Generation of scale-invariant sequential activity in linear recurrent networks

Yue Liu[1,2] and Marc W. Howard[1,2,3]

[1] *Department of Physics, Boston University, Boston, MA 02215*
[2] *Center for Systems Neuroscience, Boston University, Boston, MA 02215*
[3] *Department of Psychological and Brain Sciences, Boston University, Boston, MA 02215*

Abstract

Sequential neural activity has been observed in many parts of the brain and has been proposed as a neural mechanism for memory. The natural world expresses temporal relationships at a wide range of scales. Because we cannot know the relevant scales *a priori* it is desirable that memory, and thus the generated sequences, are scale-invariant. Although recurrent neural network models have been proposed as a mechanism for generating sequences, the requirements for *scale-invariant* sequences are not known. This paper reports the constraints that enable a linear recurrent neural network model to generate scale-invariant sequential activity. A straightforward eigendecomposition analysis results in two independent conditions that are required for scale-invariance for connectivity matrices with real, distinct eigenvalues. First, the eigenvalues of the network must be geometrically spaced. Second, the eigenvectors must be related to one another *via* translation. These constraints are easily generalizable for matrices that have complex and distinct eigenvalues. Analogous albeit less compact constraints hold for matrices with degenerate eigenvalues. These constraints, along with considerations on initial conditions, provide a general recipe to build linear recurrent neural networks that support scale-invariant sequential activity.

1 Introduction

The current state of the brain can carry memory for the past *via* history-dependent dynamics. This memory can be used to adaptively shape behavior to anticipate the future. However, the natural world has temporal relationships on a wide range of timescales (e.g. Voss and Clarke, 1975). This presents a problem in the design of the history-dependent dynamics of the brain. The world can contain behaviorally relevant predictive information over a range of scales, but we do not necessarily know the relevant timescales *a priori*. Suppose the brain's dynamics had a single characteristic scale s_o. If the world contains useful information at a scale much longer than s_o, this information would be invisible to the brain's dynamics and the system would not be able to exploit this information. Similarly, if the world contains useful information at a scale much shorter than s_o, the brain's dynamics would not represent

this information efficiently. One solution is a dynamic representation of the world that is scale-invariant across time (Howard and Shankar, 2018). Indeed, a spectrum of timescales is an essential ingredient in neural circuit models for temporal pattern recognition (Tank and Hopfield, 1987; Hopfield and Brody, 2000; Buonomano and Maass, 2009; Gütig and Sompolinsky, 2009).

There is empirical evidence suggesting that the brain in fact implements something like scale-invariance. Decades of research in cognitive psychology demonstrate that human timing and memory behavior exhibit the same properties on a wide range of timescales (Murdock, 1962; Glenberg et al., 1980; Rakitin et al., 1998; Howard et al., 2008). Sequential neural activity has been observed in many areas of the brain and is thought to have important cognitive functions in memory and decision making (MacDonald et al., 2011; Harvey et al., 2012; Howard, 2018). In light of these considerations, recently it has been proposed that "scale-invariance" is a desirable property for neural sequences (Shankar and Howard, 2012, 2013). In a scale-invariant neural sequence, the cells that are activated later have wider temporal receptive fields. More specifically, the responses of different cells have identical time courses when they are rescaled in time by their peaks. The hypothesis of scale-invariant neural sequences for time is consistent with recent electrophysiological recordings of "time cells" during a delay period when animals are performing various cognitive tasks (Pastalkova et al., 2008; Jin et al., 2009; MacDonald et al., 2011; Kraus et al., 2013; Mello et al., 2015; Salz et al., 2016; Tiganj et al., 2018). The firing fields of time cells that fire later in the delay period are wider than the firing fields of time cells that fire earlier in the delay period.

Many researchers have studied recurrent neural networks that generate sequential activity (Goldman, 2009; Rajan et al., 2016; Wang et al., 2018), but not many of these works considered scale-invariant sequences (but see Voelker and Eliasmith, 2018). In this paper we seek to identify general constraints on the network connectivity for the generation of scale-invariant neural sequences in recurrent neural networks. We study a linear network of interacting neurons. The f-I curve of many neurons are observed to be largely linear (e.g. Chance et al., 2002). The learning dynamics of linear feedforward neural networks exhibit many similarities compared to their non-linear counterparts (Saxe et al., 2013). Therefore linear neural networks provide a good model for studying systems-level properties of real neuronal circuits.

The paper is organized as follows. In Section 2 the network constraints for the generation of scale-invariant neural sequences are derived analytically. The exact constraints hold when the connectivity matrix has real, distinct eigenvalues. The modifications to generalize them to complex and distinct eigenvalues are discussed. Networks with degenerate eigenvalues are discussed in Appendices A.3 and A.4. To illustrate the mathematical result, in Section 3 two example networks with different single cell dynamics are constructed. Each of the constraints are broken to show that they are necessary for scale-invariance. In Section 4, we compare the eigenvectors and eigenvalues of a chaining model and a random recurrent network with those of the examples in Section 3. It is shown that neither of these networks satisfy the structural constraints derived from Section 2, therefore neither of them support sequential neural activity that is scale-invariant.

2 Derivation of the constraints for scale-invariance

In this section we derive the constraints on the connectivity matrix of a linear recurrent network for it to support scale-invariant activity. In Section 2.1, we start with a formal definition of scale-invariance of network activity. In Section 2.2 and Section 2.3 we will derive the two constraints on the connectivity matrix to achieve scale-invariance. Lastly in Section 2.4, we point out that in addition to the two constraints, a particular initial condition is required for the subsequent network dynamics to be scale-invariant. These constraints are sufficient and necessary for a network to generate scale-invariant activity, but only if the network connectivity matrix has real, distinct eigenvalues. We mention a straightforward modification to the results when the network has complex but distinct eigenvalues. The details are shown in Appendix A.1. We discuss the modifications to the constraints for matrices with degenerate eigenvalues in Appendices A.3 and A.4. We also show in Appendix A.5 that deviations from the constraints derived below will cause a graceful degradation in the scale-invariant property of the resulting network activity.

2.1 Formulation of the problem

We consider the autonomous dynamics of a linear recurrent network with N neurons:

$$\dot{\mathbf{x}}(t) = \mathbf{M}\mathbf{x}(t), \tag{1}$$

where $\mathbf{x}$ is an N-dimensional vector summarizing the activity of all the neurons in the network and $\mathbf{M}$ is the $N \times N$ connectivity matrix of the network. We consider the case where there is no input into the network since sequential neural activity is thought to be maintained by internal neuronal dynamics (Pastalkova et al., 2008).

Scale-invariance of the network activity means that the responses of any two neurons in the sequence are rescaled version of each other in time (Figure 1b). Mathematically, this requirement can be written in the following form:

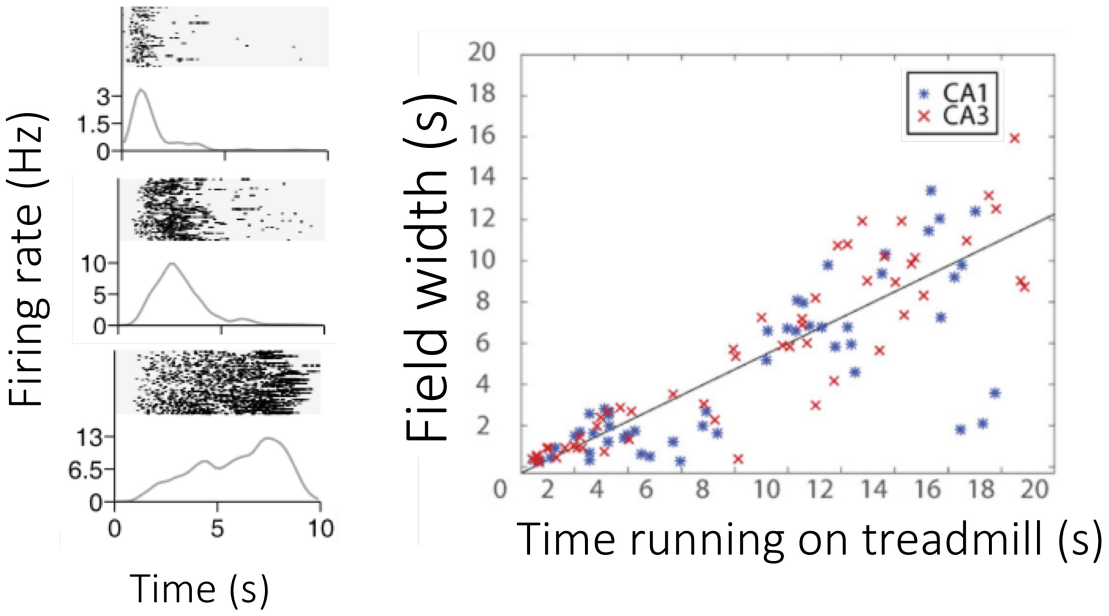
$$x_i(t) = x_j(\alpha_{ij}t), \qquad \forall i, j \in 1, 2, ..., N. \tag{2}$$

That is, for every pair of neurons i and j, their responses $x_i(t)$ and $x_j(t)$ are rescaled in time by a factor α_{ij}. As will be discussed below, this condition can only be satisfied when the network connectivity has real, distinct eigenvalues. Otherwise, only a subset of responses can be rescaled versions of each other. Therefore, in the following sections we are going to focus on connectivity matrices with real, distinct eigenvalues. We will derive two conditions on the connectivity matrix $\mathbf{M}$ necessary for Equation 2 to hold and for the network to generate scale-invariant sequential activity. When the connectivity matrix has complex, distinct eigenvalues, we will state the constraints that allow it to have two distinct scale-invariant sequences. The modifications to the constraints when the eigenvalues are degenerate are discussed in Appendices A.3 and Appendices A.4.

2.2 Constraint 1: Geometrically spaced network timescales

We start by solving Equation 1 using the standard eigendecomposition technique. We diagonalize the connectivity matrix $\mathbf{M}$ as $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ where $\mathbf{\Lambda}$ is a diagonal

a.



b.

