The Eigenvalue Value (in Neuroscience)

Georgia Christodoulou[1] & Tim P. Vogels[1,2]

[1]*Centre for Neural Circuits and Behaviour, University of Oxford, United Kingdom*
[2]Institute of Science and Technology, Austria

The connectivity of neuronal networks is thought to lie at the heart of their processing abilities. We are starting to glean their complexity by way of connectomics, yet we still lack the tools to *understand* them, or extrapolate the dynamics they may produce. However, there are powerful mathematical tools that can be used to unveil parts of this structure–function relationship. We present the underlying mathematics needed to understand network structure, and link the theory directly to its neuroscience applications. We start by presenting the archetypal linear dynamical system and show how it can be used to gain an intuition for neuronal activity using the eigenvectors and eigenvalues of the connectivity matrix. We explore how the Schur decomposition and the theory of pseudo-spectra can offer additional insight and finish with a brief review of recent studies in theoretical neuroscience that use the language of random matrix theory. We aim to take readers without a background in mathematics by the hand and walk them through the mathematical formalism by way of intuitive examples and a tutorial code. We finish with an overview of some of the most mathematically profound research papers in theoretical neuroscience.

The brain is a large dynamical system: constantly evolving over time in response to external inputs and in accordance with its own connectivity [1–11]. This connectivity lies at the heart of almost all neuronal matters, e.g., spontaneous and evoked activity, learning and plasticity, excitatory-inhibitory balance, and oscillations. Understanding the link between structure and function, or in other words, between a matrix **W** (which represents the recurrent connections between neurons in a network) and the resulting dynamical responses becomes a fundamental problem. Previous work has shed light on aspects of this structure-function relationship [11–22] but it is still impossible to carve a clear, direct path between the two. Here, we follow a simplified, middle-step approach. We use the fact that between **W** and dynamics lies the eigenspectrum: An essential property of **W** and also a visually powerful tool that is helpful for understanding computations as well as generating intuition. In the following pages we will try to cover some important results that link either the eigenspectrum with the dynamics, or the eigenspectrum with the connectivity matrix **W**; allowing us to investigate the link between structure and function. We will start with the main definitions and standard uses of the eigenspectrum in the study of dynamical systems, and discuss what can we learn about the dynamics by looking at eigenspectra. Then we will look at the limitations: the questions that we cannot answer by the spectrum alone. We will go on and discuss ways of uncovering those "unseen by the spectrum alone" properties of the dynamics. This will involve a discussion on eigenvectors and also on pseudo-spectra (interestingly, a collection of eigenspectra can solve some problems that one spectrum alone can't). Next, we will present some simple, yet very useful, fun-matrix-facts: some biologically inspired ideas related to the connectivity matrix **W** (like neuronal self-loops or neuronal gain modulation), that have important and straightforward effects on the eigenspectrum. Finally, we will end with some results from random matrix theory which, in theoretical neuroscience, are perhaps the most well known "**W**-to-spectrum" examples. Throughout the article we will focus on aspects of the eigenspectrum theory that are useful in the framework of recurrent neuronal networks. By the end of it, we will hopefully have convinced you that it should be impossible to think about neuronal systems without thinking of eigenspectra.

During the course of the article we will use some mathematical terms that might need further explanation. The formal definition of many of these terms will not be necessary in order to understand the text or logical argument surrounding them. For this reason, all such formal terms will be coloured in blue. Their formal mathematical definition will be given in the Appendix.

A simple model in neuroscience

As neuroscientists we want to understand the brain. The brain consists of billions of neurons; thought of as the fundamental building blocks of the brain's processing power. To truly understand neurons we need to build mathematical models representing and reproducing their behaviour [23]. For neurons, the Hodgkin–Huxley model is the perfect example of how useful this approach is [24].

But suppose that we are interested in modelling a network, rather than a single neuron. For this reason, we will simplify our notion of a neuron and consider neurons in our system as abstract units that respond identically to the same input. The behaviour of a neuron can be approximated by its membrane potential, which we denote by $x(t)$. We can further include an input-output nonlinear function g converting the membrane potential to firing rate. We know, that the evolution of the membrane potential over time follows an exponential decay,

i.e.,

$$\tau \frac{dx(t)}{dt} = -x(t), \tag{1}$$

namely that the membrane potential at time $t = T$ is $x(0)e^{-T/\tau}$, given the initial condition $x(0)$ [25]. Note here that τ is the neuronal time constant, affecting the timescale of the exponential decay. However, a neuron receives external time-dependant inputs $I(t)$, as well as inputs from the other neurons of the network. If the network we are modelling has n neurons, then the presynpatic input of neuron j to neuron i ($1 \leq i, j \leq n$) is equal to $w_{ij}g(x_j)$; where w_{ij} is the weight or strength of the connection from neuron j to neuron i. Note that this will be zero if there is no such connection. We can thus update the equation for the membrane potential to include those inputs:

$$\tau \frac{dx(t)}{dt} = -x(t) + \sum_{j=1}^{n} w_{ij} g(x_j(t)) + I(t). \tag{2}$$

However, to solve the above equation, i.e., to calculate $x(t)$ for any t, we need to know the values of the firing rates of all other neurons in the network, the $g(x_j)$'s inside the sum. Each such firing rate is a function of the membrane potential x_j which in turn can be modelled with the same equation. In other words, we must write (assuming for simplicity that all neurons have the same time constant):

$$\begin{aligned}
\tau \frac{dx_1(t)}{dt} &= -x_1(t) + \sum_{j=1}^{n} w_{1j} g(x_j(t)) + I_1(t) \\
\tau \frac{dx_2(t)}{dt} &= -x_2(t) + \sum_{j=1}^{n} w_{2j} g(x_j(t)) + I_2(t) \\
&\vdots \\
\tau \frac{dx_n(t)}{dt} &= -x_n(t) + \sum_{j=1}^{n} w_{nj} g(x_j(t)) + I_n(t)
\end{aligned} \tag{3}$$

All these equations need to be understood simultaneously, creating an interconnected and interdepended dynamical system. The first step towards understanding is to transition from numbers to vectors and matrices.

The linear algebra of neuroscience

The most natural way to express the dynamical system (3) from the previous section is this:

$$\tau \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = -\mathbf{x}(t) + \mathbf{W}\mathbf{g}(\mathbf{x}(t)) + \mathbf{I}(t). \tag{4}$$

Instead of looking at the n neurons as a set, we look at the vector space generated by those neurons, namely $\mathbb{R}^n$. If we had ordered the neurons of the network, then neuron 1 now corresponds to the vector $(1, 0, 0, \cdots, 0)$ in $\mathbb{R}^n$, and similarly neuron j corresponds to the vector $(0, \cdots, 1, \cdots, 0), \in \mathbb{R}^n$, with 1 in the j-th place and 0 everywhere else. We can interpret the values/coefficients of these vectors in $\mathbb{R}^n$ in different ways. They can, for example, represent membrane potentials. In the above equation, $\mathbf{x}(t)$ is now a vector representing the membrane potential of all the neurons. The matrix $\mathbf{W} \in \mathbb{R}^{n \times n}$ is the connectivity matrix between neurons. In this matrix, row i represents the connection weights from all neurons to neuron i. Similarly, columns correspond to the "output" weights. The vector $\mathbf{g}(\mathbf{x}(t))$ is typically applied element–wise, and represents the firing rates of neurons, where g is again the non-linear function approximating the input-output relation between the membrane potential and firing rate of neurons. Finally, $\mathbf{I}(t)$ is a time-dependant external input; but now it is a vector in which each coordinate corresponds to the input given to every neuron in the network.

This dynamical system represents the entire network, yet it is still very similar to the single neuron equation (2). For example, we can easily see that the network dynamics, without an external input or recurrent connections, would decay exponentially according to $\tau \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = -\mathbf{x}(t)$. Simple decay is not very interesting, but if we add the recurrent connections, i.e, $\mathbf{W}$, everything changes! Now the neurons of the network can interact with each other and as a population they are able to generate rich patterns of activity.

Let's think about $\mathbf{W}$ a bit more. Its entries represent connection strengths. At each time-step, $\mathbf{W}$ is multiplied with the vector of firing rates $\mathbf{g}(\mathbf{x})$. Thus $\mathbf{W}$ can be seen as a linear map $\mathbf{W} : \mathbb{R}^n \to \mathbb{R}^n$ taking the current firing rate as an input and producing, as an output, the effect of the recurrent connectivity on all neurons (Fig 1A left). In other words, $\mathbf{W}$ represents the soul of the system, and the network can remain interesting even if we make simplifying assumptions such as:

- $\mathbf{I}(t) = 0$, assuming there is no extra input to the network.
- $\mathbf{g}(\mathbf{x}) = \mathbf{x}$, now $\mathbf{g}$ is simply the identity function; this function is also used to approximate more complicated nonlinear functions such at $\tanh(x)$ which are approximately liner around $x = 0$ (or see also linearisation).
- $\tau = 1$, for simplicity. We can then temporally scale

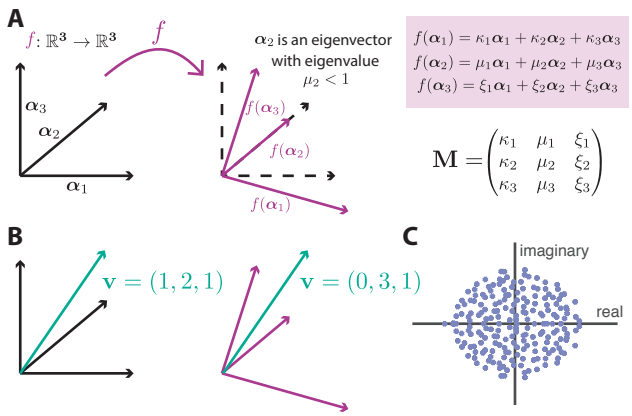

Figure 1. **Main concepts from linear algebra. A**, Left: Example of a linear map, $f : \mathbb{R}^3 \to \mathbb{R}^3$. The map can be defined by its value on three vectors $(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3)$, forming a basis of $\mathbb{R}^3$. The basis is mapped under f into three new vectors (purple). Using the two properties of a linear map (i.e., that (i) addition and (ii) scalar multiplication are preserved), we know how f acts on the entire space $\mathbb{R}^3$. Right: We can express the linear map, with respect to the above basis, as a matrix $\mathbf{M}$. In this example the basis vector $\boldsymbol{\alpha}_2$ happens to be an eigenvector, since its direction is preserved under f. Thus μ_1 and μ_3 are equal to zero, and the corresponding eigenvalue, μ_2, is less than 1 since the norm of $f(\boldsymbol{\alpha}_2)$ is smaller than the norm of $\boldsymbol{\alpha}_2$. **B**, An example of a change of basis. Here we use the map f from A. as a change of basis map, or change of coordinate system. The vector $\mathbf{v}$ is on the left represented with respect to the first basis (black vectors) and on the right with respect to the second basis (purple vectors). The coefficients for $\mathbf{v}$ are different in each case, highlighting the fact that a vector cannot be identified with any set of coefficients, unless a basis is provided for its representation. **C**, An example eigenspectrum of a 200×200 matrix, with entries taken from a uniform random distribution. Every dot represents an eigenvalue, with the horizontal axis indicating the real part and the vertical axis indicating the imaginary part.

the resulting dynamics later to achieve more realistic timescales.

