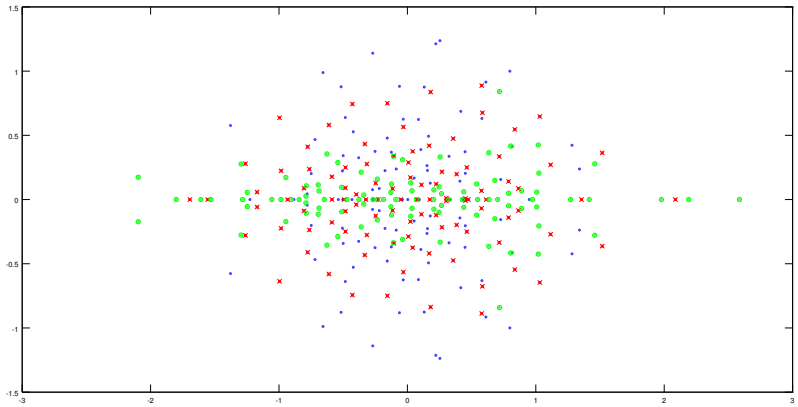
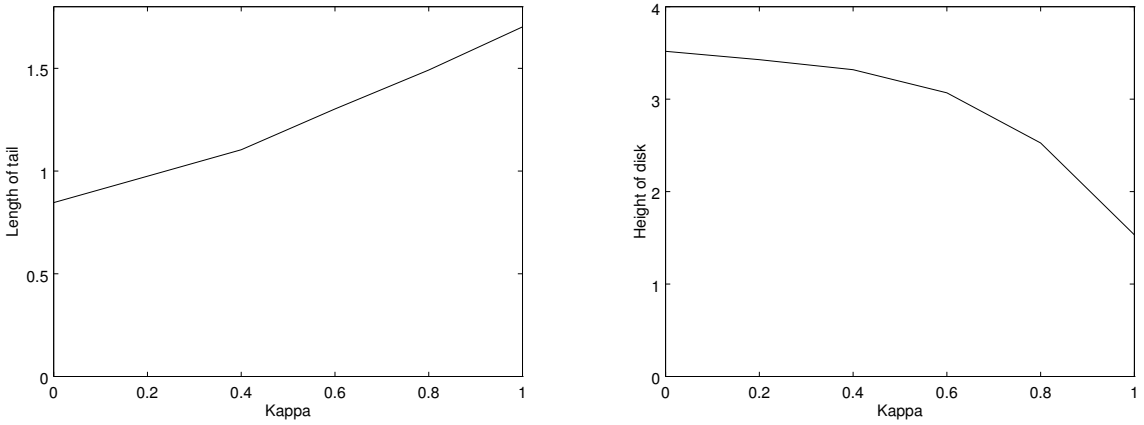


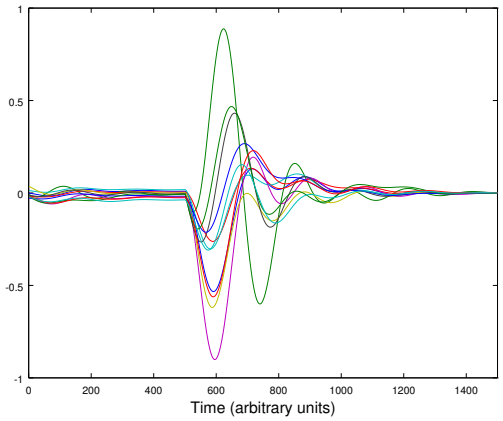
Figure 23: The eigenspectra of networks with $\kappa = 0$ (blue), $\kappa = 0.5$ (red) and $\kappa = 1$ (green) before stabilisation. The eigenvalues lie within an ellipse, with stronger symmetry in the synaptic weight matrix resulting in greater eccentricity of the eigenvalue ellipse.



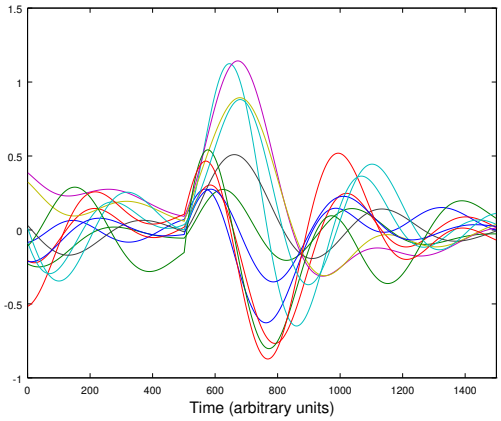
(a) Length of eigenspectrum tail

(b) Height of eigenspectrum disk

Figure 24: Trends for the variation of height of the disk-like structure and length of the negative tail with κ, for stabilised eigenspectra. The values shown are averages of 25 trials for each value of κ.