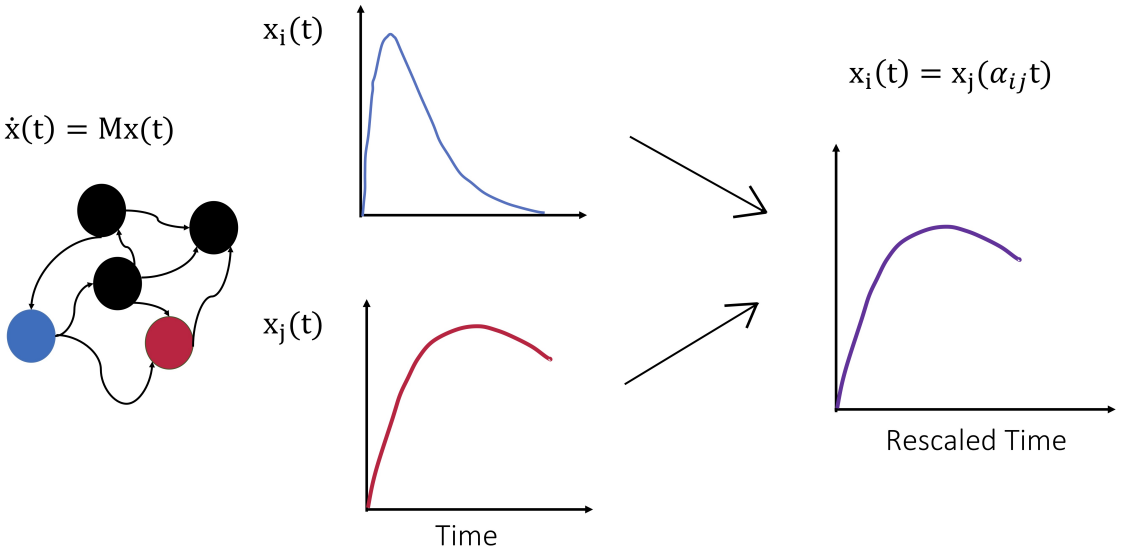


Figure 1: **Scale-invariance for neural sequences and the setup of the problem. a** Left: the raster plots (top) and trial-averaged firing rates (bottom) for three neurons from MacDonald et al. (2011). The neurons were recorded in the hippocampus of rats during a delay period when they were waiting to sample an odor. The neurons that fire later in the delay period show wider responses than neurons that fire earlier during the delay. Right: a scatter plot showing the relationship between the width of each neuron's response and the peak time at which that neuron fires for all neurons recorded from Salz et al. (2016)). Neurons were recorded in the CA3 region of the rat hippocampus during the delay period of a T-maze alteration task when the animal was running on a treadmill. **b.** In this work, we study the dynamics of a linear recurrent network. We seek constraints on the network connectivity matrix $\mathbf{M}$ (**b**, left) such that the activity of every pair of neurons (here for example neurons i and j, middle) are rescaled version of each other in time (right).

matrix consisting of the eigenvalues and $\mathbf{U}$ is a matrix whose columns are the eigenvectors of $\mathbf{M}$. The solution of Equation 1 is then a linear combination of exponential functions:

$$x_i(t) = \sum_j U_{ij} A_j e^{\lambda_j t} \equiv \sum_j \tilde{U}_{ij} e^{\lambda_j t}, \tag{3}$$

where the A_i's are constants determined by the initial condition and are absorbed into the definition of the matrix $\tilde{\mathbf{U}}$.

Imposing the scale-invariance condition (Equation 2) on Equation 3, we have

$$\sum_k \tilde{U}_{ik} e^{\lambda_k t} = \sum_k \tilde{U}_{jk} e^{\alpha_{ij} \lambda_k t} \qquad \forall i, j \in 1, 2, ..., N. \tag{4}$$

For this equation to hold, the time-dependent parts on both sides of the equation must be identical. This means that all the eigenvalues involved in the left hand side of the equation should be equal to a scaled version of the eigenvalues involved in the right hand side. In other words, for each λ_k there should exist an integer δ such that $\lambda_{k+\delta} = \alpha_{ij} \lambda_k$. The only way to achieve this is to have a geometric progression of eigenvalues (network timescales) (Figure 2a). For example, $\lambda_1 = -1$, $\lambda_2 = -2$, $\lambda_3 = -4$, etc.. Therefore we arrive at the first constraint (as depicted in Figure 2a):

Constraint 1: For a linear recurrent network whose connectivity matrix is diagonalizable and has real, distinct eigenvalues to generate scale-invariant activity, the eigenvalues must form a geometric progression.

Remarks: The above constraint needs to be slightly modified when there are complex eigenvalues, since the eigenvalues must come in complex conjugate pairs for the connectivity matrix $\mathbf{M}$ to be real. It can be shown (for details see Appendix A.1) that contrary to the definition of scale-invariance above (Equation 2), there can at best be two scale-invariant sequences, where the responses of neurons in different sequences are not rescaled versions of each other. For this to happen, all the eigenvalues must be complex[1]. Therefore Constraint 1 in the case of complex eigenvalues is as follows (See Appendix A.1 for details):

Constraint 1 for connectivity matrices with complex eigenvalues: For a linear recurrent network whose connectivity matrix is diagonalizable and has complex eigenvalues to generate two sequences of scale-invariant activity, the eigenvalues must all be complex and form two geometric progressions that are complex conjugate with each other.

In the example networks constructed in this paper, all of the eigenvalues have negative real parts to prevent unbounded growth of network activity.

2.3 Constraint 2: Translation-invariant eigenvectors

A second constraint for Equation 4 to hold is that the rows of $\tilde{\mathbf{U}}$ must satisfy a translation-invariant relationship. $\tilde{U}_{i,k+\delta} = \tilde{U}_{jk}$. This way the modes that different neurons pick out will be rescaled versions of each other. Recall from Equation 3 that $\tilde{U}_{ij} = U_{ij} A_j$. Therefore the condition above is equivalent to the columns of the

[1]When there is a mixture of real and complex eigenvalues, there will be at least 3 sequences, which is further away from the strict definition of scale-invariance in Equation 2. Therefore we do not consider this case in this paper.

matrix $\mathbf{U}$ being translation-invariant up to a constant. For example, the different columns could be $\mathbf{v_1} = [1, -1, 0, 0, 0]$, $\mathbf{v_2} = [0, 1, -1, 0, 0]$, $\mathbf{v_3} = [0, 0, 1, -1, 0]$, etc.. Notice that the columns of $\mathbf{U}$ are just the eigenvectors of $\mathbf{M}$. Therefore, we reach the second constraint (see Figure 2b for a graphical illustration):

Constraint 2: For a linear recurrent network whose connectivity matrix has real, distinct eigenvalues to generate scale-invariant activity, the eigenvectors must consist of the same motif (up to a scaling factor) at translated entries. In other words, the eigenvectors must be translation-invariant.

Remarks: When the connectivity matrix has complex eigenvectors, they must come in complex conjugate pairs to ensure that the matrix is real. In this case, it can be shown (for details see Appendix A.1) that there can at best be two scale-invariant sequences, contrary to the definition of scale-invariance above (Equation 2). The responses of neurons in different sequences are not rescaled versions of each other. For this to happen, all the eigenvalues must be complex, and Constraint 2 in the case of complex eigenvalues is as follows (see Appendix A.1 for details):

Constraint 2 for connectivity matrices with complex eigenvalues: For a linear recurrent network whose connectivity is diagonalizable and whose eigenvalues are complex to generate two sequences of scale-invariant activity, each eigenvector should either be a translated version of another eigenvector or its complex conjugate.

A motif with length L in a vector of length N can at most be translated $N - L$ times. Therefore, there will be L eigenvectors that are not translated version of the rest of the eigenvectors. In this work we are interested in large networks where $N \gg L$. Therefore this number is negligible compared to the total number of eigenvectors.

2.4 A note on initial conditions

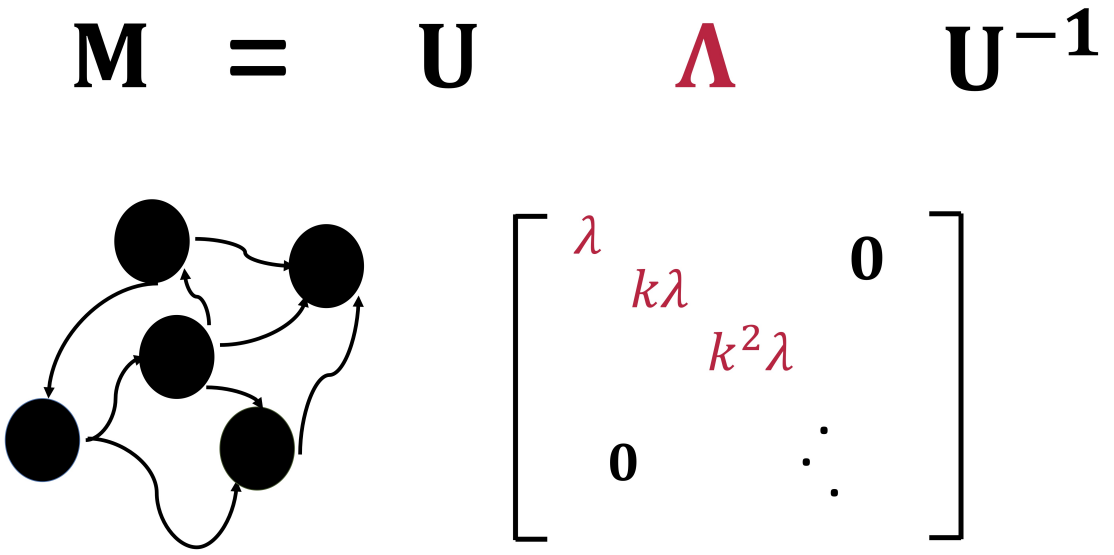
Besides the constraints on the connectivity matrix, the initial condition of the network also affects the scale-invariance of the network activity. Equation 3 requires each neuron in the scale-invariant sequence to have a specific initial condition, i.e. $x_i(0) = \sum_{j=1}^{N} \tilde{U}_{ij}$, to ensure that the dynamics that ensues are scale-invariant. This holds regardless of the eigenspectrum of the connectivity matrix.