Using these simplifying assumptions allows us to get at the essence of the network's behaviour and gain useful intuitions for more realistic cases. The dynamics now evolve as follows:

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = (\mathbf{W} - \mathbf{I}_n)\mathbf{x}(t), \tag{5}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Setting $\mathbf{M} = \mathbf{W} - \mathbf{I}_n$ we end up with

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{M}\mathbf{x}. \tag{6}$$

This equation looks very simple and compact, yet carries all the information needed to compute the behaviour of the linear neuronal network. This means we can determine each neuron's activity over time, given an initial firing rate and the connectivity structure of the network.

Spectral decomposition

From here we can solve Eq 6 in one of two ways: numerically (using computer simulations), or analytically (exactly using mathematical expressions). Here, we focus on analytical solutions so that we can best gain some intuitions and deep understanding of these systems. In practice, to determine the neurons' activities given Eq 6, we need to find the solution of the dynamical system, i.e., we need to be able to describe the system given any initial condition (i.e., the neural activity at $t = 0$). For example, if we were to solve a single differential equation,

$$\frac{dx}{dt} = kx, \tag{7}$$

we would obtain the well known exponential solution $x(t) = x(0)e^{kt}$, with $x(0)$ the initial condition [25]. In some way, the solution for the dynamical system is very similar. It is given by the matrix exponential, i.e, $\mathbf{x}(t) = e^{\mathbf{M}t}\mathbf{x}(0)$, where $\mathbf{x}(0)$ is the initial condition of the population of neurons, this time given by a vector, and the expression $e^{\mathbf{M}t}$ is a matrix, rather than a number. But by writing down the symbol of the matrix exponential, we haven't really solved anything. In fact, the matrix exponential for a matrix $\mathbf{A}$ is defined as

$$e^{\mathbf{A}} = \sum_{k=0}^{\infty} \frac{1}{k!}\mathbf{A}^k, \tag{8}$$

and involves computing all powers of the matrix, which at first glance looks computationally difficult. It turns out that this series is easily calculated when the matrix is a diagonal one.

We will thus show how to convert a general matrix $\mathbf{W}$ to a diagonal matrix. This is done by changing the coordinate system of $\mathbf{W}$ (Fig 1B). We mentioned earlier that we can view the weight matrix $\mathbf{W}$ as a linear map. This is true in general for all matrices: A matrix $\mathbf{M}$ can represent a linear map f from one vector space (e.g. $\mathbb{R}^n$) to another vector space (e.g. $\mathbb{R}^m$), with respect to a basis (a.k.a. coordinate system). We must thus pick a basis of the first space $\{a_1, a_2, \ldots, a_n\}$, and a basis for the second space $\{b_1, b_2, \ldots, b_m\}$. If $f(a_k) = \sum_{i=1}^{m} c_k^i b_i$, then the k-th column of the matrix is given by the coefficients $\{c_k^i\}_{i=1}^m$ (with the i indexed coefficient positioned on the i-th row of the matrix $\mathbf{M}$, Fig 1A right). In the case of the weight matrix $\mathbf{W}$, the basis for both spaces is generally the basis of neurons. But we can also change the coordinate system of $\mathbf{W}$ to a "better" one (we will get back to what "better" means later). To this end we have to use an invertible matrix $\mathbf{S}$, and consider the product $\mathbf{B} = \mathbf{S}^{-1}\mathbf{W}\mathbf{S}$. This new matrix $\mathbf{B}$ is a representation of the same map as $\mathbf{W}$, but with a different coordinate system, and is called **similar** to $\mathbf{W}$. We encounter such coordinate changes daily: If we think of the earth as a sphere in the three dimensional space, the natural basis pointing north, east and upwards in Madrid cannot be invariantly transferred to its antipode

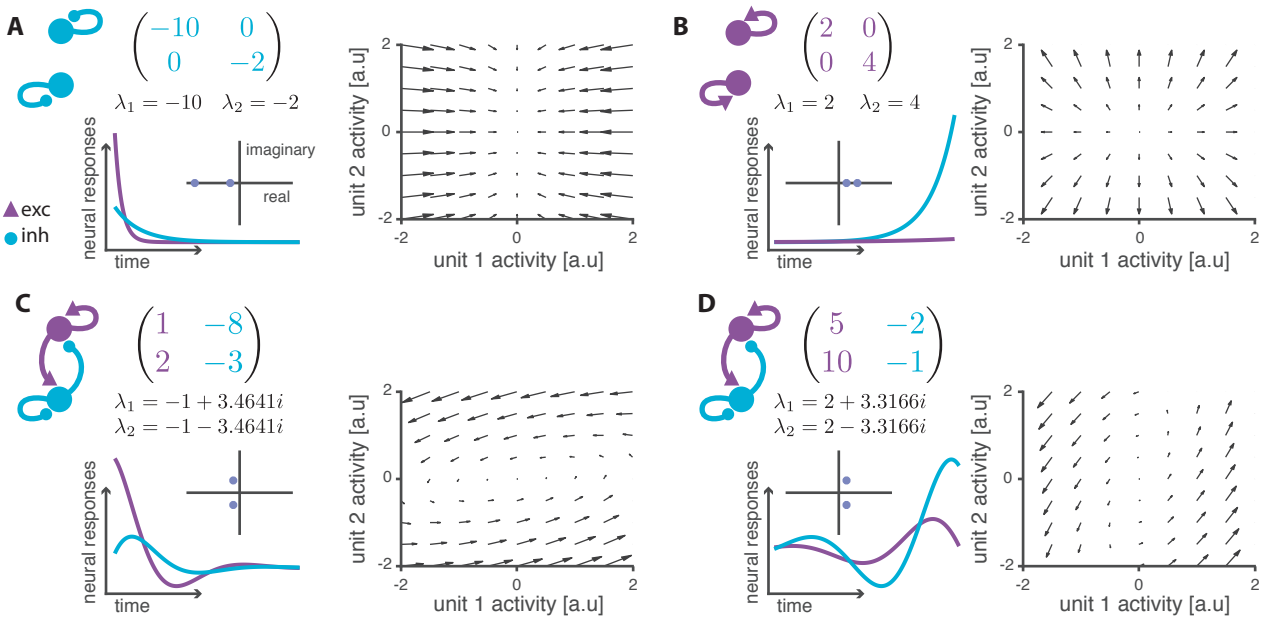

Figure 2. **Toy network examples.** **A–D**, Four examples of small neural dynamical systems, with a schematic of a network comprised of two units; the connectivity matrix; its eigenvalues λ_i, also plotted as the spectrum; the neural responses over time; and finally the vector field associated with the dynamical system. Purple and blue units are excitatory and inhibitory, respectively. **A**, A network with two uncoupled units; each unit is inhibiting itself. The matrix is diagonal and the eigenvalues are both real and negative. The dynamics here show exponential decay due to the negative real eigenvalues. The first unit decays faster because its real eigenvalue is smaller. The vector field identifies the $(0, 0)$ point as a stable fixed point. **B**, A network with two uncoupled excitatory units which excite themselves. The matrix is diagonal and the eigenvalues are both real and positive. The dynamics increase exponentially because the real eigenvalues are positive. The second unit grows faster since its eigenvalue is greater. The vector field identifies the $(0, 0)$ point as an unstable fixed point. **C**, A network with two coupled units. The matrix is fully connected and satisfies Dale's law, i.e., one unit is excitatory and the other is inhibitory. The inhibitory weights are larger in magnitude than the excitatory weights. The eigenvalues are complex, conjugate, and the real parts of the eigenvalues are negative. The real part determines whether the dynamics will decay or grow; the imaginary part determines the frequency of the oscillations. The vector field shows rotational dynamics which converge to the fixed point $(0, 0)$. **D**, A network with two coupled units that satisfy Dale's law. Here, the excitatory weights are larger in magnitude than the inhibitory weights, leading to eigenvalues with positive real parts and dynamics that grow in amplitude, making the system unstable. The vector field shows rotational dynamics diverging away from the (unstable) fixed point $(0, 0)$.

city Weber in New Zealand. In Weber, the arrows would no longer point towards the north, east and upwards directions. Instead, the upwards arrow would point directly towards the center of the earth, and the other two directions would also be rotated. To get to the new "better", i.e., more convenient, basis we need a change of coordinate system.

When the matrix we are interested in is a linear map from a space V to itself, $f : V \to V$, as is the case for $\mathbf{W}$, it's useful to spot some important features of this mapping. Most striking are the 1-dimensional invariant subspaces, i.e., the lines that stay intact after the matrix is applied to them. In the example of Madrid and Weber the lines spanned [1] by the arrows pointing straight up are invariant under the change of coordinate system, i.e., the line spanned by the upward vector (the one pointing towards the sky) in Madrid coincides with the line spanned by the upward vector in Weber, and only a sign change is needed to get from the Madrid to the Weber up–vector. In fact, every 1-dimensional subspace in a vector space is spanned by a unique vector, up to scalar multiplication. The special vectors that are invariant under a linear mapping are called eigenvectors. Eigenvectors go hand in hand with their corresponding eigenvalues. The fact that the lines are invariant, means that the only thing that the matrix can do to these vectors is change their (complex) amplitude and sign, a transformation that can be expressed by a single scalar. This scalar is the eigenvalue (Fig 1.A, vector α_2, eigenvalue μ_2).

Formally, we write this as follows: Let $\mathbf{M}$ be an $n \times n$ matrix. Let also $\mathbf{v} \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$. Then $\mathbf{v}$ is an **eigenvector** and λ is an **eigenvalue** if

$$\mathbf{M}\mathbf{v} = \lambda\mathbf{v}. \tag{9}$$

The set of all the eigenvalues is called the **eigenspectrum**, or **spectrum** (Fig 1C). It is worth noting here something that stands out in the formal definition. You may have noticed that the the eigenvalues and eigenvectors take

[1] i.e., the lines generated in the direction of the arrows

complex, rather than real values. Indeed, even though connectivity matrices of neuronal networks have only real values, the eigenvectors and eigenvalues can still be complex. However, the fact that $\mathbf{M}$ is real-valued means that the complex (or only imaginary) eigenvalues always come in conjugate pairs (reminder: the complex numbers $a + bi, a - bi$ are called conjugates). In the example of neurons and the vector space of firing rates (which in the linear case coincides with $\mathbf{x}$), a real vector $\mathbf{v} = (f_1, f_2, \ldots, f_n)$ denotes a specific firing pattern. If $\mathbf{v}$ is an eigenvector with real values, then applying $\mathbf{W}$ to this firing pattern $\mathbf{v}$ takes us back to $\mathbf{v}$, but also perhaps increases or decreases its amplitude. In other words, the relative activation of the neurons stays the same but the amplitude might be different. If the eigenvector is complex, the interpretation is not so clear but in both cases, eigenvectors help us identify subpopulations of neurons –modes– that are locked into each other, i.e., subpopulations that are always behaving consistently with each other. For the dynamical system, which describes the neuronal behaviour over time, the eigenvectors give us patterns that feedback to themselves, creating a loop in the population activity, which can either increase or diminish in amplitude and phase with time.