(a) $\kappa = 0.2$



(b) $\kappa = 0.8$

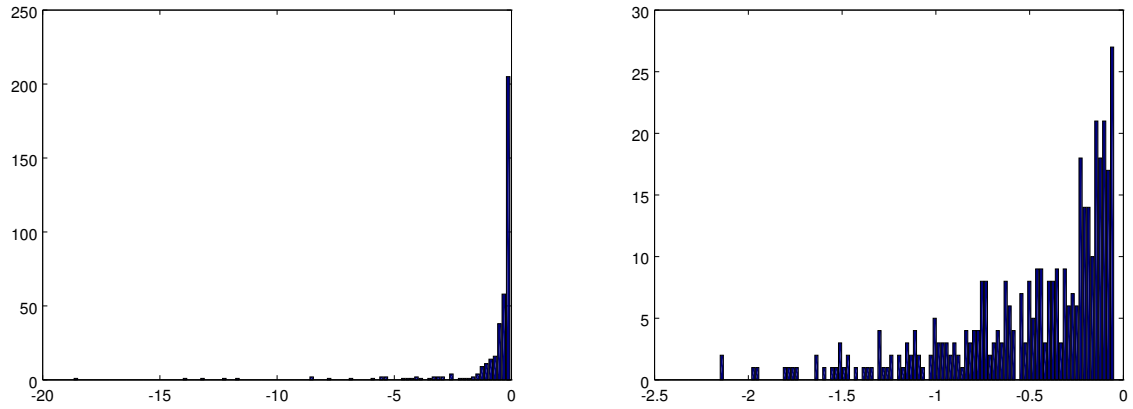
Figure 25: The increase in spectral abscissa associated with increasing κ results in more oscillatory dynamics of the stabilised networks.

κ	0	0.2	0.4	0.6	0.8	1.0
Trial correlation	0.039	0.003	0.003	0.001	0.001	0.003
Variance of inhibitory synaptic weights	3.65	3.62	4.11	3.58	2.50	0.459
Mean final $\alpha(\mathbf{W})$	0.429	0.500	0.559	0.669	0.757	0.889

Table 4: Statistical network analysis results for stabilising inhibitory networks. The initial synaptic weight matrices have been initialised with different levels of correlation.

for an uncorrelated network are near zero, with a few outlying values at large magnitudes. However, at the other end of the scale, Figure 26b displays a much lower spread of synapse strengths.

We can quantify this 'spread' by taking the variance of the non-zero synaptic strengths in the stabilising inhibitory networks, as well as analysing the average synaptic strength. From the values shown in Table 4 we see that the variance and mean values remain approximately constant up to $\kappa = 0.6$, with a reduction in the mean strength coupled with a much sharper reduction in the variance for higher values of κ. From our histogram and variance analysis we see that stabilising inhibitory networks exhibit the 'few strong, many weak' distribution of synaptic weights up to a value of around $\kappa = 0.7$.



(a) $\kappa = 0$ (b) $\kappa = 1$

Figure 26: a) The property of the non-zero stabilising inhibitory networks for uncorrelated initial weight matrices of having a small number of strong connections in a network of much weaker synapses can be visualised by a histogram.
b) A fully symmetric network doesn't display this property, instead having much more uniformly distributed synaptic weights.

The value of κ which results in a fourfold increase in reciprocal connections over random connectivity counts, as found by Song *et al.* in rat visual cortices is $\kappa = 0.75$. It is interesting to note (from Table 4) that the mean and variance of the synaptic weights remains high for $\kappa < 0.7$, but then decrease sharply, while the value of $\alpha(\mathbf{W})$ for the stabilised networks increases with κ. This transition zone is intriguingly close to the value of κ suggested by the experimental data.

6 Conclusions

Using the smoothed spectral abscissa with a gradient descent optimisation scheme we have shown that unstable matrices with a variety of different initialisation constraints can be stabilised. This is achieved by manipulating the inhibitory synapses to be tuned both in strength and location to the excitatory network dynamics. This allows networks with strong excitatory connections to be asymptotically stable, resulting in complex network dynamics exhibiting large transient amplification similar to that observed in biological experimental data.