For example, if the network has real, distinct eigenvalues, and the sum of the elements in the motif is 0 (which will be the case for the examples in Section 3), the constraint on the initial condition becomes:

$$x_i(t=0) = \begin{cases} 0, & \text{if } i \leq N - L + 1 \\ \sum_{j=1}^{N} \tilde{U}_{ij}, & \text{otherwise,} \end{cases} \tag{5}$$

where L is the length of the repeating motif. In the case where the number of neurons in the network is much larger than the length of motif ($N \gg L$), this constraint on initial condition states that most of the neurons in the network need to be inactive at $t = 0$, whereas a few active neurons act as "input nodes" and propagate their activity to the rest of the network.

a.



b.

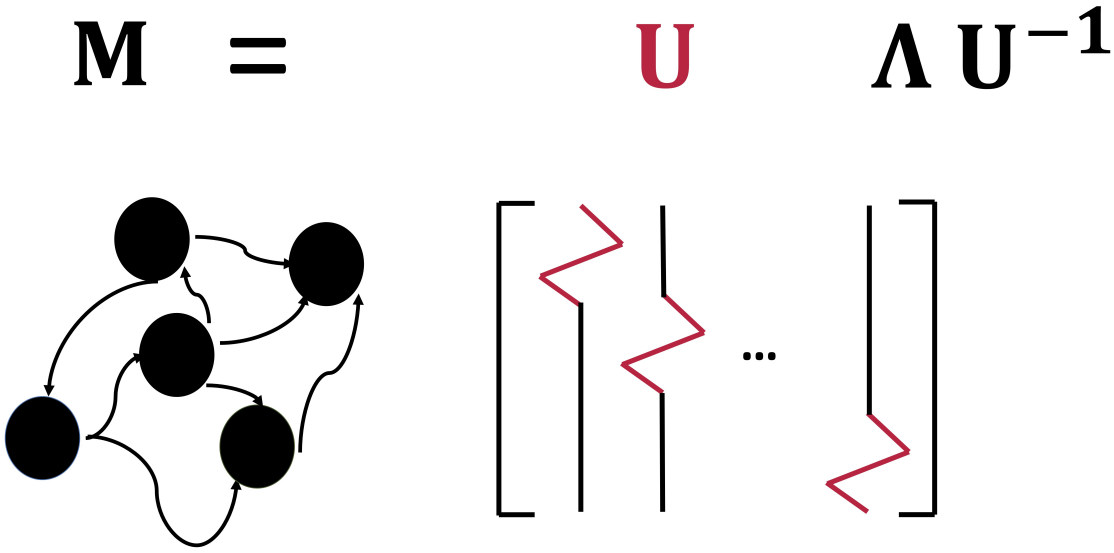


Figure 2: **Graphical illustrations of the constraints for scale-invariance in linear recurrent networks whose connectivity matrices are diagonalizable and have real, distinct eigenvalues.** To generate scale-invariant activity in linear recurrent networks whose connectivity matrices are diagonalizable and have real and distinct eigenvalues (other cases are discussed in Appendix A.1- A.4), the network connectivity must have geometrically spaced eigenvalues (**a**). Furthermore, it must have translation invariant eigenvectors (columns of the matrix **U**) that consist of the same motif (in red) at translated entries (**b**).

3 Examples

In this section we construct two example networks based on the analytical results derived above and show that they allow scale-invariant sequential dynamics. The connectivity matrices of these networks will have geometric progressions of eigenvalues and tranlation-invariant eigenvectors, as shown in Section 2. In the first example (Section 3.1), all eigenvalues are real and the neurons in the network have simple unimodal temporal receptive fields. In the second example (Section 3.2), all eigenvalues are complex, which gives rise to more complicated damped oscillatory single neuron dynamics. In this case, there will be two scale-invariant sequences, as mentioned in Section 2. We will also show that the network activity is no longer scale-invariant when either of the two constraints is violated. In Section 3.3, we will discuss the relationship of our results to a previously proposed network model that generates scale-invariant sequential activity (Shankar and Howard, 2013). In what follows, all simulations were performed in Python 3.6 using Euler's method.

3.1 Simple unimodal temporal receptive fields

In this example we constructed a network that generates sequentially-activated cells with a scale-invariant property. The network consists of $N = 10$ cells. Each cell will have a simple unimodal temporal receptive field, similar to what was observed in electrophysiological recordings of "time cells" (Pastalkova et al., 2008; Jin et al., 2009; MacDonald et al., 2011; Kraus et al., 2013; Mello et al., 2015; Salz et al., 2016; Tiganj et al., 2018).

We constructed the connectivity matrix from its eigendecomposition $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$. According to Constraint 1 (Section 2.2), the network must have geometrically spaced eigenvalues. We hence let $\mathbf{\Lambda}$ be a diagonal matrix whose diagonal elements are geometrically spaced between -0.1 and -5.12.

$$\mathbf{\Lambda} = \begin{bmatrix} -0.1 & 0 & 0 & \dots \\ 0 & -0.22 & 0 & \dots \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \dots & -5.12 \end{bmatrix} \tag{6}$$

According to Constraint 2, the eigenvectors of the connectivity matrix $\mathbf{M}$ must consist of the same motif. Equivalently, we constructed the matrix $\mathbf{U}$ such that its *rows* consist of the same motif. In this example we constructed the motif to be $(1, -1)$. Therefore the matrix $\mathbf{U}$ was given by

$$\mathbf{U} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ U_{N,1} & U_{N,2} & U_{N,3} & \dots & U_{N,N} \end{bmatrix}. \tag{7}$$

As mentioned in Section 2.3, since the motif in the eigenvectors can at most be translated $N - 2$ times, the Nth neuron will not be part of the scale-invariant sequence. Therefore, we chose the last row in $\mathbf{U}$ to be arbitrary numbers such

that $\mathbf{U}$ is invertible. In this example they were sampled from $\mathcal{N}(0,1)$. Finally the connectivity matrix of the network $\mathbf{M}$ was computed from $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$.

The simulated activity of the network is shown in Figure 3a (top left). The initial condition was specified to satisfy Equation 5. The bottom left of Figure 3a shows the network activity rescales along the time axis according to the peak times of each neuron. It is evident that the activations of the neurons are rescaled version of each other, confirming that the constraints derived above indeed lead to scale-invariant sequential activity.

We also broke each of the two constraints above and showed that the resulting activity became no longer scale-invariant (Figure 3a, right two panels). To break Constraint 1 (geometrically spaced eigenvalues), linearly spaced eigenvalues in the same range were used in constructing the matrix $\mathbf{\Lambda}$ instead of geometrically spaced eigenvalues. To break Constraint 2 (translation-invariant eigenvectors), a random vector was added to each row of the matrix $\mathbf{U}$ where each entry was sampled from $\mathcal{N}(0,1)$, making each motif different. As a result, both manipulations generated activity that was no longer scale-invariant (Figure 3a, bottom right two panels).

3.2 Complicated single-cell tuning curves

Temporal coding needs not result in a simple unimodal sequential code. Instead, it could also be embedded in the collective activity of neuronal populations where single neurons may exhibit highly complex dynamics (e.g. Machens et al., 2010). In this subsection we show that the framework we presented above is sufficiently rich to allow for more complex single cell dynamics. In the previous example, the connectivity matrix has real eigenvalues. Therefore it can only give rise to single cell dynamics that are linear combinations of exponential functions. On the other hand, in this subsection we will show that more complicated temporal dynamics can be generated if the connectivity matrix has complex eigenvalues.

Following a similar procedure as detailed in Section 3.1, the connectivity matrix with $N = 20$ neurons was set up so that all of its eigenvalues were complex, and their real and imaginary parts both formed geometric progressions with the same real common ratio (see Appendix A.2 for details). As mentioned in Section 2.2, 2.3 and in more detail in Appendix A.1, the eigenvalues and eigenvectors form complex conjugate pairs, and the network activity contains two distinct scale-invariant sequences. The simulated and rescaled activity are shown in Figure 3b (left). The neural activity exhibits more complicated dynamics but at the same time maintains scale-invariance for each sequence.

We also broke each of the two constraints using similar protocols as in the previous example. To break Constraint 1, eigenvalues with linearly spaced real and imaginary parts in the same range were used instead of geometrically spaced ones. To break Constraint 2, multiplicative noise sampled from $\mathcal{N}(0,1)$ was used, so that each matrix element U_{ij} becomes $U_{ij}(1+\epsilon_{ij})$ where each ϵ_{ij} was sampled from $\mathcal{N}(0,1)$. As shown in Figure 3b (right two panels), the resulting neural activity is no longer scale-invariant.