The eigenvectors can be used to change the coordinate system of $\mathbf{W}$ so that we can solve the dynamical system more easily. Here is something very important: Almost always [2] the eigenvectors of a matrix will be linearly independent (but not necessarily orthogonal, see below) and thus will form a basis of $\mathbb{C}^n$, i.e., we can write any vector as a linear combination of the eigenvectors. In matrix language, the fact that the eigenvectors form a basis means that if we denote by $\mathbf{V}$ the matrix whose columns are these eigenvectors, then $\mathbf{V}$ is invertible. Also, $\mathbf{V}^{-1}\mathbf{W}\mathbf{V} = \mathbf{\Lambda}$. The matrix $\mathbf{\Lambda}$ is diagonal and its entries are the eigenvalues corresponding to the eigenvectors (to see that this is the case, re-write this as $\mathbf{W}\mathbf{V} = \mathbf{\Lambda}\mathbf{V}$ and notice its parallel to the eigenvector–eigenvalue definition in Eq. 9). The fact that Λ is diagonal is very useful: powers, matrix exponentials and general matrix-computations become much easier when a matrix is given in a diagonal form.

This change–of–basis manipulation means that Equation 6 now has the following form:

$$\frac{\mathrm{d}\tilde{\mathbf{x}}}{\mathrm{d}t} = \mathbf{\Lambda}\tilde{\mathbf{x}}, \tag{10}$$

where $\mathbf{\Lambda}$ is diagonal containing the eigenvalues of $\mathbf{W}$ and $\tilde{\mathbf{x}}$ is a vector expressed with respect to the eigenvector basis, i.e., $\tilde{x} = \mathbf{V}x$. $\mathbf{\Lambda}$ is a diagonal matrix, and the matrix exponential of a diagonal matrix is easy to compute. Indeed, in this case the matrix exponential $e^{\mathbf{\Lambda}}$ is simply the diagonal matrix whose entries are the exponentials of the entries of $\mathbf{\Lambda}$, i.e., if λ_n is the $n-$th entry of $\mathbf{\Lambda}$, then e^{λ_n} is the $n-$th entry of $e^{\mathbf{\Lambda}}$.

Apart from the somewhat technical reason of having an easy–to–compute matrix exponential, another way to understand why this eigenvector basis simplifies things is to note that since $\mathbf{\Lambda}$ is a diagonal matrix, we now have n un-coupled equations to solve. This means that the solution of each differential equation in the dynamical system is now independent of the solutions of the other equations. Finally, the way to interpret the new variables $\tilde{x}_i$ is the following: Equation i represents the projection of the solution on the i-th eigenvector $\mathbf{v_i}$ and has the following form

$$\frac{d\tilde{x}_i}{dt} = \lambda_i \tilde{x}_i, \tag{11}$$

with solution: $\tilde{x}_i(t) = e^{\lambda_i t}\tilde{x}_{0,i}$, where $\tilde{x}_{0,i}$ is now the projection of the initial condition $\mathbf{x}(0)$ on the i-th eigenvector. In the example of neurons, this change of basis allows us to articulate the firing rate changes that any one population experiences *separately*. This system thus no longer represents dynamics in neuronal space, but in neuronal subgroups or mode space. Careful, though, because any one given neuron can be part of more than one mode.

If this is not neuronal space, why care?

The new coordinate system is not representing neuronal space. This is not a problem because we can always go back to neural coordinates and still use the eigenvector basis to express the solution of the system. What this means is that the solution in the initial coordinate system (of neurons!) will be a linear combination of the solutions (the $\tilde{x}_i$'s) we found above. More specifically:

$$\mathbf{x}(t) = \sum_{i=1}^{N} \tilde{x}_i(t)\mathbf{v_i}. \tag{12}$$

Therefore, it is useful to find the solutions in the coordinate system given by the eigenvectors because they can be computed and understood easily. They are just exponential functions: If the eigenvalue λ_i is real, then the solution of the i-th equation of the dynamical system is the usual exponential function (Fig 2A,B). If some of the λ_i are complex, then the situation is a little bit more complex, too. Complex eigenvalues could produce complex (i.e., 2 dimensional) solutions, even in neural space, in which *complex* firing rates are nonsensical. Fortunately, it can be shown that the real and imaginary parts of a complex solution must also satisfy the dynamical equations separately, giving rise to real solutions (Fig 2C,D).

Further, it can also be proved that real initial conditions give rise to real solutions; this follows from the conjugacy of complex eigenvalues and eigenvectors in

[2] formally: the set of matrices for which this is true is dense in the space of all matrices

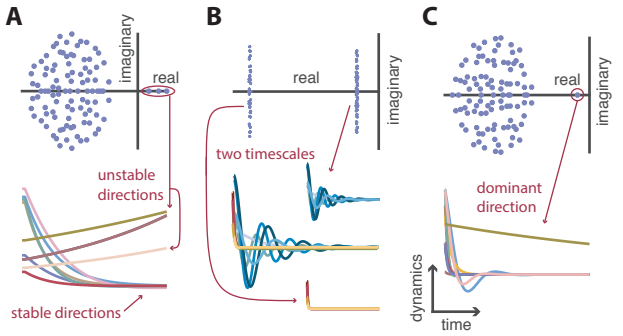

Figure 3. **What the spectrum can tell us.** Examples of dynamical features that can be extracted from the spectrum, in eigenvector coordinates. **A**, The number of eigenvalues that have real parts larger than zero (later one, see Eq.13) determine the number of unstable directions. If all eigenvalues have real parts less than zero (one) then the dynamical system is stable. **B**, Clustering of the eigenvalues, with respect to the real axis, creates a common timescale for many modes in the system. In this example, values cluster around two values and thus two timescales govern the dynamics. **C**, If one eigenvalue has real part much larger than the other eigenvalues, then this direction dominates the long term behaviour of the entire system.

real-valued matrices.For a more detailed exposition of all of the above (and many more things), see [26].

Following these arguments, one could assume that by obtaining the eigenvalues, we learn everything about the system. Indeed, the spectrum can tell us exactly what the solutions of the dynamical system look like in eigenvector coordinates. Moreover, the general solution in neuron space is a linear combination of the solution in mode space. All of the above assumptions are true but it is important to realise that the eigenvectors don't necessarily describe an orthogonal coordinate system. In other words, the spectrum alone does not provide any information about the interactions *between* the eigenvectors, i.e., how the dynamics are affected by non–orthogonal modes. So far, we have only identified the class of all similar matrices, i.e., all matrices $\mathbf{W}'$ that lead to the same diagonal matrix $\mathbf{\Lambda}$. What can be different between all those similar matrices $\mathbf{W}'$ is how the eigenvectors are oriented to each other. This information is lost in the eigenbasis coordinate system because by diagonalising a matrix $\mathbf{W}$, we implicitly treat its eigenvectors as the standard orthonormal basis. In the language of equation (12), having eigenvectors $\mathbf{v_i}$ that are not orthogonal to each other means that activity evolving along the direction of one eigenvector, $\mathbf{v_k}$, can affect the activity along the direction of another eigenvector, $\mathbf{v_j}$, if the two eigenvectors overlap. We will return to this important point in the following section, but let's focus first on what else we can learn from the spectrum alone.

What can we learn from the eigenspectrum.
Let's suppose that someone hands us a set of complex numbers and tells us "These are the eigenvalues of a matrix. Deal with it." What will we be able to say about the dynamics of the underlying system? Remember that the diagonal matrix $\mathbf{\Lambda}$ is similar to $\mathbf{M}$. It expresses dynamics with respect to the eigenbasis rather than neuronal basis. However, $\mathbf{\Lambda}$ is diagonal and hence expresses the dynamics of an uncoupled network. If the eigenvectors of $\mathbf{M}$ are not orthogonal to each other, then the dynamics expressed in the eigenbasis might be different from the neuronal dynamics. However, the spectrum, even on its own, allows us to glean some important qualities of the system:

Stability of the system: Stability of the dynamics depends on the real parts of the eigenvalues. In the general dynamical system of the form $\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{M}\mathbf{x}$, stability of the system is determined by whether the real parts of all eigenvalues are less than zero. If they are, the system is stable, i.e., the dynamics eventually return to zero after an initial perturbation. If one or more eigenvalues have real parts greater than zero, then the system is unstable and the activity of the system increases over time. In the eigenbasis the only directions that will increase are those associated with the greater-than-zero eigenvalues (Fig3A). However, in neuron coordinates, unstable eigenvalues indicate that the recurrent feedback between the neurons contributing to an eigen direction becomes larger than the natural decay of their activity, so the mode becomes 'self-sustained' (i.e., critical), or even amplifying (i.e., supercritical). If the system is not orthogonal, this activity will start to be distributed between different modes, and thus more than one modes will eventually become unstable. We would like to note here that in the case in which the dynamics are given as

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = (\mathbf{W} - \mathbf{I}_n)\mathbf{x}(t), \tag{13}$$

the system is stable if and only if all real parts of the eigenvalues of $\mathbf{W}$ are less than 1 (due to the identity matrix $\mathbf{I}_n$). We often call the identity matrix $\mathbf{I}_n$ that is subtracted from $\mathbf{W}$ the "leak term". We will formally prove the fact that the stability line moves to 1 in this case, in a later section.

The timescale of responses: The real parts of the eigenvalues provide the decay timescales along the directions of the eingenvectors. Even without knowing the eigenvectors, clustering of the eigenvalues (with respect to their real parts) can tell us that there are some directions that share similar timescales, increasing the chance for synchronised responses (Fig 3B) [27, 28].

Important directions for the dynamics: Through the spectrum we can identify the prevalent directions in the dynamics. For example, one eigenvalue with real part much larger than all others will indicate that the direction of its corresponding eigenvector dominates especially the asymptotic, long term behaviour (i.e. the behaviour that we observe when the variable t, describ-

ing time, becomes very large) of the system (Fig 3C). In eigenvector coordinates (or if the eigenvectors are orthogonal) this direction will decay much more slowly than the others following a perturbation. In neuron space, multiple neurons may share this direction in their dynamics, and thus show prolonged activity [29].

Transient behaviour

We already saw that by diagonalising a matrix we can solve the dynamical system and understand the dynamics in eigenvector coordinates. However the transformation that allows us to do this, i.e., the spectral decomposition, is not usually an orthogonal matrix. This means that the geometry of the eigenvectors of the initial matrix is not preserved. Such eigenvector geometry can be thought of as their relative position (and overlap) in space, which can be measured by their dot product.