By considering the form of the excitatory networks of the initial unstable matrix we hoped to uncover structural details about the resulting stabilising inhibitory networks. The tuned inhibitory networks place stronger synapses in locations where the post-synaptic neuron is excitatory, with little inhibition directed to other inhibitory neurons. This observation suggests that networks with a large proportion of inhibitory neurons are not utilising the stabilising 'power' of the inhibitory network efficiently. The network will have greater information capacity if a larger proportion of the network is excitatory, which suggests that using the inhibitory synapses more efficiently, and hence being able to reduce their number, will increase the capacity of a network. This agrees with experimental observations of the neuronal composition of neural networks, and an excitatory fraction of $f = 0.8$ is usually assumed for cortical network models.

This is not to say that all networks can be stabilised in this way. Stronger excitatory networks were found to be more difficult to stabilise. Using $f = 0.8$ and an initial synaptic weight matrix with no non-random connectivity considerations we found that a spectral radius of around 4 was the limit that this method could stabilise. Networks that are harder to stabilise result in a larger final spectral abscissa, and the resulting eigenspectrum is further from the 'vertical disk' distribution of eigenvalues investigated by Hennequin *et al.* (2013). The dynamics of a stabilised network depend heavily on its spectral abscissa as expected; both the duration and amplification of oscillations increase with the spectral abscissa.

Stabilising scaled versions of the same excitatory network also yielded the result that the mean inhibitory synaptic weight increases faster than the scaling factor, β. Despite this, the final matrix for larger β values was less stable. This strengthens the conclusion that the precise tuning of the inhibitory system is the more significant effect in terms of the network stabilisation, as stronger general inhibition does not have the desired stabilising effect. Stabilising networks for forms of initial synaptic weight matrix with random connectivity patterns have a structure based around a few synapses a long way above the mean inhibitory weight, among a large majority of weaker connections.

Biological data suggests that there is some level of reciprocity in synaptic connections between neurons. This reciprocity was modelled, and imposes some symmetry onto the

synaptic weight matrix. It was found that the eigenvalue distribution of the initial matrices with synaptic reciprocity included in the generation model is elliptical, with the major axis of the ellipse lying along the real axis. Increasing the level of symmetry modelled increases the eccentricity of this ellipse. The transformation of the eigenvalue distribution from a circular distribution for randomly connected networks to this elliptical structure increases the spectral abscissa of the network without increasing the average synaptic strength of the network.

The results from different levels of symmetry show that a properly tuned network, giving stable, fast, transient dynamics, will have a few strong, many weak, structure of inhibitory synapses (quantified by the variance of inhibitory synapse strengths; Table **??**). High levels of symmetry in the excitatory network increase the initial spectral abscissa, but the degradation of dynamical performance seen is disproportionate to the spectral abscissa. This is because the very high number of reciprocal connections creates much stronger local positive feedback than would be expected for randomly created network structures.

Possible further work

The obvious avenue for further research is to extend the network generation model used to initialise the excitatory networks. Song *et al.* describe multiple over-represented triplet motifs of connectivity patterns involving three neurons, which could be added. It would also add to the model to draw excitatory synaptic weights from a more accurate distribution of weights, as opposed to initialising them to a constant value, as is currently implemented.

7 Bibliography

[1] Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I., and Shenoy, K. V. (2010). *Cortical preparatory activity: representation of movement or first cog in a dynamical machine?* Neuron, 68:387400.

[2] Hennequin, G., Vogels, T. P., and Gerstner, W. (2012). *Non-normal amplification in random balanced neuronal networks.* Physical Review E, 86:011909.

[3] Hennequin, G., Vogels, T. P., and Gerstner, W. (2013). *Optimal control of transient dynamics in cortical network models.* Neuron. (Accepted)

[4] Milo R., Shen-Orr S., Itzkovitz S., Kashtan N., Chklovskii D. B., et al. (2002). *Network motifs: Simple building blocks of complex networks.* Science 298:824827

[5] Rajan, K. and Abbott, L. F. (2006). *Eigenvalue spectra of random matrices for neural networks.* Physical Review Letters, 97:188104.

[6] Song, S., Sjöström, P. J., Reigl, M., Nelson, S., Chklovskii, D. B. (2005). *Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits* PLoS Biol., 3:0508-0519

[7] Vanbiervliet, J., Vandereycken, B., Michiels, W., Vandewalle, S., and Diehl, M. (2009). *The smoothed spectral abscissa for robust stability optimization.* SIAM Journal on Optimization, 20:156171.

8 Appendix

8.1 Risk assessment

The risks associated with a computer-based project are primarily posture problems, repetitive strain injuries and tripping over trailing cables. These were identified at the start of the project, and the risks were mitigated with the use of a well adjusted desk chair, regular breaks and tidying cables away.

8.2 Code repository

The software used for this project can be found at:
https://github.com/tr325/neural-network-control