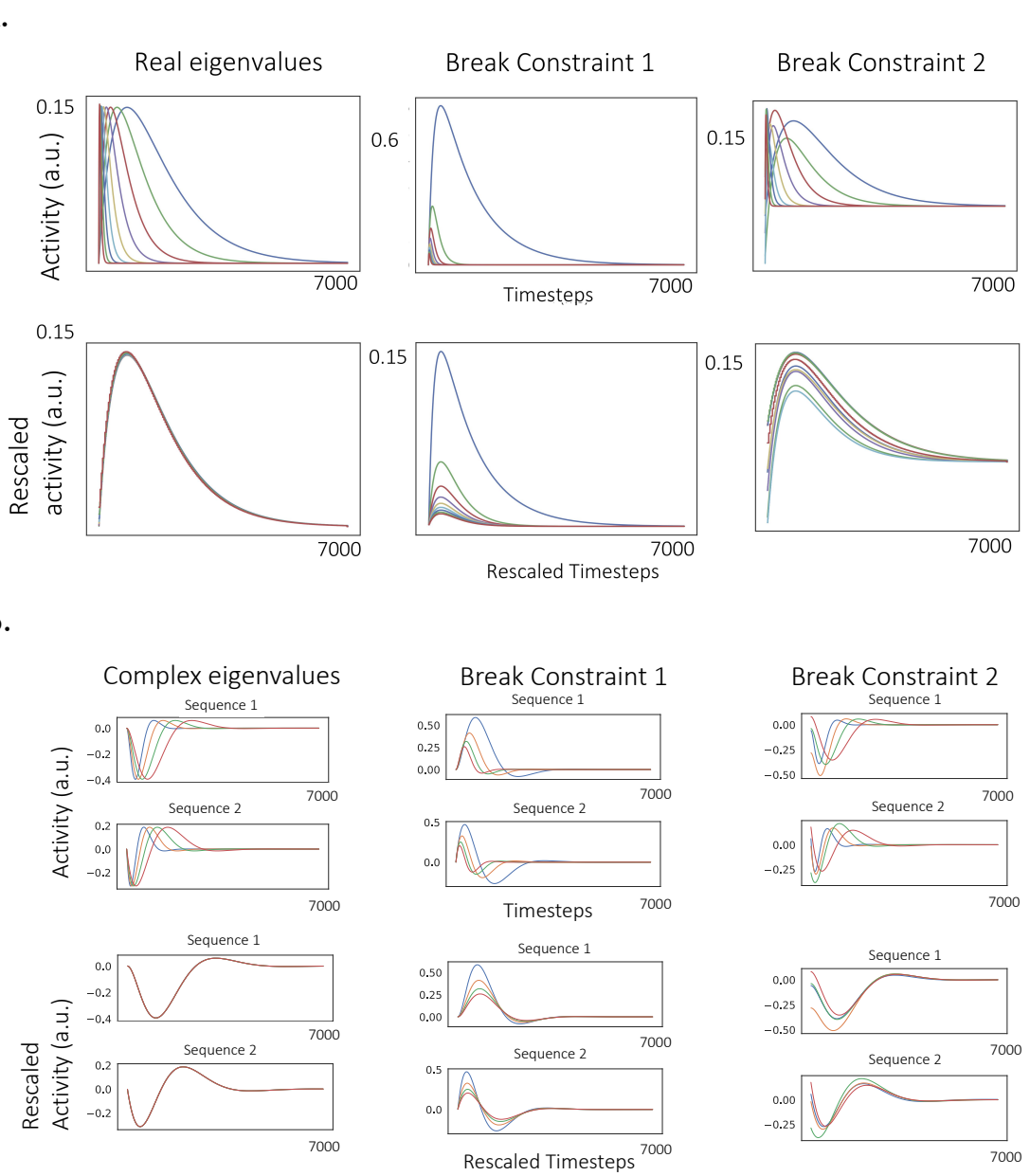


Figure 3: **Generating scale-invariant neural sequences with simple (a) and complicated (b) single neuron dynamics**. Using the analytical result, we constructed networks with specific connectivity matrices so that they generate scale-invariant sequential activity (left column). We also broke each of the two constraints derived in Section 2 and showed that the resulting network activity breaks scale-invariance (middle and right columns) **a.** The network with real eigenvalues gives rise to simple unimodal single cell temporal receptive fields (**a**, left top, each line represents the activity of one neuron in the network). The activity of different neurons overlap with each other when rescaled according to their peak times (left bottom). The network activity becomes not scale-invariant when Constraint 1 was broken by choosing linearly spaced eigenvalues or Constraint 2 was broken by adding noise to the motifs (**a**, middle and right columns, see text for details). **b.** The network with complex eigenvalues gives rise to two sequences of neuronal responses with more complex temporal dynamics (**b** left top, see text and Appendix A.2 for details). The responses of cells in each sequence are rescaled version of each other in time (**b** left bottom). When each of the two constraints is broken, the network activity becomes not scale-invariant (right two columns).

3.3 A special case: Laplace and inverse Laplace transforms

It should be noted that the geometric progression of time constants required by Constraint 1 does not necessarily have to be an emergent property of the network, but can instead be driven by physiological properties of single cells (Loewenstein and Sompolinsky, 2003; Fransén et al., 2006; Tiganj et al., 2015; Liu et al., 2019). Consequently, scale-invariant sequential activity could also be generated by feedforward networks where the neurons in the first layer receive inputs and decay exponentially with a spectrum of geometrically-spaced intrinsic time constants, and the neurons in the second layer are driven by the first layer via translation-invariant synaptic weights, implementing the eigendecomposition in Equation 3 explicitly.

One such model has been proposed by Shankar and Howard (Shankar and Howard, 2013). It is a two-layer feedforward neural network. In that model, the first layer neurons $\mathbf{F}$ encode the Laplace transform of the input and have exponentially decaying firing rates with a spectrum of decay constants.

$$\frac{dF_i(t)}{dt} = -\lambda_i F_i(t), \qquad i = 1, 2, ..., N. \tag{8}$$

The activity of the first layer constitutes a scale-invariant sequential activity (see Equation 2). It is also the basis functions that make up any general scale-invariant sequential activity (see Equation 3).

To generate sequential neural activity, the neurons in the second layer $\tilde{\mathbf{f}}$ compute the inverse Laplace transform of the first layer under the Post approximation (Post, 1930). [2]

$$\tilde{\mathbf{f}} = \mathbf{L}_k \mathbf{F}, \tag{9}$$

where $\mathbf{F}$ is the activity vector of all the first layer neurons and $\tilde{\mathbf{f}}$ is that of the second layer neurons. The matrix $\mathbf{L}_k$ is a discretized approximation of the inverse Laplace transform of the kth order (Shankar and Howard, 2013).

From Equation 8 and Equation 9, the feedfoward dynamics above is equivalent to linear recurrent dynamics involving only the second layer neurons $\tilde{\mathbf{f}}$:

$$\frac{d}{dt}\tilde{\mathbf{f}} = \mathbf{L}_k \mathbf{S} \mathbf{L}_k^{-1} \tilde{\mathbf{f}} \equiv \mathbf{M}_{\text{TILT}} \tilde{\mathbf{f}}, \tag{10}$$

where $\mathbf{S}$ is a diagonal matrix consisting of the single cell time constants λ_i's in Equation 8.

Therefore the dynamics of the neurons in the second layer are equivalent to the one generated by a linear recurrent network with connectivity matrix $\mathbf{M}_{\text{TILT}}$. Because the matrix representation of the inverse Laplace transform $\mathbf{L}_k$ is approximated by taking derivatives of nearby nodes, it has the same motif across columns (for details see Shankar and Howard, 2013). Therefore, the model in Shankar and Howard (2012), although a feedforward network, is a special case in the family of linear recurrent networks that can generate scale-invariant sequential activity. It might be the case that that the exponentially decaying basis functions are indeed maintained by a separate population of neurons, and the downstream neurons consititute a "dual"

[2] Note that this transformation is written as $\mathbf{L}_k^{-1}$ in other papers.

population. Such exponentially decaying cells have recently been identified in lateral entorhinal cortex (Tsao et al., 2018; Bright et al., 2019), whose downstream regions have been identified as locations for time cells (Pastalkova et al., 2008; MacDonald et al., 2011; Kraus et al., 2013; Salz et al., 2016).

4 Comparison with common network models

Scale-invariance puts stringent constraints on the architecture of recurrent neural networks. To illustrate this, in this section we consider two widely-used neural network models that do not generate scale-invariant sequential activity. Section 4.1 considers a simple chaining model; the following section considers a random network. We will see that the connectivity matrices for these two widely-used models violate the constraints derived above in Section 2 and that these models do not support scale-invariant sequential activity. We will compare these two networks with the two scale-invariant example networks described previously. All the networks in this section are simulated with $N = 20$ neurons.

4.1 Simple Feedforward Chaining Model

The simplest possible model to generate sequential activity is a simple feedforward chaining model, which will be analyzed in this subsection. The connectivity matrix of the studied model is non-diagonalizable, therefore the constraints derived above in Section 2.2 and 2.3 do not apply exactly. However, as discussed in Appendix A.4, the *distinct* eigenvalues should still form a geometric progression even when the connectivity is non-diagonalizable to allow scale-invariant activity. We will demonstrate that a chaining model composed of elements with the same time constant cannot meet this requirement for scale-invariant sequential activity.

In this simple model the activity of the kth neuron in a chain of N neurons obeys

$$\frac{dx_k}{dt} = \begin{cases} -x_k + x_{k-1}, & 1 < k \leq N \\ -x_k, & k = 1 \end{cases} \tag{11}$$

Note that all of these neurons have the same time constant which is here set to 1. Because an activation in the first unit spreads gradually across the network, this model generates sequential activity. Each neuron does not respond instantaneously to its input but has some finite integration time resulting in a spread of activity across time. As the activation spreads across the chain, the spread in time accumulates. However, as can be shown via the Central Limit theorem, this sequential activity is not scale-invariant because the peak time of the kth unit goes up linearly in k but the width of the peak goes up with $\sqrt{k}$ (Liu et al., 2019). Here we show that the eigenspectrum of the connectivity matrix of this simple model (Eq. 11) does not satisfy the constraints derived from Section 2.

Figure 4a (left) shows the connectivity matrix of the chaining model described in Eq. 11. This connectivity matrix has a simple motif that repeats across rows; the self-interaction term gives a -1 on the diagonal and the chaining term gives a $+1$ on the off-diagonal. This connectivity matrix is non-diagonalizable and has a single

degenerate eigenvalue of -1 (Figure 4a, middle left). According to Appendix A.4, the distinct eigenvalues of a non-diagonalizable connectivity matrix should still follow a geometric progression for the network to allow scale-invariant activity. The connectivity matrix of the chaining model violates this condition. Therefore, although the network generates sequential activity (Figure 4a, middle right), the activity of different neurons are not rescaled versions of each other (Figure 4a, right).

For contrast, Figure 4b illustrates the connectivity matrix, eigenvalues and eigenvectors for the scale-invariant network with real eigenvalues described previously in Section 3.1. In illustrating the connectivity matrix, we have ordered the neurons according to the order in which they are activated and we have only included the neurons in the sequence (the same for the complex example below). Recall that this matrix was constructed to obey the two constraints and has already been shown to generate scale-invariant sequential activity. Consequently, by construction, the eigenvalues are geometrically spaced and the eigenvectors are translated versions of one another (Figure 4b, middle left and middle panels). Although the connectivity matrix clearly has a rich structure, the rows of the connectivity matrix are certainly not translated versions of one another. The entries above the diagonal tend to be more negative, indicating that the connections from neurons later in the sequence to the neurons earlier in the sequence tend to be more inhibitory than the connections in the opposite direction (Figure 4b, leftmost panel). It is not at all obvious why this specific connectivity matrix yields scale-invariant sequential activity. This is much more clear from examining the eigenvalues and eigenvectors.