Definition: A matrix $\mathbf{M}$ is called **normal** if it commutes with its conjugate transpose $\mathbf{M}^*$, i.e., $\mathbf{M}\mathbf{M}^* = \mathbf{M}^*\mathbf{M}$, or in other words, it is normal when the dot product of all its eigenvectors is zero. In particular if $\mathbf{M} \in \mathbb{R}^{n\times n}$, then if $\mathbf{M}$ is symmetric ($\mathbf{M} = \mathbf{M}^\top$), skew-symmetric ($\mathbf{M} = -\mathbf{M}^\top$) or orthogonal ($\mathbf{M}^\top = \mathbf{M}^{-1}$) then it is normal.

Normal matrices are relatively rare, e.g., neural networks are almost never normal and have some very special properties, e.g. a matrix $\mathbf{M}$ is normal if and only if it has an orthonormal basis of eigenvectors. We will call every matrix that is not normal **non-normal** and because most neural networks are non-normal we will dedicate a section to their behaviour.

The orthonormal eigenbasis characterization explains the point we made in the previous section, i.e., that if $\mathbf{M}$ is not normal then if we only look at the spectrum, and forget the geometry of the eigenvectors, we lose important information. This is exactly because the eigenvectors of a non-normal matrix are not orthogonal to each other, but overlap. These overlaps are very important for the dynamics, because they connect the different activity modes. To highlight their importance, we compare three matrices which have exactly the same spectrum but different degrees of eigenvector overlaps (Fig 4A). The asymptotic behaviour is the same in all three cases, the system is at its only stable fixed point as $t \to \infty$, time goes to infinity. However, the transient behaviour of the system, i.e., the dynamics for some amount of time after the initialisation, is very different. To see this explicitly we can revisit the solution of the dynamical system

$$\frac{\mathrm{d}\mathbf{x}(\mathbf{t})}{\mathrm{d}t} = \mathbf{M}\mathbf{x}(\mathbf{t}). \tag{14}$$

The vector $\mathbf{x}(\mathbf{t})$ is moving in neuron space over time, changing both direction and length (its norm) with the waxing and waning of neural firing rates. When we express this dynamical system as a linear combination of the eigenvectors of $\mathbf{M}$; we obtain (12), namely:

$$\mathbf{x}(t) = \sum_{i=1}^{N} \tilde{x}_i(t)\mathbf{v_i}. \tag{15}$$

It is not obvious by the above expression how the eigenvector overlaps — their dot products — affect $\mathbf{x}(t)$. An easy way to observe it is by calculating the norm of the solution of the dynamical system, which is a measure of the overall activity of the system at any given time. Since the coefficients $\tilde{x}_i$ and eigenvectors $\mathbf{v}_i$ in Eq.15 can be complex, in order to compute the norm of $\mathbf{x}(t)$ we need to consider the complex inner product, which for two complex numbers u, w is defined as $\langle u, w\rangle = \sum_{i=1}^{N} u_i\overline{w_i}$. For the solution of the dynamical system $\mathbf{x}(t)$ we thus have the following:

$$\|\mathbf{x}(t)\| =$$

$$\sqrt{\langle \mathbf{x}(t), \mathbf{x}(t)\rangle} =$$

$$\sqrt{\langle \sum_{i=1}^{N} \tilde{x}_i(t)\mathbf{v_i}, \sum_{i=1}^{N} \tilde{x}_i(t)\mathbf{v_i}\rangle} =$$

$$\sqrt{\sum_i |\tilde{x}_i(t)|^2 + \sum_{i\neq j} \overline{\tilde{x}_i}(t)\tilde{x}_j(t)\langle \mathbf{v}_i, \mathbf{v}_j\rangle}. \tag{16}$$

Note than in the above expression we have assumed, without loss of generality, that the norm of each eigenvector is equal to 1. The last line of Equation (16) proves that indeed the dot products of the eigenvectors, i.e., $\langle \mathbf{v}_i, \mathbf{v}_j\rangle$, affect the norm of the solution significantly [30–32].

Thus, since the spectral decomposition does not immediately provide us with information about the non-normality of the matrix, it is important to express the extent of eigenvector overlaps mathematically in a different way. Two powerful methods to quantify this overlap is to look at the "Schur decomposition" of $\mathbf{M}$, and at the condition number of the eigenvector matrix $\mathbf{V}$. In the next section, we will briefly discuss both approaches.

Measures of non-normality

A "non-normality" measure quantifies how far away a matrix $\mathbf{M}$ is from being normal, i.e., how far away it is from having its eigenvectors form an orthogonal basis. The Schur decomposition is a very good way to quantify this. Remember that non-normal matrices are *similar* (in the technical definition, see above) to their diagonal form, but the matrix providing the change of coordin-

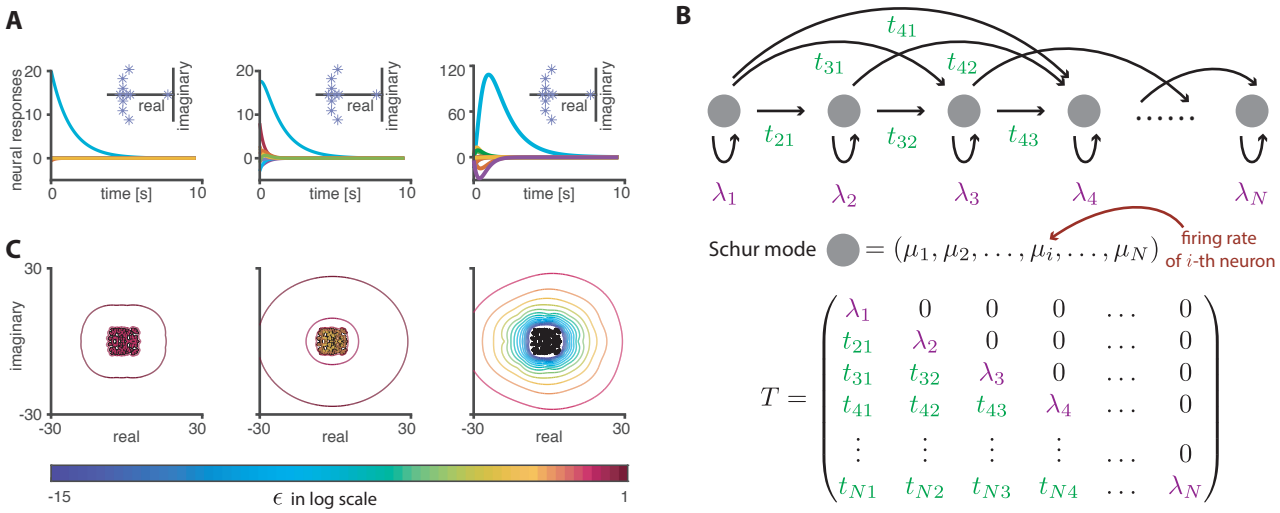

Figure 4. **What the spectrum can't tell us. A**, A system's transient behaviour cannot be deduced from the spectrum alone, e.g., for 3 matrices with identical spectra (shown in the inset). All three cases display the same asymptotic behaviour, returning to a stable fixed point at zero, but the transient behaviour can be very different. The dynamics resulting from a given initial condition [7] can show exponential decay (left), mild (middle) or strong (right) transient amplification depending on the non-normality index of each system. **B** A graphical description of the Schur decomposition which rotates the coordinate system such that the Schur modes, representing firing patterns, are connected in a feedforward manner yielding a lower triangular matrix. **C**, Three examples of matrices with identical spectra (of radius 5, shown in black, partially obscured), and different pseudospectra. The pseudospectra are computed for perturbations of norms ranging between $\epsilon = 10^{-15}$ and $\epsilon = 10$. A normal matrix (left) has very stable eigenvalues with the first detectable pseudospectrum at $\epsilon = 0.1$ (light pink). Even perturbations of norms up to $\epsilon = 10$ (dark red contour) produce spectra of radius ~ 15. In contrast, a matrix with a higher non-normality index (middle), shows a first detectable permutation already for $\epsilon = 10^{-3}$ (yellow) and perturbations of norms up to $\epsilon = 10$ result in much larger diameter spectra (dark red contour –radius ~ 30). A third matrix with very high non-normality index exhibits detectable spectra for perturbations as small as norm $\epsilon = 10^{-15}$ (dark blue contour). The last plottable eigenspectrum in the range of the axes is for $\epsilon = 0.1$ and it has a radius of ~ 30 (notice that this was the first detected perturbation in the left plot).

ates (in this case the eigenvector matrix $\mathbf{V}$) is not orthogonal. When this change of basis is not an orthogonal matrix, then the dynamics of the similar matrices can be very different. However for every matrix, there is an orthogonal change of coordinates, given by the Schur decomposition, that leads to a lower (or upper) triangular matrix, i.e., a matrix in which one of the two halves separated by the diagonal contains potentially nonzero entries (Fig 4B). In general for every matrix there are two possible Schur decompositions, the complex and the real. In the complex decomposition the entries of all matrices involved can be complex whereas in the real they are real. Here we will focus on the real Schur decomposition which is expressed formally as: $\mathbf{UMU}^\top = \mathbf{T}$, where $\mathbf{U}$ is an orthogonal matrix, $\mathbf{U}^\top = \mathbf{U}^{-1}$ is its transpose and $\mathbf{T}$ is a lower (or upper) triangular matrix. The columns of $\mathbf{U}$ are the so–called Schur vectors; they are the new basis under which the corresponding linear map can be represented by a triangular matrix, namely $\mathbf{T}$. This triangular matrix has its eigenvalues stored along the diagonal. Because these eigenvalues might be complex, in the real Schur decomposition, in which we want to restrict to real entries, we need to allow 2×2 blocks along the diagonal; in this case the diagonal entries carry the real parts of the eigenvalues and the off-diagonal entries of the 2×2 blocks carry the imaginary parts. The entries in the lower (or upper) triangular part express the coupling between the different modes, i.e. how much the activity in one mode will affect all others. As a consequence, the norm of this triangular part is a good measure for the distance of the matrix from normality. Normal matrices (in which there is no interaction between modes) have a norm of this part that is equal to zero. The Schur decomposition is not unique (intuitively, because we can express the interactions between modes in different ways), so we have to formally look at all possible Schur decompositions; the "non-normality" index is the minimum norm of the strictly lower (or upper) triangular part obtained like this ([33]).