4.2 Random recurrent networks

Nonlinear neural networks with random connectivity matrices are able to generate chaotic activity (Sompolinsky et al., 1988). With appropriately trained weights, they are also able to generate sequential neural activity similar to that obtained in actual recordings (Rajan et al., 2016). However, generic linear random neural networks without training cannot produce scale-invariant, sequential activity due to the conflict with the two constraints derived above in Section 2.

Random neural networks have eigenvalue spectrums that are uniformly distributed inside a unit disc in the complex plane (Rajan and Abbott, 2006; Girko, 1985), therefore not geometrically spaced as required by scale-invariance (Section 2.2). We simulated the activity of an instance of a linear network with a random connectivity matrix and computed its eigenvalues and eigenvectors (Figure 4c). The network dynamics is described by

$$\frac{d\mathbf{x}(t)}{dt} = -\mathbf{x}(t) + \mathbf{M}\mathbf{x}(t) = (\mathbf{M} - \mathbf{I})\mathbf{x}(t), \tag{12}$$

where each element in $\mathbf{M}$ was sampled from a Gaussian distribution with mean 0 and variance $\frac{1}{N}$.

The connectivity matrix $\mathbf{M} - \mathbf{I}$ is shown in Figure 4c (left). The eigenvalues of the connectivity matrix are approximately uniformly distributed in a unit disc centered at (-1,0) (Figure 4c, middle left), confirming the results from Rajan and Abbott (2006) and Girko (1985). The eigenvectors also apparently do not have any

translation-invariant structure (Figure 4c, middle). Consequently, the network activity does not have the scale-invariant property (Figure 4c, right two panels). In contrast, the network constructed in the same way as in Section 3.2 has two geometric progressions of complex eigenvalues (Figure 4d, middle left) that are complex conjugates of each other, and each half of the eigenvectors consist of the same motif (Figure 4d, middle). Consequently, it allows two sequences of scale-invariant activity (Figure 4d, right two panels).

5 Discussion

In this paper we study the constraints on the connectivity of a linear network for it to generate scale-invariant sequential neural activity. It is analytically shown that two conditions need to hold for the structure of the connectivity matrix when it has real, distinct eigenvalues. First of all, it must have geometrically-spaced eigenvalues. Second, its eigenvectors must contain the same motif with a translation-invariant structure. Intuitively, the same motif in different eigenvectors pick out different timescales by the same proportion, and the geometric spacing of timescales further ensures the activity of different cells are rescaled with each other. The generated activity is highly dynamic, providing a possibility for the dynamic coding of working memory (Stokes, 2015). It was also shown that a straightforward generalization leads to the constraints for networks with complex eigenvalues (for details see Appendix A.1). Analogous but more complicated constraints hold for networks with degenerate eigenvalues (Appendices A.3 and A.4).

5.1 Plausibility for a geometric progression of timescales

Geometrically spaced network eigenvalues can generically emerge from multiplicative cellular processes (Amir et al., 2012). Mechanistically, a spectrum of timescales could be generated by positive feedback in recurrent circuits. It could also be an inherent property of single neurons (Loewenstein and Sompolinsky, 2003; Fransén et al., 2006; Tiganj et al., 2015; Liu et al., 2019). *In vitro* slice experiments and modeling works have shown that a spectrum of slow timescales can be obtained in single cells by utilizing the slow dynamics of calcium-dependent currents (Loewenstein and Sompolinsky, 2003; Egorov et al., 2002; Fransén et al., 2006; Mongillo et al., 2008; Tiganj et al., 2015; Liu et al., 2019). *In vivo* experiments also showed a spectrum of timescales in cortical dynamics. Bernacchia et al. (2011) showed that in monkey prefrontal, cingulate and parietal cortex, reward modulates neural activity multiplicatively with a spectrum of time constants (Bernacchia et al., 2011).

5.2 Weber-Fechner Law

The constraints for diagonalizable matrices with real unique eigenvalues suggest a deep connection to the Weber-Fechner law. The Weber-Fechner law relates internal psychophysical scales to external physical variables. A quantity is said to obey the Weber-Fechner law if a psychological quantity p is related to a physical quantity S as $p = k \log S$ for an appropriate choice of units. The Weber-Fechner law is perhaps

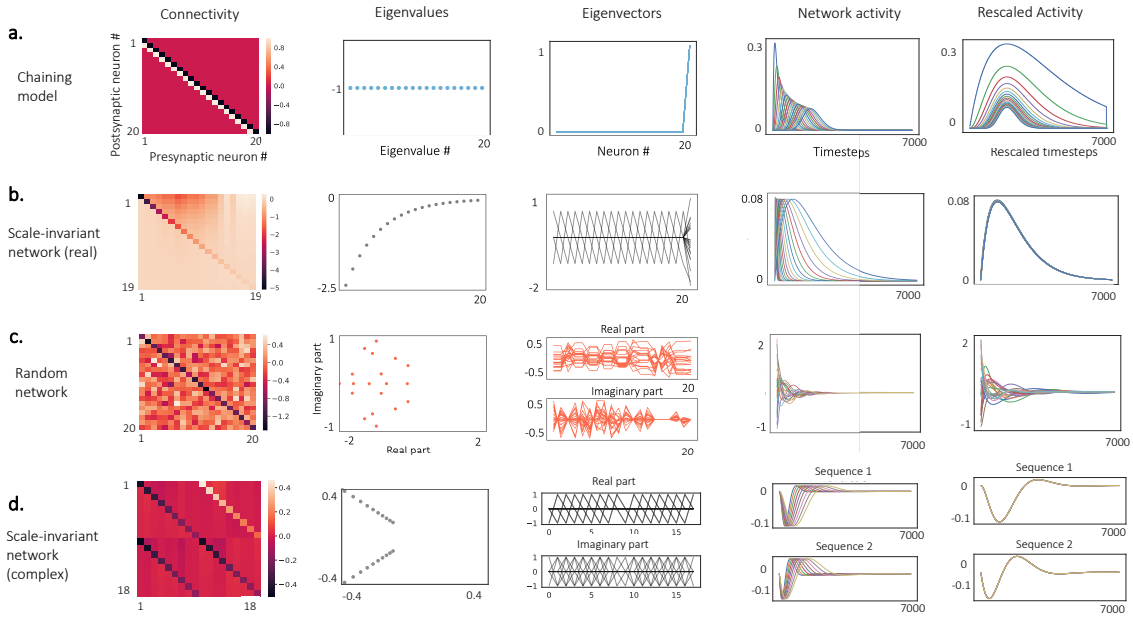


Figure 4: **Comparison with common network models. a, b.** Comparison between the simple feedforward model (**a**) and a network constructed from Section 3.1 that generates scale-invariant sequential activity (scale-invariant network [real], **b**). For the simple feedforward chaining model, its connectivity matrix is non-diagonalizable. Because the eigenvalues of its connectivity matrix are not geometrically spaced (**a**, middle left), the network activity does not have the scale-invariant property (**a**, right two panels). In contrast, the scale-invariant network has eigenvalues that are geometrically spaced (**b**, middle left) and eigenvectors that consist of the same motif (**b**, middle, aside from the boundary effect), resulting in the connectivity that consists of excitation from neurons earlier in the sequence to the ones later in the sequence and inhibition in the opposite direction (**b**, left, only neurons in the sequence are shown. Same below). Consequently, it allows scale-invariant sequential activity (**b** right two panels). **c, d.** Same comparison between an instance of a random recurrent network (**c**) and a network constructed from Section 3.2 (scale-invariant network [complex], **d**). The eigenvalues of the connectivity of a random network are not geometrically spaced (**c**, middle left). The eigenvectors do not consist of the same motif (**c**, middle). Consequently its activity is not scale-invariant (**c**, right two panels). In contrast, the connectivity of the scale-invariant network has two geometric progression of eigenvalues that are complex conjugate with each other (**d**, middle left). Its eigenvectors also contain complex conjugate pairs, each with a repeating motif (**d**, middle). Consequently the scale-invariant network can generate two scale-invariant sequences (**d**, right two panels).

the oldest quantitative relationship in psychology (Fechner, 1912) and holds at least approximately for a wide range of physical variables.

Consider a set of receptors responding to some physical variable S with a receptive field center s_i. If the receptive field centers follow a geometric progression, a logarithmic mapping between receptor number and S naturally results. Logarithmic spacing of receptive fields seems to be a common property of receptive fields in sensory systems (Merzenich et al., 1973; Van Essen et al., 1984; Schwartz, 1977). However, analogous logarithmic spacing is also observed for receptive fields over non-sensory variables such as numerosity (Nieder and Miller, 2003; Nieder and Dehaene, 2009). The receptive fields of time cells in multiple brain regions are certainly compressed (Mello et al., 2015; Kraus et al., 2013; Tiganj et al., 2018), but it has not been quantitatively established that this compression is logarithmic (Howard, 2018).

The constraints for diagonalizable matrices with unique real eigenvalues developed here are closely analogous to the requirements for a Weber-Fechner scale. The geometric progression of eigenvalues places the eigenvalues on a logarithmic scale such that the nth eigenvalue goes up exponentially with n. The requirement that the eigenvectors are translated versions of the same motif is analogous to an translation-invariant tiling along some dimension. Thus, scale-invariance in linear recurrent networks can be interpreted as a translation-invariant tiling of receptors along a logarithmic scale.