The Schur decomposition has been studied in the context of neuronal networks before [30, 31, 34, 35], where the Schur vectors from $\mathbf{U}$ are interpreted as an abstract basis that corresponds to firing patterns of the network. Indeed, every Schur mode is a vector whose i–th coordinate represents the firing rate of neuron i; thus every Schur mode represents a specific neuronal activation. Therefore, by changing the basis of the matrix to the Schur basis, we can study the activity of the neuronal network in terms of these specific firing patterns. As we saw, the matrix connecting the firing patterns (or Schur modes) is lower (upper) triangular. This means that the network propagates each such firing pattern to the next ones in a

feedforward manner, given by the strictly lower (or upper) triangular part of the Schur matrix. This is different from the eigenvector description because firing patterns do not only feed back onto themselves but feed into a new activity pattern. The diagonal entries, which are also the eigenvalues of the matrix, represent the feedback loops on these firing patterns. In other words, the Schur transformation reveals an activity related feedforward structure that is hidden in the recurrent connections. From that point of view, it makes sense that a larger feedforward norm will lead to a larger transient response, since this norm corresponds to the strength with which a signal is propagated along the network through its hidden feedforward structure (Fig 4B).

Another way to quantify network interactions is to look at the condition number $\kappa(\mathbf{V})$ of the eigenvector matrix $\mathbf{V}$. It is defined as

$$\kappa(\mathbf{V}) = \|\mathbf{V}\|\|\mathbf{V}^{-1}\|, \tag{17}$$

where the norm of the matrix $\mathbf{V}$ is given by $\|\mathbf{V}\| = \sup\{\|\mathbf{V}x\| : x \in \mathbb{R}^n, \text{with} \|x\| = 1\}$ [3]. Intuitively, this norm gives us a measure of how much the matrix can "stretch" a unit vector. The condition number takes values between $1 \leq \kappa(\mathbf{V}) \leq \infty$ and the value $\kappa(\mathbf{V}) = 1$ is possible if and only if the matrix is normal. The idea behind the condition number is the following: if the norm of $\mathbf{V}^{-1}$ is "very big", then this means that the eigenvector matrix $\mathbf{V}$ is close to being singular; i.e., some of its eigenvectors are very close to each other (they have big overlaps). To arrive at a meaningful measure that allows us to compare different matrices, we normalise $\mathbf{V}^{-1}$ by considering its product with the norm of $\mathbf{V}$ (for those interested, there are more technical details, see Trefethen and Embree [36].)

The condition number of the matrix gives an upper bound for the condition number of the individual eigenvalues (this is known as the Bauer-Fike theorem, [37]), measuring how sensitive the eigenvalues are to perturbations of the matrix. Formally the theorem states that if μ is an eigenvalue of the perturbed matrix $\mathbf{M} + \mathbf{P}$, with $\|\mathbf{P}\| = \epsilon$ then there exists an eigenvalue λ of $\mathbf{M}$ such that $|\lambda - \mu| \leq \kappa(\mathbf{V})\epsilon$, where $\kappa(\mathbf{V})$ is the condition number of the matrix $\mathbf{M}$. From this relation we can see that a small condition number, which is a signature of normal matrices, can guarantee robustness of the eigenvalues to perturbations. On the other hand, small perturbations of non-normal matrices can lead to large changes in the eigenspectra.

[3] sup stands for "supremum" which in mathematics denotes the least upper bound of a set. You can think of this informally as the maximum element of a set; the only difference is that the supremum might not belong in the set, e.g. the supremum of the open interval $(0, 1)$ is 1.

One way to visualise the effect of all possible perturbations of a matrix simultaneously is to look at its pseudo-spectrum [20, 36]. The intuition here is again that in order to truly understand the "personality of a matrix" [36], we need to understand how sensitive a matrix is to perturbation. We know that perturbation behaviour is closely related to how much eigenvalues change (see above), which in turn, affects the shape of the eigenspectrum. The ϵ-pseudospectrum describes the expected range of changes to any given eigenspectrum as a result of a perturbation by ϵ. The non-normality of a matrix will amplify the cascade of changes from a single ϵ and thus nonnormal matices will have much more interesting and telling pseudospectra than normal ones. In other words, being normal is pretty boring (we knew this already).

Formally, the ϵ-pseudospectrum of a matrix $\mathbf{M}$ is the set of all points z_{pseudo} on the complex plane for which there exists a matrix $\mathbf{E}$ with $\|\mathbf{E}\| \leq \epsilon$ such that z_{pseudo} is an eigenvalue of the perturbed matrix $\mathbf{M} + \mathbf{E}$. In other words, for every ϵ, we collect all possible eigenspectra we can obtain by adding to $\mathbf{M}$ a matrix $\mathbf{E}$, with $\|\mathbf{E}\| \leq \epsilon$. This set on the complex plane is the ϵ-pseudospectrum. The deviations of the pseudospectrum from the original spectrum thus give an intuition of how non-normal the matrix is. The pseudospectra are a very powerful tool because they can capture in a single figure not only the eigenspectrum, but also the entire perturbation behaviour of the matrix, in which the non-normality is implicit. They are thus a much more accurate description of a matrix than the eigenspectrum alone (Fig 4C).

Choosing the Dub

Up to this point we have focused on understanding how some general properties of the connectivity matrix, like the real parts of the eigenvalues or the non-normality, are linked to important dynamical features. Unfortunately, in most cases, we don't know what real connectivity matrices of the brain look like, yet (but compare Li *et al.* [38] for a surely-soon-to-be analysed full connectome of the d. melanogaster mushroom body, and Motta *et al.* [39] for a partial reconstruction of a mouse somatosensory cortex volume), so a remaining question that remains is: "How do we choose the matrix $\mathbf{W}$ in theoretical neuroscience?".

This is an important question with various answers, which depend on the scope and the approach of each researcher. A discrimination for example arises when choosing between modelling functional or structural connectivity. Some models follow strict biological constraints [40, 41], some follow weak restrictions and some are based on statistical properties of the connections [42]. In other cases, the connectivity is learned, given

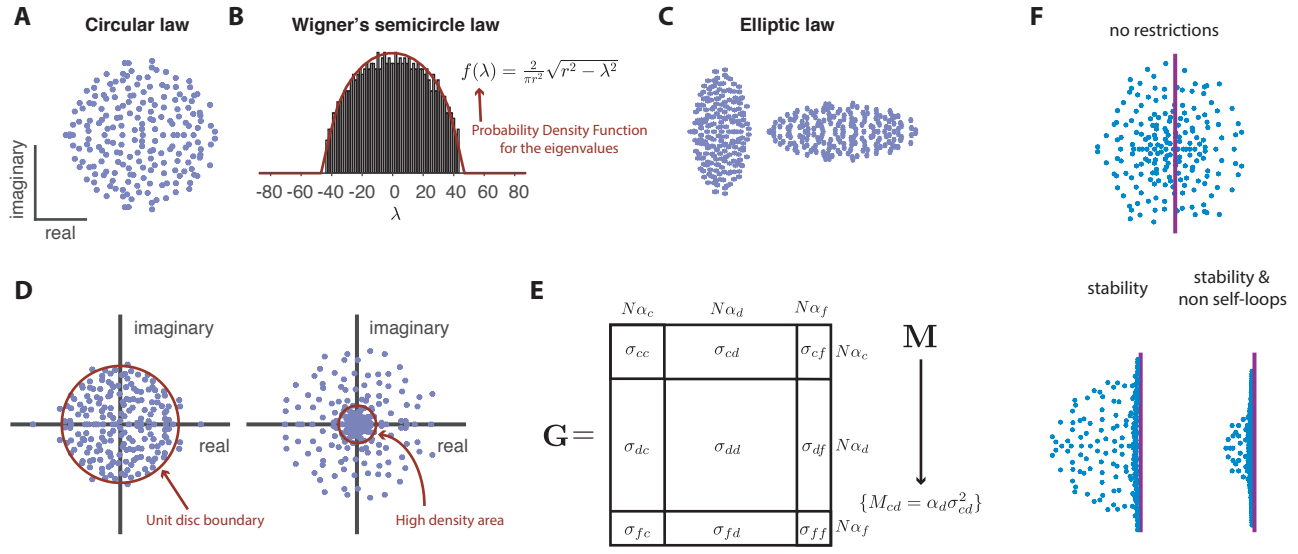

Figure 5. **Choosing W. A–C**, Eigenspectra of matrices with entries drawn from random distributions for different underlying matrix structures. **A**, Girko's circular law for random assymetric matrices, **B**, Wigner's semicircle law: the distribution of the eigenvalues of a symmetric matrix forms a semicircle, and **C**, The elliptic law for matrices with different degrees of correlations between their entries. **D**, Left: Spectrum of a random matrix of 100 excitatory and 100 inhibitory neurons. The two distributions have the same variance but different means, leading to "outliers", i.e., eigenvalues that are outside of the unit disc boundary. Right: Spectrum of a random matrix with two populations of neurons, drawn from normal distributions with zero mean and different variances. The spectrum is not uniformly distributed on a disc and there is a central high density area. **E**, The synaptic gain matrix **G**, representing the connectivity structure of a random matrix consisting of D different cell–types. Associated to **G** is a simplified matrix **M**, whose spectral abscissa determines the stability of the system. **F**, Three spectra obeying different constraints. Purple vertical line corresponds to the stability line ($x = 1$); if any eigenvalue is on the right of this line the induced dynamics are unstable. Top: a matrix with random connections and no restrictions. The eigenspectrum forms a disc. Bottom left: a matrix that is optimised to be stable, all the eigenvalues are pushed on the left of the stability line. Bottom right: a matrix that is stable and has no self–loops; the sum of the real parts must be equal to zero and therefore the eigenvalues are also pushed to the right constraining the real part of the spectrum to a limited range.

a specific functional objective through optimisation algorithms [7, 43], or shaped by plasticity rules [44].

The idea to use random connections for **W** based on statistical properties of the neuronal population is a common one in theoretical neuroscience and uses methods and tools from a field of mathematics called random matrix theory. The reason behind choosing random matrices is two-fold. First, if we model the entries of our matrix as random variables drawn from a distribution, then we know what the eigenspectrum will look like. Secondly, we can use the fact that all units of our networks are interchangeable, receiving statistically identical inputs (e.g. recurrent and/or external) to simplify the dynamical equation and understand the behaviour of the system; this is also known as "dynamical mean field analysis". The famous circular law [45, 46] is an example of what random matrix theory can contribute:

Theorem 1 (Circular law) *The spectral density of an $N \times N$ random matrix consisting of identically and independently distributed random variables with zero mean and variance $1/N$ converges, as N goes to infinity, to the uniform distribution on the unit disc.*

This result is quite surprising. In simple words it states that if we choose the entries of a matrix randomly (from any distribution with finite 1st and 2nd moments) we know what the spectrum will look like: the eigenvalues will be uniformly distributed on the unit disc (Fig 5A). This implies that sometimes the statistical properties of a system are sufficient to draw specific conclusions and that detailed knowledge of that system is not necessary.

Departing from the circular law, the matrices we are usually interested in carry extra structure. Some matrices in neuroscience might be symmetric as is the case in Hopfield networks. Symmetric matrices (with real entries) have only real eigenvalues and their probability density function always forms a semi–circle, a result known as the Wigner's semi-circle law [47, 48] (Fig 5B). In between the fully symmetric (Wigner's law) and fully asymmetric (circular law) cases, lie matrices which have some correlations between their entries. These matrices have been shown, again, to have known eigenspectra forming an ellipsoid shape [49] (Fig 5C).

Finally, matrices in theoretical neuroscience might follow specific connectivity motifs and profiles. For example, often matrices will satisfy Dale's law, i.e., neurons will either be excitatory or inhibitory,[50–52], which means that the left columns might have only positive elements, and the right columns might have only negative entries. We will review some of those cases in more

detail in the next section.

More than one neuronal population

Many times, researchers try to harvest the goods that random matrix theory has to offer, while maintaining crucial biological concepts. One such example is applying random matrix theory to networks with excitatory and inhibitory neurons [27]. In this case the entries of $\mathbf{W}$ are assumed to be drawn from two different distributions, since they represent different types of neurons. Because of this, we have in our hands two means and two variances. Typically these networks are ordered in columns in which fN number of columns represent the excitatory neurons of the network and the rest $(1-f)N$ columns the inhibitory (assume $0 < f < 1$). If we assume for simplicity that the two distributions have the same variance, $\sigma_E^2 = \sigma_I^2 = \frac{1}{N}$, but different means, $\frac{\mu_E}{\sqrt{N}} > 0$, $\frac{\mu_I}{\sqrt{N}} < 0$ (while the overall mean of the entries is still equal to zero) then the spectrum of the matrix turns out to have some outliers, i.e., some of the eigenvalues lie outside of the unit disc (Fig 5D left). This is a problem if we are interested in stable dynamics. We can't control the outliers because they are not deterministic, i.e., for every realisation of the random matrix the outliers will be in different positions. Importantly, some of them might have real values larger than one, surpassing the line of stability (note here that the line of stability is at one, rather than zero, because the dynamics have a leak term, see equation 13). However, we can stabilise the network if every row of the matrix individually sums to zero. Then the outliers vanish and the entire spectrum retreats inside the unit disc as in the case of Girko's law (see also [53]). This is very interesting because it can be linked to some ideas about excitatory-inhibitory balance. The initial assumption that the overall mean is equal to zero can be considered as a global balance between the excitatory and inhibitory weights. The second requirement, that each row sums to zero, resembles a more *detailed* balance [54] in which the input weights a neuron receives from all excitatory and inhibitory neurons in the networks sum to zero. It is interesting that this abstract form of detailed balance manages to keep the network stable.

Having two different neuronal populations could also be regarded as drawing the matrix entries from two distributions which now have the same mean (here taken to be zero) but different variances. Without loss of generality, we can assume that the fN excitatory neurons have variance $\sigma_E^2 = \frac{1}{N\alpha}$, while the $(1-f)N$ inhibitory neurons have variance $\sigma_I^2 = \frac{1}{N}$. Then the spectrum is distributed on a disc of radius equal to $\sqrt{(1-f)+f/\alpha}$, where f is the fraction of excitatory neurons. Importantly, the distribution is now not uniformly distributed on this disc, (Fig 5D right). Instead, the spectrum features a high density of values in its center. The boundary that limits this high density region is described by $\min(\sigma_E, \sigma_I)\sqrt{N}$. The inhomogenous spacing of eigenvalues can be important for the dynamics: the high density central area–depending on its radius–might have a synchronization effect on the dynamical responses because now neurons share eigenmodes whose decay times are more clustered around zero, and therefore are more similar.

The idea of constructing random matrices using more than one distribution can be extended further , e.g., to implement the assumption that different statistical properties in the entries of the matrix represent different cell types [55, 56]. If we assume that there are D different cell types, each one with a fraction α_d of neurons in it, then the connectivity matrix $\mathbf{W}$ has a block–like structure. We assume that the mean of $\mathbf{W}$ is equal to zero, $\mu(\mathbf{W}) = 0$ while the variance of each block depends on the input and output populations, and is equal to $\frac{\sigma_{cd}^2}{N}$, where c, d represent the input and output cell types. We can represent the connectivity structure using a synaptic gain matrix $\mathbf{G}$ with elements $G_{ij} = \sigma_{ij}$ arranged in D^2 blocks (Fig 5E). We can furthermore create a simplified matrix $\mathbf{M}$ by setting $M_{cd} = \alpha_d \sigma_{cd}^2$. This matrix represents the mean–population dynamics. Interestingly the stability of the system depends on the eigenspectrum of this simplified matrix $\mathbf{M}$ [55]. Indeed, the transition to instability happens exactly when the largest eigenvalue of $\mathbf{M}$ is 1 (here, again, the stability line is at 1 rather than 0 following the dynamical equation with the leak term 13). More specifically, if Λ_1 is the largest eigenvalue of $\mathbf{M}$, then the radius of the eigenspectrum of $\mathbf{W}$ is given by $\sqrt{\Lambda_1}$. This implies that even small populations of neurons can dramatically affect the stability of global dynamics.

Neuronal Self–loops

In addition to constraints regarding neuron identity, there are other connectivity constraints imposed by biology. A good example of how such constraints may affect network dynamics is that of self-loops [57]. While mathematically entirely plausible, synapses from neurons onto themselves are very rare, and it is thus sensible to set the diagonal of any $\mathbf{W}$ to 0. This very simple restriction changes the shape of the eigenspectrum in a non-trivial and surprising way. The sum of diagonal entries, i.e., the "trace" of a matrix is invariant under a change of basis. In other words the trace is a property of the linear map rather than merely a property of the basis. In fact, the trace is equal to the sum of the eigenvalues. In equations, we can write:

$$\mathrm{Tr}\,(\mathbf{W}) = \sum_{i=1}^{n} w_{ii} = \sum_{i=1}^{n} \lambda_i. \tag{18}$$

Now, the assumption that there are no self–loops in the connectivity means that the trace of any such matrix $\mathbf{W}$ must be equal to zero, which by equation 18 means that the sum of the eigenvalues of $\mathbf{W}$ is also zero. We can

write every eigenvalue as a sum of its real and imaginary parts:

$$0 = \mathrm{Tr}\,(\mathbf{W}) = \sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} \mathrm{Re}(\lambda_i) + \sum_{i=1}^{n} \mathrm{Im}(\lambda_i)\mathbf{i}. \quad (19)$$

Due to the conjugacy of the eigenvalues, we know that the sum of the imaginary parts is going to be equal to zero, i.e., $\sum_{i=1}^{n} \mathrm{Im}(\lambda_i)i = 0$. The same must be true for the sum of the real parts, i.e., $\sum_{i=1}^{n} \mathrm{Re}(\lambda_i) = 0$. This means that on top of the symmetry of the eigenspectrum with respect to the real axis, which comes from the eigenvalues' conjugacy, we also have another type of symmetry with respect to the imaginary axis. This is not symmetry in the strict sense of the term, but rather a "collective" type of symmetry imposing a balance between the sum of the positive and the sum of the negative (real part) terms.

We will see now how this simple constraint affects the eigenspectra of matrices which are stable (and note that here we are working with the linear equations). We know that stability of the system

$$\tau \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = -\mathbf{x}(t) + \mathbf{W}\mathbf{x}(t) \quad (20)$$

is equivalent to the spectral abscissa (i.e., the largest real part of the spectrum) being less than 1, $\alpha(\mathbf{W}) < 1$. This provides an upper bound for the real distribution of the spectrum. However, this bound together with the condition that $\sum_{i=1}^{n} \mathrm{Re}(\lambda_i) = 0$, means that there is a lower bound for the real distribution, too. In other words, if the real parts cannot be very big due to stability constraints, then they also can't be very small due to the zero-sum constraint. The result is that, since stability is determined by the bound at 1, the real distribution of such matrices $\mathbf{W}$ must be mainly concentrated around the $(-1, 1)$ interval, with the possibility of some negative outliers (Fig 5F).

Modulation

We have spent most of our time manifesting the importance of the connectivity matrix for the dynamics, but static connectivity is not the only thing that determines the dynamics. Neuromodulation can bring transient changes to the connectivity matrix, and consequently affect the dynamics, too. In some cases, the effects of neuromodulation can be expressed as changes in the connectivity matrix.

Formally, neuromodulation is modelled by the slope of the non-linear input output function g, translating membrane potential to firing rate, as a measure of how sensitive neurons are to their synaptic inputs. Previous experimental work has shown that the slope of this function, also referred to as the "gain", is not static but can change due to a variety of factors. For example, simultaneously increasing both excitatory and inhibitory background firing rates in a balanced manner decreases the gain modulation [58]. Additionally, depressing excitatory synaptic inputs can affect neuronal gain sensitivity [59]. The functional purpose of such gain modulation has been related to e.g. contrast invariance [60], attentional scaling [61], translation-invariant object recognition [62], auditory processing [63], the reorganisation of dynamical activity with fixed architecture [64].

Global modulation

Let us recall the main equation describing neuronal dynamics

$$\tau \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = -\mathbf{x}(t) + \mathbf{W}g(\mathbf{x}(t)). \quad (21)$$

We will begin with the case of "global", linear modulation. In this case, we are assuming that the entire network is modulated in exactly the same way, i.e., every neuron's gain function can be expressed as $g(x) = kx$, where k is the gain parameter. Therefore, we have the following:

$$\tau \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = -\mathbf{x}(t) + \mathbf{W}k\mathbf{x}(t) = -\mathbf{x}(t) + k\mathbf{W}\mathbf{x}(t). \quad (22)$$

In other words, global modulation is equivalent to multiplying every entry in the connectivity matrix $\mathbf{W}$ by k, i.e., equivalent to considering the matrix $k\mathbf{W}$ as our connectivity matrix. We can easily rewrite this as

$$\tau \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = (k\mathbf{W} - \mathbf{I_n})\mathbf{x}(t). \quad (23)$$

Naturally, it would be useful to know how such an operation affects the spectrum. For example, is the new system still stable or have the dynamics transitioned to an unstable regime? We know that for a system to be stable all eigenvalues need to have real parts less than zero. Here, the new effective connectivity matrix is not $\mathbf{W}$, but $\mathbf{M} = (k\mathbf{W} - \mathbf{I}_n)$. Is there a way to find the spectral abscissa α of $\mathbf{M}$ (i.e., $\alpha(\mathbf{M})$) as a function of $\alpha(\mathbf{W})$? Fortunately, there is a very useful tool, the spectral mapping theorem, which states the following:

Theorem 2 *Let* $\mathbf{M}$ *be a matrix from a finite dimensional space* V *to itself. Let also* p *be a polynomial. Then* $\sigma(p(\mathbf{M})) = p(\sigma(\mathbf{M}))$*, with* $\sigma(X)$ *denoting the spectrum of* X*.*

This theorem answers the above question. If $p(y) = gy - 1$ is the polynomial in question, then the theorem states that every eigenvalue λ_i in the spectrum of $\mathbf{W}$ is mapped to the eigenvalue $g\lambda_i - 1$ in the spectrum of $\mathbf{M}$. In terms of stability, if $\alpha(W)$ was the spectral abscissa of $\mathbf{W}$, then the modulated network is stable if and only if $g < 1/\alpha(W)$. Using this theorem we can see that gain modulation might not alter the *relative* shape of the eigenspectrum. However, it can dramatically alter the

stability, and hence, the dynamics of a network (Fig 6).