5.3 The eigenvectors of scale-invariant networks are localized

The second constraint we derived requires that connectivity matrices with real, distinct eigenvalues should have eigenvectors that consist of the same motif localized at different entries. It is hard to imagine how this kind of translation-invariant eigenvectors could arise generically in neural circuits. However the eigenvectors that satisfy this constraint are a special case of "localized eigenvectors", which have been studied extensively first in condensed matter physics and later in theoretical neuroscience. Anderson first argued that the eigenvectors of matrices whose elements are random and concentrated on the diagonal are exponentially localized (Anderson, P., 1958). Later numerical and analytical studies confirmed that localized eigenvectors indeed arise in neural networks in the presence of a gradient in the strength of the local interactions (Chaudhuri et al., 2014) or global inhibition, as in the case of ring attractor networks (Tanaka and Nelson, 2019). However, it should be pointed out that the procedures described in this paper do not necessarily generate local interactions where the elements M_{ij} decays with $|i-j|$, as can be seen in Figure 4b. Furthermore, our constraint on the eigenvectors is more stringent than only requiring them to be localized: the different localized "patches" need to be the same as well.

5.4 Scale-invariant sequence gives logarithmic growth in cumulative dimensionality

In a recent study, Cueva et al. (2019) showed that the cumulative dimensionality of the population activity during working memory increases with a decreasing speed (Cueva et al., 2019). This is consistent with the network activity generated by linear networks that satisfy the two constraints we derived. Translation-invariant eigenvectors ensure that each eigenmode contributes one unique dimension to the activity. Geometrically spaced eigenvalues ensure that one eigenmode would be suppressed per unit time on a logarithmic scale. Therefore the cumulative dimensionality of scale-invariant sequential activity would increase with the logarithm of time. Furthermore, notice that any affine transformation on the neural trajectory would not change the cumulative dimensionality. Therefore the same relationship for cumulative dimensionality would hold even if the eigenvectors are not translation-invariant. A geometric progression of eigenvalues is sufficient to generate linear dynamics whose cumulative dimensionality increases with the logarithm of time.

5.5 Chaining models

Goldman (2009) proposed a class of linear network models that are able to generate sequential activity. In these networks, the feedforward dynamics are constructed by building up feedforward interactions between orthogonal Schur modes and are hidden in the collective network dynamics (Goldman, 2009). Thanks to the hidden feedforward dynamics, the networks can sustain its activity far longer than the timescale constrained by the eigenvalue spectrum of the network. The example studied in Section 4.1 is the simplest model proposed in Goldman (2009). Although this simple model does not allow scale-invariant sequential activity, this does not argue that any network constructed in the way in Goldman (2009) cannot generate scale-invariant activity.

5.6 Non-normal networks

Because the eigenvetors in a scale-invariant network are translated versions of each other, they do not form an orthonormal basis. Therefore, the networks that generate scale-invariant sequential activity naturally belong to the family of non-normal networks (White et al., 2004; Ganguli et al., 2008; Goldman, 2009). Non-normal networks have been shown to have many desired properties such as extensive memory capacity where the number of timesteps over which the network state retains information about the past stimuli scales linearly with the size of the network (White et al., 2004; Ganguli et al., 2008). The current work shows that scale-invariance could be another potential computational benefit for non-normal networks.

5.7 Locally-interacting integral transform networks

Shankar studied a model where the activity of individual nodes can be effectively described by a filtered input through a set of scale-invariant kernel functions with different timescales (Shankar, 2015). It was shown that if this transformation were to

be implemented by a network involving only local (possibly non-linear) interactions between nodes with similar timescales, the forms of the kernel functions are strongly constrained: they can only be given by a linear combination of the inverse Laplace transforms (c.f. Section 3.3). In this work we consider a slightly different problem: the interactions between different nodes are constrained to be linear, but non-local interactions are also allowed since the connectivity matrix is not constrained to be local. In this case, the activity given by the inverse Laplace transform (Equation 10) covers a subspace of all possible solutions. It is still notable that these two related approaches converge on similar results.

Acknowledgements

This work was supported by ONR MURI award N00014-16-1-2832, ONR DURIP award N00014-17-1-2304 and NIBIB R01EB022864.

A Appendix

The constraints derived in Section 2 are the sufficient and necessary conditions for generating scale-invariant activity in a linear recurrent network, but only when the network connectivity matrix has real, distinct eigenvalues. In this series of appendices we discuss the modifications to the constraints when the connectivity matrix does not meet this requirement. We start out by discussing matrices with complex, distinct eigenvalues (Appendix A.1). The results from this appendix have already been stated in the main text. We then use these results to describe the setup of the example network in Section 3.2 in Appendix A.2. The case when the network connectivity has degenerate eigenvalues and is diagonalizable is discussed in Appendix A.3. The case when it is non-diagonalizable is discussed in Appendix A.4. Finally in Appendix A.5 we discuss the sensitivity of the scale-invariant property to perturbations in the initial condition and the network connectivity.

A.1 Connectivity matrices with complex eigenvalues

We consider the case when all the eigenvalues are complex. It will be become obvious below that a mixture of real and complex eigenvalues will make the network activity further deviate from being strictly scale-invariant (c.f. Equation 2), therefore we do not consider this case in this paper.

Since network connectivity matrices are real, their complex eigenvalues and eigenvectors come in conjugate pairs. Consequently, the two constraints in Section 2 should be stated in a slightly different way. The activity of cell i is still

$$x_i(t) = \sum_{j=1}^{N} U_{ij} x'_j(t), \tag{A.1}$$

where the modes are

$$x'_j(t) = a_j e^{\lambda_j t}, \tag{A.2}$$

except that the eigenvalues λ_i are now complex. Therefore the derivation in Section 2 carries over. However, since the eigenvectors form complex conjugate pairs, the matrix $\mathbf{U}$ in the diagonalization procedure $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ should be of the following form:

$$\mathbf{U} = \begin{bmatrix} a & b & 0 & \dots & a^* & b^* & 0 & \dots & 0 \\ 0 & a & b & \dots & 0 & a^* & b^* & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ U_{\frac{N}{2},1} & U_{\frac{N}{2},2} & U_{\frac{N}{2},3} & \dots & U_{\frac{N}{2},\frac{N}{2}} & U_{\frac{N}{2},\frac{N}{2}+1} & U_{\frac{N}{2},\frac{N}{2}+2} & & U_{\frac{N}{2},N} \\ U_{\frac{N}{2}+1,1} & U_{\frac{N}{2}+1,2} & U_{\frac{N}{2}+1,3} & \dots & U_{\frac{N}{2}+1,\frac{N}{2}} & U_{\frac{N}{2}+1,\frac{N}{2}+1} & U_{\frac{N}{2}+1,\frac{N}{2}+2} & \dots & U_{\frac{N}{2}+1,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ U_{N,1} & U_{N,2} & U_{N,3} & \dots & U_{N,\frac{N}{2}} & U_{N,\frac{N}{2}+1} & U_{N,\frac{N}{2}+2} & & U_{N,N} \end{bmatrix}, \tag{A.3}$$

where the motif is (a, b). Notice that the lower half of the matrix is not determined. This is due to the constraint that there has to be a complex conjugate for every eigenvector, and therefore the motif can only be translated $\frac{N}{2} - L$ times, rather than $N - L$ times as in the case of real eigenvalues (c.f. Section 2.3). The bottom half of the matrix cannot simply repeat the upper half as this will make $\mathbf{U}$ not invertible. Therefore, it is not possible to satisfy the scale-invariant condition in the main text that any pair of responses are rescaled version of each other (c.f. Equation 2). Instead, the network can only have half of its cells be scale-invariant (cells 1 to $\frac{N}{2} - L + 1$).

But we can do a little better if we allow the bottom half of the matrix to repeat the structure of the upper half, only with a different motif:

$$\mathbf{U} = \begin{bmatrix} a & b & 0 & \dots & a^* & b^* & 0 & \dots & 0 \\ 0 & a & b & \dots & 0 & a^* & b^* & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ U_{\frac{N}{2},1} & U_{\frac{N}{2},2} & U_{\frac{N}{2},3} & \dots & U_{\frac{N}{2},\frac{N}{2}} & U_{\frac{N}{2},\frac{N}{2}+1} & U_{\frac{N}{2},\frac{N}{2}+2} & & U_{\frac{N}{2},N} \\ c & d & 0 & \dots & c^* & d^* & 0 & \dots & 0 \\ 0 & c & d & \dots & 0 & c^* & d^* & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ U_{N,1} & U_{N,2} & U_{N,3} & \dots & U_{N,\frac{N}{2}} & U_{N,\frac{N}{2}+1} & U_{N,\frac{N}{2}+2} & & U_{N,N} \end{bmatrix}, \tag{A.4}$$

This way, the network will generate two different scale-invariant sequences. Cells within each sequence has responses that are rescaled version of each other, but cells in different sequences are not rescaled version of each other. This is the way the example network in Section 3.2 was constructed.

To summarize, linear recurrent networks with connectivity matrices whose eigenvalues are all complex cannot generate scale-invariant activity in the strict sense as defined by Equation 2. The closest scenario is to have two different scale-invariant sequences, each composed of half of the cells in the network. In this case, the constraint on the eigenvalues is that they form two geometric progressions that are complex conjugate with each other, with the same real common ratio for both

the real and imaginary parts. The constraint on the eigenvectors is that each of the eigenvector should either be a translated version of another eigenvector or its complex conjugate. We applied these results to construct the example network in Section 3.2, as detailed in the next appendix.

Finally, it is not hard to see that if there is a mixture of real and complex eigenvalues, the network can at best generate three distinct scale-invariant sequences. Since this scenario is even further from the strict definition of scale-invariance (c.f. Equation 2), we do not consider this case in this paper.