Neuron-specific modulation

What happens when we want to modulate each neuron in the network differently? Similarly to what we did above, this corresponds to multiplying every column of a matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$ with a different number g_j. Now the situation is far more complicated and the answer can't be deduced using the spectral mapping theorem. However, if we suppose that we start with a random matrix $\mathbf{W}$ with overall mean equal to 0, and variance σ^2/N then we can interpret the column–wise multiplication as a problem of different cell-types, and use the results from random matrix theory that we discussed earlier, [27, 55]. Indeed, multiplying each column with a scalar g_j is equivalent to assuming that the variance of this cell–type is g_j^2. In the original work, the variance of each block depended on both input and output populations. In this column–wise setting, the variance depends only on the output population. Importantly, within this framework, the radius of the new spectrum can be computed and is equal to the product of the initial radius with the mean synaptic gain, i.e., $g_{\text{new}} = \sigma\sqrt{\frac{1}{N}\sum_{k=1}^{N} g_j^2}$, [11, 65].

Nonlinear dynamics

s Before concluding, we would like to mention briefly the effects and importance of nonlinear dynamics with regard to the tools we introduced. We only considered simplified, linear dynamics as a approximation of highly complex, nonlinear neuronal network interactions. We followed this approach because the relation between the connectivity and the dynamics can only be gleaned in the linear regime. Any step towards nonlinear dynamics immediately results in multiple, often unsolvable complications, e.g., the eigenvalues can only be used to understand the network locally, near a fixed point, and often not reliably. However, biological neuronal networks are nonlinear, i.e., synapses are probabilistic, neurons spike, and network responses saturate. Linear rate networks are many abstractions away from the reality of the brain, but we believe that the connectome can still hold important information about the dynamics. We argue that a linear simplification serves as an good first step in the analysis of any system, especially the soon–expected experimental connectomes from various electron microscopy studies [38, 39]. The common practice of linearising the dynamics around fixed points to assess the behaviour of model systems using linear methods, as well as recent work converting firing to spiking networks and vice versa [66–68] lend further credence to the use of linear approaches as a first-pass method.

The most important consequence of a nonlinear system, e.g., a system with a nonlinear function like $\tanh(x)$, is the possibility of multiple fixed points. In the linear case this is not possible (as long as the connectivity matrix is invertible, which is the case for most $\mathbf{W}$). Multiple

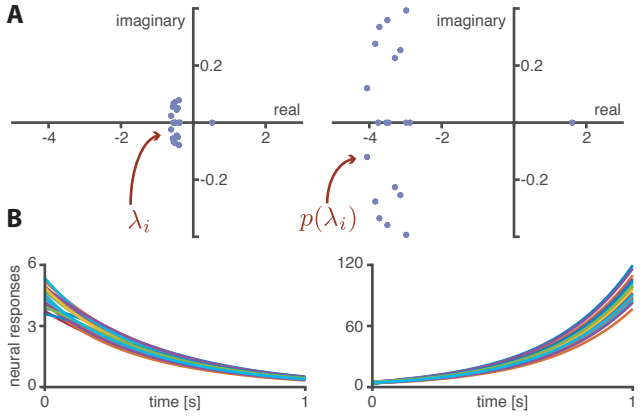

Figure 6. **Modulating the spectrum. A**, Example of modulating the spectrum by the polynomial $p(x) = 5x - 1$, according to the spectral mapping theorem. Every eigenvalue λ of the matrix M on the left has been mapped to the eigenvalue $p(\lambda)$ on the right. The *relative* shape of the spectrum is the same in both cases. **B**, The dynamics corresponding to the above eigenspectra. The spectrum on the right has crossed the stability line leading to unstable dynamics. Notice how all modes, and not only the unstable one, explode exponentially because the system is non-normal (which you can't see from the spectrum!).

fixed points make for a substantially richer palette of network responses; some initial conditions might drive the system to a certain fixed point, while a marginally different initialisation may converge to a different one [69, 70]. Moreover, unstable nonlinear dynamics can lead to the dynamical regime which in neuroscience is referred to as "chaotic". An instability now does not make the neural activity simply explode like in the linear case, because it is bound by the nonlinear function. The resulting regime is characterised by a strong mixing between neural activity modes that does not decay over time, [2].

It is interesting to know what the requirements for stability are, and how they relate to the slope of the nonlinear gain function. For example, in a network of N neurons following the dynamics

$$\tau\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = -\mathbf{x}(t) + \mathbf{W}\phi(\mathbf{x}(t)), \tag{24}$$

with $\phi(\mathbf{x}) = \tanh(g\mathbf{x})$, the non-linear input-output function with gain g and entries in $\mathbf{W}$ drawn from a Gaussian distribution (of zero mean and variance σ^2/N), the transition to chaos happens when $g\sigma = 1$, [71]. This is a nice parallel to the simplified linear system which leads to the same stability condition, $g\sigma < 1$, indicating that the linear approximation in this case can be still helpful.

Conclusion

In this article we studied the relationship between the architecture of a linear network and its dynamics, and how this relationship manifests itself through the eigenspectrum of the connectivity matrix $\mathbf{W}$. We used the language of dynamical systems in a quest to express essential, bare-bone truths of hugely complex systems

like the living brain. For example, we showed how to use perturbation theory as a means to tackle the non–normal (biologically plausible) part of a connectivity matrix and extract information beyond the spectrum itself, like the network's ability to transiently amplify certain inputs. We also showed how random matrix theory can help us deduce the shape of the eigenspectrum, along with information on stability or potential synchronisation effects of the network dynamics. We demonstrated how dynamical systems analysis can capture non-connectivity-related changes such as contextual modulation, and discussed how recent theoretical results fit into and extend the above framework. In the supplementary materials we provided a hands–on tutorial for the most important methods we introduced, allowing interested readers to explore several interesting examples.

After having familiarised themselves with the techniques mentioned here, readers should be able to peruse and enjoy a plethora of research papers on related topics. To mention just a few: Mastrogiuseppe and Ostojic [11] and Schuessler *et al.* [72] explore the dynamics of a network created as a sum of a structured low–rank matrix and a random matrix, showing how in this case there is indeed a strong link between the (low–rank) structure and the activity of a network.

Kuczala and Sharpee [73] quantify the effects of plasticity rules by studying correlated random matrices, and thus provide a way to assess the stability of more biologically plausible networks.

Aljadeff *et al.* [74] develop the modulation related results further by introducing a continuous synapse-specific gain function (depending, e.g., on the neurons' location in anatomical or functional space) and computing the eigenspectrum of the new, continuously modulated network.

Stern *et al.* [75] analyse the dynamics of a network comprised of clusters with strong local connectivity and random between–cluster connections and find the possible dynamical regimes that can arise, some of which combine features of both chaotic and multiple–fixed–point attractors.

The power of algebraic approaches for neuroscience has thus far only been shown in toy examples and simulated networks, but the coming years promise to deliver a treasure of connectome data for which we still lack sophisticated methods of analysis. Some of the techniques introduced here are borrowed from linear algebra and have been around for decades, but their usefulness holds exciting promises of holistic approaches to connectome analysis for the years to come.

Appendix Some of the main definitions from linear algebra that were used, [76–79]:

Linear map: The Yoneda Lemma in mathematics informally states the following: " *In order to understand an object, it is enough to understand its relation to all the other objects" [80].* This is exactly why functions (e.g. continuous, differentiable, algebraic etc) are used in all fields of mathematics. In linear algebra, the objects of interest are vector spaces and the appropriate maps between them are the linear maps. More specifically, let $f : V \to W$, be a function between V and W where V, W are vector spaces over some field $\mathbb{K}$. For our purposes, the field will be taken to be the set of real numbers $\mathbb{R}$, or complex numbers $\mathbb{C}$, and vector spaces V, W will always be $\mathbb{R}^n$, or $\mathbb{C}^n$, for some dimension n. The function f is a **linear map**, if $f(\alpha + \beta) = f(\alpha) + f(\beta)$, and $f(\lambda\alpha) = \lambda f(\alpha)$, for every $\alpha, \beta \in V$, $\lambda \in \mathbb{K}$ (Fig 1A left).
Basis: A basis $\mathcal{B}$ of a (finitely dimensional) vector space $\mathbf{V}$, assume for simplicity that $\mathbf{V} = \mathbb{R}^n$, is a set $\{b_1, b_2, \ldots, b_n\}$ such that:

- If $\kappa_1 b_1 + \kappa_2 b_2 + \cdots + \kappa_n b_n = 0$ for some $\kappa_1, \kappa_2, \ldots, \kappa_n \in \mathbb{R}$, then $\kappa_1 = \kappa_2 = \cdots = \kappa_n = 0$, and
- For every $\mathbf{v} \in \mathbb{R}^n$, we can find $\mu_1, \mu_2, \ldots, \mu_n$ in $\mathbb{R}$ such that $\mathbf{v} = \mu_1 b_1 + \mu_2 b_2 + \cdots + \mu_n b_n$.

Invertible matrix: A $n \times n$ square matrix $\mathbf{A}$ is invertible if there exists an $n \times n$ matrix $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I_n}$, where $\mathbf{I_n}$ is the identity matrix. The matrix $\mathbf{B}$ is often called the inverse of $\mathbf{A}$ and is denoted by $\mathbf{A}^{-1}$.
Linearization: A linear approximation of a nonlinear system that is valid in a small region around a fixed point. Formally, in order to do this, we have to change the variables of the dynamical system, so that they are now centered around the fixed point. The linearized model in term of the new variables is valid when their values are small. If the real parts of all eigenvalues of the linearized system are nonzero, then the behavior of the system near the fixed point is qualitatively the same as the behavior of the linear approximation.
Matrix exponential: Let $\mathbf{X}$ be an $n \times n$ matrix. The matrix exponential of $\mathbf{X}$, denoted by $e^{\mathbf{X}}$, is given by

$$e^{\mathbf{X}} = \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{X}^k,$$

where $\mathbf{X}^0$ is equal to $\mathbf{I}_n$, i.e., the identity matrix of dimension n.
Diagonal matrix: A (square) matrix $\mathbf{M}$ is diagonal if all elements that are not on the diagonal are zero. If m_{ij} denotes the matrix element on the i-th row and j-th column, then the diagonal entries are those for which $i = j$, i.e., denoted as m_{ii}.
Orthogonal transformation: An orthogonal transformation is a linear map $\mathbf{T} : V \to V$ on a real inner product space V, that preserves the inner product. That is, for each pair $\mathbf{u}, \mathbf{v}$ of elements of V, we have

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{Tu}, \mathbf{Tv} \rangle.$$

Since the lengths of vectors and the angles between them are defined through the inner product, orthogonal transformations preserve lengths of vectors and angles between them. In particular, orthogonal transformations map orthonormal bases to orthonormal bases.

Schur decomposition: For any square matrix $\mathbf{M}$ in $\mathbb{R}^{n\times n}$ there are two possible Schur decompositions which can provide a similarity transformation $\mathbf{U}$ of $\mathbf{M}$, to a triangular matrix $\mathbf{T}$:

- The complex Schur decomposition, in which $\mathbf{U}$ and $\mathbf{T}$ are complex. In this case $\mathbf{U}$ is unitary (i.e. $\mathbf{U}^\dagger\mathbf{U} = \mathbf{U}\mathbf{U}^\dagger = \mathbf{I_n}$) and the complex eigenvalues are the diagonal elements of the lower (or upper) triangular matrix $\mathbf{T}$.
- The real Schur decomposition, in which $\mathbf{U}$ and $\mathbf{T}$ are real. In this case $\mathbf{U}$ is orthogonal. In this case, since some eigenvalues might be complex, in order to "store" them in the real–valued triangular matrix $\mathbf{T}$ we need to allow 2×2 blocks along the diagonal. Then the diagonal entries carry the real parts of the eigenvalues and the off-diagonal entries of the 2×2 blocks carry the imaginary parts.

Software and code availability

All simulations were run in Matlab. Example code is available in the Supplementary Methods.

Acknowledgements

We thank Juan Sebastián Calderón García, Larry Abbott, Jake Stroud, Nicoletta Condruz, and James Ferguson for their valuable feedback at different stages of the manuscript. This work was supported by a Wellcome Trust Senior Research Fellowship (214316/Z/18/Z; GC, TPV).

Author contributions

G.C. and T.P.V. wrote the manuscript. G.C. performed all simulations and designed the tutorial code.

[1] T. P. Vogels, K. Rajan, and L. F. Abbott, Annu. Rev. Neurosci. **28**, 357 (2005).
[2] K. Rajan, L. Abbott, and H. Sompolinsky, Physical Review E **82**, 011903 (2010).
[3] K. V. Shenoy, M. T. Kaufman, M. Sahani, and M. M. Churchland, in *Progress in brain research*, Vol. 192 (Elsevier, 2011) pp. 33–58.
[4] K. V. Shenoy, M. Sahani, and M. M. Churchland, Annual review of neuroscience **36** (2013).
[5] R. Laje and D. V. Buonomano, Nature neuroscience **16**, 925 (2013).
[6] V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome, Nature **503**, 78 (2013).
[7] G. Hennequin, T. Vogels, and W. Gerstner, Neuron **82**, 1394 (2014).
[8] D. B. Rubin, S. D. Van Hooser, and K. D. Miller, Neuron **85**, 402 (2015).
[9] S. Denève, A. Alemi, and R. Bourdoukan, Neuron **94**, 969 (2017).
[10] G. Hennequin, Y. Ahmadian, D. B. Rubin, M. Lengyel, and K. D. Miller, Neuron **98**, 846 (2018).
[11] F. Mastrogiuseppe and S. Ostojic, Neuron **99**, 609 (2018).
[12] N. Davey, B. Christianson, and R. Adams, in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, Vol. 1 (IEEE, 2004) pp. 177–182.
[13] N. Bertschinger and T. Natschläger, Neural computation **16**, 1413 (2004).
[14] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter, PLoS Comput Biol **7**, e1002059 (2011).
[15] O. Sporns, Frontiers in computational neuroscience **5**, 5 (2011).
[16] A. Roxin, Frontiers in computational neuroscience **5**, 8 (2011).
[17] I. Vlachos, A. Aertsen, and A. Kumar, PLoS Comput Biol **8**, e1002311 (2012).
[18] O. Barak, D. Sussillo, R. Romo, M. Tsodyks, and L. Abbott, Progress in neurobiology **103**, 214 (2013).
[19] B. Doiron and A. Litwin-Kumar, Frontiers in computational neuroscience **8**, 56 (2014).
[20] Y. Ahmadian, F. Fumarola, and K. D. Miller, Physical Review E **91**, 012820 (2015).
[21] A. Litwin-Kumar, K. D. Harris, R. Axel, H. Sompolinsky, and L. Abbott, Neuron **93**, 1153 (2017).
[22] D. Martí, N. Brunel, and S. Ostojic, Physical Review E **97**, 062314 (2018).
[23] L. F. Abbott, Neuron **60**, 489 (2008).
[24] A. L. Hodgkin and A. F. Huxley, The Journal of physiology **117**, 500 (1952).
[25] G. Teschl, *Ordinary differential equations and dynamical systems*, Vol. 140 (American Mathematical Soc., 2012).
[26] V. I. Arnol'd, *Ordinary Differential Equations* (Springer-Verlag, Berlin, Heidelberg New York, 1991).
[27] K. Rajan and L. Abbott, Physical review letters **97**, 188104 (2006).
[28] R. Chaudhuri, A. Bernacchia, and X.-J. Wang, elife **3**, e01239 (2014).
[29] M. T. Schaub, Y. N. Billeh, C. A. Anastassiou, C. Koch, and M. Barahona, PLoS computational biology **11** (2015).
[30] B. K. Murphy and K. D. Miller, Neuron **61**, 635 (2009).
[31] G. Hennequin, T. P. Vogels, and W. Gerstner, Physical Review E **86**, 011909 (2012).
[32] G. Bondanelli and S. Ostojic, PLoS computational biology **16**, e1007655 (2020).
[33] L. Elsner and M. Paardekooper, Linear Algebra and its Applications **92** (1987).
[34] M. S. Goldman, Neuron **61**, 621 (2009).
[35] S. Ganguli, D. Huh, and H. Sompolinsky, Proceedings of the National Academy of Sciences **105**, 18970 (2008).
[36] L. N. Trefethen and M. Embree, *Spectra and pseudospectra: the behavior of nonnormal matrices and operators* (Princeton University Press, 2005).
[37] F. L. Bauer and C. T. Fike, Numerische Mathematik **2**, 137 (1960).
[38] F. Li, J. W. Lindsey, E. C. Marin, N. Otto, M. Dreher, G. Dempsey, I. Stark, A. S. Bates, M. W. Pleijzier, P. Schlegel, *et al.*, Elife **9**, e62576 (2020).
[39] A. Motta, M. Berning, K. M. Boergens, B. Staffler, M. Beining, S. Loomba, P. Hennig, H. Wissler, and M. Helmstaedter, Science **366** (2019).
[40] O. Shefi, I. Golding, R. Segev, E. Ben-Jacob, and A. Ayali, Physical Review E **66**, 021905 (2002).

[41] L. Hertäg and H. Sprekeler, PLoS computational biology **15**, e1006999 (2019).
[42] C. Tomm, M. Avermann, C. Petersen, W. Gerstner, and T. P. Vogels, Journal of neurophysiology **112**, 1801 (2014).
[43] D. Sussillo and L. F. Abbott, Neuron **63**, 544 (2009).
[44] E. J. Agnes, A. I. Luppi, and T. P. Vogels, Journal of Neuroscience **40**, 9634 (2020).
[45] V. L. Girko, Theory of Probability & Its Applications **29**, 694 (1985).
[46] T. Tao and V. Vu, Communications in Contemporary Mathematics **10**, 261 (2008).
[47] E. P. Wigner, Annals of Mathematics , 548 (1955).
[48] E. P. Wigner, Annals of Mathematics , 325 (1958).
[49] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, Physical review letters **60**, 1895 (1988).
[50] N. C. Spitzer, Neuron **86**, 1131 (2015).
[51] J. Zimmermann, M. Herman, and C. Rosenmund, Frontiers in synaptic neuroscience **7**, 16 (2015).
[52] S. Kim, M. L. Wallace, M. El-Rifai, A. R. Knudsen, and B. L. Sabatini, bioRxiv (2021).
[53] T. Tao, Probability Theory and Related Fields **155**, 231 (2013).
[54] G. Hennequin, E. J. Agnes, and T. P. Vogels, Annual review of neuroscience **40**, 557 (2017).
[55] J. Aljadeff, M. Stern, and T. Sharpee, Physical review letters **114**, 088101 (2015).
[56] J. Aljadeff, D. Renfrew, and M. Stern, Journal of Mathematical Physics **56**, 103502 (2015).
[57] G. Christodoulou, T. P. Vogels, and E. J. Agnes, bioRxiv (2021).
[58] F. S. Chance, L. F. Abbott, and A. D. Reyes, Neuron **35**, 773 (2002).
[59] J. S. Rothman, L. Cathala, V. Steuber, and R. A. Silver, Nature **457**, 1015 (2009).
[60] J. S. Anderson, I. Lampl, D. C. Gillespie, and D. Ferster, Science **290**, 1968 (2000).
[61] S. Treue and J. C. M. Trujillo, Nature **399**, 575 (1999).
[62] M. J. Tovee, E. T. Rolls, and P. Azzopardi, Journal of neurophysiology **72**, 1049 (1994).
[63] N. J. Ingham and D. McAlpine, Journal of Neuroscience **25**, 6187 (2005).
[64] J. P. Stroud, M. A. Porter, G. Hennequin, and T. P. Vogels, Nature neuroscience **21**, 1774 (2018).
[65] O. Harish and D. Hansel, PLoS computational biology **11**, e1004266 (2015).
[66] O. Shriki, D. Hansel, and H. Sompolinsky, Neural computation **15**, 1809 (2003).
[67] S. Ostojic and N. Brunel, PLoS Comput Biol **7**, e1001056 (2011).
[68] E. S. Schaffer, S. Ostojic, and L. F. Abbott, PLoS Comput Biol **9**, e1003301 (2013).
[69] S. H. Strogatz, *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering* (CRC press, 2018).
[70] J. J. Hopfield, Proceedings of the national academy of sciences **79**, 2554 (1982).
[71] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, Physical review letters **61**, 259 (1988).
[72] F. Schuessler, A. Dubreuil, F. Mastrogiuseppe, S. Ostojic, and O. Barak, Physical Review Research **2**, 013111 (2020).
[73] A. Kuczala and T. O. Sharpee, Physical Review E **94**, 050101 (2016).
[74] J. Aljadeff, D. Renfrew, M. Vegué, and T. O. Sharpee, Physical Review E **93**, 022302 (2016).
[75] M. Stern, H. Sompolinsky, and L. Abbott, Physical Review E **90**, 062710 (2014).
[76] I. N. Herstein and D. J. Winter, *Matrix theory and linear algebra* (Macmillan Publishing Co., Inc., 1988).
[77] G. Strang, G. Strang, G. Strang, and G. Strang, *Introduction to linear algebra*, Vol. 3 (Wellesley-Cambridge Press Wellesley, MA, 1993).
[78] J. Hefferon, "Linear algebra:(textbook)," (2009).
[79] S. Lang, *Introduction to linear algebra* (Springer Science & Business Media, 2012).
[80] S. Mac Lane, *Categories for the working mathematician*, Vol. 5 (Springer Science & Business Media, 2013).