A.2 The setup of the example network with complex eigenvalues

For the example in Section 3.2, the real and imaginary parts of the eigenvalues were both chosen to be geometrically spaced between -0.15 and -0.48 (-0.15, -0.17, ..., -0.48), and the complex conjugates were also included. Therefore, the matrix $\mathbf{\Lambda}$ in the diagonalization procedure $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ is

$$\mathbf{\Lambda} = \begin{bmatrix} -0.15(1+i) & & & & & & & & \\ & -0.17(1+i) & & & & & & & \\ & & \ddots & & & & & & \\ & & & -0.48(1+i) & & & & & \\ & & & & -0.15(1-i) & & & & \\ & & & & & -0.17(1-i) & & & \\ & & & & & & \ddots & & \\ & & & & & & & -0.48(1-i) \end{bmatrix}. \tag{A.5}$$

To satisfy the constraint that the eigenvectors form complex conjugate pairs, we constructed the motif so that the matrix $\mathbf{U}$ is in the form of Equation A.4 with two repeating motifs, where $a = 1+i$, $b = -1-i$, $c = 1-i$, $d = -1+i$. In other words the matrix $\mathbf{U}$ is

$$\mathbf{U} = \begin{bmatrix} 1+i & -1-i & 0 & \dots & 1-i & -1+i & 0 & \dots & 0 \\ 0 & 1+i & -1-i & \dots & 0 & 1-i & -1+i & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ U_{\frac{N}{2},1} & U_{\frac{N}{2},2} & U_{\frac{N}{2},3} & \dots & U_{\frac{N}{2},\frac{N}{2}} & U_{\frac{N}{2},\frac{N}{2}+1} & U_{\frac{N}{2},\frac{N}{2}+2} & & U_{\frac{N}{2},N} \\ 1-i & -1+i & 0 & \dots & 1+i & -1-i & 0 & \dots & 0 \\ 0 & 1-i & -1+i & \dots & 0 & 1+i & -1-i & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ U_{N,1} & U_{N,2} & U_{N,3} & \dots & U_{N,\frac{N}{2}} & U_{N,\frac{N}{2}+1} & U_{N,\frac{N}{2}+2} & & U_{N,N} \end{bmatrix}, \tag{A.6}$$

where the unspecified elements are chosen randomly from a uniform distribution from [0,1].

A.3 Diagonalizable connectivity matrices with degenerate eigenvalues

A matrix with degenerate eigenvalues could be diagonalizable or non-diagonalizable (defective). The modifications to the constraints are different for these two cases. In this appendix we discuss these modifications when the network connectivity is diagonalizable and in the next appendix we focus on non-diagonalizable matrices. In both cases, degeneracy in the eigenvalues decreases the number of distinct responses in the scale-invariant sequence because this number is equal to the number of *distinct* eigenvalues of the connectivity. In both this and the next appendices, we assume that the eigenvalues of the connectivity matrix are all real. It is straightforward to generalize the results to the case of complex eigenvalues based on the discussions in Appendix A.1.

For diagonalizale matrices with degenerate eigenvalues, the activity of cell i is still

$$x_i(t) = \sum_{j=1}^{N} U_{ij} x'_j(t), \tag{A.7}$$

where the modes are

$$x'_j(t) = a_j e^{\lambda_j t}, \tag{A.8}$$

except that some of the eigenvalues λ_j could be the same. Therefore, for scale-invariance to hold, the modes that contribute to the summation in Equation A.7 for any pair of cells should have eigenvalues that are a multiple of each other (e.g. for cell 1, the eigenvalues that contribute to the summation in Equation A.7 are {-1, -1, -2, -4}, then for cell 2 they could be {-2, -2, -4, -8}, etc.). A necessary condition for this to hold is that the *distinct* eigenvalues of the connectivity matrix form a geometric progression. Therefore, the constraint on the eigenvalues is as follows:

> For a linear recurrent network whose connectivity matrix is diagonalizable with degenerate, real eigenvalues to generate scale-invariant activity, the *distinct* eigenvalues must form a geometric progression (c.f. Section 2.2).

The constraint on the eigenvectors is more complicated due to the degeneracy of eigenvalues. We here offer a compact statement for a sufficient condition on the eigenvectors: During the diagonalization of the matrix $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$, if we rearrange the eigenvalues in the diagonal form of the connectivity $\mathbf{\Lambda}$ such that they form blocks of distinct eigenvalues (e.g. $\lambda_1 = -1, \lambda_2 = -2, \lambda_3 = -4, \lambda_4 = -1, \lambda_5 = -2$, etc.), then at the block level, the matrix $\mathbf{U}$ would satisfy the original constraint that its columns (i.e. the eigenvectors) consist of the same motif (c.f. Section 2.2) at translated entries. In general, the eigenvectors would at least to be localized.

A.4 Non-diagonalizable connectivity matrices

A non-diagonalizable (or defective) matrix can be transformed into a block-diagonal matrix called its Jordan normal form by a similarity transformation. We will show that when the connectivity matrix is non-diagonalizable, its *distinct* eigenvalues still

need to form a geometric progression. However, as for diagonalizable matrices with degenerate eigenvalues (Appendix A.3), the constraint on the eigenvectors is more complicated. In what follows, we assume all eigenvalues are real. It is straighforward to generalize the results to the case of complex eigenvalues based on the discussions in Appendix A.1.

To reach this conclusion, we first review some basic results about the Jordan normal form. A non-diagonalizable connectivity matrix $\mathbf{M}$ can be transformed into a block-diagonal matrix via a similarity transformation: $\mathbf{M} = \mathbf{U}^{-1}\mathbf{J}\mathbf{U}$. The columns of $\mathbf{U}$ are called the generalized eigenvectors of $\mathbf{M}$ and $\mathbf{J}$ is called the Jordan normal form of $\mathbf{M}$. $\mathbf{J}$ is a block-diagonal matrix where the diagonal elements of each block is an eigenvalue λ of the matrix $\mathbf{M}$. It also has off-diagonal elements 1 to the right of each diagonal element, i.e. $J_{i,i+1} = 1$, except for the last row of each block. Consequently, it is not hard to see that the activity of each cell is a linear combination of modes

$$x_i(t) = \sum_{\lambda k} U_{i\{\lambda k\}} a_{\lambda k} x'_{\lambda k}(t), \tag{A.9}$$

where we use $x'_{\lambda k}(t)$ to represent the mode corresponding to the kth last row of the block with eigenvalue λ. Here, the modes will not all be exponential functions, but a product of a polynomial and an exponential

$$x'_{\lambda k}(t) = \frac{1}{(k-1)!} t^{k-1} e^{\lambda t}. \tag{A.10}$$

Based on Equations A.9 and A.10 above, it is quite straightforward to come up with the constraints on scale-invariance in the case of non-diagonalizable matrices. For any pair of cells, the modes that contribute to their activity should have series of eigenvalues λ that are a multiple of each other as well as the same *sequence* of k. For example, if there are two modes contributing to the activity of cell 1, with $(k_1, \lambda_1) = (1, -1)$ and $(k_2, \lambda_2) = (2, -2)$, then the two contributing modes for cell 2 could have $(k_1, \lambda_1) = (1, -2)$ and $(k_2, \lambda_2) = (2, -4)$, and similar relationships hold for any pair of cells. As a special case, for diagonalizable connectivity matrices, the contributing modes for any cell have $k = 1$ for all λ. The necessary condition for the contributing eigenvalues for different cells to be a multiple of each other is that the distinct eigenvalues of the connectivity form a geometric progression. Therefore, the constraint on the eigenvalues is as follows:

> For a linear recurrent network whose connectivity matrix is non-diagonalizable and has real eigenvalues to generate scale-invariant activity, the *distinct* eigenvalues of the connectivity matrix should form a geometric progression (c.f. Section 2.2).

However, the constraint on the eigenvectors is harder to state in a compact manner because of the degeneracy of the eigenvalues. The best we can say is that each eigenvector would have to be localized, i.e. the range of the non-zero elements would have to be narrow.

A.5 Robustness to noise

In this appendix we show that deviations from the constraints derived in Section 2 causes a graceful degradation in scale-invariance. We performed numerical experiments where we perturbed the connectivity matrix of the network in Section 3.1 using Gaussian multiplicative noise: $M_{ij} \longrightarrow M_{ij}\,(1 + \epsilon\mathcal{N}[0,1])$ and the initial condition with Gaussian additive noise: $x_i(t=0) \longrightarrow x_i(t=0) + \epsilon\mathcal{N}(0,1)$. We chose an additive noise for the initial condition because a multiplicative noise would only rescale the initial state for the example in Section 3.1. The degree of scale-invariance was measured in terms of the ratio of the peak times between successive cells in the sequence. If the sequence is scale-invariant, this ratio should equal to the common ratio of the geometric progression of eigenvalues for all pairs of successive cells.

Figure A.1a and d show the distributions of the ratio obtained by 100 realizations of the perturbation on the connectivity and the initial condition, respectively. As can be seen, the distributions become more spread out around the noiseless values as the noise amplitudes ϵ are increased. Figure A.1b and e show that the deviation from scale-invariance as measured by the average difference from the noiseless ratio increases linearly with the noise amplitude ϵ. Figure A.1c and f show the raw and rescaled activity for two realizations of the perturbation with different noise levels. These results show that the degree of scale-invariance decreases linearly with the level of noise in both the initial condition and the connectivity matrix. Generally, a small perturbation to a matrix changes its eigenvalues and eigenvectors by an amount linear in the perturbation, and a small deviation in the initial condition of a linear dynamical system causes the subsequent trajectory to diverge from the original trajectory by an amount linear in the deviation. Therefore it is expected that the deviation from scale-invariance increases linearly with noise strength.

References

Amir, A., Oreg, Y., and Imry, Y. (2012). On relaxations and aging of various glasses. *Proceedings of the National Academy of Sciences*, 109(6):1850–1855.

Anderson, P., W. (1958). Absence of Diffusion in Certain Random Lattices. *Physical Review*, 109(5):1492–1506.

Bernacchia, A., Seo, H., Lee, D., and Wang, X. J. (2011). A reservoir of time constants for memory traces in cortical neurons. *Nature Neuroscience*, 14(3):366–72.

Bright, I. M., Meister, M. L. R., Cruzado, N. A., Tiganj, Z., Howard, M. W., and Buffalo, E. A. (2019). A temporal record of the past with a spectrum of time constants in the monkey entorhinal cortex. *bioRxiv*.

Buonomano, D. V. and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, 10(2):113–25.

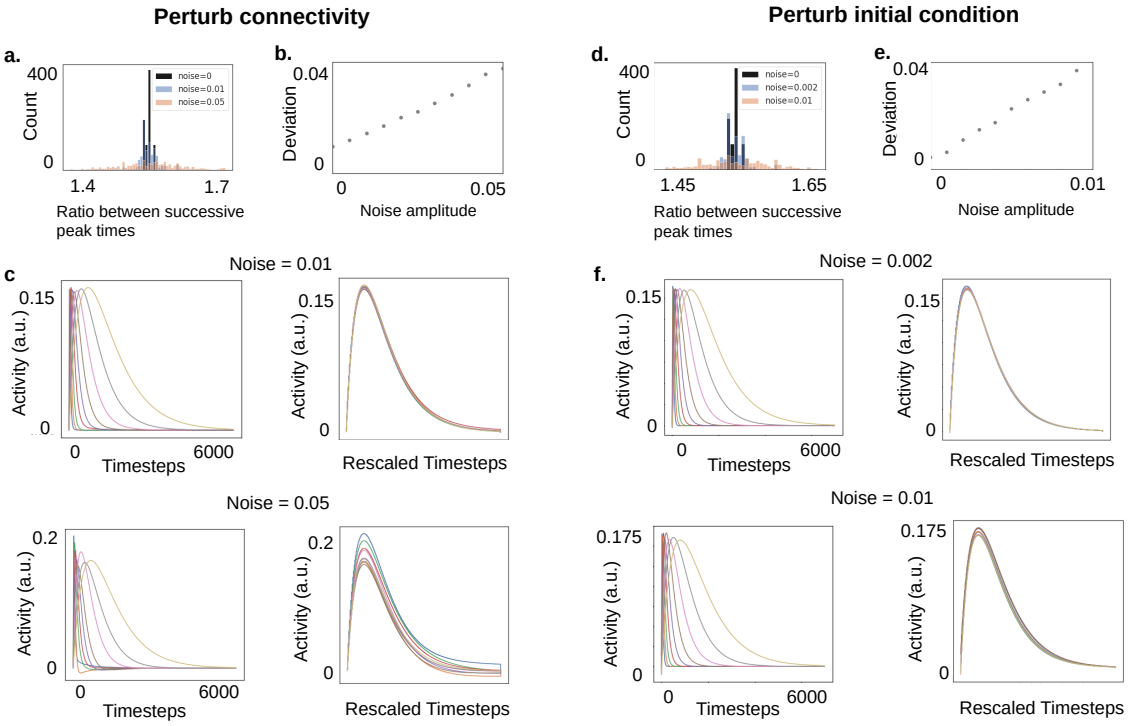


Figure A.1: **Degree of scale-invariance decreases gracefully with noise.** We perturbed the connectivity and the initial condition with Gaussian multiplicative and additive noise respectively for 100 independent runs (see texts for details) and measured the deviation from scale-invariance by the spread of the distribution of the ratio between peak times for successive cells in the sequence. For the perturbation on the connectivity matrix, the distribution of the ratio is more spread out for larger noise (**a**). The mean difference from the noiseless ratio increases linearly with the noise amplitude (**b**). **c.** The raw (left) and rescaled (right) activity for two example runs with different noise levels are shown. **d-f.** The same results for the perturbation on the initial condition. Since in both cases the deviation from scale-invariance is linear in noise strength, scale-invariance does not depend on fine-tuning the connectivity matrix or the initial condition.

Chance, F. S., Abbott, L. F., and Reyes, A. D. (2002). Gain modulation from background synaptic input. *Neuron*, 35(4):773–82.

Chaudhuri, R., Bernacchia, A., and Wang, X. J. (2014). A diversity of localized timescales in network activity. *eLife*, 3:e01239.

Cueva, C. J., Marcos, E., Saez, A., Genovesio, A., Jazayeri, M., Romo, R., Salzman, C. D., Shadlen, M. N., and Fusi, S. (2019). Low dimensional dynamics for working memory and time encoding. *bioRxiv*.

Egorov, A. V., Hamam, B. N., Fransén, E., Hasselmo, M. E., and Alonso, A. A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature*, 420(6912):173–8.

Fechner, G. (1860/1912). *Elements of Psychophysics. Vol. I.* Houghton Mifflin.

Fransén, E., Tahvildari, B., Egorov, A. V., Hasselmo, M. E., and Alonso, A. A. (2006). Mechanism of graded persistent cellular activity of entorhinal cortex layer V neurons. *Neuron*, 49(5):735–46.

Ganguli, S., Huh, D., and Sompolinsky, H. (2008). Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences of the United States of America*, 105(48):18970–18975.

Girko, V. L. (1985). Circular law. *Theory of Probability & Its Applications*, 29(4):694–706.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., and Gretz, A. L. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, 6:355–369.

Goldman, M. S. (2009). Memory without feedback in a neural network. *Neuron*, 61(4):621–634.

Gütig, R. and Sompolinsky, H. (2009). Time-warp-invariant neuronal processing. *PLoS Biology*, 7(7).

Harvey, C. D., Coen, P., and Tank, D. W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484(7392):62–8.

Hopfield, J. J. and Brody, C. D. (2000). What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proceedings of the National Academy of Sciences of the United States of America*, 98(3):1–6.

Howard, M. W. (2018). Memory as perception of the past: Compressed time inmind and brain. *Trends in Cognitive Sciences*, 22(2):124 – 136.

Howard, M. W. and Shankar, K. H. (2018). Neural scaling laws for an uncertain world. *Psychological review*, 125(1):47.

Howard, M. W., Youker, T. E., and Venkatadass, V. (2008). The persistence of memory: Contiguity effects across several minutes. *Psychonomic Bulletin & Review*, 15(PMC2493616):58–63.

Jin, D. Z., Fujii, N., and Graybiel, A. M. (2009). Neural representation of time in cortico-basal ganglia circuits. *Proceedings of the National Academy of Sciences*, 106(45):19156–19161.

Kraus, B. J., Robinson, 2nd, R. J., White, J. A., Eichenbaum, H., and Hasselmo, M. E. (2013). Hippocampal "time cells": time versus path integration. *Neuron*, 78(6):1090–101.

Liu, Y., Tiganj, Z., Hasselmo, M. E., and Howard, M. W. (2019). A neural microcircuit model for a scalable scale-invariant representation of time. *Hippocampus*, 29(3):260–274.

Loewenstein, Y. and Sompolinsky, H. (2003). Temporal integration by calcium dynamics in a model neuron. *Nature Neuroscience*, 6(9):961–7.

MacDonald, C. J., Lepage, K. Q., Eden, U. T., and Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, 71(4):737 – 749.

Machens, C. K., Romo, R., and Brody, C. D. (2010). Functional, But Not Anatomical, Separation of "What" and "When" in Prefrontal Cortex. *Journal of Neuroscience*, 30(1):350–360.

Mello, G. B., Soares, S., and Paton, J. J. (2015). A scalable population code for time in the striatum. *Current Biology*, 25(9):1113–1122.

Merzenich, M. M., Knight, P. L., and Roth, G. L. (1973). Cochleotopic organization of primary auditory cortex in the cat. *Brain Research*, 63:343–6.

Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 319(5869):1543–1546.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64:482–488.

Nieder, A. and Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, 32:185–208.

Nieder, A. and Miller, E. K. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, 37(1):149–57.

Pastalkova, E., Itskov, V., Amarasingham, A., and Buzsaki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, 321(5894):1322–7.

Post, E. (1930). Generalized differentiation. *Transactions of the American Mathematical Society*, 32:723–781.

Rajan, K. and Abbott, L. F. (2006). Eigenvalue spectra of random matrices for neural networks. *Physical Review Letters*, 97(18):188104.

Rajan, K., Harvey, C. D., and Tank, D. W. (2016). Recurrent Network Models of Sequence Generation and Memory. *Neuron*, 90(1):128–142.

Rakitin, B. C., Gibbon, J., Penny, T. B., Malapani, C., Hinton, S. C., and Meck, W. H. (1998). Scalar expectancy theory and peak-interval timing in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 24:15–33.

Salz, D. M., Tiganj, Z., Khasnabish, S., Kohley, A., Sheehan, D., Howard, M. W., and Eichenbaum, H. (2016). Time cells in hippocampal area CA3. *Journal of Neuroscience*, 36:7476–7484.

Saxe, A. M., McClelland, J. L., and Ganguli, S. (2013). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks.

Schwartz, E. L. (1977). Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25(4):181–94.

Shankar, K. H. (2015). Generic construction of scale-invariantly coarse grained memory. *Lecture Notes in Artificial intelligence*, 8955:175–184.

Shankar, K. H. and Howard, M. W. (2012). A scale-invariant internal representation of time. *Neural Computation*, 24(1):134–193.

Shankar, K. H. and Howard, M. W. (2013). Optimally fuzzy temporal memory. *Journal of Machine Learning Research*, 14:3753–3780.

Sompolinsky, H., Crisanti, A., and Sommers, H.-J. (1988). Chaos in random neural networks. *Physical review letters*, 61(3):259.

Stokes, M. G. (2015). 'activity-silent'working memory in prefrontal cortex: a dynamic coding framework. *Trends in cognitive sciences*, 19(7):394–405.

Tanaka, H. and Nelson, D. R. (2019). Non-hermitian quasilocalization and ring attractor neural networks. *Physical Review E*, 99(6):062406.

Tank, D. and Hopfield, J. (1987). Neural computation by concentrating information in time. *Proceedings of the National Academy of Sciences*, 84(7):1896–1900.

Tiganj, Z., Cromer, J. A., Roy, J. E., Miller, E. K., and Howard, M. W. (2018). Compressed timeline of recent experience in monkey lateral prefrontal cortex. *Journal of cognitive neuroscience*, pages 1–16.

Tiganj, Z., Hasselmo, M. E., and Howard, M. W. (2015). A simple biophysically plausible model for long time constants in single neurons. *Hippocampus*, 25(1):27–37.

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M.-B., and Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, 561(7721):57.

Van Essen, D. C., Newsome, W. T., and Maunsell, J. H. (1984). The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Research*, 24(5):429–48.

Voelker, A. R. and Eliasmith, C. (2018). Improving spiking dynamical networks: Accurate delays, higher-order synapses, and time cells. *Neural computation*, 30(3):569–609.

Voss, R. F. and Clarke, J. (1975). 1/f noise in music and speech. *Nature*, 258:317–318.

Wang, J., Narain, D., Hosseini, E. A., and Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. *Nature neuroscience*, 21(1):102.

White, O. L., Lee, D. D., and Sompolinsky, H. (2004). Short-term memory in orthogonal neural networks. *Physical Review Letters*, 92(14):148102.